



**HAL**  
open science

# Intégration de données génétiques à grande échelle pour l'amélioration de la prise en charge du transplanté rénal

Rokhaya Ba

► **To cite this version:**

Rokhaya Ba. Intégration de données génétiques à grande échelle pour l'amélioration de la prise en charge du transplanté rénal. Médecine humaine et pathologie. Nantes Université, 2022. Français. NNT : 2022NANU1033 . tel-03955658

**HAL Id: tel-03955658**

**<https://theses.hal.science/tel-03955658v1>**

Submitted on 25 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

NANTES UNIVERSITE

ECOLE DOCTORALE N° 605

*Biologie Santé*

Spécialité : Bioinformatique

Par

**Rokhaya BA**

**Intégration de données génétiques à grande échelle pour l'amélioration de la prise en charge du transplanté rénal**

Thèse présentée et soutenue à Nantes, le 06 Décembre 2022

Unité de recherche : UMR 1064 Centre de Recherche en Transplantation et Immunologie

## Rapporteurs avant soutenance :

Carmen Lefaucheur  
François Cornélis

Professeur des Universités – Praticien Hospitalier, Université de Paris  
Professeur des Universités – Praticien Hospitalier, Université de Clermont-Ferrand

## Composition du Jury :

Président : Magali Giral  
Examineurs : Magali Giral  
Jean-François Zagury  
Gwendaline Guidicelli

Professeur des Universités – Praticien Hospitalier, Nantes Université  
Professeur des Universités – Praticien Hospitalier, Nantes Université  
Professeur des Universités, CNAM Paris  
Praticien Hospitalier, CHU de Bordeaux

Dir. de thèse : Pierre-Antoine Gourraud  
Co-dir. de thèse : Sophie Limou

Professeur des Universités – Praticien Hospitalier, Nantes Université  
Maître de conférence, Centrale Nantes



*«C'est de la force des convictions que dépend la réussite, pas du nombre de partisans»*

Remus Lupin





# REMERCIEMENTS

---

J'aimerais remercier toutes les personnes qui de près ou de loin ont contribué à la réussite de ce projet de thèse.

Mes remerciements vont tout d'abord à mes directeurs de thèse, le Dr Sophie Limou et le Pr Pierre-Antoine Gourraud pour m'avoir offert cette opportunité et m'avoir amené jusqu'au bout de ce projet de thèse. Merci de m'avoir ouvert les portes de la recherche scientifique d'abord par le biais de mon stage et ensuite à travers ces années de thèse.

Je remercie chaleureusement mes rapporteurs, le Pr François Cornélis et le Pr Carmen Le-faucher, pour leur intérêt pour ce travail, leur disponibilité, leur lecture rigoureuse et leur contribution scientifique très enrichissante. Par la même occasion je remercie mes examinateurs les Pr Magali Giral et Jean-François Zagury et le Dr Gwendaline Guidicelli d'avoir accepté de faire partie de mon jury de thèse. Particulièrement, j'aimerais vivement remercier le Pr Giral et le Pr Zagury qui ont été en première loge de cette thèse à travers mes CSI. Merci pour vos conseils et votre apport scientifique tout au long de ces années.

Je remercie tous les membres du CR2TI qui de près ou de loin, ont contribué à ce travail. Merci aux cliniciens, techniciens, ingénieurs, chercheurs, gestionnaires, doctorants, stagiaires etc. avec lesquels j'ai pu travailler ou échanger durant ma thèse. Je remercie très chaleureusement tous les membres de l'équipe 5 nouvellement équipe 3 (anciens comme nouveaux) que j'ai pu côtoyer avec un immense plaisir. Nico, merci pour ton encadrement, ton sang froid devant les questions parfois naïves sur le HLA et la génétique des populations et ton sens du second degré à toute épreuve qui a souvent égayé ma journée. Je n'oublie pas Axelle, la maman des chats, Abel, Hélène, Léo, Ayan, Chadia, Claire et Raphaël avec qui j'ai partagé une moitié de bureau pendant 8 mois, Clémence, Morgane, Irène, Martin, Sonia. C'était un réel plaisir de travailler et d'échanger avec vous tous, j'en garderais un très bon souvenir. Particulièrement j'aimerais remercier Vincent, j'ai énormément aimé travailler avec toi et je te remercie sincèrement pour toute ton aide et ta contribution dans le projet KiT-GENIE. Tu hérites d'un bon gros bébé, soit en digne ;). J'espère que ta thèse sera à l'image de ce que tu es, brillant et prometteur.

Bien sûr je n'oublie pas les membres qui sont devenus plus que des collègues de travail mais de très bons amis et qui ont été d'un très grand soutien moral durant cette thèse. Estelle, ma vie n'a plus jamais été la même depuis que tu es partie. Ven (Vency :P), je n'oublierai jamais toutes les soirées bouffes/YouTube (je veux dire guitare) qu'on s'est faites chez moi. Éléa, Anaïs, les

journées au labo n'auraient jamais été les mêmes sans nos longues pauses gourmandes. Olivia, coach de vie, chargée de motivation ! Je me marre encore de nos rendez vous matinaux à l'arrêt de tram et de mes sprints pour ne pas le rater (ce qui était souvent le cas d'ailleurs). Sirine, on devrait se chercher une ligne d'urgence car le téléphone simple tous les jours ne suffit plus. Enfin, Magalie merci pour ta bonne humeur et ta gentillesse en toute occasion.

Je tiens à remercier ces amis qui ne sont peut être pas liés au travail mais qui ont toujours été là. D'abord mon support system au delà de toutes frontières : Cas d'urgence, Sophie mon amour, Pa Demba my very best friend, Loucar, Tabara, Falilou (tu es devenu expert en Ferret à force). Je ne saurais jamais vous remercier assez pour ce que vous êtes. Frézza et Maïra aka la Team B, je ne vous oublie pas. Adri je ne sais pas exactement où te placer mais saches que tu es mon coloc belge préféré et le plus beau des pandas.

Je ne saurais terminer sans remercier ma famille. Merci à mes frères et sœurs pour leur soutien, leur bonne humeur, leurs appels ambigus, leurs quotidiens qui me donnent souvent une bonne distraction, les fous rires etc. Ramata, tu remportes le prix de la meilleure sœur pour m'avoir donné une si belle nièce pendant ma thèse. Merci à mes chers parents qui m'ont accompagné et m'ont porté avec fierté tout le long de ma vie. Merci pour votre soutien et votre amour inconditionnels. Papa, je me suis battue pour toi. Cette thèse, c'est pour toi. Enfin, j'aimerais remercier mon mari pour l'immense soutien qu'il m'a témoigné durant cette thèse et ceci dans tous les domaines. Je ne saurais jamais te remercier assez pour ta patience, tes encouragements, tes efforts quotidiens pour me supporter et surtout ta foi en moi lorsque moi même je n'y croyais plus. Merci pour tout. Je t'aime.

# TABLE DES MATIÈRES

---

Liste des acronymes et abréviations	13
<b>I Introduction</b>	<b>15</b>
<b>1 Introduction à la transplantation rénale</b>	<b>17</b>
1.1 Historique de la transplantation . . . . .	17
1.2 Le rein . . . . .	18
1.2.1 Structure du rein . . . . .	18
1.2.2 La fonction rénale . . . . .	21
1.2.2.1 Fonction de filtration et d'épuration du sang . . . . .	21
1.2.2.2 Fonction d'équilibration et de régulateur . . . . .	21
1.2.2.3 Fonction endocrine du rein . . . . .	21
1.2.3 Mesure de la fonction rénale . . . . .	22
1.2.3.1 Mesure de la créatinine sérique . . . . .	23
1.2.3.2 Clairance urinaire de la créatinine sur les urines des 24 heures . . . . .	23
1.2.3.3 Mesure du DFG par calcul de la clairance de substances exogènes . . . . .	24
1.2.3.4 Autres mesures de la fonction rénale . . . . .	24
1.2.3.5 Estimation du Débit de Filtration Glomérulaire . . . . .	25
1.3 L'insuffisance rénale . . . . .	27
1.3.1 L'insuffisance rénale aiguë (IRA) . . . . .	27
1.4 L'insuffisance rénale chronique (IRC) . . . . .	28
1.4.1 L'insuffisance rénale chronique terminale (IRCT) . . . . .	29
1.4.1.1 La dialyse . . . . .	30
1.4.1.1.1 L'hémodialyse . . . . .	30
1.4.1.1.2 La dialyse péritonéale . . . . .	31
1.5 La transplantation . . . . .	33
1.5.1 Pré-transplantation . . . . .	33
1.5.1.1 Critères d'éligibilité à la greffe . . . . .	33
1.5.1.2 Sélection des donneurs . . . . .	34
1.5.1.2.1 Donneurs décédés . . . . .	34
1.5.1.2.2 Donneurs vivants . . . . .	35

1.5.1.3	Compatibilité Donneur/Receveur . . . . .	36
1.5.1.3.1	Le système ABO . . . . .	36
1.5.1.3.2	Le système HLA (Human Leukocyte Antigen) . . . . .	38
1.5.1.4	Attribution de l'organe . . . . .	42
1.5.2	Déroulement de la greffe . . . . .	43
1.5.3	Post-transplantation . . . . .	44
1.5.3.1	Suivi du patient . . . . .	44
1.5.3.2	Complications de la transplantation . . . . .	45
1.5.3.2.1	Les rejets . . . . .	46
1.5.3.2.2	Récidive de la maladie . . . . .	48
<b>2</b>	<b>La transplantation rénale à l'ère des «omiques»</b>	<b>51</b>
2.1	Les «omiques» en transplantation . . . . .	51
2.1.1	Définition des omiques . . . . .	51
2.1.1.1	La génomique . . . . .	51
2.1.1.2	La transcriptomique . . . . .	52
2.1.1.3	La protéomique . . . . .	53
2.1.1.4	La métabolomique . . . . .	54
2.1.1.5	L'épigénomique . . . . .	54
2.1.2	Etat de l'art des études omiques en transplantation . . . . .	55
2.1.2.1	Résumé de la revue . . . . .	55
2.2	Génomique de la transplantation rénale . . . . .	70
2.2.1	Le génome humain . . . . .	70
2.2.1.1	L'ADN . . . . .	70
2.2.1.2	Gène et génome . . . . .	71
2.2.1.3	Séquençage du génome humain . . . . .	74
2.2.1.3.1	Technologies de séquençage . . . . .	75
2.2.2	La transmission génétique . . . . .	77
2.2.2.1	Théorie de l'hérédité de Mendel . . . . .	78
2.2.2.2	Transmission génétique et maladies . . . . .	79
2.2.2.3	Les polymorphismes génétiques . . . . .	81
2.2.2.3.1	Les polymorphismes nucléotidiques simples (SNP) . . . . .	81
2.2.2.3.2	Les insertions/délétions (indels) . . . . .	82
2.2.2.3.3	Variation du nombre de copies (CNV) . . . . .	82
2.2.2.4	Notion de génétique des populations . . . . .	82
2.2.2.4.1	Équilibre d'Hardy-Weinberg . . . . .	82
2.2.2.4.2	Déséquilibres de liaison . . . . .	83
2.2.3	Les études d'association génétiques . . . . .	84

2.2.3.1	Etude gène candidat . . . . .	85
2.2.3.2	Etudes d'association à l'échelle du génome (Genome Wide Association Studies, GWAS) . . . . .	85
<b>3</b>	<b>Objectifs de la thèse</b>	<b>89</b>
3.1	Axe 1 : Consolidation d'une biobanque de paires donneur/receveur pour la transplantation rénale . . . . .	89
3.2	Axe 2 : Explorations statistiques et analyse de données génétiques à grande échelle pour une meilleure prise en charge du greffon rénal . . . . .	90
<b>II</b>	<b>Protocoles</b>	<b>91</b>
<b>1</b>	<b>Génération des données</b>	<b>93</b>
1.1	Extraction et collecte des données . . . . .	93
1.2	Génotypage . . . . .	94
1.3	Analyse primaire des données brutes . . . . .	95
1.4	Stockage des données . . . . .	96
<b>2</b>	<b>Méthodes statistiques</b>	<b>97</b>
2.1	Contrôle qualité des données . . . . .	97
2.2	Analyses en composantes principales (ACP) . . . . .	98
2.3	Imputation SNP et HLA . . . . .	100
2.3.1	Imputation SNP . . . . .	100
2.3.2	Imputation HLA . . . . .	101
2.4	Analyses GWAS . . . . .	102
<b>III</b>	<b>Résultats</b>	<b>105</b>
<b>1</b>	<b>Résultats pré-GWAS</b>	<b>107</b>
1.1	Description des données et analyses préliminaires . . . . .	107
1.1.1	La cohorte KiT-GENIE . . . . .	107
1.1.1.1	Résumé de l'article . . . . .	107
<b>2</b>	<b>Résultats GWAS</b>	<b>147</b>
2.1	Association entre variant génétique et retour en dialyse chez les receveurs et chez les donneurs KiT-GENIE . . . . .	147
2.1.1	Association entre variants génétiques et retour en dialyse pour les receveurs et les donneurs avant imputation . . . . .	147

2.1.2	Association entre variants génétiques et retour en dialyse pour les receveurs et les donneurs après imputation . . . . .	149
2.1.3	Discussion . . . . .	151
<b>IV</b>	<b>Discussion</b>	<b>153</b>
<b>1</b>	<b>L'intégration des données biomédicales en transplantation</b>	<b>154</b>
1.1	Intérêt d'une ressource de données conséquente en transplantation rénale . . . . .	154
1.1.1	Augmentation de la puissance des études . . . . .	155
1.1.2	Accès à des données diverses . . . . .	155
1.1.2.1	Importance de l'intégration de la double couche de données génomiques et phénotypique . . . . .	155
1.1.2.2	Accès à des données longitudinales . . . . .	156
1.1.2.3	Diversité populationnelle/ethnique des données . . . . .	157
1.2	Les défis de l'intégration des données . . . . .	158
<b>2</b>	<b>Analyse de données génomiques à grande échelle</b>	<b>160</b>
2.1	Importance de la pré-analyse des données . . . . .	160
2.2	Les études GWAS dans la compréhension des mécanismes de rejet/perte chronique du greffon . . . . .	161
2.2.1	Limite des études GWAS aux receveurs et aux donneurs uniquement . . . . .	161
2.2.2	Intérêt de l'analyse double génome . . . . .	162
2.2.3	Prise en compte de l'ancestralité . . . . .	163
2.2.4	Validation des études . . . . .	163
<b>V</b>	<b>Conclusion et perspectives</b>	<b>165</b>
<b>1</b>	<b>Conclusion</b>	<b>166</b>
<b>2</b>	<b>Perspectives</b>	<b>168</b>
<b>VI</b>	<b>Références bibliographiques</b>	<b>171</b>
<b>VII</b>	<b>Liste des communications scientifiques</b>	<b>211</b>
<b>1</b>	<b>Bibliographie scientifique</b>	<b>212</b>
1.1	Publications scientifiques . . . . .	212

1.2	Communications orales . . . . .	212
1.3	Posters . . . . .	212





# LISTE DES ACRONYMES ET ABRÉVIATIONS

---

1KGP	1000 Genomes Project
AA	Afro-American
ABM	Agence de la Biomedecine
ADN	Acide Désoxyribo-Nucléique
AMR	Antibody Mediated Rejection
ARN	Acide Ribo-Nucléique
CMH	Complexe Majeur d’Histocompatibilité
CNV	Copy Number Variation
CR	Call Rate
DDAC	Donneur Décédé par Arrêt Cardiaque
DDME	Donneur Décédé par Mort Encéphalique
DFG	Débit de Filtration Glomérulaire
DIVAT	Données Informatisées et Validées en Transplantation
DQC	dishQC
DSA	Donor Specific Antibody
EFI	European Federation for Immunogenetic
EUR	European
EWAS	Epigenome-Wide Association Study
GWAS	Genome-Wide Association Study
HLA	Human Leucocyte Antigens
HSCT	Hematopoietic Stem-Cell Transplantation
HSF	Hyalinose Segmentaire Focale
IBD	Identity By Descent
IBS	Identity By State
IMC	Indice de Masse Corporel
IRA	Insuffisance Rénale Aiguë
IRC	Insuffisance Rénale Chronique
IRCT	Insuffisance Rénale Chronique Terminale

## TABLE DES MATIÈRES

---

LB	Lymphocyte B
LRS	Long Read Sequencing
LT	Lymphocyte T
MAF	Minor Allelic Frequency
MDRD	Modification of Diet in Renal Disease
MRC	Maladie Rénale Chronique
MS	Mass Spectrometry
NGS	Next Generation Sequencing
NK	Natural Killer
PCR	Polymerase Chain Reaction
QC	Quality Control
RMN	Résonance Magnétique Nucléaire
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
TCMR	T-Cell Mediated Rejection
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

PREMIÈRE PARTIE

# Introduction

---



# INTRODUCTION À LA TRANSPLANTATION RÉNALE

---

## 1.1 Historique de la transplantation

Les premiers succès d'autogreffes humaines remontent au 19<sup>e</sup> siècle avec les avancées de Jacques Louis Reverdin qui effectue, en 1869, des greffes épidermiques en couvrant les surfaces sans peau de petites pièces d'épiderme humain [1]. Au début du XX<sup>e</sup> siècle, Mathieu Jaboulay et Alexis Carrel mettent au point plusieurs techniques de suture vasculaire, qui seront immédiatement appliquées à la transplantation d'organe [2, 3]. En 1908, Alexis Carrel réussit la première auto-transplantation rénale (transfert du rein sur une autre partie du corps) chez le chien avec une survie prolongée de l'animal et du greffon [2]. La réalité chirurgicale de la transplantation est alors reconnue mais les hétéro-transplantations rénales (transplantations entre deux espèces) chez l'homme échouent systématiquement. En 1933, Voronoy pratique la première transplantation rénale chez l'homme en greffant un rein de cadavre sur une jeune femme atteinte d'insuffisance rénale [4, 5] qui se solde par un échec. En 1952, la transplantation d'un rein maternel (donneur vivant) à un patient âgé de 16 ans par le groupe de Jean Hamburger à Necker ne parvient pas non plus à obtenir un fonctionnement prolongé du greffon [6] (décès du patient au 21<sup>e</sup> jour). Au début du siècle, les travaux de James Murphy puis de Williamson avaient déjà identifiés un facteur biologique inconnu à l'époque, responsable des réactions de rejet d'homogreffe, et atténué par les rayons X et le benzol [7, 8]. Mais de nombreuses années de recherche seront encore nécessaires avant la découverte de l'immunité humorale par Woglom, en 1933 [9], puis du système HLA par Jean Dausset, en 1952 [10]. Le premier succès en transplantation rénale à long terme est celui d'une transplantation entre deux jumeaux monozygotes en 1954, à Boston [11]. En 1959, deux transplantations rénales entre jumeaux dizygotes sont réussies, la première à Boston, la seconde à Paris, à l'Hôpital Necker. En 1960, l'équipe de l'hôpital Foch, de Suresnes, va plus loin, en tentant les premières transplantations entre frère et sœur non jumeaux, puis en dehors de toute parenté entre le donneur et le receveur. À chaque fois, l'immunosuppression est induite [1, 12]. En 1956 les greffes de moelle osseuse débutent [13] suivies par les premières transplantations hépatiques (1963-1969) [14], de cœur (1967) [15, 16] et de poumon (1963) [17, 18].

En 1994, l'établissement français des greffes est créé et une convention de collaboration entre France Greffe de Moelle et les Etablissements Français du Sang sera signée en 2000. En 2020, 4 417 transplantations d'organes ont été réalisées dont 400 (9%) à partir de donneurs vivants (rein, foie). Par comparaison avec 2019, il y eu une baisse de 25 % du nombre total de greffes en France en raisons du Covid-19. Cet écart a été en grande partie comblé en 2021 avec une hausse de 19,5% par rapport à 2020. Les transplantations courantes concernent aujourd'hui le rein (3251, 62%), le foie (1 224, 23%), le coeur (408, 8%), le poumon (316, 6%), le pancréas (67, 1%), la double transplantation coeur/poumon (6, 0.1%). Le rein reste à ce jour l'organe le plus transplanté dans l'histoire [19] (Table 1.1).

TABLE 1.1 – *Evolution du nombre de transplantations d'organes en France au cours des 5 dernières années [19].*

En 2021, 5 273 greffes ont été possibles grâce à 1 392 donneurs décédés et 521 donneurs vivants.  
Par rapport à 2020, il y eu une hausse de 19,3% du nombre total de greffes en France.  
Pour les greffes à partir de donneur vivant, le niveau d'activité de 2019 a été égalé.

	2017	2018	2019	2020	2021
Greffes cardiaques	467	450	425	370	408
Greffes cardio- pulmonaires	6	9	9	8	6
Greffes pulmonaires	378	372	384	283	316
Greffes hépatiques (dont à partir de donneurs vivants)	1374 (18)	1 323 (14)	1 356 (19)	1 128 (15)	1 224 (19)
Greffes rénales (dont à partir de donneurs vivants)	3 782 (611)	3 546 (537)	3 643 (509)	2 591 (385)	3 251 (502)
Greffes pancréatiques	96	78	84	34	67
Greffes intestinales	2	3	0	3	1
<b>TOTAL</b> (dont à partir de donneurs vivants)	6105 (626)	5 781 (551)	5 901 (528)	4 417 (400)	5 273 (521)

## 1.2 Le rein

### 1.2.1 Structure du rein

Les reins sont localisés dans l'espace rétropéritonéal, de part et d'autre de la colonne vertébrale sous les dernières côtes. Ils ont la forme typique d'un haricot et mesurent environ 12 cm pour 160 g. Les reins sont reliés d'une part à l'artère aorte et à la veine cave inférieure par l'artère et la veine rénale et d'autre part à la vessie par deux longs canaux, les uretères ; ils font partie de l'appareil urinaire [20] (Figure 1.1). Les reins sont entourés d'une couche de graisse

périrénale, puis d'une enveloppe de tissu conjonctif appelée fascia de Gerota. Ils sont chacun surmontés d'une glande surrénale. Chaque rein est protégé par une enveloppe dure externe ; la capsule. A l'intérieur de celle-ci se trouve le parenchyme rénal qui renferme une partie périphérique appelée le cortex et une partie centrale appelée la médulla [21, 22]. Entre les deux se trouvent environ 400 à 800 000 petites structures composées essentiellement de tubules et de glomérules, les néphrons [23] (Figure 1.1). Le néphron est l'unité fonctionnelle du rein. Il a pour rôle de filtrer les différentes substances contenues dans le sang pour ensuite réabsorber les nutriments et éliminer les déchets par l'urine. La formation de l'urine commence dans les glomérules qui filtrent les substances sanguines et l'eau pour former l'urine primaire. Ensuite, par des phénomènes de sécrétion et de réabsorption entre le fluide tubulaire et les capillaires, la composition de l'urine est modifiée dans les tubules pour aboutir à la formation de l'urine définitive. Cette urine passe ensuite dans les calices, s'écoule dans le bassinet puis transite dans les uretères pour atteindre la vessie où elle est stockée puis éliminée lors d'une miction via l'urètre [24] (Figure 1.1). Chacun d'entre nous possède normalement deux reins. Mais dans certains cas (Anomalie congénitale ou retrait d'un rein), un seul rein suffit pour vivre (5% des individus). Les reins sont des organes vitaux qui assurent les fonctions essentielles d'épurateur et de régulateur de l'organisme [22, 23, 24]



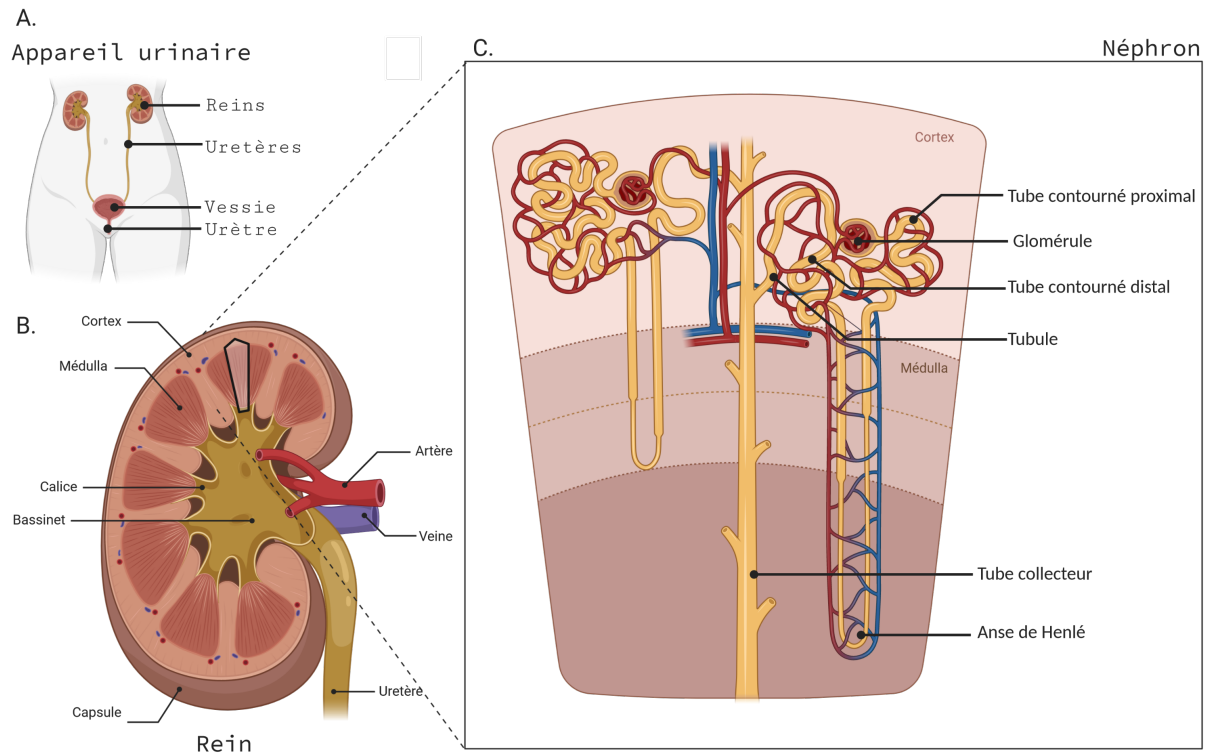


FIGURE 1.1 – *Anatomie de l'appareil urinaire et physiologie du rein.* Le rein appartient à l'appareil urinaire. Il est richement vascularisé et se divise en 2 parties : (1) le cortex, (2) la médullaire. La médullaire est composée de milliers de néphrons. Le néphron est l'unité fonctionnelle du rein. Chaque néphron comprend un glomérule et un tubule (composé de différents segments) [25]

## 1.2.2 La fonction rénale

### 1.2.2.1 Fonction de filtration et d'épuration du sang

Les reins possèdent la capacité de filtrer le sang. Ils sont parcourus par de nombreux vaisseaux sanguins qui assurent la circulation du sang dans l'organe. Le sang arrive dans chaque rein via une artère rénale qui provient d'une ramification de l'aorte. Le débit sanguin rénal représente environ 600 mL/min, soit 20 à 25 % du débit cardiaque et est transmis en quasi-totalité aux glomérules. Le sang est ensuite épuré par le rein qui trie et régule ses composants. Trois classes de substances sont filtrées au niveau du glomérule : les électrolytes (sels minéraux en circulation dans le sang, tels que le sodium, calcium, potassium, etc.), les non-électrolytes (issus du catabolisme des protéines, peptides et acides aminés, tels que le glucose ou l'urée) et l'eau. Lors de cette étape, il n'y a pas de passage des protéines qui sont trop grosses pour passer à travers les pores et sont retenues dans le sang. La réabsorption est réalisée au niveau des tubules à la suite de l'étape de filtration. Ainsi, les reins régulent la conservation ou la perte de certains constituants comme l'eau, l'urée, l'acide urique ou la créatinine. Ils éliminent également les substances étrangères comme les résidus des médicaments, dont l'accumulation serait toxique pour l'organisme. Chaque jour, les glomérules produisent environ 180 litres d'ultrafiltrat (Débit de Filtration Glomérulaire), pour un débit urinaire d'environ 1 à 2 litres/j, la différence étant réabsorbée par le tubule (diurèse). Une fois épuré, le sang repart par la veine rénale qui rejoint la veine cave inférieure et les substances non réabsorbées sont éliminées via l'urine qui sera stockée dans la vessie avant d'être évacuée [23, 24].

### 1.2.2.2 Fonction d'équilibration et de régulateur

Les reins ont une fonction de régulation du milieu intérieur afin de maintenir l'homéostasie. Ils équilibrent de façon constante les sorties rénales de certains constituants principalement l'eau et les sels minéraux (le sodium contenu dans le sel, le potassium, le calcium, le bicarbonate qui règle l'acidité du sang, le magnésium) en fonction des besoins de l'organisme et des entrées digestives. L'ajustement de chaque soluté (condition de l'homéostasie) se fait finement grâce aux phénomènes tubulaires de sécrétion et de réabsorption, sous contrôle hormonal spécifique (aldostérone pour le Na, ADH pour l'eau par exemple). Ceci permet une régulation de l'équilibre acido-basique et de la pression artérielle. Lorsque les reins fonctionnent mal, ils n'éliminent pas suffisamment les électrolytes tels que le sodium ou le potassium ce qui peut induire de l'hypertension artérielle, des œdèmes ou des complications cardiaques [23, 25].

### 1.2.2.3 Fonction endocrine du rein

De nombreuses substances à activité biologique sont synthétisées dans le rein tels que des hormones, des enzymes et des vitamines :

- **L'érythropoïétine (EPO)** L'érythropoïétine est une hormone produite par le rein plus spécifiquement par des cellules interstitielles péri-tubulaires fibroblastiques. Elle permet à l'organisme de s'adapter à différentes situations physiologiques (baisse du taux d'oxygène circulant dans les artères rénales, hémorragie, hémolyse etc.) en régulant la quantité des globules rouges pour véhiculer l'oxygène dans l'organisme. Un manque d'EPO est fréquent lorsqu'il existe une insuffisance rénale, responsable d'une anémie [23, 24].
- **Le système rénine-angiotensine-aldostérone (SRAA)** Le rein sécrète une enzyme appelée la rénine. Elle est produite au niveau de l'appareil juxta-glomérulaire et s'active par protéolyse de l'angiotensinogène circulant d'origine hépatique. Par des mécanismes de fragmentations, l'angiotensinogène est transformée en angiotensineI puis en angiotensineII grâce à la rénine et une enzyme de conversion. On parle de SRAA qui joue un rôle très important dans la régulation de la pression artérielle notamment avec des mécanismes de vasoconstrictions ou de vasodilatations [23, 24].
- **La vitamine D** Le rein participe à la production de vitamine D active appelée calcitriol et sécrétée à partir de son précurseur hépatique (25 (OH) vitamine D3). La forme active de la vitamine D augmente l'absorption digestive et rénale de calcium, et l'absorption intestinale de phosphate. Un manque de vitamine D est fréquent lorsqu'il existe une insuffisance rénale [23, 24].

### 1.2.3 Mesure de la fonction rénale

Les mesures de la fonction rénale permettent d'évaluer la capacité des reins à filtrer les déchets azotés du sang (urée, créatinine, acide urique) [26]. Ces tests contribuent à calibrer la posologie de plusieurs médicaments qui sont excrétés ou métabolisés par les reins, d'indiquer le moment approprié pour le début de la dialyse et d'évaluer le statut rénal des donneurs et des receveurs pour une transplantation de reins. Plus important encore, ces mesures sont primordiales pour la détection précoce de l'insuffisance rénale qui bien souvent évolue silencieusement (pas de symptômes visible chez le patient) jusqu'au stade de non réversibilité. Il est également important de suivre la fonction rénale chez les individus qui sont dits « à risques » (antécédents de diabète mellitus, hypertension artérielle, maladie cardiaque, cholestérol élevé, âge supérieur à 50 ans, antécédents familiaux d'insuffisance rénale etc.) [27]. Toutefois, la fonction rénale considérée comme « normale » diffère en fonction de plusieurs facteurs tels que l'âge, le sexe, le poids, l'ethnie, le régime alimentaire etc. Par exemple, un sujet avec une masse musculaire importante n'aura pas les mêmes mesures de la fonction rénale qu'un individu avec un poids normal. Cela ne signifie pas toutefois qu'il a une mauvaise fonction rénale [26]. L'évaluation de cette fonction rénale doit donc prendre en compte tous ces facteurs pour être la plus précise possible. Pour se faire, on se base sur la mesure du taux de filtration glomérulaire rénale; l'idée étant de mesurer s'il y'a une augmentation du taux de molécules présentes normalement dans le sang

et éliminées par le rein (créatinine, urée, acide urique etc.) pouvant résulter d'une mauvaise filtration rénale [28].

### 1.2.3.1 Mesure de la créatinine sérique

La créatinine, est un produit de dégradation des muscles naturellement présent dans l'organisme et éliminé uniquement par les reins. Un dosage sanguin de la créatinine (exprimé en  $\mu\text{mol/L}$ ) est souvent réalisé pour évaluer la fonction rénale. Lorsque celle-ci est dysfonctionnelle, le taux de créatinine augmente dans le sang. Toutefois cette mesure est insuffisante pour conclure sur une insuffisance rénale car le taux de créatinine sanguin est intrinsèquement lié à la masse musculaire. Les individus ayant un poids important ont tendance à avoir une créatinine plus élevée que les individus avec un poids moins élevé. Par exemple, les valeurs considérées comme « normales » chez la femme sont entre 50 et 90  $\mu\text{mol/L}$  chez la femme et entre 80 et 115  $\mu\text{mol/L}$  chez l'homme. Le taux de créatinine peut également varier en fonction de l'âge ou de l'ethnie. De ce fait, la créatininémie seule est donc un marqueur imparfait de la fonction rénale. Pour avoir une mesure plus précise, on évalue plutôt la clairance de la créatinine ou le débit de filtration glomérulaire (DFG) [26, 29].

### 1.2.3.2 Clairance urinaire de la créatinine sur les urines des 24 heures

La créatinine est librement filtrée par les glomérules, sécrétée par les tubules rénaux et éliminée dans les urines. Par une collecte de l'urine sur une période définie (habituellement sur 24 heures), il est possible d'évaluer la clairance de la créatinine basée sur une formule classique :

$$\text{Clairance de la créatinine} = \text{UCr (mg/mL)} * \text{UVol (mL/min)} / \text{PCr (mg/mL)}$$

*UCr = Créatinine urinaire, UVol = Volume d'urine sur 24 heures, PCr = Créatinine plasmatique*

Cette mesure est indépendante de la masse musculaire et la valeur obtenue est proche du DFG. Elle reste la manière la plus simple de déterminer la clairance de la créatinine afin d'évaluer la fonction rénale surtout lorsqu'elle est combinée avec une mesure des taux sériques de créatinine aux mêmes moments. Néanmoins, le taux de créatinine mesuré est surestimé car la créatinine est également sécrétée par le tubule. De plus, elle peut être imprécise car sa fiabilité dépend grandement de la précision du recueil des urines de 24 heures souvent pris en défaut (vidange vésicale incomplète, miction jetée etc.). Ces facteurs rendent cette mesure imparfaite. Elle est surtout recommandée en cas de grossesse [26].

### 1.2.3.3 Mesure du DFG par calcul de la clairance de substances exogènes

Ces mesures basées sur des méthodes radio-isotopiques ou non isotopiques sont les plus précises pour la mesure du DFG et représentent le Gold standard pour évaluer de façon exacte la filtration rénale. Elles reposent sur l'utilisation de produits injectés dans le sang et dont on mesure l'élimination urinaire. Ces techniques sont longues et coûteuses et nécessitent une certaine expertise (structures spécialisées). Plusieurs traceurs peuvent être utilisés dont l'Inuline, l'EDTA marqué au Chrome, l'Iothalamate radioactif marqué à l'iode-125, et l'Iohexol (produit de contraste iodé). Ces substances sont librement filtrées, non fixées aux protéines, non sécrétées et non réabsorbées. Leur dosage est sensible ce qui permet une mesure précise. Après injection du produit, on peut estimer la clairance dans le plasma (non filtré), dans les urines (filtré) ou bien dans le rein par scintigraphie (capter par le rein) [30].

### 1.2.3.4 Autres mesures de la fonction rénale

Il existe d'autres méthodes pour mesurer la fonction rénale basée sur le dosage de substances endogènes mais aucune de ces substances ne représente un marqueur parfait. Néanmoins ces mesures peuvent être effectuées en complément ou lorsque les autres mesures sont insuffisantes pour évaluer la fonction rénale.

#### — L'azote uréique

L'urée est un déchet du métabolisme des protéines. Elle est moins cohérente que la créatinine sérique dans l'évaluation de la fonction rénale, car elle est également affectée par de nombreux autres facteurs, tels qu'un apport élevé en protéines, un traitement par corticostéroïdes, des fluctuations de poids etc. De plus, l'urée est réabsorbée dans le néphron (à hauteur de 50%), surtout en cas d'atteinte rénale dans les premiers stades de la maladie rénale. La concentration de l'azote uréique sanguin peut néanmoins constituer un indice de la fonction rénale [26, 31].

#### — La moyenne des clairances de l'urée et de la créatinine

Cette mesure est utilisée chez les patients sous dialyse, afin d'évaluer le pourcentage restant de la fonction rénale. La moyenne des clairances de l'urée et de la créatinine est plus précise que l'une ou l'autre de ces valeurs prises individuellement, car les variations individuelles entre la quantité d'urée réabsorbée et la quantité de créatinine activement sécrétée se compensent. Bien qu'ils soient évocateurs, ces mesures ne sont toutefois pas fiables pour l'évaluation des troubles rénaux [31].

#### — L'albumine

L'albumine est une protéine présente dans le sang et qui n'est normalement pas filtrée dans l'urine. Cependant, si le rein est endommagé, il permet à une certaine quantité d'albumine de s'échapper dans l'urine. On peut donc évaluer le taux d'albumine urinaire excrété ou le rapport albumine/créatinine dans l'urine. Une albuminurie modérément

augmentée correspond à une excrétion persistante d'albumine de 30 à 300 mg/jour (20 à 200 mcg/min). Lorsque l'albumine urinaire est supérieure à 300 mg/jour ( $> 200$  mcg/min), on est dans le cas d'une protéinurie avérée. Cela peut signifier que les reins sont endommagés. L'albuminurie peut cependant être augmentée pour d'autres raisons tels qu'un diabète sucré, une hypertension, une infection urinaire, une maladie rénale chronique etc [31].

— **La cystatine C**

La cystatine C est une protéine du sang parfois mesurée en tant qu'indicateur de la fonction rénale. C'est un inhibiteur de la protéinase sérine qui est produite par toutes les cellules nucléées et filtrée par les reins. Elle est surtout utilisée comme mesure de confirmation d'une atteinte rénale spécifiquement lorsque les calculs du DFG basé sur les méthodes standards sont moins précis et/ou lorsque le DFG estimé est de 45 à 59 mL/min en l'absence de marqueurs de lésions rénales [32].

### 1.2.3.5 Estimation du Débit de Filtration Glomérulaire

Cette estimation est basée sur des formules mathématiques qui en plus du taux de créatinine dans le sang prennent en compte des facteurs comme le poids, l'âge, l'ethnie le sexe afin d'avoir une mesure plus précise. La valeur de créatinine dans le plasma est inversement liée au DFG, donc à la fonction rénale. Trois formules sont principalement utilisées pour estimer la clairance de la créatinine qui est elle-même une estimation du DFG (formule de Cockcroft) ou le DFG (formules MDRD (Modification of Diet in Renal Disease) et CKD-EPI (Chronic kidney disease epidemiology collaboration)) [26, 31].

— **La formule de cockcroft**

La formule de Cockcroft et Gault a été établie en 1976 par D.W. Cockcroft et H. Gault [33]. Elle estime la clairance de la créatinine en prenant en compte l'âge, le sexe, le poids et la créatininémie. Le résultat obtenu est en ml/min. Cette formule est simple d'utilisation, mais de moins en moins utilisée en pratique clinique pour l'estimation du DFG, car la méthode de dosage de la créatininémie sur laquelle elle se base est obsolète avec de mauvaises performances surtout chez les patients âgés ou en surpoids. Toutefois la formule conserve un intérêt clinique pour l'adaptation posologique des médicaments ayant été développés avec la formule de Cockcroft [26, 31].

*Formule de la clairance de Cockcroft normalisée :*

$$\text{Clairance} = \frac{[140 - \text{âge}(\text{années}) \times \text{poids}(\text{kg})]}{\text{créatininémie}(\mu\text{mol/L})} \times k \quad (1.1)$$

$k=1,23$  pour les hommes et  $k=1,04$  pour les femmes.

— **La formule MDRD (Modification of Diet in Renal Disease)**

La formule MDRD a été développée en 1999 [34]. Elle permet d'estimer directement le DFG indexé sur la surface corporelle. Initialement, l'équation MDRD comprenait six variables. Elle a été simplifiée en 2006 en une équation à quatre variables, normalisée pour une surface corporelle de 1,73 m<sup>2</sup> et qui inclue le sexe, l'âge, la créatininémie et l'ethnie, mais pas le poids [26, 31].

*Formule MDRD simplifiée chez l'homme (DFG estimé en mL/min/1,73 m<sup>2</sup>) :*

$$186 \times \text{créatinine } (\mu\text{mol/L}) \times 0,0113^{-1,154} \times \text{âge (années)}^{-0,203} \times k \quad (1.2)$$

*k=0,742 pour les femmes et k=1,21 pour les afro-américain.*

— **La formule CKD-EPI (Chronic kidney disease epidemiology collaboration)**

La formule CKD-EPI a été définie en 2009 [32]. Elle permet également d'évaluer le DFG en considérant la créatininémie, l'âge, le sexe et l'origine ethnique. La formule normalise le résultat pour une surface corporelle standard de 1.73m<sup>2</sup> afin de permettre la comparaison de DFG entre patients [26, 31].

*Formule du débit de filtration glomérulaire estimé en mL/min/1.73 m<sup>2</sup> :*

$$141 \times \min(C/\kappa, 1)^\alpha \times \max(C/\kappa, 1)^{-1,209} \times 0.993^{\text{âge(années)}} \quad (1.3)$$

*×1.159 pour les sujets d'origine africaine (African American) ;*

*×1.018 pour les femmes ;*

*$\kappa = 0.7$  pour les femmes ou  $0.9$  pour les hommes ;  $\alpha = -0.329$  pour les femmes ou  $-0.411$  pour les hommes ; C : Créatinine sérique en  $\mu\text{mol/L}$ .*

Il est important de noter qu'il existe d'autres formules de CKD-EPI qui ne sont pas basées sur la créatinine. Il s'agit de l'estimation du DFG basée sur la cystatine C (CKD-EPI CYST) [35], ou encore la combinaison de la créatinine et de la cystatine (CKD-EPI CYST-CREAT) [36]. L'évaluation du DFG avec la cystatine permet une mesure moins dépendante de la masse musculaire que la créatinine [37].

Les équations MDRD et CKD-EPI sont plus précises que celle de Cockcroft qui n'estime pas directement le DFG mais la clairance de la créatinine. La formule CKD-EPI est plus précise surtout pour les valeurs basses de créatininémie (taux de filtration glomérulaire inférieur à 60 mL/min pour 1,73 m<sup>2</sup>). La Haute Autorité de Santé recommande d'utiliser la formule CKD-EPI à partir d'une créatininémie dosée par méthode enzymatique, à défaut la formule MDRD

si la créatininémie est dosée par méthode colorimétrique (méthode dite de Jaffé) [27]. Toutefois, ces formules restent des estimateurs du DFG et sont donc sensibles aux facteurs pouvant faire fluctuer le DFG (dénutrition majeure, variation aiguë de la fonction rénale, surpoids, grossesse etc.). De plus, ces équations n'ont pas été validées dans certaines populations à savoir les enfants (utiliser la formule de Schwartz, spécifique de l'enfant), les populations non afro-américaines, les patients âgés de plus de 75 ans etc [26].

## 1.3 L'insuffisance rénale

Lorsque les reins n'assurent plus leur rôle de filtration et de régulateur, les toxines et l'eau s'accumulent dans l'organisme et conduisent à une insuffisance rénale qui est dite aiguë (IRA) si l'altération de la fonction rénale est réversible et chronique (IRC) lorsque le dysfonctionnement est irréversible [26, 38].

### 1.3.1 L'insuffisance rénale aiguë (IRA)

L'IRA se définit comme une altération soudaine et rapide du fonctionnement rénale qui se traduit par une accumulation des déchets organiques et des troubles de la régulation hydrominérale et acido-basique dans l'organisme [39]. L'IRA se manifeste par une augmentation de la créatininémie en quelques jours ou quelques semaines associée à une diurèse faible, inférieure à 500ml/24h (soit inférieure à 20ml/h). Du fait de cette augmentation soudaine, aucune formule d'estimation du DFG n'est utilisable en cas d'IRA. Toutefois, selon une définition universelle de l'insuffisance rénale aiguë établie par KDIGO (Kidney Disease/Improving Global Outcome 2012) [40], il est possible de poser un diagnostic et de déterminer l'évolution de l'IRA en se basant sur les taux de créatininémie et le volume de la diurèse. On distingue 3 stades :

- Le stade 1 correspond à une augmentation de la créatininémie  $> 26 \mu\text{mol/L}$  (3 mg/L) en 48 h avec une diurèse  $< 0,5 \text{ ml/kg/h}$  pendant 6 à 12h.
- Le stade 2 est atteint lorsque le taux de créatininémie double en l'espace d'une semaine avec une diurèse  $< 0,5 \text{ ml/kg/h}$  pendant plus de 12h.
- Le stade 3 correspond à une créatininémie qui triple en une semaine avec une diurèse  $< 0,3 \text{ ml/kg/h}$  pendant plus de 24 h. Ce stade est le plus avancé et peut nécessiter une dialyse d'urgence.

Les causes de l'IRA sont nombreuses. Elle peut survenir sur des reins sains suite à un traumatisme grave, une intervention médicale etc. ou être la conséquence d'une poussée aiguë dans le cadre d'une insuffisance rénale chronique. Dans la majorité des cas l'atteinte de la fonction rénale est réversible. Selon le mécanisme étiologique, les IRA se classent en trois catégories d'IRA : les IRA pré-rénales ou fonctionnelles, les IRA post-rénales ou obstructives et les IRA rénales ou organiques [41].



## 1.4 L'insuffisance rénale chronique (IRC)

L'insuffisance rénale chronique (IRC) correspond à la perte progressive et irréversible des fonctions des reins, caractérisée par un DFG (débit de filtration glomérulaire) inférieur à 60 ml/min par 1,73 m<sup>2</sup> de surface corporelle ; l'absence de réversibilité au-delà de 3 mois signe la chronicité [42]. L'IRC est souvent secondaire à une maladie chronique rénale (MRC) incluant les néphropathies vasculaires et hypertensives (25%), les néphropathies diabétiques (22%, essentiellement diabète type 2), les glomérulonéphrites chroniques (11%), les néphropathies héréditaires (8%, essentiellement polykystose rénale), les néphropathies interstitielles chroniques (moins de 5%), les néphropathies diverses (10%), les néphropathies d'origine indéterminée (16 %) [43, 44]. La MRC est définie par la présence de signes biologiques de néphropathie évolutive (protéinurie, hématurie etc.) avec ou sans altération de la fonction rénale ou par une insuffisance rénale. Suivant cette définition, l'évolution des MRC est subdivisée en 5 classes [45] (Table 1.2) :

- Stade 1 : Présence de signes de néphropathie sans altération de la fonction rénale (DFG > à 90 mL/min par 1,73 m<sup>2</sup>)
- Stade 2 : Présence de signes de néphropathie avec une réduction du DFG (89 < DFG < 60 mL/min par 1,73 m<sup>2</sup>)
- Stade 3 Insuffisance rénale modérée (59 < DFG < 30 mL/min par 1,73 m<sup>2</sup>)
- Stade 4 Insuffisance rénale sévère (29 < DFG < 15 mL/min par 1,73 m<sup>2</sup>)
- Stade 5 Insuffisance rénale terminale (DFG < 15 ml/min/1,73 m<sup>2</sup>)

TABLE 1.2 – *Les stades de la maladie rénale chronique* [43]

Stade	Description	DFG (ml/min/1,73 m <sup>2</sup> )
1	Maladie rénale chronique* avec fonction rénale normale	≥ 90
2	Maladie rénale chronique* avec insuffisance rénale légère**	60-89
3A	Insuffisance rénale légère à modérée	45-59
3B	Insuffisance rénale modérée à sévère	30-44
4	Insuffisance rénale sévère	15-29
5	Insuffisance rénale terminale	< 15

\* Avec marqueurs d'atteinte rénale : protéinurie clinique, hématurie, leucocyturie, ou anomalies morphologiques ou histologiques ou marqueurs de dysfonction tubulaire, persistant plus de 3 mois.

\*\* Un DFG compris entre 60 et 89 ml/min peut être normal chez un sujet âgé.

Le rythme de progression de l'IRC est très variable en fonction de la MRC. Aux premiers stades l'IRC est inexistante ou légère et s'intensifie aux stades 3 et 4. Les modifications biochi-

miques apparaissent très tardivement. En effet, les néphrons sains s'hypertrophient et continuent à assurer leur rôle en augmentant leur capacité de travail afin de compenser le déficit fonctionnel. Le parenchyme rénal se réduit et le rein devient progressivement fibreux. Lorsque 50 % ou plus des néphrons ne sont plus fonctionnels, les troubles cliniques commencent à apparaître (hypertension artérielle, troubles cardio-vasculaires, anémie, hyperkaliémie, dénutrition etc.). L'adaptation excessive des néphrons entraîne leur dégradation et conduit à une évolution vers les stades 4 et 5 de la maladie qui nécessitent souvent un traitement de suppléance pour les reins (dialyse ou transplantation rénale) [43]. Il est donc primordial de pouvoir diagnostiquer précocement les dysfonctionnements rénaux afin de pouvoir mettre en œuvre un traitement et un suivi dans le but de ralentir la progression de l'insuffisance rénale chronique. Le diagnostic de l'IRC repose sur la mesure du DFG. De préférence, il est préconisé d'utiliser l'équation CKD-EPI. La mesure de l'albuminurie contribue aussi à l'interprétation clinique de l'évolution de la maladie puisque ce facteur est significativement associé au risque de progression des MRC vers une insuffisance rénale terminale [46]. Chez certains individus, le risque de développer une IRC est plus élevé (hypertendus, diabétiques, individus de plus de 60 ans, individus en surpoids, individus ayant des antécédents d'uropathies, de néphropathies ou cardiovasculaires etc.). Les néphropathies diabétiques et hypertensives représentent près de la moitié des causes d'IRC. Ces atteintes rénales touchent encore plus les sujets âgés de plus de 65 ans qui représentent 17 % de la population générale et 60 % des patients commençant une dialyse de suppléance. Ces populations « à risques » doivent passer par un dépistage annuel afin d'anticiper l'apparition d'une MRC [43]. En terme de traitements, la prise en charge dépend du degré de l'IRC. Lorsque la MRC est détectée précocement (stade 1 et 2), il est possible de ralentir la progression de la maladie par la détection des facteurs de risques et la prescription de traitements en fonction du type de MRC tels que des immunosuppresseurs, les stéroïdes etc. Associé à cela, il est important de prendre en charge les facteurs de progression : contrôle strict de la pression artérielle, diminution de la protéinurie, utilisation d'inhibiteurs de l'enzyme de conversion et des récepteurs de type 1 de l'angiotensine II, prévention des épisodes d'insuffisance rénale aiguë, restriction protidique modérée et adaptée au patient, contrôle d'un diabète s'il existe et l'arrêt du tabac. Lorsque l'IRC atteint la phase terminale, les derniers recours sont la filtration externe du sang dans un rein artificiel (dialyse), ou la transplantation rénale [43, 47].

#### 1.4.1 L'insuffisance rénale chronique terminale (IRCT)

L'insuffisance rénale chronique terminale en France a une incidence de 157 par million d'habitants (pmh) et par an (10 697 cas en 2020) et une prévalence de 749 pmh et 605pmh pour la dialyse et la transplantation respectivement. Elle est 2 à 3 fois plus fréquente chez l'homme que chez la femme et est plus prévalente avec l'âge. Chez la plupart des patients en France, l'IRCT peut être traitée par la transplantation rénale ou la dialyse. En 2020, 92 199 personnes sont ainsi

traitées dont 55% par dialyse et 45% par transplantation rénale [48].

#### 1.4.1.1 La dialyse

La dialyse est une méthode de suppléance qui correspond à l'ensemble des techniques d'épuration extrarénales permettant d'assurer artificiellement les fonctions de filtration et de régulateur des reins. Cette méthode nécessite un échange entre le sang du patient et une solution de composition électrolytique proche de celle du plasma, à travers une membrane semi-perméable. Le but est de pouvoir épurer le sang des déchets azotés et de corriger, au moins partiellement, les troubles hydro-électrolytiques, phosphocalciques et acido-basiques qui résultent de la défaillance de la fonction rénale. La dialyse est généralement indiquée lorsque le DFG est inférieure à 10 ml/min/1.73 m<sup>2</sup> ou inférieure à 15 ml/min chez les sujets diabétiques, âgés et/ou dénutris. D'un point de vue clinique, il est préconisé de débiter le traitement avant l'apparition des complications graves de l'IRCT tels qu'une péricardite, une surcharge hydrosodée ou une atteinte neurologique. Il existe deux types de dialyse : l'hémodialyse basée sur une circulation sanguine extracorporelle et la dialyse péritonéale qui utilise des méthodes intracorporelles. Le choix de l'une des deux techniques de dialyse repose sur des critères médicaux (contre-indication d'une des techniques) mais également sur la situation personnelle, le mode de vie et l'autonomie du patient [47].

##### 1.4.1.1.1 L'hémodialyse

L'hémodialyse est la technique de dialyse la plus utilisée en France (> 90%) [49]. Elle consiste à l'épuration extracorporelle du sang au travers d'une membrane artificielle intégrée dans une machine appelée le dialyseur (Figure 1.2.A). Pour cette technique, un abord vasculaire particulier est nécessaire : la fistule artério-veineuse qui doit être installée des semaines avant le début de la dialyse. L'appareillage d'hémodialyse comprend [50] :

- Une pompe qui permet d'aspirer, à travers la fistule artério-veineuse, le sang chargé de déchets et de le renvoyer dans le corps une fois épuré.
- Un générateur qui produit un liquide stérile appelé dialysat dont la composition est proche du plasma humain.
- Un dialyseur composé de fibres capillaires (poils) qui jouent le rôle de surface d'échange entre le sang circulant dans les poils et le dialysat qui circule tout autour de ces poils. Le sang est donc séparé du liquide de dialyse par une très fine membrane perméable qui va permettre le passage de molécules dans les deux sens. Ainsi, les déchets passent du sang vers le dialysat. Une fois épuré, le sang retourne à nouveau dans la circulation sanguine via la fistule artério-veineuse et le dialysat chargé de déchets est jeté [50].

#### 1.4.1.1.2 La dialyse péritonéale

La dialyse péritonéale est la méthode de suppléance la moins utilisée en France (< 10%). C'est une technique intracorporelle basée sur l'utilisation d'un filtre interne particulièrement riche en vaisseaux : le péritoine (Figure 1.2.B). Ce dernier est une membrane fine qui tapisse toute la cavité abdominale ainsi que les organes qui s'y trouvent. La dialyse péritonéale consiste à mettre en contact un liquide sain (le dialysat) avec le sang en utilisant la membrane péritonéale comme paroi filtrante (dialyseur); l'idée étant de faire migrer les toxines depuis le sang vers le dialysat en traversant le péritoine [43]. Pour se faire, un cathéter est introduit dans la cavité péritonéale, sous anesthésie générale ou locale. Ceci permettra de pouvoir injecter le dialysat et le retirer une fois rempli de déchets. L'extrémité du cathéter dépasse d'une dizaine de centimètres à l'extérieur de l'abdomen et est fixée sur une tubulure adaptée pour chaque séance de dialyse. La dialyse péritonéale est généralement pratiquée à domicile soit par le patient lui-même soit par un infirmier.

En termes de performances d'épuration, la dialyse péritonéale est moins efficace que l'hémodialyse (difficultés techniques chez les patients de fort gabarit) et son utilisation est limitée à quelques années du fait de l'altération progressive des propriétés du péritoine. Néanmoins, les techniques de dialyse péritonéale permettent une plus grande autonomie du patient et s'avèrent moins contraignantes sur le plan de la vie quotidienne en comparaison avec l'hémodialyse qui est en plus du manque d'autonomie est bien plus coûteuse (25 000 à 50 000 euros/an ) et génère plus d'effets secondaires dont complications infectieuses, bactériennes ou virales (hépatites B et C par exemple). Dans tous les cas, quelle que soit la technique, la dialyse reste un traitement lourd qui impacte grandement la qualité de la vie du patient et dont le taux de mortalité reste toutefois plus importante que pour les personnes transplantées [43, 51]. Le taux de mortalité en dialyse est de 16 pour 100 patients-années alors que celui de la greffe est de 3 pour 100 patients-années depuis 2008 [48].

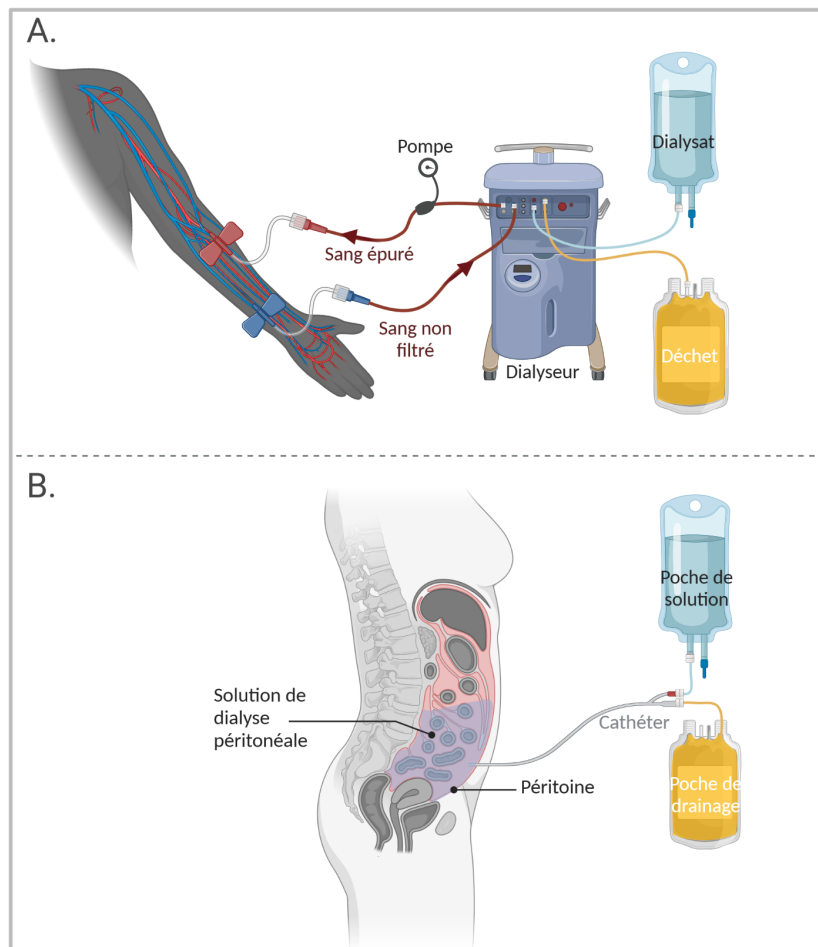


FIGURE 1.2 – **A) Représentation schématique d'un dispositif d'hémodialyse.** L'épuration du sang va se faire à travers le passage d'une membrane et par échange avec un dialysat propre. En hémodialyse, la membrane est une membrane artificielle contenue dans le dialyseur. Le dialyseur est alimenté dialysat et en sang provenant du patient. **B) Représentation schématique d'un dispositif de dialyse péritonéale.** L'épuration du sang se fait à travers une membrane naturelle, le péritoine. Un cathéter permanent permet le transfert du dialysat dans la cavité péritonéale.\*

\* Figure créée avec BioRender.com.

## 1.5 La transplantation

### 1.5.1 Pré-transplantation

Une transplantation rénale est une intervention chirurgicale qui consiste à placer un rein sain provenant d'un donneur vivant ou décédé avec rétablissement de la continuité vasculaire chez un receveur dont les fonctions rénales sont irréversiblement altérées. Le rein transplanté appelé greffon est généralement placé plus bas que la position anatomiquement normale (dans la fosse iliaque) sans nécessité de retirer les reins malades. Lorsqu'elle est possible, la transplantation est la meilleure méthode de suppléance de la fonction rénale. Elle permet une meilleure qualité de vie par rapport à la dialyse, une morbidité cardio-vasculaire moindre et une espérance de vie supérieure [52, 53]. Selon les chiffres de l'ABM [44], 91.8% des greffons sont encore fonctionnels après 1 an, 77.9% après 5 ans et 58.3% après 10 ans. Depuis 1959, année de la première greffe rénale enregistrée dans Cristal\* , un total de 94 160 greffes rénales a été enregistré : il s'agit de la greffe la plus courante en France [54]. Le nombre estimé de malades porteurs d'un greffon rénal fonctionnel est de 42 068 au 31 décembre 2020 soit une prévalence de 620 pmh [54].

En 2021, 3 251 transplantations ont été effectuées dont 502 greffons provenant de donneurs vivants [19]. Néanmoins, malgré le nombre croissant de transplantations rénales chaque année, la liste des patients en attente de greffe continue de s'allonger. En 2021, la liste de d'attente de l'ABM comptabilise 17 148 patients inscrits [44] soit plus de 4 candidats à la transplantation pour 1 greffon. 57% des malades sont greffés 3 ans après leur inscription sur le registre de greffe [44]. Il y a un manque évident de greffons rénales qui rend difficile l'accès à la transplantation pour beaucoup bien qu'elle soit la meilleure méthode de suppléance. Les patients pour qui une transplantation rénale est envisagée doivent respecter certains critères d'éligibilité. Le greffon peut provenir d'un donneur vivant ou décédé compatible (ABO et HLA) et son attribution est réglementée par l'Agence de la Biomédecine sur la base de critères d'équité et d'égalité entre les patients [44].

#### 1.5.1.1 Critères d'éligibilité à la greffe

L'indication principale en faveur de la transplantation rénale est une insuffisance rénale en phase terminale. L'inscription sur la liste de transplantation rénale peut être réalisée dès que le DFG est à 20 ml/min/1,73 m<sup>2</sup>. Dans la plupart des cas où elle est possible, elle peut être envisagée au stade 5, avant qu'un traitement par dialyse ne soit institué (greffe préemptive). C'est d'ailleurs l'option de choix de la transplantation rénale car elle permet d'éviter la contrainte et les complications de la dialyse et offre une meilleure survie pour le patient comme pour le greffon [55]. Un bilan biologique et clinique est effectué chez le patient pour détecter toute

---

\*Logiciel développé par l'ABM pour l'attribution efficace des greffons et qui recense toutes les informations relatives aux transplantations et greffes effectuées en France

contre-indications qui pourraient mettre en péril la survie du greffon ou du patient après la transplantation [52]. On distingue : les contre-indications absolues qui concernent généralement les troubles cardiovasculaires sévères, les cancers, la présence de maladies rénales avec risque important de récurrence sur le greffon etc. et les contre-indications relatives qui comprennent le diabète, l'âge, la séropositivité au VIH etc. Ce bilan permet également d'établir le typage HLA et le groupage ABO et rhésus du patient afin de pouvoir sélectionner un donneur compatible et d'évaluer l'immunisation du patient qui sera importante dans le contrôle des complications post-transplantation (rejets et perte de greffe) [55, 56]

### 1.5.1.2 Sélection des donneurs

Initialement, la sélection d'un donneur potentiel était basée sur des critères très stricts afin de réduire au maximum le risque d'altération du greffon après transplantation et d'assurer une meilleure survie du patient. Le donneur devait être un sujet jeune (moins de 50 ans), en bonne santé, sans antécédents ou facteurs de risques de comorbidités (maladies cardio-vasculaire par exemple) et sans lésions sur l'organe prélevé [57]. Dans ces conditions, la majorité des donneurs étaient des victimes d'accident de la voie publique (AVP) avec un traumatisme crânien grave évoluant vers un état de mort encéphalique. Avec la baisse très importante des décès liés aux AVP et l'augmentation progressive du nombre d'inscrits dans les listes d'attente de greffe, la pénurie d'organes est très vite devenue un problème majeur. Pour pallier à cette carence, certains critères de sélection ont été élargis et des programmes de recherche d'autres sources de greffons ont été mis en place. Ceci concerne : le prélèvement sur donneur décédé à critères élargis, le prélèvement sur donneur décédé après arrêt cardiaque, et le prélèvement à partir de donneurs vivants [57].

#### 1.5.1.2.1 Donneurs décédés

Il s'agit du type de donneurs le plus courant (84% des dons de rein en 2021) [19]. Les greffons provenant de ce type de donneurs assurent une survie de 91% à 1an, 76% à 5 ans et 56% à 10ans post-greffe [58, 59].

Les donneurs décédés peuvent être divisés en 2 catégories :

— **Les donneurs décédés en mort encéphaliques (DDME) :**

La mort encéphalique est définie comme la destruction définitive et irréversible de l'encéphale avec absence totale de conscience, d'activité motrice ou de ventilation spontanée et abolition de tous les réflexes du tronc cérébral [56]. Elle représente moins de 1 % de tous les décès (0,58 %) mais constitue la principale source de donneurs de greffe rénale (89% en 2019). On différencie les donneurs à critères standard (âgé > 50, sans facteurs de comorbidité associé) et les donneurs à critères élargis [57, 60]. Il s'agit des donneurs âgés de plus de 60 ans ou les donneurs entre 50 et 59 ans présentant des comorbidités tels

qu'une hypertension, un fonctionnement rénale réduit (créatinine  $\geq 133 \mu\text{mol/L}$ ) etc. Ces donneurs « marginaux » ont des organes de qualité inférieure et présentent un risque potentiel de retard à la reprise de fonction, un taux de non fonction primaire plus important, une durée de fonctionnement moindre et une sensibilité plus importante à l'ischémie. Les donneurs à critères élargis représentaient 2,2% en 1996 et sont plus de 30% en 2021 [59]. Le recours à ces dons de moins bonne qualité était pour pallier au manque de greffon. Pour augmenter les chances de fonctionnement du greffon et de survie du patient, le patient peut recevoir les 2 reins du même donneur (bigreffe) lorsque le DFG estimé du donneur est  $< 60 \text{ mL/min/1,73 m}^2$ .

— **Les donneurs décédés après arrêt cardiaque (DDAC) :**

Il s'agit des donneurs décédés présentant un arrêt cardiaque et respiratoire persistant. Ils ont généralement âgés de moins de 50 ans et leur décès a une origine cardiaque, sans antécédents [56]. Les DDAC sont classés en 4 catégories dites de Maastricht. On distingue : I) Les personnes qui font un arrêt cardiaque en dehors de tout contexte de prise en charge médicalisée et pour lesquelles le prélèvement d'organes ne pourra être envisagé que si la mise en œuvre de gestes de réanimation de qualité a été réalisée moins de 30 min après l'arrêt cardiaque. (II) Les personnes qui font un arrêt cardiaque en présence de secours qualifiés, aptes à réaliser un massage cardiaque et une ventilation mécanique efficaces, mais dont la réanimation ne permettra pas une récupération hémodynamique. (III) Les personnes hospitalisées pour lesquelles une décision d'un arrêt des traitements est prise en raison de leur pronostic. (IV) Les personnes décédées en mort encéphalique qui font un arrêt cardiaque irréversible au cours de la prise en charge de réanimation. Les classes I et II de Maastricht sont catégorisés comme des arrêts non contrôlés et étaient plus fréquents au début des années 2000. Les classes 3 et 4 sont des arrêts contrôlés et concernent les dons plus récents [56]. Néanmoins, ces dons présentent plus de risques que ceux issus des DDME. En effet, ces reins peuvent avoir été lésés par une ischémie chaude avant le décès et leur fonction est souvent altérée par une nécrose tubulaire aiguë. Les receveurs de greffons prélevés sur des DDAC présentent un retard de reprise de la fonction rénale plus important (80 à 90% des cas [61]). Cependant, le pronostic à 1 an est très comparable à ceux provenant des DDME (95% de survie) surtout pour la catégorie III de Maastricht (94% de survie) [56, 60].

#### 1.5.1.2.2 Donneurs vivants

Les transplantations rénales à partir de donneurs vivants sont celles qui offrent les meilleures performances en terme de survie du patient et du greffon. Elle est de 96 % à 1 an, 89% à 5 ans et 74 % à 10 ans post-greffe [58]. Les greffons issus des donneurs vivants sont de très bonne qualité en raison du bon état général du donneur, de l'absence de problèmes liés à la conservation



du greffon et des procédures de transplantation bien codifiées (durée d'ischémie froide courte, chirurgies programmées et contrôlées etc.) [56, 59].

Toutefois, les donneurs vivants sont relativement rares surtout en France. Seulement 16 % des transplantations rénales sont réalisées à partir d'un donneur vivant en 2021 contre 40 à 50 % aux États-Unis ou dans les pays scandinaves [19]. La pénurie d'organes a conduit progressivement à élargir le cercle des donneurs vivants en France. Initialement, les donneurs vivants éligibles ne concernaient que les personnes ayant un lien familial ou affectif avec le receveur (parents, enfants, fratrie, conjoints). Depuis 2011, le don du vivant est élargi à toute personne pouvant apporter la preuve d'un lien affectif étroit et stable depuis au moins deux ans avec le receveur [57].

En amont de la transplantation, un bilan est nécessaire pour s'assurer qu'il n'y a pas de contre-indication à la transplantation et pour établir la compatibilité entre le donneur et le receveur. Ce bilan comporte une évaluation de la fonction rénale, une recherche de potentielles comorbidités (hypertension, diabète etc.), une recherche d'anomalie du sédiment urinaire (protéinurie ou hématurie) etc [56, 55]. Dans le cas où il y'a une incompatibilité entre le donneur vivant et le receveur, il est possible de procéder à un don croisé : le donneur d'un couple A donne son rein au receveur du couple B et vice-versa. Les deux opérations chirurgicales sont alors engagées simultanément, tout en respectant l'anonymat entre les deux paires [56, 55]. La protection des futurs donneurs est assurée par un comité qui encadre la décision du donneur en s'assurant qu'elle est libre, éclairée consentie et ne présente pas de dangers majeurs pour la santé du candidat [56, 55].

### 1.5.1.3 Compatibilité Donneur/Receveur

La réussite d'une transplantation rénale est en grande partie régie par la compatibilité immunologique entre le patient et son donneur qui implique une compatibilité en groupe sanguin (système ABO) et en HLA (Human Leukocyte Antigen). Cette compatibilité immunologique permet de limiter le risque de développer des réactions immunitaires contre le greffons (rejets ou perte du greffon post-transplantation) [60].

#### 1.5.1.3.1 Le système ABO

Le système ABO, découvert par Karl Landsteiner en 1901 [62], est défini par la présence ou non de deux antigènes A et B, à la surface des globules rouges. Lorsque les globules rouges n'expriment pas les antigènes A ou B, des anticorps (anti A et anti B) sont produits par l'individu. En fonction de cette définition, les individus sont classés entre les groupes A, B, AB et O :

- Une personne du groupe A possède l'antigène A et produit des anticorps contre l'antigène B. Elle ne peut pas être transfusée avec du sang de groupe B ou AB.
- Une personne du groupe B possède l'antigène B et des anticorps anti-A. Elle est incompatible avec les groupe A ou AB.

- Une personne du groupe AB porte les antigènes A et B et ne produit aucun anticorps. Elle sera donc compatible avec tous les autres groupes sanguins ; on parle de receveur universel.
- Une personne du groupe O ne porte ni l'antigène A, ni l'antigène B. Elle produit par conséquent des anticorps dirigés contre ces deux groupes et ne peut être compatible qu'avec le groupe O. Par contre, les individus du groupe O peuvent donner du sang à tous les autres groupes du fait de l'absence d'antigène à la surface des cellules sanguines ; on parle de donneur universel [63]. (Figure 1.3)

En cas d'incompatibilité ABO entre les individus lors de la transfusion sanguine ou de la transplantation d'organe, un complexe anticorps-antigène se forme et conduit à une hémolyse (destruction des globules rouges) et/ou un choc transfusionnel [63]. La compatibilité ABO permet donc de limiter le risque de rejet du greffon post-transplantation. Dans le cas du don du vivant, il existe environ 36% de risque d'incompatibilité ABO entre le patient et le donneur vivant [64]. Une première solution est d'avoir recours au don croisé mais il n'est pas toujours évident de trouver rapidement deux couples compatibles ABO et HLA. De nombreux pays dont le Japon, les États-Unis et la Suisse ont développé la greffe rénale ABO incompatible (ABOi) basée sur l'élimination des anticorps anti-ABO présents chez le receveur avant la transplantation [65, 66, 67, 68]. Des conditions rigoureuses sont requises pour le succès de ce type de transplantation impliquant l'absence d'anomalies immunologiques entre le donneur et le receveur, le traitement immunosuppresseur intraveineux et per os en amont de la greffe, un taux des anticorps inférieur à 1 :8 la veille de la transplantation etc. [64, 69].

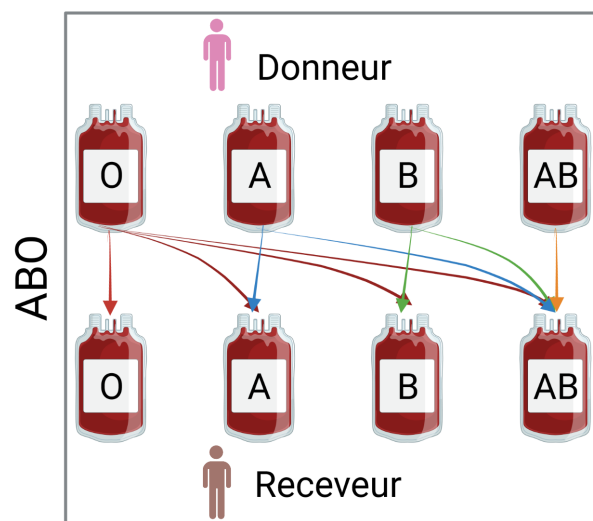


FIGURE 1.3 – *Compatibilité ABO entre donneur et receveur en fonction de leur groupe sanguin\**

### 1.5.1.3.2 Le système HLA (Human Leukocyte Antigen)

Le complexe majeur d'histocompatibilité (CMH) est un système de reconnaissance du soi qui assure la distinction entre les antigènes appartenant à l'organisme (le soi) et les antigènes étrangers (le non soi). Les antigènes peptidiques sont reconnus par les lymphocytes T (LT) grâce aux molécules du CMH présentes à la surface des cellules ; quand un antigène étranger est reconnu cela induit une réponse immunitaire [70]. Chez l'Homme, ce système est dénommé HLA (Human Leukocyte Antigen) et a été découvert par Jean Dausset en 1952 à la surface des leucocytes au cours d'expériences de transfusion et transplantation [10, 71]. La région du HLA est situé sur le bras court du chromosome 6 (6p21). Il ne que représente 0,1% du génome humain mais contient environ 1% des gènes codants [72]. On distingue deux catégories principales de molécules HLA présentes à la surface des cellules qui assurent la présentation des peptides antigéniques [73, 74] :

- **Les gènes de classe I** codent pour des molécules de surface de classe I du HLA. Parmi ces gènes, les plus polymorphiques sont les gènes dit classiques *HLA-A*, *HLA-B* et *HLA-C*. Les molécules du HLA de classe I sont présentes à la surface de toutes les cellules de l'organisme avec une expression plus importante au niveau des lymphocytes. Elles assurent la présentation des cellules du soi aux LT CD8 qui lorsqu'elles sont activées par la présence d'un corps étranger deviennent des lymphocytes cytotoxiques (LT CD8) et assurent la lyse de la cellule infectée [73, 74] (Figure 1.4).
- **Les gènes de classe II** codent pour des molécules de surface de classe II du HLA. Les plus importantes sont les gènes *HLA-DPA1*, *HLA-DPB1*, *HLA-DQB1* et *HLA-DRB1*. Les molécules du HLA de classe II sont présentes à la surface des cellules présentatrices d'antigènes (CPA) qui incluent les monocytes, les macrophages, les cellules dendritiques, les lymphocytes B (LB) et les cellules épithéliales du thymus. Les molécules de classe II vont présenter les peptides antigéniques aux LT CD4 qui activent la réponse immunitaire dite humorale traduite par la production massive par les LB, d'anticorps dirigés contre le corps étranger [73, 74] (Figure 1.4).

---

\* Figure créée avec BioRender.com.

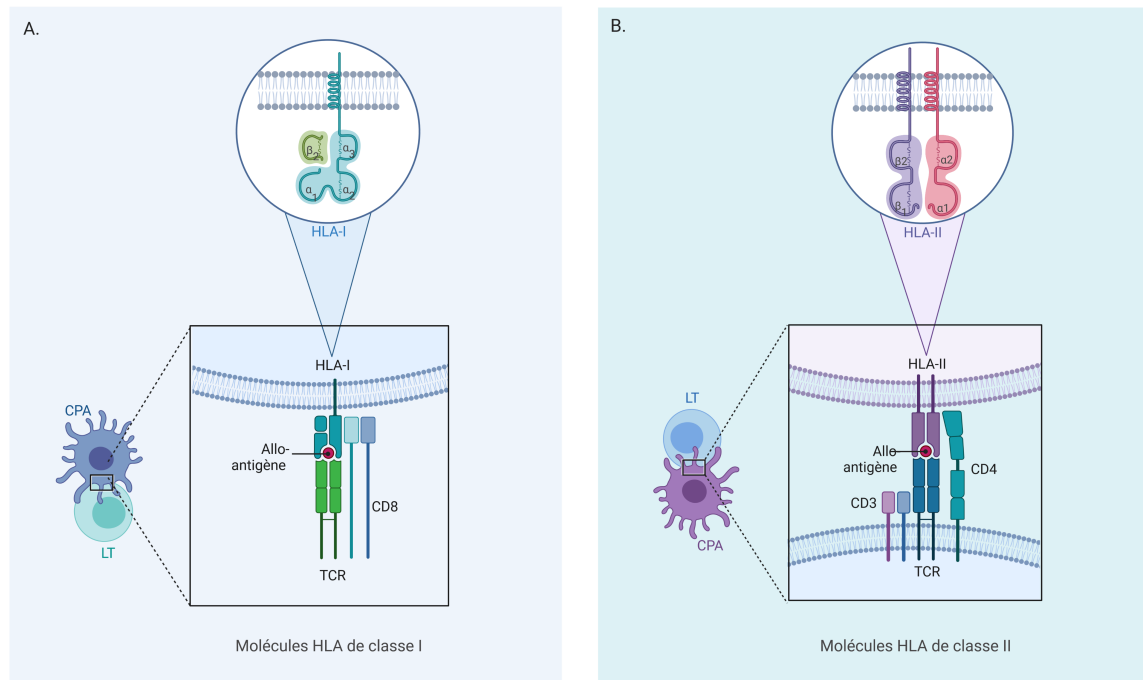


FIGURE 1.4 – **Présentation d'allo-antigène par les molécules HLA de classe I et II.**  
**A) Caractéristiques des molécules du CMH de classe I,** les molécules du HLA de classe I sont présentes sur toutes les cellules nucléées, elles ont pour fonction de présenter des peptides d'antigènes endogènes à des LT, les LT CD8 qui deviendront des LT cytotoxiques. Les molécules de classe I sont composées de deux chaînes polypeptidiques  $\alpha$  et  $\beta$ . **B) Caractéristiques des molécules du CMH de classe II** Les molécules du HLA de classe II sont présentes uniquement sur les CPA (cellules présentatrices d'antigènes), elles ont pour fonction de présenter des peptides d'antigènes exogènes à des LT, les LT CD4 qui deviendront des LT helpers. Les molécules de classe II sont également composées de deux chaînes polypeptidiques  $\alpha$  et  $\beta$ . \*

La région du HLA est extrêmement polymorphique [75], elle est dite hypervariable. Chaque individu porte deux allèles HLA, un du père et un de la mère [76]. Chaque gène HLA présente un très grand nombre d'allèles (20 076 allèles pour les gènes de classe I et 8244 allèles pour les gènes de classe II) [75]. De plus, ces allèles sont liés entre eux génétiquement, ils ne se transmettent pas de manière aléatoire à travers les générations, on parle de déséquilibre de liaison [77]. Afin d'identifier les allèles HLA portés par un individu, un typage est effectué grâce à plusieurs approches qui ont évolué au fur et à mesure des avancées technologiques. Les premières approches de typage étaient basées sur la sérologie correspondant à la caractérisation des anti-

\* Figure créée avec BioRender.com.

gènes exprimés à la surface des cellules grâce à des tests systématiques d'anti-sérum ou d'anticorps monoclonaux connus sur des lymphocytes T et B de patients (pour le HLA de classe I) ou de lymphocytes B (pour le HLA de classe II) [78]. Ces méthodes ont été succédées par des techniques de biologies moléculaires plus précises (la PCR-SSO et la PCR-SSP) qui contrairement à la sérologie, permettent une reconnaissance de séquences connues d'allèle ou de groupe d'allèles HLA permettant ainsi de différencier les allèles HLA qui étaient reconnus par les mêmes anticorps mais qui disposaient de séquences différentes [79]. Aujourd'hui, avec l'avènement des nouvelles technologies, la technique standard de typage HLA est le séquençage basé sur deux méthodes : le PCR-SBT, pour Sequence Based Typing ou séquençage de Sanger et les techniques de NGS (Next Generation Sequencing) [80, 81].

Cette évolution des techniques de typage au cours des dernières décennies a amélioré le niveau de précision impactant parallèlement la nomenclature des allèles HLA. Avec la limitation des techniques sérologiques à la détection de l'expression antigénique en surface cellulaire, la nomenclature sérologique se limitait au locus suivi d'un numéro identifiant l'antigène (ex : HLA-A2, HLA-DR4). Cela correspond à un typage dit de basse résolution [82]. Les techniques de biologie moléculaire ont permis un typage de résolution dite intermédiaire (ex : HLA-A\*02 :CNV = HLA-A\*02 :01/05/10, HLA-DRB1\*04 :AF = HLA-DRB1\*04 :01/09) qui ne permet pas de définir un allèle HLA unique mais un groupe d'allèles dont la séquence est proche, et parmi lesquels on retrouve un allèle majoritairement fréquent. Ce groupe d'allèles correspond à une liste d'ambiguïtés et sont regroupés sous une combinaison de lettres appelée code NMDP (National Marrow Donor Program) [82, 83]. Enfin, grâce au séquençage, la nomenclature allélique permet de regrouper l'ensemble des informations nécessaires permettant de caractériser chacun des allèles au niveau de la protéine synthétisée. On parle de typage à haute résolution (exemple : HLA-A\*14 :01), qui est le plus utilisé aujourd'hui en routine. Cette nomenclature identifie le gène, suivi d'un astérisque et d'un numéro à 1 champ (2 chiffres, information sur le groupe allélique) ou 2 à 4 champs (4 à 8 chiffres, information sur la protéine) [83, 84].

Le système HLA représente l'un des déterminants du rejet de greffe [77, 85, 86, 87]. Il est donc essentiel lors d'une transplantation rénale de rechercher la meilleure compatibilité possible entre un receveur et son donneur afin de limiter au maximum le risque de rejet du greffon [55, 56, 60]. En amont de la transplantation, une recherche des anticorps anti-HLA permet de mettre en évidence la présence d'éventuels anticorps dans le sérum du receveur qui seraient dirigés contre les antigènes HLA du donneur (Donor Specific Antibody ; DSA) [56, 88]. Ces anticorps sont la conséquence d'immunisations induites, entre autres, par des grossesses, des transfusions ou des transplantations antérieures [89]. Ce test appelé cross-match permet de prévenir les rejets hyperaigus qui surviennent dans les heures suivant l'acte chirurgical et qui sont délétères pour le greffon [56, 90]. La réaction de rejet contre le rein greffé est de mieux en mieux maîtrisée grâce aux traitements immunosuppresseurs [55] et le rejet aigu a drastiquement diminué ces dernières

années même en cas d'incompatibilité HLA (moins de 10% des patients ont subi un rejet aigu de greffe rénale en France en 2020) [44]. Il est important de noter que malgré la prise en compte de la compatibilité HLA entre donneurs et receveurs, les résultats de la transplantation peuvent différer d'un couple à un autre et ceci même avec un nombre d'incompatibilités HLA identiques pour les couples de donneur/receveur [91]. Classiquement, la compatibilité HLA est mesurée au niveau allélique c'est à dire en évaluant le nombre d'antigènes communs entre receveurs et donneurs au niveau des allèles. Cependant, ce système d'attribution basé sur les allèles HLA présente plusieurs limites. Parmi celles-ci, les individus présentant des génotypes HLA rares ou déjà immunisés contre un ou plusieurs allèles HLA ont des chances très réduites de bénéficier d'un donneur sur la base de la compatibilité allélique. De plus, seuls les anticorps anti-HLA pré-existants sont pris en compte pour l'attribution du greffon c'est à dire qu'il y'aura toujours un risque de développer des anticorps post-transplantation (DSA de novo). Des progrès récents ont été réalisés pour explorer la reconnaissance anticorps-antigène en étudiant les épitopes [92]. Un épitope fait référence aux acides aminés cibles spécifiques auxquels un paratope se lie (site de reconnaissance). Un épitope se compose d'une ou plusieurs entité fonctionnelle nommée épéplet qui se composent d'un ou plusieurs acides aminés polymorphes que l'on retrouve à la surface de la molécule [93]. La compatibilité entre donneur et receveur au niveau épétopique devient un sujet d'actualité dans le domaine de la transplantation, car les anticorps anti-HLA sont spécifiques des épitopes plutôt que des allèles. Entre 2 molécules HLA, les épitopes peuvent être totalement identiques (épitopes publiques), partiellement identiques (partageant certains épéplets) ou totalement différents (épitopes privés) [94]. Certains de ces épitopes sont plus immunogènes que les autres et n'auront pas le même impact sur la reconnaissance immunitaire. Il existe des épitopes cryptiques qui ne sont pas détectés par le système immunitaire et peu immunogènes. Ainsi, l'apparition d'une immunisation est due à une sensibilisation face à ces épitopes HLA. La compatibilité épétopique pourrait donc être plus précise pour évaluer la compatibilité donneur receveur et ainsi améliorer l'allocation des greffons [94]. En transplantation rénale, des études ont montré que l'anticipation de la charge épétopique permet d'identifier les patients à risque d'allosensibilisation et d'ajuster les niveaux cibles des médicaments immunosuppresseurs [95]. Plusieurs outils permettent aujourd'hui d'évaluer la compatibilité épétopique dont les plus utilisés sont HLAMatchmaker [96], HLA-EMMA [97], PIRCHE [98] etc. Plus récemment, un outil nommé HLA-EPI a été développé au sein de notre équipe et se base sur la reconnaissance directe des épéplets pour l'étude de la compatibilité HLA [99]. Toutefois l'identification des facteurs immunologiques ne suffit pas à comprendre toute la complexité des interactions donneur/receveur après la transplantation surtout dans le cas des rejets chroniques dont les mécanismes restent encore flous. Bien que les DSA soient reconnus comme la principale cause de la perte tardive de la greffe, il existerait d'autres facteurs impliqués. Des études ont montré l'implication de certains facteurs tels que les cellules NK ("Natural Killer"), mais aussi d'auto-anticorps et d'allo-anticorps non-

HLA dans le rejet chronique (cf. 1.5.3.2.1 Les rejets : Le rejet chronique p. 46) [100, 101, 102]. Ceci est bien illustré par le fait qu'il n'existe que 15% de différence de survie entre les transplantations 100% compatibles et celles incompatibles à 10 ans post-transplantation [91]. Ces avancées lèvent le voile sur la complexité des mécanismes régissant le rejet chronique et suggèrent la nécessité d'études avec des données intégratives (cliniques, biologiques, immunologiques et moléculaires) et multi-couches (génomiques, transcriptomiques, cliniques, proteomiques etc.) pour pouvoir améliorer le diagnostic et la prédiction de rejet ou perte de greffon chronique. C'est le défi que s'est lancé le projet KTD-innov (<https://www.ktdinnov.fr>) qui intègre à grande échelle l'évaluation systématique des données cliniques, biologiques, immunologiques et moléculaires des patients afin d'établir un système de prédiction du rejet fournissant une probabilité des risques individuels de rejet d'allogreffe ainsi qu'un système diagnostique de précision adressant l'activité et le stade du rejet de greffon [103, 102, 104].

#### 1.5.1.4 Attribution de l'organe

Le don d'organe est régi par des règles et des principes établis selon les lois de la bioéthique et gérée par l'agence de la biomédecine. Le don doit être consenti, gratuit et garantir l'anonymat entre le donneur et le receveur. Suivant ces principes, l'attribution d'organes se base sur deux notions fondamentale : l'utilité et l'équité du don. Cela veut dire qu'il faut tenir compte d'une part de l'urgence de la greffe ainsi que du temps d'attente pour les patients et d'autre part de l'utilisation optimale des greffons et des chances de survie du patient.

Des échelons de priorités ont été définies pour pouvoir attribuer le plus justement les greffons aux patients préalablement inscrits en liste d'attente [56].

- Un premier échelon de priorités nationales s'applique à tous les greffons et pour tous les patients. Il priorise les patients en urgence vitale dont la survie est menacée à court terme (hépatite sévère, patient sous cœur artificiel etc.), les patients ayant une difficulté d'accès particulière à la greffe (les hyperimmunisés par exemple), les patients mineurs si le donneur a moins de 18 ans et les patients 100% compatibles en HLA avec le donneur.
- Un deuxième échelon de priorités interrégionales est défini afin de favoriser les échanges de greffons entre les équipes dans l'inter région. L'attribution régionale à un patient est définie selon le score régional.
- Un dernier échelon de priorités s'applique au niveau local c'est-à-dire à l'échelle d'une équipe qui va définir les critères de sélection du receveur [56]. Lorsque deux reins sont prélevés, l'équipe de prélèvement bénéficie d'une priorité sur un des reins pour un de ses patients, et le second est proposé à l'échelon national. Dans les deux cas, l'attribution va être basée sur un score calculé pour chaque patient qui prend en compte l'ensemble des critères liés au donneur, au receveur et à l'appariement donneur/receveur. Ce score permet de classer les malades par ordre de priorité pour la greffe selon les réglementations

de l'agence de la biomédecine. Une fois le receveur attribué et le donneur prélevé, la greffe rénale peut commencer [56].

### 1.5.2 Déroulement de la greffe

L'opération de greffe rénale, qu'il s'agisse d'une greffe à partir d'un donneur décédé ou d'un donneur vivant, dure environ deux à trois heures. Lorsqu'il s'agit d'une transplantation avec un donneur décédé, l'opération ne concerne que le receveur. Le greffon est préalablement disponible et conservé à 4°C (ischémie froide) tout au long de la procédure. Cette ischémie froide qui correspond au temps pendant lequel le rein n'a plus de circulation sanguine est primordial à la réussite de la transplantation et doit être la plus courte possible. Dans le cas du donneur vivant, le rein est prélevé immédiatement avant la transplantation, ainsi il ne subit pas d'ischémie froide mais uniquement une ischémie tiède (température >4°C). D'une manière générale, la transplantation rénale va se subdiviser en 2 étapes [55, 60] :

— **Prélèvement de l'organe chez le donneur :**

L'opération est généralement effectuée en utilisant une technique minimale invasive appelée laparoscopie. Cette technique permet, surtout pour les donneurs vivants, un rétablissement plus rapide après l'intervention et une diminution des morbidités post-transplantation. Le prélèvement rénal se fait en « déconnectant » le rein de toutes les structures anatomiques qui y sont liés. Le chirurgien identifie l'ensemble des vaisseaux afférents et efférents de l'organe dont l'artère rénale, la veine rénale qui débouche dans la veine cave inférieure et l'uretère qui s'anastomose avec la vessie pour y acheminer l'urine. Ils sont ensuite clampés afin d'empêcher la perte sanguine lors de leur dissection. Le rein est ensuite retiré et placé dans une glacière contenant une solution de préservation froide (liquide physiologique). Dans le cas d'un donneur vivant, le greffon est transporté directement dans la salle d'opération du receveur. A l'issue de l'intervention, le donneur vivant est hospitalisé entre 3 à 7 jours en fonction de son évolution. Un suivi post-transplantation est réalisé au bout d'un mois puis une fois par an pour contrôler au mieux la fonction rénale du donneur [55, 56, 105].

— **Insertion de l'organe**

Le rein est inséré au niveau de l'abdomen dans la fosse iliaque sans nécessité d'une néphrectomie du/des reins malades (sauf pour les cancers et autres cas particuliers (ex. PKD sévère et douloureuse). L'insertion rénale implique une mise en communication des vaisseaux du pédicule rénal du donneur avec ceux du receveur (Figure 1.5). A la fin de l'opération, le patient est hospitalisé entre 7 à 14 jours avec un contrôle strict de la diurèse, de la fonction du greffon et de l'évolution de la cicatrice jusqu'à la reprise normale de la fonction rénale. Il est possible qu'une ou deux séances de dialyse soient nécessaires avant le retour d'un fonctionnement optimal du nouveau rein. Le patient



devra ensuite prendre un traitement immunosuppresseur le reste de sa vie pour prévenir les épisodes de rejets du greffon. Un accompagnement thérapeutique est néanmoins mis en place pour informer et accompagner le patient afin qu'il puisse se familiariser à sa nouvelle condition notamment pour le traitement médicamenteux à long terme. Un suivi périodique est également nécessaire pour pouvoir contrôler la fonction et l'évolution rénale du patient [55, 56]

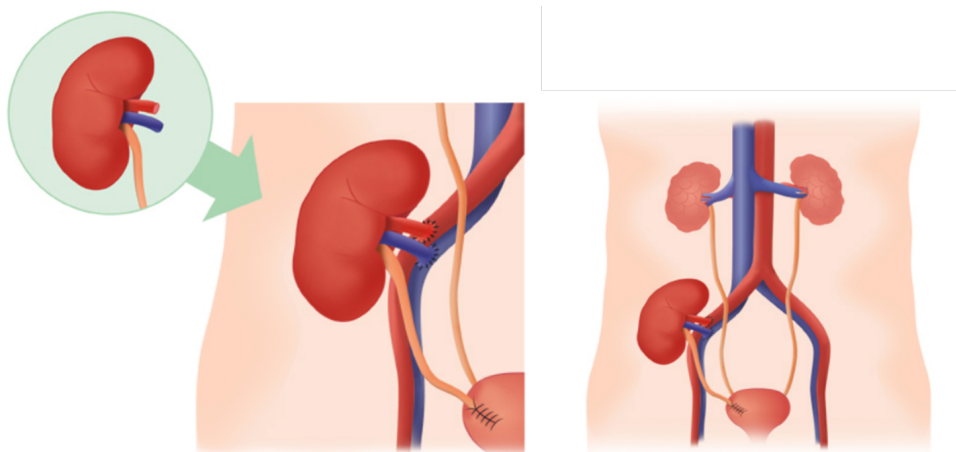


FIGURE 1.5 – *Implantation d'un nouveau greffon chez le receveur [106]. Le rein est placé dans la fosse iliaque sans forcément la nécessité d'une néphrectomie du/des reins malades. Le chirurgien connecte d'abord l'artère et la veine du rein greffé à la veine et à l'artère iliaques externes du receveur. Ensuite l'uretère du greffon est connecté à la vessie du receveur.*

### 1.5.3 Post-transplantation

#### 1.5.3.1 Suivi du patient

Après la transplantation, un suivi médical régulier est mis en place entre l'équipe de transplantation et le néphrologue afin de surveiller la fonction rénale du patient et d'anticiper toutes complications potentielles qui pourraient altérer la survie du greffon et du patient à court, moyen et long termes [107]. Au cours des trois premiers mois post-transplantation, le suivi est étroit. Les visites médicales et les examens sont très fréquents : deux fois par semaine le premier mois, puis une fois par semaine, puis deux fois par mois. Ensuite, une consultation est recommandée une fois toutes les 2 semaines de 4 à 6 mois, une fois par mois de 7 à 12 mois puis tous les 3 à 4 mois après la première année post-transplantation. Le suivi habituel comporte des analyses de sang et d'urine, un examen clinique, la prise de la tension artérielle [107]. Ces consultations ont pour objectifs de surveiller :

- La prise du traitement immunosuppresseur qui cible l'activation et la prolifération des lymphocytes T et B afin de prévenir et de traiter les rejets. En l'absence de traitement immunosuppresseur, le rejet d'un organe allogénique est inéluctable. Toutefois un suivi pharmacologique est nécessaire pour déterminer la dose adéquate pour le patient afin d'éviter les effets indésirables (complications toxiques, infectieuses et néoplasiques liées un excès d'immunosuppression) [108].
- La fonction du rein greffé avec une évaluation du DFG et de la protéinurie qui sont des facteurs très importants pour le suivi car ils permettent d'estimer l'efficacité ou la tolérance du traitement en cours. Un DFG compris entre 40 et 70 mL/min/1.73m<sup>2</sup> est considéré comme correct. Si le DFG est inférieur à 10 mL/min/1.73 m<sup>2</sup> et que des complications sur l'organe sont observées, le greffon n'est plus considéré comme fonctionnel [56, 107].
- Les risques cardio-vasculaires et les complications carcinologiques avec un contrôle de la pression artérielle et du régime alimentaire pour les complications cardiovasculaires et un dépistage précoce des cancers notamment ceux cutanés et urologiques [56, 107]. Lorsque ces facteurs ne sont pas pris en charge ou ne sont pas détectés à temps, la fonction rénale du patient peut se dégrader rapidement et conduire à la perte du greffon [56].

### 1.5.3.2 Complications de la transplantation

De nombreuses complications peuvent survenir après la transplantation comme des récives de la néphropathie initiale, des complications infectieuses, cardiovasculaires et métaboliques, digestives, néoplasiques, osseuses ou encore immunologiques. En cas de suspicion d'échec de la transplantation rénale (baisse de la fonction rénale), la biopsie du greffon reste l'analyse primordiale pour le diagnostic. Elle permet de distinguer quelles sont les causes de cet échec : rejet humoral/cellulaire aigu ou chronique, toxicité médicamenteuse, récive de la maladie initiale, glomérulonéphrite de novo, infection virale, etc [56]. Dès 1991, a été mise en place la classification de Banff [109] permettant une classification standardisée et internationale des biopsies des greffons rénaux. Elle oriente le diagnostic clinique, évalue la gravité des lésions (aigu ou chronique) et uniformise les études de recherche et les essais cliniques portant sur le diagnostic, le traitement et les résultats des greffes de rein. La classification repose sur l'analyse des lésions histologiques de l'interstitium, des tubules, des glomérules et des vaisseaux pour déterminer un score semi-quantitatif [109]. Cette classification est mise à jour 2 fois par an pour intégrer les données scientifiques les plus récentes et tenter de différencier les lésions associées au rejet chronique. On distingue 6 catégories selon les critères de Banff : (1) rein normal ou changements non spécifiques, (2) modifications médiées par les anticorps, (3) modifications "borderline" : suspicion de rejet médié par les LT aiguë, (4) rejet médié par les LT, (5) fibrose interstitielle et atrophie tubulaire, (6) autres (néphropathie à BK virus, troubles lymphoprolifératifs post-transplantation, maladie récurrente, pyélonéphrite, etc.) [109]. Dans cette partie nous allons principalement nous

intéresser à deux types de complications : les rejets et la récurrence de la maladie initiale.

### 1.5.3.2.1 Les rejets

Les premières causes de complications immunologiques post-transplantation sont caractérisées par les rejets. On distingue 2 catégories de rejets [110] :

- **Le rejet cellulaire** dont les acteurs principaux sont les LT effecteurs, est dû au développement d'une immunisation du receveur contre les antigènes HLA du donneur : on parle d'allo-reconnaissance. Elle est caractérisée par la capacité des LT à être stimulés par des molécules allo-antigènes (antigènes génétiquement différents) capables d'induire l'activation de LT allo-spécifiques et de déclencher une réponse immunitaire. Dans le cas de la transplantation rénale, cette allo-reconnaissance peut être directe ou indirecte. Elle est directe lorsqu'elle fait intervenir les CPA du donneur, apportées avec le greffon, dont les complexes CMH/peptides sont directement reconnus par les LT du receveur. L'allo-reconnaissance indirecte quant à elle fait intervenir les CPA du receveur qui vont phagocyter des peptides allo-géniques issus du greffon et présenter à leur surface un complexe CMH/peptides qui va être reconnu par les LT et entraîner leur activation. Afin de limiter ces réactions, des traitements immunosuppresseurs sont prescrits aux patients [111, 112, 113].
- **Le rejet humoral** dont les acteurs principaux sont les anticorps spécifiques anti-HLA du donneur, ou DSA est dû au développement d'immunisations induites, entre autres, par des grossesses, des transfusions ou des transplantations antérieures. On distingue 2 types de DSA : les DSA préformés (antérieur à la transplantation) impliqués dans la réponse humorale des rejets hyper-aigus, et les DSA de novo (suite à la transplantation) associés au rejet aigu tardif médié par les anticorps, au rejet chronique médié par les anticorps et à la glomérulopathie de transplantation. De manière générale, les DSA sont responsables de rejets médiés par les anticorps (AMR : Antibody Mediated Rejection). Les anticorps vont cibler les cellules endothéliales et, via une toxicité directe et une toxicité indirecte médiée par le complément, ils vont former le complexe d'attaque membranaire, attirer des cellules inflammatoires et activer la phagocytose. Ces dommages vont contribuer à la thrombose et à l'occlusion vasculaire. Il est possible d'éviter le développement des DSA en limitant les incompatibilités HLA entre donneur et receveur et en évitant les transplantations avec un donneur dont les HLA sont la cible des anticorps circulants chez le receveur (test du cross match primordial en amont de la transplantation) [114, 115].

Au sein de ces 2 catégories de rejets, 3 types de rejets peuvent être caractérisés en fonction du délai d'apparition après la transplantation et des types de lésions sur le greffon : le rejet hyper-aigu, le rejet aigu et le rejet chronique [55, 56, 110]

- **Le rejet hyper-aigu**

Le rejet hyper-aigu est un rejet à médiation principalement humorale, déclenché au moment du rétablissement de la circulation sanguine entre le greffon et le receveur. Il peut se produire dans les minutes suivant cette reconnexion sanguine (rejet hyper-aigu immédiat) ou dans la première semaine post-transplantation (rejet hyper-aigu retardé). Le mécanisme du rejet hyper-aigu est lié à la présence, chez le receveur, d'anticorps anti-HLA ou anti-ABO présents sur les cellules endothéliales du greffon. Si les anticorps anti-ABO ou les DSA circulant chez le receveur sont spécifiques des antigènes du groupe sanguin ABO ou du HLA du donneur, alors ils vont former un complexe antigène/anticorps à la surface des cellules endothéliales du greffon et entraîner une fixation et une activation du complément. Dès lors, le rejet hyper-aigu se traduit par la thrombose des vaisseaux irriguant le greffon et par un arrêt brutal et définitif de sa fonction. Toutefois, la survenue de ce type de rejet est devenue exceptionnelle depuis la pratique systématique du test de cross-match [88, 115].

— **Le rejet aigu**

On distingue deux sous-types de rejet aigu : (1) le rejet aigu humoral, et (2) le rejet aigu cellulaire. Le rejet aigu humoral se produit typiquement de manière précoce, dès le 4ème jour post-transplantation ; il peut toutefois survenir plus tardivement. Le rejet aigu cellulaire se produit généralement dans les premiers mois suivant la transplantation. Il constitue le type de rejet le plus fréquemment observé (80% des épisodes de rejet aigu). Sa survenue est favorisée par une mauvaise observance du traitement immunosuppresseur. Il existe deux voies du rejet aigu cellulaire, la reconnaissance de l'allogreffe par les LT pouvant se faire directement par reconnaissance des complexes CMH-peptides du greffon ou indirectement par l'apprêtement des antigènes du greffon sur le CMH des cellules du receveur [56, 110].

— **Le rejet chronique**

Le rejet chronique se définit par une dysfonction progressive du greffon, en l'absence de causes mécaniques ou infectieuses. Il constitue la principale cause de perte des greffons rénaux, en particulier à long terme (56% à 10 ans post-transplantation) [44, 116]. D'un point de vue clinique, ce rejet se manifeste par une protéinurie et une dysfonction du greffon. Le rejet chronique est le résultat de lésions cumulées (effet secondaire de l'ischémie/reperfusion, rejets aigus, infiltration du greffon de LT allo-réactifs, inflammation chronique) et impliquent également des mécanismes immuns (immunisation vis-à-vis des allo-antigènes, en pré- et post-transplantation : DSA) et nonimmuns, aggravés par la toxicité des traitements [56]. Malgré ces facteurs de risques connus, les facteurs immunologiques impliqués dans les mécanismes de rejet de greffes chroniques restent encore flous. Les DSA sont reconnus comme la principale cause de la perte tardive de la greffe [117]. Mais récemment, des études ont montré l'implication d'autres facteurs. En effet, une étude

avait déjà mis en évidence l'implication des cellules NK dans le rejet chronique. L'absence d'activation des cellules NK auto-induites favorise le développement d'une inflammation vasculaire du greffon qui a exactement le même impact négatif sur la survie de l'organe que l'absence d'activation de l'anti-DSA [100]. Plus récemment, d'autres études ont montré la grande diversité des mécanismes qui sous-tendent le rejet chronique notamment avec l'implication d'auto-anticorps et d'allo-anticorps non-HLA [101]. Les allo-anticorps sont dirigés contre des antigènes polymorphes qui diffèrent entre le donneur et le receveur tandis que les auto-anticorps sont des anticorps qui reconnaissent des auto-antigènes. Plusieurs gènes ont déjà été identifiés tels que *AT1R* qui code pour un récepteur de l'angiotensine II, *ICAM4* qui code pour une molécule d'adhésion intracellulaire, la protéine kinase Cz etc. Ces avancées lèvent le voile sur la complexité des mécanismes régissant le rejet chronique et suggèrent que des études avec des données multi-couches (génomiques, transcriptomiques, cliniques, proteomiques etc.) sont nécessaires pour mieux comprendre l'évolution du greffon [118].

#### 1.5.3.2.2 Récidive de la maladie

En France, les glomérulopathies primitives sont la 3<sup>ème</sup> cause d'IRCT (11%) après les néphropathies diabétiques (22%) et celles vasculaires et hypertensives (25%) [43, 44]. Certaines de ces maladies ayant entraîné l'insuffisance rénale chronique peuvent récidiver sur le rein greffé. Il s'agit essentiellement de glomérulonéphrites [56]. La récurrence de la maladie initiale est responsable de moins de 4 % de l'ensemble des échecs de transplantation. Le risque de récurrence est majeur en cas de récurrence sur une première greffe et lorsque la maladie initiale a rapidement évolué vers l'insuffisance rénale [119, 120]. Le diagnostic peut être difficile à faire car la plupart des récurrences n'ont pas de signes cliniques ou biologiques visibles. La biopsie est la meilleure option même si la différenciation entre glomérulopathies de novo, glomérulopathies transmises par le greffon, et récurrences de glomérulopathies n'est pas toujours évidente [56]. Il est possible de prescrire des traitements dans certains cas de récurrences, mais la marge de manœuvre est très limitée conduisant surtout à poser des indications très prudentes dans le choix du donneur quand une transplantation de donneur vivant est proposée. [56].

— **Le syndrome néphrotique idiopathique (SNI) avec lésions de hyalinose segmentaires et focales (HSF)**

La HSF est décrite comme un ensemble de dépôts hyalins et de lésions de scléroses segmentaires (touchant une partie du glomérule) et focales (n'atteignant que certains glomérules). Ces lésions altèrent la barrière de filtration glomérulaire, qui devient perméable aux éléments de haut poids moléculaire tels que l'albumine [121]. La HSF se décline généralement en 2 formes : la forme idiopathique (80% des cas) et la forme secondaire (20% des cas). Un des problèmes majeurs du SNI avec HSF est la récurrence sur le greffon, de 30

à 50% des cas lors d'une première greffe, à plus de 90% lors de la deuxième [122]. Cette récurrence est généralement précoce et peut intervenir dans les 6 à 12 premiers mois post-transplantation. Elle se caractérise par une protéinurie abondante (3g/24h) et une hypertension artérielle avant que n'apparaissent les lésions histologiques de hyalinose [56, 122]. Des facteurs de risque tels que l'âge du donneur ont déjà été identifiés mais l'origine de ces lésions reste indéterminée [123]. La présence d'un facteur de perméabilité circulant est l'hypothèse privilégiée. En effet, une étude a identifié le récepteur soluble de l'activateur du plasminogène de l'urokinase (*suPAR*) dans le sérum comme pouvant être le facteur circulant responsable de la récurrence HSF [124]. Plus récemment, Gallon et al. ont montré que la retransplantation chez un receveur n'ayant pas la HSF comme maladie initiale permettait la récupération rapide de la fonction d'un greffon qui était initialement dysfonctionnel chez un receveur avec la HSF comme maladie initiale [125]. Ces résultats appuient la théorie selon laquelle un facteur circulant est à l'origine de la HSF. Toutefois, l'étude sur *suPAR* n'a jamais été validée et une compréhension complète de la cascade cellulaire impliquée est encore nécessaire [126].



# LA TRANSPLANTATION RÉNALE À L'ÈRE DES «OMIQUES»

---

## 2.1 Les «omiques» en transplantation

### 2.1.1 Définition des omiques

L'essor des nouvelles techniques de séquençage qui permettent aujourd'hui de pouvoir séquencer un génome humain en une semaine a fait croître de façon considérable le volume de données à stocker et à analyser [127]. Dans le domaine biomédical, ces données générées dites « omiques » permettent la mesure et l'analyse à grande échelle de profils génétiques ou moléculaires à l'échelle d'un tissu ou d'une cellule. Au lieu d'étudier un seul gène ou une seule protéine, les approches omiques permettent une vision systémique de l'entité biologique (génome, protéome, transcriptome etc.). Par exemple, la génomique établit le profil de l'ADN (acide désoxyribonucléique), la transcriptomique mesure les transcrits; la protéomique et la métabolomique quantifient respectivement les protéines et les métabolites etc [128].

#### 2.1.1.1 La génomique

La génomique correspond à l'étude de l'ensemble du matériel génétique (appelé génome) d'une cellule ou d'un tissu. Chaque cellule de l'organisme contient une copie complète des 3 milliards de paires de bases d'ADN, qui constituent le génome humain. Au début des années 2000, le séquençage du premier génome humain a révolutionné la génomique en permettant une description détaillée de la composition du génome entier (cf. 2.2.1 Le génome humain p. 70) [129, 130]. Plus tard, l'ambition était de pouvoir séquencer un grand nombre de génomes afin d'étudier la diversité interindividuelle et d'explorer des variations et/ou des mutations génétiques susceptibles de jouer un rôle dans le développement ou la progression d'une maladie. En effet, la divergence entre deux individus est estimée à environ 0,1 % (environ 3 à 4 millions de paires de bases) et est responsable de la majeure partie de la diversité phénotypique observée au sein de la population humaine [131]. Dans ce contexte, des projets tels que 1000 Génomes (1KGP) ont vu le jour et ont permis de cartographier les variants génétiques humaines ayant une fréquence allélique (« Minor Allelic Frequency » - MAF) d'au moins 1 % au sein de 5 grandes populations (Africains,



Américains, Est et Sud Asiatiques, Européens) [132]. Cette cartographie a permis d'identifier des polymorphismes tels que les SNP (Single Nucleotide Polymorphism) qui sont au cœur des études d'associations pangénomiques (Genome-Wide Association Studies - GWAS) [133]. Ces approches permettent d'établir, par des méthodes statistiques, une association entre un SNP et un trait (ou une maladie) à l'échelle du génome. Ces études sont basées sur les technologies de génotypage qui permettent de déterminer la présence ou l'absence de SNP dans la séquence d'ADN d'un individu. Le génotypage s'effectue majoritairement à l'aide de puces à ADNs (ou « microarray ») qui consistent en de fines couches de verre dans lesquels sont disposés des oligonucléotides. Ces derniers également appelés "sondes" correspondent à des séquences d'ADN connus et permettent de détecter les variants d'intérêt pour chaque individu. Toutefois, il est également possible de génotyper des individus grâce à des approches de séquençages nouvelles générations (« Next Generation Sequencing » - NGS) qui permettent de déterminer la séquence exacte de l'ADN par opposition au génotypage qui ne cible que des séquences pré-définies. Ces techniques sont néanmoins plus coûteuses mais permettent une exploration plus complète du génome. De plus, si les puces à ADNs ne se concentrent généralement que sur les variants communs ( $MAF > 5\%$ ) du génome. Le séquençage permet d'explorer les variations rares ( $MAF < 0.5\%$ ) et communes sur l'ensemble du génome (séquençage du génome entier ou « Whole Genome Sequencing » - WGS) ou spécifiquement sur les régions codantes contenant les gènes (séquençage de l'exome entier ou « Whole Exome Sequencing » - WES). La génomique permet de lever le voile sur les relations entre génotype et phénotype. Dans le domaine de la transplantation, les études génomiques permettent de mettre en évidence la variabilité génétique inter et intra individuelle et interpopulationnelle entre donneurs et receveurs qui peuvent impacter le devenir du greffon (cf. 2.2.3 Les études d'association génétiques p. 84) [134].

### 2.1.1.2 La transcriptomique

La transcriptomique représente l'étude du transcriptome, défini comme l'ensemble des séquences d'acides ribonucléiques (ARN) produits lors de la transcription par un génome au sein d'un tissu ou d'une cellule [135]. Initialement, le transcriptome ne concernait que les ARN dites messagers (ARNm), codant pour les protéines. Mais cette définition s'est élargie et englobe dorénavant l'ensemble des ARN. Le transcriptome est constitué d'ARNm codants qui représentent environ 1 à 4% des transcrits et d'ARN non codants qui représentent, avec les régions intergéniques, le reste du génome et ne donnent pas lieu à des protéines. Les premières études transcriptomiques sont basées sur la technologie des puces à ADN. Au cours des années 2010, les puces à ADN ont laissé place au RNA-seq, une technique de séquençage nouvelle génération qui permet de capturer l'ensemble des séquences présents dans un échantillon donné qu'ils soient connus ou non. Cette technologie ne nécessite qu'une petite quantité d'ARN et aucune connaissance préalable du transcriptome. Une autre technique qui dérive directement

du RNA-seq permet d'étudier la transcription à l'échelle d'une seule cellule (ou « single-cell transcriptomics »). Le séquençage de l'ARN d'une cellule unique (scRNA-seq) permet en plus de l'étude d'une cellule individuelle, d'analyser des sous-populations de types de cellules pour un tissu d'intérêt. Les études RNA-seq permettent d'évaluer à la fois quantitativement et qualitativement le taux d'ARN différentiel pour mieux comprendre certains mécanismes phénotypiques tels que les maladies et favoriser la découverte de nouveaux transcrits. Elles permettent également d'approfondir l'étude des ARN non codants qui jusque-là étaient peu connus mais ont été prouvés comme essentiels dans de nombreux processus biologiques et maladies. De plus, grâce au développement des techniques scRNA-seq, il est possible de caractériser des populations cellulaires rares telles que les cellules tumorales circulantes, les cellules souches cancéreuses, et les cellules souches embryonnaires dans les blastocystes de mammifères [134].

### 2.1.1.3 La protéomique

La protéomique est l'étude à grande échelle des protéines, de leur structure, de leur interaction et de leur potentiel rôle au sein d'un organisme. Elle s'intéresse au protéome qui représente l'ensemble des protéines produites par une cellule, un tissu ou un organisme donné. Ce nombre de protéines varie en fonction du temps ou des conditions spécifiques (stress, maladies, variations environnementales etc.). Plusieurs techniques ont été développées pour analyser le protéome. Les plus couramment appliquées sont les techniques basées sur la spectrométrie de masse ou « Mass Spectromy » - MS) qui comprennent deux méthodes : la MS en tandem (ou « Tandem-MS ») basée sur l'utilisation de marqueurs à isotopes stables pour le marquage différentiel des protéines et les techniques basées sur la migration sur gel telle que l'électrophorèse différentielle en gel (ou « DIGE »). Cette dernière est la méthode la plus établie et la plus répandue. Elle utilise l'électrophorèse bidimensionnelle à haute résolution pour séparer les protéines de différents échantillons en parallèle. Ces protéines sont ensuite sélectionnées puis colorées de manière distincte lors de l'identification par spectrométrie de masse. Les méthodes MS étudient l'interaction des protéines et la quantification des modifications post-traductionnelles. Cependant, elles ne sont pas adaptées pour l'analyse de centaines ou de milliers d'échantillons dans un délai court (quelques jours ou quelques semaines). A cette échelle d'échantillons, les méthodes de purification par affinité sont plus adéquates. Elles utilisent des anticorps ou d'autres réactifs d'affinité (tels que les aptamères à base d'oligonucléotides) comme sondes pour la détection spécifique des protéines. Une troisième approche basée sur les puces à protéines (protein chips ou microarrays) est également utilisée pour l'identification des protéines spécialement pour la détection de la présence ou non de certaines protéines au sein d'un échantillon biologique (détection de l'affinité antigène-anticorps dans le sang par exemple) ou pour l'identification de complexes moléculaires (protéine-ADN, protéine-protéine et protéine-ligand). Les approches protéomiques permettent une analyse comparative et quantitative du répertoire de protéines au sein des tissus ou des

fluides biologiques dont le taux différentiel peut refléter une maladie sous-jacente. Ces études favorisent l'identification de nouveaux biomarqueurs potentiels pour le traitement des maladies. Par exemple, si une certaine protéine est impliquée dans une maladie, sa structure 3D fournit les informations nécessaires pour concevoir des médicaments pouvant interférer avec l'action de la protéine. En transplantation, cette approche permet, par exemple, de détecter la présence de protéines dans l'urine associée au rejet ou d'adapter un traitement immunosuppresseur en fonction des protéines produites [134].

#### **2.1.1.4 La métabolomique**

La métabolomique est l'étude à grande échelle des métabolites synthétisés dans les cellules, les biofluides, les tissus ou les organismes. Un métabolite est un composé organique de faible poids moléculaire (50 à 1500 daltons), généralement impliqué dans un processus biologique en tant que substrat ou produit (acides aminés, lipides, sucres, nucléotides, alcaloïdes, etc.). L'ensemble de ces petites molécules et leurs interactions au sein d'un système biologique constituent le métabolome et sont influencées par des facteurs génétiques et environnementaux. Le métabolome est par nature très dynamique : les molécules sont continuellement absorbées, synthétisées, dégradées et interagissent avec d'autres molécules, tant au sein des systèmes biologiques qu'entre eux, et avec l'environnement. Par conséquent, le métabolome reflète directement l'activité biochimique sous-jacente et l'état des cellules/tissus. Les technologies utilisées dans le domaine de la métabolomique sont la MS et la Résonance Magnétique Nucléaire (RMN). Ces techniques quantifient l'abondance et la structure des métabolites en se basant respectivement sur leur masse et sur les propriétés magnétiques des noyaux atomiques. Les métabolites sont d'abord séparés par chromatographie. Il existe deux méthodes : la chromatographie en phase gazeuse (GC) utilisée pour les petites molécules volatiles et la chromatographie liquide à haute performance (HPLC) qui sépare les molécules en phase liquide. Ces méthodes sont couplées avec la MS et/ou la RMN pour la détection et l'identification des métabolites. La métabolomique a pour objectif de donner l'image la plus complète de l'état cellulaire en identifiant et quantifiant son empreinte moléculaire. Des niveaux différentiels peuvent refléter une fonction métabolique anormale et être le signe d'une maladie. En raison de son lien étroit avec le phénotype, la métabolomique est un outil idéal pour la découverte de biomarqueurs [134].

#### **2.1.1.5 L'épigénomique**

L'épigénomique étudie les modifications réversibles de l'ADN ou des protéines associées à l'ADN à l'échelle du génome, qui ont un impact sur l'expression des gènes sans induire de modifications dans la séquence d'ADN elle-même. Ces modifications, notamment la méthylation de l'ADN ou l'acétylation des histones, peuvent être influencées à la fois par des facteurs génétiques et environnementaux et influencer le schéma d'expression des gènes. Récemment, les

études d'association à l'échelle de l'épigénome (ou « Epigenome-Wide Association Studies » - EWAS) permettent d'évaluer les modifications de l'ADN à un niveau global en utilisant les technologies NGS mais également les puces à ADN. Il existe une grande variété de méthodes d'analyse des modifications épigénomiques, dont le choix dépend d'un certain nombre de facteurs, notamment la qualité et le type d'échantillon de tissu disponible, le coût, les objectifs des études et les ressources informatiques/laboratoires disponibles. Ces méthodes comprennent la quantification de la méthylation globale par HPLC et l'analyse de la méthylation de l'ADN par séquençage au bisulfite du génome entier. L'analyse de la méthylation globale permet d'évaluer le statut de méthylation d'un échantillon de tissu, tandis que l'analyse de la méthylation de l'ADN par séquençage au bisulfite du génome entier donne un statut de méthylation spécifique à un site sur l'ensemble du génome. Cette approche met en évidence le rôle clé des signatures épigénétiques dans les processus biologiques, le développement des maladies et d'autres états physiopathologiques. Dans le domaine de la transplantation, une meilleure compréhension de ces modifications épigénétiques chez les donneurs et les receveurs peut conduire à l'identification de biomarqueurs [134].

### 2.1.2 Etat de l'art des études omiques en transplantation

#### Revue

#### **Surfing the Big Data Wave : omics data in transplantation**

*Rokhaya BA, MSc\* ; Estelle Geffard, PhD\* ; Venceslas Douillard, MSc ; Françoise Simon, MBA, PhD ; Laurent Mesnard, MD, PhD ; Nicolas Vince, PhD ; Pierre-Antoine Gourraud, PhD, MPH ; Sophie Limou, PhD*

Transplantation : February 2022 - Volume 106 - Issue 2 - p e114-e125 doi :  
10.1097/TP.0000000000003992 PMID : 34889882

#### 2.1.2.1 Résumé de la revue

Tant dans le domaine de la recherche que dans celui des soins, les patients, les soignants et les chercheurs sont confrontés à une augmentation considérable de la quantité de données disponibles pour l'analyse et l'interprétation, marquant ainsi « l'ère du big data ». Dans le domaine biomédical, ce changement quantitatif fait surtout référence à la -omique qui permet de mesurer et d'analyser les caractéristiques d'une entité biologique spécifique dans son ensemble. Les études omiques ont eu un impact considérable sur la recherche en transplantation. Certaines études ont souligné l'apport des omiques dans le développement de thérapies personnalisées pour éviter la perte de greffons. Cependant, l'intégration des données omiques reste un défi en termes de processus analytiques. Ces données proviennent de sources multiples. Par conséquent, elles peuvent contenir des biais et des erreurs systématiques qui peuvent être confondus avec des

informations biologiques pertinentes. Des méthodes de normalisation de correction des effets de batch ont été développés pour résoudre les problèmes liés à la qualité et à l'homogénéité des données. En parallèle, les méthodes d'imputation traitent les données manquantes. Le domaine de la transplantation représente un contexte analytique unique car l'unité statistique biologique est la paire donneur-receveur, ce qui rend les analyses omiques encore plus complexes. Des stratégies telles que les scores de risque combinés entre deux génomes qui prennent en compte de l'ancestralité génétique sont développées pour mieux comprendre les mécanismes de la greffe et affiner les interprétations biologiques. Les futures omiques seront basées sur une biologie intégrative, considérant l'analyse du système dans son ensemble et non plus l'étude d'une seule caractéristique. Cette revue résume les avancées des études omiques dans le domaine de la transplantation et présente les enjeux analytiques concernant ces approches.

*Mots-clefs : transplantation, omiques, “big data”, contrôle qualité, normalisation, effet de batch, imputation, paires donneur-receveur*

---

**Revue**

---

# Surfing the Big Data Wave: Omics Data Challenges in Transplantation

Rokhaya Ba, MSc,<sup>1,2</sup> Estelle Geffard, PhD,<sup>1</sup> Venceslas Douillard, MSc,<sup>1</sup> Françoise Simon, MBA, PhD,<sup>1,3</sup> Laurent Mesnard, MD, PhD,<sup>4,5</sup> Nicolas Vince, PhD,<sup>1</sup> Pierre-Antoine Gourraud, PhD, MPH,<sup>1</sup> and Sophie Limou, PhD<sup>1,2</sup>

**Abstract.** In both research and care, patients, caregivers, and researchers are facing a leap forward in the quantity of data that are available for analysis and interpretation, marking the daunting “big data era.” In the biomedical field, this quantitative shift refers mostly to the -omics that permit measuring and analyzing biological features of the same type as a whole. Omics studies have greatly impacted transplantation research and highlighted their potential to better understand transplant outcomes. Some studies have emphasized the contribution of omics in developing personalized therapies to avoid graft loss. However, integrating omics data remains challenging in terms of analytical processes. These data come from multiple sources. Consequently, they may contain biases and systematic errors that can be mistaken for relevant biological information. Normalization methods and batch effects have been developed to tackle issues related to data quality and homogeneity. In addition, imputation methods handle data missingness. Importantly, the transplantation field represents a unique analytical context as the biological statistical unit is the donor–recipient pair, which brings additional complexity to the omics analyses. Strategies such as combined risk scores between 2 genomes taking into account genetic ancestry are emerging to better understand graft mechanisms and refine biological interpretations. The future omics will be based on integrative biology, considering the analysis of the system as a whole and no longer the study of a single characteristic. In this review, we summarize omics studies advances in transplantation and address the most challenging analytical issues regarding these approaches.

(*Transplantation* 2022;106: e114–e125).

## INTRODUCTION

The definition of big data in the biomedical field is highly debated. In the early 2000s, a popular definition was

Received 30 April 2021. Revision received 23 September 2021.

Accepted 9 October 2021.

<sup>1</sup> Université de Nantes, Centre Hospitalier Universitaire Nantes, Institute of Health and Medical Research, Centre de Recherche en Transplantation et Immunologie, UMR 1064, Institut de Transplantation Urologie-Néphrologie, Nantes, France.

<sup>2</sup> Département Informatique et Mathématiques, Ecole Centrale de Nantes, Nantes, France.

<sup>3</sup> Mount Sinai School of Medicine, New York, NY.

<sup>4</sup> Urgences Néphrologiques et Transplantation Rénale, Hôpital Tenon, Assistance Publique-Hôpitaux de Paris, Paris, France.

<sup>5</sup> Sorbonne Université, Paris, France.

R.B. and E.G. are cofirst authors.

This work has been supported by the Government through the National Research Agency under the Future Investments program with the reference KTD-Innov (ANR-17-RHUS-0010).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 754995.

The authors declare no conflicts of interest.

R.B. and E.G. did the bibliographic research and cowrote the article. P.-A.G., S.L., and N.V. designed, supervised, and formalized the scientific content and writing of the review. V.D., L.M., and F.S. contributed to the formalization of the article. All authors have approved the final version of the article.

Correspondence: Pierre-Antoine Gourraud, PhD, MPH, Centre de Recherche en Transplantation et Immunologie UMR1064, Institut de Transplantation Urologie-Néphrologie, Centre Hospitalier Universitaire Nantes, Hôtel Dieu, 30 Blvd Jean Monnet, 44093 Nantes Cedex 01, France. (pierre-antoine.gourraud@univ-nantes.fr).

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0041-1337/20/1062-e114

DOI: 10.1097/TP.0000000000003992

founded on 3 Vs: “Big data is data that contains greater Variety, arriving in increasing Volumes and with ever-higher Velocity. This is known as the three Vs.”<sup>1</sup> Even if 17 other “Vs” have been proposed since,<sup>2</sup> big data is best described with 5 key concepts<sup>3,4</sup>: Volume represents the quantitative change that occurs with the processing of high volumes of data (2314 exabytes produced in 2020 in Healthcare<sup>5,6</sup>). Variety refers to the diversity of data types. Velocity indicates at which rate the data can be made available for analysis. Value refers to the usefulness of big data.<sup>7</sup> Veracity ensures the integrity of data that is meaningful to confidently use it for decision making.

In the biomedical field, Big data refers mostly to the -omics that permits measuring and analyzing large-scale genetic or molecular profiles of organisms or tissues. Instead of studying a single gene, protein, or metabolite, -omics approaches allow the exploration of a complete biological feature<sup>8</sup> (see Table 1). Omics data have empowered the community to investigate how complex interactions between genes and molecules may influence the phenotypes and improved medical decision making. In transplantation, some studies have emphasized the contribution of omics in developing personalized therapies to avoid graft loss while reducing comorbidities such as cancer and diabetes.<sup>15,16</sup> Additionally, the incorporation of longitudinal structured and unstructured data improves significantly graft loss and mortality prediction.<sup>17</sup> In terms of analysis, the challenge of omics data in transplantation is 2-fold:



**TABLE 1.**  
**Approaches used in omics**

Genomics	Genomics studies whole genomes of organisms that includes the complete set of DNA and genes. It is a multidisciplinary branch that focuses on the identification of genetic variants associated with disease, differential drugs effect, transplant outcome, etc. An approach called the GWAS <sup>9</sup> allows the characterization of genetic variations related to complex diseases in human populations. In the realm of transplantation, GWAS permits to highlight interethnic and interindividual genetic variability of donors and recipients (SNPs, CNVs, etc) that may impact graft function. <sup>10</sup> Associated technologies are mainly based on genotype arrays techniques and also on NGS for whole-genome sequencing and exome sequencing. Genomics studies enable a better understanding of complex phenotypes.
Transcriptomics	Transcriptomics represents the study of all transcripts (mRNA and ncRNA) produced during the process of transcription by a genome under specific circumstances or in a specific cell. It evaluates both quantitatively and qualitatively differential RNA rate to have better insight into disease mechanisms. There are several techniques used in transcriptomics field, such as quantitative PCR, DNA microarrays that estimate predetermined sequences, and RNA-Seq using high-throughput sequencing to capture all transcripts. RNA-Seq studies not only lead to the discovery of many novel isoforms of the protein-coding transcriptome, but also more importantly, these techniques empower the study of noncoding RNA that play essential roles in many biological processes, diseases, and potentially may be used as biomarkers or as therapeutic targets. <sup>10</sup> Transcriptomics analysis has enabled us to grasp the impact of gene expression depending on the organism and its consequences on human disease outcomes.
Proteomics	Proteomics attempts to identify in a global way proteins quantification, location, modification, interaction, and functions in genome scale. It is an important component of functional genomics and provides comparative and quantitative analysis of protein repertoire in tissue and bio fluids that reflect an underlying disease. <sup>10</sup> In transplantation, this approach allows, for example, to detect the presence of protein in the urine associated with rejection. Several techniques have been developed to analyze the proteome including MS methods that investigate protein interaction and quantification of posttranslational modifications and affinity purification methods used in the detection of protein interaction with other molecules. <sup>11</sup> Those approaches improved considerably in the proteomics field, especially with the development of NGS technologies.
Metabolomics	Metabolomics aims to provide a comprehensive quantification of all metabolites and small molecules in the metabolome under a given condition. The metabolic profile of a cell/tissue may result from both genetic and environmental factors. Differential levels reflect abnormal metabolic function and are often indicative of disease. Metabolomics has for objective to give the most complete picture of the cell state by identifying and quantifying metabolic imprints. Technologies used in metabolomics field are MS-based approaches, namely MS and NMR spectroscopy. Those techniques quantify small molecule abundances and structure based respectively on their mass and on the magnetic properties of atomic nuclei. <sup>12</sup>
Epigenomics	Epigenomics studies reversible changes of DNA or DNA-associated proteins on genome scale that impact gene expression without inducing mutational changes in the actual DNA sequence. These modifications including DNA methylation or histone acetylation can be influenced both by genetic and environmental factors and influence the gene expression pattern, which results in differential disease outcomes. Recently, EWAS permits to assess DNA modifications on a global level using NGS technologies. This approach highlights the key role of epigenetic signatures on biological processes, disease development, and other pathophysiologic states. <sup>11</sup> In transplantation field, a better understanding of these epigenetic modifications in both donors and recipients may lead to the identification of biomarkers, improvement in transplant recipient selection, and progress in overall prediction outcome. <sup>13</sup>
Phenomics	Phenomics is an emerging field that studies the full range of physical and biochemical traits belonging to a given organism: the phenome. Phenomics focuses on quantitative and qualitative measurements of traits and also concerns the characterization of genotype–phenotype associations. The development of phenomics capabilities requires expertise from multiple fields (biology, data sciences, engineering, etc). It is a necessary complement to genomics, offers useful insights into important outcomes such as disease and allows a better understanding of the genotype–phenotype map. <sup>14</sup>

CNV, copy number variation; EWAS, epigenome-wide association study; GWAS, genome-wide association study; MS, mass spectrometry; ncRNA, noncoding RNA; NGS, next-generation sequencing; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; RNA-Seq, RNA sequencing; SNP, single-nucleotide polymorphism.

1. Omics data come from multiple sources, require an interdisciplinary team of specialists (clinicians, specialized biologists, bioinformaticians, etc), and need in certain cases, a long period of time to be collected. This implies being very careful about the preanalytical processes because multi-source data sets contain even more systematic (biological or nonbiological) errors than a classic data set. In addition, omics data tend to have missing data, partly because of how they are collected (eg, hand-filled form) and how they are stored (multicentric databases with variables and technologies that may vary depending on the center). Procedures such as normalization, correction of batch effects, and imputation are highly needed to achieve the most reliable, complete, and responsive data set.
2. The data analysis in transplantation brings additional complexity because of the statistical unit of interest, which

may include a single genome (recipient or donor) or 2 genomes (the donor–recipient pair). Therefore, statistical methods for analysis will be different/more complex. In transplantation, strategies such as combined risk scores between 2 genomes are emerging to better understand graft mechanisms and refine biological interpretations.

In this review, we provide a nonexhaustive summary of omics studies in transplantation and address the most challenging analytical issues regarding these approaches.

## OMICS STUDIES IN TRANSPLANTATION

High-throughput analyses in transplanted patients have revealed key biomarkers<sup>18</sup> for transplantation outcomes. Monitoring biomarkers in transplantation aims to predict



rejection, treatment outcomes, and immunosuppressive therapy adaptation.<sup>19,20</sup>

Table 2 is a summary of omics studies published in the transplantation field with >70 reported articles. These studies focused on acute rejection (AR) with very little progress on chronic rejection/loss. This reflects the fact that chronic rejection/loss is more difficult to study because it involves more complex and heterogeneous mechanisms. It occurs months or years after organ or tissue transplantation and may involve changes on different levels: chronic inflammation, humoral, cellular immune reactions, etc. In addition, data related to chronic rejection/loss can be more difficult to obtain because it requires follow-up information from patients and may involve access to biopsy tissues to better capture the mechanisms underlying the progress of the graft. Usually, the appropriate sample type depends on the study (what are we trying to quantify, where? etc). In transplantation, the sources are various but what comes up most often according to Table 2 is that there are usually blood draws and also biopsy tissues. However, biopsies are more difficult to obtain, especially in large quantities. As a result, few improvements have been observed in the long-term survival of the graft<sup>97</sup> and the mechanisms of this dysfunction are still poorly understood. Overall, the studied cohorts remain modest regarding sample size and 80%

of studies do not exceed 500 samples, which is low for omics studies and power for statistical discovery. Sample size needs special attention in omics studies, as often, the statistical methods only reveal significant results when the cohort size is large.<sup>98</sup> In addition, it has been proven in genomics and metabolomics that an increasing sample size can improve the robustness and diagnostic accuracy of the markers.<sup>99</sup>

The issue that often arises is how to anticipate these questions? What cohort size could be sufficient to rule out a possible lack of statistical power and validate the presence or absence of biologically relevant results? How to balance an undersized study that may lead to missing an interesting feature because of a lack of sensitivity versus an oversized study that involves too many subjects with potentially harmful or invasive experiments that can raise ethical questions? In both cases, there will be a waste of time and money. The studies cited above suggest that sample size determination should be based on the specific purpose of the study. But that conclusion is really problematic. There are indeed ways to determine the number of subjects that may be needed for a study.<sup>100-102</sup> But it implies that we have a strong a priori about the outcome. For omics studies, analyses may assess millions of features and therefore require several simultaneous statistical tests (multitesting).

**TABLE 2.****Summary of omics studies in transplantation**

Organ type	Omics	Sample type	Sample size	Outcome	References
Kidney	Genomics	Blood, biopsies	100–3148	Acute rejection, chronic rejection, long-term allograft function, renal function, graft survival, immunosuppressor	21-30
	Transcriptomics	PBL, biopsies, whole blood, urine	10–485	Acute rejection, medium term allograft function, tolerance, renal function	31-38
	Proteomics	Biopsies, blood, urine, plasma	20–70	Acute rejection, chronic rejection, allograft survival, IFTA, chronic allograft nephropathy	39-47
	Epigenomics	Urine	88	Acute kidney injury	48
	Metabolomics	Urine, biopsies, serum	16–165	Renal function, immunosuppressive therapy, rejection	49-52
	Multomics: proteomics+genomics	Blood, biopsies, plasma	32–40	Acute rejection, chronic rejection	39, 53
Heart	Genomics	Blood, biopsies	300–1043	Left ventricular dysfunction, graft survival, CAV	54, 55
	Transcriptomics	Blood, PBMC	113–602	Acute rejection, chronic rejection, allograft loss	56-60
	Proteomics	Biopsies, plasma, peripheral blood	22–40	Acute rejection, CAV	61-63
	Epigenomics	Biopsies	42	Acute rejection	64
Lung	Genomics	Biopsies, BALF	71–226	Viral infection, acute rejection, BOS, graft survival	65, 66
	Transcriptomics	Blood	132	BOS	67
	Proteomics	BALF	7–82	Chronic lung rejection, graft survival, graft dysfunction	68-71
	Metabolomics	BALF, EVLP	29–50	Allograft injury, early transplant outcomes	72, 73
Liver	Genomics	Whole blood	80	Rejection	74
	Transcriptomics	Blood, biopsies	42–102	Allograft tolerance	75-79
	Proteomics	Blood, biopsies	9–46	Acute rejection, primary graft nonfunction	80, 81
	Metabolomics	Biopsies	78–124	Early allograft dysfunction	82, 83
HSC	Genomics	Blood, biopsies	168–1589	Acute GvHD, survival	84-86
	Transcriptomics	Blood	63	Biomarker expression	87
	Proteomics	Urine, BALF	30–412	Acute GvHD, chronic GvHD, IPS	88-90
	Metabolomics	Plasma, serum	99–114	Acute GvHD	91, 92
	Epigenomics	Blood	47–85	Acute GvHD, immune reconstitution	93-95

BALF, bronchoalveolar lavage fluid; BOS, bronchiolitis obliterans syndrome; CAV, chronic allograft vasculopathy; EVLP, ex vivo lung perfusion; GvHD, graft-vs-host disease; HSC, hematopoietic stem cell; IFTA, interstitial fibrosis and tubular atrophy; IPS, idiopathic pneumonia syndrome; PBL, peripheral blood lymphocyte; PBMC, peripheral blood mononuclear cell.

It is then really challenging to predict the sample size in advance. The least we can do to ensure a good sensitivity for a current data set is to make sure that regardless of the size of the data, we have a homogeneous data set (patient cohorts, technological platforms, etc) and that we did anticipate the identification of variables and covariates.

Regardless of the transplantation type, most omics studies (>75%) focused on genome, transcriptome, and proteome with very few investigations of the metabolome. Epigenomics and multiomics strategies have been barely investigated. Kidney omics studies are predominant (at least 1 study in each omic type and >35% of reported studies) with slightly more advances in genomics, proteomics, and transcriptomics (>20 studies). In the next section, we will further detail genomics in kidney transplantation to exemplify how omics can contribute to moving the understanding of transplantation outcomes forward.

### Genomics Studies in Kidney Transplantation

The genomics era was marked by the apogee of genome-wide association studies (GWASs)<sup>9</sup> that have identified hundreds of thousands of genetic variants (mainly single-nucleotide polymorphisms [SNPs]) associated with a trait or disease. These approaches assess the allelic frequency difference between 2 different conditions (eg, healthy versus diseased) using statistical testing. More than 5273 studies and 276 696 associated traits have been reported in the GWAS catalog to date.<sup>9</sup> In transplantation, GWASs have enabled a deeper understanding of the mechanisms underlying the interaction between donor and recipient.<sup>103</sup>

The implication of genetic factors in transplant tolerance and rejection was proven in the early 1950s, much before the first successful solid-organ transplantation in humans.<sup>104,105</sup> Recently, GWASs<sup>9</sup> have enabled an investigation of molecular mechanisms underlying kidney graft loss, including some innovative strategies considering the donor and recipient genomes at once.

First, several studies have focused on the understanding of AR. In 2016, Ghisdal et al<sup>22</sup> performed the very first GWAS of AR in kidney transplantation and identified 2 significant loci (discovery cohort: 275 cases and 503 controls and replication cohort: 313 cases and 531 controls). The first locus encompasses *PTPRO*, coding for a receptor-type tyrosine kinase essential for B cell–receptor signaling (rs10846175 and rs7976329 SNPs), and the second locus involves the ciliary *CCDC67* gene (rs10765602 SNP). Both genes were thought to be interesting candidate genes as *PTPRO* might modulate alloreactivity<sup>106</sup> (regulation of T cell–receptor and B cell–receptor signaling and cytokine production) and *CCDC67* might play a role at the immune synapse; recent evidence indicate tight similarities between the primary cilium and the immune synapse.<sup>107</sup> Two years later, Hernandez-Fuentes et al<sup>23</sup> published a study based on several cohorts that investigated the association and the interaction effects of either donor SNPs, recipient SNPs, donor–recipient SNPs on graft survival (graft failure or patient death with median follow-up time of 1886 d [~5 y]), and AR (AR event recorded within 12 mo posttransplant). Despite large sample size (discovery = 441 cases versus 941 controls, replication = 575 cases versus 2573 controls), they did not replicate Ghisdal et al's results and failed to identify any significant associations. It is important to note that these findings are only based on cohorts

of European ancestry, which suggests that their conclusions may not be transferable to other populations.

AR risk is now well managed, thanks to the continuous development of new immunosuppressive agents, as shown by the fact that <10% of patient's experienced AR in France in 2020.<sup>108</sup> However, chronic graft loss remains a challenge as there is 56% of graft survival within 10 y posttransplant even with a fully HLA compatible donor.<sup>97</sup> In 2013, O'Brien et al<sup>21</sup> performed a GWAS that included 326 renal transplants from Irish recipients with long-term follow-up. They identified 2 SNPs (rs3811321 and rs6565887) that were associated with serum creatinine levels at 5 y. However, a validation study by Pihlstrøm et al<sup>26</sup> in 2016 on 1600 kidney grafts from Caucasian recipients failed to replicate these results.

Given the lack of strong SNP associations with chronic rejection and long-term outcomes, some studies have turned to a “combined donor–recipient” GWAS analytical strategy. They simultaneously looked at the genetic variants in both individuals to summarize the incompatibilities between recipients and donors into a genetic score. Classically in GWAS, polygenic risk scores are used to evaluate the combined effect of SNP alleles with an individual phenotype (such as the risk factor to develop a disease). Here, instead of looking at 1 single genome and on combining effect size of each significant SNP, the score is calculated for a donor–recipient pair as the sum of mismatch (ie, allelic difference) contributions over a set of genomic polymorphisms across the genome.<sup>25</sup> These methods add a layer of complexity as we have to deal with 4 different sets of chromosomes.

This approach was used in 2 studies published in 2019 respectively by Reindl-Schwaighofer et al<sup>28</sup> and Steers et al.<sup>27</sup> The first study looked at nonsynonymous mismatches for transmembrane and secreted proteins and determined an a non-HLA score, based on a median value above which the risk of kidney graft loss is increased. The second study screened 50 deletions-tagging SNPs from 704 kidney allograft recipients and defined a genomic collision (risk genotype for graft rejection) as a specific donor–recipient genotype combination in which a recipient who was homozygous for a deletion-tagging allele received a transplant from a non-homozygous donor. Both recent studies confirmed that non-HLA factors are involved in chronic renal transplantation outcomes. Extending the use of polygenic risk scores in kidney transplantation, these 2 studies echoed the one by Mesnard et al,<sup>25</sup> who introduced for the first time a summing scoring system to evaluate non-HLA factors in renal graft function from genomic data (next-generation exome sequencing).

This type of strategy may benefit from integrating additional data layers (gene expression, proteins, metabolites) to better understand transplantation outcomes. Researchers have recently combined multiple omics data sets from the same set of samples to capture complexity and reconstruct molecular networks that cannot be reached with single omics resolution.<sup>53,109,110</sup> In kidney transplantation, the pursuit of biomarkers through a comprehensive approach is ongoing. A study applied unsupervised and supervised methods to explore and integrate genomics and proteomics data from kidney transplant patients (AR events within 30 d posttransplant [n=20] versus nonrejecting controls [n=20]) and to determine which genes and proteins play important roles in acute renal allograft rejection. A better separation between the 2 groups of samples was observed

in the integrative analysis (genomics and proteomics combined) than single omics analysis.<sup>53</sup>

## ANALYTICAL METHODS TO TACKLE OMICS DATA CHALLENGES

Integrating large amount of data entails specific attention to analytical processes. Figure 1 represents the workflow of a big data analysis in health. (1) Scientific questions remain the starting point. Researchers should not jump into analyzing complex big data while neglecting the critical first step of putting data into context. (2) Preanalytical steps include recruitment, collection, and storage as well as quality control (QC) methods that ensure data integrity. However, the right data may not be available or there might be information gaps. (3) The third point is the analytical process that involves downstream analyses and their interpretation. Although statistical analyses are pretty standardized in omics (eg, multivariate regression models), preanalytical data treatment dealing with confounding factors such as batch effects and incomplete data remains one of the first obstacles and a key step to obtain a reliable data set.

### Preanalytical Steps

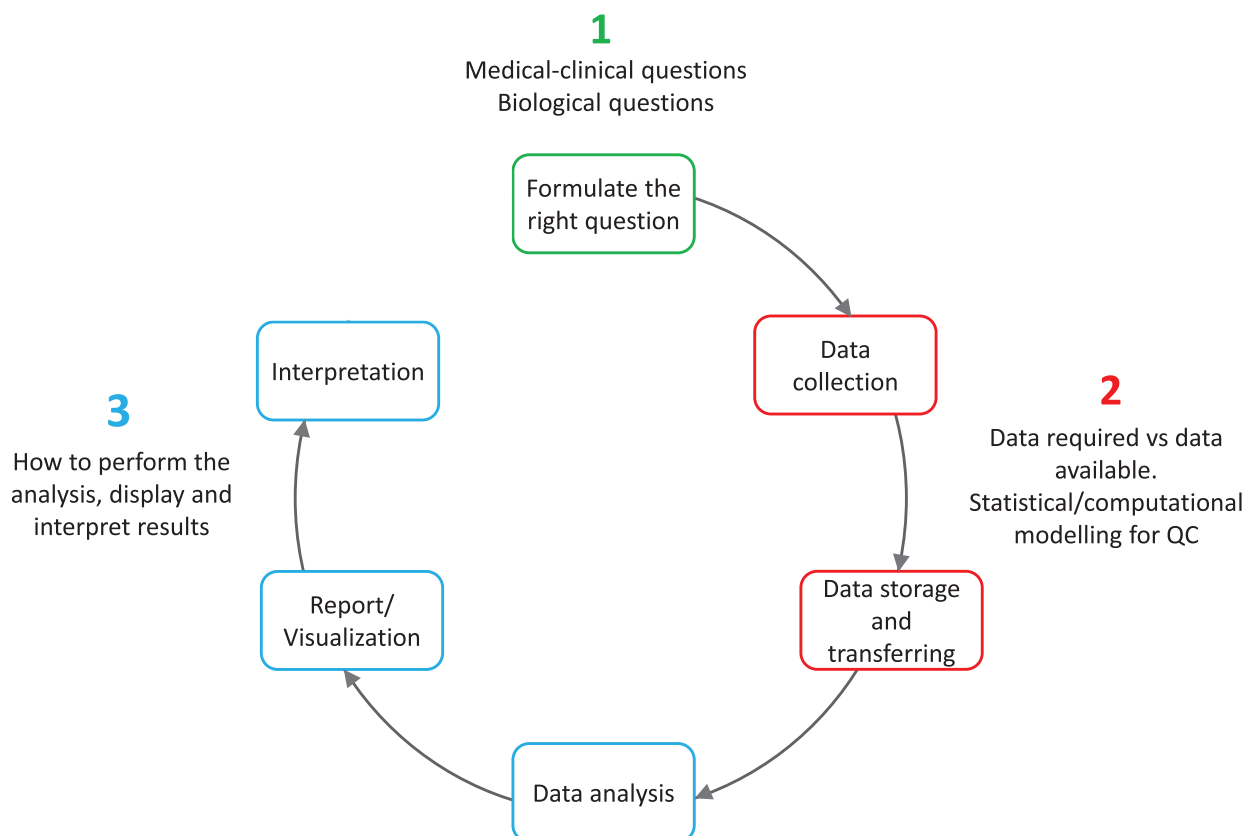
#### Quality Control of Data

Data QC is crucial to achieve the most reliable, complete, and responsive data set.<sup>111</sup> As an example, it has been stipulated that at least 8% of SNPs from single nucleotide polymorphism database would be because of artifacts from paralogues,<sup>112</sup> and next-generation sequencing

data contain systematic errors that are difficult to differentiate from true variants.<sup>113</sup> Therefore, estimating data trustworthiness becomes an essential first step to assess their relevance and limit analytical errors. Especially in the omics era, QC has encompassed various statistical methods to check for proper data generation and experimental procedures and for potential contamination before running the main analysis.<sup>114-116</sup> In transplantation, some papers have shown the importance of QC procedures in large clinical databases.<sup>117,118</sup> Generally, the standard QC analysis includes calculating quality-related measures, checking consistency among samples, detecting any experiment batch effect (technical heterogeneity), or outlying observations. Below, we focus on 2 standard pre-processing methods: normalization methods and batch effect corrections.

### Normalization

The normalization method calibrates systematic errors that may be introduced by imperfect instrument calibration. This data analysis technique adjusts global properties of measurements for individual samples so they can be appropriately comparable. Optimally, a normalization method for a given data set should be selected both on quantitative and qualitative evaluation measures.<sup>119</sup> This has been successfully applied in several transplant omic studies including transcriptomics,<sup>120,121</sup> genomics,<sup>28</sup> metabolomics,<sup>122</sup> proteomics,<sup>123</sup> etc. In transcriptomics, Wang et al<sup>120</sup> used 9 various normalization methods to define housekeeping (HKG) genes suitable for RNA



**FIGURE 1.** Workflow of big data analysis. The workflow is a process that follows a certain direction (the orientation of arrow is important). 1—Scientific questions are the starting point to process big data analysis. 2—Preanalytical processes that include the recruitment, the collect, and the storage of data. The large amount of data potentially available warrants QC methods that ensure data integrity. 3—Downstream analysis processes and the right interpretation of the results, that is, which question can be answered with our results and possibly which application can be made. QC, quality control.

sequencing analysis of kidney biopsy tissue. They show that including normalization methods in objectively derived HKG data sets will avoid errors resulting from focusing on genes with unstable expression. In genomics, studies are mainly based on GWAS methods that include a robust QC procedure.<sup>124</sup> The normalization methods are mainly based on log transformations.<sup>125</sup> Most studies related to transplantation include those steps as a standard protocol before the main downstream analysis.<sup>27,28</sup> Several normalization methods have been compared in all these studies with 2 main observations: (1) the suitability of a normalization method depends on intrinsic characteristics of the data set including variations in sample processing conditions, experimental and technical protocols, complexity of biological features, etc. (2) The best technique often depends on the type of omics. However, quantile normalization<sup>126</sup> as well as log transformation<sup>125</sup> has been largely used. Some studies offer to evaluate several normalization methods in parallel to guide the selection of the optimal method for a data set.<sup>119,127</sup>

### Correction of Batch Effects

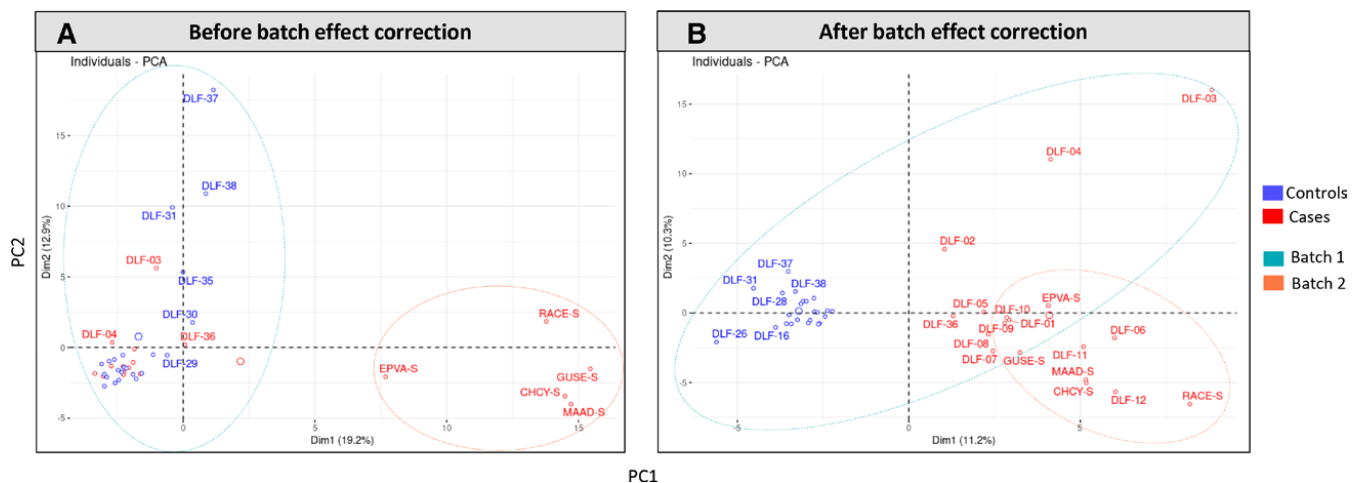
Batch effects are experimental artifacts, that is, systematic nonbiological variations between groups of samples (batches) introduced during experiments.<sup>128,129</sup> For example, batch effects may occur if experiments are performed at different times, by different persons and/or in different conditions.<sup>130</sup> The presence of batch effects has been reported among most, if not all, high-throughput technologies.<sup>128</sup> In practice, batch effects are almost inevitable.<sup>131</sup> At best, batch effects may lead to an increase in data variability, characterized by a reduction in true-positive biological signals and an inflation of false-negative signals. More critical consequences may occur when batch effects are mistaken for an outcome of interest (false positive).<sup>132</sup> For example, in 2018, Leigh et al<sup>133</sup> presented a case study where small read misalignments created false SNP detection because of a batch effect in a multiyear study.

To limit the risk of batch effects, experiments should be designed to randomly distribute each potential source of experimental variation across biological groups. In addition, computational solutions have been developed: (1) exploratory analyses using distance measures,<sup>134</sup>

clustering,<sup>135</sup> and spatial methods<sup>136</sup> to detect and quantify the impact of batch effects, as well as the effect of other technical artifacts in the data; (2) downstream statistical analyses to remove or adjust for the undesirable effects. In genomics, principal component analysis<sup>134</sup> is mostly used, especially in GWASs, to detect biases (Figure 2) and capture variability related to ancestry or phenotypic outcomes.<sup>130,137</sup> In transcriptomics, empirical Bayes<sup>138</sup> is best recommended,<sup>139</sup> but it is also largely used in other omics studies such as metabolomics. In proteomics, the combination of different methods (principal component analysis, clustering, linear regression, etc) seems to produce the best results.<sup>140</sup> All these methods have been successfully applied in transplantation studies.<sup>94,141,142</sup> Therefore, the best method depends on the type of omics and also on the size of the data set, the origin of the batch errors, and the batch effect that is hardly detectable if it is not the main source of the variation in the data set.<sup>130</sup> In any case, one should keep in mind that whatever method chosen, it can only limit the batch error and no solution can replace a good experimental design and analysis techniques.<sup>131</sup>

### Imputation

The imputation principle is to fill in missing data with a probable value representative of the data set, with the goal to reduce the modeling bias or improve accuracy of the model. The “missing data” term refers to partially available information for some variables, features, and individuals.<sup>143</sup> Imputation by definition tends to make the imputed data sets similar to the reference database including cases imputed by controls.<sup>144</sup> This requires a careful approach, as they can lead to significant bias and loss of precision in downstream analyses.<sup>145</sup> Three-category typology for missing data is defined<sup>143</sup>: (1) missing at random; (2) missing completely at random; (3) missing not at random (see Table 3). Different imputation solutions exist for filling in the missing values depending on their proportion and type.<sup>146</sup> One solution is to use simple imputations methods that replace information with values from the global characteristics of the data set (mean or median, linear regression, k-nearest,<sup>147</sup> random forest,<sup>148</sup> local regression Locally Estimated Scatterplot Smoothing,<sup>149</sup> etc). However, simple imputation usually causes standard errors to be too small, because it fails to consider uncertainty.<sup>150</sup>



**FIGURE 2.** PCA before and after batch effect correction. A, Before the correction, individuals from the same batch are grouped together. B, After the correction, individuals are grouped depending on the disease. PC, principal component; PCA, PC analysis.



**TABLE 3.****The different categories of missing data**

MAR	There might be systematic differences between the missing and observed values, but these can be entirely explained by other observed variables. It happens when the data are not missing completely random (conditionally to other variables). There are appropriate statistical methods to avoid biased analysis.
MCAR	A special case of MAR. Data are MCAR when the missing values are randomly distributed across all observations. In this case, the probability of absence is the same for all observations and the missing observations are a random subset of all observations. This probability does not depend on any external parameters independent of this variable. If the amount of MCAR data is not too important, it is reasonable to ignore cases with missing data as it will not bias the analysis. However, a loss of precision in the results is predicted.
MNAR	These data depend directly on the variable in question. These are the most problematic type of missing values as they are not randomly distributed across the observations. Furthermore, there is no information in the data that would help to choose an imputation model (the probability cannot be predicted from the variables in the model). MNAR data lead to loss of precision (inherent to all missing data) and also bias that requires the use of a sensitivity analysis.

MAR, missing at random; MCAR, missing completely at random; MNAR, missing not at random.

Multiple imputation has been arguably the most popular method for handling missing data in practice.<sup>151</sup> It consists of imputing the missing values several times and combining the results to take into account the observed variability and reduce inference errors.<sup>152</sup>

Examples of missing data imputation in an omics transplantation context are found in mass spectrometry based on metabolomics or proteomics data.<sup>153-155</sup> Comparing the performance of imputation methods for metabolomics data sets in kidney transplanted patients, several studies reported that random forest and the k-nearest neighbor method were usually outperforming other methods.<sup>153,156,157</sup> In genomics, missing data imputation is crucial. Using a reference panel, SNPs with missing data or SNPs that were not directly covered by the genotyping arrays can be predicted from the neighboring genotyped SNPs by comparing the haplotypes and linkage disequilibrium patterns between the data set of interest and the reference panel. The size and diversity of the reference panel are therefore the key for high accuracy.<sup>158</sup> International collaborative efforts, such as the HapMap<sup>159</sup> or the 1000 Genomes project,<sup>160</sup> provide a valuable data source in human diversity representative populations. Some tools such as Haplotype Reference Consortium<sup>161</sup> or Beagle<sup>162</sup> offer to predict missing SNPs from a given data set and are widely used in genomics studies related to transplantation.<sup>21,22</sup>

Finally, the current challenge for missing data imputation is the integration of multiomics.<sup>163</sup> An imputation method has recently been proposed to integrate network dimensions and multiple correlated omics data sets.<sup>164</sup>

## DISCUSSION

Omics approaches, including GWAS, are characterized by the simultaneous testing of thousands to millions of features (eg, gene expression, SNP) in parallel for association with the phenotype of interest<sup>137</sup>; it is necessary to adopt a stringent significance threshold to account for multiple testing and limit false-positive discoveries. In GWASs, Bonferroni corrections are typically applied<sup>165</sup> with a *P* value threshold of  $5 \times 10^{-8}$  for SNPs to maintain the genome-wide false-positive rate at 5%.<sup>137</sup>

However, a high expectation of statistical significance considerably reduces the power to detect associated SNPs, as most association signals do not reach the required threshold.<sup>166</sup>

One way to address this issue is to increase sample size, which is not always achievable especially in transplantation

context. Actually, sample sources such as biopsy tissues are barely available for certain organs. Recruitment and data collection require a huge investment of both time and money that is generally not manageable by most individual laboratories.<sup>167</sup> Collective efforts have been made to give access to large cohorts<sup>159,160</sup> with substantial genetic and clinical information. However, an inherent issue related to large multicenter in transplantation is the variability in donor selection and clinical protocols (immunosuppressive treatments) among centers that may impact the outcome.<sup>168</sup> Plus, the criteria used to define phenotypes such as AR and chronic allograft dysfunction might vary depending on transplant centers. This represents a critical issue in transplantation where the gold standard is the histology that can be subjective, as shown by the continual updating of the Banff classifications for AR over time.<sup>169</sup>

All of these introduce further heterogeneity between studies and have made genetic association studies difficult to determine and validate.<sup>170</sup> This may explain, alongside the lack of power in statistical analysis, the lack of reproducibility among transplantation studies.

The International Genetics and Translational Research in Transplantation Network<sup>171</sup> consortium integrates multiple cohorts of different types of allograft recipients to overcome the problem of the small impact of SNPs. The consortium homogenized genotyping methods and used common definitions of AR to perform robust meta-analyses of several types of transplantation. In addition, International Genetics and Translational Research in Transplantation Network accounted for population stratification and appropriate inclusion of covariates that lead to better identification of SNPs.<sup>172-175</sup> This last point raises a real issue that is not always properly controlled in transplantation studies. Covariates such as age, sex, and ancestry are included to avoid confounding effects from demographic factors and to account for stratification.<sup>124</sup> Genetic associations may differ across ancestries, complicating direct comparisons between groups of individuals.<sup>167</sup> In transplantation, the impact of ancestry is even more important as it implies compatibility between donor and recipient. Many studies have identified variants such as *CYP3A5* and *APOL1* that are associated with concentration of tacrolimus and greater risks of graft failure in African Americans, respectively.<sup>176,177</sup> These studies illustrate the differences between populations and emphasize the urge to better characterize such genomic diversity, including in the transplantation field. For

example, recipient–donor genomic interactions may play a critical role in graft survival. Most transplantation studies have focused on recipient SNPs, but there have been a few studies that have turned to a “combined donor–recipient” GWAS strategy.<sup>25,27,28</sup> They investigate genetic interactions between donor and recipient to summarize their incompatibilities into a genetic score (cf. section Genomics studies in kidney transplantation). These approaches are highly promising and may finally give insights into chronic graft rejection/loss. Multifactorial nature of genetics including transplantation (tetraploid 4n) genetics is often intended to be reduced into a monogenic “signature.” Similar remarks are valid for reduction of expression signature to 1 or 2 genes. This is fundamentally wrong and approaches such as “combined donor–recipient” incompatibilities scores permit to partially address this multifactorial level.

However, very few studies have sufficient numbers to be statistically powered to validate polygenic scores. This requires rigorous data collection and appropriately phenotyped recipients and donors. Importantly, these studies made arbitrary choices on the outcomes and used different phenotypes to define rejection or graft loss. Hence, there is a need of collaborative and homogeneous large cohorts to achieve robust and replicable studies. Moreover, beyond genomic appreciations, functional studies are necessary to substantiate clinical decision making with -omic research including donor-specific and anti-HLA immunization.

## CONCLUSION

Big data exploration offers new analytical opportunities, especially with the emergence of omics data both in transplantation research and care<sup>15–17</sup> to move toward personalized medicine and personalized therapies. Omics methods offer to better assess risk stratification for subcluster of phenotypes. The ambition to reach individualized medicine (at the patient level) may not be attainable at present because of the multifactorial aspect of transplantation, which involves taking into account the donor and recipient, and also because of the complex mechanisms underlying transplantation especially in chronic rejection/loss. Large international consortium including (The European Transplantation and Innovation Consortium, (<https://eu-train-project.eu>) aims to combine several types of data set in a single tool to better monitor and predict the follow-up of kidney transplanted patients.<sup>178</sup> Despite the growing breadth of available data, it is critical to ensure their homogeneity and veracity before analysis, as they can sometimes be incomplete and contain systematic biases that may be confused with relevant biological information. The very first step is to shift to standardization of procedures in clinical trials and phenotype’s definitions such as AR and chronic allograft dysfunction to perform robust meta-analyses. Plus, we presented 2 preprocessing methods (normalization and batch effect corrections) and statistical inference techniques for missing values, such as multiple imputation, to ensure a high-quality enriched data set. Proper management and analysis of omics data have revealed promising results in the transplantation field (Table 2). Good advances have been made in the omics field especially in kidney transplantation. However, as exemplified in this review by genomic studies in kidney transplantation, classic analytical strategies can sometimes lead to inconclusive results, due, for

example, to study design or power limitations.<sup>24,26</sup> Hence, challenges such as a lack of adequate sample sizes and lack of replication are yet to be managed. Importantly, the scientific community needs to proactively work to ensure that studies that are to date very euro-centered are representative of the global human population and its diversity. This underlines the need for validation studies and large collaborative research initiatives to gain power and discover novel biomarkers in transplantation. Alternative solutions for multitesting are explored including gene-based or pathway-based association to reduce the number of tests performed and also to have a functional perspective of the data and facilitate their interpretation.<sup>166</sup> Finally, this calls for innovative analytical strategies such as the computation of donor–recipient genetic incompatibility scores<sup>25,28</sup> specific to the transplantation context and the growing interest for multiomics approaches.<sup>109,179</sup> Plus, more specific guidelines are needed for the discovery and validation of clinically relevant biomarkers<sup>180</sup> and the reporting of GWASs (like Strengthening the Reporting of Genetic Association Studies for genetic associations<sup>181</sup> and Strengthening the Reporting of Immunogenomic Studies for immunogenetics<sup>182</sup> guidelines). Interestingly, addressing data organization, computational challenges, and the translational gap of -omics as a source of data for both research and clinics is insightful for enriching the use of large registries and epidemiological cohorts as well as electronic health records (EHR/EMR) and the emergence of biomedical data warehouses. These developments offer the promise to further increase in the near future our understanding of complex molecular mechanisms underlying transplantation outcomes.

## REFERENCES

1. Doug L. 3D data management: controlling data volume, velocity and variety. 2001. Available at <https://pdfcoffee.com/ad949-3d-data-management-controlling-data-volume-velocity-and-varietypdf-pdf-free.html>. Accessed October 13, 2021.
2. Panimalar SA, Shree SV, Kathrine AV. The 17 V’s of big data. *IRJET*. 2017;4:329–333.
3. Chang AC. Big data in medicine: the upcoming artificial intelligence. *Prog Pediatr Cardiol*. 2016;43:91–94.
4. De Mauro A, Greco M, Grimaldi M. A formal definition of big data based on its essential features. *Libr Rev*. 2016;65:122–135.
5. Desjardins J. How big data will unlock the potential of healthcare. 2018. Available at <https://www.visualcapitalist.com/big-data-healthcare>. Accessed October 13, 2021.
6. Stewart C. Projected growth in global healthcare data volume 2020. 2020. Available at <https://www.statista.com/statistics/1037970/global-healthcare-data-volume> Accessed October 13, 2021.
7. Huang T, Lan L, Fang X, et al. Promises and challenges of big data computing in health sciences. *Big Data Res*. 2015;2:2–11.
8. Conesa A, Beck S. Making multi-omics data accessible to researchers. *Sci Data*. 2019;6:251.
9. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*. 2017;45:D896–D901.
10. Bontha SV, Maluf DG, Mueller TF, et al. Systems biology in kidney transplantation: the application of multi-omics to a complex model. *Am J Transplant*. 2017;17:11–21.
11. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18:83.
12. Rhee EP. How omics data can be used in nephrology. *Am J Kidney Dis*. 2018;72:129–135.
13. Rakyen VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12:529–541.
14. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet*. 2010;11:855–866.

15. Stapleton CP, Conlon PJ, Phelan PJ. Using omics to explore complications of kidney transplantation. *Transpl Int*. 2018;31:251–262.
16. Sirota M, Sarwal MM. Transplantomics: toward precision medicine in transplantation research. *Transplantation*. 2017;101:1777–1782.
17. Srinivas TR, Taber DJ, Su Z, et al. Big data, predictive analytics, and quality improvement in kidney transplantation: a proof of concept. *Am J Transplant*. 2017;17:671–681.
18. Naesens M, Sarwal MM. Molecular diagnostics in transplantation. *Nat Rev Nephrol*. 2010;6:614–628.
19. Perkins D, Verma M, Park KJ. Advances of genomic science and systems biology in renal transplantation: a review. *Semin Immunopathol*. 2011;33:211–218.
20. Rozentals R, Jušinskis J, Amerika D. Biomarkers and their potential use in transplantation. *Proc Latvian Acad Sci, Section B*. 2012;66:219–268.
21. O'Brien RP, Phelan PJ, Conroy J, et al. A genome-wide association study of recipient genotype and medium-term kidney allograft function. *Clin Transplant*. 2013;27:379–387.
22. Ghisdal L, Baron C, Lebranchu Y, et al. Genome-wide association study of acute renal graft rejection. *Am J Transplant*. 2017;17:201–209.
23. Hernandez-Fuentes MP, Franklin C, Rebollo-Mesa I, et al; United Kingdom and Ireland Renal Transplant Consortium (UKIRTC) and the Wellcome Trust Case Control Consortium (WTCCC)-3. Long- and short-term outcomes in renal allografts with deceased donors: a large recipient and donor genome-wide association study. *Am J Transplant*. 2018;18:1370–1379.
24. Massart A, Ghisdal L, Viklicky O, et al. Reply to Hernandez et al.—GWAS of acute renal graft rejection. *Am J Transplant*. 2018;18:2098–2099.
25. Mesnard L, Muthukumar T, Burbach M, et al. Exome sequencing and prediction of long-term kidney allograft function. *Plos Comput Biol*. 2016;12:e1005088.
26. Pihlström HK, Mjølén G, Mucha S, et al. Single nucleotide polymorphisms and long-term clinical outcome in renal transplant patients: a validation study. *Am J Transplant*. 2017;17:528–533.
27. Steers NJ, Li Y, Drace Z, et al. Genomic mismatch at LIMS1 locus and kidney allograft rejection. *N Engl J Med*. 2019;380:1918–1928.
28. Reindl-Schwaighofer R, Heinzl A, Signorini L, et al. Mechanisms underlying human genetic diversity: consequence for anti-graft antibody responses. *Transpl Int*. 2018;31:239–250.
29. Sarwal M, Chang S, Barry C, et al. Genomic analysis of renal allograft dysfunction using cDNA microarrays. *Transplant Proc*. 2001;33:297–298.
30. Birdwell KA, Grady B, Choi L, et al. The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenet Genomics*. 2012;22:32–42.
31. Hidalgo LG, Einecke G, Allanach K, et al. The transcriptome of human cytotoxic T cells: measuring the burden of CTL-associated transcripts in human kidney transplants. *Am J Transplant*. 2008;8:637–646.
32. Mueller TF, Einecke G, Reeve J, et al. Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets. *Am J Transplant*. 2007;7:2712–2722.
33. Saint-Mezard P, Berthier CC, Zhang H, et al. Analysis of independent microarray datasets of renal biopsies identifies a robust transcript signature of acute allograft rejection. *Transpl Int*. 2009;22:293–302.
34. Scherer A, Gwinner W, Mengel M, et al. Transcriptome changes in renal allograft protocol biopsies at 3 months precede the onset of interstitial fibrosis/tubular atrophy (IF/TA) at 6 months. *Nephrol Dial Transplant*. 2009;24:2567–2575.
35. Brouard S, Mansfield E, Braud C, et al. Identification of a peripheral blood transcriptional biomarker panel associated with operational renal allograft tolerance. *Proc Natl Acad Sci U S A*. 2007;104:15448–15453.
36. Danger R, Pallier A, Giral M, et al. Upregulation of miR-142-3p in peripheral blood mononuclear cells of operationally tolerant patients with a renal transplant. *J Am Soc Nephrol*. 2012;23:597–606.
37. Suthanthiran M, Schwartz JE, Ding R, et al; Clinical Trials in Organ Transplantation 04 (CTOT-04) Study Investigators. Urinary-cell mRNA profile and acute cellular rejection in kidney allografts. *N Engl J Med*. 2013;369:20–31.
38. Verma A, Muthukumar T, Yang H, et al. Urinary cell transcriptomics and acute rejection in human kidney allografts. *JCI Insight*. 2020;5:131552.
39. Nakorchevsky A, Hewel JA, Kurian SM, et al. Molecular mechanisms of chronic kidney transplant rejection via large-scale proteogenomic analysis of tissue biopsies. *J Am Soc Nephrol*. 2010;21:362–373.
40. Quintana LF, Solé-Gonzalez A, Kalko SG, et al. Urine proteomics to detect biomarkers for chronic allograft dysfunction. *J Am Soc Nephrol*. 2009;20:428–435.
41. Kurian SM, Heilman R, Mondala TS, et al. Biomarkers for early and late stage chronic allograft nephropathy by proteogenomic profiling of peripheral blood. *PLoS One*. 2009;4:e6212.
42. Clarke W, Silverman BC, Zhang Z, et al. Characterization of renal allograft rejection by urinary proteomic analysis. *Ann Surg*. 2003;237:660–664.
43. O'Riordan E, Orlova TN, Mei J J, et al. Bioinformatic analysis of the urine proteome of acute allograft rejection. *J Am Soc Nephrol*. 2004;15:3240–3248.
44. Cibrik DM, Warner RL, Kommareddi M, et al. Identification of a protein signature in renal allograft rejection. *Proteomics Clin Appl*. 2013;7:839–849.
45. Cassidy H, Slyne J, O'Kelly P, et al. Urinary biomarkers of chronic allograft nephropathy. *Proteomics Clin Appl*. 2015;9:574–585.
46. Freue GV, Sasaki M, Meredith A, et al; Genome Canada Biomarkers in Transplantation Group. Proteomic signatures in plasma during early acute renal allograft rejection. *Mol Cell Proteomics*. 2010;9:1954–1967.
47. Bañón-Maneus E, Diekmann F, Carrascal M, et al. Two-dimensional difference gel electrophoresis urinary proteomic profile in the search of nonimmune chronic allograft dysfunction biomarkers. *Transplantation*. 2010;89:548–558.
48. Mehta TK, Hoque MO, Ugarte R, et al. Quantitative detection of promoter hypermethylation as a biomarker of acute kidney injury during transplantation. *Transplant Proc*. 2006;38:3420–3426.
49. Klepacki J, Klawitter J, Klawitter J, et al. A high-performance liquid chromatography-tandem mass spectrometry-based targeted metabolomics kidney dysfunction marker panel in human urine. *Clin Chim Acta*. 2015;446:43–53.
50. Diémé B, Halimi JM, Emond P, et al. Assessing the metabolic effects of calcineurin inhibitors in renal transplant recipients by urine metabolic profiling. *Transplantation*. 2014;98:195–201.
51. Blydt-Hansen TD, Sharma A, Gibson IW, et al. Urinary metabolomics for noninvasive detection of borderline and acute T cell-mediated rejection in children after kidney transplantation. *Am J Transplant*. 2014;14:2339–2349.
52. Klepacki J, Brunner N, Schmitz V, et al. Development and validation of an LC-MS/MS assay for the quantification of the trans-methylation pathway intermediates S-adenosylmethionine and S-adenosylhomocysteine in human plasma. *Clin Chim Acta*. 2013;421:91–97.
53. Günther OP, Shin H, Ng RT, et al. Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *OMICS*. 2014;18:682–695.
54. Khush KK, Pawlikowska L, Menza RL, et al. Beta-adrenergic receptor polymorphisms and cardiac graft function in potential organ donors. *Am J Transplant*. 2012;12:3377–3386.
55. Gallardo ME, García-Pavía P, Chamorro R, et al. Mitochondrial haplogroups associated with end-stage heart failure and coronary allograft vasculopathy in heart transplant patients. *Eur Heart J*. 2012;33:346–353.
56. Deng MC, Eisen HJ, Mehra MR, et al; CARGO Investigators. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant*. 2006;6:150–160.
57. Pham MX, Teuteberg JJ, Kfoury AG, et al; IMAGE Study Group. Gene-expression profiling for rejection surveillance after cardiac transplantation. *N Engl J Med*. 2010;362:1890–1900.
58. Crespo-Leiro MG, Stypmann J, Schulz U, et al. Performance of gene-expression profiling test score variability to predict future clinical events in heart transplant recipients. *BMC Cardiovasc Disord*. 2015;15:120.
59. Loupy A, Duong Van Huyen JP, Hidalgo L, et al. Gene expression profiling for the identification and classification of antibody-mediated heart rejection. *Circulation*. 2017;135:917–935.
60. Duong Van Huyen JP, Tible M, Gay A, et al. MicroRNAs as non-invasive biomarkers of heart transplant rejection. *Eur Heart J*. 2014;35:3194–3202.
61. De Souza AI, Wait R, Mitchell AG, et al. Heat shock protein 27 is associated with freedom from graft vasculopathy after human cardiac transplantation. *Circ Res*. 2005;97:192–198.
62. Lin D, Cohen Freue G, Hollander Z, et al; Biomarkers in Transplantation Team; Networks of Centres of Excellence, Centres of Excellence for Commercialization and Research-Prevention of Organ Failure Centre of Excellence. Plasma protein biosignatures for detection of cardiac allograft vasculopathy. *J Heart Lung Transplant*. 2013;32:723–733.



63. Cohen Freue GV, Meredith A, Smith D, et al; Biomarkers in Transplantation and the NCE CECR Prevention of Organ Failure Centre of Excellence Teams. Computational biomarker pipeline from discovery to clinical implementation: plasma proteomic biomarkers for cardiac transplantation. *Plos Comput Biol*. 2013;9:e1002963.
64. Boer K, Caliskan K, Peeters AM, et al. Thymus-derived regulatory T cells infiltrate the cardiac allograft before rejection. *Transplantation*. 2015;99:1839–1846.
65. Lewandowska DW, Schreiber PW, Schuurmans MM, et al. Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PLoS One*. 2017;12:e0177340.
66. Palmer SM, Klimecki W, Yu L, et al. Genetic regulation of rejection and survival following human lung transplantation by the innate immune receptor CD14. *Am J Transplant*. 2007;7:693–699.
67. Danger R, Royer PJ, Reboulleau D, et al; COLT and SysCLAD Consortia. Blood gene expression predicts bronchiolitis obliterans syndrome. *Front Immunol*. 2017;8:1841.
68. Kosanam H, Sato M, Batruch I, et al. Differential proteomic analysis of bronchoalveolar lavage fluid from lung transplant patients with and without chronic graft dysfunction. *Clin Biochem*. 2012;45:223–230.
69. Zhang Y, Wroblewski M, Hertz MI, et al. Analysis of chronic lung transplant rejection by MALDI-TOF profiles of bronchoalveolar lavage fluid. *Proteomics*. 2006;6:1001–1010.
70. Meloni F, Salvini R, Bardoni AM, et al. Bronchoalveolar lavage fluid proteome in bronchiolitis obliterans syndrome: possible role for surfactant protein A in disease onset. *J Heart Lung Transplant*. 2007;26:1135–1143.
71. Wolf T, Oumeraci T, Gottlieb J, et al. Proteomic bronchiolitis obliterans syndrome risk monitoring in lung transplant recipients. *Transplantation*. 2011;92:477–485.
72. Lockard C, Kirk AD, Uppal K, et al. Bile acid aspiration associated with lung chemical profile linked to other biomarkers of injury after lung transplantation. *Am J Transplant*. 2014;14:841–848.
73. Hsin MK, Zamel R, Cypel M, et al. Metabolic profile of ex vivo lung perfusate yields biomarkers for lung transplant outcomes. *Ann Surg*. 2018;267:196–197.
74. Sindi R, Higgs BW, Weeks DE, et al. Genetic variants in major histocompatibility complex-linked genes associate with pediatric liver transplant rejection. *Gastroenterology*. 2008;135:830–839, 839.e1.
75. Martínez-Llordella M, Puig-Pey I, Orlando G, et al. Multiparameter immune profiling of operational tolerance in liver transplantation. *Am J Transplant*. 2007;7:309–319.
76. Lozano JJ, Pallier A, Martínez-Llordella M, et al. Comparison of transcriptional and blood cell-phenotypic markers between operationally tolerant liver and kidney recipients. *Am J Transplant*. 2011;11:1916–1926.
77. Bohne F, Martínez-Llordella M, Lozano JJ, et al. Intra-graft expression of genes involved in iron homeostasis predicts the development of operational tolerance in human liver transplantation. *J Clin Invest*. 2012;122:368–382.
78. Li L, Wozniak LJ, Rodder S, et al. A common peripheral blood gene set for diagnosis of operational tolerance in pediatric and adult liver transplantation. *Am J Transplant*. 2012;12:1218–1228.
79. Zhao X, Li Y, Ohe H, et al. Intra-graft V $\delta$ 1  $\gamma\delta$  T cells with a unique T-cell receptor are closely associated with pediatric semiallogeneic liver transplant tolerance. *Transplantation*. 2013;95:192–202.
80. Massoud O, Heimbach J, Viker K, et al. Noninvasive diagnosis of acute cellular rejection in liver transplant recipients: a proteomic signature validated by enzyme-linked immunosorbent assay. *Liver Transpl*. 2011;17:723–732.
81. Kornasiewicz O, Bojarczuk K, Bugajski M, et al. Application of a proteomic approach to identify proteins associated with primary graft non-function after liver transplantation. *Int J Mol Med*. 2012;30:755–764.
82. Cortes M, Pareja E, García-Cañaveras JC, et al. Metabolomics discloses donor liver biomarkers associated with early allograft dysfunction. *J Hepatol*. 2014;61:564–574.
83. Faitot F, Besch C, Battini S, et al. Impact of real-time metabolomics in liver transplantation: graft evaluation and donor-recipient matching. *J Hepatol*. 2018;68:699–706.
84. Niederwieser D, O'Brien SG, Pearce KF, et al. Impact of genomic risk factors on outcome after hematopoietic stem cell transplantation for patients with chronic myeloid leukemia. *Haematologica*. 2010;95:922–927.
85. Sato-Otsubo A, Nannya Y, Kashiwase K, et al; Japan Marrow Donor Program. Genome-wide surveillance of mismatched alleles for graft-versus-host disease in stem cell transplantation. *Blood*. 2015;126:2752–2763.
86. Bari R, Hartford C, Chan WK, et al. Genome-wide single-nucleotide polymorphism analysis revealed SUFU suppression of acute graft-versus-host disease through downregulation of HLA-DR expression in recipient dendritic cells. *Sci Rep*. 2015;5:11098.
87. Snavely AC, Mitin N, Torrice C, et al. Chemotherapy and stem cell transplantation increase p16 INK4a expression, a biomarker of T-cell aging. *EBioMedicine*. 2016;11:227–238.
88. Holler E, Zander A, Schiffer E, et al. Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood*. 2007;109:5511–5519.
89. Weissinger EM, Human C, Metzger J, et al. The proteome pattern cGvHD\_MS14 allows early and accurate prediction of chronic GvHD after allogeneic stem cell transplantation. *Leukemia*. 2017;31:654–662.
90. Wu B, Wendt CH, Kumar V, et al. Proteome profiling in lung injury after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2016;22:1383–1390.
91. Michonneau D, Latis E, Dubouchet L, et al. Metabolomics profiling after allogeneic hematopoietic stem cell transplantation unravels a specific signature in human acute GVHD. *Blood*. 2018;132(suppl 1):69.
92. Wu D, Liu H, Yang B, et al. Prediction of acute GVHD and relapse by metabolic biomarkers after allogeneic hematopoietic stem cell transplantation. *JCI Insight*. 2018;3:1–9.
93. Gonza M, Fernandez AF, Colado E, et al. DNA methylation dynamics in blood after hematopoietic cell transplant. *PLoS One*. 2013;8:e56931.
94. Paul DS, Jones A, Sellar RS, et al. A donor-specific epigenetic classifier for acute graft-versus-host disease severity in hematopoietic stem cell transplantation. *Genome Med*. 2015;7:128.
95. Hardy K, Smith C, Tu WJ, et al. Epigenetic programming of T cells impacts immune reconstitution in hematopoietic stem cell transplant recipients. *Blood Adv*. 2018;2:656–668.
96. Justiz Vaillant AA, Mohseni M. Chronic transplantation rejection. In: *StatPearls*. StatPearls Publishing LLC; 2021.
97. Gondos A, Döhler B, Brenner H, et al. Kidney graft survival in Europe and the United States: strikingly different long-term outcomes. *Transplantation*. 2013;95:267–274.
98. Zhang X, Cui X, Zhu F. The effects of sample size on omics study: from the perspective of robustness and diagnostic accuracy. *IOP Conf Ser Earth Environ Sci*. 2019;252:022127.
99. Kim SY. Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC Bioinformatics*. 2009;10:147.
100. Shein-Chung C, Shao J, Wang H, et al. *Sample Size Calculations in Clinical Research*. 3rd ed. Chapman and Hall/CRC; 2020.
101. Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149–150.
102. Suresh K, Chandrashekar S. Sample size estimation and power analysis for clinical research studies. *J Hum Reprod Sci*. 2012;5:7–13.
103. Menon MC, Keung KL, Murphy B, et al. The use of genomics and pathway analysis in our understanding and prediction of clinical renal transplant injury. *Transplantation*. 2016;100:1405–1414.
104. Morris PJ. Transplantation—a medical miracle of the 20th century. *N Engl J Med*. 2004;351:2678–2680.
105. Yang JYC, Sarwal MM. Transplant genetics and genomics. *Nat Rev Genet*. 2017;18:309–326.
106. Motiwalla T, Datta J, Kutay H, et al. Lyn kinase and ZAP70 are substrates of PTPROT in B-cells: Lyn inactivation by PTPROT sensitizes leukemia cells to VEGF-R inhibitor pazopanib. *J Cell Biochem*. 2010;110:846–856.
107. Finetti F, Paccani SR, Rosenbaum J, et al. Intraflagellar transport: a new player at the immune synapse. *Trends Immunol*. 2011;32:139–145.
108. Agence de la Biomédecine. Annual report. 2020. Available at <https://rams.agence-biomedecine.fr>. Accessed October 13, 2021.
109. Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet*. 2018;19:299–310.
110. Yugi K, Kubota H, Hatano A, et al. Trans-omics: how to reconstruct biological networks across multiple 'omic' layers. *Trends Biotechnol*. 2016;34:276–290.



111. Sambourg L, Sambourg L. Décrypter les données omiques: importance du contrôle qualité. Application au cancer de l'ovaire. 2015. Available at <https://tel.archives-ouvertes.fr/tel-01168479>. Accessed October 13, 2021.
112. Musumeci L, Arthur JW, Cheung FS, et al. Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Hum Mutat*. 2010;31:67–73.
113. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7:e30619.
114. Kim GT, Kim Y, Kwon MS, et al. Quality control plot for high dimensional omics data. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine, Shenzhen, China. December 15–18, 2016;1763–1766.
115. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*. 2006;24:1151–1161.
116. Soloviev M. Quality control of expression profiling data. *J Proteomics Bioinform*. 2015;8:176–187.
117. van der Meulen JH, Jacob M, Copley L. Assessing the quality of the data in a transplant registry: the European Liver Transplant Registry. *Transplantation*. 2003;75:2164–2167.
118. Kuwatsuka Y. Quality control and assurance in hematopoietic stem cell transplantation data registries in Japan and other countries. *Int J Hematol*. 2016;103:20–24.
119. Chawade A, Alexandersson E, Levander F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J Proteome Res*. 2014;13:3114–3120.
120. Wang Z, Lyu Z, Pan L, et al. Defining housekeeping genes suitable for RNA-seq analysis of the human allograft kidney biopsy tissue. *BMC Med Genomics*. 2019;12:86.
121. van de Vrie M, Deegens JK, Eikmans M, et al. Urinary MicroRNA as biomarker in renal transplantation. *Am J Transplant*. 2017;17:1160–1166.
122. Lee SK, Kim BK, Jung HY, et al. Metabolomics study for identification of potential biomarkers of long-term survival in kidney transplantation recipients. *Transplant Proc*. 2017;49:1005–1011.
123. Song L, Fang F, Liu P, et al. Quantitative proteomics for monitoring renal transplant injury. *Proteomics Clin Appl*. 2020;14:e1900036.
124. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet*. 2011;Chapter 1:Unit1.19.
125. Feng C, Wang H, Lu N, et al. Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*. 2014;26:105–109.
126. Hicks SC, Okrah K, Paulson JN, et al. Smooth quantile normalization. *Biostatistics*. 2018;19:185–198.
127. Dillies MA, Rau A, Aubert J, et al; French StatOmique Consortium. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*. 2013;14:671–683.
128. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11:733–739.
129. Benito M, Parker J, Du Q, et al. Adjustment of systematic microarray data biases. *Bioinformatics*. 2004;20:105–114.
130. Reese SE, Archer KJ, Therneau TM, et al. A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. *Bioinformatics*. 2013;29:2877–2883.
131. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol*. 2017;35:498–507.
132. Luo J, Schumacher M, Scherer A, et al. A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *Pharmacogenomics J*. 2010;10:278–291.
133. Leigh DM, Lischer HEL, Grossen C, et al. Batch effects in a multiyear sequencing study: false biological trends due to changes in read lengths. *Mol Ecol Resour*. 2018;18:778–788.
134. Jolliffe I. Principal component analysis. In: Everitt B, Howell D, eds. *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons Ltd; 2005.
135. Nielsen F. Hierarchical clustering. In: *Introduction to HPC with MPI for Data Science. Undergraduate Topics in Computer Science*. Springer; 2016:195–211.
136. Hout MC, Papesh MH, Goldinger SD. Multidimensional scaling. *Wiley Interdiscip Rev Cogn Sci*. 2013;4:93–103.
137. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. *Nat Rev Methods Primers*. 2021;1:59.
138. Li C, Rabinovic A, Johnson WE, et al. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–127.
139. Chen C, Grennan K, Badner J, et al. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One*. 2011;6:e17238.
140. Gregori J, Villarreal L, Méndez O, et al. Batch effects correction improves the sensitivity of significance tests in spectral counting-based comparative discovery proteomics. *J Proteomics*. 2012;75:3938–3951.
141. Wang LJ, Ma XB, Xia HY, et al. Identification of biomarkers for predicting allograft rejection following kidney transplantation based on the weighted gene coexpression network analysis. *Biomed Res Int*. 2021;2021:9933136.
142. Wang Y, Wang JY, Schnieke A, et al. Advances in single-cell sequencing: insights from organ transplantation. *Mil Med Res*. 2021;8:45.
143. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–592.
144. Khankhanian P, Din L, Caillier SJ, et al. SNP imputation bias reduces effect size determination. *Front Genet*. 2015;6:30.
145. Kitchin R. Big data, new epistemologies and paradigm shifts. *Big Data Soc*. 2014;1:205395171452848.
146. Tierney NJ, Harden FA, Harden MJ, et al. Using decision trees to understand structure in missing data. *BMJ Open*. 2015;5:e007450.
147. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak*. 2016;16(suppl 3):74.
148. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28:112–118.
149. Cleveland WS, Devlin SJ, Cleveland WS, et al. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc*. 1988;83:596–610.
150. Carpenter J, Kenward M. Missing data in randomised controlled trials—a practical guide. 2007. Available at <https://researchonline.lshtm.ac.uk/id/eprint/4018500>. Accessed October 13, 2021.
151. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. John Wiley & Sons Inc.; 2002.
152. Rubin DB. Assignment to treatment group on the basis of a covariate. *J Educ Stat*. 1977;2:1–26.
153. Wei R, Wang J, Su M, et al. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep*. 2018;8:6663.
154. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 2007;26:51–78.
155. Sarwal MM, Benjamin J, Butte AJ, et al. Transplantomics and biomarkers in organ transplantation: a report from the first international conference. *Transplantation*. 2011;91:379–382.
156. Donders AR, van der Heijden GJ, Stijnen T, et al. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59:1087–1091.
157. Hrydzyszko O, Viant MR. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics*. 2012;8:161–174.
158. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010;11:499–511.
159. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449:851–861.
160. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
161. McCarthy S, Das S, Kretzschmar W, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–1283.
162. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 2018;103:338–348.
163. Fang Z. Integration and missing data handling in multiple omics studies. 2018. Available at <http://d-scholarship.pitt.edu/34596>. Accessed April 2, 2019.
164. Lin D, Zhang J, Li J, et al. An integrative imputation method based on multi-omics datasets. *BMC Bioinformatics*. 2016;17:247.
165. Napierala MA. What is the Bonferroni correction. *AAOS*. 2012;6:40.
166. Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20:467–484.
167. Yang J, Claas FHJ, Eikmans M. Genome-wide association studies in kidney transplantation: advantages and constraints. *Transpl Immunol*. 2018;49:1–4.
168. Dorr CR, Oetting WS, Jacobson PA, et al. Genetics of acute rejection after kidney transplantation. *Transpl Int*. 2018;31:263–277.

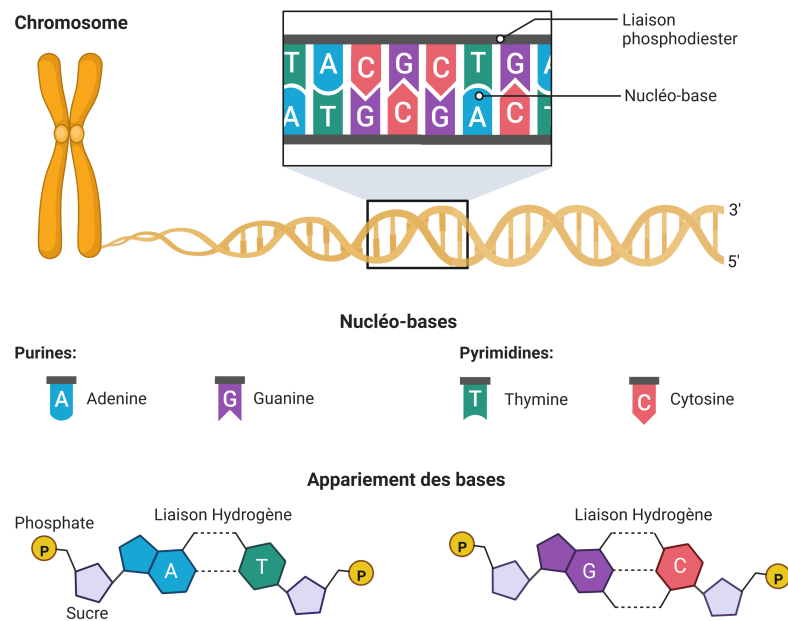
169. Jeong HJ. Diagnosis of renal transplant rejection: Banff classification and beyond. *Kidney Res Clin Pract.* 2020;39:17–31.
170. Almoguera B, Shaked A, Keating BJ. Transplantation genetics: current status and prospects. *Am J Transplant.* 2014;14:764–778.
171. Keating BJ, Van Setten J, Jacobson PA, et al. Design and implementation of the international genetics and translational research in transplantation network. *Transplantation.* 2015;99:2401–2412.
172. Shaked A, Chang BL, Barnes MR, et al. An ectopically expressed serum miRNA signature is prognostic, diagnostic, and biologically related to liver allograft rejection. *Hepatology.* 2017;65:269–280.
173. Stapleton CP, Birdwell KA, McKnight AJ, et al. Polygenic risk score as a determinant of risk of non-melanoma skin cancer in a European-descent renal transplant cohort. *Am J Transplant.* 2019;19:801–810.
174. Sanders ML, Karnes JH, Denny JC, et al. Clinical and genetic factors associated with cutaneous squamous cell carcinoma in kidney and heart transplant recipients. *Transplant Direct.* 2015;1:e13.
175. Oetting WS, Schladt DP, Dorr CR, et al; DeKAF Genomics and GEN03 Investigators. Analysis of 75 candidate SNPs associated with acute rejection in kidney transplant recipients: validation of rs2910164 in microRNA MIR146A. *Transplantation.* 2019;103:1591–1602.
176. Oetting WS, Schladt DP, Guan W, et al; DeKAF Investigators. Genomewide association study of tacrolimus concentrations in African American kidney transplant recipients identifies multiple CYP3A5 alleles. *Am J Transplant.* 2016;16:574–582.
177. Freedman BI, Pastan SO, Israni AK, et al. APOL1 genotype and kidney transplantation outcomes from deceased African American donors. *Transplantation.* 2016;100:194–202.
178. Sayadi S, Geffard E, Sudholt M, et al. Distributed contextualization of biomedical data: a case study in precision medicine. In: AICCSA 2020 – 17th International Conference on Computer Systems and Applications, Antalya, Turkey. November 2–5, 2020:1–6.
179. Cisek K, Krochmal M, Klein J, et al. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrol Dial Transplant.* 2016;31:2003–2011.
180. Kurian SM, Whisenant T, Mas V, et al. Biomarker guidelines for high-dimensional genomic studies in transplantation: adding method to the madness. *Transplantation.* 2017;101:457–463.
181. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association Studies (STREGA)—an extension of the STROBE statement. *Genet Epidemiol.* 2009; 33:581–598.
182. Hollenbach JA, Mack SJ, Gourraud PA, et al; Immunogenomics Data Analysis Working Group. A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement. *Tissue Antigens.* 2011;78:333–344.

## 2.2 Génomique de la transplantation rénale

### 2.2.1 Le génome humain

#### 2.2.1.1 L'ADN

L'ADN est un polymère composé de petites molécules répétées appelées nucléotides qui, grâce à une liaison phosphodiester, forment le brin d'ADN. Chaque nucléotide est composé d'un sucre (désoxyribose), d'un groupement phosphate et d'une des quatre nucléo-bases azotées (cytosine [C], guanine [G], adénine [A] ou thymine [T]). Un brin d'ADN est toujours lié à un autre brin selon les règles d'appariement des bases (A avec T et C avec G), formant une structure connue sous le nom de double hélice [136]. L'ADN a été isolé pour la première fois par Friedrich Miescher en 1869 [137], qui le décrit à l'époque comme une «nouvelle substance chimique» au sein des noyaux cellulaires dont la composition diffère de celle des protéines ou de toute autre molécule connue. Toutefois, ce n'est que près d'un siècle après que sa fonction et sa structure ont été entièrement déterminés, respectivement en 1944 grâce aux travaux de Avery et al. [138] et en 1953 avec les découvertes combinées de Rosalind Franklin [139, 140], Watson et Crick [141]. L'ADN cellulaire est réparti dans le noyau et dans les mitochondries. L'ADN mitochondriale est une molécule circulaire à double brin et représente une faible proportion. En effet, il se compose de 16 569 paires de bases et de 37 gènes dont 13 gènes codant pour des protéines, 22 gènes d'ARN de transfert et deux gènes d'ARN ribosomique [142, 143]. L'ADN humain est donc majoritairement nucléaire. Celui-ci est condensé sous la forme de 23 chromosomes hérités des deux parents [144] et contient environ 20.000 à 25.000 gènes [129, 130]. Toutefois, une étude très récente a réévalué et reprécisé ces chiffres sur le génome humain qui compte 3 055 milliards de paires de bases, et 19 969 gènes [145]. Les molécules d'ADN contiennent tout le matériel génétique d'un organisme. Cette information, lorsqu'elle est transcrite et traduite, produit les protéines qui sont vitales pour le fonctionnement, le développement et la reproduction d'un organisme.

FIGURE 2.1 – *Structure de l'ADN.\**

### 2.2.1.2 Gène et génome

Une grande partie de l'ADN (plus de 98 % chez l'homme) est non codante, ce qui signifie que ces sections ne sont pas impliquées dans la synthèse des protéines même si leurs rôles sont encore mal connus. La partie restante appelée partie codante représente les gènes qui sont à la base de la formation des protéines [144]. En effet, Les gènes contiennent un cadre de lecture ouvert qui peut être transcrit, en ARNm puis synthétisée en protéine. Ce processus représente le dogme central de la biologie moléculaire théorisée en 1958 par Francis Crick [146] qui décrit le flux de l'information génétique unidirectionnel comprenant 3 étapes : la réplication, la transcription et la traduction.

#### — La réplication de l'ADN

C'est le processus par lequel une molécule d'ADN double brin est copiée pour produire deux molécules d'ADN identiques lors de la division cellulaire. On parle de synthèse de l'ADN qui permet de copier exactement la même information génétique d'une cellule parent à ses cellules filles. La réplication de l'ADN est dite semi-conservative et bi-directionnelle. En effet, lorsque l'ADN parental double brin se sépare en deux, chaque brin sert de modèle pour un nouveau brin d'ADN complémentaire qui sera synthétisé

\* Figure créée avec BioRender.com.

dans les deux sens pour former deux nouvelles molécules d'ADN double brin constituées chacune d'un brin parental et d'un brin néoformé. La réplication est contrôlée par l'appariement des bases tel que décrit par Watson et Crick [141] et fait intervenir plusieurs enzymes dont les ADN polymérases. Elle comprend 3 étapes principales : l'initiation de la réplication, l'élongation ou la synthèse de l'ADN et la terminaison. Tout d'abord, l'initiation de la réplication a lieu à l'origine de réplication où se fixe une protéine dite initiatrice qui provoque une dénaturation locale des deux brins et déroule ainsi une courte portion de la double hélice d'ADN. Ensuite, une protéine appelée hélicase se fixe et rompt les liaisons hydrogène entre les bases des brins d'ADN, séparant ainsi les deux brins. Parallèlement, une autre enzyme appelée primase se fixe brièvement à chaque brin et crée une amorce qui marque le début de la zone de synthèse. Une fois l'amorce en place, l'ADN polymérase s'enroule autour de ce brin, et fixe de nouveaux nucléotides pour construire le nouveau brin complémentaire. Les nucléotides qui composent le nouveau brin sont appariés avec les nucléotides du brin matrice en respectant la correspondance A-T/G-C. On aboutit à la production de deux brins d'ADN complémentaires néoformés qui vont être liés aux brins matrices grâce à une ligase. La réplication s'arrête lorsque deux fourches de réplication se rencontrent ou lorsqu'une fourche rencontre un signal de terminaison (Figure 2.2).

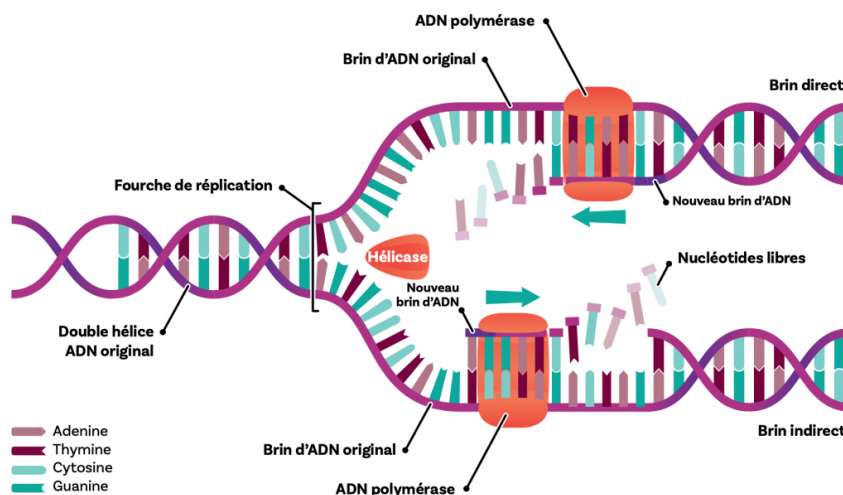


FIGURE 2.2 – *Réplication de l'ADN.* [147]

### — La transcription

Elle correspond à la synthèse de l'ARNm à partir de l'ADN. Elle se produit dans le noyau en utilisant l'ADN comme matrice pour produire l'ARN pré-messager (pré-ARNm) qui subira ensuite des modifications post-transcriptionnelles aboutissant à l'ARNm mature. La transcription débute par la séparation de l'ADN double brin grâce à l'hélicase qui

rompt les liaisons hydrogène, ce qui entraîne le déroulement d'une région de l'ADN (correspondant à un gène) exposant ainsi une série de bases. Bien que l'ADN soit une molécule à double brin, seul l'un des brins sert de modèle pour la synthèse du pré-ARNm ; ce brin est appelé le brin matrice et son brin complémentaire est appelé le brin codant. L'ARN polymérase se lie au brin matrice et transcrit la séquence dans le sens 3'-5'. La synthèse du pré-ARNm quant à elle, s'effectue dans le sens 5'-3'. Les bases de l'ADN sont échangées contre leurs bases correspondantes, sauf dans le cas de la thymine (T), à laquelle l'ARN substitue l'uracile (U) (Figure 2.3). L'ARN simple brin qui en résulte est le complément inverse de la séquence d'ADN d'origine et contient à ce stade des séquences non codantes (introns) et des séquences codantes (exons).

Afin d'arriver à maturation, la molécule de pré-ARNm subit des modifications post-transcriptionnelles telles que l'ajout d'une coiffe en 5' et d'une queue poly(A) en 3' ainsi que l'épissage des introns. Ces derniers, ne codant pas pour des protéines, sont retirés de la molécule de pré-ARNm par un complexe multiprotéique appelé spliceosome. L'ARNm mature ainsi formé est transporté hors du noyau pour être traduit en protéines [148].

#### — La traduction

Une fois dans le cytoplasme, l'ARNm va être traduit en protéines grâce à l'action des ribosomes qui sont des complexes de protéines et d'ARN dits ARN ribosomiques organisés en 2 sous unités (40S et 60S). Les ribosomes lisent la molécule d'ARNm dans le sens 5'-3' afin de synthétiser des chaînes polypeptidiques d'acides aminés pour former les protéines (Figure 2.3). Pour cela, le ribosome utilise de petites molécules, connues sous le nom d'ARN de transfert (ARNt). Il existe environ 60 types différents d'ARNt. Chaque ARNt se lie à une séquence spécifique de trois nucléotides (appelée codon) dans la molécule d'ARNm et délivre un acide aminé spécifique. Ce processus est basé sur le code génétique qui permet la correspondance entre les codons et les acides aminés. Le ribosome se fixe initialement à l'ARNm au niveau du codon dit « start » (AUG) correspondant à la méthionine afin de commencer la biosynthèse de la chaîne polypeptidique. Ensuite, il traduit la séquence d'ARNm codon par codon (translocation) et progresse le long de l'ARNm sous l'action de facteurs d'élongation. La synthèse de la protéine s'achève lorsque le ribosome atteint un codon dit « stop » (UAA, UGA ou UAG). Le ribosome se détache de la protéine et du brin d'ARN messenger. La nouvelle protéine est libérée dans la cellule.

A l'issue de la traduction, ces nouvelles protéines subissent des modifications de structure pour avoir une conformation en 3D. Ils peuvent également subir des modifications post-traductionnelles (liaisons covalentes de groupements phosphates, acétate ou méthyle par exemple). Il existe plus de 200 types connus de modifications post-traductionnelles. Ces modifications peuvent altérer l'activité des protéines, leur capacité à interagir avec



d'autres protéines et leur localisation cellulaire [149].

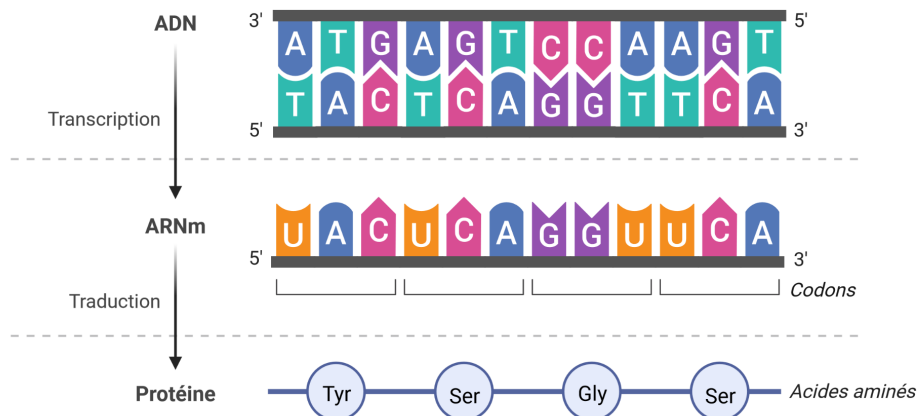


FIGURE 2.3 – *Les étapes de la synthèse des protéines.\** Lors de la transcription, l'ADN est copié en ARNm. Les bases de l'ADN sont échangées contre leurs bases correspondantes, sauf dans le cas de la thymine (T), à laquelle l'ARN substitue l'uracile (U). L'ARNm simple brin qui en résulte est le complément inverse de la séquence d'ADN d'origine. L'ARNm est transporté hors du noyau pour être traduit en protéines grâce aux ribosomes. La traduction est basée sur le code génétique qui permet la correspondance entre les codons (séquence de trois nucléotides d'ARN) et les protéines.

### 2.2.1.3 Séquençage du génome humain

Le séquençage d'un génome consiste en la détermination de la séquence nucléotidique de l'ADN présent dans chaque cellule d'un organisme donné. Dès la fin des années 1970, les techniques de la biologie moléculaire popularisée par Sanger permettaient de séquencer un brin d'ADN, c'est-à-dire de lire l'enchaînement de nucléotides d'une molécule d'ADN [150, 151]. Le séquençage du génome humain n'a été envisagée qu'au début des années 90 avec le projet génome humain qui avait pour objectif d'établir le séquençage complet de l'ADN du génome humain sur une période de 15 ans [152, 153]. Une première version brute du génome a été publiée en 2001 [129, 130] et a été très vite suivie par une séquence plus complète en 2003 [154]. Ce premier séquençage a été un tremplin pour le domaine de la génomique et constitue le point de départ de divers autres projets. A titre d'exemple, on peut citer le 1KGP [132, 155], lancé en 2008 qui s'est donné pour objectif au départ de séquencer les génomes d'un millier de personnes réparties dans 5 grandes populations afin d'améliorer la cartographie génétique du génome humain. En

\* Figure créée avec BioRender.com.

2015, le projet présentait la cartographie génétique de 2504 individus issues de diverses origines ethniques [132]. Avec l'évolution des technologies, le 1KGP compte maintenant 3 202 génomes dont 602 trios familiales (parents-enfants) séquencées avec une couverture de 30X [155]. Plus récemment, les projets sont encore plus ambitieux et se donnent le défi de séquencer jusqu'à 100.000 génomes [156]. Il est important de noter que la séquence génomique établie par le projet génome humain en 2003 n'était pas tout à fait complète avec 8% de l'ADN manquant. Il s'agissait principalement de l'ADN condensé au niveau des centromères et des télomères caractérisé par une quantité importante de séquences répétées. En 2022, grâce à l'effort combiné d'une centaine de chercheurs, le premier génome complet d'un même individu a été publié après 3 ans de travaux. Cette étude a réévalué la composition du génome et mis en évidence 3604 nouveaux gènes précédemment méconnus [145].

### 2.2.1.3.1 Technologies de séquençage

Les méthodes initiales de séquençage sont principalement basées sur les travaux de Sanger et Gilbert qui permettaient de lire des séquences d'ADN respectivement par polymérisation de l'ADN grâce à une amorce et par dégradation chimique de l'ADN basée sur les affinités des 4 bases (A,T,C,G). Ces méthodes étaient très efficaces pour lire des petites séquences mais étaient inadaptées pour le séquençage d'un génome entier qui compte des milliards de bases [150, 151]. Des méthodes qui couplent les approches de biologies moléculaires et les technologies informatiques ont été développées : la méthode de séquençage par ordonnancement hiérarchique qui consiste à fragmenter l'ADN génomique puis à classer ces fragments avant de les séquencer et la méthode globale (ou « whole-genome shotgun ») qui consiste à séquencer les fragments d'ADN génomiques aléatoirement et de les réordonner par chevauchement de séquences communes [151]. La principale différence entre ces deux principes est que l'ordonnancement hiérarchique essaie d'aligner un jeu de clones de grande taille (100 kb environ) alors que la méthode globale se base sur des fragments aléatoires de petite taille qui sont séquencés puis alignés. D'autres méthodes, notamment les techniques de génotypage ont été par la suite développées. Elles reposent principalement sur l'utilisation de puces à ADN contenant plusieurs centaines (pour les puces de première génération) à plusieurs milliers d'oligonucléotides. Le type de séquençage le plus couramment utilisé aujourd'hui est toutefois le séquençage par synthèse qui se base sur des méthodes de fluorescence [151]. Ce type de séquençage, utilise une ADN polymérase pour générer un nouveau brin d'ADN à partir d'un brin d'intérêt en incorporant des nucléotides individuels qui ont été marqués chimiquement avec un marqueur fluorescent. Le nucléotide est alors excité par une source lumineuse, et un signal fluorescent est émis et détecté. Le signal est différent selon celui des quatre nucléotides qui a été incorporé.

Le projet génome humain [152, 153], sur la base de la méthode globale de séquençage, a été réalisé grâce à l'utilisation de séquenceurs dits de première génération ou Sanger répartis dans



plusieurs laboratoires [129, 130]. Le projet a duré 13 ans (1990-2003) et a coûté 3 milliards de dollars [157]. Avec l'avènement de la biotechnologie et l'amélioration continue des méthodes de séquençages, on parle aujourd'hui de séquençage nouvelle génération (NGS) qui permettent de séquencer en quelques jours un génome humain à moins de 1000 dollars selon la technologie [157]. La différence avec les méthodes de première génération repose sur la capacité de lire en parallèle des millions de molécules d'ADN. Il existe plusieurs technologies de séquençage NGS dépendantes du fabricant : technologie 454 ou pyroséquençage chez Roche, séquençage par ligation chez Applied Biosystem, séquençage par synthèse chez Illumina etc [158, 157]. De façon général, ces technologies génèrent des séquences courtes de nucléotides (150-300 pb) appelées lectures courtes « ou short reads » qui sont par la suite assemblées pour former des séquences d'ADN plus longues comme un gène par exemple. Dans ce processus, chaque base doit être lue non pas une seule fois, mais au moins plusieurs fois dans les segments qui se chevauchent afin de garantir la précision. Il a été démontré que, malgré l'utilisation d'algorithmes bioinformatiques sophistiqués, il est souvent impossible de cartographier avec précision, ou même d'assembler, des lectures courtes provenant de régions présentant une variation, des séquences répétitives, un contenu dense en guanine-cytosine (GC) ou des séquences comportant de multiples éléments homologues au sein du génome [159]. Des nouvelles technologies basées sur des lectures longues (>10 kb) « ou long reads » sont développées et ont démontré de meilleures performances dans l'identification de séquences d'ADN auparavant difficiles à identifier [159] (Figure 2.4). Ces méthodes sont notamment développées par Pacific Biosciences (PacBio) qui a lancé le séquençage SMRT (« Single Molecule Real-Time ») basé sur l'utilisation d'une polymérase attachée au fond d'un micro-puit et qui synthétise un brin complémentaire à partir de molécules d'ADN simple brin circulaires. Cette technologie utilise la fluorescence pour le marquage des nucléotides comme dans la technologie Illumina, mais avec pour différence importante que la synthèse ne s'arrête pas après l'incorporation de chaque nucléotide, elle continue jusqu'au moment où la polymérase s'arrête spontanément créant ainsi des lectures plus longues [159].



FIGURE 2.4 – **Avantage des approches de séquençage à lectures longues (LRS).** [159] La principale différence entre le LRS et les approches NGS conventionnelles est l'augmentation significative de la longueur des lectures. Contrairement aux lectures courtes (150-300 pb), le LRS a la capacité de séquencer en moyenne plus de 10 kb en une seule lecture, nécessitant ainsi moins de lectures pour couvrir le même gène (illustré dans le panneau supérieur). Par conséquent, outre la réduction des erreurs d'alignement et de cartographie, la technologie LRS présente divers avantages par rapport aux approches à lecture courte qui peuvent avoir un impact considérable sur la génétique médicale (panneau inférieur). (1) Amélioration de la détection et de la caractérisation des grandes variations structurales (SV), dues, par exemple, à de grandes inversions ou translocations. (2) Capacité de séquencer directement, et donc avec plus de précision, de longues expansions de répétitions en tandem et des régions extrêmement riches en GC. (3) Amélioration de l'attribution des variantes génétiques aux chromosomes paternels ou maternels homologues, afin de déterminer les schémas d'hérédité, l'origine parentale des événements de novo, le mosaïcisme, l'expression spécifique des allèles et les haplotypes de risque de maladie. (4) Meilleure discrimination des gènes cliniquement pertinents de leurs pseudogènes.

### 2.2.2 La transmission génétique

Les gènes sont présents en deux copies dans les chromosomes et peuvent influencer une caractéristique particulière dans un organisme (phénotype). Les différentes formes d'un même gène sont appelées allèles. Ces allèles peuvent être différents ou identiques et proviennent des deux parents. Les gènes représentent donc l'unité de l'hérédité qui se définit comme la transmission,

au sein d'une espèce vivante ou d'une lignée de cellules, de caractéristiques d'une génération à la suivante [160]. Cette théorie de la transmission des gènes a été popularisée grâce aux travaux de Gregor Mendel sur la génétique des populations. Les mécanismes de l'hérédité comme démontré par Mendel comportent principalement un volet génétique (transmission de l'ADN nucléaire à travers les générations) [161]. Toutefois, la transmission des caractères peut se faire à travers d'autres mécanismes qui impliquent de l'ADN non nucléiques (ex : ADN mitochondriale) ou l'épigénétique qui englobe les mécanismes de modifications réversibles mais transmissibles de l'expression d'un gène [160, 161].

### 2.2.2.1 Théorie de l'hérédité de Mendel

Les travaux de Mendel sur les lois de l'hérédité constituent la base de la génétique moderne. Ils sont basés sur le principe selon lequel les traits individuels sont définis par les gènes et sont hérités de générations en générations. Pour illustrer cela, Mendel étudie la transmission de sept caractéristiques chez les pois tels que la hauteur, la couleur etc. Il sélectionne des lignées de pois présentant deux formes différentes d'un même trait (ex : grande vs petite taille) qu'il cultive sur plusieurs générations jusqu'à obtenir des lignées pures dont tous les individus possèdent toujours la même forme alternative c'est à dire un seul type d'allèle pour le caractère étudié. Il croise ensuite ces lignées pures entre elles et observe leur transmission en notant leurs caractéristiques et la proportion d'individus présentant chaque caractéristique [160].

Sur la base de ces observations, Mendel définit 3 lois sur la transmission des gènes [162, 160]

1. **La loi de dominance (ou loi d'uniformité des hybrides de première génération)**. Les traits sont contrôlés par des paires de facteurs héréditaires (les allèles) qui apparaissent sous différentes versions. Par exemple pour les petits pois, le caractère "taille" va avoir deux formes (grande ou petite) qui vont chacune être déterminées par un allèle. L'allèle qui sera phénotypiquement visible est appelé allèle dominant (grande taille) et l'allèle qui ne sera pas exprimé physiquement est appelé allèle récessif (petite taille). Lorsque l'on croise deux individus d'une même espèce homozygote ayant un caractère en commun (petite taille vs grande taille), les descendants obtenus à la première génération sont identiques relativement à ce caractère tant au niveau du phénotype que du génotype, on parle de la loi d'uniformité de première génération. Ces individus sont tous hétérozygotes pour le gène de la taille et présentent physiquement l'allèle dominant pour le caractère étudié, l'allèle récessif étant "caché". En reprenant notre exemple, tous les individus de la première génération seront de grandes tailles.
2. **La Loi de la ségrégation (ou loi de disjonction des allèles)**. Lorsque les descendants de la première génération sont recroisés entre eux, les deux phénotypes différents apparaissent selon un ratio 3 :1. L'allèle récessif est exprimé physiquement chez une mi-

norité d'individus. Plus précisément, on obtient  $3/4$  des individus qui présentent le trait dominant ( $2/4$  hétérozygotes et  $1/4$  homozygotes) et  $1/4$  des individus qui présentent le trait récessif (homozygotes). Ce phénomène est défini comme la ségrégation des caractères dans la deuxième génération.

3. **La loi sur l'assortiment indépendant (ou loi d'indépendance de la transmission des caractères)**. Les études de Mendel ont également montré que lorsque l'on étudie deux caractères (ex taille et couleur), la transmission des traits est indépendante c'est à dire que le caractère taille n'influence pas la transmission du caractère couleur. Cette règle s'applique lorsque les gènes responsables des caractéristiques se situent sur différents chromosomes ou s'ils sont éloignés sur le même chromosome. La première loi s'applique à la première génération dans la transmission de plusieurs caractères : tous les descendants sont identiques et sont hétérozygotes pour les caractères étudiés. Dans la deuxième génération, les phénotypes apparaissent selon un ratio 9 :3 :3 :1.

Les caractères ne sont pas tous exclusivement dominants ou récessifs. Dans certains cas, les allèles n'ont pas une relation de dominance pour un caractère donné. Les individus ayant ce caractère sont hétérozygotes et présentent un phénotype hybride (mélange des deux formes du caractère). Ce principe est appelé la dominance incomplète ou co-dominance [162, 160].

### 2.2.2.2 Transmission génétique et maladies

Les lois de Mendel ont permis d'établir les bases de la transmission des caractères chez l'homme. En ce sens, cela a également permis de mieux comprendre les mécanismes de transmission des maladies monogéniques ou mendéliennes. Ces maladies ont pour cause des anomalies génétiques qui ne concernent qu'un seul gène et sont transmises au sein d'une même famille selon quatre modes de transmissions : dominant ou récessif lié aux autosomes et dominant ou récessif lié aux chromosomes sexuels (X ou Y) [163].

La transmission autosomique dominante est la plus fréquente (>50% des maladies mendéliennes). La présence d'un seul allèle dominant provoque la pathologie qui se transmet de façon directe ou verticale d'une génération à une autre. Les individus portant ce genre de maladies sont généralement hétérozygotes pour la maladie, les homozygotes étant très rares [163]. Un exemple d'une maladie autosomique dominante est la polykystose rénale, une maladie génétique qui se caractérise par la formation progressive de kystes (poches de liquide) pouvant conduire à une insuffisance rénale. Elle est la conséquence de mutations génétiques sur le gène *PKD1* (80% des cas) et le gène *PKD2* (20% des cas) et touche 1/1000 personne en France. Un individu atteint de cette pathologie a une chance sur deux de transmettre la maladie à sa descendance [164].

La transmission autosomique récessive nécessite la présence de deux allèles récessifs pour que la pathologie se manifeste avec une chance sur 4 de transmettre la maladie. Les individus atteints sont exclusivement homozygotes. Un exemple d'une maladie autosomique récessive est la drépa-

nocytose qui est une maladie génétique du sang caractérisée par une anomalie de l'hémoglobine qui joue un rôle dans le transport de l'oxygène vers les tissus. La drépanocytose est due à une mutation du gène *HBB*. Pour des parents atteints de cette pathologie, chacun de leurs enfants a une chance sur 4 d'avoir la drépanocytose [165].

La transmission liée au chromosomes sexuels concerne principalement le chromosome X, le chromosome Y n'étant présente que chez les hommes en une seule copie comprenant un nombre de gènes très restreint (71 gènes [166, 167]). La transmission dominante liée au chromosome X est plus fréquente chez les femmes que chez les hommes qui héritent de l'allèle pathologique exclusivement de leur mère. La maladie se transmet de façon directe de génération en génération. Un exemple est le rachitisme vitamino résistant hypophosphatémique qui est une forme de rachitisme résistante au traitement de la vitamine D. Cette pathologie est associée à la mutation du gène *PHEX* présent sur le chromosome X [168]. Les femmes présentant la pathologie sont plus fréquentes mais les hommes porteurs de la maladie sont plus gravement atteints car ils n'ont pas d'allèle sain. La transmission récessive liée au chromosome X est quant à elle plus fréquente que celle dominante. Elle concerne plus fréquemment les hommes dans la mesure où les femmes ont deux chromosomes X tandis que les hommes n'en possèdent qu'un seul : ils sont dit hémizyotes. Comme exemple, nous pouvons citer l'hémophilie qui est une pathologie hémorragique due à l'absence ou au déficit d'un facteur de coagulation [169]. Il existe deux formes d'hémophilie liées à une mutation du chromosome X : l'hémophilie A et B qui sont dues aux mutations des gènes *F8* et *F9* respectivement. Les porteurs de l'allèle récessif sont atteints par la pathologie. L'hémophilie concerne presque exclusivement les hommes. Les femmes sont souvent conductrices de la maladie car elles possèdent deux chromosomes X et sont dans la grande majorité des cas hétérozygotes pour la maladie. Il peut arriver qu'elles héritent de deux chromosomes X portant chacun un gène muté ce qui est très rare et généralement les individus sont non viables [169].

Certaines maladies, bien que génétiques, font défaut aux lois de Mendel et ne correspondent à aucun des modes mendéliens classiques de transmission : il s'agit des maladies complexes ou multifactorielles caractérisées par l'interactions de plusieurs facteurs génétiques mais également environnementaux [170]. L'héritabilité des maladies complexes implique plusieurs gènes (déterminisme polygénique). Ces gènes pris individuellement n'influent pas sur la survenue de la maladie à l'inverse des maladies mendéliennes, mais leur action combinée peut influencer la susceptibilité à une maladie (effets additifs, effets multiplicatifs d'un grand nombre de gènes). Elles sont très fréquentes, en augmentation constante et concernent une diversité de maladies tels que les cancers, les maladies cardiovasculaires, les maladies neuro-dégénératives, les maladies auto-immunes, les maladies rénales, les diabètes etc. [171, 172].

Au vu de la complexité des maladies polygéniques, l'identification des gènes associés à ces pathologie nécessite l'adoption de stratégies et de méthodes d'études différentes. La disponibilité d'outils techniques et de nouvelles méthodologies statistiques ont nettement contribué à l'avan-

cée des études des maladies complexes à travers la réalisation d'études d'association basées sur des approches populationnelles [173]. Classiquement, les études familiales notamment pour les maladies mendéliennes nécessitent l'inclusion d'un maximum d'individus apparentés atteints et non atteints pour une maladie donnée. A l'inverse, les études populationnelles s'intéressent aux individus non apparentés atteints d'une même maladie avec l'inclusion d'un nombre au moins équivalent ou supérieur de sujets témoins (sains) et la prise en compte de leur origine ethnique et des principaux facteurs de risque connus de la maladie à l'étude. Ces approches populationnelles, ont mené à l'identification d'un grand nombre de gènes de susceptibilité pour la plupart des maladies complexes [173]. Toutefois, seule la composante génétique associée aux variants communs est étudiée. Les variants génétiques rares, ou les facteurs épigénétiques sont moins connus et représentent encore une zone d'ombre pour la majorité de ces maladies complexes [174].

### 2.2.2.3 Les polymorphismes génétiques

La diversité génétique est décrite comme étant la variété des gènes au sein d'une même espèce. Elle constitue la matière première de l'évolution car essentielle à la survie d'une espèce [175]. Au cours des dernières années, les progrès dans la génétique populationnelle et la génomique ont permis de mieux appréhender cette diversité des gènes. Des projets comme le SNP Consortium [176], le HapMap Consortium [177] et plus récemment le 1KGP [132] ont vu le jour afin de répertorier et cartographier chez l'Homme les polymorphismes génétiques tels que les insertions/délétions (indels), les variations du nombre de copies d'un segment dans le génome (CNV), et les SNP. Les polymorphismes génétiques sont définis par l'existence d'au moins deux allèles dans une population à une fréquence égale ou supérieure à 1%. Ces variations entre individus estimées à 0,1% du génome, contribuent grandement à la diversité génétique et peuvent influencer la prédisposition à certaines maladies ou être à l'origine de certains traits [178].

#### 2.2.2.3.1 Les polymorphismes nucléotidiques simples (SNP)

Un SNP définit la variation d'une seule paire de base dans la séquence nucléotidique. Ils sont génétiquement stables et majoritairement bi-alléliques. Les SNP représentent le type de variations génétiques la plus fréquente (plus de 90% de l'ensemble des polymorphismes génétiques chez l'Homme) et la plus uniformément répartie dans le génome [178, 179]. La base de données dbSNP [180] recense l'ensemble des SNP et indels connus. Du fait de leurs propriétés, les SNP permettent de dresser des cartes génétiques très denses et représentent l'actuel marqueur génétique de prédilection. L'analyse de ces marqueurs génétiques permet d'expliquer certains traits mais également de trouver des prédispositions à certaines maladies complexes des approches GWAS [133].

### 2.2.2.3.2 Les insertions/délétions (indels)

Les indels sont des polymorphismes caractérisés par des insertions et/ou des délétions de séquences courtes (<1kbs) dans le génome. Lorsque l'indel se produit à l'intérieur d'une région codante, il est décrit comme étant « in-frame » si la quantité d'ADN perdue ou gagnée est divisible par 3 correspondant à un ou plusieurs codons. Si non, il est appelé « frameshift » car le code de lecture par codons est altéré pour tous les nucléotides suivants. Les indels « frameshift » conduisent fréquemment à des codons « stop » prématurés et ont généralement un impact fonctionnel plus important que les indels « in-frame » [181]. Ces polymorphismes sont, de la même manière que les SNP, impliqués dans l'apparition de traits et de maladies chez l'Homme [182, 183]. Ils sont d'ailleurs couramment identifiés dans la plupart des cancers, bien qu'ils soient moins fréquents que les SNP [184]. .

### 2.2.2.3.3 Variation du nombre de copies (CNV)

Les CNVs sont décrits comme étant des variations du nombre de copies d'une région génomique selon les individus [185]. Ces variations sont des séquences supérieures à 1kb ayant subi une délétion, une duplication ou une inversion de fragments d'ADN et sont estimées à environ 9,5% du génome [186]. Les CNVs sont impliqués dans l'évolution et l'adaptation des espèces. En effet, la variation du nombre de copies d'un gène influence la quantité de protéines produites ce qui implique des changements phénotypiques. En 2006, une étude s'est penchée sur la question pour mieux comprendre les conséquences des CNVs sur la régulation de l'expression des gènes ou l'altération de la séquence codante [187]. Les CNVs sont référencés dans la base de données dbVar [188].

### 2.2.2.4 Notion de génétique des populations

#### 2.2.2.4.1 Équilibre d'Hardy-Weinberg

L'équilibre de Hardy-Weinberg [189] est l'un des principes fondamentaux en génétique des populations basé sur le comportement des fréquences alléliques et génotypiques pour un polymorphisme bi-allélique (les SNP). Il stipule que ces fréquences alléliques restent constantes dans une population et au fil des générations, sous certaines hypothèses :

1. La population d'étude est conséquente ou infinie.
2. Les générations sont non chevauchantes.
3. Le croisement entre individus est aléatoire (panmixie)
4. L'absence de sélection, de mutation ou de migration des populations est avérée.

Selon ces hypothèses, le modèle s'établit de la manière suivante :

Dans le cas d'un SNP A bi-allélique, d'allèles  $a_1$  et  $a_2$ , et de fréquences alléliques respectives  $f_{a_1}$  et  $f_{a_2}$ , on observe que :

- les fréquences alléliques et génotypiques sont constantes au fil des générations
- les fréquences génotypiques suivent la distribution suivante :

$$\begin{cases} f_{a_1 a_1} &= f_{a_1}^2 \\ f_{a_1 a_2} &= 2f_{a_1} f_{a_2} \\ f_{a_2 a_2} &= f_{a_2}^2 \end{cases}$$

Lorsque chacune des hypothèses est respectée, on dit que le modèle est à l'équilibre. Dans le cas où celles-ci sont invalidées, on parle de déviation de l'équilibre d'Hardy-Weinberg qui peut être due à des biais démographiques (migration de population, réduction importante et rapide de l'effectif d'une population, consanguinité etc.) ou liée à des événements de sélection naturelle [190].

#### 2.2.2.4.2 Déséquilibres de liaison

La recombinaison génétique correspond aux brassages intra et interchromosomiques conduisant à l'apparition aléatoire de gènes ou de caractères héréditaires chez les individus. Ce phénomène se produit par enjambement des chromosomes homologues durant la méiose. Lorsque deux polymorphismes sont suffisamment éloignés dans le génome, ils peuvent être transmis de manière indépendante du fait de la recombinaison. Si nous considérons des loci polymorphes indépendants, les fréquences des recombinaisons d'allèles possibles sur un chromosome dans une population, correspondent alors au produit des fréquences de ces allèles, c'est l'équilibre de liaison.

Soit deux locus  $A$  et  $B$ , d'allèles  $A/a$  et  $B/b$ , et de fréquences  $f_A/f_a$  et  $f_B/f_b$ . Pour 4 combinaisons possibles d'allèles, les fréquences de ces combinaisons sont établies comme suit :

$$\begin{cases} f_{AB} &= f_A f_B + D \\ f_{Ab} &= f_A f_b - D \\ f_{aB} &= f_a f_B - D \\ f_{ab} &= f_a f_b + D \end{cases}$$

Cependant, lorsque la distance qui sépare deux loci chromosomiques est trop faible, la probabilité pour que les polymorphismes soient transmis de manière indépendante diminue. Ces



polymorphismes présentent donc une association non aléatoire et seront transmis ensemble. Les fréquences des combinaisons d'allèles possibles sont alors différentes du produit des fréquences alléliques, c'est le déséquilibre de liaison. D'autres facteurs, tels que la pression de sélection peuvent également contribuer à une interdépendance. On parle donc plus généralement de déséquilibre gamétique.

Le déséquilibre de liaison peut se mesurer par la valeur du coefficient  $D$  de déviation entre les fréquences des combinaisons observées et celles attendues sous l'hypothèse d'indépendance entre les loci. Il s'établit comme suit :

$$\begin{cases} f_{AB} &= f_A f_B \\ f_{Ab} &= f_A f_b \\ f_{aB} &= f_a f_B \\ f_{ab} &= f_a f_b \end{cases}$$

Lorsque  $D = 0$ , il n'y a pas de déséquilibre gamétique observable. On utilise alors une valeur de  $D$  normalisée :  $D'$  variant entre -1 et 1 :

$$D' = \begin{cases} \frac{D}{\min(f_{AB}, f_{ab})} \text{ pour } D \geq 0 \\ \frac{D}{\min(f_{Ab}, f_{Ba})} \text{ pour } D \leq 0 \end{cases}$$

Il existe une troisième mesure qui est celle la plus largement utilisée et qui se base sur un coefficient de corrélation  $r^2$ , défini comme :

$$r^2 = \frac{D^2}{f_A f_B f_a f_b}$$

$r^2$  varie entre 0 et 1 ;  $r^2 = 1$  indique un déséquilibre total entre les polymorphismes.

### 2.2.3 Les études d'association génétiques

La cartographie des polymorphismes génétiques a permis entre autres choses de pouvoir lever des zones d'ombres dans les études transdisciplinaires reliant variations génétiques et maladies [133]. Ces études ont pour but d'étudier statistiquement les liens entre un polymorphisme

génétique et un phénotype particulier (expression du gène, fonction de la protéine associée ou événement clinique) [133].

### 2.2.3.1 Etude gène candidat

Au début des années 2000, les études d'associations s'intéressaient majoritairement à des polymorphismes de gènes. Ces études consistaient à caractériser les variants génétiques dans des gènes dits « candidats » déjà décrits dans la littérature et dont la fonction pourrait jouer un rôle dans le risque d'apparition d'une pathologie [191]. Le choix des gènes candidats était basé sur une bonne connaissance de la physiopathologie, de la fonction du gène ainsi que de ses interactions avec d'autres variants impliqués dans la même voie métabolique et qui pourraient être associés à la pathologie étudiée. L'identification des variants génétiques se fait par séquençage direct de l'ADN de sujets atteints et de sujets non atteints pour le/les gènes ciblés. Ensuite, par des méthodes statistiques, on détermine si le variant génétique est plus fréquent chez les sujets présentant la maladie que chez les témoins [191]. De façon générale, chaque variant génétique ne possède qu'un effet minime sur la variation du trait phénotypique et donc sur la survenue ou non de la maladie. De plus, en l'absence d'informations fonctionnelles objectives (essais expérimentaux biologiques en laboratoire) sur l'impact précis du variant génétique sur la fonction du gène et sur la maladie, on ne parlera que d'association significative au sens statistique entre la présence du variant génétique et la présence de la maladie [192]. Ces approches gène candidat ont mis en évidence de nombreux variants génétiques impliqués dans des maladies [192].

En transplantation rénale, certaines équipes ont étudiés des gènes candidats chez le receveur tels que le récepteur de type 1 à l'angiotensine II, impliqué dans les voies de signalisation de l'hypertension artérielle [193] ou encore *MICA* (Major-histocompatibility-complex class- I-related chain A), un gène de la région *HLA* codant pour des glycoprotéines de surface [194]. Toutefois, cette stratégie est limitée par le fait de ne pouvoir cibler et analyser que les gènes codants déjà décrits dans la littérature, dont la fonction est connue et qui présente un rapport à priori évident avec la pathologie. Ce qui exclut tous les gènes ou protéines inconnus ou dont le rôle est insoupçonné. Des études d'associations plus récentes qui se font à l'échelle du génome et sans a priori ont permis de mieux caractériser le rapport entre variant génétique et maladie [191, 192].

### 2.2.3.2 Etudes d'association à l'échelle du génome (Genome Wide Association Studies, GWAS)

L'ère de la génomique a été marquée par l'apogée des études GWAS qui ont permis d'identifier des centaines de milliers de SNP associés à un trait ou à une maladie. Ces approches évaluent, à l'aide de tests statistiques, la différence de fréquence allélique entre populations par rapport à un phénotype donné [195].

Les études GWAS ont révolutionné le domaine de la génétique des maladies complexes au cours

de la dernière décennie. Depuis la publication de la première étude GWAS sur la dégénérescence maculaire liée à l'âge en 2005, plus de 5 675 études et 202 969 SNP associés ont été rapportés dans le catalogue GWAS à ce jour [133]. Ces associations ont permis de mieux comprendre les mécanismes de la susceptibilité aux maladies (par l'identification de nouveaux gènes causaux) et de faire progresser les soins cliniques et la médecine personnalisée (identification de biomarqueurs et de nouvelles cibles médicamenteuses, prédiction du risque et l'optimisation des thérapies en fonction du génotype etc.) [195]. Parmi les maladies et traits étudiés en GWAS, on retrouve notamment l'anorexie mentale [196], le trouble dépressif majeur [197], les cancers et les sous-types de cancers [198], le diabète de type 2 [199], les maladies inflammatoires de l'intestin [200], l'insomnie [201], l'indice de masse corporelle (IMC) [202] etc. Les loci identifiés en GWAS impliquent souvent des gènes dont la fonction était inconnue ou dont la pertinence était jusqu'alors insoupçonnée. Le suivi expérimental de ces loci peut conduire à la découverte de nouveaux mécanismes biologiques sous-jacents à la maladie. Par exemple, le rôle de l'autophagie dans la maladie de Crohn a été mis en évidence avec l'identification de SNP associés au risque de maladie dans les gènes *ATG16L1* (rs2241880) et *IRGM* (rs1000113) [203]. Deux autres exemples de l'application des découvertes en GWAS à la biologie sont (1) l'association entre le gène *CFH* et la dégénérescence maculaire, qui met en jeu le système du complément de l'immunité innée [204] (2) et l'association entre le locus du complexe majeur d'histocompatibilité et la schizophrénie, qui met en évidence le rôle de l'activité du composant 4 du complément [204, 205]. Dans le domaine de la transplantation, les études GWAS ont permis de mieux comprendre les interactions entre donneur et receveur [206]. L'implication des facteurs génétiques dans la tolérance et le rejet des greffes a été prouvée au début des années 1950, bien avant la première transplantation réussie d'un organe solide chez l'homme [207]. Récemment, les études d'association ont permis d'étudier les mécanismes moléculaires qui régissent la perte du greffon rénal notamment avec le développement de nouvelles stratégies innovantes qui prennent en compte les génomes du donneur et du receveur simultanément.

Plusieurs études ont porté sur la compréhension du rejet aigu (RA). En 2016, Ghisdal et al. [208] ont réalisé la toute première GWAS du rejet aigu en transplantation rénale et ont identifié deux loci significatifs : (1) *PTPRO*, codant pour un récepteur de la tyrosine kinase, essentiel à la signalisation du récepteur des cellules B (SNP rs10846175 et rs7976329) et (2) *CCDC67* codant pour le gène ciliaire *CCDC67* (SNP rs10765602). Ces deux gènes ont été considérés comme des gènes candidats intéressants car *PTPRO* pourrait moduler l'alloréactivité [209] (régulation de la signalisation *TCR* et *BCR* et de la production de cytokines) et le gène *CCDC67* pourrait jouer un rôle au niveau de la synapse immunitaire ; des études récentes indiquent des similitudes étroites entre le cilium primaire et la synapse immunitaire [210]. Deux ans plus tard, Hernandez-Fuentes et al. [211] publient une étude basée sur plusieurs cohortes et ont déterminé les potentiels associations des SNP du donneur, des SNP du receveur et des SNP donneur-receveur combiné

avec la survie du greffon (échec de la greffe ou décès du patient avec une durée médiane de suivi d'environ 5 ans) et le rejet aigu (événement de RA enregistré dans les 12 mois suivant la transplantation). Malgré la taille importante de leurs cohortes (découverte = 441 cas contre 941 témoins, réplication = 575 cas contre 2573 témoins), ils n'ont pas réussi à répliquer les résultats de Ghisdal et al. et n'ont identifié aucune association significative. Il est important de noter que ces résultats ne sont basés que sur des cohortes d'ascendance européenne, ce qui suggère que leurs conclusions ne sont pas transposables à d'autres populations. Le risque de RA est désormais bien géré, grâce au développement continu de nouveaux agents immunosuppresseurs (10 % de RA en France en 2020) [212]. Cependant, la perte chronique du greffon reste un défi, car le taux de survie du greffon dans les 10 ans suivant la transplantation est de 56 %, même avec un donneur HLA compatible à 100% [116]. En 2013, O'Brien et al. [213] ont réalisé une GWAS incluant 326 transplantations rénales de receveurs irlandais avec un suivi à long terme. Ils ont identifié deux SNP (rs3811321 et rs6565887) qui étaient associés aux taux de créatinine sérique à 5 ans. Cependant, une étude de validation réalisée par Pihlstrøm et al. [214] en 2016 sur 1 600 greffons rénaux provenant de receveurs caucasiens n'a pas permis de reproduire ces résultats.

Compte tenu de l'absence d'associations SNP significatifs avec le rejet/perde chronique, certaines études se sont tournées vers une stratégie GWAS « double génome ». Cette stratégie est basée sur l'étude simultanément des variants génétiques chez les donneurs et les receveurs à travers un score qui calcule la somme des incompatibilités entre receveurs et donneurs. Classiquement, dans les GWAS, les scores de risque polygénique sont utilisés pour évaluer l'effet combiné des SNP d'un organisme avec un phénotype individuel (tel que le facteur de risque de développer une maladie). Ici, au lieu d'examiner un seul génome et de combiner la taille de l'effet de chaque SNP significatif, le score est calculé pour une paire donneur-receveur comme la somme des contributions de mismatch (c'est-à-dire la différence allélique) sur un ensemble de polymorphismes génomiques à travers le génome [215]. Ces méthodes ajoutent une couche de complexité car on passe de  $2n$  chromosomes à  $4n$  chromosomes différents. Cette approche a été utilisée dans deux études publiées en 2019 respectivement par Reindl-Schwaighofer et al. [216] et Steers et al. [217]. La première étude a examiné les variants non synonymes c'est-à-dire qui impliquent des changements en acide aminés pour les protéines transmembranaires et sécrétées et a déterminé un score mHLA, basé sur une valeur médiane au-dessus de laquelle le risque de perte de greffe rénale est accru. La seconde étude a criblé 50 SNP de marquage de délétions chez 704 receveurs d'allogreffes rénales et a défini une collision génomique (génotype à risque de rejet de greffe) comme une combinaison génotype spécifique donneur-receveur dans laquelle un receveur homozygote pour un allèle de marquage de délétions a reçu une greffe d'un donneur non homozygote. Ces deux études récentes ont confirmé que les facteurs non HLA sont impliqués dans les résultats chroniques de la transplantation rénale. Elles font écho à celle de Mesnard et

al. [215], qui ont introduit pour la première fois un système de score additif pour évaluer les facteurs non HLA dans la fonction du greffon rénal à partir de données génomiques (séquençage de l'exome de nouvelle génération).

Ce type de stratégie peut bénéficier de l'intégration de couches de données supplémentaires (expression génique, protéines, métabolites) pour mieux comprendre les résultats de la transplantation. Des études ont récemment combiné plusieurs types de données omiques provenant des mêmes échantillons afin de reconstruire les réseaux moléculaires qui sont inaccessibles avec une seule couche de données omiques [218, 219]. Dans le domaine de la transplantation rénale, la recherche de biomarqueurs par des méthodes multi-omiques commence à être utilisée. Une étude a appliqué des méthodes non supervisées et supervisées pour explorer et intégrer les données génomiques et protéomiques de patients ayant subi une transplantation rénale (événements de rejet aigu dans les 30 jours suivant la transplantation (n=20) par rapport à des contrôles (n=20)) et pour déterminer quels gènes et protéines jouent un rôle important dans le rejet aigu d'allogreffe rénal. Une meilleure séparation entre les deux groupes d'échantillons a été observée dans l'analyse intégrative (génomique et protéomique combinées) que dans l'analyse omique simple [220].

# OBJECTIFS DE LA THÈSE

---

Les études GWAS ont permis de mieux comprendre les mécanismes d'interactions entre donneur et receveur. Malgré ces avancées, de nombreuses questions demeurent. En effet ces études ont été principalement menées sur des receveurs, à l'exception de quelques études menées sur des paires de donneurs et de receveurs (paires D/R) [221, 216]. La taille des cohortes est encore très modeste et malgré les progrès significatifs dans la gestion post-transplantation, notamment grâce au développement des traitements immunosuppresseurs, le rejet et la perte chronique du greffon restent des défis majeurs. En effet, il y a 56% de survie du greffon dans les 10 ans post-transplantation et ceci inclut les transplantations HLA compatibles à 100%. De plus, il n'y a que 15% de différence dans la survie à 10 ans post-transplantation entre les donneurs HLA compatibles et incompatibles [91]. Ces résultats suggèrent que au-delà du HLA, d'autres facteurs génétiques pourraient être impliqués dans la transplantation rénale. L'objectif de ma thèse est de contribuer à l'identification de ces facteurs génétiques afin de (1) mieux appréhender les interactions entre donneur et receveur notamment celles impliqués dans la compatibilité (2) améliorer la compréhension des mécanismes de rejet/perte chronique (3) affiner la prise en charge des transplantés rénaux via l'identification de potentiels biomarqueurs non-invasifs et/ou de nouvelles cibles thérapeutiques. L'originalité de notre approche repose sur le fait qu'on disposera d'une cohorte homogène et conséquente avec des données de suivi à long terme qui vont nous permettre de pouvoir aller au-delà de ce qui été fait jusqu'à présent. Mes travaux vont se diviser en deux axes :

## 3.1 Axe 1 : Consolidation d'une biobanque de paires donneur/-receveur pour la transplantation rénale

Le premier axe va consister à l'élaboration d'une ressource de données complète et homogène pour les études génomiques en transplantation rénale appelée KiT-GENIE (Kidney Transplantation - GENomics Investigation of Essential clinical concerns). L'ambition est d'intégrer une triple couche de données (cliniques, biologiques et génomiques) afin d'obtenir l'une des plus grandes cohortes monocentrique de paires donneurs/receveurs à ce jour. Les objectifs sont :

- L'extraction des données cliniques de transplantation et de suivi des patients et de leurs

donneurs à partir de registres médicaux déjà existants.

- La collecte des échantillons d'ADN pour chaque paire donneur/receveur
- Le génotypage, l'analyse primaire et le contrôle qualité des données
- Le stockage propre des données cliniques, biologiques et génétiques

### 3.2 Axe 2 : Explorations statistiques et analyse de données génétiques à grande échelle pour une meilleure prise en charge du greffon rénal

Le second axe représente la partie analytique des données. Mon projet est de m'appuyer sur les données de KiT-GENIE pour l'identification de variants génétiques d'intérêt impliqués dans la transplantation en utilisant des approches génétiques à grande échelle. Les objectifs sont :

- Explorations statistiques des données afin de pouvoir établir une description précise et complète de la cohorte
- Analyse préliminaires des données génétiques
- Validation des résultats de la littérature actuelle avec des analyses GWAS simple sur les donneurs et les receveurs séparément
- Analyse GWAS sur les paires donneur/receveur en simultanée
- Elaboration d'un score polygénique de risque de perte de greffon entre donneur et receveur

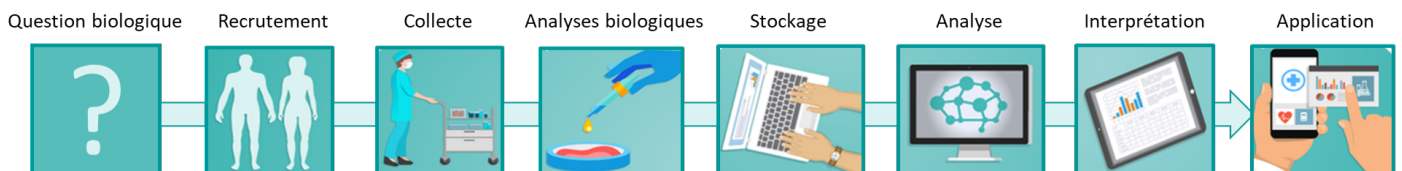


FIGURE 3.1 – *Workflow des objectifs de la thèse*

DEUXIÈME PARTIE

# Protocoles

---





# GÉNÉRATION DES DONNÉES

---

## 1.1 Extraction et collecte des données

Les données utilisées pour l'élaboration de KiT-GENIE proviennent de DIVAT-Nantes (Données Informatisées et VALidées en Transplantation, [www.divat.fr](http://www.divat.fr)). DIVAT est une cohorte française multicentrique qui fournit une ressource de données riche en transplantation en accord avec la CNIL (commission nationale informatique et des libertés) afin d'améliorer les recherches axées sur les résultats cliniques, les stratégies thérapeutiques et les questions de santé publique ayant trait à la transplantation rénale et pancréatique [222, 223, 224]. La cohorte regroupe huit centres : Nantes, Nancy, Montpellier, Toulouse, Necker (Paris), Lyon, Saint-Louis (Paris) et Nice qui collaborent pour collecter et échanger des données centralisées dans une base de données à Nantes.

Environ un tiers des transplantations rénales réalisées aujourd'hui en France sont incluses dans DIVAT (30 000 données de patients recueillies depuis 1990 dans huit centres différents). Les données sont basées sur les dossiers médicaux existants, y compris les données relatives aux receveurs, aux donneurs, à la transplantation et au suivi. Les données sont mises à jour à chaque visite de suivi post-transplantation (à 3 mois, 6 mois et 12 mois post-transplantation, puis tous les ans). Les données sont collectées depuis l'inclusion du patient jusqu'à la perte du greffon (ou arrêt du suivi). Elles concernent (1) les données du receveur telles que les caractéristiques individuelles (âge, sexe, IMC (Indice de Masse Corporelle) etc.), la maladie initiale ayant conduit à la greffe, la sérologie virale, les antécédents médicaux, le typage HLA etc. (2) les informations relatives au donneur, y compris les caractéristiques individuelles, le type de donneur (vivant ou décédé), la cause du décès, la sérologie virale, le typage HLA, les analyses de sang (urée, créatinine), etc. (3) et les données de la transplantation : date, lieu, type, temps d'ischémie, retour de la fonction rénale, nombre d'incompatibilités HLA-ABDR, etc. Au cours du suivi post-transplantation, les paramètres suivants sont recueillis : caractéristiques individuelles, créatinine sérique, taux de filtration glomérulaire estimé (DFG), protéinurie sur 24 heures, traitements immunosuppresseurs, épisodes de rejet aigu, infections, complications, retour en dialyse ou décès, informations sur les biopsies etc. Ces données cliniques ont été extraites afin de pouvoir collecter les échantillons d'ADN pour chaque paires donneur-receveur provenant du centre nantais. La collecte a été réalisée sur la base des échantillons disponibles dans la bio-collection DIVAT déjà existante et des

échantillons stockés à l’Etablissement Français du Sang (EFS). Les échantillons ont été aliquotés dans des tubes adaptés à la conservation à long terme et stockés à  $-20^{\circ}\text{C}$ . Un numéro unique propre à KiT-GENIE a été attribué à chaque échantillon d’ADN afin de garantir la traçabilité des échantillons et la récupération rapide des informations individuelles. Les échantillons d’ADN ont été quantifiés pour obtenir des informations sur le volume, l’absorbance (230, 260, 280, 260/230, 260/280) et la concentration afin d’avoir des informations biologiques complètes sur la bio-collection et de faciliter la sélection des échantillons pour le génotypage qui nécessite une bonne qualité et quantité d’ADN. Nous récupérons l’ADN avec une concentration d’au moins  $10\text{ ng}/\mu\text{L}$  et  $20\ \mu\text{L}$  de volume. La qualité des échantillons a été évaluée par migration sur gel d’agarose. Les échantillons dont la quantité est inférieure aux seuils normaux ou avec une qualité non conforme sont considérés comme détériorés et sont écartés de la bio-collection.

## 1.2 Génotypage

Le génotypage est un processus qui permet de déterminer des variants génétiques (SNP) à des positions spécifiques dans le génome d’un individu [225]. La méthode repose sur des puces de génotypage qui peuvent mesurer des milliers de variants à la fois. Pour le projet KiT-GENIE, nous avons suivi les directives d’Axiom® 2.0 de ThermoFisher pour tout le processus. Les individus ont été génotypés à l’aide de la puce Axiom PMRA (Precision Medicine Research Array) qui contient plus de 900 000 variants y compris des variants rares ainsi que de variants immunologiques, fonctionnels, des variants communs du cancer, etc.

Les échantillons d’ADN ont été soigneusement randomisés (pour le sexe, le statut D/R, le statut de donneur décédé ou vivant) dans les plaques afin de limiter l’impact de l’effet de lot ou « batch effect ». De plus, nous avons introduit des échantillons contrôles et des doublons afin d’évaluer la qualité et la reproductibilité de l’expérience. Le protocole de génotypage dure 5 jours pour une plaque de 96 puits (Figure 1.1). Les 3 premiers jours correspondent aux étapes d’amplification, de fragmentation et de resuspension de l’ADN. Une analyse de contrôle qualité (QC) au troisième jour est effectuée afin de requantifier l’ADN et de vérifier l’intégrité des échantillons. Enfin, les plaques sont chargées dans le GeneTitan pour les étapes d’hybridation et de scan qui correspond à la mesure de l’intensité de fluorescence spécifique à chaque allèle afin d’attribuer correctement les génotypes. Le GeneTitan fournit des données brutes qui seront soumises à une analyse de contrôle qualité primaire. Dans le cadre de KiT-GENIE, le génotypage a été réalisé par nos soins sur la plateforme Geno-BiRD et la plateforme de l’institut Curie. Le génome de référence était le GRCh37/Hg19.

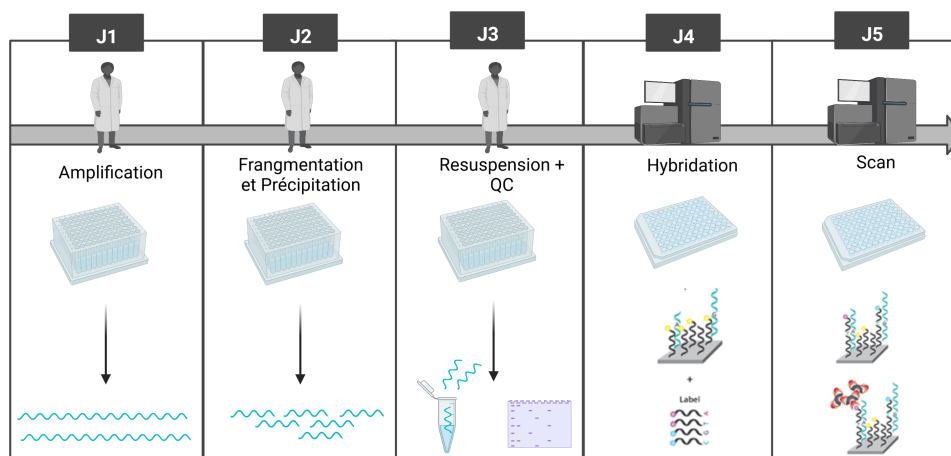


FIGURE 1.1 – *Les différentes étapes du protocole de génotypage.\** Le protocole de génotypage dure 5 jours pour une plaque de 96 puits. Les 3 premiers jours se font en paillasse et correspondent à la préparation de l'ADN pour l'étape d'hybridation. L'ADN est amplifié puis fragmenté afin de récupérer des petits brins qui seront remis en suspension. Au 3e jour, un QC est effectué pour vérifier la conformité des échantillons. Les plaques sont ensuite chargées dans le GeneTitan au 4e et 5e jour pour les étapes d'hybridation et de lecture. Le GeneTitan fournit en sortie des données brutes qui passeront par une analyse primaire de contrôle qualité.

### 1.3 Analyse primaire des données brutes

Cette analyse de QC a été réalisée avec AxAs (Axiom Analysis Suite) proposé par ThermoFisher qui évalue en 3 étapes essentielles, la qualité des plaques, des individus et des variants génétiques afin de fournir des données génomiques propres et exploitables (Figure 1.2).

Le QC des individus est basé sur la mesure du dishQC (DQC) et du Call Rate (CR) qui évaluent respectivement le chevauchement entre les deux pics homozygotes (AT versus GC) et mesurent le pourcentage de SNP correctement assignés à un génotype dans un run de génotypage. Les échantillons présentant une valeur  $DQC \leq 0,82$  et un  $CR \leq 97\%$  ont été supprimés. Sur cette base, les plaques ayant un taux d'échantillons qui ont passés les seuils du  $CR \leq 98,5\%$  ne seront pas considérées pour l'étape suivante. Le QC des SNP est basé sur plusieurs métriques dont le CR, le FLD (« Fisher's linear discriminant »), l'hétérozygotie (HetsO) et l'homozygotie (HomRO). Les mesures CR (seuil  $\geq 95$ ) et FLD (seuil  $\geq 3,6$ ) assurent la qualité des SNP génotypés en vérifiant l'assignation correcte des SNP dans les clusters (homozygotes et hétérozygotes) et leur séparation. HetsO (seuil  $\geq 95$ ) et HomRO (seuil  $\geq -0,9$ ;  $0,3$ ;  $0,6$  selon le génotype du SNP)

\* Figure créée avec BioRender.com.

évaluent le bon étiquetage des clusters hétérozygotes et homozygotes.

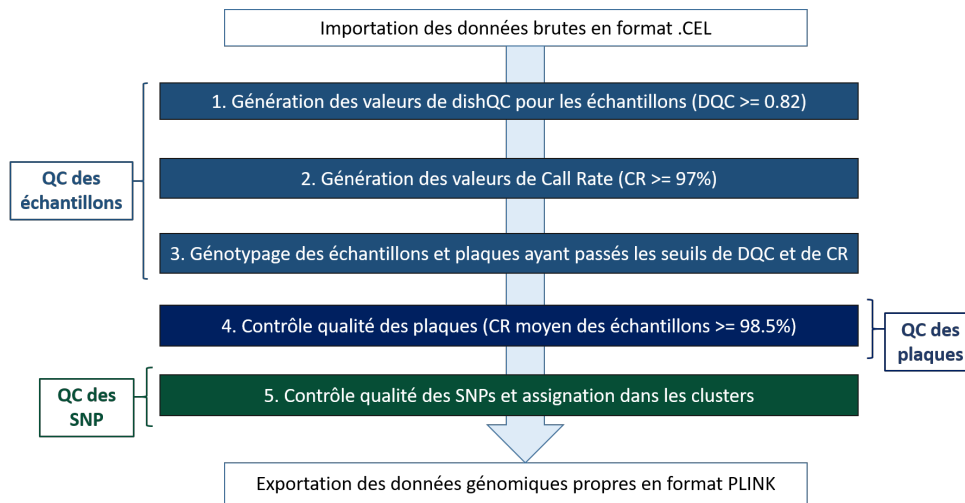


FIGURE 1.2 – *Les différentes étapes de l'analyse primaire\**

## 1.4 Stockage des données

L'intégration des données génomiques aux multiples données biologiques et cliniques de DIVAT dans le cadre du projet KiT-GENIE passe par la mise en place d'une structure de stockage propre, homogène et adaptée à la récupération facile des informations. Pour ce projet, les données seront stockées dans une base de données avec différentes tables contenant les données cliniques, biologiques et génétiques. L'utilisation d'une base de données nécessite deux entités : la base de données et ce qu'on appelle le SGBD (Système de Gestion de Base de Données). Ce dernier permet entre autres choses la communication avec la base de données, la construction et la gestion des tables. La base de données a été mise en place sous le système de gestion de base de données relationnelle et objet PostgreSQL [226]. La gestion de la base de données est assurée par pgAdmin qui est le SGBD de base de PostgreSQL, et qui propose une interface permettant une gestion plus facile des tables. La base données KiT-GENIE est hébergée par la plateforme de génomique et bio-informatique de l'université de Nantes. Elle contient à ce jour toutes les informations concernant le receveur, le donneur, la transplantation, le suivi, les antécédents etc. ainsi que les informations relatives aux échantillons d'ADN, au génotypage et aux analyses de contrôle qualité. L'intégration des données génétiques est en cours de réalisation au sein de l'équipe.

\* Figure créée avec BioRender.com.

# MÉTHODES STATISTIQUES

---

## 2.1 Contrôle qualité des données

Avant toute analyse GWAS, il convient de vérifier rigoureusement la qualité des données génomiques afin d'éviter les résultats biaisés qui conduisent souvent à une mauvaise interprétation. Ce contrôle qualité est basé sur les recommandations standards pour les analyses GWAS et concerne principalement un contrôle qualité sur les SNP et un contrôle qualité sur les individus [195, 227]. Dans le cadre de KiT-GENIE, nous avons effectué ce QC à l'aide du logiciel PLINK [228] en respectant les étapes suivantes :

- **Le taux de données manquantes.** Les SNP présentant un taux de données manquantes de plus de 2% ont été éliminés car un pourcentage trop élevé de données manquantes peut indiquer un biais technique remettant en question la fiabilité de la caractérisation des génotypes (« variant calling ») pour ces SNP en particulier. De la même manière, les individus ayant un taux de données manquantes de plus de 5% ont été exclus.
- **L'exclusion des indels et des doublons.** Les variants de type insertion/délétion ont été exclus car ils peuvent interférer lors des étapes d'analyses en aval telles que l'imputation. Afin de s'assurer de l'indépendance de nos individus, les doublons ainsi que les individus qui sont génétiquement trop proches ont été éliminés à l'aide d'un score de proximité génétique disponible dans PLINK et qui est basé sur les principes d'IBS (« identity by state ») et d'IBD (« identity by descent ») [229]. L'IBS est définie par l'existence de séquences nucléotidiques identiques entre deux individus. Lorsque ces séquences en communs ont été hérités d'un ancêtre commun sans recombinaison, on parle alors d'IBD qui traduit un lien de descendance entre individus [230]. Les segments d'ADN qui sont IBD sont par définition IBS, mais les segments qui ne sont pas IBD peuvent quand même être IBS en raison de mutations similaires chez différents individus ou de recombinaisons sans modifications de certains segments de l'ADN qui ne sont pas forcément dû à une descendance commune [231]. Ces scores génétiques, permettent d'identifier les séquences nucléotidiques partagées entre deux individus, et d'estimer leur degré de parenté. Au-delà d'un certain seuil (0.75 pour nos données), on peut considérer que les deux génomes comparés appartiennent soit à des jumeaux monozygotes (vrais jumeaux) soit au même individu. Ces individus sont exclus.

- **Vérification du sexe et du HLA.** Une fonction de PLINK permet d'inférer le sexe des individus à partir des données génétiques issues des chromosomes sexuels. Cela nous permet de comparer le sexe génotypé vs le sexe enregistré dans la base de données dans le but de s'assurer qu'il n'y a pas eu de mélange d'échantillons lors du génotypage. Nous avons détecté grâce à cela quelques incohérences chez certains individus qui ont pu être retracées et corrigées dans le cas où c'était possible ou tout simplement supprimées. Nous avons également eu recours à une imputation de la région *HLA* [232] grâce à HIBAG [233] afin de pouvoir comparer les allèles inférés avec ceux renseignés dans la base de données. Cette comparaison a été effectuée pour les 6 allèles des gènes *HLA-A*, *B*, et *DR*. Les individus ayant des différences  $HLA \geq 3/6$  incompatibilités par rapport au donneur ou qui présentaient 2/6 différences *HLA* mais avec une incohérence au niveau du sexe ont été éliminés.
- **Fréquence de l'Allèle Mineur (MAF).** Ce filtre permet de contrôler les erreurs d'assignation dues à des défauts de détection des SNP de fréquences trop faibles. En effet, ces SNP sont plus difficiles à génotyper car leur intensité lors du génotypage peut être mal détectée. Les fréquences faibles produisent de petits îlots d'un ou deux individus pour les homozygotes portant l'allèle mineur qui risquent d'être faussement assimilés au groupe le plus proche, les hétérozygotes. De plus, ce filtre permet de se concentrer sur les variants communs pour l'étude GWAS. Plusieurs seuils peuvent être utilisés selon la taille de la population d'étude mais les valeurs sont en général fixées entre 1% et 5%. Pour nos données, les SNP avec une MAF inférieure à 1% ont été exclus pour ne garder que les variants qui sont communs dans la population générale.
- **Équilibre d'Hardy-Weinberg (HWE).** Le principe d'HWE [190] repose sur le fait qu'en l'absence de conditions spécifiques (stratification de la population, apparemment entre individus ou erreurs systématiques de génotypage), les variants ont tendance à conserver les mêmes fréquences alléliques d'une génération à une autre au sein d'une population (cf. 2.2.2.4.1 Équilibre d'Hardy-Weinberg p. 82). Pour éviter une déviation de l'HWE, on utilise un test statistique implémenté dans PLINK qui permet d'exclure les SNP ayant une p-value  $<$  au seuil de  $10^{-6}$ .

## 2.2 Analyses en composantes principales (ACP)

L'analyse en composantes principales (ACP) [234] est une méthode statistique non supervisée développée pour l'analyse des données dites multivariées et qui permet de réduire la dimensionnalité de ces données tout en conservant leur variabilité. L'ACP synthétise les variables d'origine en seulement quelques nouvelles variables appelées composantes principales ou dimensions grâce à des combinaisons linéaires. En utilisant quelques composantes, chaque échantillon

peut être représenté par un nombre réduit de variables plutôt que par les valeurs de milliers de variables. Chaque composante peut alors être interprétée comme la direction, non corrélée aux composantes précédentes, qui maximise la variance (l'inertie) des échantillons lorsqu'elle est projetée sur la composante. L'ACP permet de visualiser les similitudes et les différences entre les échantillons d'un jeu de données et de déterminer s'ils peuvent être séparés en groupes. Elle est utilisée dans de nombreux domaines (finances, économie, environnement, infographie etc.) notamment dans le domaine biomédical [234]. En génomique, l'analyse en composantes principales est principalement utilisée, notamment dans les études GWAS, pour détecter les potentiels erreurs systématiques sur les données et saisir la variabilité liée à l'ancestralité [195].

En effet, les mouvements populationnels au cours de l'évolution, ont favorisés l'apparition de mutations génétiques propre à une population données et qui sont liées à ce qu'on appelle la stratification géographique. Il a été mis en évidence que l'une des principales causes de la variance des données entre individus dans les études génomiques provient de la répartition des origines géographiques de ces individus [235]. Ces études s'intéressent aux facteurs causaux indépendamment des effets de population. Pour pouvoir prendre en compte la spécificité génétique d'une population et contrôler cette variance des données dans les études, nous utilisons l'ACP qui permet de stratifier.

Dans le cadre de KiT-GENIE, l'analyse en composantes principales a permis d'identifier la diversité populationnelle des données en se basant sur les données de 1KGP. Quelques étapes préliminaires de QC ont été appliquées afin d'éviter des résultats biaisés. D'abord, les SNP qui sont systématiquement en déséquilibre de liaison [228] donc corrélées ont été éliminées. S'ils ne sont pas supprimés, ces SNP seront considérés comme des variables indépendantes pour l'ACP mais auraient tendance à être réassociés à cause de la colinéarité entre les variables, inhérente à l'analyse. Ceci aura un impact sur le nombre de variables mais également sur l'inertie qui sera faussement augmentée. Ensuite, les variants des chromosomes X et Y ainsi que les variants de la région HLA ont été éliminés d'une part pour éviter un regroupement basé sur le genre (homme/femme) après projection et d'autre part pour éviter une trop grosse variabilité entre les SNP. Enfin seuls les variants avec une MAF > 10% ont été conservés.

Les analyses préliminaires ont été réalisées grâce à l'outil PLINK [228] qui permet le traitement et l'analyse des données génétiques. L'ACP a été effectuée grâce à EIGENSOFT [236] avec comme population de référence les individus issus du projet 1KG [132] alignés sur la version GRCh37 du génôme.

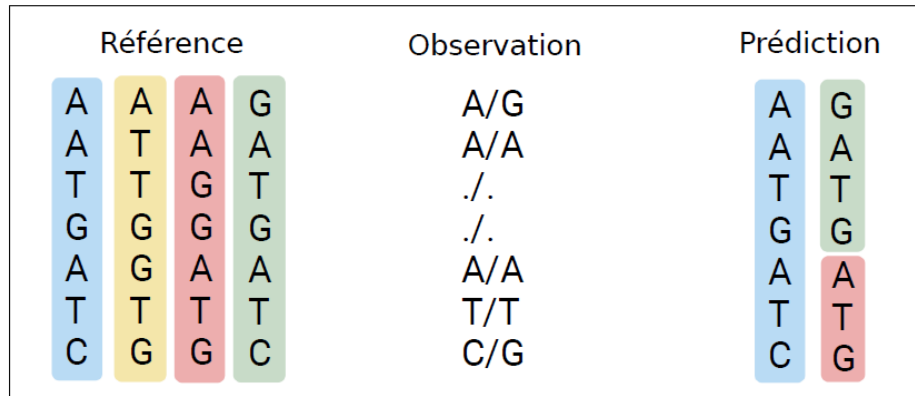


## 2.3 Imputation SNP et HLA

### 2.3.1 Imputation SNP

En utilisant un panel de référence, les SNP avec des données manquantes ou les SNP qui n'étaient pas directement couverts par les puces de génotypage peuvent être prédits (imputés) afin d'augmenter la couverture du génome, et ainsi maximiser la puissance statistique de l'étude [237]. L'imputation des SNP est basée sur la re-identification des segments d'haplotypes communs entre les génomes à imputer et un génome de référence. A partir des SNP génotypés, l'information manquante est identifiée et complétée en se référant à un panel de référence comportant tous les SNP notamment ceux en déséquilibre de liaison avec les SNP initialement génotypés (Figure 2.1). Pour chaque SNP imputé, un score de prédiction (la post-probabilité) est calculé afin d'estimer la fiabilité de la prédiction statistique (ex : si la post-probabilité est égale à 0.8, il y a 20% de chance que la prédiction soit fautive) [238].

Plusieurs outils ont déjà été développés pour inférer les génotypes manquants à partir des données de référence. Les données de KiT-GENIE ont été imputées en utilisant Beagle [239]. Cette approche se base sur les modèles de Markov cachés [240] qui sont des méthodes de décision bayésienne permettant d'attribuer une séquence inconnue à la classe qui a la plus grande probabilité de l'avoir engendrée. Les données KiT-GENIE ont été imputées en utilisant respectivement les données de la phase 30X GRCh38 (couverture de génotypage x30) et de la phase 3 (couverture de génotypage x3) du projet 1KG [132] qui comprend 7 325 7620 SNP. Par ailleurs, un nettoyage préliminaire spécifique des données non imputées est nécessaire afin notamment d'aligner les SNP génotypés avec le panel de référence. Cette préparation est basée sur les recommandations préconisées dans la documentation de Beagle [239] et a été effectuée grâce à l'outil PLINK [228].

FIGURE 2.1 – *Principe général de l'imputation SNP\**

### 2.3.2 Imputation HLA

De la même manière que les SNP, il est possible d'inférer les génotypes HLA à partir de données SNP. Dans la région du CMH, les allèles HLA sont en déséquilibre de liaison avec certains SNP c'est-à-dire qu'ils sont systématiquement transmis ensemble, ce qui permet une corrélation SNP-HLA. Ces SNP, lorsqu'ils sont à disposition, permettent de pouvoir prédire les génotypes HLA associés [232]. L'imputation SNP-HLA repose sur des panels de référence d'individus dont les SNP, les génotypes HLA et leurs potentielles associations sont connus. En utilisant ces panels de référence avec des méthodes appropriées, il est possible de refaire le lien entre les SNP disponibles et les allèles HLA manquants à l'aide d'algorithmes de machine learning. Suivant ce principe, l'imputation se subdivise en deux parties : l'apprentissage et la prédiction. L'apprentissage est réalisé avec des données d'entraînement (SNP et HLA connus), qui une fois assimilées par l'algorithme permettent de créer un modèle ou panel de référence. En appliquant ce dernier à des individus dont les SNP seulement sont connus, il est possible de prédire les allèles HLA manquants [232, 241, 242]. Des programmes utilisant différentes approches statistiques ont été développées pour l'imputation. Parmi ceux-là, SNP2HLA [243] et HLA\*IMP [244] ont été les premiers à être publiés et se basent sur des modèles de Markov cachés. Plus récemment, des nouvelles méthodes tels que HIBAG [233], DEEP-HLA [245], CookHLA [246] et HLA-IMP\*03 [247] ont vu le jour.

Pour les individus KiT-GENIE, les allèles HLA ont été imputés d'une part pour s'affranchir de toute erreur d'identification pouvant résulter d'une mauvaise manipulation lors du génotypage et d'autre part pour avoir la possibilité d'explorer ultérieurement les associations au niveau des

\* Figure créée avec BioRender.com.

allèles HLA corrélés à des SNP préidentifiés comme intéressants (associations significatives avec un trait donné). Ce qui peut être biologiquement plus informatif qu’une simple association SNP. L’imputation a été effectuée en utilisant HIBAG [233] qui fournit des prédictions robustes avec une haute précision et permet de fournir des panels de référence spécifiques à une population pour des centaines d’individus tout en permettant la construction de panels de référence personnalisés. L’algorithme d’HIBAG fonctionne sur la base de l’attribut bagging, une technique de machine learning reposant sur le rééchantillonnage (bootstrapping) des individus dans plusieurs sous-espaces appelés bags ou classifieurs (Figure 2.2). Ces classifieurs sélectionnent aléatoirement un sous-ensemble de SNP pour prédire le HLA et les fréquences d’haplotype estimées à partir du panel de référence. La probabilité estimée d’un génotype est donc calculée sur chacun des classifieurs puis moyennée : cette moyenne indique le génotype prédit.

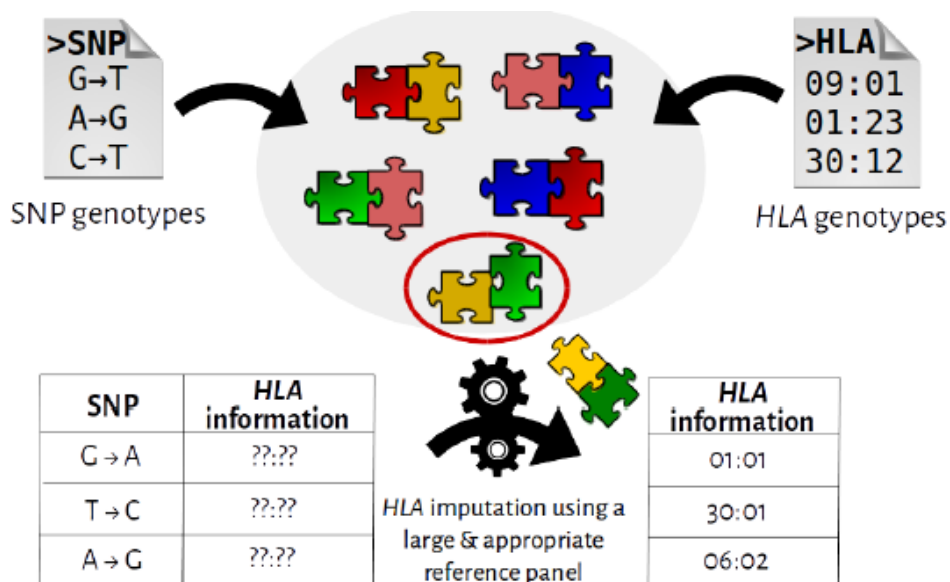


FIGURE 2.2 – *Principe général de l’imputation HLA\**

## 2.4 Analyses GWAS

Les études GWAS ont pour objectif d’identifier des potentielles associations entre la fréquence d’un allèle ou d’un génotype et un phénotype d’intérêt à l’échelle du génome. Les GWAS s’appuient sur des modèles de régression [248, 249] qui comparent la distribution des génotypes de chaque SNP en fonction de la variable phénotypique étudiée [195]. Ainsi, lorsque le phénotype est continu, le modèle utilisé est une régression linéaire et lorsque le phénotype est binaire, il

\* Figure créée avec BioRender.com.

s'agira d'une régression logistique [250].

La régression logistique est un modèle de régression binomiale qui vise à expliquer une variable expliquée binaire c'est-à-dire à deux modalités. Considérons une variable cible  $Y=1$  (où  $Y$  ne peut que prendre deux valeurs 0 ou 1) et des variables explicatives  $X = x_1, x_2, \dots, x_n$ . La probabilité a posteriori que  $Y=1$  sachant la valeur prise par  $X$  est notée  $p(1/X)$  :

$$\text{Logit de } p(1/X) = \ln\left(\frac{p(1/X)}{1-p(1/X)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

Dans le cadre du projet, notre analyse s'appuie sur une régression logistique définie avec une variable expliquée à deux modalités : l'échec ou la survie du greffon chez les receveurs et les donneurs européens séparément. Ces modèles ont été corrigés pour toutes les variables potentiellement confondantes identifiées et pour l'ancestralité (2 premiers axes de l'ACP) selon l'équation suivante :

$$\text{Pheno} \sim \text{Modèle}_{\text{SNP}} + \text{covariables} + \text{Eigen1} + \text{Eigen2}$$

Pour chaque SNP, un modèle de régression logistique a été appliqué afin de tester son association avec le phénotype intérêt. Cette association est évaluée à l'aide d'un test statistique, généralement un t-test (test de student) [251], portant sur le coefficient du SNP dans le modèle de régression. Le t-test aura pour hypothèse nulle : le coefficient de la régression est égal à zéro c'est-à-dire qu'il y'a une association. Ceci permet d'obtenir une p-value (entre 0 et 1) permettant d'évaluer la significativité (classiquement au seuil de 0.05) de l'association entre le SNP et le phénotype.

Etant donné que des millions de tests indépendants sont réalisés (à l'échelle du nombre de SNP testés), il est probable d'obtenir des p-values très proches de 0 par hasard plutôt que du fait d'une réelle association avec un SNP [252]. La correction de Bonferroni [253] qui consiste à modifier le seuil de significativité en le divisant par le nombre de tests indépendants effectués est appliquée afin de tenir en compte l'effet du multiple testing. Cette correction est établie selon la formule :

$$\alpha_{\text{Bonferroni}} = \alpha/N \quad (\alpha = \text{seuil de significativité}; N \text{ nombre de tests indépendants})$$

Le seuil de significativité est classiquement fixé à  $5 \cdot 10^{-8}$  pour un génome entier afin de maintenir le taux de faux positif à 5%.



TROISIÈME PARTIE

# Résultats

---



# RÉSULTATS PRÉ-GWAS

Cette partie concerne les résultats relatifs à la génération des données et aux analyses pré-GWAS. Elle détaille le montage et la description détaillée des données cliniques, biologiques et génétiques de la biocollection KiT-GENIE mais également les analyses préliminaires qui ont été effectuées en amont de l'analyse GWAS telles que l'exploration de l'ancestralité de notre population d'étude avec une caractérisation des proportions pour chaque groupe d'individus. Ces résultats ont fait l'objet d'un article soumis à EJHG.

## 1.1 Description des données et analyses préliminaires

### 1.1.1 La cohorte KiT-GENIE

#### Article

#### **KiT-GENIE, the French Genetic Biobank of Kidney Transplantation**

*Rokhaya BA, MSc ; Axelle Durand, MSc ; Vincent Mauduit, MSc ; Christine Chauveau, PhD ; Stéphanie Bernardet-Le Bas, PhD ; Sonia Salle, BSc ; Pierrick Guérif, BSc ; Clémence Petit, MD ; Venceslas Douillard, MSc ; Olivia Rousseau, MSc ; Clarisse Kerleau, MPH ; Nicolas Vince, PhD ; Magali Giral, MD, PhD ; Pierre-Antoine Gourraud, PhD, MPH ; Sophie Limou, PhD*

Soumis à EJHG : statut accepté avec révisions mineurs

#### 1.1.1.1 Résumé de l'article

La transplantation rénale est le seul traitement de suppléance pour l'insuffisance rénale chronique une maladie grave qui touche 10-15% de la population. L'insuffisance rénale entraîne une détérioration progressive de la fonction de filtration du rein qui lorsqu'elle n'atteint plus les 15% conduit à la phase terminale qui ne peut être traitée que de deux manières : la dialyse ou la transplantation rénale. Toutefois, la transplantation rénale reste à ce jour le traitement qui assure la meilleure survie et la meilleure qualité de vie pour le patient. Cependant, les organes sont une ressource rare et le nombre de patients sur liste d'attente n'a cessé d'augmenter au cours des dernières décennies. Le développement de traitements immunosuppresseurs efficaces et l'amélioration des procédures chirurgicales ont considérablement amélioré le taux de survie



des greffons à un an (~92%) au cours des 30 dernières années, mais le rejet et l'échec chroniques restent un défi majeur. En effet, la survie du greffon est de ~60% à 10 ans post-transplantation et il n'y a qu'une différence de 15 % dans la survie entre les donneurs HLA compatibles et incompatibles ; ce qui peut suggérer que d'autres facteurs génétiques, au-delà du HLA, pourraient être impliqués dans les résultats de la transplantation. Dans ce contexte, plusieurs études GWAS ont ainsi été réalisées pour étudier les variants génétiques (SNP, CNV) ayant un impact sur la fonction du rein, le rejet ou l'échec de la transplantation. Malgré ces avancées, de nombreux problèmes subsistent. En effet, les résultats concernent le plus souvent le rejet aigu, la taille des cohortes est très modeste ( $300 < n < 1\ 400$ ) et ces études ont été principalement menées sur des receveurs uniquement en raison de la disponibilité limitée de l'ADN du donneur et du receveur. De nouvelles approches ont été développées pour étudier les marqueurs non-HLA impliqués dans la transplantation rénale en considérant à la fois le donneur et le receveur. Au lieu d'étudier un seul organisme, ces études évaluent un score génétique qui calcule les incompatibilités entre receveurs et donneurs. Ces études génomiques sont des preuves de concepts et suggèrent que des efforts internationaux de méta-analyses, ainsi que des cohortes importantes et homogènes sont nécessaires pour mieux appréhender le rôle des facteurs de risque génétiques dans la transplantation rénale. C'est dans ce contexte que le projet KiT-GENIE a été mis en place. Il a pour but de consolider la très riche et homogène cohorte française DIVAT de donneurs et receveurs de rein (D/R) avec l'ajout d'une couche de données génomiques dans le but d'explorer les facteurs moléculaires impliqués en transplantation rénale. Pour cela, nous avons extrait à partir de la base de donnée DIVAT-Nantes, les données cliniques de 5 946 individus (D/R) et avons collecté des échantillons d'ADN auprès de 4 723 individus. A l'issue du génotypage, 4 217 individus présentant des données génomiques et cliniques précises ont été inclus dans la biobanque dont 1 945 D, 2 272 R et 1 969 paires D/R. Nous avons obtenu une reproductibilité du génotypage de 99,9 %, ce qui souligne la robustesse de nos analyses. Cette robustesse a été évaluée grâce à l'introduction de doublons et d'échantillons de contrôle au cours du génotypage. Après des contrôles qualité stricts et l'imputation des SNP, le jeu de données final englobe 4 217 individus, 8M de SNP et plus de 200 variables cliniques. KiT-GENIE fournit une représentation précise de la pratique clinique de la transplantation rénale à Nantes entre 2002 et 2018, avec un enrichissement en donneurs vivants (17%) et en receveurs souffrant de HSF (4%). Les receveurs étaient majoritairement des hommes (63%) avec un âge moyen de 51 ans et 86% ont été transplantés pour la première fois durant la période d'étude. Enfin, nous avons effectué une analyse exploratoire de l'ancestralité de nos données. La grande majorité des individus ont une ascendance européenne (93% des receveurs et 98% des donneurs). Néanmoins, les receveurs sont plus diversifiés que les donneurs, avec notamment 3,4% d'individus d'ascendance africaine chez les receveurs et 0,6% chez les donneurs. Grâce à l'intégration de données génétiques à grande échelle, KiT-GENIE est l'une des plus grandes biocollections de paires D/R ( $N = 1\ 969$

paire) avec des informations de suivi des patients pouvant aller jusqu'à 25 ans. Elle offre une occasion unique d'étudier les facteurs immunogénétiques et génétiques, ainsi que les interactions donneur-receveur et les mécanismes impliqués dans le rejet, la survie du greffon, la récurrence de la maladie primaire et d'autres comorbidités. Trois études génomiques s'appuyant sur cette ressource de données riches sont déjà en cours : (1) une première étude qui examine l'interaction génétique entre D/R (2) une deuxième étude basée sur la récurrence de la HSF et (3) une troisième étude basée sur les donneurs vivants et l'évolution de leur fonction rénale après le don.

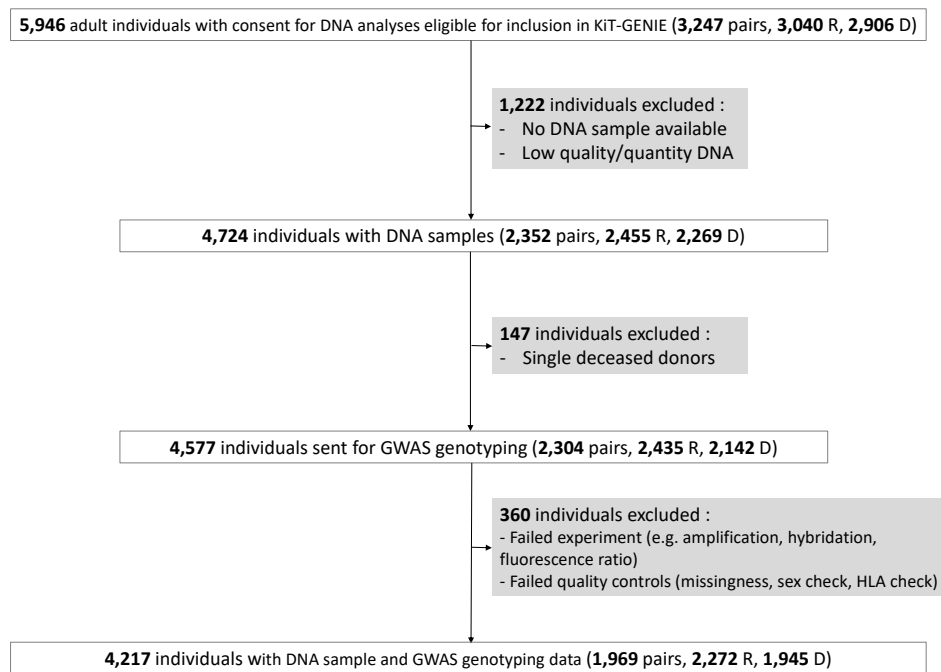


FIGURE 1.1 – *Les différentes étapes du traitement des données individuelles pour la biocollection KiT-GENIE*

---

**Article 1**

---

# 1 **KiT-GENIE, the French Genetic Biobank of**

## 2 **Kidney Transplantation**

3 Rokhaya BA, MSc<sup>1</sup> ; Axelle Durand, MSc<sup>1</sup> ; Vincent Mauduit, MSc<sup>1</sup> ; Christine Chauveau,  
4 PhD<sup>1</sup> ; Stéphanie Le Bas-Bernardet, PhD<sup>1</sup> ; Sonia Salle, BSc<sup>1</sup> ; Pierrick Guérif, BSc<sup>2</sup> ; Martin  
5 Morin, PhD<sup>1</sup>; Clémence Petit, MD<sup>1,2</sup> ; Venceslas Douillard, MSc<sup>1</sup> ; Olivia Rousseau, MSc<sup>1</sup> ;  
6 Gilles Blancho, MD, PhD<sup>1,2</sup>; Clarisse Kerleau, MPH<sup>2</sup> ; Nicolas Vince, PhD<sup>1</sup> ; Magali Giral,  
7 MD, PhD<sup>1,2</sup> ; Pierre-Antoine Gourraud, PhD, MPH<sup>1</sup> ; Sophie Limou, PhD<sup>1</sup>

8

9 <sup>1</sup>Nantes Université, Ecole Centrale Nantes, CHU Nantes, INSERM, Centre for Research in  
10 Transplantation and Translational Immunology, UMR 1064, F-44000 Nantes, France

11 <sup>2</sup>CHU Nantes, Nantes Université, Service de Néphrologie-Immunologie Clinique, ITUN, F-44000  
12 Nantes, France.

13

14 Correspondence:

15 Sophie Limou, PhD

16 CR2TI UMR1064 - ITUN - CHU Nantes Hôtel Dieu

17 30 bd Jean Monnet

18 F-44093 Nantes Cedex 01, France

19 Phone number: + 33 2.44.76.82.71

20 Email: [sophie.limou@univ-nantes.fr](mailto:sophie.limou@univ-nantes.fr)

21

22 The authors declare no conflicts of interest.

1 **Abstract**

2 KiT-GENIE is a monocentric DNA biobank set up to consolidate the very rich and  
3 homogeneous DIVAT French cohort of kidney donors and recipients (D/R) in order to explore  
4 the molecular factors involved in kidney transplantation outcomes. We collected DNA samples  
5 for kidney transplantations performed in Nantes, and we leveraged GWAS genotyping data for  
6 securing high-quality genetic data with deep SNP and HLA annotations through imputations  
7 and for inferring D/R genetic ancestry. Overall, the biobank included 4,217 individuals  
8 (n=1,945 D + 2,272 R, including 1,969 D/R pairs), 8M SNPs and over 200 clinical variables.  
9 KiT-GENIE represents an accurate snapshot of kidney transplantation clinical practice in  
10 Nantes between 2002 and 2018, with an enrichment in living kidney donors (17%) and  
11 recipients with focal segmental glomerulosclerosis (4%). Recipients were predominantly male  
12 (63%), of European ancestry (93%), with a mean age of 51yo and 86% experienced their first  
13 graft over the study period. D/R pairs were 93% from European ancestry, and 95% pairs  
14 exhibited at least one HLA allelic mismatch. The mean follow-up time was 6.7yrs with a  
15 hindsight up to 25yrs. Recipients experienced biopsy-proven rejection and graft loss for 16.6%  
16 and 21.3%, respectively. KiT-GENIE constitutes one of the largest kidney transplantation  
17 genetic cohorts worldwide to date. It includes homogeneous high-quality triple data layer  
18 (clinical, biological and genetic) for donors and recipients, hence offering a unique opportunity  
19 to investigate immunogenetic and genetic factors, as well as donor-recipient interactions and  
20 mismatches involved in rejection, graft survival, primary disease recurrence and other  
21 comorbidities.

22 **Keywords:** kidney transplantation, donor-recipient pair, genomics, clinical database, biobank,  
23 GWAS data analysis

# 1 **Introduction**

2 Chronic kidney disease is a public health condition affecting 10-15% of the population  
3 worldwide<sup>1</sup>. People with chronic kidney disease gradually lose their kidney filtering function  
4 and progress to end-stage kidney disease<sup>2</sup>. Kidney transplantation is the best renal replacement  
5 strategy for end-stage kidney disease patients regarding patient survival and quality of life  
6 compared to dialysis<sup>3-5</sup>. However, organs are a scarce resource and patients on waiting list have  
7 been steadily increasing over the last decades. In France, 3,251 kidney transplantations were  
8 performed in 2021<sup>6</sup> for 17,148 patients on the waiting list, *i.e.* over a 5:1 ratio of transplant  
9 candidates per available organ<sup>7</sup>. In addition, 54.5% of patients were transplanted 3 years after  
10 registration on the waiting list<sup>7</sup>. The development of effective immunosuppressive treatments  
11 and improved surgical procedures have significantly improved the 1yr graft survival rate  
12 (~92%) over the past 30 years, but with no major impact on the long-term graft survival, which  
13 still plateaus at ~60% 10 years after kidney transplantation<sup>7,8</sup>. Chronic kidney allograft rejection  
14 and failure therefore remain a major challenge and a critical public health concern considering  
15 the scarcity of organs. Interestingly, there is only a 15% difference in survival at 10 years post-  
16 transplantation between compatible and incompatible HLA donors<sup>9</sup>, which may suggest that  
17 other genetic factors, beyond HLA, could be involved in transplant outcomes.

18 Large-scale genetic studies using genome-wide association study (GWAS) genotyping<sup>10</sup> or  
19 next-generation sequencing can respectively assess the role of common and rare genetic  
20 variants spread across the genome in disease outcomes<sup>11,12</sup>. Genomics and other large-scale  
21 omics, including transcriptomics, have promoted the development of biobanks such as the UK  
22 biobank<sup>13</sup>, containing well-annotated and characterized biological samples (*e.g.* cells, tissues,  
23 blood, DNA) and their related databases<sup>14,15</sup>. These biocollections have contributed to a better  
24 understanding of human diseases and the discovery of new biomarkers<sup>16-18</sup>. In the context of  
25 kidney transplantation, several GWASs have thereby been performed to investigate genetic

1 variants (single nucleotide polymorphisms [SNP] or copy number variants [CNV]) impacting  
2 kidney allograft function, rejection or failure<sup>19–26</sup>. All genomic studies but three<sup>20,24,26</sup> were only  
3 conducted on recipients due to the limited availability of both donor and recipient DNAs.  
4 Overall, the cohort sizes were modest ( $300 < n < 1,400$ ), except for one with close to 2,000 pairs  
5 through a multicentric design<sup>20</sup>, and they failed to identify any significant reproducible donor  
6 or recipient SNP association. However, one study underlined the global importance of non-  
7 HLA factors for kidney allograft survival by combining donor-recipient (D/R) SNP allelic  
8 mismatches impacting the amino-acid sequence of cell-surface proteins<sup>26</sup>. Finally, acute  
9 rejection and graft survival were the most studied endpoints but the outcomes definition can be  
10 heterogeneous from one centre to another. There is therefore an important need for validation  
11 studies. These genomic studies have paved the way for the future of kidney transplantation  
12 genomics and suggest that international meta-analysis efforts<sup>27</sup>, as well as large and  
13 homogeneous cohorts are warranted to better disentangle the role of genetic risk factors in  
14 kidney transplantation.

15 Here, we describe the genetic characterisation and the clinical annotations of the KiT-GENIE  
16 (Kidney Transplantation - GENomics Investigation of Essential clinical outcomes) French  
17 biobank. KiT-GENIE gathers DNA samples, GWAS genotyping data and clinical data for both  
18 donors and recipients from the Nantes kidney transplantation clinical centre. KiT-GENIE was  
19 built onto the well-established DIVAT (*Données Informatisées et VALidées en Transplantation*)  
20 clinical database, which contains the monitoring of medical records for kidney transplantations,  
21 by collecting DNA samples and generating GWAS genotyping for 1,969 D/R pairs. KiT-  
22 GENIE is thus one of the largest and most homogeneous biobanks with D/R pairs worldwide  
23 to date. This major resource for kidney transplantation research offers a unique opportunity to  
24 explore the genetic architecture of different post-transplant outcomes in order to improve their  
25 understanding, prevention, management, prediction and treatment.

# 1 **Subjects and Methods**

## 2 *Description of the DIVAT clinical database*

3 We established the KiT-GENIE DNA biobank to consolidate the existing DIVAT clinical  
4 database and biocollection of kidney recipients from Nantes (France). DIVAT is a French  
5 multicentric observational and prospective cohort providing a unique resource for kidney  
6 transplantation-related epidemiological studies focusing on clinical outcomes, therapeutic  
7 strategies and public health issues<sup>28-30</sup>. It gathers data from eight clinical centres in order to  
8 increase statistical power and better represent the French population and diversity of clinical  
9 practice: Nantes, Nancy, Montpellier, Toulouse, Necker (Paris), Lyon, Saint-Louis (Paris) and  
10 Nice. Data from each participating centre are centralized in a database located and managed in  
11 Nantes ([www.divat.fr](http://www.divat.fr), CNIL n°914184, ClinicalTrials.gov recording NCT02900040). About  
12 one third of kidney transplantations performed today in France are included in DIVAT, with a  
13 total of 30,000 patients' data collected since 1990. The collected data are based on medical  
14 records and include more than 200 variables related to the recipient, the donor, the  
15 transplantation and the patient follow-up. Routine follow-up visits are performed at 3 months  
16 and 6 months post-transplantation and at each transplant anniversary. Additional visits may  
17 occur in case of complications. Data are collected from patient's inclusion until loss of graft or  
18 follow-up. Baseline variables encompass (1) recipient-related variables, including individual  
19 characteristics (*e.g.* age, sex, body mass index [BMI]), the initial disease leading to  
20 transplantation, viral serology, medical history, dialysis history, previous transplantations, or  
21 HLA typing ; (2) donor-related variables, including individual characteristics, the donor type  
22 (living or deceased), cause of death, viral serology, HLA typing, or blood assays (urea,  
23 creatinine) ; and (3) transplantation-related variables, including date, location, ischemia time,  
24 renal recovery time, or number of *HLA-A*, *-B*, and *-DR* allelic incompatibilities. Follow-up  
25 variables include recipient characteristics, serum creatinine, estimated glomerular filtration rate



1 (eGFR), 24-hour proteinuria, immunosuppressive treatments, acute rejection episodes,  
2 infections, complications, return to dialysis, and biopsy conclusions.

### 3 ***KiT-GENIE inclusion criteria***

4 KiT-GENIE is a monocentric DNA biobank focusing on kidney transplantations performed in  
5 Nantes between 2000 and 2018. It was built onto the DIVAT Nantes clinical database and  
6 consists of three layers: (1) DNA samples of donors and recipients, (2) GWAS genomic data,  
7 and (3) the aforementioned clinical data related to transplantation and patient follow-up.

8 First, we requested an extraction from the DIVAT Nantes database for all adult kidney  
9 transplantations performed in Nantes between 2000 and 2018 for whom we could collect  
10 consent and authorisation for DNA analyses. All paediatric patients (age<18yo at the time of  
11 transplantation) were excluded. We then extended our recruitment time period (1984-2020) to  
12 specifically enrich our cohort in living kidney donors and in recipients suffering from focal  
13 segmental glomerulosclerosis (FSGS) as primary kidney disease. Overall, a total of 5,946 adult  
14 individuals were eligible for DNA investigations (n=2,906 D + 3,040 R, including 3,247 D/R  
15 pairs, see Figure 1A).

### 16 ***KiT-GENIE DNA samples collection***

17 Based on these graft medical records, we collected DNA samples for donors and recipients  
18 from the DIVAT biocollection (DNA extraction from PBMC samples) and from the stored  
19 samples located in the local French blood bank/HLA typing lab. We collected DNA samples  
20 for 4,724 out of 5,946 individuals (79%), as DNA was missing or of low quality/quantity for  
21 1,222 individuals (Figure 1A). In particular, we could not retrieve most transplanted patients  
22 from 2000 and 2001 due to a progressive transition from paper-based to electronic biological  
23 records in the local blood bank. Each DNA was assessed for volume, concentration, purity  
24 (260/230 and 260/280 optical density absorbance ratios) and degradation by migration on an  
25 agarose gel (see Table S1). Indeed, GWAS genotyping requires good quality and quantity of

1 DNA, as defined by an absence of smear on the agarose gel, 260/230 and 260/280 ratios  
2 between 1.5 and 2, a volume greater than 20 $\mu$ L and a concentration greater than 10ng/ $\mu$ L.  
3 Overall, we were able to retrieve more recipients than donors (n=2,455 vs 2,269). Multiple  
4 samples were in some cases available for recipients from follow-up visits or multiple  
5 transplantation events. However, samples were most often unique for donors and sometimes  
6 depleted, especially for deceased donors coming from a different clinical centre. Each DNA  
7 sample from the KiT-GENIE biocollection has been assigned a unique and specific  
8 identification number to ensure sample traceability and retrieval of individual information, and  
9 has been aliquoted into tubes suitable for long-term conservation and stored at -20°C.  
10 For the genotyping analysis, 147 deceased donors for whom we could not retrieve a DNA  
11 sample for the recipient were excluded from the study. However, we decided to include all  
12 recipients whether in a pair or not, as clinical information is still available in the DIVAT  
13 database. A total of 4,577 individuals were hence submitted to GWAS genotyping including  
14 2,435 R, 2,142 D and 2,304 D/R pairs (Figure 1A).

#### 15 ***GWAS genotyping, data processing and imputations***

16 All DNA samples were genotyped using the Axiom PMRA (Precision Medicine Research  
17 Array) chip (ThermoFisher, Waltham MA, USA), which covers more than 900K genetic  
18 variants and is enriched for functional coding variants, including in the *HLA* and *KIR* genes, for  
19 clinically pathogenic variants and additional relevant genes for cancer and immunology  
20 research<sup>31</sup>. We followed the Axiom® 2.0 ThermoFisher guidelines for the genotyping protocol  
21 that encompasses 5 days for each 96-well plate. DNA samples were carefully randomly  
22 distributed (for sex, D/R status, deceased/living donor status) in plates to limit the impact of  
23 batch effect. In addition, we genotyped control samples and duplicates to assess experimental  
24 quality and replicability. DNA were first amplified and fragmented before hybridisation on the  
25 genotyping array. Allelic-specific fluorescence intensity ratio was then measured by the

1 GeneTitan scanner to correctly assign genotypes. Finally, the primary technological quality  
2 control (QC) steps were performed with AxAs (Axiom Analysis Suite) to only retain high-  
3 quality individuals (dishQC >0.82 and call rate >97%), plates (mean call rate >98.5%) and  
4 genetic variants (n=831,213 SNPs after QC, see Figure 1B). The SNP QC metrics evaluate the  
5 correct genotype assignment and ensure a proper separation of homozygous and heterozygous  
6 clusters (call rate >95%, Fisher's linear discrimination >3.6 and heterozygosity >95%).  
7 Further QC steps were performed with Plink<sup>32</sup> to exclude low confidence variants and  
8 individuals (*e.g.* missingness >2%, allelic frequency >1%, deviation from the Hardy-Weinberg  
9 equilibrium >10<sup>-6</sup>, see Figure 1B). From this dataset (n=374,743 SNPs), we imputed missing  
10 SNPs and HLA alleles with Beagle<sup>33,34</sup> and HIBAG<sup>35</sup>, respectively. Using the 1,000 Genomes  
11 Project (1KGP) reference panel<sup>36</sup>, SNPs with missing data or SNPs not directly covered by the  
12 genotyping arrays can be statistically inferred from the neighbouring genotyped SNPs by  
13 comparing the haplotypes and linkage disequilibrium patterns between the reference panel with  
14 the dataset of interest. Similarly, we can predict HLA alleles from SNPs data using a reference  
15 panel containing HLA alleles and SNP data<sup>37</sup>. Finally, we also determined the genetic sex of  
16 each individual from genetic variants located on sexual chromosomes.

17 To ensure the quality of KiT-GENIE genotyping data, we performed some preliminary analyses  
18 to (1) assess the genotyping error rate from duplicate samples, and (2) assess proper sample  
19 management process by comparing imputed sex and HLA alleles with sex and HLA types from  
20 clinical records. Finally, we performed a principal component analysis (PCA) from independent  
21 SNPs (Figure 1B) to detect potential biases and capture ancestry-related variability. To  
22 characterize the genetic ancestry of our KiT-GENIE individuals, we overlapped our data with  
23 data from the 1KGP<sup>36</sup> that catalogues common variants in 2,504 individuals from 5 large  
24 reference genetic populations (African, American, East Asian, European and South Asian).

## 1 **Results**

2 The KiT-GENIE biobank encompasses DNA samples with accurate GWAS genomic (8 million  
3 SNPs) and clinical (>200 variables) data for 4,217 individuals (n=2,272 R + 1,945 D, including  
4 1,969 pairs; Figure 1).

### 5 ***KiT-GENIE individuals characteristics***

6 The KiT-GENIE biobank incorporates clinical data related to the donor and recipient baseline  
7 demographics, to the transplantation and to the patient follow-up. Table S2 lists all these  
8 variables along with their respective level of missingness. Table 1 presents summary statistics  
9 for major clinical variables for the 4,217 genotyped individuals. Donors and recipients are  
10 predominantly male with 57.9% and 63.2%, respectively. The mean age is 51yo for both donors  
11 and recipients and their mean BMI is 25.4 and 24.4, respectively. During our study period, a  
12 large majority of recipients enrolled in KiT-GENIE (85.7%) received their first transplant,  
13 while some experienced their 2<sup>nd</sup> to 5<sup>th</sup> graft. 88.4% recipients underwent a kidney-only organ  
14 transplantation, while 11.3% got combined kidney-pancreas grafts. Only 4.4% grafts presented  
15 full HLA match (*HLA-A*, *-B* and *-DR* genes) between donor and recipient, while half the grafts  
16 exhibited 4 to 6 mismatches. Indeed, significantly more D/R pairs exhibited at least 3 *HLA*  
17 allelic mismatches ( $p=4.81 \times 10^{-10}$ , Figure S1). Overall, the distribution of *HLA* mismatches  
18 remained stable over the study period with no correlation observed between mismatch number  
19 and graft year ( $p>0.05$ , Figure S1). The mean follow-up time was 80months (*i.e.* 6.7yrs) post-  
20 transplantation. Due to inclusion criteria, most patients were transplanted between 2000 and  
21 2018 (Figure 2), but we also included 8 patients transplanted between 1984 and 1999, whose  
22 primary disease was FSGS and who underwent a subsequent transplant after the 2000s, and 23  
23 patients transplanted after 2018 who received a graft from a living kidney donor. Indeed, our  
24 biobank has been enriched in living kidney donors (17.5%) and FSGS patients (4.1%, Table 1).  
25 As previously mentioned, the lower number of grafts reported in 2000-2001 (Figure 2) does not

1 reflect a lower clinical activity but rather a difficulty to retrieve DNA samples for that time  
2 period with a progressive transition from paper-based to electronic biological records. Overall,  
3 the KiT-GENIE biobank thus provides an accurate snapshot of the transplantation activity in  
4 the Nantes clinical hospital between 2002 and 2018 (Figure 2).

5 Biopsy-proven rejections occurred in 16.6% KiT-GENIE recipients (Table 1), including 8.1%  
6 of biopsy-proven acute rejection. Graft failure (defined as return to dialysis or pre-emptive  
7 graft) happened for 21.3%. We observed some variability in the occurrence of rejection and  
8 graft failure events over the study period spanning over 15yrs (Figure S2). For example, we  
9 observed rejection rates up to 24% and 22% in 2003 and 2007 and as low as 10% and 12% in  
10 2002 and 2014, respectively. Similarly, we can note a graft failure rate as high as 44% and 33%  
11 in 2004 and 2008 and as low as 18% and 9% in 2010 and 2014, respectively.

### 12 ***Robustness and quality of KiT-GENIE genetic data***

13 First, we assessed the genotyping robustness and quality of our genetic data. The introduction  
14 of duplicates (n=58) and control samples (n=17) in the genotyping experiments has  
15 demonstrated a reproducibility rate of 99.9%, which corresponds to the advertised Axiom  
16 PMRA performance. In addition, we evaluated the error rate along the whole process of KiT-  
17 GENIE building (DNA extraction, sample handling, normalisation and genotyping) by  
18 comparing genetically-inferred sex and HLA alleles with medical records. We reported sex and  
19 HLA discrepancies for 119 individuals, corresponding to an error rate of 2.6%. To avoid pairing  
20 genotypes with incorrect phenotypes, we excluded these individuals from the final dataset  
21 (n=2,272 R and 1,945 D, including 1,969 pairs).

22 In a second step, we ran a careful standard QC protocol for SNPs and imputed over 8 million  
23 SNPs with accurate genotypes ( $r^2 > 0.8$ ) from 374K genotyped SNPs (Figure 1B). Before  
24 imputation, we observed an enrichment in rare variants in the dataset (Figure 1C dark grey) in  
25 accordance with the PMRA design. After imputation, the number of SNPs with a minor allele

1 frequency (MAF) >1% has significantly increased across the whole MAF spectrum (Figure 1C,  
2 light grey). All these high-quality genetic variants are available for subsequent association  
3 testing.

4 Finally, we compared *HLA* allele information available before and after imputation (Figure S3).  
5 We observed a considerable gain of annotation for the *HLA-C* gene (from 18% before  
6 imputation to 100% after imputation). Importantly, HLA imputation preserved the distribution  
7 of allelic diversity while increasing the available HLA information. As expected<sup>38-40</sup>, the *HLA-*  
8 *B* locus exhibits a higher level of polymorphism compared to other *HLA* genes such as *HLA-*  
9 *DQ* for example (Figure S3).

#### 10 ***Genetic ancestry and ancestry matching within pairs***

11 We performed a PCA to detect population stratification and capture ancestry-related variability  
12 in our genetic dataset. We first overlapped KiT-GENIE samples with the 1KGP reference  
13 populations<sup>36</sup> and determined the proportions of KiT-GENIE individuals in each large  
14 population using clusterisation methods<sup>41</sup> (Figure 3). The PCA showed a good separation of the  
15 1KGP individuals into 5 clusters representing the 5 reference ancestral populations (African,  
16 American, East Asian, European, and South Asian). The KiT-GENIE individuals spread across  
17 all reference populations, with a more compact distribution of donors centred on Europeans  
18 compared to recipients (Figures 3A and 3B). Indeed, a very large majority of individuals are of  
19 European ancestry with 93.4% of recipients and 98.2% of donors (Figures 3C and 3D), and  
20 recipients are genetically more diverse than donors with up to 3.4% recipients of African  
21 ancestry vs. 0.6% in donors.

22 When focusing on the genetic matching within D/R pairs (Figures 4A and 4B), 92.5% pairs  
23 (n=1,822) are composed of donor and recipient of European ancestry (zoom in on Figure 4C).  
24 Consequently, 147 pairs involve at least one non-European individual (Figure 4D) including  
25 135 pairs (6.8%) composed of one individual of European ancestry paired with a non-European

1 individual, and only 12 pairs (<1%) are composed of donor and recipient of non-European  
2 ancestry.

### 3 **Discussion**

4 The KiT-GENIE biobank is unique in France and a valuable resource for investigating genetic  
5 and immunogenetic factors impacting kidney transplantation-related outcomes thanks to its 3-  
6 layer design: DNA samples, large-scale genetic data and clinical data. It was built onto the  
7 DIVAT cohort, which represents a large and rich clinical data resource on kidney  
8 transplantation since the 1990s empowering us to investigate numerous phenotypes beyond  
9 rejection and graft survival. We consolidated a homogeneous and consistent monocentric  
10 cohort that accurately captures a snapshot of kidney transplantation clinical practice and  
11 medical records in Nantes over a 15yr period. In addition, DIVAT follow-up data are  
12 prospective and already provide us a hindsight up to 20-25 years for some patients, which is  
13 longer than most previous studies<sup>23,24,42</sup> and which offers an opportunity for revisiting the  
14 genetic data along accumulation of clinical visits.

15 KiT-GENIE is one of the few cohorts to contain both deceased and living (up to 17.5%) kidney  
16 donors thus offering a unique opportunity to study donor-recipient genetic interactions and  
17 mismatches. Most previous genomic analyses focused on deceased donors<sup>20,43,44</sup>, and  
18 comparing living and deceased donors could help better understanding the D/R interactions in  
19 the context of non-HLA mismatches. Indeed, living kidney donors represented 15% (n= 390)  
20 of kidney donations in 2020 in France<sup>7</sup> and 63% (n= 247) were genetically related to the  
21 recipient (*e.g.* parent, sibling, child, cousin)<sup>7</sup>. In addition, it provides an opportunity to further  
22 investigate the long-term kidney function decline and risk for chronic kidney disease after  
23 nephrectomy<sup>45,46,47</sup>. KiT-GENIE is also enriched in recipients suffering from focal segmental  
24 glomerulosclerosis (FSGS) as primary kidney disease. Idiopathic nephrotic syndrome with  
25 FSGS lesions is a glomerular nephropathy that recurs shortly after a first transplant in 30-50%

1 of cases, and in >90% after a subsequent transplant. A circulating factor has been suspected for  
2 numerous years without any definitive conclusions and there is an urgent need for alternative  
3 strategies to better understand the complex molecular mechanisms at stake and to define non-  
4 invasive predictive biomarkers for recurrence risk stratification<sup>48</sup>. KiT-GENIE therefore offers  
5 a unique opportunity to further investigate those mechanisms.

6 The KiT-GENIE genomic data were generated with the same experimental pipeline  
7 (genotyping chip, technological protocols, analytical pipeline) to ensure data homogeneity, and  
8 underwent robust quality controls to guarantee the highest data quality. In particular, our  
9 genotyping pipeline was highly reproducible (>99.9%). Additional checks on sex and HLA  
10 matching (genetically-inferred vs. medical records) prevented us from sample mishandling.  
11 Indeed, the establishment of KiT-GENIE involved the search for samples that were stored more  
12 than 20 years ago along with paper-based biological records, which increased the risk of human  
13 errors and underlines the importance of digitisation in biomedicine, along with robust QC.  
14 Beyond ensuring successful data conformity and sample traceability, SNP to HLA allele  
15 imputation also contributed to significantly enriching the immunogenetic genotypes within  
16 KiT-GENIE. Indeed, the HLA information reported within DIVAT clinical records over 20  
17 years suffers from incomplete data especially for *HLA-C* (84.2% and 78.6% missingness in  
18 recipients and donors, respectively - Table S2) as D/R matching has long been based on *HLA-*  
19 *A*, *-B*, and *-DR* loci only. SNP to HLA imputation retrospectively filled the missing gaps and  
20 provided homogenized HLA annotations for immunogenetic association testing and for  
21 evaluating HLA mismatches on 5 *HLA* genes instead of 3 genes. Finally, projecting SNP genetic  
22 data from KiT-GENIE and 1KGP individuals empowered us to describe the genetic ancestry of  
23 donors and recipients and to prevent population stratification, as collecting ancestry background  
24 is forbidden in France (law n°78-17, article 6 from January 6<sup>th</sup>, 1978).



1 The monocentric nature of KiT-GENIE represents its main limitation. Although it ensures  
2 homogeneous clinical practice and collection of medical records, there is a risk that results  
3 identified in our study might not be fully generalizable outside of Nantes. To limit this risk, we  
4 have established collaborations to replicate and validate our findings into independent cohorts.  
5 We also plan on opening the recruitment to other DIVAT or external collaborative centres. In  
6 early 2022, KiT-GENIE has joined the iGeneTRaiN international collaborative effort  
7 (International Genetics and Translational Research in Transplantation Network,  
8 <http://igenetrain.org/>), which is the largest genetic initiative focusing on better understanding  
9 transplant rejections and complications to date<sup>49,50</sup>. We have been contributing to the ongoing  
10 global meta-analysis effort to define genetic factors associated with kidney and other solid  
11 organ rejection and graft failure. Meta-analyses and patient recruitment will increase the  
12 statistical power for discoveries, while potentially increasing the genetic diversity of the study  
13 population. Indeed, KiT-GENIE is mostly composed of individuals of European ancestry  
14 (>90%), which might restrict the generalisation of our genetic associations in non-European  
15 populations.

16 To conclude, KiT-GENIE is one of the largest DNA biocollections with kidney donors and  
17 recipients (n=4,217, including 1,969 pairs) worldwide to date. Our cohort holds both clinical  
18 and genetic information of D/R pairs, which represents a valuable resource for studying the  
19 pathogenesis of kidney transplantation outcomes and complications. Several GWAS studies are  
20 currently underway in our team, but the rich DIVAT clinical database and the KiT-GENIE  
21 DNA biocollection will provide additional key opportunities for future research. The HLA and  
22 non-HLA genetic associations identified in KiT-GENIE will contribute (1) to better  
23 understanding major kidney transplantation clinical issues, (2) to revealing potential novel  
24 therapeutic targets to further characterize through functional studies, and (3) to identifying

1 predictive biomarkers that can stratify individual risk early before symptoms appearance and  
2 can improve clinical management (personalized and predictive medicine).

3

#### 4 **Data Availability Statement**

5 The KiT-GENIE data are available from our CR2TI team. Data and summary statistics are  
6 available upon reasonable request following approval from the KiT-GENIE steering committee  
7 and ethics committee to ensure data protection and privacy in compliance with French and  
8 European laws. Collaborations are encouraged through specific research projects using the KiT-  
9 GENIE data or through enriching the existing cohort with new patients. Potential collaborators  
10 are invited to contact the primary investigator Sophie Limou: [sophie.limou@univ-nantes.fr](mailto:sophie.limou@univ-nantes.fr).

11

#### 12 **References**

- 13 1. Hill NR, Fatoba ST, Oke JL, et al. Global Prevalence of Chronic Kidney Disease – A  
14 Systematic Review and Meta-Analysis. Remuzzi G, ed. *PLoS One*.  
15 2016;11(7):e0158765. doi:10.1371/journal.pone.0158765
- 16 2. Dalrymple LS, Katz R, Kestenbaum B, et al. Chronic Kidney Disease and the Risk of  
17 End-Stage Renal Disease versus Death. *J Gen Intern Med*. 2011;26(4):379-385.  
18 doi:10.1007/s11606-010-1511-x
- 19 3. Garcia GG, Harden P, Chapman J. The Global Role of Kidney Transplantation. *Kidney*  
20 *Blood Press Res*. 2012;35(5):299-304. doi:10.1159/000337044
- 21 4. Tonelli M, Wiebe N, Knoll G, et al. Systematic Review: Kidney Transplantation  
22 Compared With Dialysis in Clinically Relevant Outcomes. *Am J Transplant*.  
23 2011;11(10):2093-2109. doi:10.1111/j.1600-6143.2011.03686.x
- 24 5. Wong G, Howard K, Chapman JR, et al. Comparative Survival and Economic Benefits  
25 of Deceased Donor Kidney Transplantation and Dialysis in People with Varying Ages  
26 and Co-Morbidities. van Baal PHM, ed. *PLoS One*. 2012;7(1):e29591.  
27 doi:10.1371/journal.pone.0029591
- 28 6. Agence de la Biomédecine. *Press Communication 2021.*; 2022. [https://presse.agence-  
29 biomedecine.fr/une-hausse-de-193-des-greffes-dorganes-en-2021-grace-a-la-  
30 mobilisation-et-lengagement-de-tous-les-professionnels-de-la-chaine-du-don-a-la-  
31 greffe/](https://presse.agence-biomedecine.fr/une-hausse-de-193-des-greffes-dorganes-en-2021-grace-a-la-mobilisation-et-lengagement-de-tous-les-professionnels-de-la-chaine-du-don-a-la-greffe/) accessed 08/21/2022.
- 32 7. Agence de la Biomédecine. *Annual Report 2020 - Kidney.*; 2021. [https://rams.agence-  
33 biomedecine.fr/greffe-renale-0](https://rams.agence-biomedecine.fr/greffe-renale-0) accessed 21/08/2022.
- 34 8. Gondos A, Döhler B, Brenner H, Opelz G. Kidney Graft Survival in Europe and the

- 1 United States. *Transplant J.* 2013;95(2):267-274. doi:10.1097/TP.0b013e3182708ea8
- 2 9. Sasaki N, Idica A. The HLA-matching effect in different cohorts of kidney transplant  
3 recipients: 10 years later. *Clin Transpl.* 2010:261-282.  
4 <http://www.ncbi.nlm.nih.gov/pubmed/21696046>.
- 5 10. Levey AS, Coresh J. Chronic kidney disease. *Lancet.* 2012;379(9811):165-180.  
6 doi:10.1016/S0140-6736(11)60178-5
- 7 11. Vivian Tam, Nikunj Patel. Benefits and limitations of genome-wide association studies.  
8 *Nat Rev Genet.* 2019:467-484.
- 9 12. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. *Annu Rev*  
10 *Genomics Hum Genet.* 2016;17(1):95-115. doi:10.1146/annurev-genom-083115-  
11 022413
- 12 13. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep  
13 phenotyping and genomic data. *Nature.* 2018;562(7726):203-209. doi:10.1038/s41586-  
14 018-0579-z
- 15 14. Hanif Z, Sufiyan N, Patel M, Akhtar MZ. Role of biobanks in transplantation. *Ann Med*  
16 *Surg.* 2018;28:30-33. doi:10.1016/j.amsu.2018.02.007
- 17 15. Cambon-Thomsen A, Ducournau P, Gourraud P-A, Pontille D. Biobanks for Genomics  
18 and Genomics for Biobanks. *Comp Funct Genomics.* 2003;4(6):628-634.  
19 doi:10.1002/cfg.333
- 20 16. Yu Z, Jin J, Tin A, et al. Polygenic Risk Scores for Kidney Function and Their  
21 Associations with Circulating Proteome, and Incident Kidney Diseases. *J Am Soc*  
22 *Nephrol.* 2021;32(12):3161-3173. doi:10.1681/ASN.2020111599
- 23 17. Xu X, Eales JM, Jiang X, et al. Contributions of obesity to kidney health and disease:  
24 insights from Mendelian randomization and the human kidney transcriptomics.  
25 Antoniadou C, ed. *Cardiovasc Res.* December 2021. doi:10.1093/cvr/cvab357
- 26 18. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular  
27 disease risk prediction using automated machine learning: A prospective study of  
28 423,604 UK Biobank participants. Aalto-Setälä K, ed. *PLoS One.*  
29 2019;14(5):e0213653. doi:10.1371/journal.pone.0213653
- 30 19. Ghisdal L, Baron C, Lebranchu Y, et al. Genome-Wide Association Study of Acute  
31 Renal Graft Rejection. *Am J Transplant.* 2017;17(1):201-209. doi:10.1111/ajt.13912
- 32 20. Hernandez-Fuentes MP, Franklin C, Rebollo-Mesa I, et al. Long- and short-term  
33 outcomes in renal allografts with deceased donors: A large recipient and donor  
34 genome-wide association study. *Am J Transplant.* 2018;18(6):1370-1379.  
35 doi:10.1111/ajt.14594
- 36 21. Massart A, Ghisdal L, Viklicky O, Naesens M, Abramowicz D, Abramowicz M. Reply  
37 to Hernandez et al. - GWAS of acute renal graft rejection. *Am J Transplant.*  
38 2018;18(8):2098-2099. doi:10.1111/ajt.14877
- 39 22. Pihlstrøm HK, Mjøen G, Mucha S, et al. Single Nucleotide Polymorphisms and Long-  
40 Term Clinical Outcome in Renal Transplant Patients: A Validation Study. *Am J*  
41 *Transplant.* 2017;17(2):528-533. doi:10.1111/ajt.13995
- 42 23. Reindl-Schwaighofer R, Heinzl A, Signorini L, Thaunat O, Oberbauer R. Mechanisms  
43 underlying human genetic diversity: consequence for allograft antibody responses.  
44 *Transpl Int.* 2018;31:239-250. doi:10.1111/tri.13059

- 1 24. Steers NJ, Li Y, Drace Z, et al. Genomic Mismatch at LIMS1 Locus and Kidney  
2 Allograft Rejection . *N Engl J Med*. 2019;380(20):1918-1928.  
3 doi:10.1056/nejmoa1803731
- 4 25. Ba R, Geffard E, Douillard V, et al. Surfing the Big Data Wave: Omics Data  
5 Challenges in Transplantation. *Transplantation*. 2022;106(2):e114-e125.  
6 doi:10.1097/TP.0000000000003992
- 7 26. Reindl-Schwaighofer R, Heinzl A, Kainz A, et al. Contribution of non-HLA  
8 incompatibility between donor and recipient to kidney allograft survival: genome-wide  
9 analysis in a prospective cohort. *Lancet*. 2019;393(10174):910-917.  
10 doi:10.1016/S0140-6736(18)32473-5
- 11 27. Keating BJ, Van Setten J, Jacobson PA, et al. Design and implementation of the  
12 international genetics and translational research in transplantation network.  
13 *Transplantation*. 2015;99(11):2401-2412. doi:10.1097/TP.0000000000000913
- 14 28. Masset C, Kerleau C, Garandeau C, et al. A third injection of the BNT162b2 mRNA  
15 COVID-19 vaccine in kidney transplant recipients improves the humoral immune  
16 response. *Kidney Int*. 2021;100(5):1132-1135. doi:10.1016/j.kint.2021.08.017
- 17 29. Foucher Y, Lorent M, Albano L, et al. Renal transplantation outcomes in obese  
18 patients: a French cohort-based study. *BMC Nephrol*. 2021;22(1):79.  
19 doi:10.1186/s12882-021-02278-1
- 20 30. Dujardin A, Chesneau M, Dubois F, et al. Clinical and immunological follow-up of  
21 very long-term kidney transplant recipients treated with calcineurin inhibitors indicates  
22 dual phenotypes. *Kidney Int*. 2021;99(6):1418-1429. doi:10.1016/j.kint.2020.09.036
- 23 31. Applied Biosystems™. Axiom™ Precision Medicine Research Array. ThermoFisher.  
24 <https://www.thermofisher.com/order/catalog/product/902981#/902981>.
- 25 32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome  
26 association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-  
27 575. doi:10.1086/519795
- 28 33. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-  
29 Generation Reference Panels. *Am J Hum Genet*. 2018;103(3):338-348.  
30 doi:10.1016/j.ajhg.2018.07.015
- 31 34. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale  
32 sequence data. *Am J Hum Genet*. 2021;108(10):1880-1890.  
33 doi:10.1016/j.ajhg.2021.08.005
- 34 35. Zheng X, Shen J, Cox C, et al. HIBAG - HLA genotype imputation with attribute  
35 bagging. *Pharmacogenomics J*. 2014;14(2):192-200. doi:10.1038/tpj.2013.18
- 36 36. Gibbs RA, Boerwinkle E, Doddapaneni H, et al. A global reference for human genetic  
37 variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
- 38 37. Douillard V, Castelli EC, Mack SJ, et al. Approaching Genetics Through the MHC  
39 Lens: Tools and Methods for HLA Research. *Front Genet*. 2021;12(December):1-13.  
40 doi:10.3389/fgene.2021.774916
- 41 38. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression,  
42 interaction, diversity and disease. *J Hum Genet*. 2009;54(1):15-39.  
43 doi:10.1038/jhg.2008.5
- 44 39. Olson E, Geng J, Raghavan M. Polymorphisms of HLA-B: influences on assembly and  
45 immunity. *Curr Opin Immunol*. 2020;64:137-145. doi:10.1016/j.coi.2020.05.008

- 1 40. Douillard V, Castelli EC, Mack SJ, et al. Approaching Genetics Through the MHC  
2 Lens: Tools and Methods for HLA Research. *Front Genet.* 2021;12.  
3 doi:10.3389/fgene.2021.774916
- 4 41. Nielsen F. Hierarchical Clustering. In: ; 2016:195-211. doi:10.1007/978-3-319-21903-  
5 5\_8
- 6 42. Mesnard L, Muthukumar T, Burbach M, et al. Exome Sequencing and Prediction of  
7 Long-Term Kidney Allograft Function. *PLoS Comput Biol.* 2016;12(9):1-15.  
8 doi:10.1371/journal.pcbi.1005088
- 9 43. O'Brien RP, Phelan PJ, Conroy J, et al. A genome-wide association study of recipient  
10 genotype and medium-term kidney allograft function. *Clin Transplant.*  
11 2013;27(3):379-387. doi:10.1111/ctr.12093
- 12 44. Divers J, Ma L, Brown WM, et al. Genome-wide association study for time to failure  
13 of kidney transplants from African American deceased donors. *Clin Transplant.*  
14 2020;34(6). doi:10.1111/ctr.13827
- 15 45. Lam NN, Lloyd A, Lentine KL, et al. Changes in kidney function follow living donor  
16 nephrectomy. *Kidney Int.* 2020;98(1):176-186. doi:10.1016/j.kint.2020.03.034
- 17 46. Mjøen G, Hallan S, Hartmann A, et al. Long-term risks for kidney donors. *Kidney Int.*  
18 2014;86(1):162-167. doi:10.1038/ki.2013.460
- 19 47. Muzaale AD, Massie AB, Wang M-C, et al. Risk of end-stage renal disease following  
20 live kidney donation. *JAMA.* 2014;311(6):579-586. doi:10.1001/jama.2013.285141
- 21 48. Rudnicki M. FSGS Recurrence in Adults after Renal Transplantation. *Biomed Res Int.*  
22 2016;2016:1-7. doi:10.1155/2016/3295618
- 23 49. International Genetics & Translational Research in Transplantation Network  
24 (iGeneTRAiN). Design and Implementation of the International Genetics and  
25 Translational Research in Transplantation Network. *Transplantation.*  
26 2015;99(11):2401-2412. doi:10.1097/TP.0000000000000913
- 27 50. Fishman CE, Mohebnasab M, van Setten J, et al. Genome-Wide Study Updates in the  
28 International Genetics and Translational Research in Transplantation Network  
29 (iGeneTRAiN). *Front Genet.* 2019;10:1084. doi:10.3389/fgene.2019.01084

30

1 **Acknowledgements.** We thank everyone who helped in the design, collection, genotyping  
2 experiments, data cleaning and analyses. We are especially indebted to all donors and recipients  
3 who participated in this study, to the clinical staff and physicians who care and manage the  
4 patients on a daily basis and helped recruiting the patients, and to research associates who  
5 participated in the data collection. Clinical data were collected from the French DIVAT  
6 multicentric prospective cohort of kidney and/or pancreatic transplant recipients by focusing on  
7 the Nantes centre (www.divat.fr, n° CNIL 914184, ClinicalTrials.gov identifier:  
8 NCT02900040). The analysis and interpretation of these data are the responsibility of the  
9 authors. We would like to acknowledge the local blood bank and HLA typing lab  
10 (*Etablissement Français du Sang* [EFS] Nantes) for facilitating the access to DNA samples.  
11 Finally, we thank the GenoBiRD and Curie genomic platforms for technical support and GWAS  
12 genotyping.

13 **Author contribution.** RB, AD and SL were involved in the design, conception,  
14 implementation and data collection of KiT-GENIE. RB, AD, CC, SLBB, SS, PG, and CP were  
15 involved in DNA collection and normalisation. RB, AD, VM, MM, VD, OR and SL were  
16 involved in quality controls, statistical description, imputation of missing data, PCA analysis  
17 and ancestry analysis. CK realized clinical data extraction from the DIVAT database. MG and  
18 GB provided clinical guidance and ensured access to the DIVAT clinical data. RB and SL wrote  
19 the article. SL, PAG and NV supervised the building of KiT-GENIE, revised and formalized  
20 the scientific content of the article and mentored RB. All authors have read and approved the  
21 final version of the manuscript.

22 **Ethical approval.** Written consent was obtained from each participant for access to clinical  
23 data and participation to DNA analyses (DIVAT n° CNIL 914184). Patients and/or the public  
24 were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

25

1 **Fundings.** The KiT-GENIE cohort has been funded by several entities: (1) The *Etoiles*  
2 *Montantes* funding by the Pays de la Loire region (n°2018-09998), (2) The IRCT Dialyse  
3 research project by the Société Francophone de Néphrologie, Dialyse et Transplantation  
4 (SFNDT), (3) The *Greffe* research project by French *Agence de la Biomédecine* (ABM,  
5 n°18GREFFE014). In addition, this work has benefited from government support through the  
6 National Research Agency (ANR) under the future investment program (n°ANR-17-RHUS-  
7 0010) and from the European Union's Horizon 2020 Research and Innovation Programme  
8 (n°754995). Finally, we thank the Roche Pharma, Novartis, Astellas, Chiesi, Sandoz and Sanofi  
9 laboratories for supporting the DIVAT cohort as the CENTAURE Foundation  
10 (<http://www.fondation-centaure.org>). The authors have declared no conflict of interest.

1 **Figure legends**

2 **Figure 1: Building the KiT-GENIE DNA biocollection and GWAS SNP data**

3 **A Establishment of the KiT-GENIE DNA biocollection.** 5,946 individuals were initially  
4 extracted from DIVAT Nantes for investigation of DNA samples. 1,222 people were excluded  
5 due to missing or deteriorated DNA samples, and DNA was collected for 4,723 individuals.  
6 Single deceased D for whom R DNA could not be retrieved (n=147) were excluded from  
7 GWAS genotyping. As a result, 4,577 individuals were genotyped, but 360 individuals were  
8 then excluded due to failed experiments or GWAS QC procedures. Finally, 4,217 individuals  
9 with accurate genomic and clinical data were included in the KiT-GENIE biocollection. It is  
10 important to note that some recipients received multiple grafts and that some deceased donors  
11 have provided an organ to two different recipients, therefore impacting the number of pairs. **B**  
12 **Analytical steps of GWAS genotyping data processing.** The Axiom PMRA GWAS  
13 genotyping chip covers 902,506 SNPs. 831,213 SNPs passed the primary technological QC  
14 steps. Following the standard GWAS guidelines, 161,185 independent variants were used to run  
15 the PCA excluding SNPs on sexual chromosomes, from the highly polymorphic HLA region,  
16 or in high LD level. 374,743 frequent high-quality SNPs were included in the imputation  
17 pipeline to generate a total of 8 million SNPs with high accuracy. **C Common SNP enrichment**  
18 **through SNP imputation.** Before imputation, we observed an enrichment in rare functional  
19 genetic variants (MAF<1%) in accordance with the Axiom PMRA GWAS genotyping chip  
20 design. After imputation, we significantly increased the number of SNPs with a MAF >1%  
21 available for subsequent association testing.

22 QC, Quality control; PCA, Principal component analysis; LD, linkage disequilibrium; MAF, Minor allele  
23 frequency; HWE, Hardy-Weinberg equilibrium; 1KGP, 1000 Genomes Project

24

25 **Figure 2: Distribution of KiT-GENIE recipients according to the year of graft**



1 KiT-GENIE patients were mostly transplanted between 2002 and 2018. Due to the enrichment  
2 in FSGS recipients (circles) and LKD (triangles), we have 8 patients transplanted between 1984  
3 and 1999, whose primary disease was FSGS and who underwent a subsequent transplant after  
4 the 2000s, and 23 patients transplanted after 2018 who received a graft from a LKD. In addition,  
5 the progressive transition from paper-based to electronic medical records in the early 2000s led  
6 to a low number of DNA samples retrieved in 2000-2001.

7 Focal Segmental GlomeruloSclerosis, FSGS; Living Kidney Donor (LKD)

8

9 **Figure 3: PCA projection of KiT-GENIE individuals with the 1,000 Genomes Project**  
10 **reference individuals**

11 **A** PCA projection of KiT-GENIE recipients (black stars) with 1KGP individuals from 5 large  
12 reference populations (colour-coded circles). **B** PCA projection of KiT-GENIE donors (black  
13 stars) with 1KGP individuals from 5 large reference populations (color-coded circles). **C**  
14 Proportion of KiT-GENIE recipients in each large reference population with their  
15 corresponding colour code. **D** Proportion of KiT-GENIE donors in each large reference  
16 population with their corresponding colour code.

17 PCA, principal component analysis; 1KGP, 1,000 Genomes Project; AFR, African; AMR, American; EAS, East  
18 Asian; SAS, South Asian; EUR, European

19

20 **Figure 4: Genetic ancestry matching between donors and recipients**

21 **A** Distribution of genetic ancestry within the D/R pairs. The EUR/EUR pairs are predominant  
22 (92.53%). **B** Genetic ancestry matching for all D/R pairs projected in the 3D space composed  
23 of the first three principal components. **C** Genetic ancestry matching for the EUR/EUR D/R  
24 pairs (n=1,822). **D** Genetic ancestry matching for D/R pairs composed of at least one non-EUR  
25 individuals (n=147).

26 AFR, African; AMR, American; EAS, East Asian; SAS, South Asian; EUR, European

1 **Table 1 Individuals and graft clinical characteristics**

<b>Individuals baseline characteristics (n = 4,217)</b>	
<b>Recipients (n = 2,272)</b>	
<b>Gender</b> (male)	1,436 (63.2%)
<b>Age</b> (years)	18-87 yo (mean 51 yo)
<b>Body Mass Index</b> (kg/m)	14–2 - 50.3 (mean 24.4)
<b>Primary disease</b>	
Unknown aetiology	211 (9.3%)
Chronic glomerulonephritis	583 (25.7%)
Chronic interstitial nephritis and malformative uropathy	909 (40.0%)
Renal vascular disease	191 (8.4%)
Diabetes	378 (16.6%)
<b>Focal segmental glomerulosclerosis (FSGS)</b>	91 (4.1%)
<b>Renal replacement therapy</b>	
Pre-emptive transplantation	470 (20.7%)
Peritoneal dialysis	198 (8.7%)
Haemodialysis	1,599 (70.5%)
<b>Donors (n = 1,945)</b>	
<b>Gender</b> (male)	1,126 (57.9%)
<b>Age</b> (years)	4-90 yo (mean 51 yo)
<b>Body Mass Index</b> (kg/m)	13–9 - 58.8 (mean 25.4)
<b>Donor type</b>	
Living	340 (17.5%)
Deceased	1,605 (82.5%)

<b>Inclusion criteria</b>	
Standard criteria donors	954 (49.4%)
Extended criteria donors	647 (33.5%)
Missing values	344 (17.1%)
<b>Kidney function</b>	
Serum creatinine ( $\mu\text{mol/l}$ )	20-813 (mean 85.6)
Urea ( $\text{mmol/l}$ )	0-38 (mean 3.6)
Proteinuria	
absence	1400 (71.9%)
presence	365 (18.7%)
missing values	180 (9.2%)
<b>Graft baseline characteristics (n = 2,272)</b>	
<b>Year of graft</b>	1984-2020
<b>Rank</b>	
1	1,947 (85.7%)
2 - 5	325 (14.3%)
<b>Type</b>	
Kidney	2,009 (88.4%)
Kidney-Pancreas	257 (11.3%)
Double kidneys, Kidney-Liver, Kidney-Heart	6 (< 1%)
<b>Cold Ischemia (hours)</b>	0.5 – 48 (mean 16)
<b>Delayed renal graft function (days)</b>	0 – 86 (mean 6.9)
<b>Biopsy-proven rejection</b>	377 (16.6%)
<b>Graft failure</b>	485 (21.3%)
<b>HLA incompatibilities (<i>HLA-A, -B, and -DR</i>)</b>	
0 (full match)	105 (4.4%)

1 - 3	1,007 (45.0%)
4 - 6	1,153 (50.2%)
Missing values	7 (0.4%)
<b>Follow-up time (months)*</b>	0 – 300 (mean 80)

1

2

### 3 **Supplementary material content**

4 **Table S1:** DNA quality and quantity metrics.

5 **Table S2:** Missingness of KiT-GENIE clinical variables related to recipients, donors,  
6 transplantation and pre/post transplantation events.

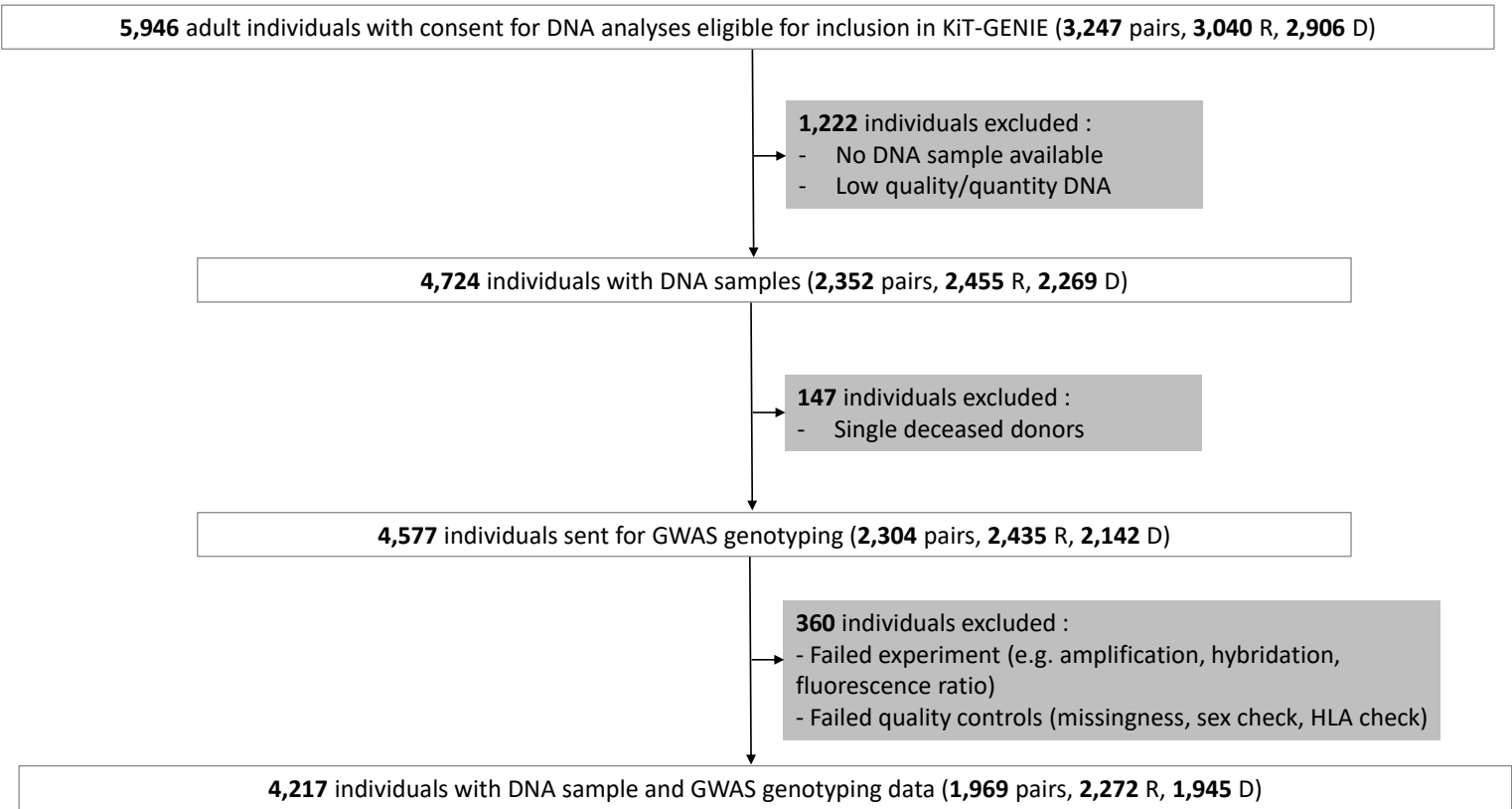
7 **Figure S1:** Evolution of mismatch number for typed *HLA-A*, *-B* and *-DR* genes over the study  
8 period.

9 **Figure S2:** Proportions of rejection and graft loss events over the study period.

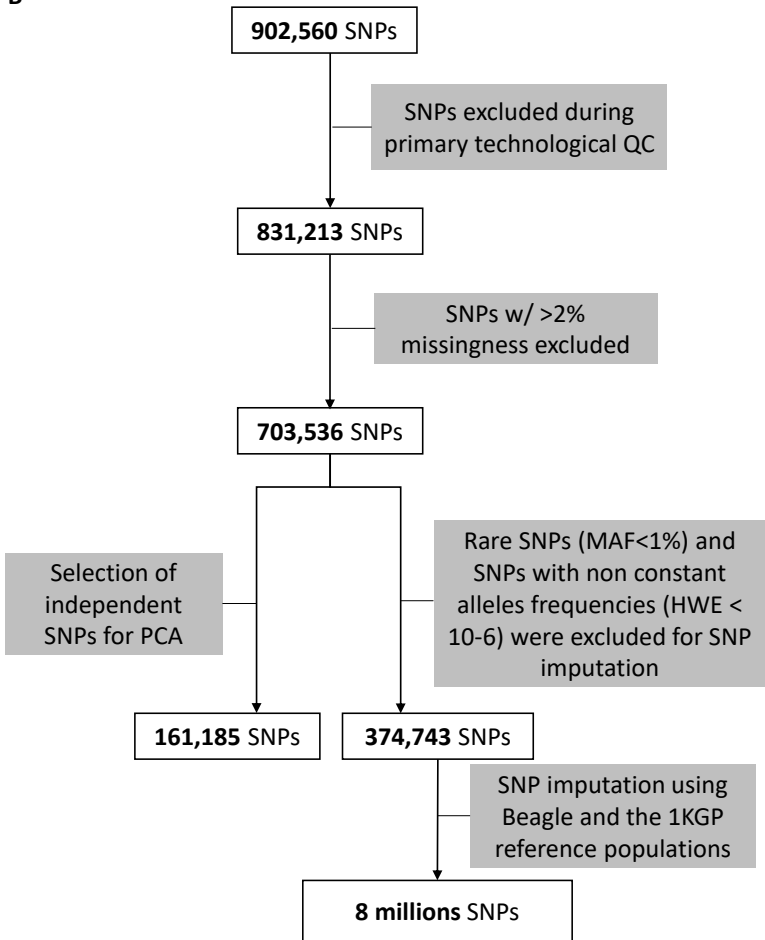
10 **Figure S3:** Distribution of *HLA* alleles in KiT-GENIE individuals according to imputed vs.  
11 typed *HLA-A*, *-B*, *-C*, *-DR* and *-DQ* alleles.

12

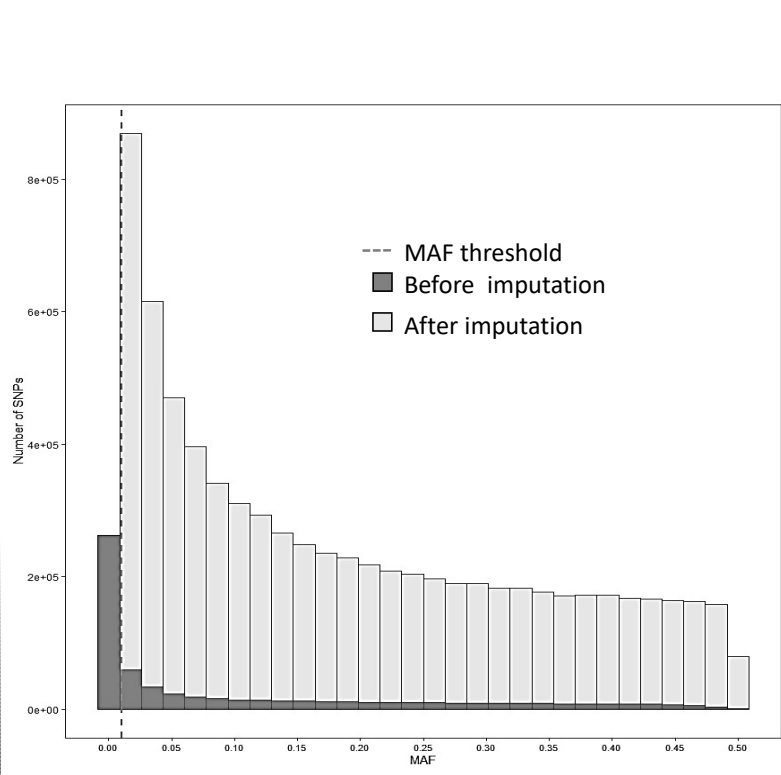
**A**

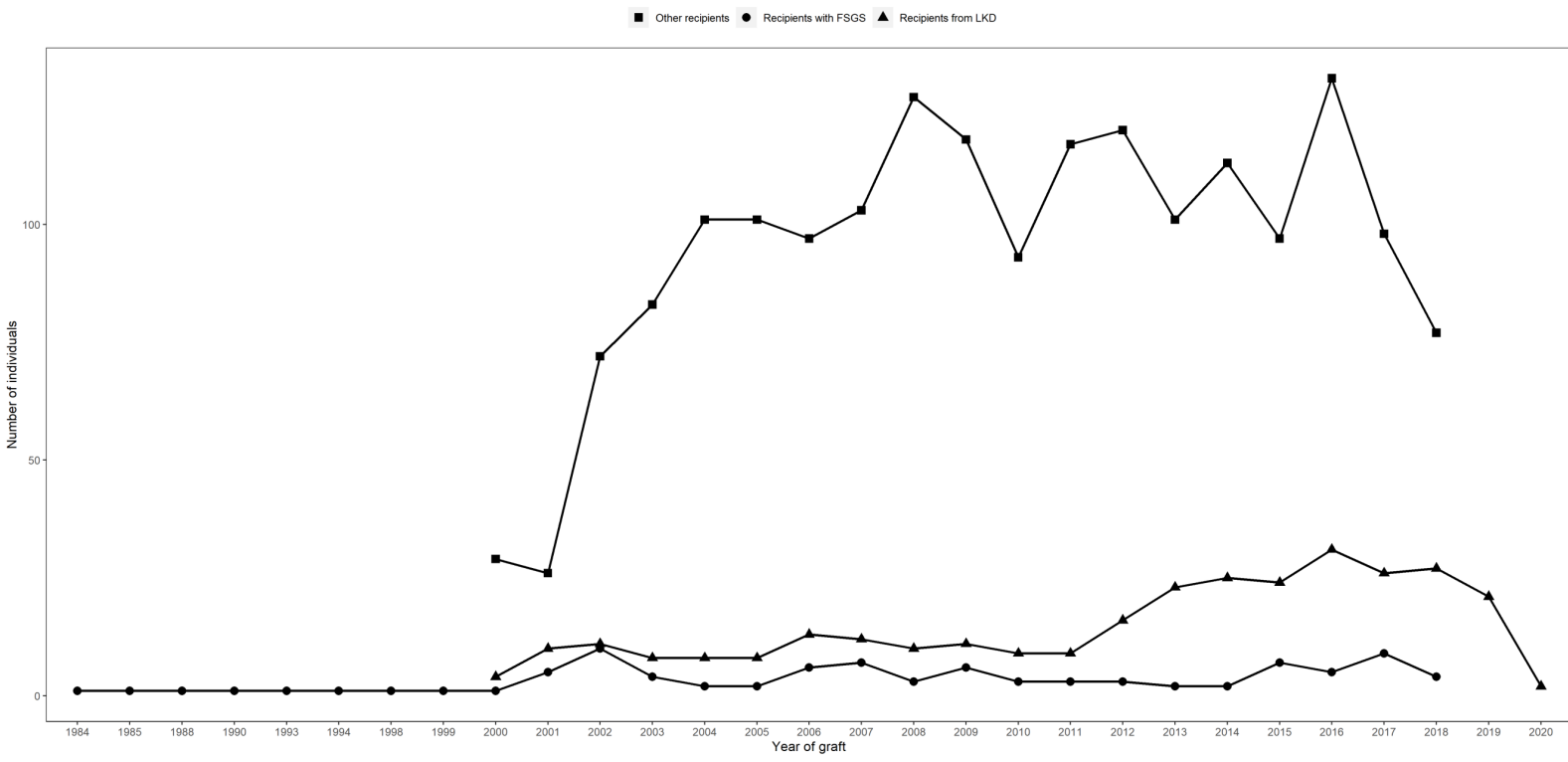


**B**



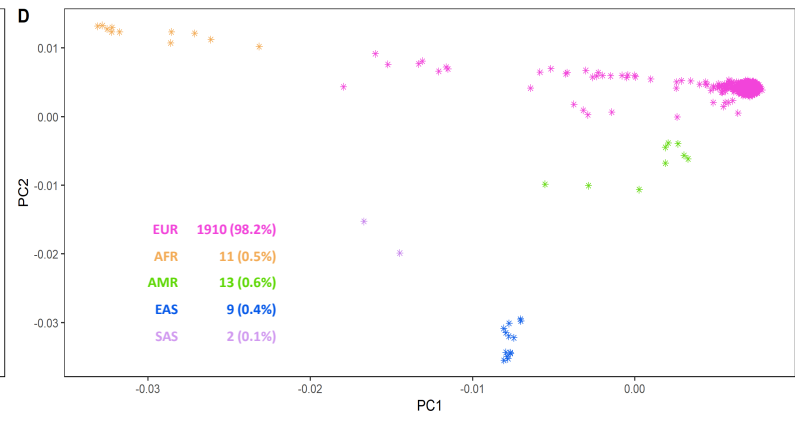
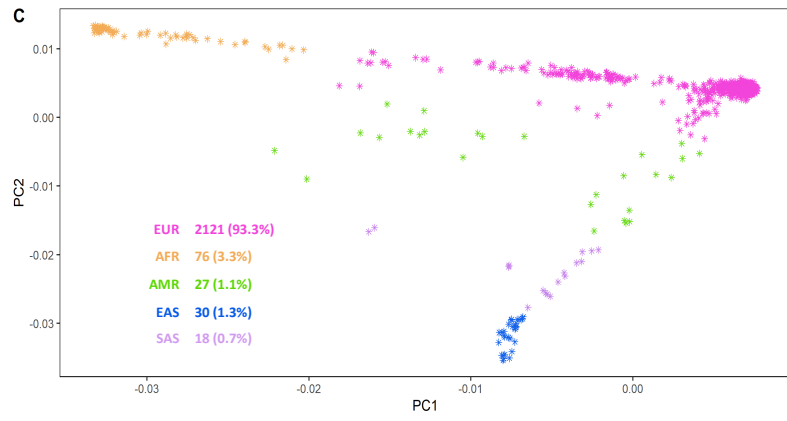
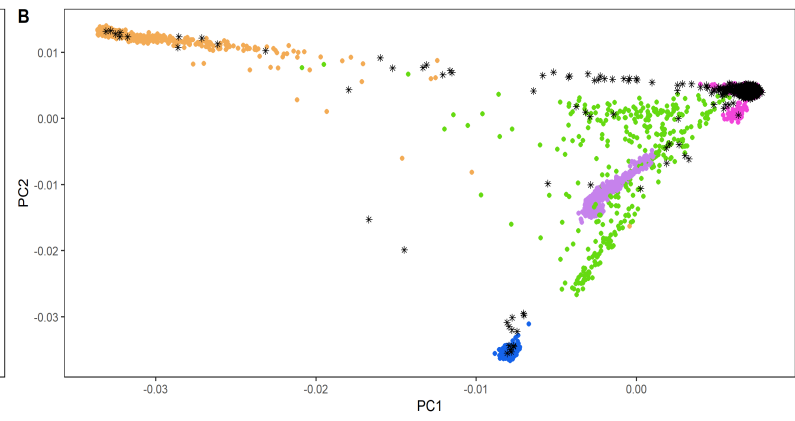
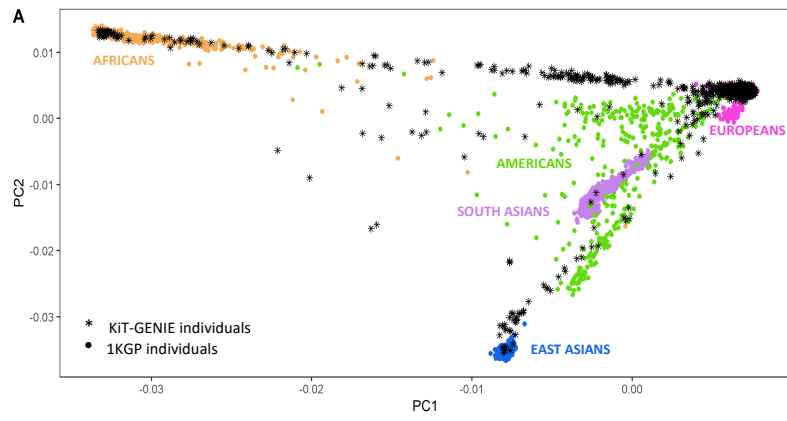
**C**



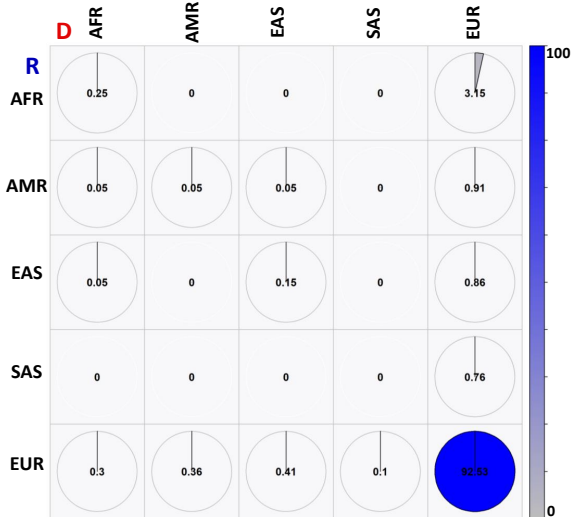


**Recipients**

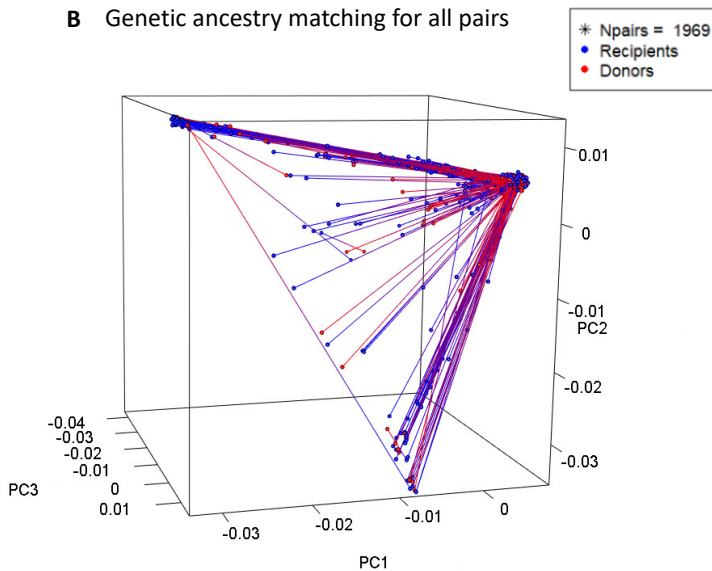
**Donors**



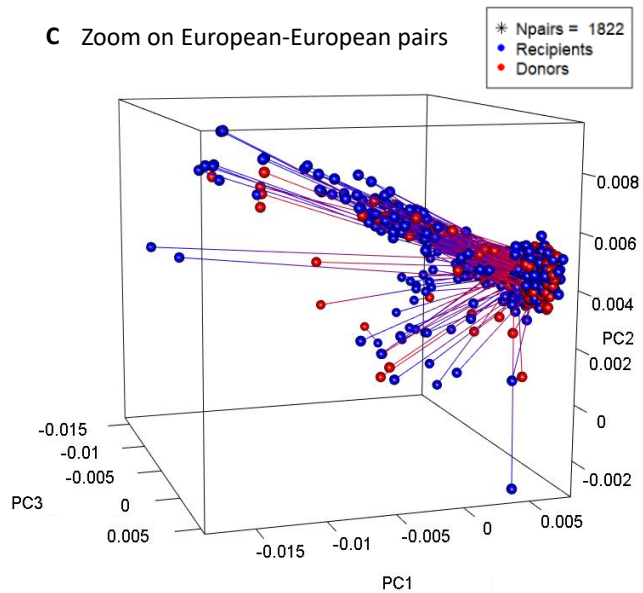
**A** Distribution of genetic ancestry within D/R pairs



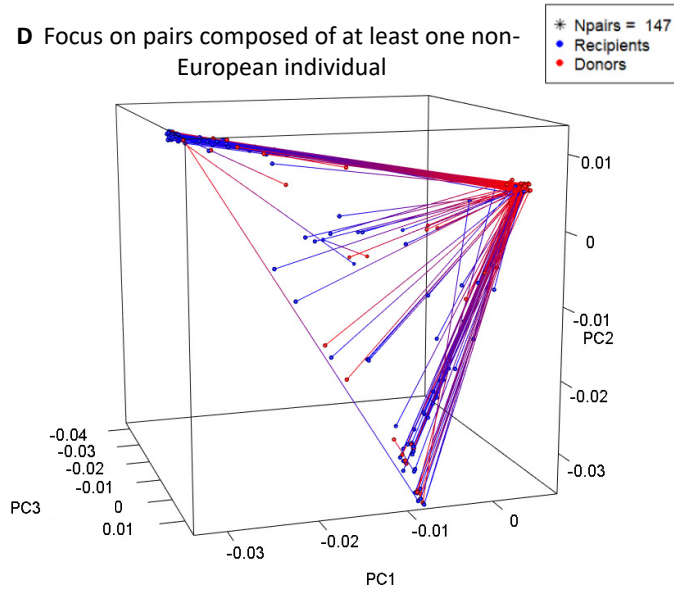
**B** Genetic ancestry matching for all pairs



**C** Zoom on European-European pairs



**D** Focus on pairs composed of at least one non-European individual





## Supplementary file

### KiT-GENIE, the French Genetics of Kidney Transplantation Study

**Table S1: DNA quality and quantity metrics**

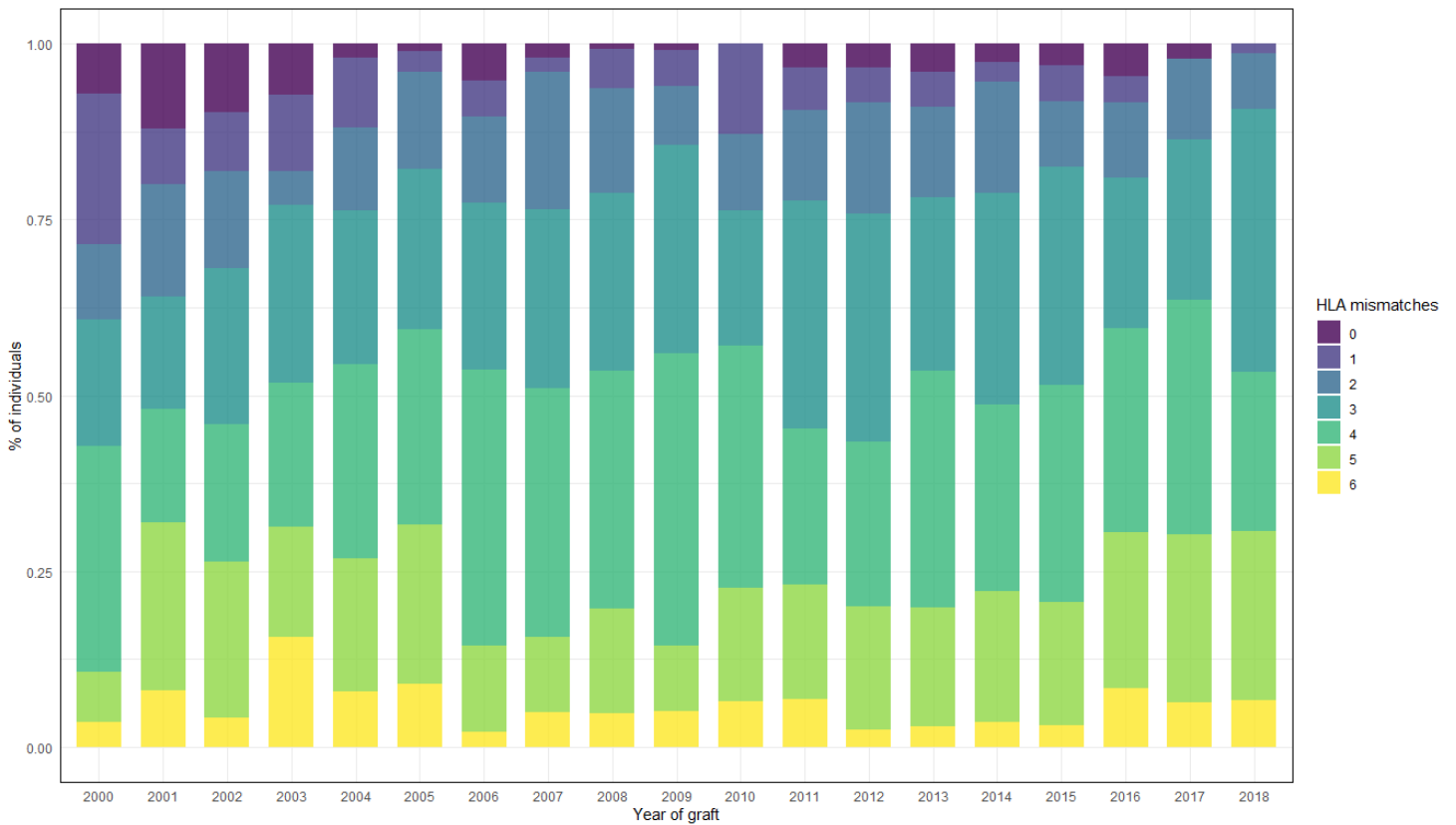
	<b>KiT-GENIE DNA biocollection</b> (n=4,724)
<b>DNA Volume (µl)</b> Mean (SD) Range	110.37 (61.11) 10.00 - 300.00
<b>Concentration (ng/µl)</b> Mean (SD) Range	256.64 (291.77) 11.03 - 3,707.85
<b>260/230 ratio</b> Mean (SD) Range	1.55 (0.72) 0.10 - 3.10
<b>260/280 ratio</b> Mean (SD) Range	1.86 (0.12) 1.00 - 3.30

**Table S2: Data missingness of KiT-GENIE clinical variables related to recipients, donors, transplantation and pre/post transplantation events**

<b>Description</b>	<b>Missing value (%)</b>
<b>Recipient characteristics at baseline</b>	
Age	0
Sex	0
Initial disease	0.16
Height	0.16
HLA-A1 typing	0.16
HLA-A2 typing	0.16
HLA-B1 typing	0.16
HLA-B2 typing	0.16
HLA-DR1 typing	0.16
HLA-DR2 typing	0.16
HLA-C1 typing	84.18
HLA-C2 typing	84.18
HLA-DP1 typing	87.51
HLA-DP2 typing	87.51
HLA-DQa1 typing	98.53
HLA-DQa2 typing	98.53
HLA-DQb1 typing	1.80
HLA-DQb2 typing	1.80
Blood type	0.16
Weight	0.22
BMI	0.22
HCV serology	0.22
HBs antigens	0.27
CMV serology	0.38
HIV antibodies	0.55
Pneumocystis Prophylactic treatment	0.60
Pneumocystis Prophylactic treatment	0.60
CMV Prophylactic treatment	0.87
CMV Prophylactic treatment	0.87
EBV serology	0.93
HBs antibodies	1.69
Transfusion before blood collection	2.73
Time from transplant to last follow-up	6.60
Number of follow-up months	6.60
<b>Donor characteristics</b>	
Sex	0
Donor type	0
Cadaver type	0
CMV serology	0
HBs antigens	0
Blood type	0
Age	0.05
HBc antibodies	0.05
HLA-A1 typing	0.05
HLA-A2 typing	0.05
HLA-B1 typing	0.05
HLA-B2 typing	0.05
HLA-DR1 typing	0.05
HLA-DR2 typing	0.05

HLA-C1 typing	78.56
HLA-C2 typing	78.56
HLA-DP1 typing	81.07
HLA-DP2 typing	81.07
HLA-DQa1 typing	98.42
HLA-DQa2 typing	98.42
HLA-DQb1 typing	2.13
HLA-DQb2 typing	2.13
HCV serology	0.11
Cardiac arrest prior to blood collection	0.16
EBV serology	0.16
HIV antibodies	0.16
Creatinine level before blood collection	0.27
Blood pressure	3.44
Height	5.78
Weight	5.78
BMI	5.78
Proteinuria	9.06
Hematuria	11.89
Transfusion before blood collection	12.17
HBs antibodies	13.97
Urea	16.58
Cause of death	0
Inclusion criteria	16.97
Prescription of vasopressor drug	17.46
History of diabetes	18
History of dyslipidemia	21.82
Last diuresis before blood collection	21.99
Smoking habits	55.54
<b>Transplantation characteristics</b>	
Transplant ID	0
Recipient ID	0
Graft rank	0
Year	0
Center	0
Graft type	0
Rank	0
Cold ischemia	0
Purification techniques	0.16
Number of donor/recipient matches on locus A	0.16
Number of donor/recipient matches on locus B	0.16
Number of donor/recipient matches on locus DR	0.16
Number of donor/recipient incompatibilities on locus A	0.16
Number of donor/recipient incompatibilities on locus B	0.16
Number of donor/recipient incompatibilities on locus DR	0.16
Number of ABO incompatibilities between donor and recipient	0.16
Time from registration to transplantation	0.27
Kidney Type	0.38
Recovery of renal function	1.69
Pre- or post-transplant probe	2.13
Number of graft arteries	2.29
Location of graft implantation	3.27
Graft infusion fluid	3.55
Post-Transplant Dialysis	3.60
Post-Transplant Dialysis in 2 classes	3.60

Graft preservation fluid	4.31
Perfusion machine	16.80
Time from dialysis to transplantation	0.07
Hospital discharge	31.70
Warm Ischemia	57.45
Time from dialysis to transplantation (if several kidney transplants)	9.63
Transplantectomy	94.65
Time between transplantation and transplantectomy	0
<b>Pre-transplant events</b>	
First transplantation	0
Third transplantation	0
Second transplantation	0
Fourth transplantation	0
Prior pregnancies	18.10
Number of prior pregnancies	18.70
Year of first graft	0
First graft failure	0.57
Duration of first graft	3.12
Cause of first graft failure	8.36
First transplantectomy	20.68
Time between first transplantation and transplantectomy	1.15
Year of second graft	0
Second graft failure	0
Duration of second graft	0
Cause of second graft failure	6.06
Second transplantectomy	25.76
Time between second transplantation and transplantectomy	1
Third graft failure	0
Year of third graft	0
Duration of third graft	0
Cause of third graft failure	14.29
Third transplantectomy	28.57
Time between third transplantation and transplantectomy	0
Year of fourth graft	0
Fourth graft failure	0
Duration of fourth graft	0
Cause of fourth graft failure	0
Fourth transplantectomy	0
Time between fourth transplantation and transplantectomy	0
<b>Post-transplant events</b>	
Rejection	0
Number of rejection	0
Return to dialysis	0
Death	0
Death with a functional kidney	0
Patient outcome	0
Occurrence of a post-transplant event	0
Time between transplantation and a post-transplant event	0
Loss of sight	0
Graft failure	26.35
Time between transplantation and graft failure	81.23
Cause of graft failure	81.23
Time between transplant and death	81.56
Cause of death	86.20

**A****B**

	F value	P(>F)
HLA mismatches	48.15	<b>4.81e-10</b>
Year of graft	1.21	0.27
HLA mismatches:Year of graft	0.47	0.97

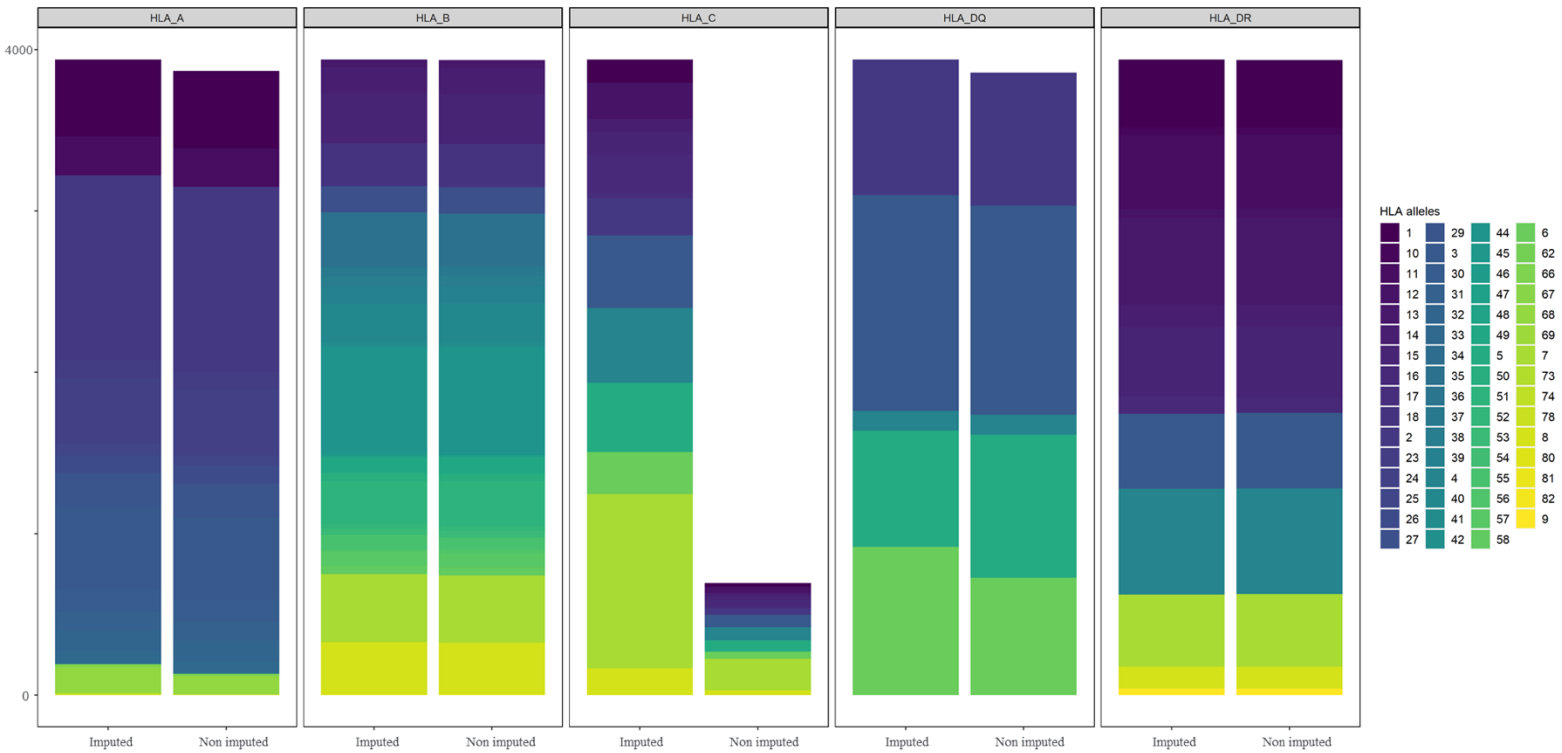
**Figure S1: Evolution of mismatch number for typed *HLA-A*, *-B* and *-DR* genes over the study period.**

**A** The proportion of *HLA* mismatches between 2000 and 2018 varies between 0 (full match for the 6 *HLA* alleles) to 6 (full mismatch for the 6 *HLA* alleles), and transplants with 4 *HLA* mismatches were predominant in KiT-GENIE. **B** A two-factor Anova comparison of the proportion of individuals according to the number of *HLA* mismatches and the year of graft revealed that *HLA* mismatches are constantly distributed over the years between 2000 and 2018 ( $p=0.27$ ). Overall, there were less individuals with 0-2 *HLA* incompatibilities than with 3-6 *HLA* incompatibilities ( $p=4.81e-10$ ). Finally, there was no correlation between the number of mismatches and the graft year ( $p=0.97$ ).



**Figure S2: Proportions of rejection and graft loss events over the study period.**

**A** Proportion of individual who experienced biopsy-proven rejection for each year between 2002 and 2018. This time period represents an accurate snapshot of the DIVAT-Nantes center transplantation activity. **B** Proportion of individual who experienced graft loss (defined as return to dialysis or pre-emptive graft) for each year between 2002 and 2018.



**Figure S3: Distribution of *HLA* alleles in KiT-GENIE individuals according to imputed vs. typed *HLA-A*, *-B*, *-C*, *-DR* and *-DQ* alleles.**

We compared the *HLA* alleles distribution in KiT-GENIE before (typed or non-imputed) and after *HLA* imputation. The imputation preserved the allelic diversity and distribution while increasing the available information. Importantly, considerable *HLA-C* information was gained with up to 80% added knowledge. As expected, *HLA-B* alleles are more diverse compared to other genes such as *HLA-DQ*.

# RÉSULTATS GWAS

---

Cette partie présente les GWAS préliminaires qui ont été menées sur les receveurs et les donneurs séparément afin de déterminer s'il existe des associations entre variant génétique et survie du greffon. Ces résultats ont été générés dans le cadre du projet de stage d'un étudiant avec qui j'ai collaboré au cours de ma dernière année de thèse et qui prends la suite du projet KiT-GENIE pour son doctorat. Ces résultats n'ont pas encore été publiés dans un journal scientifique mais ont fait l'objet d'un poster lors du congrès annuel de la Société Francophone de Transplantation 2021 à Genève.

## 2.1 Association entre variant génétique et retour en dialyse chez les receveurs et chez les donneurs KiT-GENIE

Les GWAS présentées ci-dessous répondent à l'hypothèse suivante : Existe-t-il une association significative entre les variants génétiques du donneur ou du receveur et le retour en dialyse à plus d'un an post-transplantation. Ces GWAS n'ont été effectuées que sur des individus d'ancestralité européennes. Pour chaque type d'individus (receveurs et donneurs), une GWAS a été effectuée avant imputation et après imputation. Les résultats sont présentés sous la forme d'un QQplot qui permet de visualiser la déviation des p-values mesurées en fonction des p-values attendues sous l'hypothèse nulle de non-association et d'un Manhattan plot qui représente les p-values calculées pour chaque SNP selon leur position chromosomique. Le seuil de significativité est fixé à  $5.10^{-8}$  et correspond au seuil de bonferroni et le seuil de suggestivité est arbitrairement fixé à  $10^{-6}$ . Un point se trouvant entre les deux lignes correspond à un SNP potentiellement associé au phénotype sans que l'association ne puisse être considérée comme significative.

### 2.1.1 Association entre variants génétiques et retour en dialyse pour les receveurs et les donneurs avant imputation

La figure 2.1 montre les résultats de GWAS avant l'imputation des données SNP (358 101 SNP). Lorsque l'on s'intéresse aux QQplots, nous observons une corrélation normale entre les p-values théoriques et celles mesurées (pas de SNP qui se démarque) avec toutefois une légère déflation chez les receveurs. Aucun des points ne dépasse le seuil de significativité pour les rece-



veurs comme pour les donneurs. Chez les receveurs, nous pouvons observer un point qui dépasse le seuil de suggestivité au niveau du chromosome 11. Ceci suggère qu'il y a potentiellement un signal dans cette région du génome. Pour les donneurs, des signaux dépassant le seuil de suggestivité sont détectés dans les chromosomes 1 et 10.

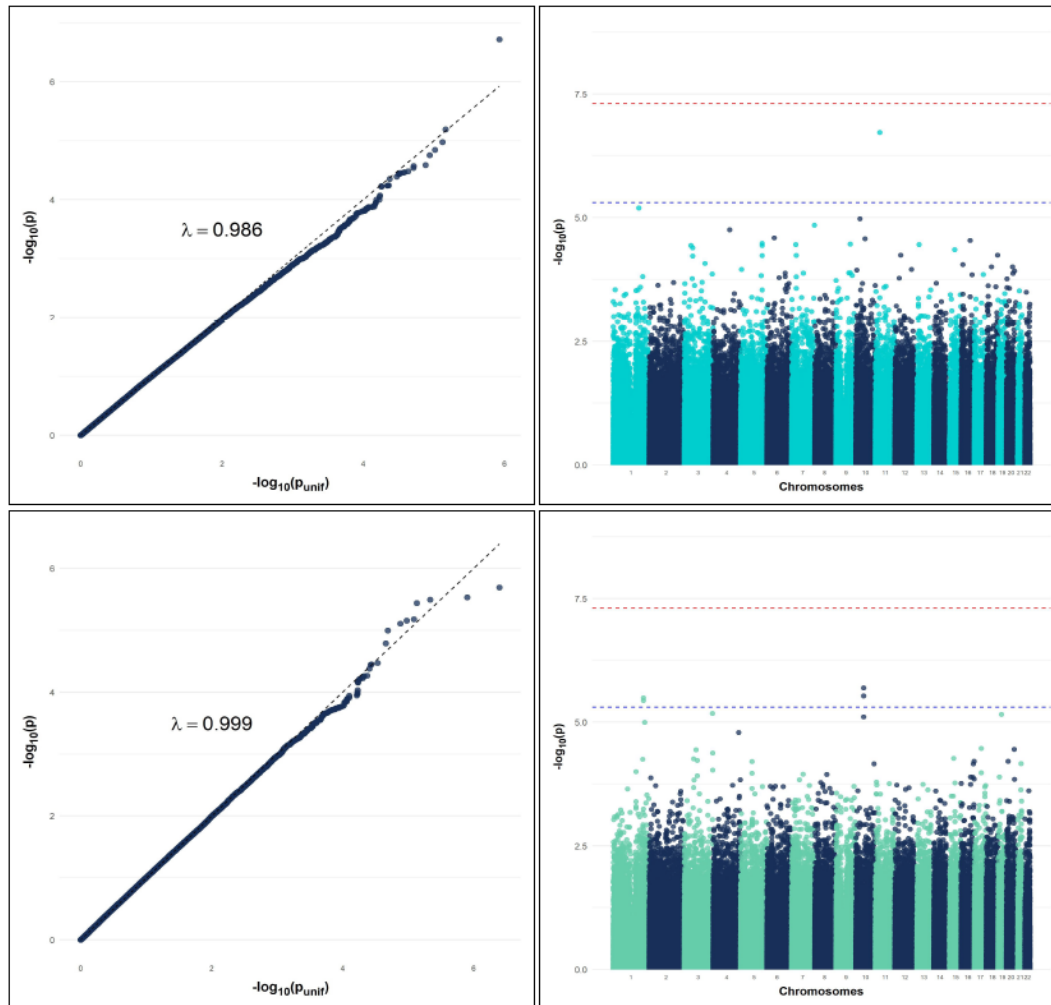


FIGURE 2.1 – Association entre variants génétiques et retour en dialyse au delà de 1 an post-transplantation pour les receveurs et les donneurs avant imputation. La figure montre les résultats GWAS avant imputation pour les receveurs (panel supérieur) et pour les donneurs (panel inférieur) représentés sous forme de QQplots (panels à gauche) et de Manhattan plots (panels à droite). La ligne rouge représente le seuil de significativité (fixé à  $5.10^{-8}$ ) et la ligne bleue, le seuil de suggestivité fixé ici arbitrairement à  $10^{-6}$ .

### 2.1.2 Association entre variants génétiques et retour en dialyse pour les receveurs et les donneurs après imputation

La figure 2.2 montre les résultats post-imputation avec Beagle qui a généré un total de 7 434 516 SNP. Les QQplots obtenus sont assez similaires à ceux avant imputation avec des coefficients qui sont comparables. Aucun des points ne dépasse le seuil de significativité pour les receveurs comme pour les donneurs. Toutefois, nous retrouvons les signaux préalablement observés avec une présence plus marquée illustrée par l'apparition de trainées. Les trainées représentent des SNP en déséquilibre de liaison, et donc corrélés, qui peuvent être associés au phénotype d'intérêt. De plus, de nouveaux signaux sont apparus avec l'imputation dans les chromosomes 1, 3 et 11 pour les receveurs et dans les chromosomes 1, 8 et 10 pour les donneurs. Ces derniers se situent néanmoins plus loin du seuil de significativité que pour les receveurs.

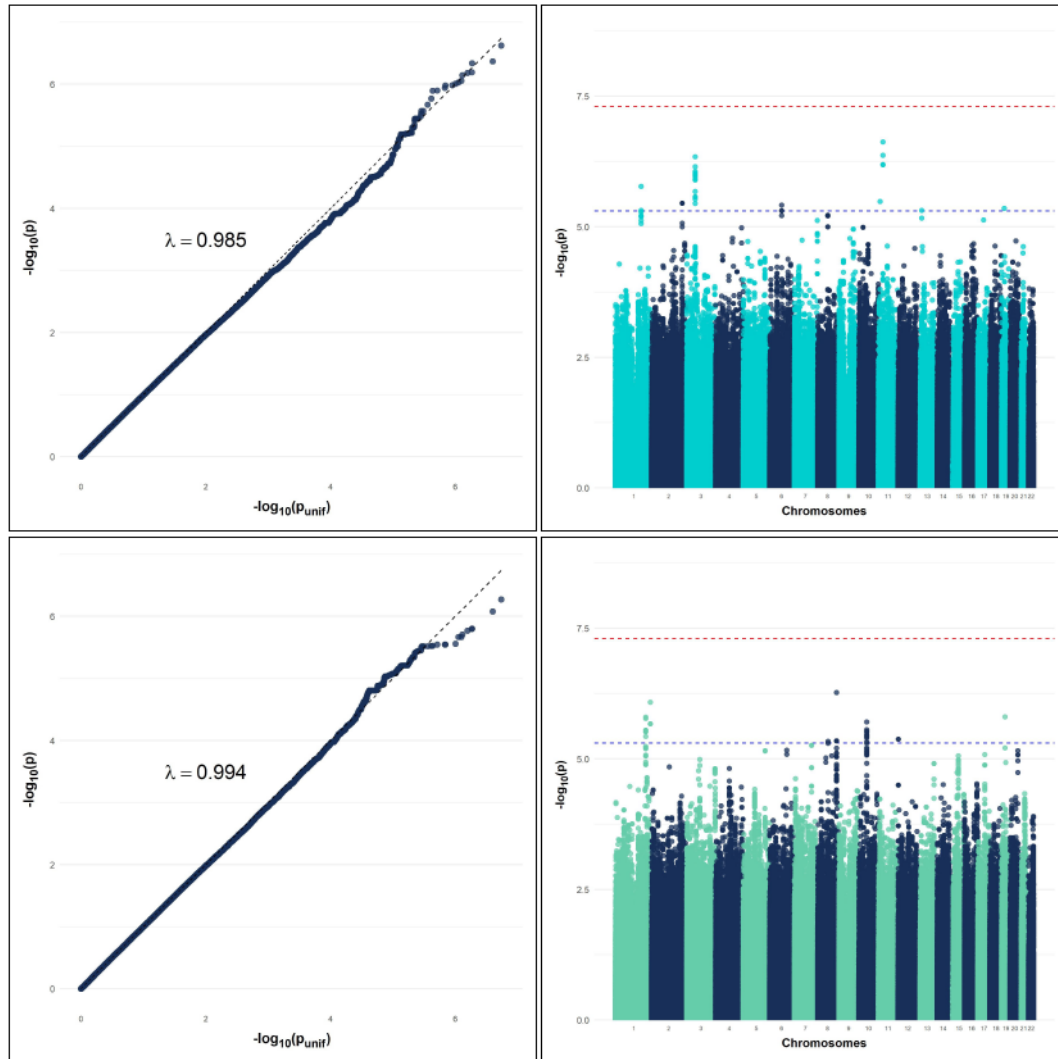


FIGURE 2.2 – *Association entre variants génétiques et retour en dialyse au delà de 1 an post-transplantation pour les receveurs et les donneurs après imputation.* La figure montre les résultats GWAS avant imputation pour les receveurs (panel supérieur) et pour les donneurs (panel inférieur) représentés sous forme de QQplots (panels à gauche) et de Manhattan plots (panels à droite). La ligne rouge représente le seuil de significativité (fixé à  $5.10^{-8}$ ) et la ligne bleue, le seuil de suggestivité fixé ici arbitrairement à  $10^{-6}$ .

### 2.1.3 Discussion

Les résultats présentés ci-dessus ne montre aucune association significative entre variants génétiques et le retour en dialyse à un an post-transplantation. Toutefois, un des signaux identifiés chez les receveurs à savoir le locus 11p14.1 ( $p=4,27 \times 10^{-7}$ , OR 1,79) a déjà été associé avec une pathologie rénale dans la littérature [254]. En effet, Pattaro et al. ont identifié en 2012 une association entre une diminution de l'eGFR et le SNP rs3925584 présent sur le locus que nous avons identifié. Ce SNP se trouve dans une région intergénique, entre les gènes *MPPED2* et *DCDC1*. L'étude montre qu'un knockdown (réduction de l'expression du gène) de *MPPED2* était associé à un développement défectueux du rein chez le zebrafish [254]. Ces résultats peuvent suggérer la possibilité d'une association entre cette région du génome et la survie du greffon rénal. Même si nos résultats ne sont pas significatifs, ils restent quand même très cohérents avec les études GWAS précédemment menées en transplantation rénale. En effet, elles restent sur la même lignée que les conclusions de Hernandez-Fuentes et al. [211] qui avec une cohorte conséquente (2094 paires D/R) ont tenté sans succès de répliquer les résultats de Ghisdal et al. [208] (cf. 2.2.3.2 Etudes d'association à l'échelle du génome (Genome Wide Association Studies, GWAS) p. 85). L'étude, comme la nôtre, n'a identifié aucune association significative ni chez les receveurs ni chez les donneurs pour l'échec de greffe. Il est tout de même important de noter que les études de Hernandez-Fuentes et al. et de Ghisdal et al. ne se sont pas basées sur les mêmes critères pour définir leurs phénotypes d'intérêt (rejet chronique, échec chronique, fonction rénale).

Ce manque de signaux statistiquement significatifs dans les résultats précédents et particulièrement dans nos analyses GWAS pourrait d'abord s'expliquer par un manque de puissance statistique. En effet, les approches GWAS, nécessitent une certaine taille de cohorte pour la détection statistique des signaux. Notre étude s'est concentrée sur les individus d'ancestralité européenne avec une dichotomisation entre receveurs et donneurs ce qui a réduit le nombre d'individus par rapport à la taille globale de KiT-GENIE ( $N = 4\,217$  individus). Par ailleurs, l'approche GWAS pourrait être inadaptée pour la taille de notre cohorte. Une augmentation de la population pourrait permettre d'améliorer la puissance statistique et par conséquent, de renforcer le signal que nous avons identifié. Mais cette hypothèse semble être réfutée par l'étude de Hernandez-Fuentes qui malgré la taille conséquente de leur cohorte n'ont pas trouvé d'associations significatives. Ensuite, un manque de spécificité en ce qui concerne le choix des phénotypes d'intérêt pourrait expliquer ces résultats. Dans le cas de notre étude, nous nous sommes intéressés au retour en dialyse à un an post-transplantation. Ce phénotype gagnerait à être affiné. Par exemple, la prise en compte du type de rejet (cellulaire ou humoral) et du nombre de rejet pourrait révéler des associations. La limite qu'on pourrait néanmoins rencontrer, surtout dans le cas de notre cohorte, serait un nombre restreint d'individus. En effet, les transplantations effectuées à Nantes présentent un taux de rejet assez faible (16.6% de rejets dans la cohorte

globale incluant 8.1% de rejets aigus). Une dichotomisation entre rejet chronique cellulaire et humoral affecterait la puissance de l'étude.

Finalement, peut être que ces résultats négatifs suggèrent simplement un besoin d'explorer de nouvelles approches qui impliquent la prise en compte des génomes du receveurs et du donneur en simultanée [215, 216, 217] (cf. 2.2.3.2 Etudes d'association à l'échelle du génome (Genome Wide Association Studies, GWAS) p. 85). Etant donné que la transplantation crée un contexte génétique particulier avec la coexistence de deux génomes au sein d'un même individu, analyser les receveurs et les donneurs séparément ne permet pas de tenir compte de toute la complexité des mécanismes impliqués dans la diminution de la fonction rénale. Il serait donc intéressant de développer des approches qui prennent en compte les génomes des receveurs et des donneurs simultanément. Enfin, ces résultats préliminaires ne sont, à l'heure actuelle, applicables que chez les européens, une analyse avec le reste des individus de KiT-GENIE est nécessaire.

QUATRIÈME PARTIE

# Discussion

---

# L'INTÉGRATION DES DONNÉES BIOMÉDICALES EN TRANSPLANTATION

---

## 1.1 Intérêt d'une ressource de données conséquente en transplantation rénale

En 2021, 5 patients sont en attente d'une transplantation rénale pour 1 greffon disponible [44]. L'un des enjeux majeurs de la transplantation rénale est de faire face à cette pénurie d'organes. Les moyens mis en place impliquent de trouver un équilibre entre une meilleure attribution des organes et une limitation de la perte du greffon une fois transplanté. Ce dernier point requiert entre autres choses une bonne compréhension des interactions entre donneurs et receveurs. Les études omiques ont permis de lever le voile sur certaines de ces questions comme illustré dans la partie 2 de l'introduction (cf. 2 La transplantation rénale à l'ère des «omiques» p. 51). En effet, la première étape durant ma thèse était, d'effectuer une revue non exhaustive de l'état de l'art actuelle des études omiques en transplantation notamment sur la génomique en transplantation rénale. Cette revue montre que dans le cas spécifique de la transplantation rénale, les études menées sont pour la majeure partie focalisées sur la protéomique, la génomique et la transcriptomique avec surtout des progrès sur la compréhension du rejet aigu [128]. Les cohortes étudiées restent très modestes et le rejet/perte chronique du greffon reste un défi. Notre problématique de base repose sur l'identification de facteurs non-HLA potentiellement impliqués dans les mécanismes de rejet/perte chronique du greffon. Cette identification passe par l'exploration simultanée des génomes du receveur et du donneur grâce à des approches GWAS. Des précédentes approches notamment celle de l'équipe de Mesnard et al. [215] avaient déjà mis en évidence l'existence de ces facteurs non HLA impliqués dans la fonction rénale. Mais cette étude n'était focalisée que sur les variants codants du génome. Plus important, le nombre de paires étudiées était très restreint (53 paires) ce qui constituait une limitation à la puissance statistique de l'étude. Il a été prouvé en génomique et en métabolomique que l'augmentation croissante du nombre d'échantillon dans une étude peut améliorer la robustesse et la précision du diagnostic des marqueurs [255]. Ces études requièrent donc une collection de données importante. Particulièrement dans ce travail, la biobanque KiT-GENIE constitue

une source de données conséquente de donneur-receveur, consolidée par une base de données cliniques, biologiques et génomiques.

### 1.1.1 Augmentation de la puissance des études

La mise en place de KiT-GENIE avait pour ambition de constituer l'une des plus grandes biobanques homogène de paires donneur-receveur en lien avec la transplantation rénale. L'intérêt d'avoir une grande cohorte permettait d'abord de limiter les problèmes liés à la puissance statistique. En effet, les études d'associations notamment les GWAS, se caractérisent par le test simultané de millions de variants (par exemple les SNP) pour identifier des associations potentielles avec un phénotype d'intérêt [256]. Pour mieux contrôler ces tests multiples et limiter les faux positifs, on applique la correction de Bonferroni [253] qui établit un seuil de significativité  $P = 5 \times 10^{-8}$  pour les SNP. Cette correction permet de maintenir un taux de faux positifs égal à 5% à l'échelle du génome [256] mais réduit considérablement la puissance de détection des SNP associés, car la plupart des signaux d'association n'atteignent pas le seuil requis [195]. Une façon de résoudre ce problème est d'augmenter la taille de la cohorte d'étude. KiT-GENIE donne accès à une vaste collection de données contenant des informations génétiques et cliniques substantielles pour les donneurs et les receveurs. Parallèlement, KiT-GENIE contribue à l'effort mondial de méta-analyse pour définir les facteurs génétiques associés au rejet et à l'échec de la greffe en rejoignant le consortium international iGeneTRAiN (International Genetics and Translational Research in Transplantation Network, <http://igenetrain.org/>) [257] en début 2022. Le consortium iGeneTRAiN intègre plusieurs cohortes de différents types de receveurs d'allogreffes afin d'identifier les facteurs génétiques associés au succès des transplantations rénales. Il s'agit de la plus grande initiative génétique visant à mieux comprendre les rejets et les complications des greffes à ce jour [257, 258]. Ce type d'effort facilite la mise à disposition, pour le monde scientifique, de ressources indispensables à une meilleure compréhension des mécanismes de la transplantation. Cela permet de mener des études plus robustes avec une puissance statistique conséquente et des données diverses mais également d'avoir l'opportunité de valider ces études avec d'autres cohortes.

### 1.1.2 Accès à des données diverses

#### 1.1.2.1 Importance de l'intégration de la double couche de données génomiques et phénotypique

KiT-GENIE s'appuie sur les données de DIVAT contenant des données biologiques, cliniques et épidémiologiques relatives à plus de 30 000 transplantations à ce jour. Ces données sont soigneusement collectées tout au long du suivi médical de chaque patient transplanté. La base de données compte plus de 200 variables phénotypiques. Pour DIVAT-Nantes, KiT-GENIE enrichit



cette base, en rajoutant une troisième couche de données génomiques qui va permettre d'augmenter la diversité des données et par la même occasion accroître leur potentiel. Ces dernières années, des ressources qui combinent les données issues des dossiers médicaux électroniques (EMR) aux données génotypiques sont développées. A titre d'exemple, nous pouvons citer la UK Biobank qui contient des informations génétiques et cliniques détaillées sur un demi-million de patients [259]. Une étude menée sur 423 604 participants de la UK-biobank a permis la mise au point d'un modèle d'apprentissage statistique automatique qui prédit le risque de maladie cardiovasculaire [260]. L'intégration de données cliniques et génomiques au sein d'une même cohorte permet d'approfondir l'exploration des associations génotype-phénotype notamment dans l'étude de maladies complexes. Ces dernières résultent souvent de l'interaction entre la génétique et l'environnement [261]. Par exemple dans le cas du VIH ou encore de l'obésité, des facteurs environnementaux peuvent altérer la méthylation de l'ADN et affecter des gènes potentiellement impliqués dans ces maladies [262, 263]. Les études GWAS peuvent bénéficier de l'information clinique afin d'explorer plus amplement certains phénotypes notamment ceux ayant trait aux facteurs environnementaux. Par exemple, dans le cas du VIH, l'intégration les facteurs tels que le mode d'infection, l'évolution du virus, le tabagisme etc. en plus de l'information génétique a été révélateur pour les études et a conduit à l'identification de facteurs de risques associés au VIH [264, 265]. En transplantation rénale, l'accès à des données sur les traitements immunologiques, les biopsies et les informations sur le devenir du greffon en simultanée permet d'avoir une vision plus globale des mécanismes ayant conduit à un échec ou un rejet de greffe [266]. Dans le cas de notre étude, la richesse des données permet d'une part une prise en compte des covariables pouvant introduire un biais dans l'analyse tels que l'âge, le sexe, le BMI qui sont des facteurs de risques connus du rejet de greffe [267, 268, 269]. D'autre part, la disponibilité des informations telles que la durée de la greffe, le nombre de rejets, les antécédents de maladies, les informations sur le donneur, les traitements immunosuppresseurs etc. contribuent certes à enrichir les résultats GWAS mais surtout à pouvoir explorer plusieurs autres phénotypes d'intérêt liés à la transplantation rénale dans la même cohorte d'étude (récidive de la maladie, fonction rénale du donneur).

### 1.1.2.2 Accès à des données longitudinales

L'intégration des données cliniques notamment l'inclusion des registres patients donne également accès à des données de suivi. En transplantation rénale, Prezelin-Reydit et al. [270] ont montré, grâce à l'intégration de données temporelles de suivi, que la greffe préemptive rénale était associée à un risque plus faible d'échec de la transplantation (défini comme le décès, le retour en dialyse ou la retransplantation.) qu'importe la durée de la dialyse.

KiT-GENIE, grâce à l'intégration des données cliniques de DIVAT, comporte pour chaque patient des données de suivi pouvant aller jusqu'à 25 ans (moyenne de suivi = 6,7 ans). En 2010,

une étude faite sur la base des données de DIVAT, a montré l'intérêt d'avoir des données de suivi en développant un score de prédiction de la survie du greffon à long terme [271]. Ce score nommé KTFS (Kidney Transplantation Failure Score) prends en compte 8 facteurs de risque d'échec de greffe connus et précédemment décrits dans la littérature (âge, sexe, nombre de transplantations, créatininémie (3mois et 12 mois), protéinurie (12 mois), créatininémie du donneur, rejet aigu) et évalue, chez les patients greffés dans les 1 an, un score de prédiction de la survie du greffon à 8 ans post-transplantation [271]. Plus récemment, cette même équipe a développé un modèle dynamique de prédiction du risque d'échec du greffon pour un patient. Ce modèle inclut en plus des caractéristiques de base précédemment citées, des mesures répétées de la créatinine sérique (SCr) qui pourraient être mises à jour en continu pendant le suivi du patient. La disponibilité des données longitudinales ouvre donc les portes à une médecine prédictive et personnalisée pour chaque patient. Dans le cadre de KiT-GENIE, l'enrichissement en données génomiques, en plus des données de suivi permettrait de pousser encore plus loin ces analyses et de développer un score qui prendra en compte les facteurs de risques cliniques connus mais également des facteurs génétiques pour la prédiction de la survie ou du rejet de greffe.

### 1.1.2.3 Diversité populationnelle/ethnique des données

La diversité des données issues de ressources massives implique également une diversité populationnelle dans certains cas. Bien que des projets nationaux tels que UK biobank [259] ou OPTN (Organ Procurement and Transplantation Network) [272] concentrent principalement des individus d'ascendance européenne, ces ressources permettent néanmoins d'avoir accès à des données populationnelles issues d'individus de différentes ethnies. Par exemple, l'OPTN compte 346 908 (38%) [273] de greffes sur des patients qui n'ont pas une ancestralité européenne. De plus, le projet 1KG [274] fournit un catalogue des variations génétiques communes humaines dans 5 populations de référence (Africains, Américains, Sud et Est Asiatiques, Européens). La disponibilité de ces données permet de pouvoir explorer la diversité génétique des populations de différentes ancestralités, leur l'évolution et d'évaluer les différentes susceptibilités à certaines maladies ou traits. Par exemple, plusieurs études ont déterminé une plus grande susceptibilité à la maladie rénale chez les populations d'ancestralité africaine. En effet, des variants décrits dans le gène *APOL1* sont associés à un risque accru de maladies rénales [275, 276, 277, 278, 279]. Ce risque est considérablement accentué pour les suds africains [280] et les afro-américains ayant des néphropathies associées au VIH [281] (Odd ratio de 89 et 29 respectivement). Ces variants sont exclusifs des personnes d'ascendance africaine et ont été naturellement sélectionné chez ces populations car ils confèrent une résistance à la trypanosomiase humaine africaine ("maladie du sommeil africaine"), en plus d'autres maladies infectieuses [282]. La découverte de ce facteur de risque génétique illustre parfaitement l'importance de l'ancestralité génétique dans le risque de maladie. Cela a notamment permis des avancées significatives dans la compréhension de la

pathophysiologie des maladies rénales [283, 284].

L'accès à des données diverses permet également de pouvoir valider les études génomiques et confirmer que les résultats peuvent être généralisés à d'autres populations. Bien souvent, les résultats des études génomiques se basent sur des données qui sont majoritairement ou entièrement issues d'individus avec une ancestralité européenne. Ces résultats, appliqués à une population non européenne peuvent différer ou être complètement biaisés. En 2015, une étude a montré qu'un taux faible de tacrolimus, qui est l'immunosuppresseur de prédilection en transplantation rénale [285], augmente considérablement le risque de rejet aigu chez les Afro-Américains (AA). Les AA ont 70% de chance de ne pas atteindre les concentrations minimales thérapeutiques moyennes de tacrolimus au cours de la première année suivant la transplantation comparé aux autres populations (non AA). Mais contrairement aux non-AA, les AA qui atteignent des concentrations thérapeutiques de tacrolimus présentent des taux de rejet aigu nettement inférieurs, ce qui n'a pas été démontré chez les populations non AA [286]. Cette étude met en évidence les différences ethniques dans la relation concentration-effet du tacrolimus sur les mécanismes de rejets et suggère une meilleure adaptation des posologies de tacrolimus en fonction des populations. En clinique, un bon exemple réside dans la mise en place et la validation des formules de mesures du DFG (formules MDRD et CKD-EPI). Ces formules ont été mises au point en prenant en compte un facteur ethnique pour les populations noires. Néanmoins, ce facteur ethnique a été établi sur des données de sujets AA. Ce qui remet en doute sa validité sur des sujets noirs qui ne sont pas de ce groupe (africains par exemple). Une étude de 2016 menée en Côte d'Ivoire [287] a montré que la formule MDRD surestime le DFG lorsqu'on prend en compte les facteurs ethniques. De plus, la performance des équations MDRD et CKD-EPI est significativement meilleure, en population générale, sans le «facteur AA» habituellement proposé. Ils ont conclu que le «facteur AA» était inapplicable en Afrique de l'ouest [287].

Ces exemples montrent l'intérêt d'avoir accès à une cohorte diverse pour les études. Les données KiT-GENIE sont pour l'instant majoritairement européennes (plus de 90% pour les donneurs et les receveurs) avec l'ambition d'intégrer des données issues d'autres centres afin d'augmenter la diversité génétique de la population d'étude et de pouvoir effectuer des méta-analyses à l'échelle de sous-populations ancestrales.

## 1.2 Les défis de l'intégration des données

L'intégration des données en transplantation passe par des collaborations multicentriques qui permettent de pouvoir centraliser des données provenant de plusieurs régions/pays. C'est notamment le cas pour DIVAT qui compte 8 centres à l'échelle de la France. Un problème inhérent à cette intégration de données multicentriques réside dans la diversité des définitions et des pratiques entre les centres. En effet, il existe très souvent des variabilités inter-centres dans les

protocoles cliniques (traitements immunosuppresseurs par exemple), la sélection des donneurs, les paramètres mesurés lors du suivi des patients, l'extraction et la collecte des échantillons etc. [288]. De plus, les critères utilisés pour définir les phénotypes tels que les rejets aigus ou les dysfonctionnements chroniques du greffon peuvent varier selon les centres de transplantation [289]. En transplantation, le gold-standard pour décrire et évaluer les mécanismes de rejets aigus et chroniques est l'histologie. Mais les mesures basées sur l'histologie peuvent être subjectives comme le montre la mise à jour continue de la classification de Banff pour la rejet aigu au fil du temps [289]. Cette disparité peut impacter les recherches et les analyses qui sont faites sur la base de ces données. En premier lieu, cela rend l'intégration des données plus fastidieuse. En effet, le stockage des données de recherche dans des structures informatiques adaptées demande une certaine homogénéité dans les données pour faciliter leur requêtage et leur analyse. Cette homogénéité implique par exemple des protocoles cliniques similaires afin de pouvoir comparer les mêmes variables d'intérêt chez tous les individus. Ensuite, cette disparité peut impacter le choix des critères d'inclusion pour les études. Par exemple, des définitions claires des mécanismes tels que les rejets aigus ou chroniques permettraient de pouvoir mieux choisir les critères d'inclusion pour une étude dont le phénotype d'intérêt porte sur la survie par exemple. Enfin, cette disparité peut affecter les résultats des analyses car elle peut impliquer des données qui ne sont pas distribuées de façon normale ; cela peut être le cas lorsque, en fonction des centres, les unités de mesures sont différentes pour une variable donnée (créatininémie, protéinurie etc.). Tous ces éléments, s'ils ne sont pas bien contrôlés introduisent une hétérogénéité supplémentaire entre les études et peuvent rendre les études d'associations génétiques difficiles à mener et à valider. Cela peut expliquer, outre le manque de puissance des analyses statistiques, le manque de reproductibilité des études sur les transplantations. D'un point de vue analytique, le traitement des données génomiques nécessite une rigueur absolue dans les protocoles et ceci tout au long du processus d'analyse. Par exemple, la mise en place de KiT-GENIE a nécessité des contrôles qualités avant pendant et après le génotypage avec des protocoles référencés et clairs. Les données brutes générées ont été rigoureusement traitées grâce à AxAs pour avoir des données robustes et de haute qualité pour les analyses GWAS. Ces étapes sont encore plus importantes lorsqu'il s'agit de données provenant de différents centres. Toutefois, des projets comme iGENETRAiN [221] dont fait maintenant partie KiT-GENIE font des efforts dans ce sens. Le consortium a homogénéisé les méthodes de génotypage et utilisé des définitions communes pour le rejet afin de réaliser des méta-analyses robustes pour plusieurs types de transplantations.

# ANALYSE DE DONNÉES GÉNOMIQUES À GRANDE ÉCHELLE

---

L'intégration d'une grande quantité de données implique d'être très prudent lorsqu'il s'agit de processus analytique. Ce dernier suit généralement un schéma qui s'articule sur trois axes : la problématique, la préanalyse et l'analyse des données. Le point de départ est la question biologique. Avant même de considérer les données, il est primordial de s'interroger sur ce que l'on cherche à savoir. Bien souvent, on se lance dans l'analyse des données dans toute leur complexité en négligeant l'étape critique de la mise en contexte des données. Cette étape n'est toutefois pas simple lorsque l'on met en place une cohorte telle que KiT-GENIE qui a été établie de sorte à ne pas être spécifique à une seule question biologique. Néanmoins, l'axe de réflexion restait le même et portait sur les facteurs moléculaires qui régissent les mécanismes de la transplantation rénale. Cette question bien que plus large, représentait un point de départ solide pour nos analyses et nous a permis d'être plus exhaustifs dans le choix des critères d'inclusions afin d'anticiper l'étude de plusieurs phénotypes d'intérêt. Ensuite viennent les étapes de préanalyses et d'analyses des données qui représentent des phases critiques en traitement de données génomiques.

## 2.1 Importance de la pré-analyse des données

La pré-analyse permet de s'assurer que les protocoles suivis tout au long du processus de collecte et de génotypage n'ont pas induit de biais ou d'erreurs systématiques. Les bonnes pratiques d'analyse commencent à l'état expérimental c'est à dire que la manière dont les échantillons ont été collectés et manipulés a un impact sur la qualité des données produites. Pour le projet KiT-GENIE, plusieurs étapes ont été nécessaires à la pré-analyse. D'abord, tous les échantillons d'ADN collectés ont été mesurés et quantifiés afin d'établir un premier bilan de la qualité des échantillons. Ensuite pour la préparation au génotypage, les ADNs ont été soigneusement normalisés à un volume et une concentration spécifique pour éviter les erreurs de mesures lors du génotypage qui peuvent résulter de quantités d'ADN différentiels entre les échantillons. De plus, nous nous sommes assurés de bien randomiser nos échantillons en fonction du sexe, du type d'individus (D ou R), et du type de donneurs afin d'éviter les effets de lots c'est à dire des individus de même sexe ou de même type qui se retrouvent dans le même lot et peuvent

donc biaiser l'analyse GWAS. Les données KiT-GENIE ont également été génotypées avec la même technologie à savoir la puce PMRA en respectant le protocole préconisé par Affymetrix. Les données brutes générées ont subi un contrôle qualité rigoureux à trois échelles : un QC des individus, un QC des plaques et un QC des variants. La conformité des données a été vérifiée en faisant une comparaison du sexe et du HLA des individus génotypés versus les individus présents dans la base de données.

Toutes ces étapes nous ont permis d'assurer des données robustes et qualitatives pour les analyses GWAS. Toutefois, nous avons eu recours à deux plateformes de génotypage pour l'ensemble des données de KiT-GENIE. Malgré une technologie et des protocoles similaires, les techniciens n'étaient pas les mêmes et surtout les analyses n'ont pas été faites sur un intervalle de temps continu. Ce délai est en partie dû au Covid 2019. Cette disparité pourrait avoir introduit du bruit dans nos données. L'analyse en composante principale permettait de s'assurer que les données n'avaient pas d'effets de lots ni de valeurs aberrantes et sinon pouvoir les corriger.

## 2.2 Les études GWAS dans la compréhension des mécanismes de rejet/perte chronique du greffon

### 2.2.1 Limite des études GWAS aux receveurs et aux donneurs uniquement

Les études GWAS ont permis de mieux appréhender les mécanismes qui sous tendent la fonction rénale chez l'homme comme illustrés par les nombreuses études sur *APOL1* [275, 277] ou encore sur l'effet des taux de tacrolimus dans le rejet du greffon [290, 286]. Néanmoins, les études d'associations sur les mécanismes de rejets aigus ou la survie du greffon à long terme sont arrivées un peu plus tard, en 2013 pour les mécanismes chroniques [213] et en 2016 pour le rejet aigu [208]. D'autres études ont très vite suivies [214, 211] notamment celle de Hernandez-Fuentes et al. [211] qui disposaient d'une cohorte conséquente de 2094 paires donneurs-receveurs et qui ont effectué 3 types de GWAS (receveurs seuls, donneurs seuls et D/R) pour évaluer la survie du greffon. Toutefois aucun résultat n'a été validé pour toutes ces études. Les associations décrites par les études de Ghisdal et al. sur le rejet aigu ou encore O'Brien et al. sur la fonction du greffon à long terme n'ont pas pu être répliquées par aucune autre étude malgré une augmentation de la cohorte d'étude pour les travaux qui ont suivis. Il est important de noter que pour toutes ces études, à l'exception de celle de Hernandez-Fuentes, les GWAS n'ont été menées que sur les receveurs seuls. Nous avons également mené une étude GWAS sur des receveurs seuls ( $n = 1291$ ) et des donneurs seuls ( $n = 1183$ ) qui n'ont donné aucune association significative. Ces résultats étaient quelque peu attendus car ils restent cohérents avec les précédentes études de réplifications qui n'avaient identifié aucune association. Néanmoins, chez les receveurs, un signal approchant le seuil de significativité a cependant été identifié au niveau du locus 11p14.1 ( $p=4,27 \times 10^{-7}$ , OR 1,79), qui a précédemment été associé avec la fonction rénale.

Nous pourrions penser qu'une augmentation de la puissance statistique permettrait de renforcer ce signal mais l'étude de Hernandez-Fuentez semble dire le contraire car ils ont mené leur étude sur un plus grand nombre de receveurs et n'ont pas trouvé d'associations. Toutefois, l'argument selon laquelle leur étude n'est pas une réplique exacte de la notre pourrait être avancé dans ce cas là et constituerait un bon argument pour explorer ce signal avec plus d'individus. C'est d'ailleurs, un des éléments de réponse donnée par l'équipe de Ghisdal et al. [291] afin d'expliquer cette disparité de résultats entre les deux études.

### 2.2.2 Intérêt de l'analyse double génome

À ce jour, il n'y a eu qu'une seule étude GWAS combinant l'analyse de donneurs et de receveurs de rein réalisée par Hernandez-Fuentes et al. [211]. Aucun signal statistiquement significatif en dehors du HLA n'a été trouvé dans aucun des modèles testés. Cette étude était limitée par le manque d'homogénéité dans la définition des phénotypes testés. Ces derniers ont été évalués rétrospectivement à partir des données des registres provenant de différents centres [292]. Plus récemment, les études de Reindl-Schwaighofer et al. en 2018 [216] et de Steers et al. en 2019 [217] explorent des nouvelles méthodes basées respectivement sur le rôle des mésappariements à l'échelle du génome dans les variants faux-sens (substitutions non synonymes) et l'hypothèse de la "collision génomique" selon laquelle le risque de rejet peut être accru chez les receveurs, homozygotes pour des variants de perte de fonction, ayant reçu un greffon provenant de donneurs non homozygotes [292]. Ces deux études récentes ont confirmé que les facteurs non-HLA sont impliqués dans les résultats chroniques de la transplantation rénale. En élargissant l'utilisation des scores de risque polygénique en transplantation rénale, elles font écho à celle de Mesnard et al. [215] qui ont introduit pour la première fois un système de score de sommation pour évaluer les facteurs non-HLA dans la fonction du greffon rénal à partir de données génomiques (séquençage de l'exome de nouvelle génération). Les études GWAS de paires donneur-receveur représentent une approche non biaisée qui pourrait être utilisée pour identifier des associations en testant formellement divers types d'interactions donneur-receveur à l'échelle du génome. Ces études peuvent bénéficier de l'intégration de couches de données supplémentaires (expression génétique, protéines, métabolites) pour mieux comprendre les résultats de la transplantation. Étant donné la complexité et l'hétérogénéité des résultats du rejet, de grandes cohortes avec des données cliniques complètes, l'ADN du donneur et du receveur, et du sérum prélevé de préférence au moment du rejet sont nécessaires pour le succès de telles études. C'est l'effort que nous avons mis en place avec KiT-GENIE afin d'avoir une ressource de données complète, homogène avec des informations cliniques exhaustives et une couche de données génomiques qui vient enrichir la base de données. Nos résultats actuels sont limités à l'exploration des donneurs et des receveurs seuls. Nous comptons aller plus loin avec la réalisation d'une analyse simultanée des génomes

du donneur et du receveur qui pourrait identifier de potentiels facteurs de risque pour la perte du greffon.

### 2.2.3 Prise en compte de l'ancestralité

L'analyse en composante principale nous a permis d'identifier des potentiels valeurs aberrantes dans notre cohorte d'études mais également de voir comment nos individus étaient distribués en terme d'ancestralité. Cela nous a permis de voir que les individus transplantés à Nantes étaient majoritairement européens même s'il subsiste une certaine diversité (plus de 7% des receveurs sont non européens). Cette diversité bien que intéressante peut poser problème dans les analyses. Il a été montré à plusieurs reprises que certains loci de risque à des maladies présentent des différences ethniques considérables en termes de fréquence et/ou d'ampleur de l'effet d'un variant. La réalisation d'études GWAS dans divers groupes ethniques a d'ailleurs révélé une hétérogénéité dans la susceptibilité génétique à la maladie [279].

Par ailleurs, les analyses seraient plus robustes avec une dichotomisation propre des individus en fonction de leur ancestralité. Le problème qui se pose reste la puissance. Les individus non européens sont trop peu nombreux pour être séparés en sous groupes et trop nombreux pour être complètement éliminés des analyses. Une première solution est de les enlever de l'analyse principale et de les utiliser comme cohorte de validation. Mais encore une fois, cette solution diminue la puissance statistique de nos analyses. Nous pouvons prendre en exemple notre analyse axée que sur les européens ( $n=2\ 474$  versus  $4\ 417$  pour toute la cohorte). Une deuxième solution est de faire une analyse globale de tous les individus en prenant bien soin d'impliquer l'ancestralité comme covariable. Mais cette solution peut être insuffisante pour limiter le bruit dans nos analyses c'est à dire que cette disparité en terme d'ancestralité peut intensifier ou affaiblir un signal génétique et impliquer des résultats qui ne sont pas fiables. La meilleure alternative reste l'intégration de données plus diverses pour pouvoir mener des méta-analyses robustes avec des résultats qui peuvent être généralisés dans plusieurs populations. Cette intégration passe par des collaborations entre chercheurs, la mise en place de ressources de données massives et diversifiées, l'accès facilité à des données provenant d'autres études afin de contribuer à l'effort mondial de compréhension des mécanismes de la transplantation.

### 2.2.4 Validation des études

La validation des études en transplantation et plus largement dans le domaine biomédical revêt une importance primordiale pour s'assurer de la crédibilité des résultats. En transplantation rénale, la validation des résultats reste un grand challenge. En effet, plusieurs études qui initialement semblaient fournir des résultats prometteurs avec l'identification de variants associées à la survie ou au rejet de greffe n'ont pas pu être répliqués car les résultats se sont avérés par la suite être des faux positifs. Cela semble être le cas de l'étude de O'Brien et al. [213] dans le



rapport de Philstrøm et al. [214] où deux SNP associés à la créatinine à 5 ans et à la survie du greffon chez les receveurs de rein n'ont pas pu être validés dans une cohorte indépendante. De la même manière, l'étude de Ghisdal et al. en 2016 [208] sur le rejet aigu n'a pas pu être validée même avec une cohorte conséquente et en utilisant plusieurs modèles statistiques. Néanmoins, Ghisdal et al. ont attribué ces résultats négatifs à une disparité de méthodes entre les deux études ce qui les rend donc incomparables selon eux. Dans tous les cas, aucune autre étude n'a pu confirmer et valider les associations précédemment trouvés en GWAS autant sur les rejets aigus que sur la survie du greffon à long terme. Ceci inclut nos GWAS sur les donneurs et les receveurs séparés. Un autre point important à souligner est le côté très eurocentré des cohortes pour toutes ces études. Par conséquent, les résultats nécessitent également une validation sur d'autres populations pour que les résultats puissent être appliqués à la population générale.

CINQUIÈME PARTIE

# Conclusion et perspectives

---

# CONCLUSION

---

L'objectif scientifique qui a motivé mon projet de thèse était l'identification de facteurs génétiques au delà du HLA, pouvant potentiellement jouer un rôle dans les mécanismes qui régissent le rejet ou la perte de greffon à long terme. Dans le but de répondre à ce questionnement, nous avons préconisé une approche GWAS sur des données de paires donneurs-receveurs (D/R) afin de :

- valider/répliquer les études GWAS qui ont déjà été menées sur les receveurs et les donneurs séparément
- établir un alloscore polygénique de risque qui estime le nombre d'incompatibilités allogéniques entre donneurs et receveurs dans le sillage de l'étude menée par Mesnard et al. en exome [215].

Etant donné la taille modeste des cohortes D/R précédemment étudiées, l'originalité de notre approche repose sur l'accès aux données de la cohorte française DIVAT de transplantés rénaux (>30000 patients transplantés rénaux depuis 1990), le but étant d'établir une ressource de données homogènes et conséquentes de D/R.

A partir d'une extraction de DIVAT-Nantes, nous avons sélectionné, collecté, génotypé, nettoyé et analysé les données de près de 2000 paires D/R pour consolider la biobanque KiT-GENIE qui compte à ce jour une triple couche de données cliniques, biologiques et génétiques pour chaque paire donneur-receveur. Ce travail s'enrichit d'une part de la qualité et l'homogénéité des données et d'autre part d'une description fine des individus et des informations cliniques. En effet, j'ai établi un contrôle qualité des données rigoureux tout au long des analyses. Ce contrôle qualité implique, les bonnes pratiques de laboratoires lors des phases de génotypage, une analyse primaire des données brutes rigoureuse selon les recommandations d'Affymetrix, le contrôle des erreurs systématiques tels que les effets de lots et une comparaison des données de la base avec les données génétiques pour le sexe et le HLA pour assurer une conformité parfaite des données. KiT-GENIE dispose de données robustes et fiables pour l'analyse. J'ai également établi la "carte d'identité" de la cohorte KiT-GENIE en fournissant une description démographique claire de la population d'étude et en mettant en exergue la richesse des phénotypes présents dans la base de données. Au delà des aspects descriptifs, mes travaux ont portés sur l'analyse des données génomiques de la cohorte. J'ai d'abord mené une exploration de l'ancestralité des individus KiT-GENIE qui a soulevé des questionnements intéressantes sur l'intérêt de la prise en compte

---

de l'ethnicité en transplantation. Enfin, sur la base des travaux de stage d'un de nos étudiants de Master, nous avons réalisé une analyse génomique sur 7,4M de polymorphismes (SNP) pour rechercher les associations avec le retour en dialyse au-delà d'un an post-transplantation chez les receveurs et les donneurs d'origine européenne pour des transplantations rénales réalisées à Nantes après 2000. L'analyse a été réalisée séparément pour les donneurs (201 cas vs. 982 contrôles) et pour les receveurs (238 cas vs. 1053 contrôles). Aucune association statistiquement significative n'a été détectée chez les receveurs comme chez les donneurs. Cependant, l'analyse des receveurs détecte un signal approchant le seuil de significativité au niveau du locus 11p14.1 ( $p=4,27 \times 10^{-7}$ , OR 1,79), qui a précédemment été associé avec la fonction rénale.

Les objectifs du projet KiT-GENIE n'ont pas été entièrement atteints à l'issue de ma thèse. Néanmoins, mes travaux ont apporté plusieurs réponses. D'abord, la conception et la mise en place de la biocollection a représenté un travail colossal mais extrêmement important pour le futur de ce projet. En effet, le montage de KiT-GENIE a nécessité un investissement important en temps mais également en ressources financières. Ceci est assez représentatif des projets génomiques de cette envergure surtout lorsque la conception de la biobanque n'a pas été pensée pour un seul projet mais pour répondre à plusieurs problématiques relatives à un domaine. Par ailleurs, la mise en place de la ressource doit alors être le plus qualitatif possible. J'ai débuté ma thèse lorsque le projet KiT-GENIE était en phase de conception. J'ai participé à toute la procédure de montage : de l'extraction des données informatiques à la collecte des échantillons d'ADN en passant par le processus de génotypage. KiT-GENIE représente aujourd'hui une ressource majeure en transplantation de par sa richesse et la qualité des données disponibles. La biobanque a été façonnée de sorte de pouvoir tacler des questionnements primordiaux en transplantation ; sa triple couche de données permet aujourd'hui de pouvoir étudier les interactions entre donneurs et receveurs, d'explorer plus amplement certains phénotypes relatifs aux receveurs (maladies récidivantes, réponses aux immunosuppresseurs, analyses longitudinales etc.) et d'améliorer la compréhension sur le devenir de la fonction rénale des donneurs particulièrement pour les dons du vivant. Mon travail a permis de préparer le terrain et de permettre la genèse de nouveaux projets détaillés dans la partie perspectives. En deuxième lieu, les analyses préliminaires des données ont permis d'ouvrir des questionnements sur la diversité génétique réelle des patients de DIVAT-Nantes et l'utilité d'impliquer l'éthnie dans les transplantations. Par rapport à notre analyse, cela permet un meilleur contrôle des biais qui peuvent être liés à l'ancestralité des individus. Enfin, les résultats GWAS préliminaires éclairent sur deux choses. D'abord, elles valident le manque de signaux significatifs dans les études menées sur les receveurs ou les donneurs séparément ce qui n'est pas contradictoire avec la littérature actuelle. Ensuite elles confirment la nécessité de l'analyse simultanée des génomes du donneur et du receveur qui pourrait être déterminant dans l'identification de potentiels facteurs de risque pour la perte du greffon.

# PERSPECTIVES

---

La cohorte KiT-GENIE représente une ressource précieuse pour l'étude des complications des transplantations rénales, ainsi que les conséquences chez les donneurs vivants de rein. Plusieurs études GWAS liées aux résultats des transplantations rénales sont actuellement en cours au sein de l'équipe. Ces projets seront enrichis par la biobanque KiT-GENIE.

1. Dans la suite directe de mon projet de thèse, une étude est actuellement en cours dans l'équipe pour explorer plus amplement les interactions-donneurs receveurs et tenter d'identifier des facteurs génétiques non-HLA pouvant impacter la survie ou le rejet du greffon. Nous avons au cours de ma thèse, réalisé deux GWAS, séparément chez les donneurs et les receveurs, qui n'avaient pas générés de signaux significatifs pour les deux analyses. Plus récemment, une analyse comparative des mismatches génétiques entre receveurs et donneurs a été menée avec une séparation des paires ayant un donneur vivant versus un donneur décédé. Ces résultats préliminaires font l'objet d'explorations et de validations supplémentaires. Les étapes suivantes seront d'effectuer une exploration longitudinales de la survie du greffon grâce aux données de suivi en utilisant des modèles de régression multivariés de Cox. Enfin, l'étape finale portera sur la mise en place de scores de risque polygénique afin de mieux saisir le risque génétique global pour une transplantation rénale à partir des SNP du donneur, des SNP du receveur et des mésappariements donneur-receveur. Une nouvelle thèse est actuellement en cours pour prendre en charge ces analyses et fournir des résultats prometteurs dans le domaine de la transplantation.
2. Deux autres projets sont également en cours et concernent respectivement la récurrence HSF et l'étude de la fonction rénale post-don des donneurs vivants.
  - Le syndrome néphrotique idiopathique avec lésions HSF présente un risque élevé de récurrence après une transplantation rénale sans qu'aucun facteur prédictif de stratification du risque n'ait été identifié à ce jour. Nous avons exploré les associations génétiques avec la HSF et sa récurrence après transplantation dans la sous-population européenne KiT-GENIE afin de mieux caractériser les voies moléculaires sous-jacentes.
  - Les donneurs vivants, après un don de rein, ont la possibilité de reprendre une vie normale quelques semaines après l'opération sans conséquence sur leur qualité de vie quotidienne. Néanmoins, certaines études ont précédemment rapporté un risque

---

accru de déclin de la fonction rénale et de maladie rénale par rapport à la population générale [293, 294]. D'autres études sur les donneurs vivants sont donc nécessaires pour évaluer leur fonction rénale à long terme après une néphrectomie et pour identifier les facteurs de risque de déclin accéléré de la fonction rénale. Nous avons étudié les déterminants génétiques de la fonction rénale chez les donneurs vivants dans le cadre de KiT-GENIE. Nous prévoyons également d'intégrer les facteurs génétiques de risque identifiés aux facteurs démographiques et cliniques préopératoires pour prédire le déclin de la fonction rénale après néphrectomie à court (1 an) et à long (5-10 ans) terme.

Les associations génétiques et immunogénétiques identifiées dans KiT-GENIE contribueront (1) à mieux comprendre les principaux défis cliniques liés à la transplantation rénale, (2) à révéler de nouvelles cibles thérapeutiques potentielles à caractériser davantage par des études fonctionnelles, et (3) à identifier des biomarqueurs prédictifs qui peuvent stratifier le risque individuel avant l'apparition des symptômes et améliorer la gestion clinique (médecine personnalisée et prédictive).



SIXIÈME PARTIE

# Références bibliographiques

---





# RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [1] Pierre Mordant. La transplantation d'organe. Petite histoire d'une grande aventure. *Info Respiration*, (82) :27–31, 2007.
- [2] Alexis Carrel and C. C. Guthrie. Functions of a Transplanted Kidney. *Science*, 22(563) :473–473, oct 1905.
- [3] Mathieu Jaboulay. Greffe de reins au pli du coude par soudures artérielles et veineuses. *Lyon Méd*, 107(39) :575–83, 1906.
- [4] Edouard Matevossian, Hans Kern, Norbert Hüser, Dietrich Doll, Yurii Snopok, Jörg Nährig, Jennifer Altomonte, Inga Sinicina, Helmut Friess, and Stefan Thorban. Surgeon Yurii Voronoy (1895-1961) - a pioneer in the history of clinical transplantation : in Memoriam at the 75th Anniversary of the First Human Kidney Transplantation. *Transplant International*, 22(12) :1132–1139, dec 2009.
- [5] David M. Hume, John P. Merrill, Benjamin F. Miller, and George W. Thorn. EXPERIENCES WITH RENAL HOMOTRANSPLANTATION IN THE HUMAN : REPORT OF NINE CASES 1. *Journal of Clinical Investigation*, 34(2) :327–382, feb 1955.
- [6] L MICHON, J HAMBURGER, N OECONOMOS, P DELINOTTE, G RICHEL, J VAYSSE, and B ANTOINE. [An attempted kidney transplantation in man : medical and biological aspects]. *La Presse medicale*, 61(70) :1419–23, nov 1953.
- [7] Carl S. Williamson. Some Observations on the Length of Survival and function of Homogenous Kidney Transplants : Preliminary Report. *Journal of Urology*, 10(4) :275–287, oct 1923.
- [8] Carl S. Williamson. Further Studies on the Transplantation of the Kidney. *Journal of Urology*, 16(4) :231–254, oct 1926.
- [9] William H. Woglom. Absorption of the protective agent from rats resistant to a transplantable sarcoma. *American Journal of Cancer*, 17(4) :873–893, 1933.
- [10] Dominique J. Charron. Jean Dausset (1916–2009). *Nature*, 460(7253) :338–338, jul 2009.
- [11] John P. Merrill. SUCCESSFUL HOMOTRANSPLANTATION OF THE HUMAN KIDNEY BETWEEN IDENTICAL TWINS. *Journal of the American Medical Association*, 160(4) :277, jan 1956.
- [12] Marine Fontaine. Histoire des greffes et des transplantations d'organes \* A short history of graft and transplantation. pages 493–498, 2018.

- [13] E. Donnall Thomas, Harry L. Lochte, Wan Ching Lu, and Joseph W. Ferrebee. Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. *New England Journal of Medicine*, 257(11) :491–496, sep 1957.
- [14] Thomas E. Starzl, Lawrence J. Koep, Charles G. Halgrimson, J. Hood, Gerhard P.J. Schroter, K.A. Porter, and Richard Weil. Fifteen years of clinical liver transplantation. *Gastroenterology*, 77(2) :375–388, aug 1979.
- [15] C N Barnard. The operation. A human cardiac transplant : an interim report of a successful operation performed at Groote Schuur Hospital, Cape Town. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 41(48) :1271–4, dec 1967.
- [16] Emma Buchanan. The operation : A human cardiac transplant : An interim report of a successful operation performed at Groote Schuur Hospital, Cape Town. Author : C N Barnard. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 107(12) :1041–1044, nov 2017.
- [17] James D. Hardy, Watts R. Webb, Martin L. Dalton, and George R. Walker. Lung Homotransplantation in Man. *JAMA*, 186(12), dec 1963.
- [18] Bruck A. Reitz, John L. Wallwork, Sharon A. Hunt, John L. Pennock, Margaret E. Billingham, Philip E. Oyer, Edward B. Stinson, and Norman E. Shumway. Heart-Lung Transplantation. *New England Journal of Medicine*, 306(10) :557–564, mar 1982.
- [19] Agence de la Biomédecine. Press communication 2021. Technical report, 2022.
- [20] John Feehally. *Comprehensive Clinical Nephrology*. Elsevier Health Sciences, 2010.
- [21] Dee Unglaub Silverthorn. *Physiologie humaine : une approche intégrée*. Pearson Education France, 2007.
- [22] M. A. Wallace. Anatomy and physiology of the kidney. *AORN journal*, 68(5) :799–800, nov 1998.
- [23] CUEN Manuel de Néphrologie 9e édition. *Eléments de la physiologie rénale*.
- [24] Bernard Lacour. Physiologie du rein et bases physiopathologiques des maladies rénales. *Revue Francophone des Laboratoires*, 2013(451) :25–37, apr 2013.
- [25] Estelle Geffard. *Développement d’outils de médecine de précision pour accompagner la prise de décision médicale en transplantation*. Thèse de bioinformatique, Université de Nantes - CRTI eq5, 2020.
- [26] CUEN Manuel de Néphrologie 8e édition. *Elevation de la créatininémie*. 2018.
- [27] HAS. Diagnostic de l’insuffisance rénale chronique chez l’adulte. Technical report, 2006.
- [28] B. Dussol. Méthodes d’exploration de la fonction rénale : intérêt et limites des formules permettant d’estimer la fonction rénale. *Immuno-analyse Biologie Spécialisée*, 26(1) :6–12, feb 2011.

- [29] R D Perrone, N E Madias, and A S Levey. Serum Creatinine as an Index of Renal Function : New Insights into Old Concepts. *Clinical Chemistry*, 38(10) :1933–1953, oct 1992.
- [30] Brod, Jan and Jonas H. Sirota. The renal clearance of endogenous “creatinine” in man. *The Journal of Clinical Investigation*, 27(5) :645–54, 1948.
- [31] Geetha Maddukuri and MD. *Evaluation du patient présentant un trouble rénal*. Le manuel MSD, 2020.
- [32] Andrew S. Levey, Lesley A. Stevens, Christopher H. Schmid, Yaping (Lucy) Zhang, Alejandro F. Castro, Harold I. Feldman, John W. Kusek, Paul Eggers, Frederick Van Lente, Tom Greene, and Josef Coresh. A New Equation to Estimate Glomerular Filtration Rate. *Annals of Internal Medicine*, 150(9) :604, may 2009.
- [33] Donald W. Cockcroft and Henry Gault. Prediction of Creatinine Clearance from Serum Creatinine. *Nephron*, 16(1) :31–41, 1976.
- [34] Andrew S. Levey. A More Accurate Method To Estimate Glomerular Filtration Rate from Serum Creatinine : A New Prediction Equation. *Annals of Internal Medicine*, 130(6) :461, mar 1999.
- [35] Kevin Damman, Pim van der Harst, Tom D J Smilde, A A Voors, Gerjan Navis, Dirk J van Veldhuisen, and Hans L Hillege. Use of cystatin C levels in estimating renal function and prognosis in patients with chronic systolic heart failure. *Heart*, 98(4) :319–324, feb 2012.
- [36] Lesley A. Inker, Christopher H. Schmid, Hocine Tighiouart, John H. Eckfeldt, Harold I. Feldman, Tom Greene, John W. Kusek, Jane Manzi, Frederick Van Lente, Yaping Lucy Zhang, Josef Coresh, and Andrew S. Levey. Estimating Glomerular Filtration Rate from Serum Creatinine and Cystatin C. *New England Journal of Medicine*, 367(1) :20–29, jul 2012.
- [37] D Kervella, S Lemoine, F Sens, E Pouliquen, L Dubourg, F Guebre-Egziabher, and L Juillard. Intérêt de la cystatine c dans la mesure de la fonction rénale dans le syndrome cardiorénal de type 2. *Néphrologie & Thérapeutique*, 11(5) :258, 2015.
- [38] Anna Malkina and MD. *Présentation de l’insuffisance rénale*. Le manuel MSD, 2020.
- [39] Anna Malkina. *Lésion rénale aiguë (insuffisance rénale aiguë)*. Le manuel MSD, 2020.
- [40] Arif Khwaja. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. *Nephron*, 120(4) :c179–c184, aug 2012.
- [41] CUEN Manuel de Néphrologie 8e édition. *Insuffisance Rénale Aiguë*. 2018.
- [42] Bénédicte Stengel, Cécile Couchoud, Catherine Helmer, Carole Loos-Ayav, and Michèle Kessler. Épidémiologie de l’insuffisance rénale chronique en France. *La Presse Médicale*, 36(12) :1811–1821, dec 2007.

- [43] CUEN Manuel de Néphrologie 8e édition. *Insuffisance rénale chronique et maladies rénales chroniques*. 2018.
- [44] Agence de la Biomédecine. Annual report 2020 - Kidney. Technical report, 2021.
- [45] B. Dussol. Différents stades de l'insuffisance rénale chronique : recommandations. *Immuno-analyse Biologie Spécialisée*, 26(2) :55–59, apr 2011.
- [46] Bernard Lacour and Ziad Massy. Diagnostic, suivi biologique de l'insuffisance rénale chronique et prise en charge de l'insuffisance rénale chronique terminale. *Revue Francophone des Laboratoires*, 2013(451) :59–73, apr 2013.
- [47] C Jungers, P., Man, N. K., Joly, D., Legendre. *L'insuffisance rénale chronique : prévention et traitement*. Lavoisier, 2011.
- [48] R.E.I.N. Annual Report. Technical report, 2020.
- [49] Agence de la Biomédecine. Le Réseau Épidémiologie et Information en Néphrologie - Journée mondiale du rein. Technical report, Agence de la Biomedecine, 2022.
- [50] Mohamed Nasr, Mohamed Hadj Ammar, Sami Khammouma, Nasr Ben Dhia, and Anouar Ghachem. L'hémodialyse et son impact sur la qualité de vie. *Néphrologie Thérapeutique*, 4(1) :21–27, feb 2008.
- [51] L. Aimee Hechanova and MD. *Dialyse péritonéale*. Le manuel MSD, 2020.
- [52] Marie Matignon, Karine Dahan, Ghislaine Fruchaud, Vincent Audard, Philippe Grimbert, and Philippe Lang. Transplantation rénale : indications, résultats, limites et perspectives. *La Presse Médicale*, 36(12) :1829–1834, dec 2007.
- [53] Guillermo Garcia Garcia, Paul Harden, and Jeremy Chapman. The Global Role of Kidney Transplantation. *Kidney and Blood Pressure Research*, 35(5) :299–304, 2012.
- [54] Fondation du rein. Le don d'organes. Technical report, 2020.
- [55] Christophe Legendre. *La transplantation rénale*. Lavoisier, 2011.
- [56] CUEN Manuel de Néphrologie 8e édition. *Transplantation d'organes*. 2018.
- [57] Jean-Christophe Danober. *Prélèvements et greffes à partir de donneurs de 75 ans et plus, mythe ou réalité ? Description et analyse à partir d'une étude de dix ans en France*. Publié sur hal - hal-01731961, Université de lorraine, 2014.
- [58] Agence de la Biomédecine. Greffe d'organes données générales et méthodes. Technical report, 2021.
- [59] Thierry Hauet. Des situations complexes : l'exemple de la transplantation rénale. Technical report, Springer Healthcare, Poitiers, 2017.
- [60] Martin Hertl and PhD , MD. *Transplantation rénale*. Le manuel MSD, 2020.

- [61] Ana I. Sánchez-Fructuoso, Maria Marques, Dolores Prats, José Conesa, Natividad Calvo, M. Jesús Pérez-Contín, Jesús Blazquez, Cristina Fernández, Ervigio Corral, Francisco Del Río, Jose R. Núñez, and Alberto Barrientos. Victims of Cardiac Arrest Occurring Outside the Hospital : A Source of Transplantable Kidneys. *Annals of Internal Medicine*, 145(3) :157, aug 2006.
- [62] Thomas Bertsch, Jochen Lüdecke, Werner Antl, and Lydia W M Nausch. Karl Landsteiner : The Discovery of the ABO Blood Group System and its Value for Teaching Medical Students. *Clinical laboratory*, 65(6), jun 2019.
- [63] Eiji Hosoi. Biological and clinical aspects of ABO blood group system. *The Journal of Medical Investigation*, 55(3,4) :174–182, 2008.
- [64] K Hadaya. [ABO incompatible renal transplantation]. *Revue medicale suisse*, 8(346) :1310–3, jun 2012.
- [65] Kota Takahashi. A new concept of accommodation in ABO-incompatible kidney transplantation. *Clinical Transplantation*, 19(s14) :76–85, aug 2005.
- [66] Kota Takahashi and Kazuhide Saito. ABO-incompatible kidney transplantation. *Transplantation Reviews*, 27(1) :1–8, jan 2013.
- [67] Stefan Zschiedrich, Albrecht Kramer-Zucker, Bernd Jänigen, Maximilian Seidl, Florian Emmerich, Przemyslaw Pisarski, and Tobias B. Huber. An update on ABO-incompatible kidney transplantation. *Transplant International*, 28(4) :387–397, apr 2015.
- [68] Nephro-Blog. Greffe rénale ABO incompatible, 2011.
- [69] Christian Morath, Martin Zeier, Bernd Döhler, Gerhard Opelz, and Caner Süsal. ABO-Incompatible Kidney Transplantation. *Frontiers in Immunology*, 8, mar 2017.
- [70] Myriam Labalette, Siamak Bahram, and Marie Christine Béné. Le complexe majeur d’histocompatibilité humain (HLA). Polymorphisme et présentation des antigènes aux lymphocytes T. Technical report, 2015.
- [71] Jean Dausset and Felix T. Rapaport. The HLA Story. In *Immunology*, pages 111–120. Elsevier, 1995.
- [72] Takashi Shiina, Kazuyoshi Hosomichi, Hidetoshi Inoko, and Jerzy K Kulski. The HLA genomic loci map : expression, interaction, diversity and disease. *Journal of Human Genetics*, 54(1) :15–39, jan 2009.
- [73] Corrine M.C. Heijmans, Natasja G. Groot, and Ronald E. Bontrop. Comparative genetics of the major histocompatibility complex in humans and nonhuman primates. *International Journal of Immunogenetics*, 47(3) :243–260, jun 2020.
- [74] Kenneth L. Rock, Eric Reits, and Jacques Neefjes. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends in Immunology*, 37(11) :724–737, nov 2016.

- [75] James Robinson, Anup R. Soormally, James D. Hayhurst, and Steven G.E. Marsh. The IPD-IMGT/HLA Database – New developments in reporting HLA variation. *Human Immunology*, 77(3) :233–237, mar 2016.
- [76] Diogo Meyer, Vitor R. C. Aguiar, Bárbara D. Bitarello, Débora Y. C. Brandt, and Kelly Nunes. A genomic perspective on HLA evolution. *Immunogenetics*, 70(1) :5–27, jan 2018.
- [77] A. E. Kennedy, U. Ozbek, and M. T. Dorak. What has GWAS done for HLA and disease associations? *International Journal of Immunogenetics*, 44(5) :195–211, oct 2017.
- [78] Paul I. Terasaki, Domenico Bernoco, Min Sik Park, Gungor Ozturk, and Yuichi Iwaki. Microdroplet Testing for HLA-A, -B, -C, and -D Antigens : The Philip Levine Award Lecture. *American Journal of Clinical Pathology*, 69(2) :103–120, feb 1978.
- [79] F. Jordan, A. J. McWhinnie, S. Turner, N. Gavira, A. A. Calvert, S. A. Cleaver, R. H. Holman, J. M. Goldman, and J. A. Madrigal. Comparison of HLA-DRB1 typing by DNA-RFLP, PCR-SSO and PCR-SSP methods and their application in providing matched unrelated donors for bone marrow transplantation. *Tissue Antigens*, 45(2) :103–110, feb 1995.
- [80] Christina E. M. Voorter, Fausto Palusci, and Marcel G. J. Tilanus. Sequence-Based Typing of HLA : An Improved Group-Specific Full-Length Gene Sequencing Approach. pages 101–114. 2014.
- [81] D. De Santis, D. Dinauer, J. Duke, H. A. Erlich, C. L. Holcomb, C. Lind, K. Mackiewicz, D. Monos, A. Moudgil, P. Norman, P. Parham, A. Sasson, and R. J. N. Allcock. 16 th IHIW : Review of HLA typing by NGS. *International Journal of Immunogenetics*, 40(1) :72–76, feb 2013.
- [82] J.-M. Tiercy, S. G. E. Marsh, G. M. Th. Schreuder, E. Albert, G. Fischer, and R. Wassmuth. Guidelines for nomenclature usage in HLA reports : ambiguities and conversion to serotypes. *European Journal of Immunogenetics*, 29(3) :273–274, jun 2002.
- [83] S. G. E. Marsh, E. D. Albert, W. F. Bodmer, R. E. Bontrop, B. Dupont, H. A. Erlich, M. Fernández-Viña, D. E. Geraghty, R. Holdsworth, C. K. Hurley, M. Lau, K. W. Lee, B. Mach, M. Maiers, W. R. Mayr, C. R. Müller, P. Parham, E. W. Petersdorf, T. Sasazuki, J. L. Strominger, A. Svejgaard, P. I. Terasaki, J. M. Tiercy, and J. Trowsdale. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, 75(4) :291–455, apr 2010.
- [84] Carolyn Katovich Hurley. Naming HLA diversity : A review of HLA nomenclature. *Human Immunology*, 82(7) :457–465, jul 2021.
- [85] Calliope A. Dendrou, Jan Petersen, Jamie Rossjohn, and Lars Fugger. HLA variation and disease. *Nature Reviews Immunology*, 18(5) :325–339, may 2018.

- [86] Sreeram V. Ramagopalan, Narelle J. Maugeri, Lahiru Handunnetthi, Matthew R. Lincoln, Sarah-Michelle Orton, David A. Dymont, Gabriele C. DeLuca, Blanca M. Herrera, Michael J. Chao, A. Dessa Sadovnick, George C. Ebers, and Julian C. Knight. Expression of the Multiple Sclerosis-Associated MHC Class II Allele HLA-DRB1\*1501 Is Regulated by Vitamin D. *PLoS Genetics*, 5(2) :e1000369, feb 2009.
- [87] Andrej Košmrlj, Elizabeth L. Read, Ying Qi, Todd M. Allen, Marcus Altfeld, Steven G. Deeks, Florencia Pereyra, Mary Carrington, Bruce D. Walker, and Arup K. Chakraborty. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature*, 465(7296) :350–354, may 2010.
- [88] Christian Morath, Gerhard Opelz, Martin Zeier, and Caner Süsal. Clinical Relevance of HLA Antibody Monitoring after Kidney Transplantation. *Journal of Immunology Research*, 2014 :1–5, 2014.
- [89] Edward J. Filippone and John L. Farber. Humoral Immune Response and Allograft Function in Kidney Transplantation. *American Journal of Kidney Diseases*, 66(2) :337–347, aug 2015.
- [90] W M Howell, V Carter, and B Clark. The HLA system : immunobiology, HLA typing, antibody screening and crossmatching techniques. *Journal of Clinical Pathology*, 63(5) :387–390, may 2010.
- [91] Nori Sasaki and Adam Idica. The HLA-matching effect in different cohorts of kidney transplant recipients : 10 years later. *Clinical transplants*, pages 261–82, 2010.
- [92] Renaud Snanoudj and Christophe Legendre. Nouveaux aspects de la compatibilité HLA en transplantation. *Néphrologie Thérapeutique*, 12 :S125–S130, apr 2016.
- [93] P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, 329(6139) :512–518, oct 1987.
- [94] A. R. Tambur and F. H. J. Claas. HLA Epitopes as Viewed by Antibodies : What Is it All About ? *American Journal of Transplantation*, 15(5) :1148–1154, may 2015.
- [95] Renaud Snanoudj, Nassim Kamar, Elisabeth Cassuto, Sophie Caillard, Marie Metzger, Pierre Merville, Antoine Thierry, Isabelle Jollet, Philippe Grimbert, Dany Anglicheau, Marc Hazzan, Gabriel Choukroun, Bruno Hurault De Ligny, Bénédicte Janbon, Vincent Vuiblet, Anne Devys, Yann Le Meur, Michel Delahousse, Emmanuel Morelon, Elodie Bailly, Sophie Girerd, Kahina Amokrane, Christophe Legendre, Alexandre Hertig, Eric Rondeau, and Jean-Luc Taupin. Epitope load identifies kidney transplant recipients at risk of allosensitization following minimization of immunosuppression. *Kidney International*, 95(6) :1471–1485, jun 2019.



- [96] Rene J. Duquesnoy, Steve Takemoto, Peter de Lange, Ilias I. N. Doxiadis, Geziena M. Th. Schreuder, Guido G. Persijn, and Frans H. J. Claas. HLAmatchmaker : a molecularly based algorithm for histocompatibility determination. III. Effect of matching at the HLA-A,B amino acid triplet level on kidney transplant survival1. *Transplantation*, 75(6) :884–889, mar 2003.
- [97] Cynthia S. M. Kramer, Johan Koster, Geert W. Haasnoot, Dave L. Roelen, Frans H. J. Claas, and Sebastiaan Heidt. HLA-EMMA : A user-friendly tool to analyse HLA class I and class II compatibility on the amino acid level. *HLA*, 96(1) :43–51, jul 2020.
- [98] Kirsten Geneugelijk, Matthias Niemann, Julia Drylewicz, Arjan van Zuilen, Irma Joosten, Wil Allebes, Arnold van der Meer, Luuk Hilbrands, Marije Baas, Erik Hack, Franka van Reekum, Marianne Verhaar, Elena Kamburova, Michiel Bots, Marc Seelen, Jan-Stefan Sanders, Bouke Hepkema, Annechien Lambeck, Laura Bungener, Marcel Tilanus, Christien Voorter, Joris Vanderlocht, Lotte Wieten, Elly Duijnhoven, Marielle Gelens, Maarten Christiaans, Frans van Ittersum, Azam Nurmohamed, Neubury Lardy, Wendy Swelsen, Karlijn van der Pant, Neelke van der Weerd, Ineke ten Berge, Frederike Bemelman, Andries Hoitsma, Paul van der Boog, Johan de Fijter, Michiel Betjes, Sebastiaan Heidt, Dave Roelen, Frans Claas, Henny Otten, and Eric Spierings. OR41 PIRCHE-II : A novel tool to identify permissible HLA mismatches in kidney transplantation. *Human Immunology*, 78 :39, sep 2017.
- [99] Estelle Geffard, Léo Boussamet, Alexandre Walencik, Florent Delbos, Sophie Limou, Pierre-Antoine Gourraud, and Nicolas Vince. HLA-EPI : A new EPIisode in exploring donor/recipient epitopic compatibilities. *HLA*, 99(2) :79–92, feb 2022.
- [100] Alice Koenig, Chien-Chia Chen, Antoine Marçais, Thomas Barba, Virginie Mathias, Antoine Sicard, Maud Rabeyrin, Maud Racapé, Jean-Paul Duong-Van-Huyen, Patrick Brunel, Alexandre Loupy, Sébastien Dussurgey, Stéphanie Ducreux, Vannary Meas-Yedid, Jean-Christophe Olivo-Marin, Héléna Paidassi, Romain Guillemain, Jean-Luc Taupin, Jasper Callemeyn, Emmanuel Morelon, Antonino Nicoletti, Béatrice Charreau, Valérie Dubois, Maarten Naesens, Thierry Walzer, Thierry Defrance, and Olivier Thaumat. Missing self triggers NK cell-mediated chronic vascular rejection of solid organ transplants. *Nature Communications*, 10(1) :5350, dec 2019.
- [101] Emilie Lebraud, Maëva Eloudzeri, Marion Rabant, Baptiste Lamarthée, and Dany Anglicheau. Microvascular Inflammation of the Renal Allograft : A Reappraisal of the Underlying Mechanisms. *Frontiers in Immunology*, 13, mar 2022.
- [102] Carmen Lefaucheur, Kevin Louis, Aurélie Philippe, Alexandre Loupy, and P. Toby Coates. The emerging field of non-human leukocyte antigen antibodies in transplant medicine and beyond. *Kidney International*, 100(4) :787–798, oct 2021.

- [103] Gillian Divard, Marc Raynaud, Vasishta S. Tatapudi, Basmah Abdalla, Elodie Bailly, Maureen Assayag, Yannick Binois, Raphael Cohen, Huanxi Zhang, Camillo Ulloa, Kamila Linhares, Helio S. Tedesco, Christophe Legendre, Xavier Jouven, Robert A. Montgomery, Carmen Lefaucheur, Olivier Aubert, and Alexandre Loupy. Comparison of artificial intelligence and human-based prediction and stratification of the risk of long-term kidney allograft failure. *Communications Medicine*, 2(1) :150, nov 2022.
- [104] Marc Raynaud, Olivier Aubert, Gillian Divard, Peter P. Reese, Nassim Kamar, Daniel Yoo, Chen Shan Chin, Élodie Bailly, Matthias Buchler, Marc Ladrière, Moglie Le Quintrec, Michel Delahousse, Ivana Juric, Nikolina Basic-Jukic, Marta Crespo, Helio Tedesco Silva, Kamilla Linhares, Maria Cristina Ribeiro de Castro, Gervasio Soler Pujol, Jean Philippe Empana, Camilo Ulloa, Enver Akalin, Georg Böhmig, Edmund Huang, Mark D. Stegall, Andrew J. Bentall, Robert A. Montgomery, Stanley C. Jordan, Rainer Oberbauer, Dorry L. Segev, John J. Friedewald, Xavier Jouven, Christophe Legendre, Carmen Lefaucheur, and Alexandre Loupy. Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence : an observational, international, multicohort study. *The Lancet Digital Health*, 3(12) :e795–e805, 2021.
- [105] Centre de transplantation d’organes - CHU de Lausanne. Déroulement de la greffe, 2019.
- [106] Centre de transplantation d’organes - CHU de Lausanne. Déroulement de la greffe, 2019.
- [107] Haute Autorité de Santé - HAS. Suivi ambulatoire de l’adulte transplanté rénal au-delà de 3 mois après transplantation. Technical Report 3, 2008.
- [108] Lionel Couzi, Bruno Moulin, Marie-Pascale Morin, Laetitia Albano, Michel Godin, Benoit Barrou, Eric Alamartine, Emmanuel Morelon, Sandrine Girardot-Seguin, Laurence Mendes, David Misdrahi, Elisabeth Cassuto, and Pierre Merville. Factors Predictive of Medication Nonadherence After Renal Transplantation. *Transplantation Journal*, 95(2) :326–332, jan 2013.
- [109] Candice Roufosse, Naomi Simmonds, Marian Clahsen-van Groningen, Mark Haas, Kammi J. Henriksen, Catherine Horsfield, Alexandre Loupy, Michael Mengel, Agnieszka Perkowska-Ptasińska, Marion Rabant, Lorraine C. Racusen, Kim Solez, and Jan U. Becker. A 2018 Reference Guide to the Banff Classification of Renal Allograft Pathology. *Transplantation*, 102(11) :1795–1814, nov 2018.
- [110] Brian J. Nankivell and Stephen I. Alexander. Rejection of the Kidney Allograft. *New England Journal of Medicine*, 363(15) :1451–1462, oct 2010.
- [111] David S Game and Robert I Lechler. Pathways of allorecognition : implications for transplantation tolerance. *Transplant Immunology*, 10(2-3) :101–108, aug 2002.
- [112] P E Harris, R Cortesini, and N Suci-Foca. Indirect allorecognition in solid organ transplantation. *Reviews in immunogenetics*, 1(3) :297–308, 1999.

- [113] M Refik Gökmen, Giovanna Lombardi, and Robert I Lechler. The importance of the indirect pathway of allorecognition in clinical transplantation. *Current Opinion in Immunology*, 20(5) :568–574, oct 2008.
- [114] MR Clatworthy, M Espeli, N Torpey, and KGC Smith. The generation and maintenance of serum alloantibody. *Current Opinion in Immunology*, 22(5) :669–681, oct 2010.
- [115] Shamila Mauiyyedi, Marta Crespo, A. Bernard Collins, Eveline E. Schneeberger, Manuel A. Pascual, Susan L. Saidman, Nina E. Tolkoff-Rubin, Winfred W. Williams, Francis L. Delmonico, A. Benedict Cosimi, and Robert B. Colvin. Acute Humoral Rejection in Kidney Transplantation : II. Morphology, Immunopathology, and Pathologic Classification. *Journal of the American Society of Nephrology*, 13(3) :779–787, mar 2002.
- [116] Adam Gondos, Bernd Döhler, Hermann Brenner, and Gerhard Opelz. Kidney Graft Survival in Europe and the United States. *Transplantation Journal*, 95(2) :267–274, jan 2013.
- [117] Paul I. Terasaki. Humoral Theory of Transplantation. *American Journal of Transplantation*, 3(6) :665–673, jun 2003.
- [118] Claire Tinel, Baptiste Lamarthée, Jasper Callemeyn, Elisabet Van Loon, Virginia Sauvaget, Lise Morin, Laïla Aouni, Marion Rabant, Wilfried Gwinner, Pierre Marquet, Maarten Naesens, and Dany Anglicheau. Integrative Omics Analysis Unravels Microvascular Inflammation-Related Pathways in Kidney Allograft Biopsies. *Frontiers in Immunology*, 12, nov 2021.
- [119] Penelope J. Allen, Steve J. Chadban, Jonathan C. Craig, Wai H. Lim, Richard D.M. Allen, Philip A. Clayton, Armando Teixeira-Pinto, and Germaine Wong. Recurrent glomerulonephritis after kidney transplantation : risk factors and allograft outcomes. *Kidney International*, 92(2) :461–469, aug 2017.
- [120] P. Niaudet. Récidive de la maladie initiale après transplantation rénale, 2006.
- [121] Vivette D. D’Agati, Frederick J. Kaskel, and Ronald J. Falk. Focal Segmental Glomerulosclerosis. *New England Journal of Medicine*, 365(25) :2398–2411, dec 2011.
- [122] Michael Rudnicki. FSGS Recurrence in Adults after Renal Transplantation. *BioMed Research International*, 2016 :1–7, 2016.
- [123] Agathe Pardon, Vincent Audard, Sophie Caillard, Bruno Moulin, Dominique Desvaux, Bouteina Bentaarit, Philippe Remy, Dil Sahali, Françoise Roudot-Thoraval, Philippe Lang, and Philippe Grimbert. Risk factors and outcome of focal and segmental glomerulosclerosis recurrence in adult renal transplant recipients. *Nephrology Dialysis Transplantation*, 21(4) :1053–1059, apr 2006.
- [124] Changli Wei, Shafic El Hindi, Jing Li, Alessia Fornoni, Nelson Goes, Junichiro Sageshima, Dony Maiguel, S Ananth Karumanchi, Hui-Kim Yap, Moin Saleem, Qingyin Zhang, Boris

- Nikolic, Abanti Chaudhuri, Pirouz Daftarian, Eduardo Salido, Armando Torres, Moro Sa-lifu, Minnie M Sarwal, Franz Schaefer, Christian Morath, Vedat Schwenger, Martin Zeier, Vineet Gupta, David Roth, Maria Pia Rastaldi, George Burke, Phillip Ruiz, and Jochen Reiser. Circulating urokinase receptor as a cause of focal segmental glomerulosclerosis. *Nature Medicine*, 17(8) :952–960, aug 2011.
- [125] Lorenzo Gallon, Joseph Leventhal, Anton Skaro, Yashpal Kanwar, and Antonio Alvarado. Resolution of Recurrent Focal Segmental Glomerulosclerosis after Retransplantation. *New England Journal of Medicine*, 366(17) :1648–1649, apr 2012.
- [126] Rutger J. H. Maas, Jeroen K. J. Deegens, and Jack F. M. Wetzels. Serum suPAR in patients with FSGS : trash or treasure? *Pediatric Nephrology*, 28(7) :1041–1048, jul 2013.
- [127] Richard A. Gibbs. The Human Genome Project changed everything. *Nature Reviews Genetics*, 21(10) :575–576, oct 2020.
- [128] Rokhaya Ba, Estelle Geffard, Venceslas Douillard, Françoise Simon, Laurent Mesnard, Nicolas Vince, Pierre-Antoine Gourraud, and Sophie Limou. Surfing the Big Data Wave : Omics Data Challenges in Transplantation. *Transplantation*, 106(2) :e114–e125, feb 2022.
- [129] J. Craig Venter, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, Mark Yandell, Cheryl A. Evans, Robert A. Holt, Jeannine D. Gocayne, Peter Amanatides, Richard M. Ballew, Daniel H. Huson, Jennifer Russo Wortman, Qing Zhang, Chinnappa D. Kodira, Xiangqun H. Zheng, Lin Chen, Marian Skupski, Gangadharan Subramanian, Paul D. Thomas, Jinghui Zhang, George L. Gabor Miklos, Catherine Nelson, Samuel Broder, Andrew G. Clark, Joe Nadeau, Victor A. McKusick, Norton Zinder, Arnold J. Levine, Richard J. Roberts, Mel Simon, Carolyn Slayman, Michael Hunkapiller, Randall Bolanos, Arthur Delcher, Ian Dew, Daniel Fasulo, Michael Flanigan, Liliana Florea, Aaron Halpern, Sridhar Hannenhalli, Saul Kravitz, Samuel Levy, Clark Mobarry, Knut Reinert, Karin Remington, Jane Abu-Threideh, Ellen Beasley, Kendra Biddick, Vivien Bonazzi, Rhonda Brandon, Michele Cargill, Ishwar Chandramouliswaran, Rosane Charlab, Kabir Chaturvedi, Zuoming Deng, Valentina Di Francesco, Patrick Dunn, Karen Eilbeck, Carlos Evangelista, Andrei E. Gabrielian, Weiniu Gan, Wangmao Ge, Fangcheng Gong, Zhiping Gu, Ping Guan, Thomas J. Heiman, Maureen E. Higgins, Rui-Ru Ji, Zhaoxi Ke, Karen A. Ketchum, Zhongwu Lai, Yiding Lei, Zhenya Li, Jiayin Li, Yong Liang, Xiaoying Lin, Fu Lu, Gennady V. Merkulov, Natalia Milshina, Helen M. Moore, Ashwinikumar K Naik, Vaibhav A. Narayan, Beena Neelam, Deborah Nusskern, Douglas B. Rusch, Steven Salzberg, Wei Shao, Bixiong Shue, Jingtao Sun, Zhen Yuan Wang, Aihui Wang, Xin Wang, Jian Wang, Ming-Hui Wei, Ron Wides, Chunlin Xiao, Chunhua Yan, Alison Yao, Jane Ye, Ming Zhan, Weiqing Zhang, Hongyu Zhang, Qi Zhao, Liansheng Zheng, Fei Zhong, Wenyan Zhong, Shiaoping C. Zhu,

Shaying Zhao, Dennis Gilbert, Suzanna Baumhueter, Gene Spier, Christine Carter, Anibal Cravchik, Trevor Woodage, Feroze Ali, Huijin An, Aderonke Awe, Danita Baldwin, Holly Baden, Mary Barnstead, Ian Barrow, Karen Beeson, Dana Busam, Amy Carver, Angela Center, Ming Lai Cheng, Liz Curry, Steve Danaher, Lionel Davenport, Raymond Desilets, Susanne Dietz, Kristina Dodson, Lisa Doup, Steven Ferriera, Neha Garg, Andres Gluecksmann, Brit Hart, Jason Haynes, Charles Haynes, Cheryl Heiner, Suzanne Hladun, Damon Hostin, Jarrett Houck, Timothy Howland, Chinyere Ibegwam, Jeffery Johnson, Francis Kalush, Lesley Kline, Shashi Koduru, Amy Love, Felecia Mann, David May, Steven McCawley, Tina McIntosh, Ivy McMullen, Mee Moy, Linda Moy, Brian Murphy, Keith Nelson, Cynthia Pfannkoch, Eric Pratts, Vinita Puri, Hina Qureshi, Matthew Reardon, Robert Rodriguez, Yu-Hui Rogers, Deanna Romblad, Bob Ruhfel, Richard Scott, Cynthia Sitter, Michelle Smallwood, Erin Stewart, Renee Strong, Ellen Suh, Reginald Thomas, Ni Ni Tint, Sukyee Tse, Claire Vech, Gary Wang, Jeremy Wetter, Sherita Williams, Monica Williams, Sandra Windsor, Emily Winn-Deen, Keriellen Wolfe, Jayshree Zaveri, Karena Zaveri, Josep F. Abril, Roderic Guigo, Michael J. Campbell, Kimmen V. Sjolander, Brian Karlak, Anish Kejariwal, Huaiyu Mi, Betty Lazareva, Thomas Hatton, Apurva Narechania, Karen Diemer, Anushya Muruganujan, Nan Guo, Shinji Sato, Vineet Bafna, Sorin Istrail, Ross Lippert, Russell Schwartz, Brian Walenz, Shibu Yooseph, David Allen, Anand Basu, James Baxendale, Louis Blick, Marcelo Caminha, John Carnes-Stine, Parris Caulk, Yen-Hui Chiang, My Coyne, Carl Dahlke, Anne Deslattes Mays, Maria Dombroski, Michael Donnelly, Dale Ely, Shiva Esparham, Carl Fosler, Harold Gire, Stephen Glanowski, Kenneth Glasser, Anna Glodek, Mark Gorokhov, Ken Graham, Barry Gropman, Michael Harris, Jeremy Heil, Scott Henderson, Jeffrey Hoover, Donald Jennings, Catherine Jordan, James Jordan, John Kasha, Leonid Kagan, Cheryl Kraft, Alexander Levitsky, Mark Lewis, Xiangjun Liu, John Lopez, Daniel Ma, William Majoros, Joe McDaniel, Sean Murphy, Matthew Newman, Trung Nguyen, Ngoc Nguyen, Marc Nodell, Sue Pan, Jim Peck, Marshall Peterson, William Rowe, Robert Sanders, John Scott, Michael Simpson, Thomas Smith, Arlan Sprague, Timothy Stockwell, Russell Turner, Eli Venter, Mei Wang, Meiyuan Wen, David Wu, Mitchell Wu, Ashley Xia, Ali Zandieh, and Xiaohong Zhu. The Sequence of the Human Genome. *Science*, 291(5507) :1304–1351, feb 2001.

- [130] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, Roel Funke, Diane Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie LeVine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian, Dudley Wyman, Jane Rogers, John Sulston, Rachael Ainscough, Stephan

Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Showkeen, Sarah Sims, Robert H. Waterston, Richard K. Wilson, LaDeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx, Sandra W. Clifton, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan-Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher, Marvin Frazier, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson, George M. Weinstock, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki, Todd Taylor, Jean Weissenbach, Roland Heilig, William Saurin, Francois Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert, Patrick Wincker, André Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien, Andreas Rump, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Hong Mei Lee, JoAnn Dubois, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang, Jun Gu, Leroy Hood, Lee Rowen, Anup Madan, Shizen Qin, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola, Michael J. Proctor, Bruce A. Roe, Feng Chen, Huaqin Pan, Juliane Ramser, Hans Lehrach, Richard Reinhardt, W. Richard McCombie, Melissa de la Bastide, Neilay Dedhia, Helmut Blöcker, Klaus Hornischer, Gabriele Nordsiek, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu-Chuan Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kasprzyk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran, Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, Jörg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw-Pyng Yang, Ru-Fang

Yeh, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld, Kris A. Wetters-  
trand, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood, David R. Cox,  
Maynard V. Olson, Rajinder Kaul, Christopher Raymond, Nobuyoshi Shimizu, Kazuhiko  
Kawasaki, Shinsei Minoshima, Glen A. Evans, Maria Athanasiou, Roger Schultz, Aristides  
Patrinos, and Michael J. Morgan. Initial sequencing and analysis of the human genome.  
*Nature*, 409(6822) :860–921, feb 2001.

- [131] Bertrand Jordan. La génomique et la diversité humaine. *Cahiers de l'Urmis*, (20), jun  
2021.
- [132] Adam Auton, Gonçalo R. Abecasis, David M. Altshuler, Richard M. Durbin, Gonçalo R.  
Abecasis, David R. Bentley, Aravinda Chakravarti, Andrew G. Clark, Peter Donnelly,  
Evan E. Eichler, Paul Flicek, Stacey B. Gabriel, Richard A. Gibbs, Eric D. Green, Mat-  
thew E. Hurles, Bartha M. Knoppers, Jan O. Korb, Eric S. Lander, Charles Lee, Hans  
Lehrach, Elaine R. Mardis, Gabor T. Marth, Gil A. McVean, Deborah A. Nickerson, Je-  
nette P. Schmidt, Stephen T. Sherry, Jun Wang, Richard K. Wilson, Richard A. Gibbs,  
Eric Boerwinkle, Harsha Doddapaneni, Yi Han, Viktoriya Korchina, Christie Kovar, San-  
dra Lee, Donna Muzny, Jeffrey G. Reid, Yiming Zhu, Jun Wang, Yuqi Chang, Qiang Feng,  
Xiaodong Fang, Xiaosen Guo, Min Jian, Hui Jiang, Xin Jin, Tianming Lan, Guoqing Li,  
Jingxiang Li, Yingrui Li, Shengmao Liu, Xiao Liu, Yao Lu, Xuedi Ma, Meifang Tang,  
Bo Wang, Guangbiao Wang, Honglong Wu, Renhua Wu, Xun Xu, Ye Yin, Dandan Zhang,  
Wenwei Zhang, Jiao Zhao, Meiru Zhao, Xiaole Zheng, Eric S. Lander, David M. Altshuler,  
Stacey B. Gabriel, Namrata Gupta, Neda Gharani, Lorraine H. Toji, Norman P. Gerry,  
Alissa M. Resch, Paul Flicek, Jonathan Barker, Laura Clarke, Laurent Gil, Sarah E. Hunt,  
Gavin Kelman, Eugene Kulesha, Rasko Leinonen, William M. McLaren, Rajesh Radha-  
krishnan, Asier Roa, Dmitriy Smirnov, Richard E. Smith, Ian Streeter, Anja Thormann,  
Iliana Toneva, Brendan Vaughan, Xiangqun Zheng-Bradley, David R. Bentley, Russell Gro-  
cock, Sean Humphray, Terena James, Zoya Kingsbury, Hans Lehrach, Ralf Sudbrak, Mar-  
cus W. Albrecht, Vyacheslav S. Amstislavskiy, Tatiana A. Borodina, Matthias Lienhard,  
Florian Mertes, Marc Sultan, Bernd Timmermann, Marie-Laure Yaspo, Elaine R. Mardis,  
Richard K. Wilson, Lucinda Fulton, Robert Fulton, Stephen T. Sherry, Victor Ananiev,  
Zinaida Belaia, Dimitriy Beloslyudtsev, Nathan Bouk, Chao Chen, Deanna Church, Ro-  
bert Cohen, Charles Cook, John Garner, Timothy Hefferon, Mikhail Kimelman, Chunlei  
Liu, John Lopez, Peter Meric, Chris O'Sullivan, Yuri Ostapchuk, Lon Phan, Sergiy Pono-  
marov, Valerie Schneider, Eugene Shekhtman, Karl Sirotkin, Douglas Slotta, Hua Zhang,  
Gil A. McVean, Richard M. Durbin, Senduran Balasubramaniam, John Burton, Petr Da-  
necek, Thomas M. Keane, Anja Kolb-Kokocinski, Shane McCarthy, James Stalker, Michael  
Quail, Jeanette P. Schmidt, Christopher J. Davies, Jeremy Gollub, Teresa Webster, Brant  
Wong, Yiping Zhan, Adam Auton, Christopher L. Campbell, Yu Kong, Anthony Mar-

cketta, Richard A. Gibbs, Fuli Yu, Lilian Antunes, Matthew Bainbridge, Donna Muzny, Aniko Sabo, Zhuoyi Huang, Jun Wang, Lachlan J. M. Coin, Lin Fang, Xiaosen Guo, Xin Jin, Guoqing Li, Qibin Li, Yingrui Li, Zhenyu Li, Haoxiang Lin, Binghang Liu, Ruibang Luo, Haojing Shao, Yinlong Xie, Chen Ye, Chang Yu, Fan Zhang, Hancheng Zheng, Hongmei Zhu, Can Alkan, Elif Dal, Fatma Kahveci, Gabor T. Marth, Erik P. Garrison, Deniz Kural, Wan-Ping Lee, Wen Fung Leong, Michael Stromberg, Alistair N. Ward, Jiantao Wu, Mengyao Zhang, Mark J. Daly, Mark A. DePristo, Robert E. Handsaker, David M. Altshuler, Eric Banks, Gaurav Bhatia, Guillermo del Angel, Stacey B. Gabriel, Giulio Genovese, Namrata Gupta, Heng Li, Seva Kashin, Eric S. Lander, Steven A. McCarroll, James C. Nemes, Ryan E. Poplin, Seungtae C. Yoon, Jayon Lihm, Vladimir Makarov, Andrew G. Clark, Srikanth Gottipati, Alon Keinan, Juan L. Rodriguez-Flores, Jan O. Korbel, Tobias Rausch, Markus H. Fritz, Adrian M. Stütz, Paul Flicek, Kathryn Beal, Laura Clarke, Avik Datta, Javier Herrero, William M. McLaren, Graham R. S. Ritchie, Richard E. Smith, Daniel Zerbino, Xiangqun Zheng-Bradley, Pardis C. Sabeti, Ilya Shlyakhter, Stephen F. Schaffner, Joseph Vitti, David N. Cooper, Edward V. Ball, Peter D. Stenson, David R. Bentley, Bret Barnes, Markus Bauer, R. Keira Cheetham, Anthony Cox, Michael Eberle, Sean Humphray, Scott Kahn, Lisa Murray, John Peden, Richard Shaw, Eimear E. Kenny, Mark A. Batzer, Miriam K. Konkel, Jerilyn A. Walker, Daniel G. MacArthur, Monkol Lek, Ralf Sudbrak, Vyacheslav S. Amstislavskiy, Ralf Herwig, Elaine R. Mardis, Li Ding, Daniel C. Koboldt, David Larson, Kai Ye, Simon Gravel, Anand Swaroop, Emily Chew, Tuuli Lappalainen, Yaniv Erlich, Melissa Gymrek, Thomas Frederick Willems, Jared T. Simpson, Mark D. Shriver, Jeffrey A. Rosenfeld, Carlos D. Bustamante, Stephen B. Montgomery, Francisco M. De La Vega, Jake K. Byrnes, Andrew W. Carroll, Marianne K. DeGorter, Phil Lacroute, Brian K. Maples, Alicia R. Martin, Andres Moreno-Estrada, Suyash S. Shringarpure, Fouad Zakharia, Eran Halperin, Yael Baran, Charles Lee, Eliza Cerveira, Jaeho Hwang, Ankit Malhotra, Dariusz Plewczynski, Kamen Radew, Mallory Romanovitch, Chengsheng Zhang, Fiona C. L. Hyland, David W. Craig, Alexis Christoforides, Nils Homer, Tyler Izatt, Ahmet A. Kurdoglu, Shripad A. Sinari, Kevin Squire, Stephen T. Sherry, Chunlin Xiao, Jonathan Sebat, Danny Antaki, Madhusudan Gujral, Amina Noor, Kenny Ye, Esteban G. Burchard, Ryan D. Hernandez, Christopher R. Gignoux, David Haussler, Sol J. Katzman, W. James Kent, Bryan Howie, Andres Ruiz-Linares, Emmanouil T. Dermitzakis, Scott E. Devine, Gonçalo R. Abecasis, Hyun Min Kang, Jeffrey M. Kidd, Tom Blackwell, Sean Caron, Wei Chen, Sarah Emery, Lars Fritsche, Christian Fuchsberger, Goo Jun, Bingshan Li, Robert Lyons, Chris Scheller, Carlo Sidore, Shiya Song, Elzbieta Sliwerska, Daniel Taliun, Adrian Tan, Ryan Welch, Mary Kate Wing, Xiaowei Zhan, Philip Awadalla, Alan Hodgkinson, Yun Li, Xinghua Shi, Andrew Quitadamo, Gerton Lunter, Gil A. McVean, Jonathan L. Marchini, Simon



Myers, Claire Churchhouse, Olivier Delaneau, Anjali Gupta-Hinch, Warren Kretzschmar, Zamin Iqbal, Iain Mathieson, Androniki Menelaou, Andy Rimmer, Dionysia K. Xifara, Taras K. Oleksyk, Yunxin Fu, Xiaoming Liu, Momiao Xiong, Lynn Jorde, David Witherspoon, Jinchuan Xing, Evan E. Eichler, Brian L. Browning, Sharon R. Browning, Fereydoun Hormozdiari, Peter H. Sudmant, Ekta Khurana, Richard M. Durbin, Matthew E. Hurles, Chris Tyler-Smith, Cornelis A. Albers, Qasim Ayub, Senduran Balasubramaniam, Yuan Chen, Vincenza Colonna, Petr Danecek, Luke Jostins, Thomas M. Keane, Shane McCarthy, Klaudia Walter, Yali Xue, Mark B. Gerstein, Alexej Abyzov, Suganthi Balasubramaniam, Jieming Chen, Declan Clarke, Yao Fu, Arif O. Harmanci, Mike Jin, Donghoon Lee, Jeremy Liu, Ximeng Jasmine Mu, Jing Zhang, Yan Zhang, Yingrui Li, Ruibang Luo, Hongmei Zhu, Can Alkan, Elif Dal, Fatma Kahveci, Gabor T. Marth, Erik P. Garrison, Deniz Kural, Wan-Ping Lee, Alistair N. Ward, Jiantao Wu, Mengyao Zhang, Steven A. McCarroll, Robert E. Handsaker, David M. Altshuler, Eric Banks, Guillermo del Angel, Giulio Genovese, Chris Hartl, Heng Li, Seva Kashin, James C. Nemes, Khalid Shakir, Seungtae C. Yoon, Jayon Lihm, Vladimir Makarov, Jeremiah Degenhardt, Jan O. Korbel, Markus H. Fritz, Sascha Meiers, Benjamin Raeder, Tobias Rausch, Adrian M. Stütz, Paul Flicek, Francesco Paolo Casale, Laura Clarke, Richard E. Smith, Oliver Stegle, Xiangqun Zheng-Bradley, David R. Bentley, Bret Barnes, R. Keira Cheetham, Michael Eberle, Sean Humphray, Scott Kahn, Lisa Murray, Richard Shaw, Eric-Wubbo Lameijer, Mark A. Batzer, Miriam K. Konkel, Jerilyn A. Walker, Li Ding, Ira Hall, Kai Ye, Phil Lacroute, Charles Lee, Eliza Cerveira, Ankit Malhotra, Jaeho Hwang, Dariusz Plewczynski, Kamen Radew, Mallory Romanovitch, Chengsheng Zhang, David W. Craig, Nils Homer, Deanna Church, Chunlin Xiao, Jonathan Sebat, Danny Antaki, Vineet Bafna, Jacob Michaelson, Kenny Ye, Scott E. Devine, Eugene J. Gardner, Gonçalo R. Abecasis, Jeffrey M. Kidd, Ryan E. Mills, Gargi Dayama, Sarah Emery, Goo Jun, Xinghua Shi, Andrew Quitadamo, Gerton Lunter, Gil A. McVean, Ken Chen, Xian Fan, Zechen Chong, Tenghui Chen, David Witherspoon, Jinchuan Xing, Evan E. Eichler, Mark J. Chaisson, Fereydoun Hormozdiari, John Huddleston, Maika Malig, Bradley J. Nelson, Peter H. Sudmant, Nicholas F. Parrish, Ekta Khurana, Matthew E. Hurles, Ben Blackburne, Sarah J. Lindsay, Zemin Ning, Klaudia Walter, Yujun Zhang, Mark B. Gerstein, Alexej Abyzov, Jieming Chen, Declan Clarke, Hugo Lam, Ximeng Jasmine Mu, Cristina Sisú, Jing Zhang, Yan Zhang, Richard A. Gibbs, Fuli Yu, Matthew Bainbridge, Danny Challis, Uday S. Evani, Christie Kovar, James Lu, Donna Muzny, Uma Nagaswamy, Jeffrey G. Reid, Aniko Sabo, Jin Yu, Xiaosen Guo, Wangshen Li, Yingrui Li, Renhua Wu, Gabor T. Marth, Erik P. Garrison, Wen Fung Leong, Alistair N. Ward, Guillermo del Angel, Mark A. DePristo, Stacey B. Gabriel, Namrata Gupta, Chris Hartl, Ryan E. Poplin, Andrew G. Clark, Juan L. Rodriguez-Flores, Paul Flicek, Laura Clarke, Richard E. Smith, Xiangqun Zheng-Bradley,

Daniel G. MacArthur, Elaine R. Mardis, Robert Fulton, Daniel C. Koboldt, Simon Gravel, Carlos D. Bustamante, David W. Craig, Alexis Christoforides, Nils Homer, Tyler Izatt, Stephen T. Sherry, Chunlin Xiao, Emmanouil T. Dermitzakis, Gonçalo R. Abecasis, Hyun Min Kang, Gil A. McVean, Mark B. Gerstein, Suganthi Balasubramanian, Lukas Habegger, Haiyuan Yu, Paul Flicek, Laura Clarke, Fiona Cunningham, Ian Dunham, Daniel Zerbino, Xiangqun Zheng-Bradley, Kasper Lage, Jakob Berg Jaspersen, Heiko Horn, Stephen B. Montgomery, Marianne K. DeGorter, Ekta Khurana, Chris Tyler-Smith, Yuan Chen, Vincenza Colonna, Yali Xue, Mark B. Gerstein, Suganthi Balasubramanian, Yao Fu, Donghoon Kim, Adam Auton, Anthony Marcketta, Rob Desalle, Apurva Narechania, Melissa A. Wilson Sayres, Erik P. Garrison, Robert E. Handsaker, Seva Kashin, Steven A. McCarroll, Juan L. Rodriguez-Flores, Paul Flicek, Laura Clarke, Xiangqun Zheng-Bradley, Yaniv Erlich, Melissa Gymrek, Thomas Frederick Willems, Carlos D. Bustamante, Fernando L. Mendez, G. David Poznik, Peter A. Underhill, Charles Lee, Eliza Cerveira, Ankit Malhotra, Mallory Romanovitch, Chengsheng Zhang, Gonçalo R. Abecasis, Lachlan Coin, Haojing Shao, David Mittelman, Chris Tyler-Smith, Qasim Ayub, Ruby Banerjee, Maria Cerezo, Yuan Chen, Thomas W. Fitzgerald, Sandra Louzada, Andrea Massaia, Shane McCarthy, Graham R. Ritchie, Yali Xue, Fengtang Yang, Richard A. Gibbs, Christie Kovar, Divya Kalra, Walker Hale, Donna Muzny, Jeffrey G. Reid, Jun Wang, Xu Dan, Xiaosen Guo, Guoqing Li, Yingrui Li, Chen Ye, Xiaole Zheng, David M. Altshuler, Paul Flicek, Laura Clarke, Xiangqun Zheng-Bradley, David R. Bentley, Anthony Cox, Sean Humphray, Scott Kahn, Ralf Sudbrak, Marcus W. Albrecht, Matthias Lienhard, David Larson, David W. Craig, Tyler Izatt, Ahmet A. Kurdoglu, Stephen T. Sherry, Chunlin Xiao, David Haussler, Gonçalo R. Abecasis, Gil A. McVean, Richard M. Durbin, Senduran Balasubramanian, Thomas M. Keane, Shane McCarthy, James Stalker, Aravinda Chakravarti, Bartha M. Knoppers, Gonçalo R. Abecasis, Kathleen C. Barnes, Christine Beiswanger, Esteban G. Burchard, Carlos D. Bustamante, Hongyu Cai, Hongzhi Cao, Richard M. Durbin, Norman P. Gerry, Neda Gharani, Richard A. Gibbs, Christopher R. Gignoux, Simon Gravel, Brenna Henn, Danielle Jones, Lynn Jorde, Jane S. Kaye, Alon Keinan, Alastair Kent, Angeliki Kerasidou, Yingrui Li, Rasika Mathias, Gil A. McVean, Andres Moreno-Estrada, Pilar N. Ossorio, Michael Parker, Alissa M. Resch, Charles N. Rotimi, Charmaine D. Royal, Karla Sandoval, Yeyang Su, Ralf Sudbrak, Zhongming Tian, Sarah Tishkoff, Lorraine H. Toji, Chris Tyler-Smith, Marc Via, Yuhong Wang, Huanming Yang, Ling Yang, Jiayong Zhu, Walter Bodmer, Gabriel Bedoya, Andres Ruiz-Linares, Zhiming Cai, Yang Gao, Jiayou Chu, Leena Peltonen, Andres Garcia-Montero, Alberto Orfao, Julie Dutil, Juan C. Martinez-Cruzado, Taras K. Oleksyk, Kathleen C. Barnes, Rasika A. Mathias, Anselm Hennis, Harold Watson, Colin McKenzie, Firdausi Qadri, Regina LaRocque, Pardis C. Sabeti, Jiayong Zhu, Xiaoyan Deng, Pardis C. Sabeti, Danny Asogun, Onikepe

- Folarin, Christian Happi, Omonwunmi Omoniwa, Matt Strelau, Ridhi Tariyal, Mumintou Jallow, Fatoumatta Sisay Joof, Tumani Corrah, Kirk Rockett, Dominic Kwiatkowski, Jaspal Kooner, Trâ'n Tnh Hiê'n, Sarah J. Dunstan, Nguyen Thuy Hang, Richard Fonnier, Robert Garry, Lansana Kanneh, Lina Moses, Pardis C. Sabeti, John Schieffelin, Donald S. Grant, Carla Gallo, Giovanni Poletti, Danish Saleheen, Asif Rasheed, Lisa D. Brooks, Adam L. Felsenfeld, Jean E. McEwen, Yekaterina Vaydylevich, Eric D. Green, Audrey Duncanson, Michael Dunn, Jeffery A. Schloss, Jun Wang, Huanming Yang, Adam Auton, Lisa D. Brooks, Richard M. Durbin, Erik P. Garrison, Hyun Min Kang, Jan O. Korbel, Jonathan L. Marchini, Shane McCarthy, Gil A. McVean, and Gonçalo R. Abecasis. A global reference for human genetic variation. *Nature*, 526(7571) :68–74, oct 2015.
- [133] Jacqueline MacArthur, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, Aoife McMahon, Annalisa Milano, Joannella Morales, Zoe May-Pendlington, Danielle Welter, Tony Burdett, Lucia Hindorff, Paul Flicek, Fiona Cunningham, and Helen Parkinson. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45(D1) :D896–D901, 2017.
- [134] S. V. Bontha, D. G. Maluf, T. F. Mueller, and V. R. Mas. Systems Biology in Kidney Transplantation : The Application of Multi-Omics to a Complex Model. *American Journal of Transplantation*, 17(1) :11–21, 2017.
- [135] G Piétu, R Mariage-Samson, N A Fayein, C Matingou, E Eveno, R Houlgatte, C Decraene, Y Vandenbrouck, F Tahiri, M D Devignes, U Wirkner, W Ansorge, D Cox, T Nagase, N Nomura, and C Auffray. The Genexpress IMAGE knowledge base of the human brain transcriptome : a prototype integrated resource for functional and computational genomics. *Genome research*, 9(2) :195–209, feb 1999.
- [136] Andrew Travers and Georgi Muskhelishvili. DNA structure and function. *FEBS Journal*, 282(12) :2279–2295, jun 2015.
- [137] Ralf Dahm. Discovering DNA : Friedrich Miescher and the early years of nucleic acid research. *Human Genetics*, 122(6) :565–581, jan 2008.
- [138] Oswald T. Avery, Colin M. Macleod, and Maclyn McCarty. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *Journal of Experimental Medicine*, 79(2) :137–158, 1944.
- [139] ROSALIND E. FRANKLIN and R. G. GOSLING. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature*, 172(4369) :156–157, jul 1953.
- [140] A. Klug. Rosalind Franklin and the discovery of the structure of DNA. *Nature*, 219(5156) :808–844, aug 1968.

- 
- [141] J. D. WATSON and F. H. C. CRICK. Molecular Structure of Nucleic Acids : A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356) :737–738, apr 1953.
- [142] Rebecca L. Cann, Mark Stoneking, and Allan C. Wilson. Mitochondrial DNA and human evolution. *Nature*, 325(6099) :31–36, jan 1987.
- [143] T Pulkes and M.G Hanna. Human mitochondrial DNA diseases. *Advanced Drug Delivery Reviews*, 49(1-2) :27–43, jul 2001.
- [144] W Makalowski. The human genome structure and organization. *Acta Biochimica Polonica*, 48(3) :587–598, sep 2001.
- [145] Sergey Nurk, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, Nicolas Altemose, Lev Uralsky, Ariel Gershman, Sergey Aganezov, Savannah J. Hoyt, Mark Diekhans, Glennis A. Logsdon, Michael Alonge, Stylianos E. Antonarakis, Matthew Borchers, Gerard G. Bouffard, Shelise Y. Brooks, Gina V. Caldas, Nae-Chyun Chen, Haoyu Cheng, Chen-Shan Chin, William Chow, Leonardo G. de Lima, Philip C. Dishuck, Richard Durbin, Tatiana Dvorkina, Ian T. Fiddes, Giulio Formenti, Robert S. Fulton, Arkarachai Functammasan, Erik Garrison, Patrick G. S. Grady, Tina A. Graves-Lindsay, Ira M. Hall, Nancy F. Hansen, Gabrielle A. Hartley, Marina Haukness, Kerstin Howe, Michael W. Hunkapiller, Chirag Jain, Miten Jain, Erich D. Jarvis, Peter Kerpedjiev, Melanie Kirsche, Mikhail Kolmogorov, Jonas Korf, Milinn Kremitzki, Heng Li, Valerie V. Maduro, Tobias Marschall, Ann M. McCartney, Jennifer McDaniel, Danny E. Miller, James C. Mullikin, Eugene W. Myers, Nathan D. Olson, Benedict Paten, Paul Peluso, Pavel A. Pevzner, David Porubsky, Tamara Potapova, Evgeny I. Rogaev, Jeffrey A. Rosenfeld, Steven L. Salzberg, Valerie A. Schneider, Fritz J. Sedlazeck, Kishwar Shafin, Colin J. Shew, Alaina Shumate, Ying Sims, Arian F. A. Smit, Daniela C. Soto, Ivan Sović, Jessica M. Storer, Aaron Streets, Beth A. Sullivan, Françoise Thibaud-Nissen, James Torrance, Justin Wagner, Brian P. Walenz, Aaron Wenger, Jonathan M. D. Wood, Chunlin Xiao, Stephanie M. Yan, Alice C. Young, Samantha Zarate, Urvashi Surti, Rajiv C. McCoy, Megan Y. Dennis, Ivan A. Alexandrov, Jennifer L. Gerton, Rachel J. O’Neill, Winston Timp, Justin M. Zook, Michael C. Schatz, Evan E. Eichler, Karen H. Miga, and Adam M. Phillippy. The complete sequence of a human genome. *Science*, 376(6588) :44–53, apr 2022.
- [146] F. H. CRICK. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12 :138–163, 1958.
- [147] Jardins de France. L’analyse PCR : tout le monde en parle, on vous l’explique.
- [148] J Lucas-Lenard and F Lipmann. Protein Biosynthesis. *Annual Review of Biochemistry*, 40(1) :409–448, jun 1971.
- [149] S Clancy and W Brown. Translation : Dna to mrna to. *Nature Education*, 2008.

- [150] Artem E. Men, Peter Wilson, Kirby Siemering, and Susan Forrest. Sanger DNA Sequencing. In *Next Generation Genome Sequencing*, pages 1–11. Wiley-VCH Verlag GmbH Co. KGaA, Weinheim, Germany.
- [151] *Séquençage*, pages 59–65. Springer Paris, Paris, 2006.
- [152] James D. Watson. The human genome project : Past, present, and future. *Science*, 248(4951) :44–49, apr 1990.
- [153] Mark P. Sawicki, Ghassan Samara, Michael Hurwitz, and Edward Passaro. Human Genome Project. *The American Journal of Surgery*, 165(2) :258–264, feb 1993.
- [154] Zahra Abdellah, Alireza Ahmadi, Shahana Ahmed, Matthew Aimable, Rachael Ainscough, Jeff Almeida, Claire Almond, Andrew Ambler, Karen Ambrose, Kerrie Ambrose, Robert Andrew, Daniel Andrews, Neil Andrews, Heather Andrews, Abel Ureta-Vidal, Imre Vastrik, Caleb Webber, Ryan J. Weber, Cara Woodwark, Francis S. Collins, Mark Guyer, Jane Peterson, Adam Felsenfeld, Kris Wetterstrand, and Sandra Kamholz. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011) :931–945, oct 2004.
- [155] Marta Byrska-Bishop, Uday S. Evani, Xuefang Zhao, Anna O. Basile, Haley J. Abel, Allison A. Regier, André Corvelo, Wayne E. Clarke, Rajeeva Musunuri, Kshithija Nagulapalli, Susan Fairley, Alexi Runnels, Lara Winterkorn, Ernesto Lowy, Paul Flicek, Soren Germer, Harrison Brand, Ira M. Hall, Michael E. Talkowski, Giuseppe Narzisi, Michael C. Zody, Evan E. Eichler, Jan O. Korbel, Charles Lee, Tobias Marschall, Scott E. Devine, William T. Harvey, Weichen Zhou, Ryan E. Mills, Tobias Rausch, Sushant Kumar, Can Alkan, Fereydoun Hormozdiari, Zechen Chong, Yu Chen, Xiaofei Yang, Jiadong Lin, Mark B. Gerstein, Ye Kai, Qihui Zhu, Feyza Yilmaz, and Chunlin Xiao. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*, 185(18) :3426–3440.e19, sep 2022.
- [156] Jeffrey D. Wall, Eric W. Stawiski, Aakrosh Ratan, Hie Lim Kim, Changhoon Kim, Ravi Gupta, Kushal Suryamohan, Elena S. Gusareva, Rikky Wenang Purbojati, Tushar Bhargale, Vadim Stepanov, Vladimir Kharkov, Markus S. Schröder, Vedam Ramprasad, Jennifer Tom, Steffen Durinck, Qixin Bei, Jiani Li, Joseph Guillory, Sameer Phalke, Analabha Basu, Jeremy Stinson, Sandhya Nair, Sivasankar Malaichamy, Nidhan K. Biswas, John C. Chambers, Keith C. Cheng, Joyner T. George, Seik Soon Khor, Jong-Il Kim, Belong Cho, Ramesh Menon, Thiramsetti Sattibabu, Akshi Bassi, Manjari Deshmukh, Anjali Verma, Vivek Gopalan, Jong-Yeon Shin, Mahesh Pratapneni, Sam Santhosh, Katsushi Tokunaga, Badrul M. Md-Zain, Kok Gan Chan, Madasamy Parani, Purushothaman Natarajan, Michael Hauser, R. Rand Allingham, Cecilia Santiago-Turla, Arkasubhra Ghosh, Santosh Gopi Krishna Gadde, Christian Fuchsberger, Lukas Forer, Sebastian Schoenherr, Herawati Sudoyo, J. Stephen Lansing, Jonathan Friedlaender, George Koki, Murray P. Cox,

- Michael Hammer, Tatiana Karafet, Khai C. Ang, Syed Q. Mehdi, Venkatesan Radha, Viswanathan Mohan, Partha P. Majumder, Somasekar Seshagiri, Jeong-Sun Seo, Stephan C. Schuster, and Andrew S. Peterson. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, 576(7785) :106–111, dec 2019.
- [157] Ayman Grada and Kate Weinbrecht. Next-Generation Sequencing : Methodology and Application. *Journal of Investigative Dermatology*, 133(8) :1–4, aug 2013.
- [158] Chandra Shekhar Pareek, Rafal Smoczynski, and Andrzej Tretyn. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4) :413–435, nov 2011.
- [159] Tuomo Mantere, Simone Kersten, and Alexander Hoischen. Long-Read Sequencing Emerging in Medical Genetics. *Frontiers in Genetics*, 10, may 2019.
- [160] Samantha Fowler, Rebecca Roush, James Wise, et al. *Concepts of biology*. OpenStax College, Rice University, 2013.
- [161] Jean Gayon. From Mendel to epigenetics : History of genetics. *Comptes Rendus Biologies*, 339(7-8) :225–230, jul 2016.
- [162] William Bateson and Gregor Mendel. *Mendel's principles of heredity*. Courier Corporation, 2013.
- [163] OrphaSchool. Transmission des maladies génétiques.
- [164] Ying-Cai Tan, Jon Blumenfeld, and Hanna Rennert. Autosomal dominant polycystic kidney disease : Genetics, mutations and microRNAs. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1812(10) :1202–1212, oct 2011.
- [165] Jean-Antoine Ribeil, Salima Haccin-Bey-Abina, Emmanuel Payen, Alessandra Magnani, Michaela Semeraro, Elisa Magrin, Laure Caccavelli, Benedicte Neven, Philippe Bourget, Wassim El Nemer, Pablo Bartolucci, Leslie Weber, Hervé Puy, Jean-François Meritet, David Grevent, Yves Beuzard, Stany Chrétien, Thibaud Lefebvre, Robert W. Ross, Olivier Negre, Gabor Veres, Laura Sandler, Sandeep Soni, Mariane de Montalembert, Stéphane Blanche, Philippe Leboulch, and Marina Cavazzana. Gene Therapy in a Patient with Sickle Cell Disease. *New England Journal of Medicine*, 376(9) :848–855, mar 2017.
- [166] Akhlaq A Maan, James Eales, Artur Akbarov, Joshua Rowland, Xiaoguang Xu, Mark A Jobling, Fadi J Charchar, and Maciej Tomaszewski. The Y chromosome : a blueprint for men's health? *European Journal of Human Genetics*, 25(11) :1181–1188, nov 2017.
- [167] Stacy Colaco and Deepak Modi. Genetics of the human Y chromosome and its association with male infertility. *Reproductive Biology and Endocrinology*, 16(1) :14, dec 2018.
- [168] Ingrid A. Holm, Anne E. Nelson, Bruce G. Robinson, Rebecca S. Mason, Deborah J. Marsh, Christopher T. Cowell, and Thomas O. Carpenter. Mutational Analysis and Genotype-Phenotype Correlation of the PHEX Gene in X-Linked Hypophosphatemic Rickets. *The Journal of Clinical Endocrinology Metabolism*, 86(8) :3889–3899, aug 2001.

- [169] Paul Batty and David Lillicrap. Advances and challenges for hemophilia gene therapy. *Human Molecular Genetics*, 28(R1) :R95–R101, oct 2019.
- [170] Cécile LIBIOULLE and Vincent BOURS. Les maladies complexes : l'importance de la génétique. *Revue médicale de Liege*, 67(5-6), 2012.
- [171] François Rousseau and Nathalie Laffamme. Human molecular genetics : From monogenic to polygenic or complex disorders. *Medecine/Sciences*, 19(10) :950–954, oct 2003.
- [172] Arno G. Motulsky. Genetics of complex diseases. *Journal of Zhejiang University SCIENCE B*, 7(2) :167–168, feb 2006.
- [173] Eddie Cano-Gamez and Gosia Trynka. From GWAS to Function : Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*, 11, may 2020.
- [174] Evan E. Eichler, Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M. Leal, Jason H. Moore, and Joseph H. Nadeau. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*, 11(6) :446–450, jun 2010.
- [175] Hans Ellegren and Nicolas Galtier. Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7) :422–433, jul 2016.
- [176] Ravi Sachidanandam, David Weissman, Steven C. Schmidt, Jerzy M. Kakol, Lincoln D. Stein, Gabor Marth, Steve Sherry, James C. Mullikin, Beverley J. Mortimore, David L. Willey, Sarah E. Hunt, Charlotte G. Cole, Penny C. Coggill, Catherine M. Rice, Zemin Ning, Jane Rogers, David R. Bentley, Pui-Yan Kwok, Elaine R. Mardis, Raymond T. Yeh, Brian Schultz, Lisa Cook, Ruth Davenport, Michael Dante, Lucinda Fulton, LaDeana Hillier, Robert H. Waterston, John D. McPherson, Brian Gilman, Stephen Schaffner, William J. Van Etten, David Reich, John Higgins, Mark J. Daly, Brendan Blumenstiel, Jennifer Baldwin, Nicole Stange-Thomann, Michael C. Zody, Lauren Linton, Eric S. Lander, and David Altshuler. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822) :928–933, feb 2001.
- [177] Kelly A. Frazer, Dennis G. Ballinger, David R. Cox, David A. Hinds, Laura L. Stuve, Richard A. Gibbs, John W. Belmont, Andrew Boudreau, Paul Hardenbol, Suzanne M. Leal, Shiran Pasternak, David A. Wheeler, Thomas D. Willis, Fuli Yu, Huanming Yang, Changqing Zeng, Yang Gao, Haoran Hu, Weitao Hu, Chaohua Li, Wei Lin, Siqi Liu, Hao Pan, Xiaoli Tang, Jian Wang, Wei Wang, Jun Yu, Bo Zhang, Qingrun Zhang, Hongbin Zhao, Hui Zhao, Jun Zhou, Stacey B. Gabriel, Rachel Barry, Brendan Blumenstiel, Amy Cargano, Matthew Defelice, Maura Faggart, Mary Goyette, Supriya Gupta, Jamie Moore, Huy Nguyen, Robert C. Onofrio, Melissa Parkin, Jessica Roy, Erich Stahl, Ellen Winchester, Liuda Ziaugra, David Altshuler, Yan Shen, Zhijian Yao, Wei Huang, Xun Chu, Yungang He, Li Jin, Yangfan Liu, Yayun Shen, Weiwei Sun, Haifeng Wang, Yi Wang,

Ying Wang, Xiaoyan Xiong, Liang Xu, Mary M.Y. Waye, Stephen K.W. Tsui, Hong Xue, J. Tze Fei Wong, Luana M. Galver, Jian Bing Fan, Kevin Gunderson, Sarah S. Murray, Arnold R. Oliphant, Mark S. Chee, Alexandre Montpetit, Fanny Chagnon, Vincent Ferretti, Martin Leboeuf, Jean François Olivier, Michael S. Phillips, Stéphanie Roumy, Clémentine Sallée, Andrei Verner, Thomas J. Hudson, Pui Yan Kwok, Dongmei Cai, Daniel C. Koboldt, Raymond D. Miller, Ludmila Pawlikowska, Patricia Taillon-Miller, Ming Xiao, Lap Chee Tsui, William Mak, Qiang Song You, Paul K.H. Tam, Yusuke Nakamura, Takahisa Kawaguchi, Takuya Kitamoto, Takashi Morizono, Atsushi Nagashima, Yozo Ohnishi, Akihiro Sekine, Toshihiro Tanaka, Tatsuhiko Tsunoda, Panos Deloukas, Christine P. Bird, Marcos Delgado, Emmanouil T. Dermitzakis, Rhian Gwilliam, Sarah Hunt, Jonathan Morrison, Don Powell, Barbara E. Stranger, Pamela Whittaker, David R. Bentley, Mark J. Daly, Paul I.W. De Bakker, Jeff Barrett, Yves R. Chretien, Julian Maller, Steve McCarroll, Nick Patterson, Itsik Pe'Er, Alkes Price, Shaun Purcell, Daniel J. Richter, Paridis Sabeti, Richa Saxena, Stephen F. Schaffner, Pak C. Sham, Patrick Varilly, Lincoln D. Stein, Lalitha Krishnan, Albert Vernon Smith, Marcela K. Tello-Ruiz, Gudmundur A. Thorisson, Aravinda Chakravarti, Peter E. Chen, David J. Cutler, Carl S. Kashuk, Shin Lin, Gonçalo R. Abecasis, Weihua Guan, Yun Li, Heather M. Munro, Zhaohui Steve Qin, Daryl J. Thomas, Gilean McVean, Adam Auton, Leonardo Bottolo, Niall Cardin, Susana Eyheramendy, Colin Freeman, Jonathan Marchini, Simon Myers, Chris Spencer, Matthew Stephens, Peter Donnelly, Lon R. Cardon, Geraldine Clarke, David M. Evans, Andrew P. Morris, Bruce S. Weir, Todd A. Johnson, James C. Mullikin, Stephen T. Sherry, Michael Feolo, Andrew Skol, Houcan Zhang, Ichiro Matsuda, Yoshimitsu Fukushima, Darryl R. MacEr, Eiko Suda, Charles N. Rotimi, Clement A. Adebamowo, Ike Ajayi, Toyin Aniagwu, Patricia A. Marshall, Chibuzor Nkwodimmah, Charmaine D.M. Royal, Mark F. Leppert, Missy Dixon, Andy Peiffer, Renzong Qiu, Alastair Kent, Kazuto Kato, Norio Niikawa, Isaac F. Adewole, Bartha M. Knoppers, Morris W. Foster, Ellen Wright Clayton, Jessica Watkin, Donna Muzny, Lynne Nazareth, Erica Sodergren, George M. Weinstock, Imtaz Yakub, Bruce W. Birren, Richard K. Wilson, Lucinda L. Fulton, Jane Rogers, John Burton, Nigel P. Carter, Christopher M. Clee, Mark Griffiths, Matthew C. Jones, Kirsten McLay, Robert W. Plumb, Mark T. Ross, Sarah K. Sims, David L. Willey, Zhu Chen, Hua Han, Le Kang, Martin Godbout, John C. Wallenburg, Paul L'Archevêque, Guy Bellemare, Koji Saeki, Hongguang Wang, Daochang An, Hongbo Fu, Qing Li, Zhen Wang, Renwu Wang, Arthur L. Holden, Lisa D. Brooks, Jean E. McEwen, Mark S. Guyer, Vivian Ota Wang, Jane L. Peterson, Michael Shi, Jack Spiegel, Lawrence M. Sung, Lynn F. Zacharia, Francis S. Collins, Karen Kennedy, Ruth Jamieson, and John Stewart. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164) :851–861, oct 2007.



- [178] Somaia Ismail and Mona Essawi. Genetic polymorphism studies in humans. *Middle East Journal of Medical Genetics*, 1(2) :57–63, jul 2012.
- [179] C.A Hartman. What are single nucleotide polymorphisms (SNPs), 2016.
- [180] S. T. Sherry. dbSNP : the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1) :308–311, jan 2001.
- [181] Stephan J. Sanders and Christopher E. Mason. The Newly Emerging View of the Genome. In *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*, pages 3–26. Elsevier, 2016.
- [182] J. M. Mullaney, R. E. Mills, W. S. Pittard, and S. E. Devine. Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics*, 19(R2) :R131–R136, oct 2010.
- [183] Maoxuan Lin, Sarah Whitmire, Jing Chen, Alvin Farrel, Xinghua Shi, and Jun-tao Guo. Effects of short indels on protein structure and function in human genomes. *Scientific Reports*, 7(1) :9313, dec 2017.
- [184] Sebastian Boegel, John C. Castle, Julia Kodysh, Timothy O'Donnell, and Alex Rubinsteyn. Bioinformatic methods for cancer neoantigen prediction. pages 25–60. 2019.
- [185] A John Iafrate, Lars Feuk, Miguel N Rivera, Marc L Listewnik, Patricia K Donahoe, Ying Qi, Stephen W Scherer, and Charles Lee. Detection of large-scale variation in the human genome. *Nature Genetics*, 36(9) :949–951, sep 2004.
- [186] Mehdi Zarrei, Jeffrey R. MacDonald, Daniele Merico, and Stephen W. Scherer. A copy number variation map of the human genome. *Nature Reviews Genetics*, 16(3) :172–183, mar 2015.
- [187] P. Stankiewicz and J.R. Lupski. The Genomic Basis of Disease, Mechanisms and Assays for Genomic Disorders. In *Genome and Disease*, pages 1–16. KARGER, Basel, 2006.
- [188] Jeffrey R. MacDonald, Robert Ziman, Ryan K. C. Yuen, Lars Feuk, and Stephen W. Scherer. The Database of Genomic Variants : a curated collection of structural variation in the human genome. *Nucleic Acids Research*, 42(D1) :D986–D992, jan 2014.
- [189] Curt Stern. The Hardy-Weinberg law. *Science*, 97(2510) :137–138, feb 1943.
- [190] Oliver Mayo. A Century of Hardy–Weinberg Equilibrium. *Twin Research and Human Genetics*, 11(3) :249–256, jun 2008.
- [191] Thomas Hoffman. *Association entre polymorphisme de gènes de l'immunité et événements cliniques post-transplantation rénale*. Sciences de la vie et de la santé, Université de tours, 2009.
- [192] F.Vasseur. Les maladies génétiques "complexes". Technical report, Collège National des Enseignants et Praticiens de Génétique Médicale, 2012.

- [193] Duska Dragun, Dominik N. Müller, Jan Hinrich Bräsen, Lutz Fritsche, Melina Nieminen-Kelhä, Ralf Dechend, Ulrich Kintscher, Birgit Rudolph, Johan Hoebeke, Diana Eckert, Istvan Mazak, Ralph Plehm, Constanze Schönemann, Thomas Unger, Klemens Budde, Hans-Hellmut Neumayer, Friedrich C. Luft, and Gerd Wallukat. Angiotensin II Type 1–Receptor Activating Antibodies in Renal-Allograft Rejection. *New England Journal of Medicine*, 352(6) :558–569, feb 2005.
- [194] Yizhou Zou, Peter Stastny, Caner Süsal, Bernd Döhler, and Gerhard Opelz. Antibodies against MICA Antigens and Kidney-Transplant Rejection. *New England Journal of Medicine*, 357(13) :1293–1300, sep 2007.
- [195] Vivian Tam and Nikunj Patel. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, pages 467–484, 2019.
- [196] Hunna J. Watson, Zeynep Yilmaz, Laura M. Thornton, Christopher Hübel, Jonathan R. I. Coleman, Hélène A. Gaspar, Julien Bryois, Anke Hinney, Virpi M. Leppä, Manuel Mattheisen, Sarah E. Medland, Stephan Ripke, Shuyang Yao, Paola Giusti-Rodríguez, Ken B. Hanscombe, Kirstin L. Purves, Roger A. H. Adan, Lars Alfredsson, Tetsuya Ando, Ole A. Andreassen, Jessica H. Baker, Wade H. Berrettini, Ilka Boehm, Claudette Boni, Vesna Boraska Perica, Katharina Buehren, Roland Burghardt, Matteo Cassina, Sven Cichon, Maurizio Clementi, Roger D. Cone, Philippe Courtet, Scott Crow, James J. Crowley, Unna N. Danner, Oliver S. P. Davis, Martina de Zwaan, George Dedoussis, Daniela Degortes, Janiece E. DeSocio, Danielle M. Dick, Dimitris Dikeos, Christian Dina, Monika Dmitrzak-Weglarz, Elisa Docampo, Laramie E. Duncan, Karin Egberts, Stefan Ehrlich, Geòrgia Escaramís, Tõnu Esko, Xavier Estivill, Anne Farmer, Angela Favaro, Fernando Fernández-Aranda, Manfred M. Fichter, Krista Fischer, Manuel Föcker, Lenka Foretova, Andreas J. Forstner, Monica Forzan, Christopher S. Franklin, Steven Gallinger, Ina Giegling, Johanna Giuranna, Fragiskos Gonidakis, Philip Gorwood, Monica Gratacos Mayora, Sébastien Guillaume, Yiran Guo, Hakon Hakonarson, Konstantinos Hatzikotoulas, Joanna Hauser, Johannes Hebebrand, Sietske G. Helder, Stefan Herms, Beate Herpertz-Dahlmann, Wolfgang Herzog, Laura M. Huckins, James I. Hudson, Hartmut Imgart, Hidetoshi Inoko, Vladimir Janout, Susana Jiménez-Murcia, Antonio Julià, Gursharan Kalsi, Deborah Kaminská, Jaakko Kaprio, Leila Karhunen, Andreas Karwautz, Martien J. H. Kas, James L. Kennedy, Anna Keski-Rahkonen, Kirsty Kiezebrink, Youl-Ri Kim, Lars Klareskog, Kelly L. Klump, Gun Peggy S. Knudsen, Maria C. La Via, Stephanie Le Hellard, Robert D. Levitan, Dong Li, Lisa Lilenfeld, Bochao Danae Lin, Jolanta Lissowska, Jurjen Luykx, Pierre J. Magistretti, Mario Maj, Katrin Mannik, Sara Marsal, Christian R. Marshall, Morten Mattingsdal, Sara McDevitt, Peter McGuffin, Andres Metspalu, Ingrid Meulenbelt, Nadia Micali, Karen Mitchell, Alessio Maria Monteleone, Palmiero Monteleone, Melissa A. Munn-Chernoff, Benedetta Nacmias, Marie Navratilova, Ioanna Ntalla, Julie K. O’Toole,

- Roel A. Ophoff, Leonid Padyukov, Aarno Palotie, Jacques Pantel, Hana Papezova, Dalila Pinto, Raquel Rabionet, Anu Raevuori, Nicolas Ramoz, Ted Reichborn-Kjennerud, Valdo Ricca, Samuli Ripatti, Franziska Ritschel, Marion Roberts, Alessandro Rotondo, Dan Rujescu, Filip Rybakowski, Paolo Santonastaso, André Scherag, Stephen W. Scherer, Ulrike Schmidt, Nicholas J. Schork, Alexandra Schosser, Jochen Seitz, Lenka Slachtova, P. Eline Slagboom, Margarita C. T. Slof-Op 't Landt, Agnieszka Slopian, Sandro Sorbi, Beata Świątkowska, Jin P. Szatkiewicz, Ioanna Tachmazidou, Elena Tenconi, Alfonso Tortorella, Federica Tozzi, Janet Treasure, Artemis Tsitsika, Marta Tyszkiewicz-Nwafor, Konstantinos Tziouvas, Annemarie A. van Elburg, Eric F. van Furth, Gudrun Wagner, Esther Walton, Elisabeth Widen, Eleftheria Zeggini, Stephanie Zerwas, Stephan Zipfel, Andrew W. Bergen, Joseph M. Boden, Harry Brandt, Steven Crawford, Katherine A. Halmi, L. John Horwood, Craig Johnson, Allan S. Kaplan, Walter H. Kaye, James E. Mitchell, Catherine M. Olsen, John F. Pearson, Nancy L. Pedersen, Michael Strober, Thomas Werge, David C. Whiteman, D. Blake Woodside, Garret D. Stuber, Scott Gordon, Jakob Grove, Anjali K. Henders, Anders Juréus, Katherine M. Kirk, Janne T. Larsen, Richard Parker, Liselotte Petersen, Jennifer Jordan, Martin Kennedy, Grant W. Montgomery, Tracey D. Wade, Andreas Birgegård, Paul Lichtenstein, Claes Noring, Mikael Landén, Nicholas G. Martin, Preben Bo Mortensen, Patrick F. Sullivan, Gerome Breen, and Cynthia M. Bulik. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature Genetics*, 51(8) :1207–1214, aug 2019.
- [197] Craig L Hyde, Michael W Nagle, Chao Tian, Xing Chen, Sara A Paciga, Jens R Wendland, Joyce Y Tung, David A Hinds, Roy H Perlis, and Ashley R Winslow. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, 48(9) :1031–1036, sep 2016.
- [198] Amit Sud, Ben Kinnersley, and Richard S. Houlston. Genome-wide association studies of cancer : current insights and future perspectives. *Nature Reviews Cancer*, 17(11) :692–704, nov 2017.
- [199] Wei Zhao, Asif Rasheed, Emmi Tikkanen, Jung-Jin Lee, Adam S Butterworth, Joanna M M Howson, Themistocles L Assimes, Rajiv Chowdhury, Marju Orho-Melander, Scott Damrauer, Aeron Small, Senay Asma, Minako Imamura, Toshimasa Yamauch, John C Chambers, Peng Chen, Bishwa R Sapkota, Nabi Shah, Sehrish Jabeen, Praveen Surendran, Yingchang Lu, Weihua Zhang, Atif Imran, Shahid Abbas, Faisal Majeed, Kevin Trindade, Nadeem Qamar, Nadeem Hayyat Mallick, Zia Yaqoob, Tahir Saghir, Syed Nadeem Hasan Rizvi, Anis Memon, Syed Zahed Rasheed, Fazal-ur-Rehman Memon, Khalid Mehmood, Naveeduddin Ahmed, Irshad Hussain Qureshi, Tanveer us Salam, Wasim Iqbal, Uzma Malik, Narinder Mehra, Jane Z Kuo, Wayne H-H Sheu, Xiuqing Guo, Chao A Hsiung, Jyh-Ming J Juang, Kent D Taylor, Yi-Jen Hung, Wen-Jane Lee, Thomas Quer-

- termous, I-Te Lee, Chih-Cheng Hsu, Erwin P Bottinger, Sarju Ralhan, Yik Ying Teo, Tzung-Dau Wang, Dewan S Alam, Emanuele Di Angelantonio, Steve Epstein, Sune F Nielsen, Børge G Nordestgaard, Anne Tybjaerg-Hansen, Robin Young, Marianne Benn, Ruth Frikke-Schmidt, Pia R Kamstrup, J Wouter Jukema, Naveed Sattar, Roelof Smit, Ren-Hua Chung, Kae-Woei Liang, Sonia Anand, Dharambir K Sanghera, Samuli Ripatti, Ruth J F Loos, Jaspal S Kooner, E Shyong Tai, Jerome I Rotter, Yii-Der Ida Chen, Philippe Frossard, Shiro Maeda, Takashi Kadowaki, Muredach Reilly, Guillaume Pare, Olle Melander, Veikko Salomaa, Daniel J Rader, John Danesh, Benjamin F Voight, and Danish Saleheen. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nature Genetics*, 49(10) :1450–1457, oct 2017.
- [200] Katrina M de Lange, Loukas Moutsianas, James C Lee, Christopher A Lamb, Yang Luo, Nicholas A Kennedy, Luke Jostins, Daniel L Rice, Javier Gutierrez-Achury, Sun-Gou Ji, Graham Heap, Elaine R Nimmo, Cathryn Edwards, Paul Henderson, Craig Mowat, Jeremy Sanderson, Jack Satsangi, Alison Simmons, David C Wilson, Mark Tremelling, Ailsa Hart, Christopher G Mathew, William G Newman, Miles Parkes, Charlie W Lees, Holm Uhlig, Chris Hawkey, Natalie J Prescott, Tariq Ahmad, John C Mansfield, Carl A Anderson, and Jeffrey C Barrett. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics*, 49(2) :256–261, feb 2017.
- [201] Philip R. Jansen, Kyoko Watanabe, Sven Stringer, Nathan Skene, Julien Bryois, Anke R. Hammerschlag, Christiaan A. de Leeuw, Jeroen S. Benjamins, Ana B. Muñoz-Manchado, Mats Nagel, Jeanne E. Savage, Henning Tiemeier, Tonya White, Joyce Y. Tung, David A. Hinds, Vladimir Vacic, Xin Wang, Patrick F. Sullivan, Sophie van der Sluis, Tinca J. C. Polderman, August B. Smit, Jens Hjerling-Leffler, Eus J. W. Van Someren, and Danielle Posthuma. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics*, 51(3) :394–403, mar 2019.
- [202] Loic Yengo, Julia Sidorenko, Kathryn E Kemper, Zhili Zheng, Andrew R Wood, Michael N Weedon, Timothy M Frayling, Joel Hirschhorn, Jian Yang, and Peter M Visscher. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry. *Human Molecular Genetics*, 27(20) :3641–3649, oct 2018.
- [203] Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145) :661–678, jun 2007.
- [204] Aswin Sekar, Allison R. Bialas, Heather de Rivera, Avery Davis, Timothy R. Hammond, Nolan Kamitaki, Katherine Tooley, Jessy Presumey, Matthew Baum, Vanessa Van Doren, Giulio Genovese, Samuel A. Rose, Robert E. Handsaker, Mark J. Daly, Michael C. Carroll,

- Beth Stevens, and Steven A. McCarroll. Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589) :177–183, feb 2016.
- [205] Hreinn Stefansson, Roel A. Ophoff, Stacy Steinberg, Ole A. Andreassen, Sven Cichon, Dan Rujescu, Thomas Werge, Olli P. H. Pietiläinen, Ole Mors, Preben B. Mortensen, Engilbert Sigurdsson, Omar Gustafsson, Mette Nyegaard, Annamari Tuulio-Henriksson, Andres Ingason, Thomas Hansen, Jaana Suvisaari, Jouko Lonnqvist, Tiina Paunio, Anders D. Børglum, Annette Hartmann, Anders Fink-Jensen, Merete Nordentoft, David Hougaard, Bent Norgaard-Pedersen, Yvonne Böttcher, Jes Olesen, René Breuer, Hans-Jürgen Möller, Ina Giegling, Henrik B. Rasmussen, Sally Timm, Manuel Mattheisen, István Bitter, János M. Réthelyi, Brynja B. Magnusdottir, Thordur Sigmundsson, Pall Olason, Gisli Masson, Jeffrey R. Gulcher, Magnus Haraldsson, Ragnheidur Fossdal, Thorgeir E. Thorgeirsson, Unnur Thorsteinsdottir, Mirella Ruggeri, Sarah Tosato, Barbara Franke, Eric Strengman, Lambertus A. Kiemeney, Ingrid Melle, Srdjan Djurovic, Lilia Abramova, Vasily Kaleda, Julio Sanjuan, Rosa de Frutos, Elvira Bramon, Evangelos Vassos, Gillian Fraser, Ulrich Ettinger, Marco Picchioni, Nicholas Walker, Timi Touloupoulou, Anna C. Need, Dongliang Ge, Joeng Lim Yoon, Kevin V. Shianna, Nelson B. Freimer, Rita M. Cantor, Robin Murray, Augustine Kong, Vera Golimbet, Angel Carracedo, Celso Arango, Javier Costas, Erik G. Jönsson, Lars Terenius, Ingrid Agartz, Hannes Petursson, Markus M. Nöthen, Marcella Rietschel, Paul M. Matthews, Pierandrea Muglia, Leena Peltonen, David St Clair, David B. Goldstein, Kari Stefansson, and David A. Collier. Common variants conferring risk of schizophrenia. *Nature*, 460(7256) :744–747, aug 2009.
- [206] Barbara Murphy Madhav C. Menon, Karen L. Keung and Philip J. O’Connell. THE USE OF GENOMICS AND PATHWAY ANALYSIS IN OUR UNDERSTANDING AND PREDICTION OF CLINICAL RENAL TRANSPLANT INJURY. 100(7) :1405–1414, 2016.
- [207] Joshua Y.C. Yang and Minnie M. Sarwal. Transplant genetics and genomics. *Nature Reviews Genetics*, 18(5) :309–326, 2017.
- [208] L. Ghisdal, C. Baron, Y. Lebranchu, O. Viklický, A. Konarikova, M. Naesens, D. Kuypers, M. Dinic, E. Alamartine, G. Touchard, T. Antoine, M. Essig, J. P. Rerolle, P. Merville, J. L. Taupin, Y. Le Meur, A. Grall-Jezequel, F. Glowacki, C. Noël, C. Legendre, D. Anglicheau, N. Broeders, W. Coppieters, E. Docampo, M. Georges, Z. Ajarchouh, A. Massart, J. Racapé, D. Abramowicz, and M. Abramowicz. Genome-Wide Association Study of Acute Renal Graft Rejection. *American Journal of Transplantation*, 17(1) :201–209, jan 2017.
- [209] Tasneem Motiwala, Jharna Datta, Huban Kutay, Satavisha Roy, and Samson T. Jacob. Lyn kinase and ZAP70 are substrates of PTPROt in B-cells : Lyn inactivation by PTPROt

- sensitizes leukemia cells to VEGF-R inhibitor pazopanib. *Journal of Cellular Biochemistry*, 110(4) :846–856, apr 2010.
- [210] Francesca Finetti, Silvia Rossi Paccani, Joel Rosenbaum, and Cosima T. Baldari. Intraflagellar transport : a new player at the immune synapse. *Trends in Immunology*, 32(4) :139–145, apr 2011.
- [211] Maria P. Hernandez-Fuentes, Christopher Franklin, Irene Rebollo-Mesa, Jennifer Molton, Florence Delaney, Esperanza Perucha, Caragh Stapleton, Richard Borrows, Catherine Byrne, Gianpiero Cavalleri, Brendan Clarke, Menna Clatworthy, John Feehally, Susan Fuggle, Sarah A. Gagliano, Sian Griffin, Abdul Hammad, Robert Higgins, Alan Jardine, Mary Keogan, Timothy Leach, Iain MacPhee, Patrick B. Mark, James Marsh, Peter Maxwell, William McKane, Adam McLean, Charles Newstead, Titus Augustine, Paul Phelan, Steve Powis, Peter Rowe, Neil Sheerin, Ellen Solomon, Henry Stephens, Raj Thuraiingham, Richard Trembath, Peter Topham, Robert Vaughan, Steven H. Sacks, Peter Conlon, Gerhard Opelz, Nicole Soranzo, Michael E. Weale, and Graham M. Lord. Long- and short-term outcomes in renal allografts with deceased donors : A large recipient and donor genome-wide association study. *American Journal of Transplantation*, 18(6) :1370–1379, jun 2018.
- [212] Agence de la Biomédecine. RAMS 2019 ORGANES - greffe rénale. Technical report, 2019.
- [213] Robert P. O’Brien, Paul J. Phelan, Judith Conroy, Patrick O’Kelly, Andrew Green, Mary Keogan, Derek O’Neill, Susan Jennings, Carol Traynor, Jillian Casey, Mark McCormack, Ronan Conroy, Anthony Chubb, Sean Ennis, Denis C. Shields, Gianpiero L. Cavalleri, and Peter J. Conlon. A genome-wide association study of recipient genotype and medium-term kidney allograft function. *Clinical Transplantation*, 27(3) :379–387, may 2013.
- [214] H. K. Pihlstrøm, G. Mjøen, S. Mucha, G. Haraldsen, A. Franke, A. Jardine, B. Fellström, H. Holdaas, and E. Melum. Single Nucleotide Polymorphisms and Long-Term Clinical Outcome in Renal Transplant Patients : A Validation Study. *American Journal of Transplantation*, 17(2) :528–533, feb 2017.
- [215] Laurent Mesnard, Thangamani Muthukumar, Maren Burbach, Carol Li, Huimin Shang, Darshana Dadhania, John R. Lee, Vijay K. Sharma, Jenny Xiang, Caroline Suberbielle, Maryvonnick Carmagnat, Nacera Ouali, Eric Rondeau, John J. Friedewald, Michael M. Abecassis, Manikkam Suthanthiran, and Fabien Campagne. Exome Sequencing and Prediction of Long-Term Kidney Allograft Function. *PLoS Computational Biology*, 12(9) :1–15, 2016.
- [216] Roman Reindl-Schwaighofer, Andreas Heinzl, Lorenzo Signorini, Olivier Thaunat, and Rainer Oberbauer. Mechanisms underlying human genetic diversity : consequence for antigrift antibody responses. *Transplant International*, 31 :239–250, 2018.

- [217] Nicholas J. Steers, Yifu Li, Zahida Drace, Justin A. D’Addario, Clara Fischman, Lili Liu, Katherine Xu, Young-Ji Na, Y. Dana Neugut, Jun Y. Zhang, Roel Sterken, Olivia Balderes, Drew Bradbury, Nilgun Ozturk, Fatih Ozay, Sanya Goswami, Karla Mehl, Jaclyn Wold, Fatima Z. Jelloul, Mersedeh Rohanizadegan, Christopher E. Gillies, Elena-Rodica M. Vasilescu, George Vlad, Yi-An Ko, Sumit Mohan, Jai Radhakrishnan, David J. Cohen, Lloyd E. Ratner, Francesco Scolari, Katalin Susztak, Matthew G. Sampson, Silvia Deaglio, Yasar Caliskan, Jonathan Barasch, Aisling E. Courtney, Alexander P. Maxwell, Amy J. McKnight, Iuliana Ionita-Laza, Stephan J.L. Bakker, Harold Snieder, Martin H. de Borst, Vivette D’Agati, Antonio Amoroso, Ali G. Gharavi, and Krzysztof Kiryluk. Genomic Mismatch at LIMS1 Locus and Kidney Allograft Rejection . *New England Journal of Medicine*, 380(20) :1918–1928, 2019.
- [218] Konrad J. Karczewski and Michael P. Snyder. Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5) :299–310, 2018.
- [219] Katsuyuki Yugi, Hiroyuki Kubota, Atsushi Hatano, and Shinya Kuroda. Trans-Omics : How To Reconstruct Biochemical Networks Across Multiple ‘Omic’ Layers. *Trends in Biotechnology*, 34(4) :276–290, apr 2016.
- [220] Gabriela V. Cohen Freue, Mayu Sasaki, Anna Meredith, Oliver P. Günther, Axel Bergman, Mandeep Takhar, Alice Mui, Robert F. Balshaw, Raymond T. Ng, Nina Opushneva, Zsuzsanna Hollander, Guiyun Li, Christoph H. Borchers, Janet Wilson-McManus, Bruce M. McManus, Paul A. Keown, and W. Robert McMaster. Proteomic Signatures in Plasma during Early Acute Renal Allograft Rejection. *Molecular Cellular Proteomics*, 9(9) :1954–1967, sep 2010.
- [221] Brendan J. Keating, Alexandre C. Pereira, Michael Snyder, and Brian D. Piening. Applying genomics in heart transplantation. *Transplant International*, 31(3) :278–290, 2018.
- [222] Christophe Masset, Clarisse Kerleau, Claire Garandeau, Simon Ville, Diego Cantarovich, Maryvonne Hourmant, Delphine Kervella, Aurélie Houzet, Cécile Guillot-Gueguen, Irène Guihard, Magali Giral, Jacques Dantal, and Gilles Blancho. A third injection of the BNT162b2 mRNA COVID-19 vaccine in kidney transplant recipients improves the humoral immune response. *Kidney International*, 100(5) :1132–1135, nov 2021.
- [223] Y. Foucher, M. Lorent, L. Albano, S. Roux, V. Pernin, M. Le Quintrec, C. Legendre, F. Buron, E. Morelon, S. Girerd, M. Ladrière, D. Glotz, C. Lefaucher, C. Kerleau, J. Dantal, J. Branchereau, and M. Giral. Renal transplantation outcomes in obese patients : a French cohort-based study. *BMC Nephrology*, 22(1) :79, dec 2021.
- [224] Amaury Dujardin, Mélanie Chesneau, Florian Dubois, Richard Danger, Linh Bui, Clarisse Kerleau, Pierrick Guérif, Sophie Brouard, and Jacques Dantal. Clinical and immunological

- follow-up of very long-term kidney transplant recipients treated with calcineurin inhibitors indicates dual phenotypes. *Kidney International*, 99(6) :1418–1429, jun 2021.
- [225] Pui-Yan Kwok. Methods for Genotyping Single Nucleotide Polymorphisms. *Annual Review of Genomics and Human Genetics*, 2(1) :235–258, sep 2001.
- [226] Regina O Obe author. *PostgreSQL : up and running : a practical guide to the advanced open source database / Regina O. Obe and Leo S. Hsu*. O’Reilly Media, Inc., 2015.
- [227] Geraldine M Clarke, Carl A Anderson, Fredrik H Pettersson, Lon R Cardon, Andrew P Morris, and Krina T Zondervan. Basic statistical analysis in genetic case-control studies. *Nature Protocols*, 6(2) :121–133, feb 2011.
- [228] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3) :559–575, sep 2007.
- [229] Eric L. Stevens, Greg Heckenberg, Elisha D. O. Roberson, Joseph D. Baugher, Thomas J. Downey, and Jonathan Pevsner. Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State. *PLoS Genetics*, 7(9) :e1002287, sep 2011.
- [230] Evan L. Sticca, Gillian M. Belbin, and Christopher R. Gignoux. Current Developments in Detection of Identity-by-Descent Methods and Applications. *Frontiers in Genetics*, 12, sep 2021.
- [231] Elizabeth A Thompson. Identity by Descent : Variation in Meiosis, Across Genomes, and in Populations. *Genetics*, 194(2) :301–326, jun 2013.
- [232] Diogo Meyer and Kelly Nunes. HLA imputation, what is it good for ? *Human Immunology*, 78(3) :239–241, 2017.
- [233] X Zheng, J Shen, C Cox, J C Wakefield, M G Ehm, M R Nelson, and B S Weir. HI-BAG—HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal*, 14(2) :192–200, apr 2014.
- [234] Markus Ringnér. What is principal component analysis ? *Nature Biotechnology*, 26(3) :303–304, mar 2008.
- [235] John Novembre, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens, and Carlos D. Bustamante. Genes mirror geography within Europe. *Nature*, 456(7218) :98–101, nov 2008.
- [236] D Shriner. Investigating population stratification and admixture using eigenanalysis of dense genotypes. *Heredity*, 107(5) :413–420, nov 2011.



- [237] Bryan N. Howie, Peter Donnelly, and Jonathan Marchini. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, 5(6) :e1000529, jun 2009.
- [238] P. I.W. de Bakker, M. A.R. Ferreira, X. Jia, B. M. Neale, S. Raychaudhuri, and B. F. Voight. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human Molecular Genetics*, 17(R2) :R122–R128, oct 2008.
- [239] Brian L. Browning, Ying Zhou, and Sharon R. Browning. A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, 103(3) :338–348, sep 2018.
- [240] L. Rabiner and B. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1) :4–16, 1986.
- [241] Kelly Nunes, Xiuwen Zheng, Margareth Torres, Maria Elisa Moraes, Bruno Z. Piovezan, Gerlandia N. Pontes, Lilian Kimura, Juliana E.P. Carnavalli, Regina C. Mingroni Netto, and Diogo Meyer. HLA imputation in an admixed population : An assessment of the 1000 Genomes data as a training set. *Human Immunology*, 2016.
- [242] Venceslas Douillard, Erick C. Castelli, Steven J. Mack, Jill A. Hollenbach, Pierre Antoine Gourraud, Nicolas Vince, and Sophie Limou. Approaching Genetics Through the MHC Lens : Tools and Methods for HLA Research. *Frontiers in Genetics*, 12(December) :1–13, 2021.
- [243] Xiaoming Jia, Buhm Han, Suna Onengut-Gumuscu, Wei-Min Chen, Patrick J. Concannon, Stephen S. Rich, Soumya Raychaudhuri, and Paul I.W. de Bakker. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS ONE*, 8(6) :e64683, jun 2013.
- [244] Alexander T. Dilthey, Loukas Moutsianas, Stephen Leslie, and Gil McVean. HLA\*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics*, 27(7) :968–972, apr 2011.
- [245] Tatsuhiko Naito, Ken Suzuki, Jun Hirata, Yoichiro Kamatani, Koichi Matsuda, Tatsushi Toda, and Yukinori Okada. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nature Communications*, 12(1) :1639, dec 2021.
- [246] Seungho Cook, Wanson Choi, Hyunjoon Lim, Yang Luo, Kunhee Kim, Xiaoming Jia, Soumya Raychaudhuri, and Buhm Han. Accurate imputation of human leukocyte antigens with CookHLA. *Nature Communications*, 12(1) :1264, dec 2021.
- [247] Allan Motyer, Damjan Vukcevic, Alexander Dilthey, Peter Donnelly, Gil McVean, and Stephen Leslie. Practical Use of Methods for Imputation of HLA Alleles from SNP Genotype Data. *bioRxiv*, page 091009, 2016.
- [248] Gülden Kaya Uyamık and Neşe Güler. A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 106 :234–240, dec 2013.

- [249] John Ludbrook. Linear regression analysis for comparing two measurers or methods of measurement : But which regression ? *Clinical and Experimental Pharmacology and Physiology*, 37(7) :692–699, mar 2010.
- [250] Rohan L. Fernando and Dorian Garrick. Bayesian Methods Applied to GWAS. pages 237–274. 2013.
- [251] Elise Whitley and Jonathan Ball. Statistics review 5 : Comparison of means. *Critical care (London, England)*, 6(5) :424–8, oct 2002.
- [252] Duncan J Murdoch, Yu-Ling Tsai, and James Adcock. P -Values are Random Variables. *The American Statistician*, 62(3) :242–245, aug 2008.
- [253] Matthew A Napierala. What is the Bonferroni correction. *AAOS*, 6 :40, 2012.
- [254] Cristian Pattaro, Anna Köttgen, Alexander Teumer, Maija Garnaas, Carsten A. Böger, Christian Fuchsberger, Matthias Olden, Ming-Huei Chen, Adrienne Tin, Daniel Taliun, Man Li, Xiaoyi Gao, Mathias Gorski, Qiong Yang, Claudia Hundertmark, Meredith C. Foster, Conall M. O’Seaghdha, Nicole Glazer, Aaron Isaacs, Ching-Ti Liu, Albert V. Smith, Jeffrey R. O’Connell, Maksim Struchalin, Toshiko Tanaka, Guo Li, Andrew D. Johnson, Hincó J. Gierman, Mary Feitosa, Shih-Jen Hwang, Elizabeth J. Atkinson, Kurt Lohman, Marilyn C. Cornelis, Åsa Johansson, Anke Tönjes, Abbas Dehghan, Vincent Chouraki, Elizabeth G. Holliday, Rossella Sorice, Zoltan Kutalik, Terho Lehtimäki, Tõnu Esko, Harshal Deshmukh, Sheila Ulivi, Audrey Y. Chu, Federico Murgia, Stella Trompet, Medea Imboden, Barbara Kollerits, Giorgio Pistis, Tamara B. Harris, Lenore J. Launer, Thor Aspelund, Gudny Eiriksdottir, Braxton D. Mitchell, Eric Boerwinkle, Helena Schmidt, Margherita Cavalieri, Madhumathi Rao, Frank B. Hu, Ayse Demirkan, Ben A. Oostra, Mariza de Andrade, Stephen T. Turner, Jingzhong Ding, Jeanette S. Andrews, Barry I. Freedman, Wolfgang Koenig, Thomas Illig, Angela Döring, H.-Erich Wichmann, Ivana Kolcic, Tatijana Zemunik, Mladen Boban, Cosetta Minelli, Heather E. Wheeler, Wilmar Igl, Ghazal Zaboli, Sarah H. Wild, Alan F. Wright, Harry Campbell, David Ellinghaus, Ute Nöthlings, Gunnar Jacobs, Reiner Biffar, Karlhans Endlich, Florian Ernst, Georg Homuth, Heyo K. Kroemer, Matthias Nauck, Sylvia Stracke, Uwe Völker, Henry Völzke, Peter Kovacs, Michael Stumvoll, Reedik Mägi, Albert Hofman, Andre G. Uitterlinden, Fernando Rivadeneira, Yuri S. Aulchenko, Ozren Polasek, Nick Hastie, Veronique Vitart, Catherine Helmer, Jie Jin Wang, Daniela Ruggiero, Sven Bergmann, Mika Kähönen, Jorma Viikari, Tiit Nikopensius, Michael Province, Shamika Ketkar, Helen Colhoun, Alex Doney, Antonietta Robino, Franco Giulianini, Bernhard K. Krämer, Laura Portas, Ian Ford, Brendan M. Buckley, Martin Adam, Gian-Andri Thun, Bernhard Paulweber, Margot Haun, Cinzia Sala, Marie Metzger, Paul Mitchell, Marina Ciullo, Stuart K. Kim, Peter Vollenweider, Olli Raitakari, Andres Metspalu, Colin Palmer, Paolo Gasparini, Mario Pirastu, J. Wouter Jukema, Nicole M. Probst-Hensch, Florian Kronenberg, Daniela

- Toniolo, Vilmundur Gudnason, Alan R. Shuldiner, Josef Coresh, Reinhold Schmidt, Luigi Ferrucci, David S. Siscovick, Cornelia M. van Duijn, Ingrid Borecki, Sharon L. R. Kardia, Yongmei Liu, Gary C. Curhan, Igor Rudan, Ulf Gyllensten, James F. Wilson, Andre Franke, Peter P. Pramstaller, Rainer Rettig, Inga Prokopenko, Jacqueline C. M. Witteman, Caroline Hayward, Paul Ridker, Afshin Parsa, Murielle Bochud, Iris M. Heid, Wolfram Goessling, Daniel I. Chasman, W. H. Linda Kao, and Caroline S. Fox. Genome-Wide Association and Functional Follow-Up Reveals New Loci for Kidney Function. *PLoS Genetics*, 8(3) :e1002584, mar 2012.
- [255] Xiaoyu Zhang, Xuejiao Cui, and Feng Zhu. The effects of sample size on omics study : from the perspective of robustness and diagnostic accuracy. *IOP Conference Series : Earth and Environmental Science*, 252 :022127, jul 2019.
- [256] Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R. Martin, Hilary C. Martin, Tuuli Lappalainen, and Danielle Posthuma. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1) :59, dec 2021.
- [257] International Genetics Translational Research in Transplantation Network (iGene-TRAiN). Design and Implementation of the International Genetics and Translational Research in Transplantation Network. *Transplantation*, 99(11) :2401–12, nov 2015.
- [258] Claire E Fishman, Maede Mohebnasab, Jessica van Setten, Francesca Zanoni, Chen Wang, Silvia Deaglio, Antonio Amoroso, Lauren Callans, Teun van Gelder, Sangho Lee, Krzysztof Kiryluk, Matthew B Lanktree, and Brendan J Keating. Genome-Wide Study Updates in the International Genetics and Translational Research in Transplantation Network (iGene-TRAiN). *Frontiers in genetics*, 10 :1084, 2019.
- [259] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O’Connell, Adrian Cortes, Samantha Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly, and Jonathan Marchini. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726) :203–209, oct 2018.
- [260] Ahmed M. Alaa, Thomas Bolton, Emanuele Di Angelantonio, James H. F. Rudd, and Mihaela van der Schaar. Cardiovascular disease risk prediction using automated machine learning : A prospective study of 423,604 UK Biobank participants. *PLOS ONE*, 14(5) :e0213653, may 2019.
- [261] David J. Hunter. Gene–environment interactions in human diseases. *Nature Reviews Genetics*, 6(4) :287–298, apr 2005.
- [262] Lei Chen, Shiqi Zhang, Xiaoyong Pan, XiaoHua Hu, Yu-Hang Zhang, Fei Yuan, Tao Huang, and Yu-Dong Cai. HIV infection alters the human epigenetic landscape. *Gene Therapy*, 26(1-2) :29–39, feb 2019.

- [263] Kyung E. Rhee, Suzanne Phelan, and Jeanne McCaffery. Early Determinants of Obesity : Genetic, Epigenetic, and In Utero Influences. *International Journal of Pediatrics*, 2012 :1–9, 2012.
- [264] Daniëlle van Manen, Angélique B van ‘t Wout, and Hanneke Schuitemaker. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology*, 9(1) :70, dec 2012.
- [265] Beryl A Koblin, Marla J Husnik, Grant Colfax, Yijian Huang, Maria Madison, Kenneth Mayer, Patrick J Barresi, Thomas J Coates, Margaret A Chesney, and Susan Buchbinder. Risk factors for HIV infection among men who have sex with men. *AIDS*, 20(5) :731–739, mar 2006.
- [266] Hani Oweira, Ali Ramouz, Omid Ghamarnejad, Elias Khajeh, Sadeq Ali-Hasan-Al-Saegh, Rajan Nikbakhsh, Christoph Reißfelder, Nuh Rahbari, Arianeb Mehrabi, and Mahmoud Sadeghi. Risk Factors of Rejection in Renal Transplant Recipients : A Narrative Review. *Journal of Clinical Medicine*, 11(5) :1392, mar 2022.
- [267] Mathilde Lemoine, Dimitri Titeca Beauport, Thierry Lobbedez, Gabriel Choukroun, Bruno Hurault de Ligny, Marc Hazzan, Dominique Guerrot, and Dominique Bertrand. Risk Factors for Early Graft Failure and Death After Kidney Transplantation in Recipients Older Than 70 Years. *Kidney International Reports*, 4(5) :656–666, may 2019.
- [268] Gholamhossein Naderi, Amin Azadfar, Seyed Reza Yahyazadeh, Fatemeh Khatami, and Seyed Mohammad Kazem Aghamir. Impact of the donor-recipient gender matching on the graft survival from live donors. *BMC Nephrology*, 21(1) :5, dec 2020.
- [269] Jesse D. Schold, Joshua J. Augustine, Anne M. Huml, Richard Fatica, Saul Nurko, Alvin Wee, and Emilio D. Poggio. Effects of body mass index on kidney transplant outcomes are significantly modified by patient characteristics. *American Journal of Transplantation*, 21(2) :751–765, feb 2021.
- [270] Mathilde Prezelin-Reydit, Christian Combe, Jérôme Harambat, Christian Jacquelinet, Pierre Merville, Lionel Couzi, and Karen Leffondré. Prolonged dialysis duration is associated with graft failure and mortality after kidney transplantation : results from the French transplant database. *Nephrology Dialysis Transplantation*, 34(3) :538–545, 03 2018.
- [271] Yohann Foucher, Pascal Daguin, Ahmed Akl, Michèle Kessler, Marc Ladrière, Christophe Legendre, Henri Kreis, Lionel Rostaing, Nassim Kamar, Georges Mourad, Valérie Garrigue, François Bayle, Bruno H. de Ligny, Mathias Büchler, Carole Meier, Jean P. Daurès, Jean-Paul Soulillou, and Magali Giral. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney International*, 78(12) :1288–1294, dec 2010.

- [272] K. L. Lentine, J. M. Smith, A. Hart, J. Miller, M. A. Skeans, L. Larkin, A Robinson, K. Gauntt, A. K. Israni, R. Hirose, and J. J. Snyder. OPTN/SRTR 2020 Annual Data Report : Kidney. *American Journal of Transplantation*, 22(S2) :21–136, mar 2022.
- [273] OPTN. OPTN National Data, 2019 <https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/> accessed 10/15/22.
- [274] 1KG consortium Auton. A global reference for human genetic variation. *Nature*, 526(7571) :68–74, 2015.
- [275] Michelle E. Tarver-Carr, Neil R. Powe, Mark S. Eberhardt, Thomas A. LaVeist, Raynard S. Kington, Josef Coresh, and Frederick L. Brancati. Excess Risk of Chronic Kidney Disease among African-American versus White Subjects in the United States : A Population-Based Study of Potential Explanatory Factors. *Journal of the American Society of Nephrology*, 13(9) :2363–2370, sep 2002.
- [276] Giulio Genovese, David J. Friedman, Michael D. Ross, Laurence Lecordier, Pierrick Uzureau, Barry I. Freedman, Donald W. Bowden, Carl D. Langefeld, Taras K. Oleksyk, Andrea L. Uscinski Knob, Andrea J. Bernhardt, Pamela J. Hicks, George W. Nelson, Benoit Vanhollebeke, Cheryl A. Winkler, Jeffrey B. Kopp, Etienne Pays, and Martin R. Pollak. Association of Trypanolytic ApoL1 Variants with Kidney Disease in African Americans. *Science*, 329(5993) :841–845, aug 2010.
- [277] David J. Friedman and Martin R. Pollak. Genetics of kidney failure and the evolving story of APOL1. *Journal of Clinical Investigation*, 121(9) :3367–3374, sep 2011.
- [278] Afshin Parsa, W.H. Linda Kao, Dawei Xie, Brad C. Astor, Man Li, Chi-yuan Hsu, Harold I. Feldman, Rulan S. Parekh, John W. Kusek, Tom H. Greene, Jeffrey C. Fink, Amanda H. Anderson, Michael J. Choi, Jackson T. Wright, James P. Lash, Barry I. Freedman, Akinlolu Ojo, Cheryl A. Winkler, Dominic S. Raj, Jeffrey B. Kopp, Jiang He, Nancy G. Jensvold, Kaixiang Tao, Michael S. Lipkowitz, and Lawrence J. Appel. APOL1 Risk Variants, Race, and Progression of Chronic Kidney Disease. *New England Journal of Medicine*, 369(23) :2183–2196, dec 2013.
- [279] Barry I. Freedman, Carl D. Langefeld, Kelly K. Andringa, Jennifer A. Croker, Adrienne H. Williams, Neva E. Garner, Daniel J. Birmingham, Lee A. Hebert, Pamela J. Hicks, Mark S. Segal, Jeffrey C. Edberg, Elizabeth E. Brown, Graciela S. Alarcón, Karen H. Costenbader, Mary E. Comeau, Lindsey A. Criswell, John B. Harley, Judith A. James, Diane L. Kamen, S. Sam Lim, Joan T. Merrill, Kathy L. Sivils, Timothy B. Niewold, Neha M. Patel, Michelle Petri, Rosalind Ramsey-Goldman, John D. Reville, Jane E. Salmon, Betty P. Tsao, Keisha L. Gibson, Joyce R. Byers, Anna K. Vinnikova, Janice P. Lea, Bruce A. Julian, and Robert P. Kimberly. End-Stage Renal Disease in African Americans With Lupus Nephritis Is Associated With APOL1. *Arthritis Rheumatology*, 66(2) :390–396, feb 2014.

- [280] Jeffrey B. Kopp, George W. Nelson, Karmini Sampath, Randall C. Johnson, Giulio Genovese, Ping An, David Friedman, William Briggs, Richard Dart, Stephen Korbet, Michele H. Mokrzycki, Paul L. Kimmel, Sophie Limou, Tejinder S. Ahuja, Jeffrey S. Berns, Justyna Fryc, Eric E. Simon, Michael C. Smith, Howard Trachtman, Donna M. Michel, Jeffrey R. Schelling, David Vlahov, Martin Pollak, and Cheryl A. Winkler. APOL1 Genetic Variants in Focal Segmental Glomerulosclerosis and HIV-Associated Nephropathy. *Journal of the American Society of Nephrology*, 22(11) :2129–2137, nov 2011.
- [281] Alex N. Kasembeli, Raquel Duarte, Michèle Ramsay, Pulane Mosiane, Caroline Dickens, Thérèse Dix-Peek, Sophie Limou, Efe Sezgin, George W. Nelson, Agnes B. Fogo, Stewart Goetsch, Jeffrey B. Kopp, Cheryl A. Winkler, and Saraladevi Naicker. APOL1 Risk Variants Are Strongly Associated with HIV-Associated Nephropathy in Black South Africans. *Journal of the American Society of Nephrology*, 26(11) :2882–2890, nov 2015.
- [282] Russell Thomson, Giulio Genovese, Chelsea Canon, Daniella Kovacsics, Matthew K. Higgins, Mark Carrington, Cheryl A. Winkler, Jeffrey Kopp, Charles Rotimi, Adebowale Adeyemo, Ayo Doumatey, George Ayodo, Seth L. Alper, Martin R. Pollak, David J. Friedman, and Jayne Raper. Evolution of the primate trypanolytic factor APOL1. *Proceedings of the National Academy of Sciences*, 111(20), may 2014.
- [283] B. A. Julian, R. S. Gaston, W. M. Brown, A. M. Reeves-Daniel, A. K. Israni, D. P. Schladt, S. O. Pastan, S. Mohan, B. I. Freedman, and J. Divers. Effect of Replacing Race With Apolipoprotein L1 Genotype in Calculation of Kidney Donor Risk Index. *American Journal of Transplantation*, 17(6) :1540–1548, jun 2017.
- [284] Carmen A. Peralta, Kirsten Bibbins-Domingo, Eric Vittinghoff, Feng Lin, Myriam Fornage, Jeffrey B. Kopp, and Cheryl A. Winkler. APOL1 Genotype and Race Differences in Incident Albuminuria and Renal Function Decline. *Journal of the American Society of Nephrology*, 27(3) :887–893, mar 2016.
- [285] Haley M. Gonzales, John W. McGillicuddy, Vinayak Rohan, Jessica L. Chandler, Satish N. Nadig, Derek A. Dubay, and David J. Taber. A comprehensive review of the impact of tacrolimus inpatient variability on clinical outcomes in kidney transplantation. *American Journal of Transplantation*, 20(8) :1969–1983, aug 2020.
- [286] David J. Taber, Mulugeta G. Gebregziabher, Tittle R. Srinivas, Kenneth D. Chavin, Prabhakar K. Baliga, and Leonard E. Egede. African-American Race Modifies the Influence of Tacrolimus Concentrations on Acute Rejection and Toxicity in Kidney Transplant Recipients. *Pharmacotherapy : The Journal of Human Pharmacology and Drug Therapy*, 35(6) :569–577, jun 2015.
- [287] Éric Sagou Yayo, Mireille Aye, Jean-Louis Konan, Arlette Emième, Marie-Laure Atoungbre, Appolinaire Gnionsahé, Étienne Cavalier, Dagui Monnet, and Pierre Delanaye.

- Inadéquation du facteur ethnique pour l'estimation du débit de filtration glomérulaire en population générale noire-africaine : résultats en Côte d'Ivoire. *Néphrologie Thérapeutique*, 12(6) :454–459, nov 2016.
- [288] Casey R. Dorr, William S. Oetting, Pamala A. Jacobson, and Ajay K. Israni. Genetics of acute rejection after kidney transplantation. *Transplant International*, 31(3) :263–277, mar 2018.
- [289] Hyeon Joo Jeong. Diagnosis of renal transplant rejection : Banff classification and beyond. *Kidney Research and Clinical Practice*, 39(1) :17–31, mar 2020.
- [290] W. S. Oetting, D. P. Schladt, W. Guan, M. B. Miller, R. P. Remmel, C. Dorr, K. Sanghavi, R. B. Mannon, B. Herrera, A. J. Matas, D. R. Salomon, P.-Y. Kwok, B. J. Keating, A. K. Israni, and P. A. Jacobson. Genomewide Association Study of Tacrolimus Concentrations in African American Kidney Transplant Recipients Identifies Multiple CYP3A5 Alleles. *American Journal of Transplantation*, 16(2) :574–582, feb 2016.
- [291] Annick Massart, Lidia Ghisdal, Ondrej Viklicky, Maarten Naesens, Daniel Abramowicz, and Marc Abramowicz. Reply to Hernandez et al. - GWAS of acute renal graft rejection. *American Journal of Transplantation*, 18(8) :2098–2099, aug 2018.
- [292] Francesca Zanoni and Krzysztof Kiryluk. Genetic background and transplantation outcomes : insights from genome-wide association studies. *Current Opinion in Organ Transplantation*, 25(1) :35–41, feb 2020.
- [293] Geir Mjøen, Stein Hallan, Anders Hartmann, Aksel Foss, Karsten Midtvedt, Ole Øyen, Anna Reisæter, Per Pfeffer, Trond Jenssen, Torbjørn Leivestad, Pål-Dag Line, Magnus Øvrehus, Dag Olav Dale, Hege Pihlstrøm, Ingar Holme, Friedo W Dekker, and Hallvard Holdaas. Long-term risks for kidney donors. *Kidney international*, 86(1) :162–7, jul 2014.
- [294] Abimereki D Muzaale, Allan B Massie, Mei-Cheng Wang, Robert A Montgomery, Maureen A McBride, Jennifer L Wainright, and Dorry L Segev. Risk of end-stage renal disease following live kidney donation. *JAMA*, 311(6) :579–86, feb 2014.

SEPTIÈME PARTIE

# Liste des communications scientifiques

---



# BIBLIOGRAPHIE SCIENTIFIQUE

---

## 1.1 Publications scientifiques

**Ba R**, Geffard E, Douillard V, Simon F, Mesnard L, Vince N, Gourraud PA, Limou S. Surfing the Big Data Wave : Omics Data Challenges in Transplantation. *Transplantation*. 2022 Feb 1 ;106(2) :e114-e125. doi : 10.1097/TP.0000000000003992. PMID : 34889882.

**Ba R**, Durand A, Mauduit V, Chauveau C, Le Bas-Bernardet S, Salle S, Guérif P, Morin M, Petit C, Douillard V, Rousseau O, Blancho G, Kerleau C, Vince N, Giral M, Gourraud PA, Limou S. Cohort Profile : the French Genetics of Kidney Transplantation Study (KiT-GENIE). *Soumis à EJHG (en revisions)*.

## 1.2 Communications orales

**Ba R**, Vince N, Geffard E, Malguid D, Lanza M, Gourraud PA, Limou S. Burrowing functional and immunogenetic information through the 1000 Genomes project with Ferret v3.0. *European Immunogenetics and Histocompatibility Conference (EFI) - Lisbon, Portugal, 2019*.

Petit C, Dantal J, Durand A, **Ba R**, Rousseau O, Mauduit V, Morin M, Gatault P, Garrouste C, Couzi L, Mesnard L, Vince N, Limou S. First genome-wide association study of post-transplantation FSGS recurrence. *Société Francophone de Transplantation - Genève 2021*

## 1.3 Posters

**Ba R**, Vince N, Geffard E, Malguid D, Lanza M, Gourraud PA, Limou S. Burrowing functional and immunogenetic information through the 1000 Genomes Project with Ferret v3.0. *JOBIM annual bioinformatic meeting - Nantes, France, 2019*.

Geffard E, **Ba R**, Limou S, Brouard S, Loupy A, Vince N, Gourraud PA. De-centralized database : new challenges to design innovative contextualization algorithms. *European Immunogenetics and Histocompatibility Conference (EFI) - Lisbon, Portugal, 2019*.

Durand A, Geffard E, **Ba R**, Limou S, Brouard S, Loupy A, Vince N, Gourraud PA. De-centralized database : new challenges to design innovative contextualization algorithms. *JOBIM*

annual bioinformatic meeting - Nantes, France, 2019.

Mauduit V, Durand A, Ba R, Morin M, Petit C, Gourraud PA, Vince N, Limou S. Integrating large-scale genetic data to study kidney graft survival. Société Francophone de Transplantation - Genève 2021



**Titre :** Intégration de données génétiques à grande échelle pour l'amélioration de la prise en charge du transplanté rénal

**Mots clés :** Transplantation rénale, Paires donneur-receveur, Génomique, Biobanque, Analyses GWAS

**Résumé :** La transplantation rénale est le meilleur traitement de suppléance pour l'insuffisance rénale chronique terminale. Cependant, des défis majeurs liés au rejet et l'échec chronique restent à relever. L'objectif de ma thèse est de contribuer à l'identification de facteurs génétiques pour mieux comprendre les mécanismes de rejet/perte chroniques et améliorer la prise en charge des transplantés rénaux. Nous avons mis en place KiT-GENIE, l'une des plus grandes cohortes génétiques de transplantation rénale à ce jour qui s'organise sur une triple couche de données homogènes (cliniques, biologiques et génétiques) de donneurs et receveurs (D/R). KiT-GENIE inclut 4 217 individus (n=1 945 D et 2 272 R dont 1 969 paires D/R), 8M de SNPs et plus de 200 variables cliniques.

L'analyse statistique de ces données a permis d'établir la « carte d'identité » de KiT-GENIE avec une description précise et complète des individus et des variables de la cohorte.

Les résultats des GWAS menées sur les receveurs et les donneurs séparément ne montrent aucune association statistiquement significative entre variants et retour en dialyse au-delà de 1 an post-transplantation. Néanmoins, un signal approchant le seuil de significativité a été identifié dans le génome des receveurs au niveau du locus 11p14.1 qui a précédemment été associé avec la fonction rénale. Pour aller plus loin, la réalisation d'une méta-analyse ou l'analyse simultanée des génomes du donneur et du receveur pourraient identifier de potentiels facteurs de risque pour la perte du greffon.

KiT-GENIE offre l'opportunité unique d'approfondir l'étude des interactions donneurs-receveurs mais également d'explorer d'autres phénotypes d'intérêts tels que la fonction rénale du donneur ou encore la récurrence de certaines maladies rénales sur le greffon.

**Title :** Integration of genomic data towards care improvement for kidney transplanted patients

**Keywords :** Kidney transplantation, Donor-recipient pairs, Genomics, Biobank, GWAS data analysis

**Abstract :** Kidney transplantation is the best renal replacement strategy for end-stage renal disease. However, major challenges related to chronic rejection and failure remain. The objective of my thesis is to identify genetic factors underlying chronic rejection/failure and to improve patients care. We have consolidated KiT-GENIE, one of the largest kidney transplant genetic cohorts to date containing homogeneous high-quality triple data layer (clinical, biological and genetic) for donors and recipients (D/R). KiT-GENIE includes 4,217 individuals (n=1,945 D and 2,272 R, including 1,969 D/R pairs), 8M SNPs, and over 200 clinical variables.

Using statistical methods, we establish the "cohort profile" of KiT-GENIE with a precise and complete description of individuals and clinical variables.

GWAS analyses were conducted on recipients and donors separately. The results showed no significant association between variants and return to dialysis beyond 1 year post-transplantation. However, a signal approaching significance was identified in the recipients' genome at the 11p14.1 locus that has previously been associated with kidney function. To go further, a meta-analysis or simultaneous analysis of donor and recipient genomes could identify potential risk factors for graft loss.

KiT-GENIE offers a unique opportunity to further study donor-recipient interactions but also to explore other phenotypes of interest such as donor renal function or recurrence of initial diseases on the graft.