



HAL
open science

Analyse génomique de la coinfection par le virus VIH et VHC

Damien Ulveling

► **To cite this version:**

Damien Ulveling. Analyse génomique de la coinfection par le virus VIH et VHC. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Conservatoire national des arts et métiers - CNAM, 2016. Français. NNT : 2016CNAM1066 . tel-01508852

HAL Id: tel-01508852

<https://theses.hal.science/tel-01508852>

Submitted on 14 Apr 2017

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.

ÉCOLE DOCTORALE Sciences des Métiers de l'Ingénieur
Laboratoire Génomique Bioinformatique et Applications

THÈSE présentée par :

Damien ULVELING

Soutenue le : **28 Juin 2016**

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Spécialité : **Bioinformatique**

**Analyses génomiques de la coinfection par
les virus VIH et VHC**

THÈSE dirigée par :

M. ZAGURY Jean-François

Professeur, CNAM, Paris

RAPPORTEURS :

M. AMBROISE Christophe

Professeur, Université d'Evry Val d'Essonne, Evry

M. HEZODE Christophe

Professeur, Université Paris Est-Créteil, Créteil

Président du jury :

Mme. SALMON Dominique

Professeur, Hôpital Cochin, Paris

JURY :

Mme. DOMINGUEZ Stéphanie

Docteur, AP-HP Henri-Mondor, Créteil

Mme. LE CLERC Sigrid

Docteur, CNAM, Paris

« La science consiste à passer d'un étonnement à un autre. »

– Aristote –

Remerciements

Cette thèse est l'aboutissement d'un long travail, durant lequel j'ai rencontré d'extraordinaires personnes qui m'ont permis d'évoluer tant sur le plan professionnel que sur le plan personnel. Je m'apprête à remercier ceux sans qui ce manuscrit n'aurait pas vu le jour.

Mes premiers remerciements sont pour le Professeur Jean-François Zagury qui m'a offert l'opportunité de réaliser ce travail de thèse et m'a accordé sa confiance tout au long de ces années.

Je tiens à remercier le Professeur Christophe Ambroise et le Professeur Christophe Hézode d'avoir accepté de juger mon travail et d'être rapporteur de cette thèse.

Je tiens à remercier le Professeur Dominique Salmon, le Dr Stéphanie Dominguez et le Dr Sigrid Le Clerc de me faire l'honneur d'examiner mon travail.

Je remercie encore le Dr Stéphanie Dominguez sans qui ce travail de thèse n'existerait pas et qui au quotidien a pris le temps de m'expliquer l'aspect clinique de mon travail.

Je souhaite exprimer ma gratitude au groupe d'étude ANRS CO13 HEPAVIH qui m'a fait confiance pour l'analyse des données.

J'aimerais également remercier le Dr Laurent Abel et le Dr Aurélie Cobat de l'institut Imagine avec qui j'ai collaboré et qui m'ont grandement aidé pour réaliser ce travail.

Ces années de thèse n'auraient jamais été les mêmes sans le soutien des personnes du laboratoire Génomique Bioinformatique et Applications du CNAM. Tout d'abord, les piliers du GBA sans qui la vie de cette équipe ne serait que désorganisation : Christiane (merci d'avoir toujours été présente pour moi), Juanjo et Jean-Louis (pour ces moments de co-voiturage lors des séminaires), sans oublier Janine.

Je tiens également à remercier les personnes qui m'ont le plus influencé au quotidien, j'ai nommé la « dream team » Génomique : Sigrid sans qui mon travail n'aurait été qu'un « brouillon de science », Vincent qui m'a permis d'atteindre une certaine compréhension de la statistique corrélée à un amour caché du rap, Toufik, le bonheur du GBA, Cédric pour sa précieuse aide malgré nos nombreux désaccords lors de « débats publics », Josselin pour son flegme anglais et sa bienveillance lors des moments difficiles, et enfin Marc pour sa collection de chocolat.

Je n'oublie pas les anciens génomiciens, Lieng pour sa créativité débordante, Sophie de m'avoir défendu lors de mon premier Ermenonville, Olivier pour ses bons mots d'humeur et enfin Pierre pour l'alcool de prune.

L'équipe du « Drug Design » a aussi été très présente, Matthieu le grand manitou de la protéine mais aussi du Hell Fest !, Nathalie, ma voisine de bureau, pour ses moments de bonne humeur quotidienne et qui m'a supporté/aidé durant toutes ces années, Charly pour sa « présence tranquille », Daniela, Jérémy et Benjamin les derniers arrivés, sans oublier Hélène pour sa culture et Nesrine l'unique pour sa bienveillance dès notre première rencontre.

Vient ensuite l'équipe « Cytokine », Hadley toujours partante pour un moment convivial, Lucille d'être aussi forte dans la prédiction de résultats footballistiques, Barbara, la joyeuse nouvelle, Hervé pour sa capacité à débattre de tout, et enfin Gaby le Coluche des présentations.

J'ai une pensée particulière pour le Dr Claire Francastel et le Dr Florent Hubé qui m'ont mis le pied à l'étrier et toujours soutenu et aidé dans mes entreprises scientifiques et professionnelles.

Je souhaite remercier mes amis, sans qui la vie serait moins drôle, Darius mon frère Polonais, Estelle pour ces soirées restaurants réconfortantes, Hugo, Tomp et Thierry pour les répétitions improvisées, mais aussi Julien et Nono qui ont eu le plaisir de goûter avant moi la joie d'être docteur.

Je remercie également toute ma famille, de m'avoir permis de réaliser mes rêves depuis toujours et de m'avoir poussé lors des moments difficiles. Plus particulièrement mes parents, sans qui je n'en serais pas là et qui m'ont aidé à grandir au quotidien malgré des étapes toujours plus hautes.

Mes derniers remerciements sont pour la personne la plus importante de ma vie, Cécile, sainte patronne de ma musique, qui fait battre mon cœur depuis près de 8 ans et qui a toujours été présente. Tu as enduré mes moments difficiles mais tu es toujours resté mon premier soutien. Je t'aime.

Mes derniers mots ne seront pas des remerciements mais un message d'amour, à toi ma fille : Je ne te connais pas encore, mais une chose est sûre, dès notre premier contact, par échographie interposée, j'ai été ému comme jamais, et ma vie a subi le plus merveilleux des bouleversements.

Résumé

Plus de 170 millions d'individus sont infectés par le VHC dans le monde et 37 millions par le VIH. La coinfection VIH/VHC est fréquente et représente un élément clé de la prise en charge des patients infectés par le VIH. Depuis l'arrivée des HAART, les maladies du foie sont devenues la cause principale de mortalité chez les patients coinfectés VIH/VHC. L'évolution naturelle et le pronostic de l'hépatite C sont plus sévères en cas de coinfection par le VIH du fait d'une fibrose accélérée et d'une évolution rapide vers la cirrhose et ses complications. Certains facteurs accélérant la fibrose hépatique sont clairs aujourd'hui comme: l'absence de recours au traitement anti-VHC, la réplication active du VHC et la consommation excessive d'alcool. De plus, il existe de plus en plus de preuves que les variants génétiques contribuent à la fibrose hépatique chez les patients mono-infectés par le VHC, mais cet aspect a été peu étudié dans la coinfection VIH/VHC.

Durant ma thèse, j'ai eu accès aux données d'un échantillon de 494 patients coinfectés génotypés issu de la cohorte ANRS CO13 HEPAVIH. L'histoire naturelle du VIH et du VHC y est renseignée de manière très détaillée et le suivi clinique des patients permet d'avoir des informations précises sur l'état de fibrose hépatique. J'ai pu alors réaliser deux études d'association « génome-entier » pour identifier des polymorphismes associés à la sévérité de la fibrose à l'aide de données complètes de 292 patients. La première étude a mis en évidence une association entre la quantification de l'élasticité hépatique par Fibroscan® et un locus, également répliqué dans la mono-infection par le VHC. Cette association a permis d'identifier deux gènes impliqués dans des mécanismes de maintien de structure et de signalisation cellulaire (CAV3) mais aussi dans la réplication du VHC (RAD18). La seconde étude a identifié deux associations significatives en comparant deux groupes de scores METAVIR (F0F1F2 vs F3F4), en particulier dans le gène CTNND2 qui est impliqué dans un réseau d'interaction associé à des mécanismes moléculaires lié à des maladies hépatiques.

Ces deux études sont en cours de publication dans des revues scientifiques internationales à comité de lecture. Ces nouvelles perspectives dans la compréhension des mécanismes de fibrose dans le contexte de la coinfection VIH/VHC pourraient aider à l'identification de nouvelles cibles pour la création de médicaments ou de tests diagnostiques afin d'améliorer les soins des patients.

Mots clés : GWAS, SNP, coinfection VIH/VHC, fibrose hépatique, maladie du foie

Résumé en anglais

Over 170 million people worldwide are infected by HCV and 37 million by HIV. Both viruses share the same modes of transmission, and HIV/HCV coinfection is common and represents a key element in the management of patients infected with HIV. Since the appearance of HAART, liver diseases have become the leading cause of death in HIV/HCV coinfecting patients. The natural history and prognosis of hepatitis C are more severe in case of coinfection with HIV due to accelerated rate of fibrosis progression and rapid progression to cirrhosis and its complications. Factors accelerating liver fibrosis are known today such as the lack of recourse to anti-HCV treatment, active HCV replication and excessive alcohol consumption. There is increasing evidence that genetic variants contribute to liver fibrosis in HCV monoinfection, but this aspect has been little studied in HIV/HCV coinfection.

I have exploited the genotype information from 494 coinfecting patients from the cohort ANRS CO13 HEPAVIH. These patients are very-well documented regarding the history of their HIV/HCV infection and are very carefully followed-up, especially regarding the status of liver fibrosis. I have performed two genome-wide association studies to identify polymorphisms associated with the severity of fibrosis from complete data of 292 patients. The first study has dealt with the quantification of liver stiffness by Fibroscan® and an association with the 3p25 region has been identified, also replicated in monoinfection HCV. Two genes involved in cell signaling and structure of holding mechanisms (CAV3) but also in HCV replication (RAD18) appear as good candidates. The second study has unraveled two significant associations by comparing the METAVIR score group (F0F1F2 vs F3F4), especially in the CTNND2 gene implicated in a network of interactions with molecular mechanisms involved in liver diseases.

These results are under publications in peer-review international scientific journals. These new insights into the molecular mechanisms of liver fibrosis in patients with HIV/HCV co-infection may help to define new targets for drug development or new diagnostic tests, to improve patient care.

Keywords : GWAS, SNP, HIV/HCV coinfection, liver fibrosis, liver disease

Table des matières

Remerciements	1
Résumé	3
Résumé en anglais	4
Table des matières	5
Liste des tableaux	10
Liste des figures	11
Liste des abréviations	14
Première partie Introduction.....	17
1. GENETIQUE et MALADIES.....	19
1.1 L'Acide Désoxyribonucléique	19
1.2 Rôle dans le vivant	21
1.3 Les polymorphismes génétiques	22
1.3.1 Les polymorphismes chromosomiques	22
1.3.2 Les séquences répétées en tandem	23
1.3.3 Les indels.....	23
1.3.4 Les polymorphismes mono-nucléotidique	24
1.3.5 Les Copy Number Variations.....	24
1.4 Notions de génétique des populations	24
1.4.1 Modèle de Hardy-Weinberg.....	24
1.4.2 Déséquilibre gamétique et déséquilibre de liaison	26
1.4.3 Haplotypes.....	30
1.4.4 Structure des populations	32
1.5 Epidémiologie génétique.....	33
1.5.1 Polymorphismes génétiques et pathologies.....	33
1.5.2 Types d'études génétiques et génomique	34
2 LE SYNDROME D'IMMUNO-DEFICIENCE ACQUISE.....	38
2.1 La maladie	38
2.1.1 Epidémiologie du SIDA et modes de transmission du VIH-1	39
2.1.2 Evolution clinique et biologique	39
2.2 Le virus de l'immunodéficience humaine	41
2.2.1 Organisation génomique et structurale du VIH-1	41
2.2.2 Le cycle de réplication virale	43

2.2.3	Tropisme cellulaire.....	44
2.3	Mécanismes de pathogenèse du SIDA	44
2.3.1	Effets directs.....	44
2.3.2	Effets indirects.....	45
2.4	Les traitements actuels	46
2.5	Facteurs génétiques de l'hôte et SIDA	47
3	L'HEPATITE C	50
3.1	La maladie	50
3.1.1	Epidémiologie et modes de transmission du VHC.....	51
3.2	Le virus de l'hépatite C	52
3.2.1	Organisation structurale et génomique du VHC	52
3.2.2	Cycle réplcatif du VHC.....	55
3.2.3	Variabilité génétique du VHC.....	57
3.2.4	Tropisme cellulaire.....	59
3.3	Histoire naturelle de l'infection par le VHC	60
3.4	Mécanismes de pathogenèse liés au VHC.....	61
3.4.1	Réponses immunitaires innées et adaptatives au VHC	62
3.4.2	Mécanismes pathogéniques associés à l'hépatite C chronique	65
3.4.3	Mécanisme de fibrose lié à l'hépatite C	66
3.4.4	Cirrhose et hépatocarcinome cellulaire (HCC)	68
3.5	Les traitements de l'hépatite C.....	69
3.6	Facteurs génétiques de l'hôte et l'hépatite C	71
3.6.1	Guérison sous traitement ou spontanée	71
3.6.2	Progression de la maladie.....	72
3.6.3	Hépatocarcinogenèse.....	73
4	COINFECTION PAR LES VIRUS DU VIH ET VHC	75
4.1	Epidémiologie	75
4.2	Histoire naturelle de la coinfection	75
4.2.1	Effet de la coinfection VIH/VHC sur l'histoire naturelle du VHC.....	75
4.2.2	Effet de la coinfection par le VHC sur les maladies associées au SIDA	76
4.2.3	Les mécanismes d'interaction VIH/VHC	77
4.3	Traitement des individus coinfectés VIH/VHC	78
4.4	Hépatotoxicité des thérapies par HAART.....	79

5	OBJECTIFS DE MA THESE	80
	Deuxième partie Matériels et méthodes	81
1	OUTILS GENOMIQUES ET BIOINFORMATIQUE	83
1.1	Bases de données bioinformatiques	83
1.1.1	dbSNP.....	83
1.1.2	Le projet HapMap	83
1.1.3	Le projet 1000 Genomes	85
1.2	Le génotypage et les puces à ADN	86
1.3	Reconstruction des haplotypes	87
1.4	Imputation	88
2	ETUDES D'ASSOCIATION GENOME ENTIER	91
2.1	Population étudiée	91
2.1.1	La cohorte ANRS CO 13 HEPAVIH.....	91
2.2	Phénotypes analysés	92
2.2.1	Méthodes d'évaluation de la fibrose hépatique.....	92
2.2.2	Définition des phénotypes analysés	95
2.3	Covariables utilisées dans l'analyse statistique.....	95
2.4	Génotypage.....	96
2.5	Contrôle qualité du génotypage.....	96
2.6	Etude de la stratification de la population	97
2.7	Imputation	97
2.8	Analyses statistiques	97
2.9	Approfondissement des associations trouvées	98
2.9.1	Déséquilibre de liaison	98
2.9.2	Exploration des résultats par bioinformatique	99
2.9.3	Réplication dans des cohortes de patients monoïnfectés VHC	99
	Troisième partie Résultats	101
1	Etude « génome entier » de la cohorte ANRS CO13 HEPAVIH de patients coïnfectés par le VHC et le VIH : à la recherche d'associations liées à la quantification de la fibrose.....	104
2	Etude « génome entier » de la cohorte ANRS CO13 HEPAVIH de patients coïnfectés par le VIH et le VHC : à la recherche d'associations liées à l'impact de la gravité de fibrose par comparaison de groupes de score METAVIR F0F1F2 et F3F4.....	151
3	RESULTATS ADDITIONNELS	190

3.1	Etude d'épistasie	190
3.2	Réplication de signaux obtenus par d'autres GWAS sur des maladies du foie	192
	Quatrième partie Discussion et perspectives.....	194
1	BILAN DES DEUX ETUDES GENETIQUES REALISEES SUR LA COHORTE DE COINFECTION ANRS CO13 HEPAVIH	196
1.1	Travaux portant sur l'élasticité hépatique mesurée par Fibroscan®	196
1.1.1	Rappel des résultats	196
1.1.2	Interprétation biologique	197
1.1.3	Comparaison avec la littérature	198
1.2	Travaux portant sur l'impact de la gravité de fibrose entre les groupes METAVIR F0F1F2 et F3F4.....	199
1.2.1	Rappel des résultats	199
1.2.2	Interprétation biologique	199
1.2.3	Comparaison avec la littérature	200
1.3	Discussion sur les résultats additionnels	201
1.3.1	Etude d'épistasie	201
1.3.2	Réplication des signaux obtenus par d'autres GWAS sur des maladies du foie.....	202
1.4	Comparaison des deux études génétiques « génome entier » réalisées sur la cohorte ANRS CO13 HEPAVIH	203
1.4.1	Comparaison entre les deux études	203
1.4.2	Interprétation des différences observées	205
1.4.3	Intérêt de la cohorte de patients coinfectés ANRS CO13 HEPAVIH.....	206
2	CRITIQUE DES METHODES UTILISEES	209
2.1	Etudes « génome entier »	209
2.1.1	Intérêt	209
2.1.2	Limites.....	210
3	PERSPECTIVES.....	212
3.1	Réplifications et méta-analyse dans la coinfection.....	212
3.2	Approches multi-marqueurs	212
3.2.1	Haplotypes.....	212
3.2.2	Epistasie	213
3.3	Analyse des CNVs	213
3.4	Analyses d'eQTLs.....	214

3.5	Nouvelles technologies de séquençage ou NGS	215
3.6	Biologie des systèmes	216
3.7	Autres phénotypes à explorer	218
	Cinquième partie Conclusion	220
1	Conclusion.....	222
	Bibliographie.....	223
	Liste des publications.....	250
	Liste des communications orales.....	251
	Posters	251

Liste des tableaux

Tableau 1. Classification METAVIR (344)	94
Tableau 2. Listes des gènes et des SNPs associés dans le SIDA	190
Tableau 3. Moyennes de fibrose en kPa, en fonction des génotypes du SNP rs61183828 et du SNP rs3092960 (CCR5)	191
Tableau 4. Moyennes de fibrose en kPa, en fonction des génotypes du SNP rs61183828 et du SNP rs3092960 (CCL5)	191
Tableau 5. Listes des SNPs associés à des maladies hépatiques, répliqués dans notre étude d'association génétique à sqrtLS	192
Tableau 6. Listes des SNPs associés à des maladies hépatiques, répliqués dans notre étude d'association génétique de comparaison des groupes METAVIR F0F1F2 et F3F4.....	193
Tableau 7. Ensemble des SNPs et des p-valeurs associés à la sévérité de la fibrose dans chacune des deux études génétiques. Par trait quantitatif (sqrtLS) ou par « cas/témoins » (F0F1F2 vs F3F4).....	203

Liste des figures

- Figure 1. Portion de double hélice d'ADN. La direction de chaque brin (violet et vert) est indiquée par les simples flèches dans le sens 5' vers 3'. Les flèches bidirectionnelles indiquent les petits et grands sillons. Les nucléotides sont appariés selon leur complémentarité (A/T et C/G). 20
- Figure 2. Représentation de la synthèse des protéines. La séquence d'ADN est transcrite en ARN messager. L'ARN messager est ensuite traduit en protéine. 21
- Figure 3 Exemples de polymorphismes chromosomiques. (a) Translocation entre 2 chromosomes. (b) Inversion au sein d'un même chromosome. (c) Anomalie du nombre de chromosomes : exemple de la trisomie 21. 23
- Figure 4.Exemples d'insertion/délétion. La séquence 1 fait office de référence. La séquence 2 comporte une insertion de deux nucléotides (en rouge) entre le G et le T (en bleu) par rapport à la séquence 1. La séquence 3 illustre une délétion des deux nucléotides G et T (en bleu sur la séquence 1) par rapport à la séquence de référence. 23
- Figure 5. Exemple de SNP. Le nucléotide G (en bleu) de la séquence 1 est remplacé par un C (en bleu) dans la séquence 2..... 24
- Figure 6. Vitesse de disparition du déséquilibre gamétique pour différentes valeurs du taux de recombinaison entre deux loci. Plus le taux de recombinaison est grand, plus le déséquilibre disparaît rapidement..... 28
- Figure 7. Exemple d'haplotypes composés de 3 SNPs. Adapté de (12). 30
- Figure 8. Exemple de genèse d'haplotypes dans une région de 3 SNPs 31
- Figure 9. Résumé statistique de données génétiques de 1387 individus Européens grâce à une analyse en composante principale (ACP). Chaque label de couleur représente les individus des pays et les ronds pleins avec label sont la médiane des axes PC1 et PC2 des individus par pays. La carte en haut à droite permet de relier les labels des pays (18). 32
- Figure 10. Découverte d'un variant associé à un phénotype. Le variant détecté lors d'une étude de génétique épidémiologique n'est pas nécessairement le variant causal de la maladie mais peut être en déséquilibre avec celui-ci. Adapté de (26). 34
- Figure 11. Exemple de transmission d'une maladie par étude de 3 SNPs (A, B et C). Arbre généalogique incluant, en rouge, les individus malades et en blanc, les individus sains. Le SNP C, par l'allèle c1, est présent uniquement chez les individus malades, suggérant que le locus est impliqué dans l'apparition de la maladie..... 35

Figure 12. Représentation schématique d'une association génétique par étude 'cas/témoins' entre un SNP et une pathologie présente chez les cas. La significativité est évaluée par la différence de répartition des allèles entre les cas et les témoins.	36
Figure 13. Profil d'évolution de l'infection par le VIH-1.	40
Figure 14. Structure génomique du VIH-1. En plus des trois gènes de « structure » (gag, pol et env), 6 gènes codent pour des protéines régulatrices (vif, vpr, vpu, tat, rev et nef)	41
Figure 15. Structure schématique d'une particule virale VIH-1.	42
Figure 16. Cycle de réplication du VIH-1 (Pour la science, n°377, mars 2009).....	43
Figure 17. Liste de gènes candidats trouvés en association avec le SIDA.....	49
Figure 18. Prévalence du VHC dans le monde (96).....	51
Figure 19. Le génome du VHC et ses protéines virales. A. Schéma du génome du VHC et composition de la polyprotéine traduite puis clivée en protéines structurales ou non structurales. B. Fonctions des protéines virales.	53
Figure 20. Représentations structurales du VHC. A. Organisation schématique du VHC. B. Image de particules virales par microscopie électronique montrant l'hétérogénéité morphologique des virions (112).	54
Figure 21. Représentation schématique du cycle répliatif du VHC. Adapté de (114).	55
Figure 22. Arbre d'évolution des principaux génotypes du VHC. Adapté de (136).	58
Figure 23. Histoire naturelle de l'infection par le VHC. L'évolution vers le carcinome hépatocellulaire prend environ 30ans. Adapté de http://yourhealtheducator.com/hepatitis-inflammation-liver-cells/	60
Figure 24. Histoire naturelle du VHC et les causes biologiques induisant des pathologies liées à l'infection. Adapté de (175).	62
Figure 25. Schématisation d'un lobule hépatique. Regroupement de cellules en unité fonctionnelle du foie. L'espace porte est l'ensemble composé de l'artère hépatique, de la veine porte ainsi que du canal biliaire. (Image Servier Medical Art)	65
Figure 26. Le foie normal présente des cellules étoilées (HSC) quiescentes et une matrice extracellulaire (MEC) de faible densité dans l'espace peri-sinusoïdal. Lors de lésions chroniques, les HSC prolifèrent et sont activées. Cela favorise la production d'une MEC de haute densité qui à long terme abouti à la fibrose (199).....	67
Figure 27. Variations génétiques de l'hôte et autres facteurs associés à l'évolution clinique de l'infection par le VHC. Les gènes en gras ont été identifiés par GWAS alors qu'en police normale par approche "gène candidat". Adapté de (250).....	71

Figure 28. Interaction entre le VIH et le VHC dans les hépatocytes et les HSC, contribuant à la fibrose hépatique. Le VIH active plusieurs mécanismes qui augmentent la fibrose, la réplication du VHC et l'induction de l'apoptose (306). Abréviations: CCR5, C-C chemokine receptor type 5; COL1A1, type 1 collagen; CXCR4, C-X-C chemokine receptor type 4; DR, death receptor; HSC, hepatic stellate cell; MCP-1, monocyte chemoattractant protein 1; ROS, reactive oxygen species; TGF- β 1, transforming growth factor beta-1; TIMP-1, tissue inhibitor of metalloproteinases; TRAIL, TNF-related apoptosis-inducing ligand.	77
Figure 29. Notion de "tagSNP" (12). (a) Identification de 3 SNPs (en couleur) dans une portion chromosomique. (b) Reconstruction des haplotypes constitués de 20 SNPs dont les trois identifiés précédemment. (c) Détermination de 3 tagSNPs dont la connaissance suffit pour identifier les 4 haplotypes de la population. Par exemple, un profil G-T-C pour ces 3 tagSNPs correspond toujours à l'haplotype 3.	85
Figure 30. Schématisation des étapes du génotypage selon le procédé Illumina.....	87
Figure 31. Problématique de l'haplotypage	88
Figure 32. Représentation schématique du problème d'imputation. Les génotypes manquants de la puce seront imputés à l'aide des haplotypes d'un panel de référence. Adapté de (342).	89
Figure 33. Echelle des valeurs d'élasticité hépatique en kPa (2.5, 7, 9.5, 12.5, 75) selon le stade de fibrose et sa correspondance en score METAVIR (stade F du vert au rouge).	94
Figure 34. Classification des marqueurs génétiques associés aux maladies en fonction de leur fréquence et de leur pénétrance dans la maladie	210

Liste des abréviations

ADN	Acide DésoxyriboNucléique
ANRS	Agence Nationale de Recherché sur le SIDA et les hépatites
ARN	Acide RiboNucléique
CEU	Caucasiens Européens de l'Utah (Etats-Unis)
CMSP	Cellule Mononuclée du Sang Périphérique
CNV	Copy Number Variant
DAA	Direct Acting Antiretroviral
eQTL	expression Quantative Trait Locus
FPR	Fibrosis Progression Rate
GWAS	Genome Wide Association Study
HAART	Highly Active AntiRetroviral Therapy
HCC	HépatoCarcinome Cellulaire
HLA	Human Leucocyte Antigen
HSC	Hepatic Stellate Cell
IFN	InterFeroN
IL	InterLeukine
IRES	Internal Ribosome Entry Site
kPa	kilo Pascal (unité de mesure)
LD	Linkage Disequilibrium
lncRNA	long non-coding RNA
LPS	LipoPolySaccharide
LTR	Long Terminal Repeat
MAF	Minor Allele Frequency
MEC	Matrice ExtraCellulaire
MHC	Major Histocompatibility Complex
NGS	Next Generation Sequencing
NK	Natural Killer
pb	paire de bases
PBH	Ponction Bio-Hépatique
PEG-IFN	PEGylated-InterFeroN
RBV	RiBaVirine
SIDA	Syndrome d'ImmunoDéficiency Acquis

SNP	Single Nucleotide Polymorphism
sqrtLS	square-root of Liver Stiffness
SVR	Sustained Virological Response
UTR	UnTranslated Region
VHC	Virus de l'Hépatite C
VIH	Virus de l'Immunodéficience Humaine

Première partie
Introduction

1. GENETIQUE et MALADIES

Pourquoi s'intéresser à la génétique ? Depuis bien longtemps, l'être humain a cherché à comprendre le fonctionnement de la nature et de son corps. C'est donc naturellement que cette science, assez récente, est devenue indispensable pour comprendre les propriétés des êtres vivants. Depuis Mendel au 19e siècle, le pionnier de la génétique, jusqu'au séquençage complet des 3 milliards de bases du génome humain en 2003, la génétique a acquis une place importante dans la résolution des problèmes de société d'aujourd'hui. La compréhension de certaines maladies génétiques comme le diabète, ou encore l'amélioration génétique dans le domaine agricole (soja, maïs), bien que controversée, en sont de parfait exemples.

La génétique peut être définie comme étant la science qui étudie l'hérédité et les gènes. En biologie, l'hérédité est l'ensemble des propriétés que les êtres vivants transmettent à leurs descendants par le biais de la reproduction. Ces propriétés se manifestent par l'intermédiaire des caractères héréditaires qui ont la particularité de s'exprimer avec régularité au cours des générations. L'étude de l'hérédité passe par la compréhension des propriétés du matériel génétique. Il faut donc étudier les propriétés des acides nucléiques (l'acide désoxyribonucléique, ADN, et l'acide ribonucléique, ARN) à différents niveaux.

La génétique progresse d'une manière qui lui est propre, fondée sur l'analyse de variations, naturelles ou induites, des gènes dans une population d'organismes. Cette variation constitue la matière première de la génétique et en son absence aucune analyse génétique n'est possible. La génétique étudie tous les aspects du gène. L'étude de la transmission des gènes de génération en génération est dite génétique de la transmission (ou encore génétique formelle). L'étude de la structure et des fonctions des gènes est dite génétique moléculaire. L'étude du comportement des gènes dans des populations est dite génétique des populations. Toutefois, cette subdivision de la génétique en plusieurs disciplines est arbitraire et on observe souvent des recouvrements entre ces différentes disciplines.

1.1 L'Acide Désoxyribonucléique

Le premier ADN a été identifié et isolé en 1869 à partir du noyau de leucocytes par le Suisse Friedrich Miescher (1). Sa structure en double hélice n'a cependant été mise en évidence qu'en

1953 par le Britannique Francis Crick et l'Américain James Watson (2) à partir des données expérimentales de diffraction de rayons X obtenues par les Britanniques Rosalind Franklin et Maurice Wilkins.

L'ADN est le support de l'information génétique qui permet le développement et le fonctionnement des êtres vivants. Il constitue une matrice stable, répliquative et transmissible. Les molécules d'ADN sont formées de deux brins antiparallèles complémentaires. Chaque brin est dit polynucléotidique, c'est-à-dire qu'il est formé d'un enchaînement de nucléotides, lesquels sont constitués d'une base azotée ; A pour l'adénine, T pour la thymine, C pour la cytosine et G pour la guanine ; liée à un désoxyribose lui-même lié à un groupe phosphate. Les nucléotides polymérisés sont liés les uns aux autres par des liaisons covalentes entre le désoxyribose d'un nucléotide et le groupe phosphate du nucléotide suivant, formant ainsi le squelette de l'ADN. L'aspect antiparallèle complémentaire de l'ADN est relatif à la direction de la polymérisation de chaque brin et à la complémentarité inter-brin entre les bases azotées C et G ou A et T, grâce à des liaisons hydrogènes (Figure 1).

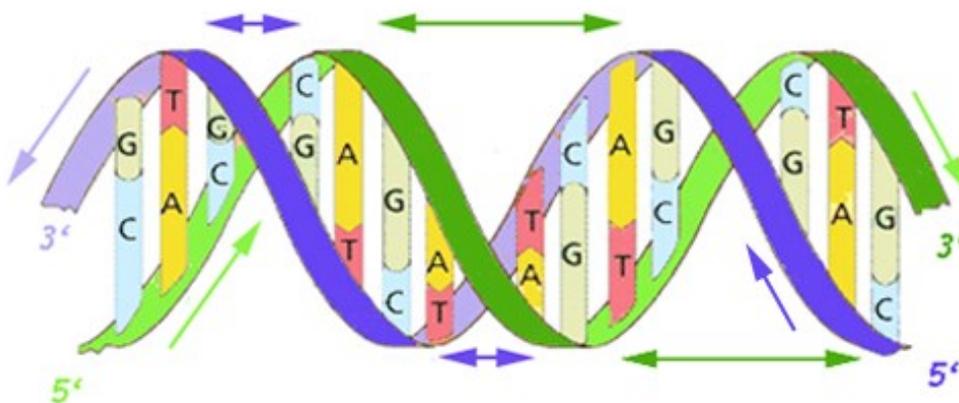


Figure 1. Portion de double hélice d'ADN. La direction de chaque brin (violet et vert) est indiquée par les simples flèches dans le sens 5' vers 3'. Les flèches bidirectionnelles indiquent les petits et grands sillons. Les nucléotides sont appariés selon leur complémentarité (A/T et C/G).

Dans les cellules eucaryotes, l'ADN est réparti au sein du noyau en molécules compactes appelées chromosomes subdivisées sur leur longueur en milliers de régions fonctionnelles appelées gènes. L'être humain possède 23 paires de chromosomes : 22 paires d'autosomes numérotées de 1 à 22 par taille décroissante et une paire de chromosomes sexuels (XX pour les

femmes et XY pour les hommes). Lors de la reproduction, les chromosomes sont brassés et transmis à la descendance. L'ensemble des chromosomes est appelé génome et son étude est appelée génétique. Une partie de cette discipline se focalise sur l'ensemble des chromosomes et s'appelle génomique.

1.2 Rôle dans le vivant

Le dogme central de la biologie veut que l'information génétique stockée dans les molécules d'ADN, plus précisément les gènes, soit traduite en protéines par l'intermédiaire de molécules d'ARN (Figure 2), confinées à un rôle d'intermédiaire. Chez les eucaryotes, les gènes ont des régions codantes et non codantes (nommées exons et introns respectivement). La région du génome transcrite en ARN messenger (ARNm) est complémentaire à la région génique cible. L'ARNm est une copie des séquences d'ADN cibles à l'exception du T qui est remplacé par la base nucléique Uracile (U). Dans le but de garder uniquement les séquences codantes des ARNm, une étape d'épissage est nécessaire avant leurs traductions en protéine. La lecture de l'ARNm par la machinerie de traduction se fait de manière séquentielle, par triplets de nucléotides de l'ARNm, appelés codons, qui définissent l'enchaînement des acides aminés formant les protéines. En effet, à chaque codon correspond un acide aminé, selon une table de correspondance appelée le code génétique.

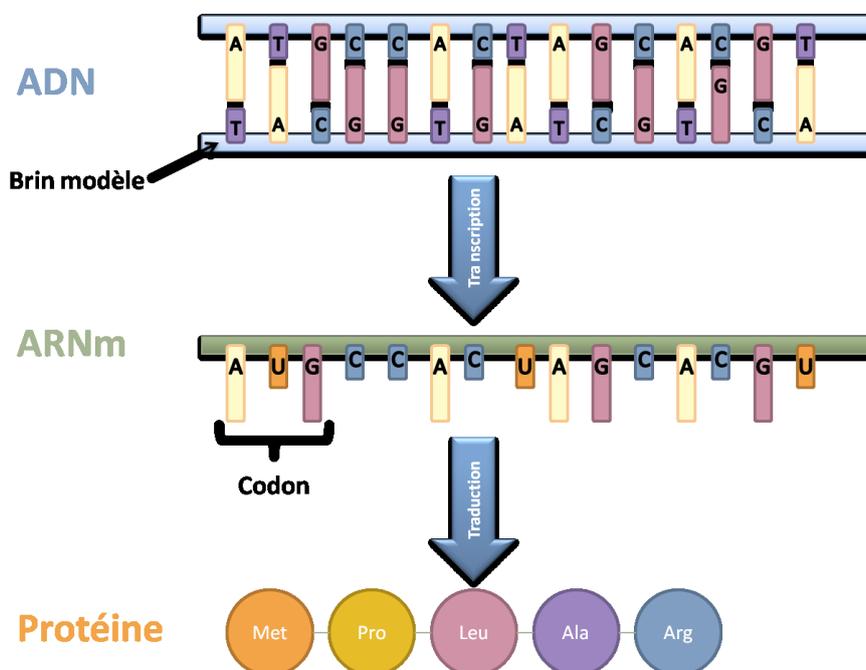


Figure 2. Représentation de la synthèse des protéines. La séquence d'ADN est transcrite en ARN messenger. L'ARN messenger est ensuite traduit en protéine.

Le génome humain contient environ 25 000 gènes codant pour des protéines, représentant ainsi moins de 3% de la transcription d'une cellule (3). Les 97% restants sont un territoire largement inexploré avec seulement une petite fraction caractérisée. En effet, de nouvelles unités de transcription, dans les régions dites non codantes du génome, sont de plus en plus mises en lumière et révèlent des fonctions importantes et nécessaires dans des mécanismes de régulation de la synthèse des protéines (3).

1.3 Les polymorphismes génétiques

Un polymorphisme génétique est une différence observée sur un même locus chromosomique chez des individus d'une même espèce. Chacune des versions de ce locus est appelée un allèle. Généralement, les polymorphismes apparaissent dans les cellules germinales en raison notamment d'erreurs dans la réplication de l'ADN et sont donc transmissibles d'une génération à l'autre. Ces modifications peuvent se produire à l'échelle d'un chromosome entier ou d'un nucléotide uniquement. Les différents allèles d'un gène peuvent soit n'avoir aucune conséquence sur la fonction de ce gène, soit en affecter la fonction selon trois modalités : perte de fonction, maintien partiel de la fonction avec interférences ou gain de fonction.

Chez les espèces diploïdes, c'est-à-dire possédant des paires de chromosomes, la combinaison des deux allèles observés à un locus particulier est appelée le génotype. Un individu diploïde dont les deux allèles d'un locus sont différents est dit hétérozygote. S'ils sont identiques, l'individu est dit homozygote pour l'allèle observé.

1.3.1 Les polymorphismes chromosomiques

Les polymorphismes chromosomiques sont des variations structurales résultant d'événements de translocation (échange réciproque de segment d'ADN entre des chromosomes non homologues), d'inversion (renversement bout à bout d'un segment du chromosome), de fusion ou de fission de fragments chromosomiques (Figure 3). Des anomalies dans le nombre de chromosomes peuvent également être observées (Figure 3). Ces variations ne sont pas nécessairement liées à des anomalies phénotypiques.

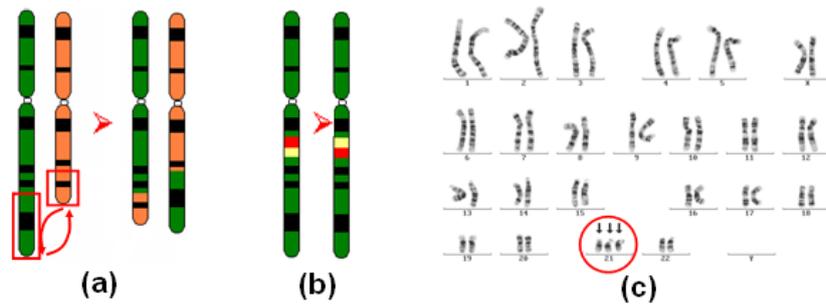


Figure 3 Exemples de polymorphismes chromosomiques. (a) Translocation entre 2 chromosomes. (b) Inversion au sein d'un même chromosome. (c) Anomalie du nombre de chromosomes : exemple de la trisomie 21.

1.3.2 Les séquences répétées en tandem

Les séquences répétées représentent 50% du génome humain. Elles peuvent être hautement ou moyennement répétitives ou encore sous forme de séquences uniques. Au sein des séquences hautement répétées, trois classes se distinguent : 1- les séquences microsatellites sont des motifs de 1 à 5 paires de bases (pb) répétés de 2 à 50 fois consécutivement. 2- les séquences minisatellites sont des motifs entre 15 et 100 pb répétés entre 15 et 50 fois. 3- les séquences 'grand bloc' satellites sont constituées de motifs plus grands (α : 171, β : 168, et γ : 220 pb respectivement) répétés un grand nombre de fois, localisés très majoritairement au niveau des centromères et des télomères.

1.3.3 Les indels

Le terme indel désigne une insertion ou une délétion d'un ou plusieurs nucléotides dans une séquence d'ADN relatif à une séquence de référence (Figure 4). La taille de l'insertion ou de la délétion est variable.

Séquence 1	...ATGCTAGATGTGCCATAACG...
Séquence 2	...ATGCTAGATGCATGCCATAACG...
Séquence 3	...ATGCTAGATGCCATAACG...

Figure 4. Exemples d'insertion/délétion. La séquence 1 fait office de référence. La séquence 2 comporte une insertion de deux nucléotides (en rouge) entre le G et le T (en bleu) par rapport à la séquence 1. La séquence 3 illustre une délétion des deux nucléotides G et T (en bleu sur la séquence 1) par rapport à la séquence de référence.

1.3.4 Les polymorphismes mono-nucléotidique

Le ‘Single Nucleotide Polymorphism’ (SNP) est la plus petite forme de polymorphisme correspondant au changement d’un seul nucléotide à un locus particulier. Ainsi les SNPs constituent la variation la plus commune des génomes et explique près de 90% de la diversité génétique observée entre des individus. D’après le projet 1000 Genomes (4), environ 40 millions de SNPs ont déjà été identifiés sur l’ensemble du génome. Leur caractérisation étant simple, les SNPs sont devenus des marqueurs génétiques privilégiés. Ainsi, des cartographies denses et précises de ces polymorphismes ont été développées ces dernières années, avec comme exemple la base de données dbSNP (5).

Séquence 1	...ATGCTAGATGTGCCATAACG...
Séquence 2	...ATGCTAGATCTGCCATAACG...

Figure 5. Exemple de SNP. Le nucléotide G (en bleu) de la séquence 1 est remplacé par un C (en bleu) dans la séquence 2.

1.3.5 Les Copy Number Variations

Les Copy Number Variations (CNVs) sont des polymorphismes découverts récemment et représentent près de 12% du génome humain. Ils sont définis comme étant des séquences d’ADN d’une longueur supérieure à 1 kilobase (kb), soit 1000 nucléotides, répétées un nombre de fois variable d’un individu à l’autre. Les CNVs sont le résultat d’événements d’insertion, délétion et duplication. Ce nouveau type de polymorphisme suscite un grand intérêt au sein de la communauté scientifique et ouvre de nouvelles perspectives en génétique épidémiologique puisqu’il peut modifier le niveau d’expression d’un gène et être à l’origine de pathologies (6).

1.4 Notions de génétique des populations

1.4.1 Modèle de Hardy-Weinberg

En 1908, indépendamment, Hardy - un mathématicien anglais - et Weinberg - un médecin allemand - démontraient la loi qui porte leur nom et qui constitue la base de toute la génétique des populations. Cette démonstration, sous certaines conditions, enseigne que les fréquences des allèles et des génotypes d’un polymorphisme bi-allélique au sein d’une population restent

constantes au cours du temps (7, 8). Ce résultat est connu comme étant l'équilibre de Hardy-Weinberg.

Lorsque cet équilibre n'est pas respecté et qu'une déviation significative est observée, il convient d'en comprendre les causes. Elles peuvent être, entre autres, démographiques (migration de population, réduction importante et rapide de l'effectif d'une population, écart à la panmixie) ou liées à des événements de sélection naturelle.

1.4.1.1 Enoncé de l'équilibre de Hardy-Weinberg

Considérons une population diploïde P dont la reproduction est sexuée et un polymorphisme S bi-allélique dont les allèles sont notés A et a. Notons respectivement f_A et f_a les fréquences de l'allèle A et de l'allèle a à la génération t .

Si les hypothèses suivantes sont vérifiées :

- La population P est d'effectif infini ou très grand.
- Les générations sont non chevauchantes.
- La panmixie est respectée dans P : les individus de la génération $t+1$ sont obtenus par des croisements aléatoires entre les individus de la génération t , donc par tirages aléatoires de deux chromosomes dans ceux de P à la génération t .
- Il n'y a pas de nouvelle mutation au locus étudié, de phénomène de sélection naturelle et de migration.

alors, la fréquence des génotypes à la génération $t+1$ est donnée par :

$$\begin{cases} f_{AA} = f_A^2 \\ f_{Aa} = 2f_A f_a \\ f_{aa} = f_a^2 \end{cases}$$

où f_{AA}, f_{Aa} et f_{aa} désignent respectivement les fréquences des génotypes AA, Aa et aa.

Etant donné que la somme des fréquences alléliques à la génération t est égale à 1, on vérifie facilement que la somme des fréquences génotypiques à la génération $t+1$ est égale à 1. Enfin, si g_A et g_a désignent respectivement les fréquences de A et a à la génération $t+1$:

$$\begin{cases} g_A = f_{AA} + \frac{1}{2}f_{Aa} = f_A^2 + f_A f_a = f_A(f_A + f_a) = f_A \\ g_a = f_{aa} + \frac{1}{2}f_{Aa} = f_a^2 + f_A f_a = f_a(f_A + f_a) = f_a \end{cases}$$

Par récurrence, on en déduit que les fréquences des allèles A et a restent inchangées au cours des générations.

1.4.2 Déséquilibre gamétique et déséquilibre de liaison

L'étude de la composition génétique d'une population et de son évolution pour un seul gène est très restrictive, d'autant plus que de nombreux caractères ou de nombreuses pathologies se révèlent multigénique.

L'étude de l'évolution simultanée de plusieurs gènes devient rapidement si complexe qu'elle ne permet pas de mise en équation comme pour l'équilibre de Hardy-Weinberg, mais s'avère d'un grand intérêt. En effet, l'analyse de cette situation permet d'introduire le concept de déséquilibre gamétique, important pour élaborer une cartographie des gènes, analyser l'origine de certaines mutations mais également pour l'élaboration de nouveaux marqueurs diagnostiques de risque génétique.

1.4.2.1 Définition du déséquilibre gamétique

Lorsque l'étude de la diversité génétique se focalise sur un gène, la fréquence d'un allèle est égale à la fréquence du gamète portant cet allèle. Cette égalité ne tient plus quand deux gènes ou plus sont étudiés simultanément.

Considérons :

- un gène A dont les allèles sont A_1 de fréquence p et A_2 de fréquence q
- un gène B dont les allèles sont B_1 de fréquence u et B_2 de fréquence v

Quatre types de gamètes différents portant chacun une combinaison d'allèles de chaque gène sont potentiellement observables :

- le gamète (A_1, B_1) de fréquence f_{11}
- le gamète (A_1, B_2) de fréquence f_{12}
- le gamète (A_2, B_1) de fréquence f_{21}
- le gamète (A_2, B_2) de fréquence f_{22}

Nécessairement, il existe une relation entre les fréquences gamétiques et les fréquences alléliques de chaque gène, bien que celle-ci ne soit pas évidente.

Supposons que les allèles des deux gènes soient « réunis » indépendamment les uns les autres et aléatoirement dans les gamètes. Cette hypothèse conduit alors à une situation appelée équilibre gamétique et les fréquences de chaque gamète sont égales au produit des fréquences des allèles qu'il porte :

$$\begin{cases} f_{11} = pu \\ f_{12} = pv \\ f_{21} = qu \\ f_{22} = qv \end{cases}$$

Toutefois, même si la population est à l'équilibre de Hardy-Weinberg pour chacun de ces deux gènes, cette situation d'équilibre gamétique ne revêt ni un caractère obligatoire ni un caractère courant. Le non-respect de ces égalités est appelé le déséquilibre gamétique, défini comme la différence entre la fréquence réelle d'un gamète et sa fréquence théorique à l'équilibre :

$$\begin{array}{ll} \text{pour } (A_1, B_1) & \Delta_{11} = f_{11} - pu \\ (A_1, B_2) & \Delta_{12} = f_{12} - pv \\ (A_2, B_1) & \Delta_{21} = f_{21} - qu \\ (A_2, B_2) & \Delta_{22} = f_{22} - qv \end{array}$$

Tous les mécanismes supposés inexistant dans le modèle d'Hardy-Weinberg, tels que les migrations de population ou les mutations, peuvent être à l'origine d'un déséquilibre gamétique.

1.4.2.2 Evolution du déséquilibre gamétique

Lorsqu'il existe un déséquilibre gamétique dans une population, le mode de reproduction panmictique associé aux recombinaisons intergéniques qui se produisent au moment de la méiose tend à faire diminuer ce déséquilibre. Ce phénomène de recombinaison résulte de la ségrégation indépendante des loci portés par des chromosomes différents ou des crossing-over qui se produisent entre loci d'un même chromosome.

En utilisant les mêmes notations qu'au paragraphe précédent, il est possible de quantifier le temps nécessaire à la disparition du déséquilibre gamétique entre les gènes A et B en fonction du taux de recombinaison r entre ces deux gènes.

Notons $f_{11, i-1}$ la fréquence du gamète (A_1, B_1) à la génération $i-1$. A la génération i , ces gamètes seront issus de deux phénomènes :

- les gamètes (A_1, B_1) de la génération $i-1$ qui n'ont pas recombiné. La probabilité d'occurrence de cet événement est $(1 - r)$.
- les nouveaux gamètes (A_1, B_1) issus de recombinaisons associant l'allèle A_1 (de fréquence p) et l'allèle B_1 (de fréquence u).

La fréquence de ce gamète à la génération i s'écrit alors :

$$f_{11,i} = (1 - r)f_{11,i-1} + rpu$$

Cette équation, après avoir retranché pu à ses deux membres, équivaut à :

$$\Delta_{11,i} = (1 - r)\Delta_{11,i-1}$$

d'où,

$$\Delta_{11,i} = (1 - r)^i \Delta_0$$

Cette dernière équation prouve que le déséquilibre gamétique tend vers 0 avec le temps. La vitesse de cette décroissance dépend seulement de r et donc de la liaison génétique, existante ou non entre les gènes A et B. Ainsi, si les gènes ne sont pas liés, ce déséquilibre gamétique va disparaître très rapidement. En revanche, si les gènes sont fortement liés génétiquement (r faible), le déséquilibre pourra perdurer au cours du temps (Figure 6). Le complexe majeur d'histocompatibilité est un exemple de déséquilibre persistant au fil des générations.

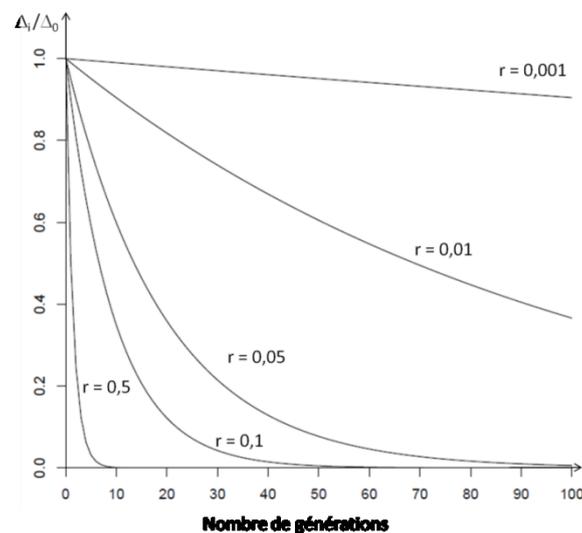


Figure 6. Vitesse de disparition du déséquilibre gamétique pour différentes valeurs du taux de recombinaison entre deux loci. Plus le taux de recombinaison est grand, plus le déséquilibre disparaît rapidement.

1.4.2.3 Déséquilibre de liaison

Un déséquilibre gamétique entre deux loci peut perdurer en raison d'une importante liaison génétique entre ceux-ci. Les Anglo-Saxons ont alors qualifié cette situation de *linkage disequilibrium*, traduit en français par déséquilibre de liaison (LD). Ainsi, le déséquilibre de liaison entre deux loci suggère qu'il existe une liaison génétique et un déséquilibre génétique.

Cependant, cette formule est souvent employée à tort. En effet, une liaison génétique entre deux gènes peut exister sans qu'un déséquilibre gamétique ne puisse être observé, et inversement. L'observation d'un déséquilibre gamétique n'est pas nécessairement l'indication d'une liaison génétique. Cette confusion est fréquente dans de nombreuses études d'épidémiologie génétique.

Dans la suite de cette thèse, par souci de simplicité, nous assumons volontairement la confusion entre ces deux notions. De plus, nous travaillons sur de petites régions chromosomiques dans lesquelles la liaison génétique semble pertinente.

1.4.2.4 Les différentes mesures du déséquilibre de liaison

Plusieurs mesures du déséquilibre de liaison ont été développées à partir de la comparaison des fréquences alléliques et des fréquences des couples d'allèles. En utilisant les mêmes notations que précédemment, la première mesure introduite est notée D (9):

$$\begin{cases} f_{11} = pu + D \\ f_{12} = pv - D \\ f_{21} = qu - D \\ f_{22} = qv + D \end{cases}$$

L'équilibre gamétique se traduit par un coefficient D égal à 0. Cette mesure n'est que très rarement utilisée en pratique. En effet, elle est dépendante des fréquences alléliques et la comparaison de cette mesure pour deux couples de polymorphismes n'est pas aisée. Afin de pallier cet inconvénient, une mesure normalisée, notée D' , a été proposée (10):

$$D' = \begin{cases} \frac{D}{\min(pu, qv)} & \text{pour } D \geq 0 \\ \frac{D}{\min(pv, qu)} & \text{pour } D \leq 0 \end{cases}$$

Ce coefficient D' varie entre -1 et 1. Lorsqu'il vaut 0, l'équilibre de liaison est vérifié. Quand il est égal à 1 ou -1, cela signifie qu'une ou deux combinaisons d'allèles ne sont pas observées au sein de la population. Cependant, cette mesure est encore une fois dépendante des

fréquences. La mesure la plus utilisée permettant de s'affranchir de cette dépendance aux fréquences est le r^2 (11):

$$r^2 = \frac{D^2}{pquv}$$

Ce coefficient varie de 0 à 1 : un r^2 égal à 0 indique une situation d'équilibre alors qu'un r^2 égal à 1 indique une situation de déséquilibre de liaison total. Le déséquilibre de liaison total désigne la situation où les allèles des deux SNPs sont parfaitement corrélés et systématiquement co-transmis. Dans ce cas, la connaissance du génotype d'un SNP détermine totalement le génotype de l'autre SNP.

1.4.3 Haplotypes

Un haplotype est défini comme la combinaison d'allèles de deux SNPs ou plus sur le même chromosome (Figure 7)

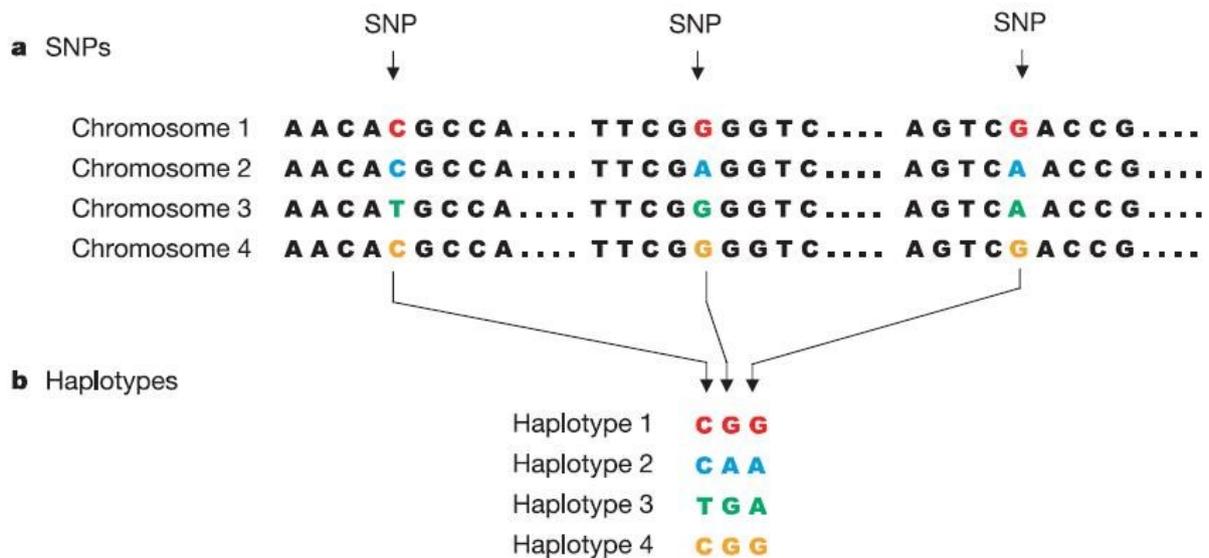


Figure 7. Exemple d'haplotypes composés de 3 SNPs. Adapté de (12).

La notion de déséquilibre de liaison est adaptée pour l'étude simultanée de deux loci mais n'est pas la plus pertinente lorsqu'une région chromosomique contenant plus de deux polymorphismes est considérée. Il est naturellement possible d'évaluer le LD en considérant les paires de SNPs de la région mais cette méthode ne permet pas de capturer pleinement la structure complexe des corrélations entre les allèles de tous les SNPs de la région. Dans une telle situation, l'étude des haplotypes est privilégiée.

Généralement, les haplotypes sont créés par une succession de mutations dont les combinaisons d'allèles sont brassées ou non par différentes recombinaisons (Figure 8). La disparition ou non de ces haplotypes dépend de divers paramètres tels que la dérive génétique, la sélection naturelle ou les migrations.

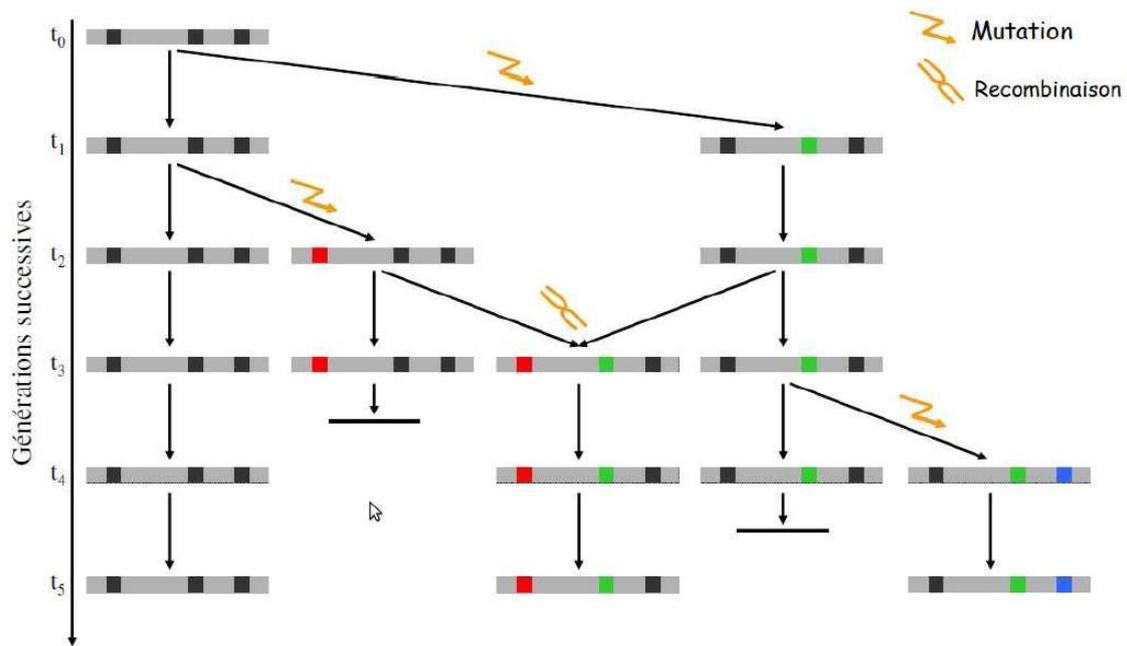


Figure 8. Exemple de genèse d'haplotypes dans une région de 3 SNPs

Le nombre de combinaisons observables croît avec le nombre de SNPs considérés et dépend également des taux de recombinaison de la région chromosomique. En effet, en s'intéressant à n SNPs, il est possible d'observer jusqu'à 2^n haplotypes et de forts taux de recombinaison dans la région étudiée favorisent l'apparition de nouvelles combinaisons. Cependant, des études ont montré que les taux de recombinaison ne sont pas uniformes le long du génome et que les forts taux de recombinaison se concentrent généralement dans de petites régions séparées par des zones plus ou moins grandes dans lesquelles ces taux sont très faibles (13, 14). Ces variations dans la distribution des taux de recombinaison expliquent la présence de « blocs » d'haplotypes. Au sein de ces blocs, le LD entre les SNPs est fort et la diversité haplotypique est relativement limitée du fait de l'absence de recombinaisons (15). Au contraire, les déséquilibres de liaison entre deux blocs sont faibles et la diversité haplotypique est plus importante à cheval sur plusieurs blocs puisque les recombinaisons brassent les combinaisons d'allèles.

La notion d'haplotype est utilisée pour l'étude de nombreux phénomènes et notamment l'évolution démographique de l'espèce humaine. En effet, leur analyse permet de retracer l'histoire migratoire des populations ou de repérer des événements de sélection naturelle. Les

haplotypes ont également été utilisés en génétique épidémiologique afin de détecter des combinaisons d'allèles impliquées dans la susceptibilité à certaines pathologies (16, 17).

1.4.4 Structure des populations

En génétique, la stratification en sous-populations est due aux migrations ancestrales sur plusieurs générations. La structure des populations est basée sur les différences de fréquences alléliques entre des sous-populations dont l'origine ancestrale diffère. Ces différences peuvent être observées à l'échelle des continents mais aussi à celle de pays voisins comme en Europe (Figure 9). Lors d'une analyse génétique d'une population, la stratification doit être prise en compte car elle peut devenir un facteur de confusion avec pour conséquence des associations observées n'étant pas liées au phénotype étudié mais à une structure migratoire différente au sein de la population analysée.

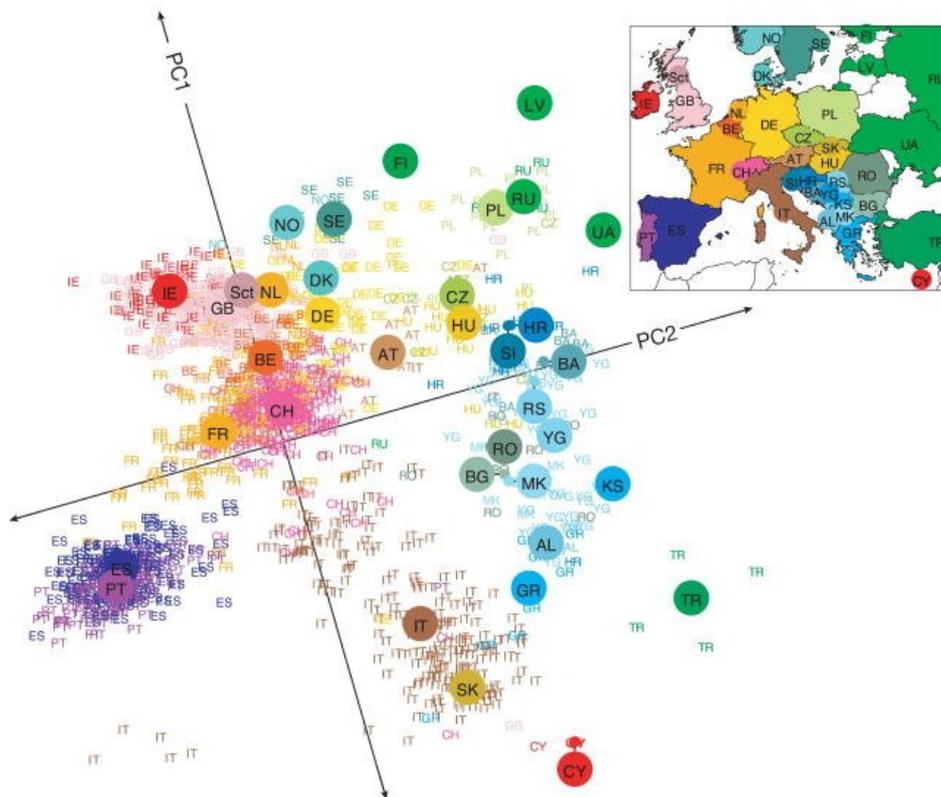


Figure 9. Résumé statistique de données génétiques de 1387 individus Européens grâce à une analyse en composante principale (ACP). Chaque label de couleur représente les individus des pays et les ronds pleins avec label sont la médiane des axes PC1 et PC2 des individus par pays. La carte en haut à droite permet de relier les labels des pays (18).

1.5 Epidémiologie génétique

Les avancées technologiques en génétique moléculaire au cours de la dernière décennie ont ouvert une nouvelle ère pour les études génétiques, plus particulièrement chez l'homme, et de nombreuses données sur les variations de l'ADN se sont accumulées. Ces études permettent d'appréhender la compréhension des maladies, grâce à l'analyse de leurs transmissions au sein des populations et des familles. Cet aspect de la génétique est nommé génétique épidémiologique et joue un rôle important dans le processus de mise en cause d'un ou plusieurs gènes dans une maladie. L'objectif principal de ces études est, à court terme, de comprendre des mécanismes de biologie fondamentale mais aussi de développer et mettre en œuvre des outils préventifs, diagnostiques et thérapeutiques sur le long terme.

1.5.1 Polymorphismes génétiques et pathologies

De nombreux polymorphismes génétiques ont été associés à des maladies. En effet, un polymorphisme peut changer l'information portée par un gène. Par conséquent, l'expression ou encore la structure des protéines sont impactées engendrant des pertes totales ou partielles de fonctions biologiques. Pas moins de 6000 maladies seraient d'origine génétique. Parmi elles, on peut citer des maladies autosomiques récessives, donc transmises par les deux parents à la fois. C'est notamment le cas de la mucoviscidose et la drépanocytose, causées par une mutation des deux copies d'un gène, respectivement, CFTR (19) ou HBB (20) et de la maladie de Huntington, associée à des répétitions en grand nombre d'un microsatellite situé dans le gène HTT (21).

Cependant, de nombreuses maladies communes comme les cancers, les maladies auto-immunes ou les myopathies, ne sont pas causées par une mutation génétique seule. Dans le cas des myopathies, il existe environ une centaine de gènes impliqués, tous importants pour la fabrication et le fonctionnement des muscles (22) et chaque type de myopathie correspond à une déficience génétique différente. Néanmoins, dans le cas des cancers, l'identification de tels gènes est difficile car ils sont nombreux, dans diverses voies biologiques et interagissent entre eux, ainsi qu'avec des facteurs environnementaux. De plus, depuis la naissance de l'épigénétique, science qui permet d'étudier la couche régulatrice de l'expression des gènes par modification des protéines de structure de l'ADN, il semblerait que certaines maladies génétiques comme le cancer soit aggravées par ces modifications. Cela complexifie les phénotypes à analyser et la découverte des composantes génétiques (23, 24).

Depuis la fin du XX^{ème} siècle de nombreuses maladies infectieuses ont émergé et continueront de le faire, représentant pour l'homme un de ses principaux fléaux. Dès le XIX^{ème} siècle, il a été observé une grande variabilité de résistance à ces maladies selon les individus, la même pathologie pouvant être mortelle chez les uns et bénigne ou asymptomatique chez les autres. Près d'un quart des décès dans le monde, soit près de quinze millions par an, est encore directement lié aux maladies infectieuses (25). À partir de 1930, de nombreuses études épidémiologiques ont montré que les facteurs génétiques de l'hôte jouent un rôle majeur dans la susceptibilité ou la résistance aux infections.

1.5.2 Types d'études génétiques et génomique

Différents types d'études ont permis la découverte de facteurs génétiques impliqués dans des pathologies. La conception d'une telle étude doit prendre en compte deux paramètres : la population considérée et les régions chromosomiques étudiées. De même, la nature du caractère étudié, appelée phénotype, conduit à des analyses différentes.

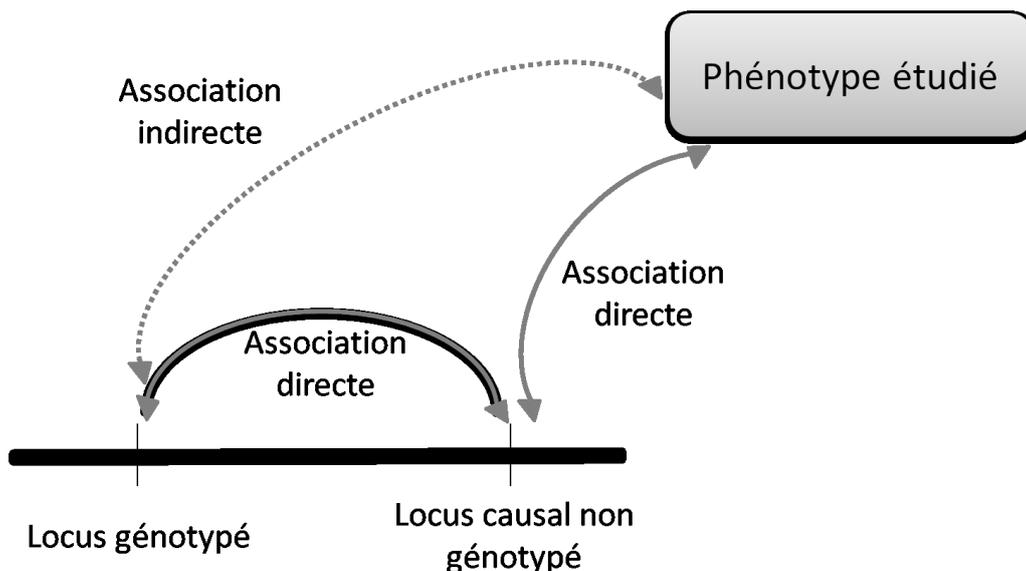


Figure 10. Découverte d'un variant associé à un phénotype. Le variant détecté lors d'une étude de génétique épidémiologique n'est pas nécessairement le variant causal de la maladie mais peut être en déséquilibre avec celui-ci. Adapté de (26).

Bien que les résultats obtenus avec ces analyses aient permis la découverte de nombreuses liaisons ou associations avec des maladies, il est nécessaire de garder à l'esprit que les polymorphismes mis en évidence lors des analyses de génétique ne sont pas nécessairement les variants causaux de la maladie. Ceci peut s'expliquer par exemple par la non inclusion du

variant causal dans l'étude ou parce que le polymorphisme causal est d'une nature différente des marqueurs étudiés. Ces analyses peuvent mettre en exergue une association entre la maladie et un variant en déséquilibre de liaison avec le variant causal (Figure 10). Ainsi, les conclusions de telles études doivent être prudentes et prendre en compte cette possibilité.

1.5.2.1 Etudes de liaison

Les études de liaison s'intéressent à la co-ségrégation des allèles dans les cas d'héritabilité d'une maladie au cours des générations au sein des familles (Figure 11).

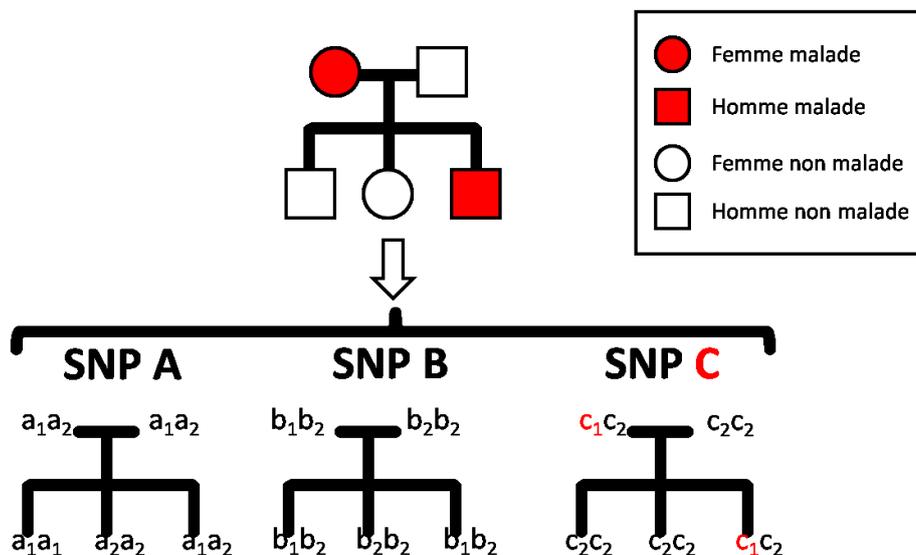


Figure 11. Exemple de transmission d'une maladie par étude de 3 SNPs (A, B et C). Arbre généalogique incluant, en rouge, les individus malades et en blanc, les individus sains. Le SNP C, par l'allèle c_1 , est présent uniquement chez les individus malades, suggérant que le locus est impliqué dans l'apparition de la maladie.

Ce type d'étude familiale a notamment permis la découverte de facteurs génétiques responsables de maladies monogéniques telles que la mucoviscidose (27, 28) ou la maladie de Huntington (29-31). Ces études ont abouti à des résultats intéressants concernant des maladies infectieuses telles que la lèpre (32). Elles restent cependant limitées pour la détection des facteurs génétiques impliqués dans les maladies multifactorielles, comme le SIDA par exemple, pour lesquelles chaque facteur n'explique qu'une fraction du phénotype et les corrélations familiales sont plus difficiles à caractériser. De plus, ces pathologies ne touchent pas nécessairement plusieurs membres d'une même famille et la collecte d'information familiale perd donc de son intérêt. Pour ce type de maladies, les généticiens ont recours aux études d'association.

1.5.2.2 Etudes d'association

Les études d'associations visent à déterminer la répartition des allèles en fonction du phénotype étudié au sein de populations de sujets sans relation de parenté. Les méthodes visant à calculer ces répartitions dépendent de la nature du phénotype étudié, et se divisent en études transversales et longitudinales. Les études transversales se focalisent, classiquement, sur un phénotype qualitatif dichotomique. Ces études sont généralement composées d'un groupe 'cas' d'individus atteints par une pathologie et d'un groupe 'témoins' d'individus sains, On appelle ces études 'cas/témoins' (Figure 12). Néanmoins, le phénotype peut être quantitatif reflétant, à un temps t , l'analyse d'un critère d'une pathologie (*e.g.* hématokrite pour la drépanocytose, glycémie pour le diabète). Les études longitudinales sont composées de cohortes de patients dont l'évolution d'une pathologie est suivie au cours du temps.

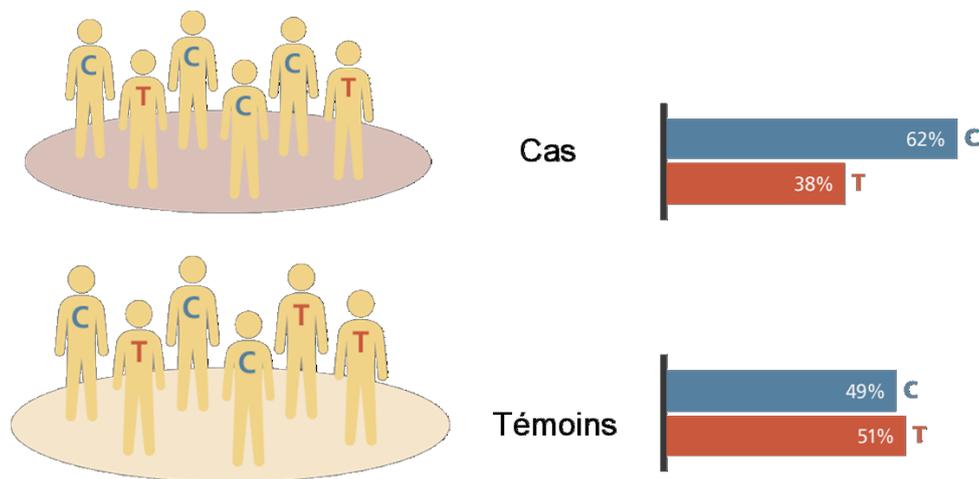


Figure 12. Représentation schématique d'une association génétique par étude 'cas/témoins' entre un SNP et une pathologie présente chez les cas. La significativité est évaluée par la différence de répartition des allèles entre les cas et les témoins.

Dans les deux cas, l'association entre un allèle et le phénotype se fait par une analyse statistique (*e.g.* test exact de Fisher, régression logistique ou linéaire pour les études transversales, et régression linéaire ou Kaplan-Meier pour les études longitudinales). La significativité d'une association est évaluée par le calcul statistique d'une p -valeur, qui représente la probabilité que le résultat observé soit dû au hasard.

a) Approche « gène-candidat »

Les approches « gène-candidat » consistent à sélectionner un ensemble de gènes pertinents pour intervenir dans l'étiologie de la maladie étudiée (33) et à évaluer leurs polymorphismes directement par association. Le choix des gènes peut être guidé par des *a priori* biologiques tels que la fonction ou l'appartenance à une voie métabolique associée à une maladie, ou encore sur la base de la localisation dans une région chromosomique d'intérêt, suggérée par une précédente étude de liaison ou d'association. Ce type d'approche repose donc sur des *a priori* et n'est pas adaptée pour explorer de manière exhaustive les causes génétiques d'une maladie. Du fait des limitations du séquençage et du génotypage, ce fut longtemps la seule source d'exploration génétique des maladies. Elle reposait donc sur une connaissance préalable de la maladie et une connaissance approfondie des mécanismes moléculaires associés. Par exemple, dans le cas du SIDA, l'approche initiale a été de génotyper des gènes de l'immunité, des gènes connus ou suspectés d'avoir un rôle dans la pathogenèse du VIH-1 (34).

b) Approche « génome entier »

Au contraire des études « gène-candidat », les approches « génome entier » ne se concentrent pas uniquement sur certains gènes mais sur l'ensemble des polymorphismes communs du génome. Ce type d'étude s'affranchit de tout *a priori* biologique et a donc pour but de repérer de nouveaux facteurs impliqués dans des maladies multifactorielles. Cependant, elles ne permettent pas obligatoirement la compréhension complète du rôle joué par les marqueurs repérés dans la pathogenèse de la maladie. Des études plus spécifiques (et notamment de type « gène candidat ») sont ensuite nécessaires pour améliorer cette compréhension.

L'amélioration des techniques de génotypage a largement contribué au développement de cette stratégie. Ainsi, plusieurs milliers d'études « génome entier » ont été réalisées au cours de la dernière décennie, comme le démontre la base de données GWAS Catalog recensant les résultats de ce type d'analyses (35). Les études d'association « génome entier » (Genome Wide Association Study ou GWAS) seront détaillées plus précisément dans la partie Matériels et Méthodes.

2 LE SYNDROME D'IMMUNO-DEFICIENCE ACQUISE

2.1 La maladie

Les premiers cas de Syndrome d'Immuno-DéficiencE AcquisE (SIDA) ont été décrits en 1981 aux Etats-Unis par le Docteur Michael Gottlieb. Il observa des symptômes rares chez quatre jeunes homosexuels de Los Angeles reliés à des maladies tels que des pneumocystoses et candidoses buccales (36). Ces symptômes sont associés à une perturbation du système immunitaire avec des manifestations comme un amaigrissement et de la fièvre. Dès 1982, les examens biologiques révélèrent, chez l'ensemble des patients, une immunodépression caractérisée par une chute significative du taux de lymphocytes T CD4⁺. Cette caractéristique donna naissance à l'appellation actuelle de 'syndrome d'immuno-déficiencE acquisE' ou SIDA.

Le virus du SIDA fut isolé pour la première fois en 1983 par l'équipe du Pr Montagnier à partir de lymphocytes T issus d'un ganglion de patient atteint de lymphadénopathie généralisée (37). Ce virus a été initialement dénommé LAV (LymphAdenopathy Virus) mais le terme VIH-1 (Virus de l'Immuno-déficiencE 1 Humaine) sera adopté par la communauté scientifique en 1986. Le test de dépistage sérologique mis au point par l'équipe du Pr Gallo en 1984 a permis d'établir que le VIH-1 était l'agent étiologique du SIDA, en démontrant que tous les patients atteints du SIDA étaient porteurs d'anticorps dirigés contre ce virus (38).

Un virus proche du VIH-1 fut découvert en 1986 par l'équipe du Pr Clavel. Ce virus, appelé VIH-2 (39), existe essentiellement à l'état endémique en Afrique de l'ouest, mais il existe aussi d'autres sites de présence sporadique ailleurs dans le monde, notamment les pays entretenant des liens commerciaux avec les pays d'Afrique de l'ouest. Très apparenté sur le plan morphologique au VIH-1, le VIH-2 est cependant décrit comme moins pathogène que le VIH-1 car il est caractérisé par un temps de latence plus long avant l'apparition du syndrome d'immuno-déficiencE, une charge virale plus faible, un taux de progression vers le SIDA plus faible, et des risques de transmission (notamment mère-enfant) plus faibles (40, 41).

2.1.1 Epidémiologie du SIDA et modes de transmission du VIH-1

Jusqu'en 2015, plus de 39 millions de morts ont été recensées, et le SIDA reste un des problèmes majeurs de santé publique de notre époque (dernier rapport de l'ONU-SIDA, www.unaids.org, 2015). Néanmoins, l'épidémie du VIH n'affecte que très peu la population générale, en dehors de groupes dits à risque.

Fin 2014, environ 37 millions de personnes vivant avec le VIH dans le monde étaient dénombrés, dont 2 millions de nouvelles infections. Il existe trois voies d'entrée du VIH : la voie sexuelle, la voie sanguine et la transmission verticale. Le SIDA est une maladie sexuellement transmissible et c'est d'ailleurs la principale voie de contamination à travers le monde. La découverte, dès 1982, de cas de SIDA chez des héroïnomanes par injection intraveineuse et chez des patients hémophiles a permis d'identifier le mode de transmission par voie sanguine. Le VIH peut être également transmis par la mère à son enfant pendant la grossesse, le travail et l'accouchement (transmission périnatale) ou par l'allaitement, c'est ce qu'on appelle la transmission verticale.

2.1.2 Evolution clinique et biologique

La progression de l'infection par le VIH-1 peut être suivie à l'aide de deux indicateurs biologiques évoluant en sens opposé : la charge virale et le nombre de lymphocytes T CD4⁺ dans le sang. Elle se distingue par trois phases successives : la primo-infection, la phase de latence clinique puis la phase symptomatique ou phase SIDA (Figure 13).

2.1.2.1 Phase de primo-infection

Elle peut être asymptomatique ou symptomatique. Lorsqu'elle est symptomatique (50% à 70% des cas), elle se manifeste par des signes généraux peu spécifiques (fièvre, pharyngite, fatigue, myalgies, courbatures, éruptions cutanées, etc...). Cependant, dans certains cas, des symptômes plus évocateurs de l'infection VIH, tels qu'une lymphodénopathie généralisée ou éruption fébrile, peuvent être constatés. Les premiers signes de primo-infection apparaissent en moyenne 20 jours après la contamination. La primo-infection peut durer 3 à 6 semaines, cette phase correspond à une multiplication intense du virus provoquant alors une chute brutale du nombre de cellules CD4⁺. Par la suite le nombre de CD4⁺ augmente et la charge virale diminue.

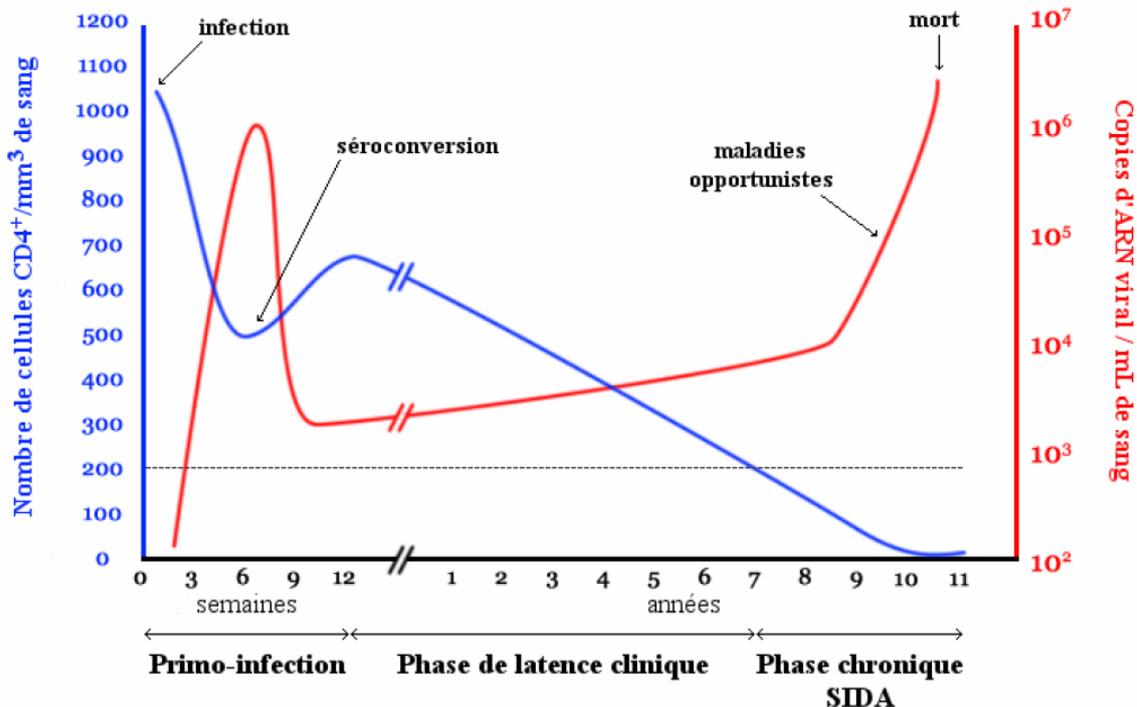


Figure 13. Profil d'évolution de l'infection par le VIH-1.

2.1.2.2 Phase de latence clinique

Une période de latence clinique suit la primo-infection, cette phase est caractérisée par un équilibre entre la production virale et son élimination. En conséquence, le sujet infecté ne présente aucun signe clinique lié à l'infection (42). La durée de cette phase est variable, pouvant aller de quelques mois à plusieurs dizaines d'années, avec une moyenne entre 7 et 10 ans sans traitement. La période précédant la phase SIDA est définie par une augmentation de la virémie et une détérioration progressive du système immunitaire attestée par la chute du taux de lymphocytes T CD4⁺. Selon la classification CDC1993 (<http://www.cdc.gov>), lorsque le seuil de 200 cellules CD4⁺/mm³ de sang est atteint, le sujet entre dans le stade symptomatique.

2.1.2.3 Phase symptomatique ou SIDA

La phase symptomatique, qui correspond à la phase SIDA à proprement parler, se manifeste principalement par des infections opportunistes sévères. La moyenne de survie des patients parvenus au stade SIDA est brève, de l'ordre de deux ans.

2.2 Le virus de l'immunodéficience humaine

Le VIH-1 est un rétrovirus appartenant à la sous-famille des lentivirus, caractérisée par une longue période d'incubation. Les rétrovirus se distinguent par la présence de la transcriptase inverse, enzyme virale responsable de la rétrotranscription (ou transcription inverse) de leur génome d'ARN en ADN. Ainsi, l'ADN viral est intégré sous forme de provirus dans le génome de la cellule hôte. Le provirus est stable et se réplique grâce à la machinerie cellulaire en même temps que l'ADN de l'hôte.

Sur le plan évolutif, les rétrovirus sont générateurs de diversité, du fait des fréquentes erreurs commises lors de la rétrotranscription (43) et de leur intégration dans le génome de l'hôte (44-46). Cette dernière étape peut conduire : (i) à l'incorporation de portions du génome de l'hôte lors de la génération de nouveaux virus, (ii) à l'altération de l'activation/inactivation des gènes situés à proximité du site d'intégration (fort pouvoir oncogène des rétrovirus) (44) et (iii) à leur cooptation complète ou partielle dans les cellules germinales hôtes (HERV, Human Endogenous Retrovirus) (45).

2.2.1 Organisation génomique et structurale du VIH-1

2.2.1.1 Structure génétique du VIH-1

Le génome du VIH-1 est petit, environ 9 kilobases, et se compose de deux molécules identiques d'ARN monocaténaire. Sa structure sous forme LTR-gag-pol-env-LTR est commune à tous les rétrovirus (Figure 14). Les gènes viraux sont au nombre de 9 et sont codant pour 15 protéines.

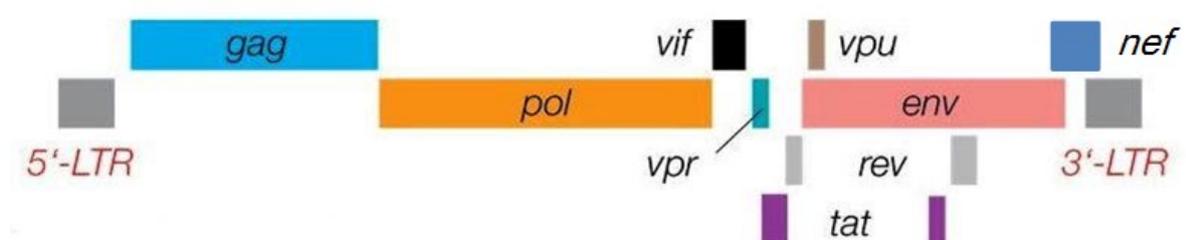


Figure 14. Structure génomique du VIH-1. En plus des trois gènes de « structure » (gag, pol et env), 6 gènes codent pour des protéines régulatrices (vif, vpr, vpu, tat, rev et nef)

Ces 9 gènes peuvent être divisés en trois classes (47) :

- les gènes codant pour les protéines de structure (*gag, pol et env*), *gag* (**g**roup-specific **a**ntigen) pour les protéines de capsid et matrice ainsi que les protéines associées à l'ARN viral, *pol* (**p**olymerase) pour les enzymes de transcriptase inverse, protéase virale et intégrase, puis *env* pour l'**e**nveloppe virale (gp120 et gp41).
- les gènes codant pour les protéines régulatrices (*tat, rev et nef*), *tat* pour **t**rans-**a**ctivateur de **t**ranscription, *rev* pour **r**égulateur de l'expression des protéines **v**irales et *nef* pour **n**egative regulation **f**actor.
- les gènes codant pour des protéines accessoires (*vpu, vpr et vif*), non indispensables à la réplication virale mais nécessaires à la maturation et au relargage des particules virales infectieuses (48).

Les deux extrémités du génome sont bordées par des séquences nucléotidiques répétées, les LTR pour Long Terminal Repeats. Ces séquences sont importantes pour la transcription du génome proviral, notamment pour la fixation de facteurs de transcription (49).

2.2.1.2 Structure protéique du VIH-1

Le virus se présente en microscopie électronique sous forme d'une particule sphérique de 80 à 120 nm, comportant une enveloppe lipidique et une nucléocapside excentrée et cylindrique. La nucléocapside contient : le génome viral, les protéines qui assurent la cohésion de la structure et les protéines dont la fonction enzymatique est nécessaire à la réplication virale (Figure 15).

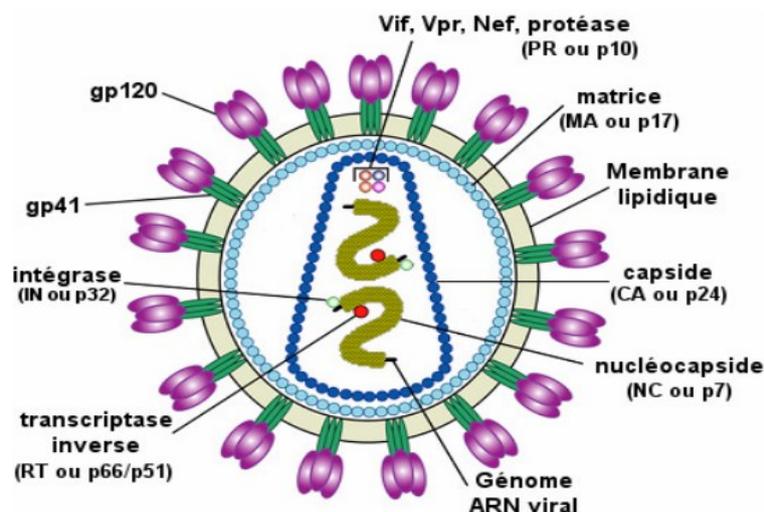


Figure 15. Structure schématique d'une particule virale VIH-1.

2.2.2 Le cycle de réplication virale

Le mode d'action du virus consiste à envahir les cellules hôtes et à réquisitionner leur machinerie afin de produire davantage de copies virales. La Figure 16 représente les étapes du cycle de réplication virale.

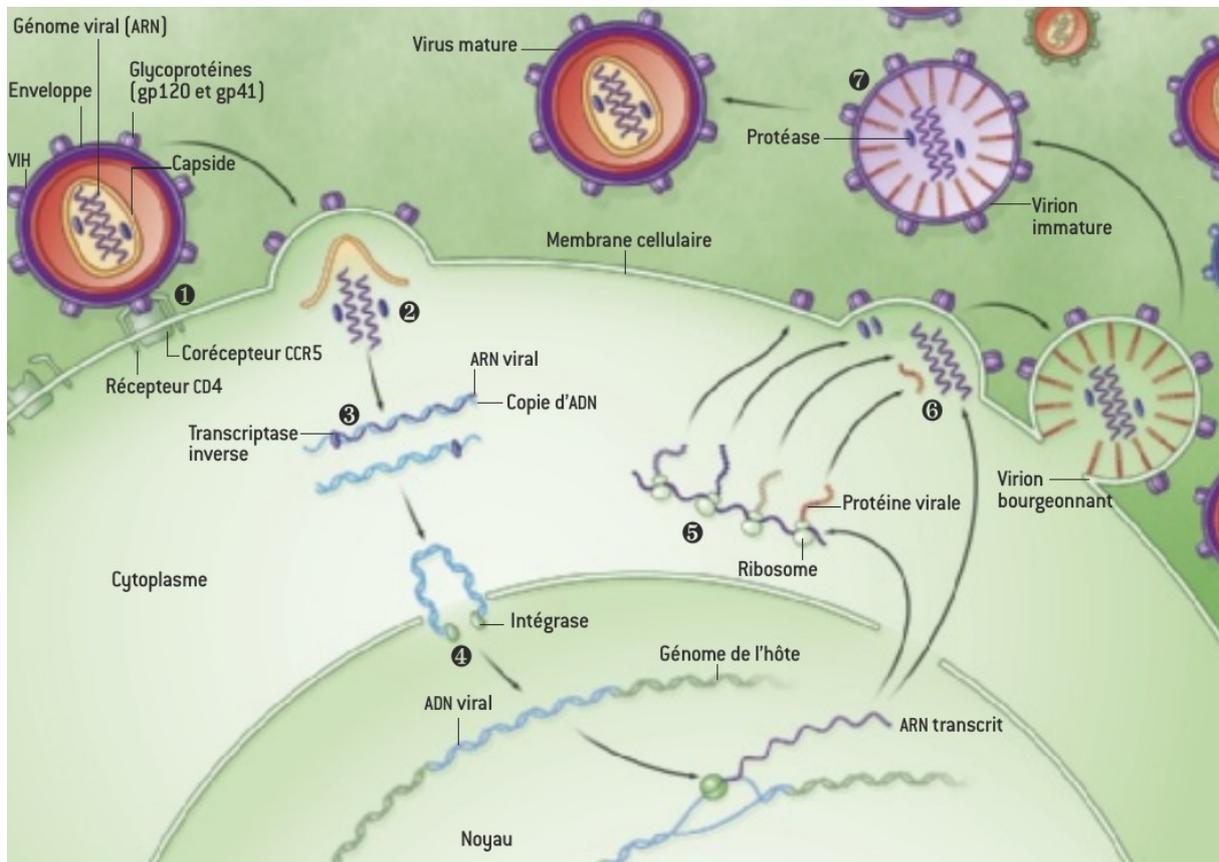


Figure 16. Cycle de réplication du VIH-1 (*Pour la science*, n°377, mars 2009).

La première étape (1) permet au virus de se fixer aux récepteurs de surfaces CD4 et CCR5 de la cellule hôte, grâce à ses glycoprotéines de surface gp120 et gp41. Dans un deuxième temps (2), la fusion du virus avec la membrane de la cellule hôte permet la libération dans le cytoplasme du contenu du virus où la transcriptase inverse copie l'ARN du génome du virus en ADN double brin (3). Cette étape est hautement mutagène pour le virus et ce phénomène est responsable de la diversité génétique des futures particules virales. L'intégrase insère l'ADN viral dans l'ADN hôte (4) et grâce à la machinerie cellulaire de l'hôte, les gènes viraux sont transcrits en ARN qui sont dirigés vers le cytoplasme pour être traduits en protéines virales ou en néo-génome viral, nécessaire à la formation des néo-virus (5). Ces nouvelles molécules migrent par la suite vers la membrane cellulaire où elles s'assemblent en nouveaux virus (6). Chez ces copies immatures, les protéases du VIH permettent de cliver les polyprotéines en

protéines fonctionnelles afin de d'aboutir à des virions « matures », prêts à infecter de nouvelles cellules (7).

2.2.3 Tropisme cellulaire

Les cellules sensibles à l'infection VIH sont essentiellement celles exprimant à leur surface les molécules CD4, CCR5 et CXCR4, nécessaires à l'entrée du virus. Ces cellules incluent : i) la sous population CD4 (dont la déplétion progressive au cours du temps est caractéristique de l'infection VIH); ii) les cellules de la lignée monocytaire-macrophagique (réservoir principal du virus); iii) les cellules dendritiques; iv) les cellules de Langerhans de l'épiderme, du sang et des muqueuses. De même les lymphocytes B, T CD8⁺ et les cellules Natural Killer (NK) peuvent être infectés par le VIH, mais la démonstration *in vivo* reste à confirmer (50, 51). Cependant, l'expression en surface des molécules nécessaires à l'entrée du VIH n'implique pas forcément une infection, les adipocytes par exemple, expriment CD4, CCR5 et CXCR4 et ne sont pas sensibles à l'infection *in vivo* (52, 53). Il est à noter que le récepteur CCR5 sert de récepteur pour les souches R5 monocyto-tropiques du VIH-1 (54).

Des cellules telles que certains précurseurs hématopoïétiques, les fibroblastes et certaines cellules intestinales et nerveuses, ne présentent pas les récepteurs nécessaires à l'entrée du VIH à leur surface et s'avèrent pourtant sensibles à l'infection. Il existe donc des mécanismes de pénétration du virus dans la cellule, différents de celui mentionné précédemment. C'est notamment le cas de l'entrée dans les cellules (T et macrophages notamment) du VIH complexé avec des anticorps *via* les récepteurs du complément et le fragment Fc (55). Il faut cependant noter que ces processus d'entrée restent limités, comparés au processus médié par CD4.

2.3 Mécanismes de pathogenèse du SIDA

L'infection par le VIH entraîne un déficit de l'immunité cellulaire marqué par une baisse de la population lymphocytaire T CD4⁺, mais également par de nombreuses altérations du système immunitaire, dont les mécanismes ne sont toujours pas complètement élucidés. D'une manière générale, les mécanismes moléculaires de la physiopathogenèse du SIDA restent encore en grande partie, à ce jour, un mystère, et les hypothèses sur ces mécanismes sont multiples.

2.3.1 Effets directs

La baisse progressive de la population de lymphocytes T CD4⁺ pourrait s'expliquer par une lyse directe des cellules infectées (56, 57). Cependant alors que environ 1% des cellules CD4⁺

exprime le génome viral chez les sujets encore asymptomatiques, ce taux n'atteint que 10% chez les sujets symptomatiques. Il est donc difficile d'expliquer la déplétion des lymphocytes T CD4⁺ par le seul effet cytopathogène. Il est possible, également, que les lymphocytes T CD4⁺ infectés fusionnent avec des lymphocytes T CD4⁺ non infectés, ce qui étendrait l'effet cytopathogène, avec pour conséquence la formation de cellules géantes multinucléées ou syncytium. Une observation *in vitro* montre que les syncytiums sont rapidement lysés (58, 59), mais l'existence d'un tel mécanisme *in vivo* reste à démontrer (60).

2.3.2 Effets indirects

Les effets cytopathogènes directs seuls n'étant pas suffisants pour expliquer la baisse du taux de lymphocytes T CD4⁺, les recherches se sont élargies à des hypothèses d'effets indirects du virus qui causerait un déficit immunitaire. La lymphopénie CD4⁺ et l'épuisement du système immunitaire pourraient être la résultante de mécanismes tels que ceux décrit ci-après :

- Les cellules CD4⁺ infectées sont la cible de puissantes réponses cytotoxiques spécifiques car elles expriment à leur surface des antigènes VIH. Néanmoins, ces antigènes peuvent être solubles et adsorbés par des cellules non-infectées qui deviennent des cibles des mécanismes de cytotoxicité (61, 62). De plus, la gp120, antigène viral, possède des similitudes structurales avec les molécules HLA (*human leucocyte antigen* en anglais), et donne lieu à des réactions auto-immunes de type allogénique (63).
- L'infection par le VIH engendrerait aussi une activation chronique et pathologique du système immunitaire accompagnée d'apoptose. Les superantigènes d'origine viral sont capable d'activer les cellules T de façon non spécifique (64, 65). Cette stimulation, en produisant des cytokines pro-inflammatoires, induirait la différenciation et la mort des lymphocytes T. De plus, elle favoriserait la production de particules virales par des lymphocytes infectés, mécanismes contribuant à la déplétion progressive en lymphocytes T (66, 67). Plusieurs études ont suggéré que la pathogenèse et le stade de progression de l'infection VIH pouvaient être corrélés à une activation de l'apoptose des cellules CD4⁺ infectées mais aussi des non-infectées (effet 'bystander') (68, 69).
- Les protéines virales telles que Nef, Tat et Vpu peuvent induire une diminution de l'expression des molécules HLA de classe I (70-72) et ainsi protéger les cellules infectées de la destruction par le système immunitaire.

- Le déficit fonctionnel des CD4⁺ se caractérise par un défaut de production d'interleukine 2 (IL-2) entraînant une anergie de ces cellules. De plus, l'absence de production d'interféron gamma (IFN- γ) met en défaut la fonction auxiliaire Th1 des lymphocytes CD4⁺ (associée à une forte inflammation et à l'immunité cellulaire) alors que l'absence de sécrétion d'IL-4, IL-5, IL-6, IL-10 et IL-3 concerne les Th2 (activatrices de l'immunité humorale). L'IFN- γ joue, normalement, un rôle primordial dans les réponses antivirales et contribue à la mise en place de réponses immunitaires tout en s'opposant à la mise en place des réponses Th2. Inversement, IL-4 et IL-10, secrétées par les Th2, s'opposent à la mise en place des réponses Th1 (50). L'infection par le VIH induit un dérèglement de cette balance cytokinique Th1/Th2 favorisant les cellules Th2 dans lesquelles le VIH se réplique préférentiellement (73, 74).
- Plus récemment, il a été démontré que l'infection par le VIH donne lieu à une attaque massive des lymphocytes T CD4⁺ situés dans les muqueuses gastro-intestinales. Cette agression altère la fonction de barrière que joue l'intestin et facilite en conséquence la translocation microbienne intestinale (75), permettant la présence dans la circulation sanguine de lipopolysaccharides (LPS), constituants majeurs de la paroi de certaines bactéries, ainsi que des composés toxiques, le tout capable de stimuler très fortement le système immunitaire (76, 77). Cela pourrait expliquer l'état d'activation chronique du système immunitaire. De plus, il semblerait que les lymphocytes T CD4⁺ de l'intestin soit une cible privilégiée du VIH puisqu'ils expriment immédiatement le corécepteur CCR5 par rapport à ceux de la circulation sanguine (75).

2.4 Les traitements actuels

Depuis l'introduction de l'AZT (azidothymidine) en 1987 (78), de nombreux progrès ont été réalisés dans la thérapie antirétrovirale notamment avec l'utilisation de la trithérapie. Lors du développement de médicaments antiviraux, il est important d'interrompre le cycle viral sans tuer la cellule hôte. Ainsi, le premier médicament anti-VIH visait la rétrotranscriptase virale, la transcription inverse étant un processus absent des cellules eucaryotes. Les inhibiteurs de transcriptases virales sont séparés en deux classes : NRTI (Nucleoside Reverse Transcriptase Inhibitor) et NNRTI (Non Nucleoside Reverse Transcriptase Inhibitor). Les NRTI sont des inhibiteurs compétitifs de la transcriptase inverse, ils se fixent au niveau du site actif et sont reconnus comme des nucléotides. Ils sont donc incorporés dans l'ADN et provoquent l'arrêt de la synthèse provirale. L'AZT (Zidovudine[®]) et l'Abacavir (Ziagen[®]) font partis de cette famille

d'inhibiteurs NRTI. Les NNRTI sont des inhibiteurs non compétitifs de la transcriptase virale c'est à dire qu'ils ne ciblent pas le site actif de l'enzyme.

Par la suite, sont apparus les inhibiteurs de la protéase virale. Ces molécules anti-VIH, telles que : le Rintovir (Norvir[®]) ou l'Indinavir (Crixivan[®]), les deux premiers médicaments de cette classe mis sur le marché, miment un peptide de liaison au niveau du site actif de la protéase. Pour ces deux classes d'inhibiteurs, il a été rapidement constaté une pression de sélection, avec l'émergence de virus mutants résistants à ces molécules antivirales (entre quelques semaines et quelques mois) aboutissant à l'échec des monothérapies. Cette observation a mené au développement de thérapies combinées dès le début des années 1990 avec les trithérapies. Les trithérapies actuelles sont composées de deux NRTI combinés à un NNRTI ou à un inhibiteur de protéase.

Plus récemment, de nouvelles classes d'antirétroviraux sont apparues. Le Maraviraoc (Celsentri[®]) est une petite molécule bloquant l'interaction entre le corécepteur de chimiokine CCR5 (79) et la protéine virale gp120, qui inhibe ainsi l'infection des cellules par les virus R5. Le T20 (Fuezon[®]) bloque les changements conformationnels de la protéine gp41 nécessaire à la réalisation de l'étape de fusion entre le virus et la cellule cible. Enfin, des anti-intégrases ont vu le jour, tel que l'Isentress (Raltégravir[®]) (80).

Tous ces traitements ont permis d'améliorer la qualité et l'espérance de vie des patients infectés par le VIH. Cependant, comme évoqué précédemment, ils ne permettent pas l'éradication du virus. De plus, ces traitements sont à l'origine de nombreux effets secondaires indésirables, de virus résistants, et enfin, ces traitements sont très onéreux. Toutes ces considérations encouragent le développement de nouvelles thérapies anti-VIH, le graal étant ici bien entendu la mise en place d'un vaccin prophylactique ou curatif

2.5 Facteurs génétiques de l'hôte et SIDA

Les études génétiques contribuent à l'amélioration des connaissances sur les maladies et à une meilleure compréhension des mécanismes de pathogenèse afin de développer des stratégies visant à enrayer l'infection et/ou l'évolution de la maladie. Dans le cadre du SIDA, les premières études ont ciblé des gènes participant au cycle de la réplication virale et à l'immunité de l'hôte. Cette approche « gène candidat » a permis de décrire de nombreuses associations génétiques dans l'infection du VIH (Figure 17) (81, 82).

Un exemple de ces études est le cas du gène *CCR5*. Le récepteur de chimiokines *CCR5* est un corécepteur du VIH pour son entrée dans la cellule. Une délétion de 32 pb a été identifiée au niveau de la région codante (34, 83) : cette délétion aboutit à la synthèse d'une protéine tronquée, non fonctionnelle et non transportée en surface de la cellule. L'allèle *CCR5-Δ32* n'a été observé que dans les populations d'origine européenne. Les individus homozygotes pour la délétion ($\Delta32/\Delta32$) sont protégés de l'infection par les souches R5 et représentent ~1-2% de la population européenne. Les individus hétérozygotes $\Delta32/WT$ présentent une expression de *CCR5* en surface cellulaire diminuée, une réplication virale et une progression vers le SIDA ralenties. Un autre polymorphisme de *CCR5* a été découvert au niveau du promoteur : l'allèle *CCR5-P1* est un haplotype composé de 10 SNPs qui augmenterait l'expression de *CCR5* et qui est associé à une progression plus rapide vers le SIDA (84, 85).

Parmi les gènes de l'immunité, les gènes HLA jouent un rôle majeur dans la pathogenèse du VIH-1. En effet, les différents allèles HLA-I vont induire des présentations des antigènes viraux aux cellules T CD8⁺ différentes et initier des réponses lymphocytaires T cytotoxiques de force variable. Les molécules HLA de classe I gouvernent ainsi la réponse immunitaire aux antigènes viraux. La région HLA du chromosome 6 a donc été étudiée de façon extensive et de nombreuses associations avec la progression ont été révélées, notamment grâce aux travaux de Kaslow, Carrington, Hendel et Magierowska (86-88).

Depuis une dizaine d'années, la recherche sur le SIDA a pu bénéficier des nouvelles approches « génome entier » et s'affranchir des à priori biologiques. Une première étude par Fellay *et al.*, parue en 2007, a montré des associations génétiques entre les SNPs rs2395029 situé dans le gène *HCP5* et rs9264942 situé dans le gène *HLA-C*, et une charge virale plus faible (89). Le SNP rs3869068 situé dans le gène *ZNRD1*, quant à lui, a été trouvé associé à l'aggravation de la progression dans la maladie.

Dans le laboratoire GBA du Professeur Zagury, deux études de GWAS ont été publiées en 2009, par analyse de patients à profils de progression extrêmes vers le SIDA (90, 91). Limou *et al.* ont répliqué les principaux signaux trouvés par Fellay *et al.* grâce à l'étude de patients au profil non-progresseurs à long terme vis-à-vis de la maladie. Dans l'article de Le Clerc *et al.* de nouveaux signaux ont été observés, localisés dans des gènes tels que *PRMT6*, *SOX5* ou encore *TGFBRAP1*, en comparant des individus témoins à des progresseurs rapides vers le SIDA.

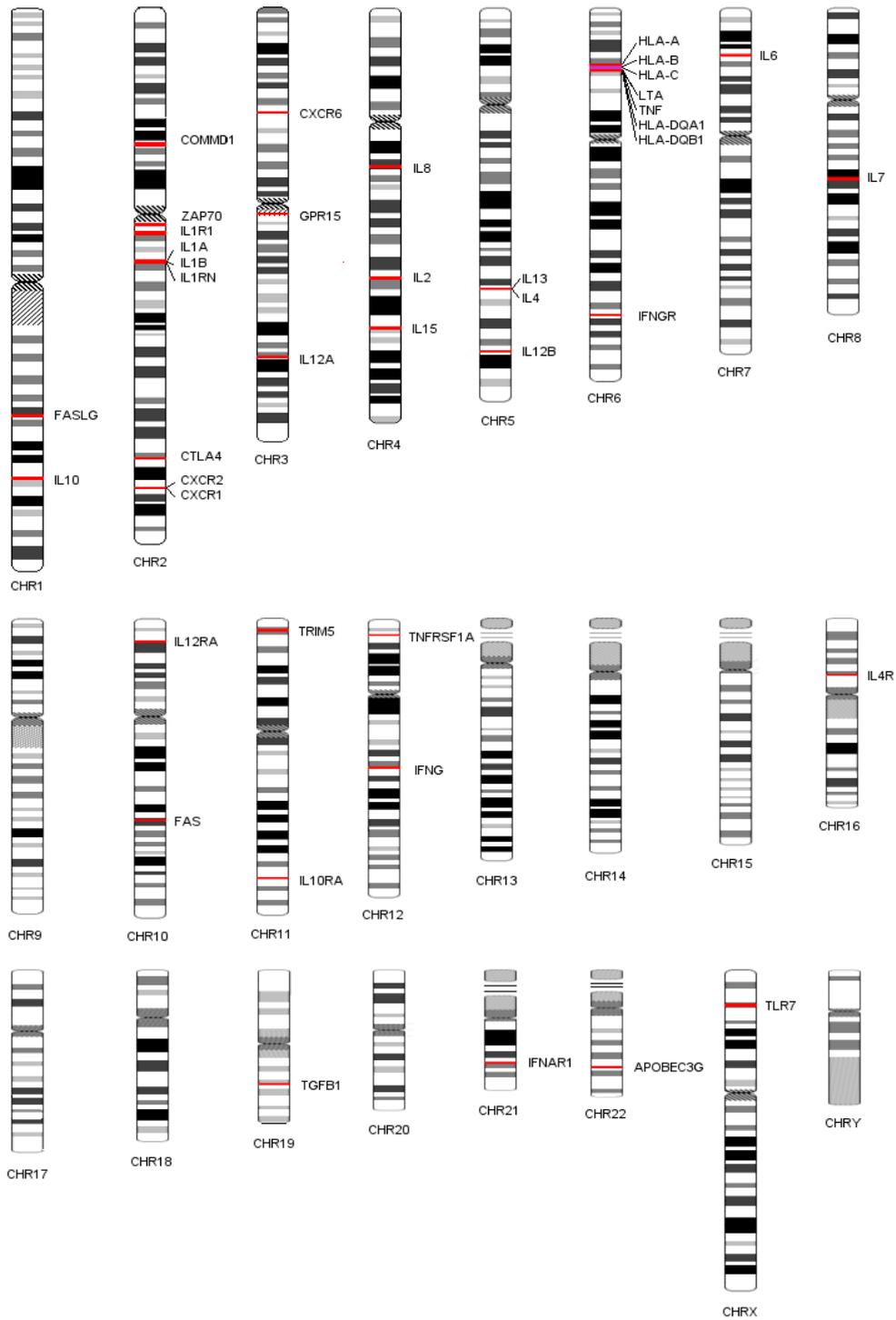


Figure 17. Liste de gènes candidats trouvés en association avec le SIDA.

3 L'HEPATITE C

3.1 La maladie

Dans les années 1960, le virus de l'hépatite A (VHA) et le virus de l'hépatite B (VHB) étaient les seuls agents pathogènes connus comme étant responsables d'hépatites virales. Entre 1970 et 1988, plusieurs cas d'hépatites furent observés chez des patients ayant reçu une transfusion sanguine. Pendant cette période, Harvey J. Alter, responsable de la section des maladies infectieuses au département de médecine transfusionnelle des National Institutes of Health (NIH), a démontré avec son équipe que la plupart des cas d'hépatite post-transfusionnelles n'étaient dus ni au VHA ni au VHB (92). Malgré cette découverte, les efforts de recherche coordonnés au niveau international pour identifier le virus responsable de cette maladie, initialement baptisée « hépatite non A non B » (NANBH en anglais), sont restés sans résultat. Ce n'est qu'en 1989, grâce à des techniques de clonages moléculaires, que le virus de l'hépatite C (VHC) fut découvert par isolement partiel et séquençage de son génome (93).

Avant la découverte de l'agent étiologique, le virus contaminait jusqu'à 90% des hémophiles transfusés de l'époque, du fait de l'insuffisance des contrôles des produits transfusés. En plus des infections post-transfusionnelles, un risque de transmission nosocomiale ou un manque d'hygiène lors de séances d'acupuncture, de tatouage ou encore de percement d'oreilles augmentaient la propagation du virus.

Après 1990, une loi concernant le contrôle des produits de transfusion réduira considérablement l'incidence de la maladie grâce à des dépistages du virus, sans toutefois l'annuler puisqu'une grande partie des donneurs de sang ne connaissaient pas leur statut de séropositif.

De nos jours, parmi les séropositifs VHC, 60 à 80 % sont des toxicomanes ou des transfusés d'avant 1990, le reste étant des personnes ayant des concentrations de transaminases anormalement élevées témoignant d'une agression hépatique. C'est pourquoi les politiques de dépistage se focalisent actuellement sur les trois T : Toxicomanes, Transfusés avant 1990 et Transaminases élevées.

3.1.1 Epidémiologie et modes de transmission du VHC

3.1.1.1 Prévalence de l'infection

Le nombre de personnes infectées par le VHC à travers le monde est estimé à plus de 170 millions et environ 500 000 patients meurent chaque année d'une maladie hépatique liée au VHC (94). Avec 3 à 4 millions de personnes nouvellement infectées chaque année, le VHC représente un problème majeur de santé publique. De manière globale (Figure 18), avec un pourcentage de 4.7%, le Moyen-Orient possède la plus haute prévalence d'hépatite C, devant l'Afrique (3.2%), l'Europe (2.3%) et les Amériques (1.5%) (95).

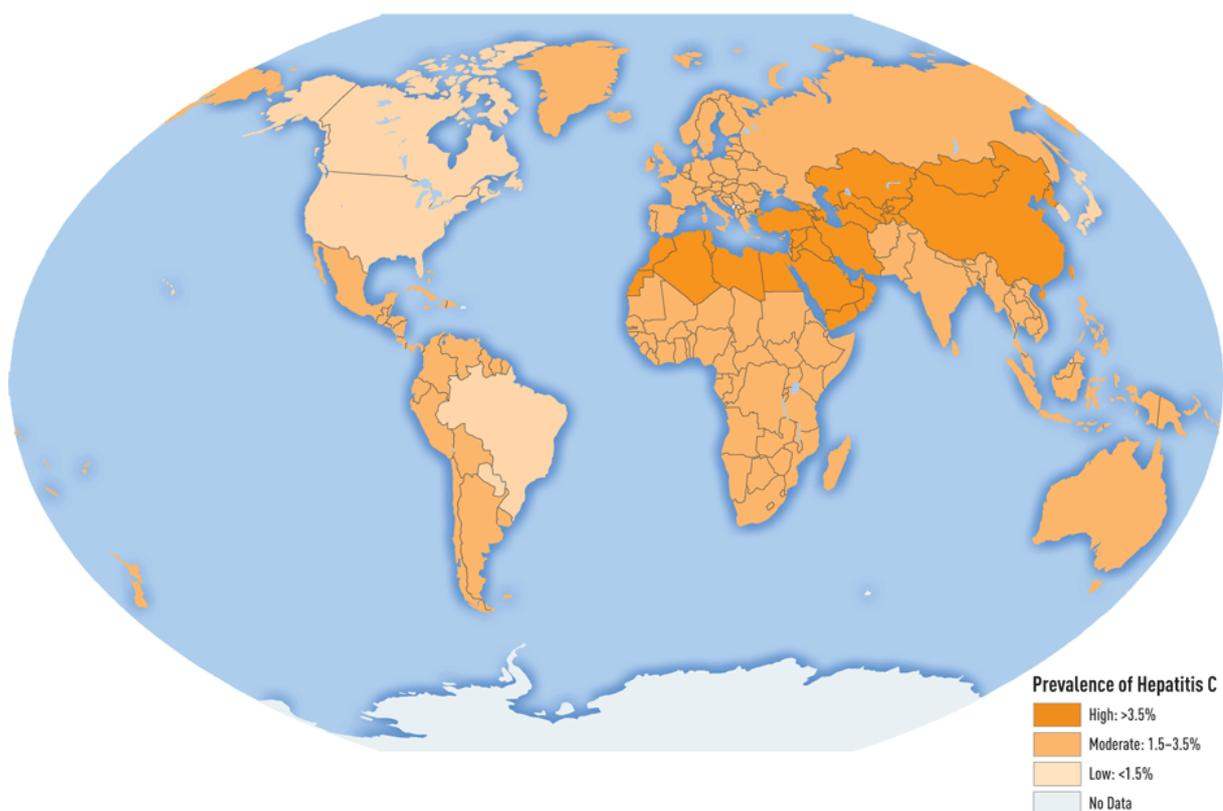


Figure 18. Prévalence du VHC dans le monde (96).

Cependant, d'importantes variabilités sont observées en fonction des pays. Ces différences peuvent s'expliquer par une augmentation de sous-populations à risque ou encore par l'absence de mise en place de mesures de prévention. Par exemple, des taux très élevés d'hépatite C sont observés en Chine à cause d'une augmentation de l'usage de drogues injectées, ainsi qu'en Egypte à cause d'une contamination en masse lors d'une campagne anti-schistosomes (95, 97). De plus, l'absence de données issues des pays africains (71%) combinée aux dons sanguins peu contrôlés et un dépistage très limité, sous-estiment probablement les chiffres réels (98).

3.1.1.2 Modes de contamination

Le VHC se transmet essentiellement par voie parentérale, avec du sang ou des produits sanguins labiles contaminés. Les deux sources majeures de contaminations sont la transfusion sanguine ou l'injection intraveineuse de drogues avec du matériel contaminé. Avant l'instauration du dépistage systématique des dons sanguins au début des années 1990, de nombreux patients ont été contaminés par le VHC lors de transfusions sanguines (99).

La toxicomanie est le mode de contamination le plus fréquent. Le partage de seringues infectées par du sang contaminé est associé à une prévalence de l'ordre de 80% chez les toxicomanes (100). Malgré les mesures mises en place, telles que le dépistage ou la distribution de seringues stériles, le risque de transmission du VHC reste important du fait d'une relative résistance du virus dans le milieu extérieur (101).

La transmission nosocomiale est également une source potentielle d'infection par le VHC. Les risques *via* les actes chirurgicaux comme des transplantations d'organe ou de tissus sont désormais extrêmement faibles grâce au dépistage systématique des donneurs (102). Néanmoins, des actes médicaux, tels que des hémodialyses, soins bucco-dentaires ou examens invasifs (endoscopies) peuvent être à l'origine d'une contamination (103).

Bien que la transmission sexuelle représente une minorité de cas (104, 105), ces dernières années le VHC a émergé comme une cause majeure de mortalité chez les patients coinfectés avec le VIH et cela non lié au SIDA (106). Ce phénomène prend de plus en plus d'ampleur dans les pays en voie de développement ainsi que chez les hommes homosexuels infectés par le VIH (107).

Enfin, la transmission du virus de la mère à l'enfant au cours de la grossesse varie d'un taux de 4 à 7% (108). Le risque de transmission dépend du niveau de la charge virale et ce risque est majoré de 2-3 fois en cas de coinfection VIH/VHC.

3.2 Le virus de l'hépatite C

3.2.1 Organisation structurale et génomique du VHC

Le VHC est classé dans le genre *Hepacivirus* de la famille des *Flaviridae*, qui comporte les genres *Flavivirus*, *Pestivirus* et *Pegivirus*. Le VHC est un virus enveloppé à ARN monocaténaire de polarité positive.

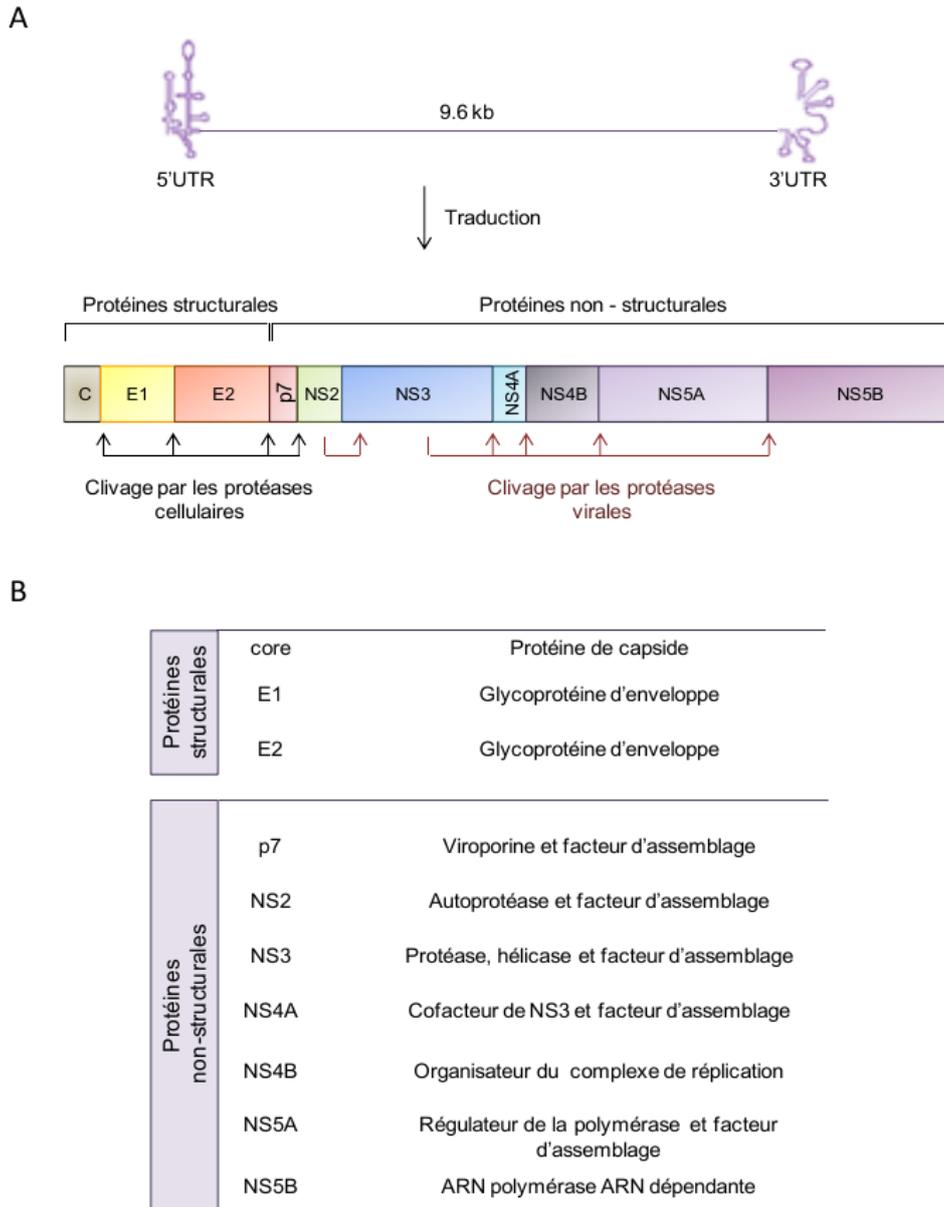


Figure 19. Le génome du VHC et ses protéines virales. A. Schéma du génome du VHC et composition de la polyprotéine traduite puis clivée en protéines structurales ou non structurales. B. Fonctions des protéines virales.

Le génome de 9,6 kb est composé d'une région 5' UTR comprenant une région IRES (*internal ribosome entry site*) permettant l'initiation de la traduction d'un ARN messager de manière interne, un cadre ouvert de lecture codant pour une polyprotéine virale et une région 3' UTR (109). Cette polyprotéine précurseur, d'environ 3 000 acides aminés, est clivée de manière co/post-traductionnelle par des protéases virales et cellulaires en protéines structurales (core,

E1 et E2) et non structurales (p7, NS2, NS3, NS4A/B et NS5A/B) (109, 110) (Figure 19A). La Figure 19B résume les fonctions des différentes protéines VHC.

La particule virale mesure entre 55 et 65 nm de diamètre (111). Les glycoprotéines E1 et E2 du VHC sont insérées dans une enveloppe lipidique dérivant de l'hôte. La nucléocapside est composée de la protéine de capsid core et de l'ARN du VHC (Figure 20A). Le virus circule en association avec des lipoprotéines, formant des lipo-viro-particules. Ces associations résultent en une importante variabilité morphologique du VHC, avec des densités hétérogènes allant de 1,03 à 1,10 g/ml (112, 113) (Figure 20B).

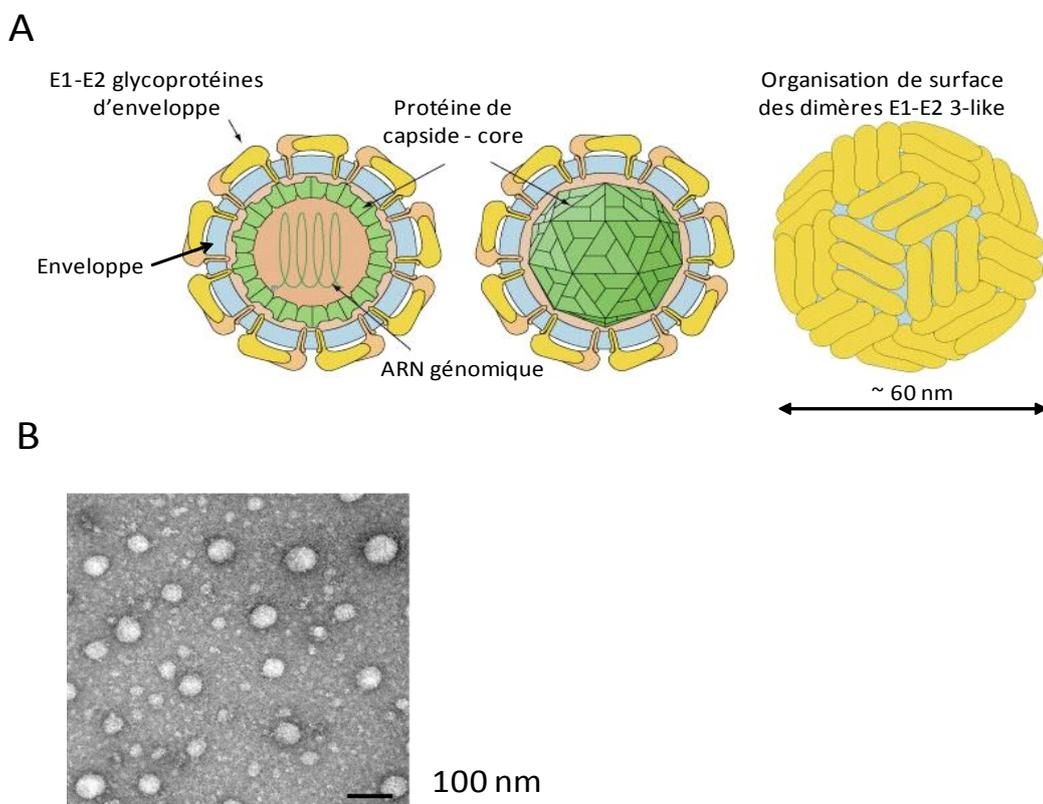


Figure 20. Représentations structurales du VHC. A. Organisation schématique du VHC. B. Image de particules virales par microscopie électronique montrant l'hétérogénéité morphologique des virions (112).

3.2.2 Cycle répliatif du VHC

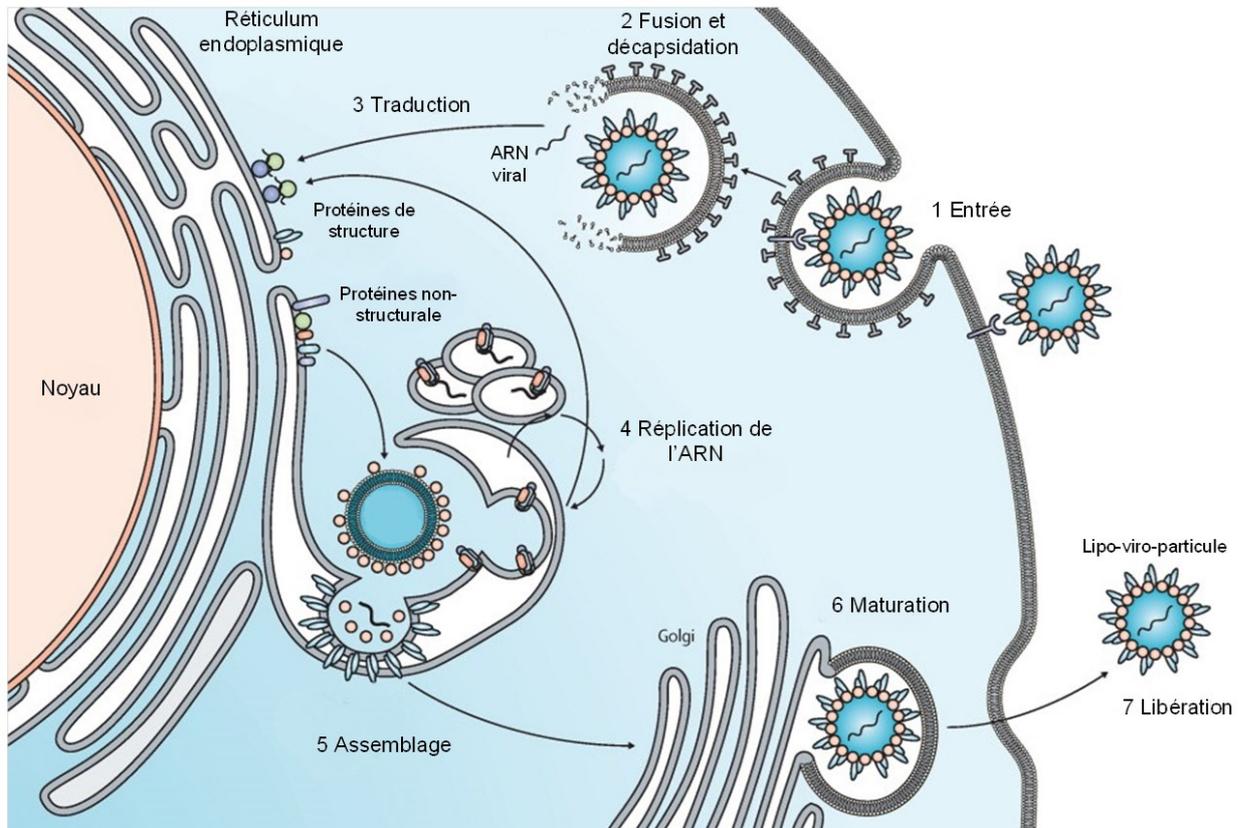


Figure 21. Représentation schématique du cycle répliatif du VHC. Adapté de (114).

3.2.2.1 Entrée du VHC dans la cellule

Le VHC se multiplie dans les hépatocytes. L'entrée du virus dans la cellule (Figure 21) se fait par un mécanisme d'endocytose grâce à des interactions entre ses protéines d'enveloppe E1 et E2 et des récepteurs des lipoprotéines de faible densité (R-LDL) mais aussi des facteurs d'entrée. Les facteurs d'entrée jouent un rôle très important dans l'infection par le VHC. En effet, les cellules non hépatiques, qui normalement ne permettent pas l'entrée du VHC, deviennent permissives si les quatre récepteurs humains du VHC, présentés ci-après, sont co-exprimés (115).

a) Récepteur scavenger type B classe 1 (SR-BI/Cla-1)

Ce récepteur de lipoprotéines est ubiquitaire, mais présente une expression plus importante au niveau hépatique et surrénal (116). Il joue un rôle dans l'attachement et le transfert lipidique en association avec des lipoprotéines de haute, faible ou de très faible densité, et participe au métabolisme du cholestérol (117). Un réarrangement de la bicouche lipidique cellulaire a été

observé suite à la liaison E2/SR-BI, permettant de rapprocher le virus des autres facteurs d'entrée (118). Grâce à la fonction d'endocytose de ses ligands, SR-BI pourrait également faciliter l'entrée du VHC (119).

b) Récepteur CD81

Cette glycoprotéine fait partie de la famille ubiquitaire des tétraspanines, qui sont impliquées dans la régulation de la prolifération ou de l'adhérence cellulaire mais aussi dans d'autres voies de signalisation (120). CD81 est le premier récepteur du VHC à avoir été identifié, et semble jouer différents rôles dans le cycle viral, notamment au niveau de la réplication (121, 122). Des cellules normalement non permissives au VHC et modifiées pour exprimer CD81 permet leur infection (123). Chez des souris dont le foie a été humanisé, des anti-CD81 injectés en prophylaxie empêchent totalement l'infection par le VHC démontrant son aspect indispensable à l'entrée du virus (124).

c) Protéine de jonction transmembranaire claudine-1 (CLDN1)

Les hépatocytes sont des cellules polarisées dont la partie apicale formant la lumière des canaux biliaires est séparée de la partie basolatérale en contact avec le sang par les jonctions serrées (125). CLDN1 se situe au niveau de ces jonctions, et participe à l'étanchéité de celles-ci, se liant à la *zonula occludens* et au cytosquelette. Si CLDN1 est exprimée principalement dans la partie apicale des cellules, il existe aussi des molécules de CLDN1 au niveau basolatéral de façon co-localisée avec SR/B1 et CD81 (126). CLDN1 est le troisième récepteur du VHC et comme CD81, la diminution de son expression ou son blocage par des anticorps inhibe l'entrée du virus, tandis que le VHC parvient à pénétrer dans des cellules non permissives lorsque CLDN1 y est exprimée (127, 128). CLDN6 et CLDN9, appartenant également à la famille des claudines, participent aussi à l'entrée du virus (129).

d) Protéine de jonction transmembranaire occludine (OCLN)

Cette molécule est le 4ème facteur d'entrée du VHC. En effet, il a été montré que la présence conjointe de SR-BI, CD81 et CLDN1, sans OCLN n'est pas suffisante pour permettre l'infection virale, et que la diminution de l'expression d'OCLN dans des hépatocytes permissifs inhibe l'entrée du virus. De plus, l'introduction de CD81 et OCLN humaines dans des hépatocytes murins surexprimant SR-BI et CLDN1 murines permettent l'infection et la production de virions VHC (130). OCLN possède les mêmes caractéristiques de localisation

(jonctions serrées, principalement apicales), et d'interactions (protéines de la *zonula occludens* et du cytosquelette) que CLDN1 (131). L'activité d'OCLN est régulée par son état de phosphorylation, qui est perturbé lors de l'infection virale VHC, déstabilisant les jonctions serrées, et favorisant l'entrée du virus dans les hépatocytes (132). Les sites de phosphorylation permettent à OCLN d'interagir avec différentes kinases comme Src, EGFR ou ERK (133).

D'autres facteurs d'entrée de moindre importance ont été mis à jour au cours des dernières années, comme les récepteurs tyrosine kinase. Différentes kinases sont également impliquées dans l'entrée du virus en tant que co-facteurs, telles que EphA2 et EGFR (134).

3.2.2.2 Synthèse des protéines virales et réplication de l'ARN du VHC

Après la fusion du virus dans la cellule et sa décapsidation, le génome viral peut être directement traduit au niveau des ribosomes cytoplasmiques (Figure 21). L'IRES situé en partie 5' non codante du génome (voir Introduction section 3.2.1) permet la fixation du génome dans le réticulum aux sous unités ribosomales pour assurer la transcription de l'ARN viral et la traduction en une longue polyprotéine. La maturation de cette polyprotéine est réalisée par la protéase NS3 du VHC et par des protéases cellulaires. La polymérase virale (ARN polymérase ARN dépendante virale, NS5B) permet la synthèse de l'ARN viral pour les nouveaux virions au sein du complexe de réplication formé notamment par NS4B attaché aux membranes du réticulum endoplasmique (135).

3.2.2.3 Assemblage et sécrétion des virions

Les ARNs néoformés sont ensuite encapsidés par interaction avec les protéines de capsid core. La nucléocapside bourgeonne au contact d'E1 et E2 retenues dans la membrane du réticulum endoplasmique pour former le virion. Ce dernier est ensuite libérée par exocytose (Figure 21) ou lyse des cellules infectées (113).

3.2.3 Variabilité génétique du VHC

Il existe une variabilité génétique du VHC liée à son fort taux de réplication et à l'absence d'activité exonucléasique correctrice de l'ARN polymérase. Le taux de mutations du génome du VHC est estimé entre 10^{-3} et 10^{-4} substitutions par site par cycle répliatif. Cette variabilité s'exprime au niveau du génome par l'existence d'au moins sept génotypes majeurs et plusieurs sous-types du VHC (Figure 22), plus particulièrement au niveau de la protéine d'enveloppe E2/NS1 (dans des zones dites « hypervariables »).

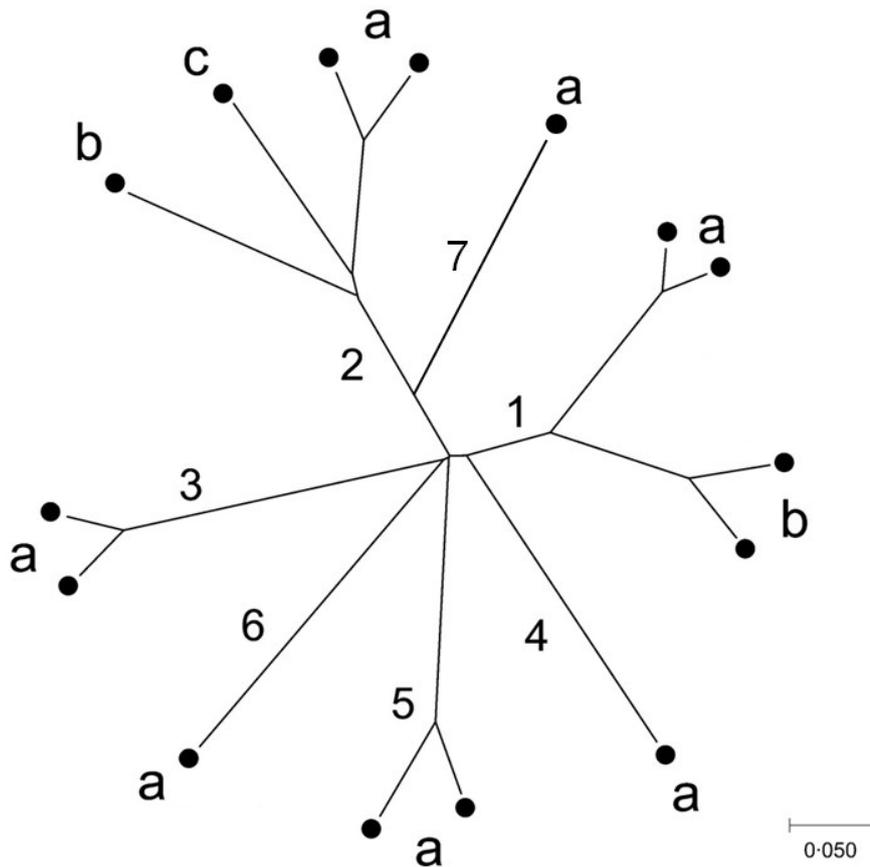


Figure 22. Arbre d'évolution des principaux génotypes du VHC. Adapté de (136).

Les génotypes sont définis par une divergence minimale de 30% de leurs séquences nucléotidiques alors que les sous-types ont une divergence de seulement 20% (136). Bien que présentant des caractéristiques relativement similaires de persistance, de dynamique de transmission et de développement de la maladie, les génotypes du VHC peuvent se distinguer par leur évolution séparée dans différentes régions géographiques. Le génotype 1 est le plus décrit en Amérique, en Europe ainsi qu'au Japon, suivis des génotypes 3 et 2. Les génotypes 3 et 6 sont répandus en Asie du Sud et du Sud-Est. Les génotypes 4 et 5 sont couramment retrouvés en Afrique mais se propagent actuellement en Europe (137-139). Enfin, le génotype 7 est retrouvé chez des patients en Afrique mais reste rare et d'importance clinique mineure (140). De manière intéressante, les différents sous-types sont majoritairement identifiés dans des groupes à risques distincts, tels qu'une prédominance des génotypes 1a et 3a dans les populations toxicomanes en Europe (141).

Chez le malade infecté, le VHC circule sous la forme d'un mélange de variants viraux apparentés définissant une distribution en « quasi-espèces ». Plusieurs « quasi-espèces » peuvent être isolées chez un même malade et cette hétérogénéité évolue au cours de la maladie

(142). Ce phénomène confère au VHC un avantage considérable en termes de survie, car la présence simultanée chez un même individu de multiples séquences virales et la rapidité de génération de nouveaux variants permettent la sélection rapide des mutants les mieux adaptés à l'environnement immunologique mis en place par l'hôte. Il est sans doute un élément essentiel de persistance virale favorisant l'action d'autres facteurs de persistance davantage liés à la régulation immunologique.

3.2.4 Tropisme cellulaire

Le tropisme du VHC est principalement hépatique. Le foie est la première cible de l'infection VHC et plusieurs millions de copies d'ARN du VHC par gramme sont retrouvées dans le tissu infecté. Néanmoins, l'ARN du VHC et/ou ses protéines ont été détectés dans d'autres organes de patients infectés suggérant que l'hépatite C pourrait avoir des réservoirs extra-hépatiques. Ces détections sont corrélées avec des manifestations extra-hépatiques après infection (143).

Les mécanismes définissant l'hépatotropisme ne sont pas encore bien décrits. L'infection par le VHC est liée à plusieurs facteurs cellulaires de l'hôte et ce à chaque étape du cycle du virus (voir Introduction section 3.2.2). L'entrée du VHC nécessite au moins la présence de quatre facteurs en surface cellulaire : CD81, SR-BI, CLDN1 et OCLN, mais aucun de ces facteurs n'est spécifique du foie. Il semblerait que les niveaux d'expression de ces récepteurs à la surface des hépatocytes soient importants pour l'entrée du VHC (144-147).

L'infection par le VHC est associée à des symptômes extra-hépatiques comme la fatigue et des dysfonctionnements cognitifs (148). Ces symptômes suggèrent que l'infection par le VHC pourrait se produire ailleurs que dans son principal organe cible et atteindre des cellules neuronales *via* un neurotropisme du VHC mal connu. Des preuves de la présence de l'ARN du VHC dans le cerveau ont vite vu le jour avec le développement de techniques de détection comme la RT-PCR (149-151). De plus, plusieurs études ont évalué et identifié une expression des facteurs d'entrée de l'hôte (CD81, SR-BI, CLDN1 et OCLN) dans des lignées cellulaires dérivées du cerveau (152, 153).

Il semblerait que des cellules mononuclées du sang périphérique (CMSP) soient aussi un réservoir potentiel pour le VHC. Des protéines virales ainsi que l'ARN du VHC ont été détectés dans des monocytes, des cellules dendritiques mais aussi dans des lymphocytes B et T (154-159). Toutefois, d'autres études argumentent que la détection de l'infection par le VHC dans les CMSPs est le résultat d'artefacts expérimentaux dus aux techniques utilisées (160-162). La

présence d'ARN viral pourrait être due à la capacité naturelle d'adsorption de certaines cellules (163) mais des études ont montré une répllication du VHC dans les CMSPs (164, 165). Cela montre que le sujet de la répllication du VHC dans les CMSPs reste controversé (166) et donne matière à discussion.

3.3 Histoire naturelle de l'infection par le VHC

L'hépatite C est une maladie progressive (Figure 23): l'infection par le VHC entraîne obligatoirement une phase aiguë, et l'organisme peut éliminer spontanément le virus dans 10 à 40% des cas pendant cette phase. Dans la majorité des cas, la maladie persiste donc et devient chronique, on parle d'hépatite C chronique.

L'hépatite C aiguë évolue de manière asymptomatique chez la majorité des individus (167). L'infection aiguë symptomatique implique des symptômes non spécifiques tels que malaise, fatigue, myalgies, jaunisse et nausées. On observe aussi dans ce cas une augmentation des enzymes du foie comme l'alanine amino-transférase (ALAT) et l'aspartate amino-transférase (ASAT). Par ailleurs, il est intéressant de noter que la majorité des cas de guérisons spontanées sont observés chez les patients atteints d'une infection aiguë symptomatique (168, 169). Dans de très rares cas, moins de 1% des patients, on peut observer une hépatite fulminante qui conduit généralement à la mort du patient.

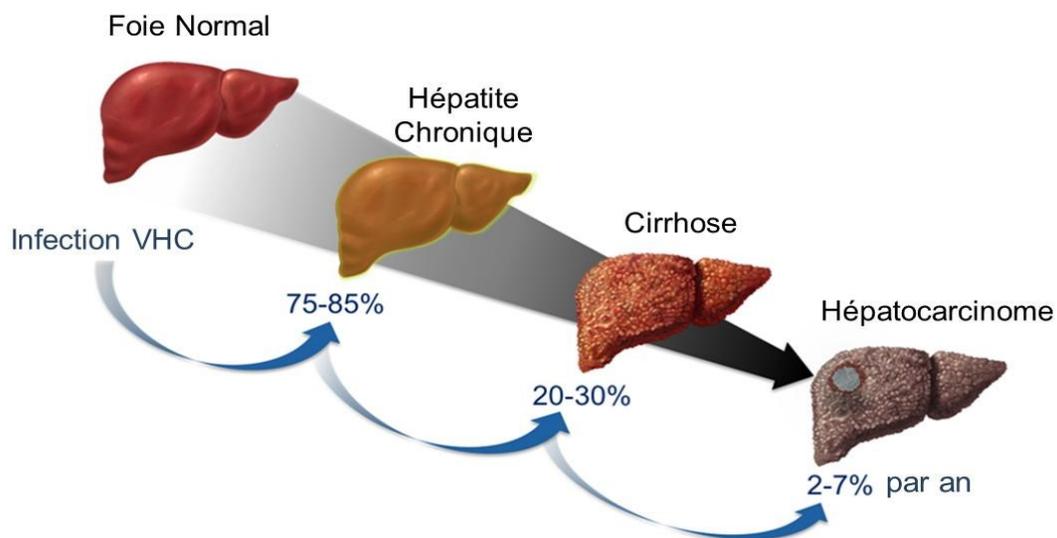


Figure 23. Histoire naturelle de l'infection par le VHC. L'évolution vers le carcinome hépatocellulaire prend environ 30ans. Adapté de <http://yourhealtheducator.com/hepatitis-inflammation-liver-cells/>.

Après la primo infection, le VHC va persister chez environ 75% des individus malgré la réponse immunitaire cellulaire et humorale (170). L'infection chronique par le VHC est définie par la présence de l'ARN du VHC 6 mois après la date estimée de l'infection. On observe alors au cours du temps des lésions d'infection chronique menant à l'apparition d'une fibrose. Il a cependant été observé une élimination spontanée de l'infection chronique par le VHC dans 0,5% -0,74% des cas (171). Mais dans 20 à 30% des cas l'infection chronique peut amener à des complications graves comme le développement de cirrhose du foie (172) qui peut conduire dans 1-5% des cas à un hépatocarcinome cellulaire (HCC) (Figure 23).

Il est important de noter que plusieurs facteurs favorisent la progression vers la fibrose comme l'insulino-résistance et la stéatose hépatique (accumulation d'adipocytes intra-hépatocytaires) (173). Le genre masculin, un âge supérieur à 50 ans, et surtout la surconsommation d'alcool ou une coinfection par le VIH ont aussi été associés à une progression accélérée de la fibrose hépatique et une augmentation de l'incidence de la cirrhose et du carcinome hépatocellulaire (170).

En conséquence de l'infection par le VHC, des complications ou manifestations extra-hépatiques du VHC peuvent apparaître. Elles sont peu connues et pourtant très importantes si l'on considère que pratiquement tous les organes peuvent être affectés par ce virus (148). Par exemple, le risque d'infarctus du myocarde et d'accident vasculaire cérébral est augmenté, indépendamment des facteurs de risques bien connus tels que le tabagisme, l'hypertension ou encore l'excès de cholestérol. Des troubles métaboliques peuvent aussi intervenir, augmentant l'apparition de diabète de type 2 et donc l'insulino-résistance favorisant la fibrose. Des atteintes rénales sévères ont aussi été observées, ainsi que la survenue de lymphomes ou des désordres touchant le système nerveux central (148).

3.4 Mécanismes de pathogenèse liés au VHC

Des progrès considérables ont été réalisés dans la connaissance du virus, de son épidémiologie et des pathologies associées, mais les mécanismes de persistance/pathogenèse restent encore mal connus. Malgré tout, plusieurs modèles *in vivo* et *in vitro* ont récemment vu le jour (174) et permettent d'améliorer la compréhension des mécanismes immunitaires, des interactions hôte/virus ou de la pathogenèse associée aux différents stades induit par l'infection par le VHC (Figure 24).

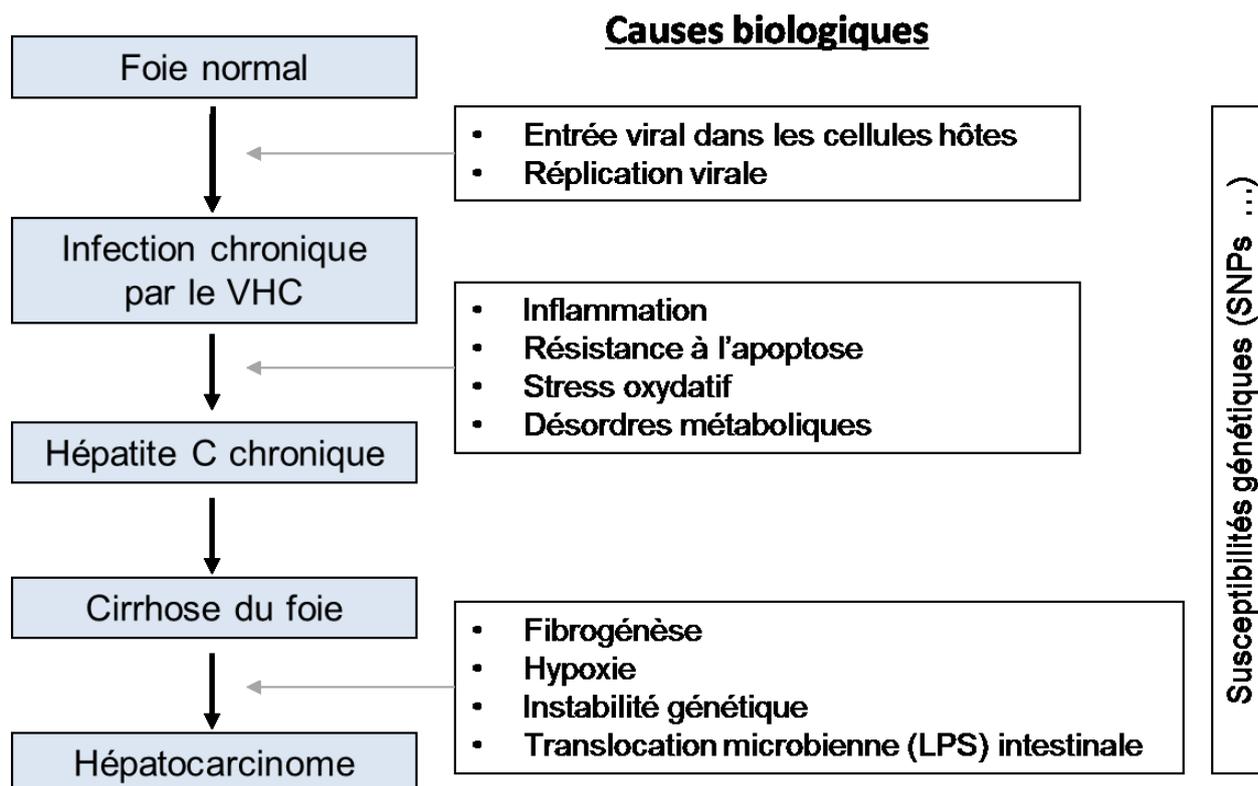


Figure 24. Histoire naturelle du VHC et les causes biologiques induisant des pathologies liées à l'infection. Adapté de (175).

3.4.1 Réponses immunitaires innées et adaptatives au VHC

Comme pour la grande majorité des infections, une réponse immunitaire innée de l'hôte est mise en place. Cela se traduit par une activation de cellules NK et de la voie de production d'IFN de type I et III. Cette réponse innée précède et prépare la mise en place de la réponse adaptative qui se traduit par la maturation des lymphocytes B en plasmocytes sécrétoires d'anticorps spécifiques anti-VHC : « la réponse humorale » ; mais aussi abouti à l'activation de la division lymphocytaires T cytotoxiques : « la réponse cellulaire ».

3.4.1.1 La réponse immunitaire innée

La réponse immunitaire innée est la première ligne de défense contre le VHC et se met en place environ deux semaines après infection lors de la phase précoce de la phase aiguë de la maladie. Son rôle est de détecter le VHC, d'initier des réponses antivirales locales dans les cellules infectées et de recruter des cellules immunitaires à la médiation des réponses adaptatives. Les récepteurs TLR-3 et RIG-I reconnaissent les composants du virus appelés PAMPs (*pathogen-associated molecular patterns* en anglais). TLR-3 reconnaît l'ARN double brin alors que RIG-

I reconnaît les extrémités 5' et 3' non codantes du génome du VHC. La conséquence est l'activation de voies de signalisations par ces deux récepteurs, induisant la sécrétion par les cellules infectées d'IFN- β de type I et d'IFN- λ de type III. Ceux-ci activent des mécanismes d'apoptose *via* la voie des caspases (176). La production d'IFN de type I et III pourrait moduler l'apoptose par plusieurs mécanismes dont la voie JAK/STAT (177). Il semblerait que le virus ait évolué pour échapper à l'immunité innée. Les protéines virales NS3-NS4A jouent un rôle central dans le processus de l'évasion virale en bloquant la protéine RIG-I et l'induction de la mort cellulaire. En outre, les protéines du VHC sont antagonistes à la voie de signalisation JAK/STAT induite par les IFN de type I et III. Un autre mécanisme de l'évasion viral repose sur la modulation de l'expression des gènes stimulés par ces IFN, supprimant ainsi les réponses immunitaires innées (178).

En plus de l'induction de la réponse antivirale dans les cellules, les IFN de type I et III sont principalement responsables de l'activation/régulation des cellules de l'immunité telles que les cellules NK qui jouent un rôle important dans le contrôle de l'infection virale. Les NK sont à l'origine d'effets cytolytiques directs des cellules infectées médiés par les perforines ou le TRAIL (*TNF related apoptosis-inducing ligand* en anglais) et d'effets indirects non-cytolytiques médiés par les IFN- γ . Mais là encore, le virus a développé des mécanisme pour échapper à l'immunité innée *via* la protéine E2 qui interagit avec le récepteur CD81 des cellules NK pour altérer leur fonctionnement (179).

3.4.1.2 La réponse immunitaire adaptative

La réponse immunitaire innée participe aussi à la mise en place de la réponse adaptative qui est beaucoup plus spécifique et durable. La réponse adaptative intervient par deux mécanismes (humorale et cellulaire) et sa détection se fait à partir de la quatrième semaine post-infection environ.

a) La réponse adaptative cellulaire

Les IFN sécrétés par les éléments de la réponse innée induisent les cellules dendritiques, qui sont utilisées comme cellules présentatrices d'antigènes (CPA). Suite à la détection des CPA, la réponse adaptative est mise en place. Interviennent, alors, des lymphocytes T auxiliaires CD4⁺ et des lymphocytes T cytotoxiques CD8⁺.

L'action de la réponse CD4⁺ participe à la clairance du virus pendant la phase de l'hépatite aiguë. En effet, une réponse CD4⁺ de type Th1, faisant intervenir les voies de l'IFN- γ et de l'IL-2, est soutenue, rapide, forte et peut aboutir à la clairance du virus. A l'opposé, une réponse CD4⁺ de type Th2 est beaucoup plus retardée, faible et même transitoire, faisant intervenir les voies de l'IL-4 et de l'IL-10 et ne permettant pas à bloquer le passage vers la chronicité (180).

La réponse CD8⁺ vise à inactiver et éliminer le virus par une action cytolytique directe des lymphocytes mais aussi par un mécanisme non cytolytique *via* la sécrétion de cytokines anti-inflammatoire. Toutefois, il a déjà été montré que l'infection est capable d'inhiber la prolifération des CD8⁺, leur fonction cytotoxique mais aussi et surtout leur capacité à sécréter des cytokines (181). Les CD8⁺ sont aussi capables de reconnaître différents épitopes des protéines virales (principalement NS3 et NS5B). Selon l'épitope ciblé par la réponse cytotoxique, la clairance du virus est plus ou moins variable et la coordination entre les deux composants CD4⁺ et CD8⁺ est nécessaire afin de pouvoir établir une réponse suffisante à la clairance virale (182). De puissantes réponses lymphocytaires CD4⁺ multi-épitopes spécifiques et CD8⁺ ont ainsi été observées chez les patients guérissant spontanément lors de la phase aiguë de la maladie (183).

b) La réponse adaptative humorale

La réponse humorale commence à s'établir pendant la phase aiguë de l'hépatite C, à partir de la quatrième semaine post-infection. L'installation de la réponse humorale se traduit par la sécrétion d'anticorps anti-VHC ciblant différentes protéines virales. Cependant, la capacité de neutralisation de ces anticorps n'est pas claire *in vivo*. En effet, certains patients chroniquement infectés possèdent des anticorps neutralisants. Cette observation peut être expliquée par l'existence d'anticorps non-neutralisants dirigés contre les mêmes épitopes que les anticorps neutralisants, ou des épitopes chevauchants, et qui pourraient empêcher leur activité neutralisante (184). De plus, le virus est capable de se soustraire à cette réponse humorale en raison de son processus de réplication d'erreurs qui se traduit par une grande variabilité dans la région E1E2, cible d'une grande majorité des anticorps neutralisants (185). Par conséquent, le VHC persiste malgré la détection des anticorps anti-VHC neutralisants chez les patients atteints d'hépatite C chronique.

Le rôle de la réponse humorale dans la lutte contre l'infection par le VHC reste cependant très important. En effet, une étude portant sur une cohorte de patients infectés par le VHC lors

d'injection d'immunoglobuline contaminée, indique qu'un développement précoce d'anticorps neutralisants peut contribuer à la clairance de l'infection par le VHC alors qu'une réponse tardive peut conduire à une infection chronique (186). De plus, la présence d'anticorps neutralisants a été associée à une faible charge virale, ce qui indique leur rôle dans le contrôle de la virémie VHC (187).

3.4.2 Mécanismes pathogéniques associés à l'hépatite C chronique

L'hépatite C chronique est caractérisée sur le plan histologique par la présence d'infiltrats lymphoïdes dans les espaces portes (ou espaces de Kiernan) des lobules hépatiques (Figure 25), souvent sous la forme d'agrégats ou de follicules lymphoïdes. Ceux-ci sont souvent situés à proximité des canaux biliaires qui peuvent alors être lésés. Les lésions hépatocytaires associent typiquement une nécrose cellulaire périportale et une inflammation lobulaire modérée. Une stéatose macrovésiculaire modérée associée est également fréquente.

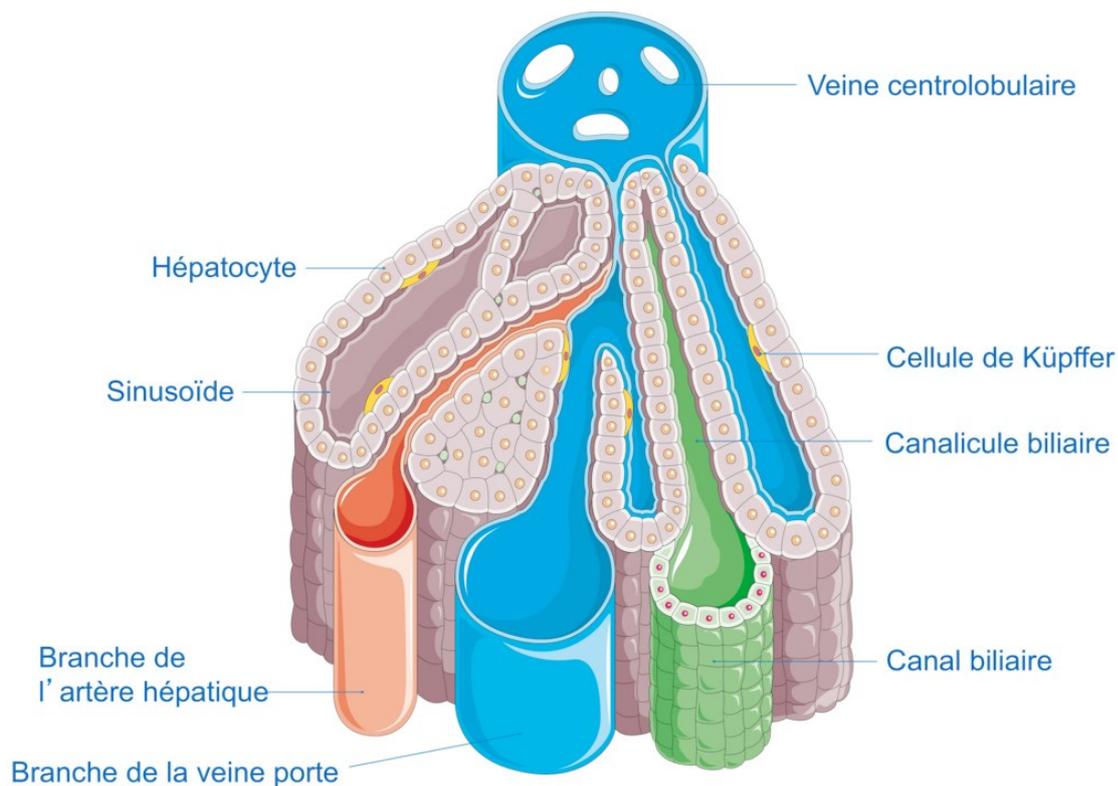


Figure 25. Schématisation d'un lobule hépatique. Regroupement de cellules en unité fonctionnelle du foie. L'espace porte est l'ensemble composé de l'artère hépatique, de la veine porte ainsi que du canal biliaire. (Image Servier Medical Art)

Plusieurs arguments expérimentaux pointent le rôle de la réponse immune cellulaire dans le processus de destruction progressive du tissu hépatique caractérisant l'hépatite C chronique chez l'homme (188).

En l'absence d'action cytopathique directe du VHC, les réponses immunes jouent un rôle de premier plan dans l'apparition et l'évolution des lésions hépatiques de l'hépatite C chronique. Des lymphocytes T CD4⁺, principalement de type Th1, et des lymphocytes T cytotoxiques s'accumulent au niveau du foie (voir Introduction section 3.4.1). La lyse hépatocytaire induite par ce mécanisme serait largement dépendante du type de cytokines secrétées (189-191). Certaines études suggèrent que le virus altère directement la production de cytokines et de facteurs de croissance dans le tissu hépatique des malades infectés (192, 193). Ainsi, le déséquilibre dans l'expression des cytokines pro-inflammatoires et anti-inflammatoires au cours de l'hépatite C chronique pourrait soit aboutir à une réponse immunitaire qualitativement inadaptée à l'élimination virale (194), soit être à l'origine d'une réaction inflammatoire non spécifique excessive supportant l'évolution en fibrose de la maladie (189).

3.4.3 Mécanisme de fibrose lié à l'hépatite C

La fibrose hépatique est la principale complication de l'hépatite C chronique. Son terme évolutif, la cirrhose, est responsable de la morbidité et de la mortalité de l'hépatite C (195). Le foie réagit à l'agression par le VHC par la mise en place d'une réaction inflammatoire donnant lieu à de la fibrogenèse. La fibrogenèse est un processus dynamique caractérisé par la synthèse de molécules constitutives de la matrice extracellulaire (MEC), ensemble complexe de molécules protéiques (collagène, élastine), glycoprotéiques (fibronectine, laminine, etc...) et de protéoglycanes, organisée en réseaux tridimensionnels interconnectés (196, 197).

Il s'agit d'un mécanisme non spécifique d'organe qui se pérennise tant que l'agent agresseur reste présent dans le foie et qui vise à limiter l'extension de la réaction inflammatoire. Il s'agit donc au début d'un mécanisme physiologique bénéfique pour l'organisme et qui fait partie intégrante du mécanisme de cicatrisation. La fibrose est quant à elle un mécanisme pathologique consécutif à une fibrogenèse prolongée liée à la persistance du VHC dans le foie. La fibrose est caractérisée par le dépôt anarchique des molécules de la MEC et leur organisation en polymères complexes et peu solubles aboutissant à la perturbation de l'architecture hépatique (198).

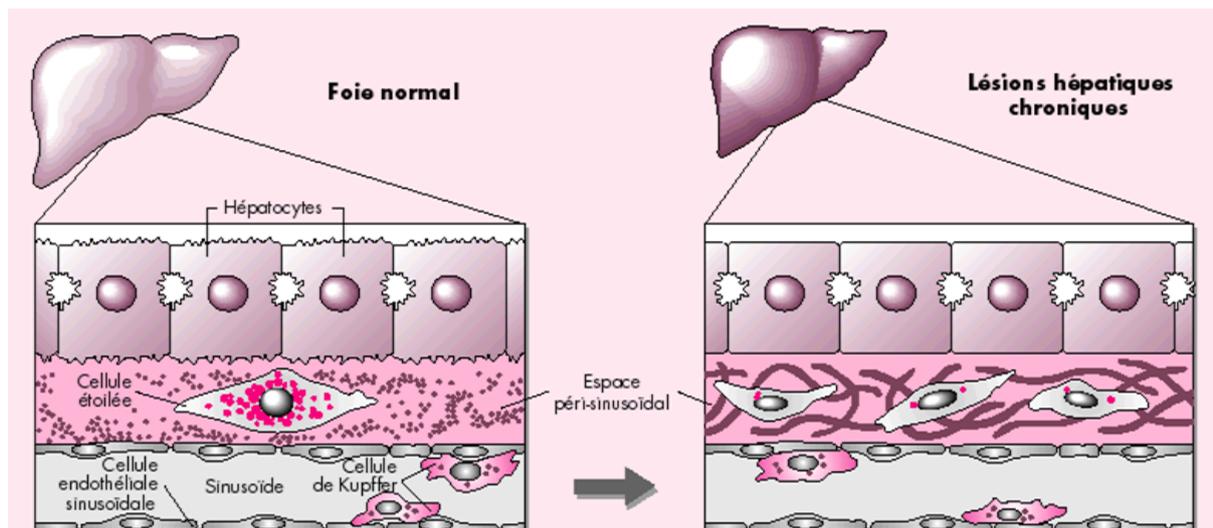


Figure 26. Le foie normal présente des cellules étoilées (HSC) quiescentes et une matrice extracellulaire (MEC) de faible densité dans l'espace peri-sinusoidal. Lors de lésions chroniques, les HSC prolifèrent et sont activées. Cela favorise la production d'une MEC de haute densité qui à long terme aboutit à la fibrose (199).

Il est établi que la plupart des cellules hépatiques sont susceptibles de participer à la synthèse matricielle (200). C'est ainsi le cas des hépatocytes (201, 202), des cellules épithéliales biliaires (203) et des cellules endothéliales sinusoidales (201). Cependant, le rôle majeur de la fibrogenèse est joué par la cellule étoilée hépatique (HSC, *hepatic stellate cell* en anglais).

Une des hypothèses serait que les HSC expriment des récepteurs d'entrée du VHC (204), ce qui les rendrait permissives à une infection *in vivo* par le VHC, mais cela n'a pas encore été établi et une étude récente a montré, *in vitro*, la non-infection de ces cellules par le VHC (205). Sous l'influence de cytokines, notamment par le transforming growth factor beta (TGF- β) produit par les cellules de Kupffer lors de dommages hépatiques, les HSC acquièrent des propriétés myofibroblastiques et se mettent à synthétiser des constituants de la MEC (Figure 26) (206). La prolifération des HSC est quant à elle soumise à l'influence du platelet-derived growth factor (PDGF) aussi produit lors de dommages hépatiques (207). La synthèse de cytokines, lors de la réaction inflammatoire chronique due au VHC, a donc pour action avec le temps, la prolifération et la différenciation des HSC favorisant l'accumulation exagérée de MEC au sein du parenchyme hépatique.

3.4.4 Cirrhose et hépatocarcinome cellulaire (HCC)

La cirrhose est le stade final des maladies chroniques du foie. C'est une condition pathologique définie par une architecture hépatique dérégulée et une régénération nodulaire parenchymateuse résultant d'une fibrose progressive. Après 10 à 15 années, on parle de cirrhose décompensée lorsque l'adaptation des mécanismes pour compenser les effets nuisibles de la pathologie n'est plus suffisamment efficace. Le foie n'est alors plus capable d'assurer ses fonctions normales et des symptômes graves apparaissent comme des hémorragies œsophagiques ou du tube digestif, des ascites (épanchement liquidien intra-abdominal) ou encore des œdèmes. Le développement de la cirrhose est aussi la condition la plus commune prédisposant pour le HCC. Une fois la cirrhose établie, le risque de développement de HCC augmente de 1 à 6% par an (208).

Les mécanismes de progression vers le HCC ne sont pas bien connus. Les protéines du VHC peuvent indirectement conduire à une instabilité génétique des cellules infectées *via* la suppression des systèmes de réparation de l'ADN, à l'apparition de cassures d'ADN, à une augmentation de la fréquence de mutation et à des réarrangements chromosomiques (209). En effet, la protéine core du VHC, capable d'induire le HCC chez des souris transgéniques (210), est impliquée dans des voies de signalisation cellulaires, d'activation de transcription, d'apoptose, de métabolisme des lipides ou encore de transformation (211). D'autre part, la protéine E2 peut interagir avec le système immunitaire et inhibe l'action des cellules T et NK afin de promouvoir la prolifération et la survie des cellules malignes (212, 213). Enfin, la protéine NS3 ayant des activités de protéase, hélicase à ARN et NTPase peut induire la carcinogénèse par interaction avec p21 et p53, deux oncogènes (214, 215).

D'autres mécanismes pourraient aussi intervenir dans le développement d'un HCC. Ainsi, il a été observé que les télomères sont plus courts chez certains patients avec une infection VHC chronique ou en rémission, cela pouvant causer des instabilités génétiques, des aneuploïdies et des transformations malignes (216, 217). Des analyses de pertes d'hétérozygoties ont aussi montré, de façon significative chez des patients avec un HCC primaire, plusieurs loci chromosomiques abritant des gènes suppresseurs de tumeur potentiels tels que *TP53* et *IGF2R* (218-220). Il a aussi été identifié des mutations somatiques lors d'induction de HCC par le VHC dans des gènes tels que *TP53*, la catenin beta-1 (*CTNNB1*) et le phosphoinositide-3-kinase-catalytic-alpha (*PIK3CA*) (221). Les mutations dans *TP53* affectent en aval les gènes liés au cycle cellulaire ou encore à la prolifération. Les mutations dans *CTNNB1* sont présentes dans 33% des cas de HCC induit par le VHC (222).

3.5 Les traitements de l'hépatite C

Le traitement standard et seul traitement de l'infection par le VHC était, jusque fin 2011, la bithérapie par interféron pégylé et ribavirine (PEG-IFN/RBV). Historiquement, le traitement de l'hépatite C chronique était l'interféron alpha simple mais qui avait comme désavantage majeur de nécessiter plusieurs injections par semaine. La pégylation (PEG) de l'interféron, c'est-à-dire la fixation d'un résidu polyéthylène glycol inerte, a permis de contourner ce problème en conférant au produit une demi-vie plus longue et permettant de réduire la fréquence d'injection (223). Cependant, il a été démontré que l'origine ethnique des patients a une influence sur la réponse au traitement PEG-IFN/RBV. Ainsi, des polymorphismes fréquents chez les européens dans le gène *IL28B* favorisent deux fois plus une réponse virologique soutenue (SVR, *sustained virological response* en anglais) après traitement par rapport aux patients d'origine africaine (224).

Les IFN- α 2a et 2b agissent au niveau de la réponse innée (225). Toutes les étapes du cycle viral (voir Introduction section 3.2.2) peuvent être inhibées par les IFN. Ils n'agissent pas directement sur le virus mais indirectement par induction de la transcription de gènes codant pour des protéines médiatrices des effets antiviraux (226). Cela va d'une activité antivirale comme l'induction de la dégradation d'ARN simple brin *via* l'activation de la 2'5'A synthetase/RNaseL (227), à une action immuno-modulatrice *via* notamment l'activation de cellules T CD8 cytotoxiques et NK, une augmentation de l'expression de molécules HLA de classe I sur les cellules infectées et d'une inhibition de l'apoptose des cellules T (228, 229).

La ribavirine (RVB) appartient à la classe thérapeutique des analogues nucléosidiques de synthèse, très important dans le traitement de l'hépatite C. Plus précisément, la RBV est un analogue de la guanosine, synthétisé pour la première fois en 1970 (230). Son activité antivirale à spectre large a été rapportée en 1972 (231). Sa forme aérosol a été approuvée dans le traitement pour le virus respiratoire syncytial chez les enfants en détresse respiratoire dans le milieu des années 1980, avant la découverte du virus de l'hépatite C en 1989. De nombreuses propriétés antivirales *in vitro* contre de nombreux virus ont été démontrées (232-235). Même si son activité antivirale reste modérée *in vivo* à l'encontre du VHC (236), sa synergie avec le PEG-IFN est indiscutable sans que ce mécanisme d'action ait été élucidé (237). Bien que la RBV joue clairement un rôle dans la réponse au traitement contre le VHC, le mécanisme antiviral reste controversé. Il existe plusieurs mécanismes proposés pour l'action de la RBV (238-240) comme l'inhibition de la polymérase virale ou du coiffage des ARNs ou encore de

façon indirecte un effet immuno-modulateur. En effet, il a été prouvé qu'elle réduit, *in vitro*, la prolifération lymphocytaire induite par les agents mitogènes tels que le VHC. Son effet immuno-modulateur provient de sa capacité à réguler la réponse immune en diminuant la réponse Th2 au profit de la réponse Th1, augmentant ainsi la production d'IFN γ (voir Introduction section 3.4.1). Cette immuno-modulation augmente la clairance virale alors que la réponse Th2 augmente le risque de passage à la chronicité (236, 237, 241).

Depuis 2011, le traitement de l'hépatite C a connu une avancée majeure avec le développement de nouveaux antiviraux à action directe (DAA pour *direct acting antiviral*) dirigés contre les protéines du VHC. La prise en charge des patients infectés par le génotype 1 du VHC a grandement été améliorée par la combinaison du traitement PEG-IFN/RBV avec le télaprévir ou le bocéprévir, deux DAAs. Cette combinaison a permis d'atteindre des taux de SVR de 70 à 75% avec le télaprévir et de 68% avec le bocéprévir (242-244). Bien qu'efficaces, ces molécules présentent néanmoins d'importantes limites : des effets secondaires majeurs (rash cutané et d'anémie sévère nécessitant la co-prescription de facteurs de croissance hématopoïétique, voir des transfusions itératives), un coût élevé et l'émergence de mutations résistantes en cas d'échec (compromettant les stratégies ultérieures). Tous ces éléments restent un frein à la généralisation de ces traitements.

La mise à disposition en 2014 d'une nouvelle génération de DAA très efficace et disposant d'un bon profil de tolérance a laissé entrevoir une guérison possible de la majorité des personnes porteuses d'infection chronique par le VHC. En effet, après un traitement de 12 ou 24 semaines, 90% de réponse soutenue virologique sont observés (245, 246). Cette guérison est associée à une réduction des complications, des cirrhoses, des cancers et des transplantations. Après le premier de cette nouvelle classe thérapeutique : le sofosbuvir, un inhibiteur de polymérase NS5B, de nombreux autres produits ont vu le jour comme le siméprévir et le daclatasvir, tous deux inhibiteurs du complexe NS5A (245). Encore, ces traitements sont très coûteux et ne sont pas disponibles pour tous les patients infectés. Le sofosbuvir n'est ainsi remboursé que pour les malades avec une atteinte hépatique sévère.

D'autres approches sont en voie de développement comme cibler les récepteurs cellulaires nécessaires à l'entrée du virus (voir Introduction section 3.2.2.1) par des molécules ou des anticorps spécifiques (242-244, 247-249).

3.6 Facteurs génétiques de l'hôte et l'hépatite C

L'avancé clinique de la maladie diffère grandement entre les individus infectés par le VHC. Les variations virales et/ou celles de l'hôte ainsi que les facteurs environnementaux influencent cette progression (Figure 27). Depuis quelques années et jusqu'à récemment, plusieurs GWAS ont identifié des facteurs génétiques de l'hôte impliqués dans la progression de la maladie ou dans l'efficacité des traitements. La principale découverte a été l'association de variations génétiques dans *IFNL3-IFNL4* (aussi nommé *IL28B*), tout comme le locus *HLA* de classe II avec la clairance spontanée, lors de l'hépatite C aiguë. D'autres variants génétiques ont été associés à l'activité de l'hépatite ou à la progression de l'hépatite C chronique. Ainsi, des gènes liés à l'apoptose comme *RNF7*, *TULP1* et *MERK* ont été associés à une progression de la fibrose, et des variants dans *DEPDC5* et *MICA* ont été associés à des HCC lié au VHC.

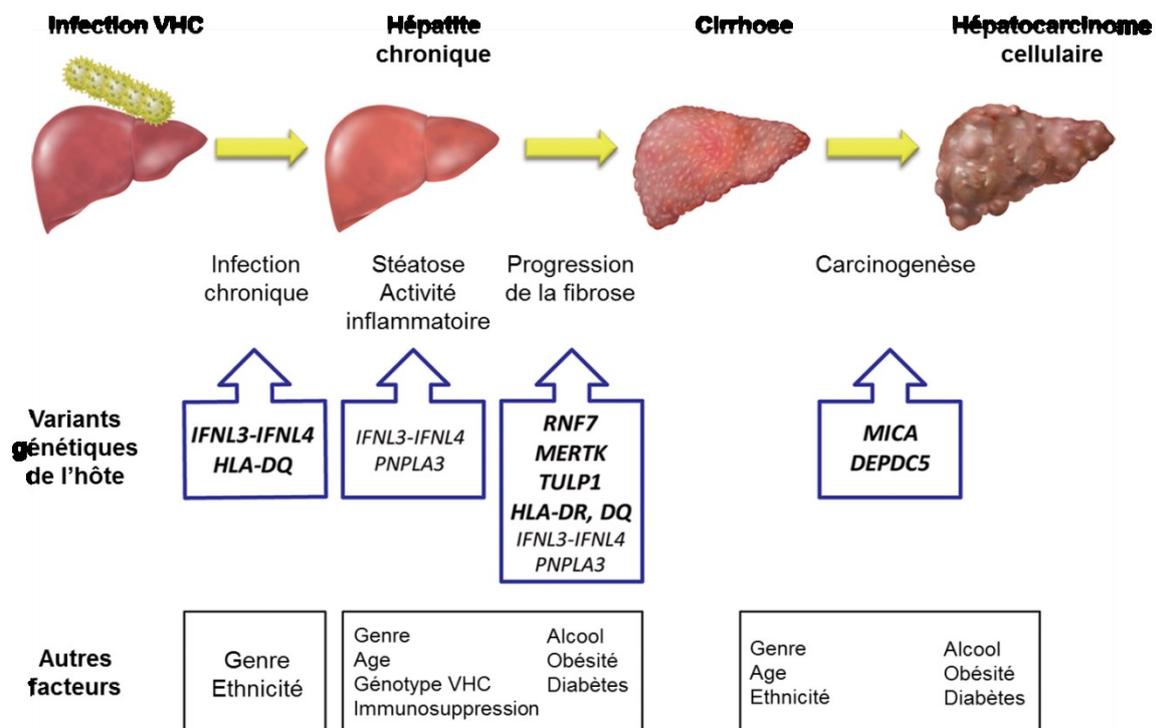


Figure 27. Variations génétiques de l'hôte et autres facteurs associés à l'évolution clinique de l'infection par le VHC. Les gènes en gras ont été identifiés par GWAS alors qu'en police normale par approche "gène candidat". Adapté de (250).

3.6.1 Guérison sous traitement ou spontanée

En 2009, quatre groupes indépendants ont publiés simultanément des GWAS testant le rôle génétique dans la réponse au traitement PEG-IFN/RBV chez des patients infectés

chroniquement par le VHC (224, 251-253). Les quatre études ont montré une forte association entre des polymorphismes dans/près du locus *IFNL3-IFNL4* et la clairance VHC induite par le traitement, les patients homozygotes pour l'allèle favorable ayant un taux de deux à trois fois plus élevé pour une SVR. Les données se sont accumulées en regard de la significativité des polymorphismes du locus *IFNL3-IFNL4* non seulement dans la réponse aux traitements mais aussi avec la clairance spontanée lors de la phase aiguë de l'infection par le VHC. Une GWAS de 347 patients clairés spontanément, 567 infectés chroniques et de 448 coinfectés VIH/VHC a permis d'identifier le rs8099917. Ce SNP est situé en amont de 9kb d'*IFNL3* et en dehors d'*IFNL4* (251). Des études « gène candidat » ont aussi été réalisées comme celle de Thomas et *al.* en 2009 (254) qui a étudié le cas du rs12979860 dans la clairance spontanée du VHC, situé à 3kb en aval d'*IFNL3* dans le premier intron d'*IFNL4*. La forte association entre le rs12979860 et la guérison spontanée a été observée au sein d'Européens et d'Afro-américains (254). Il semblerait que la fréquence génotypique d'*IFNL3-IFNL4* varie parmi la population mondiale impliquant probablement un taux de guérison variable selon les ethnies (255-257). Récemment, un nouveau polymorphisme dans le premier exon d'*IFNL4* a été associé à la clairance (258). Il s'agit du rs368234815 (TT ou G) en fort LD avec le rs12979860, où l'allèle G, favorisant la guérison, permet un cadre de lecture pour la traduction de la protéine IFN- λ 4 de type III similaire à IFN- λ 3.

En 2013, une GWAS entre la guérison spontanée du VHC et un large nombre d'individus d'origine européenne et africaine a observé le rs12979860 (*IFNL4*), mais aussi, associé de façon indépendante, le rs4273729 près du locus *HLA-DQB1* (259). Une association entre le *HLA* de classe II et la clairance spontanée avait déjà été rapportée par analyse « gène candidat » (260, 261).

3.6.2 Progression de la maladie

L'infection chronique amène à la fibrose hépatique et par conséquent à la cirrhose et au HCC. Cependant, l'activité inflammatoire et la progression vers la cirrhose et le HCC varient selon les individus. Plusieurs études ont déjà montré que l'âge, le genre, le génotype VHC, la prise d'alcool, l'obésité, la résistance à l'insuline, les diabètes ou encore l'immunosuppression influencent et favorisent la progression de la maladie vers la cirrhose et le HCC. Plusieurs études « gène candidat » ont révélé des associations entre la progression de la fibrose et des marqueurs génétiques impliqués dans la séquence de plusieurs gènes de cytokine et de chimiokines comme *IL6*, *CCL2*, *CCL8*, *TNF* et le *TGFB1* (262-266). Il a aussi été rapporté que

le *HLA* de classe II influence l'activité inflammatoire hépatique et la progression vers l'infection chronique (267, 268). Le locus *IFNL3-IFNL4* semble aussi jouer un rôle dans la progression de la maladie, comme ses associations avec un risque de fibrose sévère (269), ou une haute activité inflammatoire et une progression de la fibrose (270).

En 2008, une GWAS a mis en avant une association entre la stéatose hépatique non-alcoolique (NASH) et le gène de la patatin-like phospholipase domain-containing protein 3 (*PNPLA3*). Le rs738409, variant de *PNPLA3*, code pour un isoforme protéique (I148M), influençant la stéatose sans affecter la résistance à l'insuline (271). Ce polymorphisme a aussi été associé avec des maladies hépatiques dues à l'alcool et des cirrhoses alcooliques (272). Lors d'une étude avec la stéatose hépatique et la progression de la fibrose chez des patients européens infectés chroniques, l'allèle G homozygote du rs738409 augmente les risques de stéatose avec un Odd-Ratio (OR) de 2,55 et de fibrose (OR = 3,13) (273).

En 2012, Patin *et al.* ont réalisé une GWAS pour identifier des facteurs génétiques associés à la progression de la fibrose hépatique au sein d'une cohorte de patients européens infectés VHC très bien caractérisés (274). Le rs16851720, dans le gène *RNF7* qui code pour un antioxydant protégeant de l'apoptose, a été significativement identifié, tout comme les rs4374383 et rs9380516 situés dans ou près de *MERTK* et *TULP1* qui sont impliqués dans la phagocytose cellulaire apoptotique par les macrophages.

3.6.3 Hépatocarcinogénèse

Plusieurs études ont identifié des facteurs génétiques prédictifs du HCC lié au VHC. Par approche « gène candidat », plusieurs gènes comme *EGF*, *TP53* ou encore *MDM2* ont été associés à l'hépatocarcinogénèse (275-277). Par analyses « génome entier », les gènes *DEPDC5* et *MICA* ont été identifiés.

En 2011, une GWAS de patients japonais infectés par le VHC, avec ou sans HCC, a révélé un variant intronique, le rs1012068, localisé dans le locus *DEPDC5*, associé significativement avec la susceptibilité au HCC (OR = 1,96) (278). De façon intéressante, le niveau d'expression de *DEPDC5* est plus important dans des tissus tumoraux que dans du tissu non-tumoral sans être relié aux génotypes du rs1012068. Cependant, l'allèle à risque (G) est significativement plus fréquent chez les hommes présentant un ratio > 5 de *DEPDC5* ARNm tissu tumoral/tissu normal, alors que ce n'est pas le cas chez les femmes (278). La fonction de *DEPDC5* n'est toujours pas connue mais il semblerait que la protéine soit impliquée dans des signaux de

transduction intracellulaire selon l'homologie entre les domaines DEP de DEPDC5 et ceux de Dishevelled-1.

Une autre étude en « cas/témoins », toujours composée de patients japonais, ayant un HCC lié au VHC ou chroniquement infecté par le VHC sans HCC, a montré une association entre le rs2596542, situé en 5' de la région du MHC class I polypeptide-related sequence A (*MICA*) et le HCC lié au VHC (OR = 1.39). Plus précisément, l'association de *MICA* concerne la progression vers le HCC depuis l'infection chronique et non la susceptibilité de la chronicité de la maladie (279). *MICA* est une protéine membranaire qui agit comme un ligand pour les cellules NK groupe 2D (NKG2D). Cette interaction *MICA*/NKG2D déclenche la cytotoxicité médiée par les NK vers les cellules infectées. C'est aussi une molécule importante dans la surveillance immunitaire de la sénescence cellulaire (280). Le niveau d'expression de *MICA* est plus faible chez les individus présentant un allèle A du rs2596542. Cela pourrait expliquer la faible activation des cellules NK et du système immunitaire chez les patients infectés chroniques possédant un allèle A du rs2596542 et donc une progression vers le HCC (279). Cette association a été répliquée chez des patients caucasiens mais avec un effet inverse à la première GWAS, l'allèle A du rs2596542 protégeant du développement d'HCC (281). En affinant cette dernière analyse, un nouveau polymorphisme, le rs2244546 situé dans *HCP5* en amont de *MICA*, a été identifié comme étant un fort prédicteur de HCC lié au VHC dans cette cohorte de caucasiens. Il est à noter que le rs1012068, associé avec le HCC chez les japonais, n'est pas retrouvé chez les caucasiens (281).

4 COINFECTION PAR LES VIRUS DU VIH ET VHC

4.1 Epidémiologie

L'infection par le VHC est fréquemment retrouvée chez les patients atteints par le VIH. Le nombre d'individus coinfectés est estimé à près de 7 millions à travers le monde (282). La coinfection est devenue un évènement plus qu'anecdotique. Les deux virus partagent des voies communes de transmission (voies parentérale, sexuelle et verticale), et l'infection chronique par le VHC émerge comme la cause majeure de comorbidité, après complications liées au SIDA, chez les individus infectés par le VIH ayant un accès aux thérapies par HAART (*highly active antiretroviral therapy* en anglais) (283).

Le VHC se transmet 10 fois plus efficacement que le VIH *via* une exposition parentérale, et les patients coinfectés sont souvent en premier touchés par le VHC (284, 285). Avant 1990, l'injection de drogue intraveineuse et les transfusions sanguines répétées ont toujours été identifiées comme étant les facteurs de risque les plus importants (286).

Concernant la transmission par voie sexuelle, le VHC est transmis moins efficacement que le VIH (104, 105, 107). Malgré tout, une explosion mondiale d'infection par voie sexuelle du VHC est apparue au cours des dernières années chez des patients contaminés par le VIH. En effet, énormément de cas ont été rapportés parmi la communauté homosexuelle masculine aux Etats-Unis et en Europe (287). Ces infections résultent d'une agrégation de plusieurs facteurs de risque non reliés à la toxicomanie et à un comportement sexuel à haut risque (288).

4.2 Histoire naturelle de la coinfection

4.2.1 Effet de la coinfection VIH/VHC sur l'histoire naturelle du VHC

La coinfection par le VIH a un impact significatif sur le cycle répliatif du VHC ainsi que sur l'histoire naturelle de l'infection par le VHC. Des charges virales plasmatiques VHC plus élevées ont été montrées parmi les patients coinfectés VIH/VHC (289, 290) et la virémie a été inversement corrélée aux taux de CD4⁺ (291). Une augmentation d'ARN VHC hépatique a aussi été observée chez les individus coinfectés VIH/VHC (292). Cet effet serait dû à l'immunodépression liée au VIH ou bien à une interaction entre les deux virus. Depuis

l'introduction des thérapies par HAART, les individus coinfectés montrent une légère augmentation paradoxale des niveaux d'ARN VHC sériques (293, 294), suggérant que l'immunosuppression et un effet direct du VIH ne sont pas les seuls facteurs impliqués.

L'impact majeur de la coinfection VIH/VHC sur l'évolution de l'infection liée au VHC est une accélération de la progression des maladies hépatiques. La coinfection montre une mortalité plus élevée que lors des cas de monoinfection (295). Avant l'ère des thérapies par HAART, la mortalité des coinfectés était due au SIDA masquant d'autres causes potentielles. Cependant, depuis l'introduction des thérapies VIH efficaces, la mortalité liée au SIDA a fortement diminué et la comorbidité associée aux maladies du foie a émergé comme la cause principale de décès chez les patients infectés par le VIH et spécialement parmi les coinfectés VIH/VHC (296).

Les complications en phase terminale de maladies hépatiques ou un HCC sont associés à la principale cause de mortalité parmi les patients coinfectés VIH/VHC (297). Plusieurs études ont démontré un degré d'inflammation hépatique et des taux de progression de fibrose plus élevés lors de la coinfection VIH/VHC par rapport à la monoinfection VHC (284), et les biopsies hépatiques de patients coinfectés montrent des fibroses avancées dès la phase aiguë de l'infection par le VHC (298). Il a aussi clairement été démontré que la coinfection augmente le risque de déclarer une cirrhose de deux fois par rapport à l'infection VHC seule (299).

4.2.2 Effet de la coinfection par le VHC sur les maladies associées au SIDA

Le VHC ne semble pas avoir d'impact majeur sur l'évolution du SIDA. Avant l'ère des thérapies par HAART, aucun effet du VHC n'a pu être observé sur la progression du SIDA (300). Depuis, plusieurs études ont cherché à identifier des effets mais les rapports étaient souvent contradictoires (301). Selon The Swiss Cohort Study, l'infection par le VHC accélère la progression du SIDA et la mort des patients (302) alors qu'une étude américaine n'a pu trouver de différences entre des patients coinfectés et monoinfectés VIH (303). Cependant, les deux études se basaient sur des cohortes dont la date de séroconversion VIH était inconnue et l'analyse présentait des biais potentiels (304). Il est aussi difficile de savoir si la coinfection VIH/VHC est associée à la restauration des propriétés des cellules CD4⁺ par la thérapie par HAART. Encore une fois, les études réalisées sont contradictoires, et montrent souvent des effets faibles (305).

4.2.3 Les mécanismes d'interaction VIH/VHC

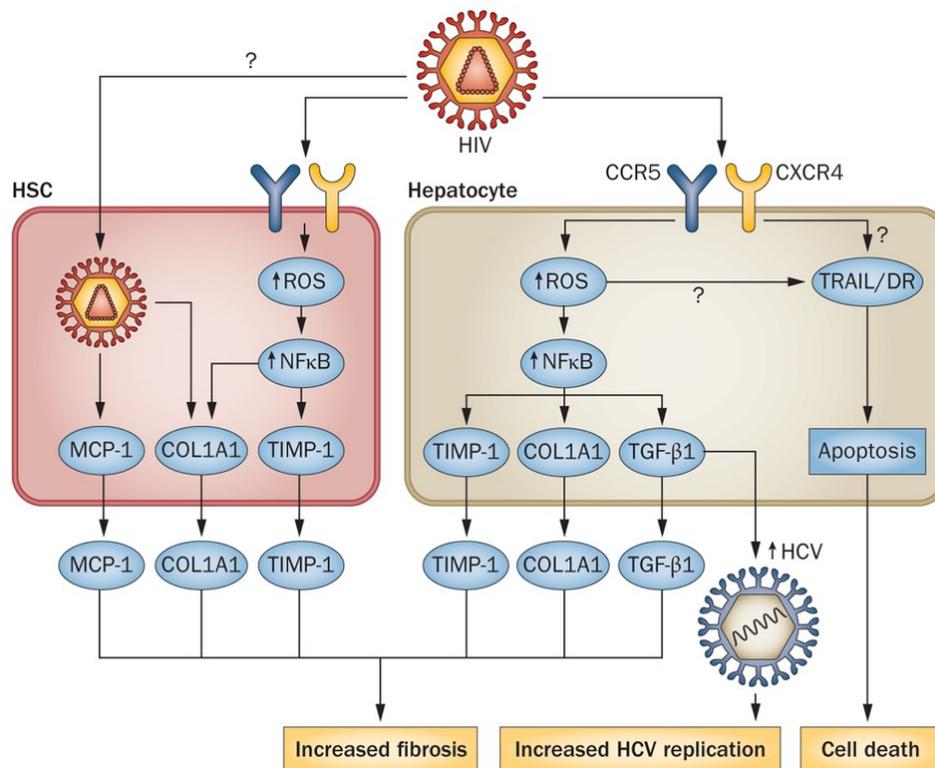


Figure 28. Interaction entre le VIH et le VHC dans les hépatocytes et les HSC, contribuant à la fibrose hépatique. Le VIH active plusieurs mécanismes qui augmentent la fibrose, la réplication du VHC et l'induction de l'apoptose (306). Abréviations: CCR5, C-C chemokine receptor type 5; COL1A1, type 1 collagen; CXCR4, C-X-C chemokine receptor type 4; DR, death receptor; HSC, hepatic stellate cell; MCP-1, monocyte chemoattractant protein 1; ROS, reactive oxygen species; TGF-β1, transforming growth factor beta-1; TIMP-1, tissue inhibitor of metalloproteinases; TRAIL, TNF-related apoptosis-inducing ligand.

Plusieurs mécanismes pathogéniques pourraient expliquer l'accélération du taux de progression des maladies hépatiques chez les individus coinfectés (Figure 28). Bien que le VIH ne se transmette pas directement aux hépatocytes, il a été montré que le VIH améliore la réplication du VHC dans des cellules hépatiques *in vitro*. Cet effet pourrait être médié par l'interaction entre la gp120 du VIH et les corécepteurs CCR5/CXCR4 sur les hépatocytes, par l'intermédiaire du TGF-β (307). Le TGF-β est un médiateur clé dans le processus de fibrose du foie, car il est l'une des cytokines les plus profibrogénique. Le VIH peut également favoriser la fibrogenèse par l'induction de la production d'espèces réactives d'oxygène par les hépatocytes et les HSC, par l'intermédiaire de la voie de NFκ-B dépendante ; cet effet est renforcé en présence du VHC. Le VIH peut également induire l'apoptose des hépatocytes par une sensibilité accrue au TRAIL

(*TNF-related apoptosis-inducing ligand* en anglais) induisant l'apoptose (308). L'activation systémique du système immunitaire lors de l'infection VIH a été associée à l'activation des HSC, qui ont un rôle central dans le développement de la fibrose (voir Introduction section 3.4.3). L'infection directe des HSC par le VIH a été documentée, bien que le mécanisme exact reste inconnu (309). De plus, l'activation des HSC altère l'activité anti-fibrotique des cellules NK, observée lors de l'infection au VIH, en raison de la perte de fonction des cellules CD4⁺ (310).

L'infection au VIH a été associée à des taux sanguins élevés d'hepcidine, plus que ceux observés lors de la monoinfection VHC et la coinfection VIH/VHC (311, 312). L'hepcidine est une hormone peptidique qui régule l'homéostasie du fer. L'augmentation des réserves de fer hépatique dans la monoinfection VHC et la coinfection VIH/VHC reste à prouver. Néanmoins, le fer hépatique a été démontré comme étant stimulant des HSC et affecte négativement la progression de la fibrose parmi les individus monoinfectés par le VHC (313).

Les réponses immunitaires anti-VHC sont compromises lors de la coinfection par le VIH. Un taux faible de CD4⁺ engendre des réponses immunitaires T CD8⁺ spécifiques du VHC largement atténuées (314). La population virale d'infection VHC semble plus diversifiée dans les cas de coinfections VIH/VHC. Ceci reflète la faible pression de sélection par le système immunitaire affaibli. De ce fait, la forte hétérogénéité en quasi-espèces pourrait affecter négativement la réponse au traitement par le PEG-IFN (315).

Le VIH conduit à une dérégulation et à une activation chronique du système immunitaire. La baisse des CD4⁺ dans le tissu lymphoïde associé à l'intestin se produit relativement tôt dans l'évolution de l'infection par le VIH et conduit à une translocation microbienne à travers la muqueuse intestinale avec la forte présence de LPS dans la circulation sanguine (75). Ces niveaux élevés en LPS circulant ont été associés à un risque plus élevé de développement d'une cirrhose lors de la coinfection VIH/VHC (316).

4.3 Traitement des individus coinfectés VIH/VHC

La diminution de la mortalité des patients coinfectés par le VIH, suite à l'utilisation généralisée des HAART après 1996, a été contemporaine de l'amélioration du traitement anti-VHC. Ces avancées majeures ont été validées par des essais thérapeutiques pour les patients coinfectés VIH/VHC. Plusieurs essais thérapeutiques ont été concluants avec la combinaison PEG-IFN/RBV, thérapie anti-VHC, chez les individus coinfectés. A noter que les taux de guérison SVR fluctuaient selon le génotype VHC et étaient toujours inférieurs aux patients monoinfectés

VHC (317). De façon similaire à ce qui est observé chez les mono-infectés VHC, une réponse virologique précoce (négativation de la charge virale VHC à la douzième semaine de traitement), chez les individus co-infectés VIH/VHC, prédit une SVR associée à une non-progression, une amélioration hépatique (318) et la réduction significative de la décompensation hépatique ou d'un HCC (319).

Plus récemment, les nouveaux traitements par DAA contre le VHC ont démontré une efficacité absolument identique chez les co-infectés VIH/VHC et chez les mono-infectés VHC. Une guérison virologique soutenue est obtenue dans plus de 90% des cas même chez des patients cirrhotiques. Seule une vigilance est requise quant aux nombreuses interactions entre les HAART et DAAs chez ces patients co-infectés (320).

4.4 Hépatotoxicité des thérapies par HAART

Plusieurs rapports d'hépatotoxicité ont émergé après l'avènement des thérapies par HAART. Tous faisaient observations d'une association avec un risque élevé d'hépatotoxicité dû aux HAART (321, 322). A priori, la plupart des cas d'élévation enzymatique ne sont pas associés à des symptômes cliniques et se résolvent avec la modification du traitement. Cependant, de graves complications peuvent apparaître et une sévère toxicité hépatique est observable (définie par un taux de ALAT 10 fois supérieur à la normale) (323).

Les mécanismes d'hépatotoxicité liés aux HAART ne semblent pas être si différents entre la co-infection VIH/VHC et la mono-infection VIH (324), mis à part le développement potentiel de lésions hépatiques suite à la restauration immunitaire et l'aggravation des dommages immuno-médiés des hépatocytes infectés VHC (325). Les facteurs de risque d'hépatotoxicité liés aux HAART, parmi les co-infectés VIH/VHC, incluent le degré de fibrose préexistante et l'infection par le génotype 3 du VHC (324). Certaines molécules telles que la névirapine, la stavudine, la didanosine ont été particulièrement incriminées mais sont désormais très peu ou très exceptionnellement prescrites. L'élimination de l'infection VHC par le traitement PEG-INF/RBV, et donc l'atteinte de la réponse SVR, est associée à une réduction des risques d'hépatotoxicité liés aux HAART (326), donnant un argument supplémentaire à traiter ces patients pour le VHC le plus tôt possible.

5 OBJECTIFS DE MA THESE

J'ai rejoint l'équipe du Pr Zagury au cours de l'année 2011 en qualité d'ingénieur pour un projet d'une année financé par le SIDACTION en collaboration avec le laboratoire « Interaction hôte/virus » (Inserm, U1016, Hôpital Cochin, Paris) dirigé par le Dr Emiliani. Le projet consistait en l'analyse bioinformatique de données obtenues par CLIP-seq (*Cross-Linking Immunoprecipitation* suivi de séquençage haut-débit) afin d'identifier des miRNA mobilisés lors de l'infection par le VIH-1, ainsi que leurs cibles pour comprendre les voies touchées (327).

En parallèle de ce projet, j'ai pu me familiariser avec les approches génomiques mises en œuvre au laboratoire GBA et acquérir la capacité d'analyser des données de génotypage. A la fin de mon contrat SIDATION, fin 2012, le Pr Zagury m'a proposé de réaliser une thèse au sein du laboratoire GBA et j'ai décidé de m'engager afin d'enrichir mes connaissances et mon expérience en génomique.

Mon projet de thèse a consisté à analyser par GWAS la cohorte ANRS CO 13 HEPAVIH comprenant des patients coinfectés par le VIH et le VHC. L'objectif principal de ce travail était l'identification des facteurs génétiques impliqués dans le développement des symptômes liés à la coinfection VIH/VHC et notamment la progression de la fibrose hépatique bien documentée sur le plan clinique dans cette cohorte. Ce projet avait l'intérêt de pouvoir mettre en valeur des interactions possibles entre les virus VHC et VIH pour des phénotypes déjà bien étudiés en mono-infection pour chacun d'entre eux. Mon projet de thèse s'est déroulé selon les étapes suivantes :

- Exploration des données démographiques de la cohorte ANRS CO 13 HEPAVIH afin de définir les phénotypes et les covariables.
- Exploitation de données génomiques issues des puces de génotypage par étude d'association « génome entier ».
- Identification et interprétation biologique d'associations génétiques liées à la fibrose hépatique dans la coinfection VIH/VHC.

Deuxième partie
Matériels et méthodes

1 OUTILS GENOMIQUES ET BIOINFORMATIQUE

1.1 Bases de données bioinformatiques

La première GWAS générant des associations significatives a été publiée en 2005 (328). Depuis, le nombre de GWAS publiées chaque année n'a cessé d'augmenter et on recense plus de 1 900 études publiées (35). Cette croissance rapide s'explique notamment par l'émergence de plusieurs bases de données utilisées pour réaliser des GWAS.

1.1.1 dbSNP

La base de données « SNP database », connue sous le nom de dbSNP, est une base de données publique qui a pour but de répertorier les variations génétiques recensées chez plusieurs espèces animales dont l'être humain (5). Les SNPs ne sont pas les seuls polymorphismes à être renseignés dans cette base, les indels ou les microsatellites y sont également inclus. Au total, plus de 60 millions de polymorphismes sont actuellement décrits dans cette banque. Cette base de données représente une source d'information majeure pour la communauté scientifique puisque le but de ce catalogue est d'annoter de façon précise les polymorphismes découverts dans le génome. Ainsi, lorsqu'un SNP est répertorié dans la base de données, des informations concernant sa localisation, ses conséquences fonctionnelles possibles ou encore les fréquences alléliques sont disponibles. De plus, lorsque des publications relatives à un SNP existent, la fiche dbSNP de ce site les recense et propose de liens vers celles-ci. Par conséquent, cette base de données constitue un outil majeur dans les études génomiques en permettant de connaître rapidement et facilement les caractéristiques d'un polymorphisme.

1.1.2 Le projet HapMap

Les haplotypes constituent une information essentielle dans les études génétiques. Ainsi, en 2002, un consortium international a entrepris une cartographie des haplotypes dans plusieurs populations : le projet HapMap (12, 329). Ce projet s'est déroulé en trois phases.

-Phase I du projet, environ 1 million de SNPs génotypés chez 270 individus originaires de 4 populations différentes :

- 90 nigériens (Yoruba d'Ibadan) formant 30 trios (deux parents et un enfant),

- 90 résidents des Etats-Unis originaires d'Europe du Nord et d'Europe de l'Ouest formant 30 trios,
- 45 japonais non apparentés de la région de Tokyo,
- 45 chinois non apparentés de la région de Pékin

-Phase II du projet, augmentation du nombre de SNPs génotypés : environ 3,1 millions de SNPs chez les mêmes individus (330).

-Phase III du projet, augmentation de la densité de SNPs génotypés et du nombre d'individus, dont certains issus de populations non étudiées dans les phases précédentes (331) :

- Pour les populations originelles du projet, 90 nouveaux individus africains, 90 nouveaux individus d'origine européenne et 45 individus pour chacune des deux populations asiatiques ont été rajoutés au projet. Environ 5 millions de polymorphismes, dont des CNVs, ont été génotypés pour ces individus.
- Sept nouvelles populations (des afro-américains, des chinois résidant aux Etats-Unis, des mexicains résidant aux Etats-Unis, des toscans, des kenyans de la tribu Luhya, des kenyans masaïs et des indiens Gujarati vivant aux Etats-Unis) ont été ajoutées au projet. Environ 1,4 millions de SNPs ont été génotypés pour ces individus.

L'intérêt du projet HapMap est multiple. Tout d'abord, les données obtenues par le projet, et en particulier les génotypes, ont été mises gratuitement à la disposition de la communauté scientifique constituant une source d'information importante. De plus, il a permis d'identifier et de caractériser une grande partie des SNPs fréquents (dont la fréquence de l'allèle mineur est supérieure à 5%) dans plusieurs populations. Ainsi, le projet permet de connaître la fréquence de ces SNPs au sein des différentes populations génotypées. Enfin, l'inclusion de trios dans certaines populations a permis une cartographie des haplotypes. Etant donné l'existence de blocs haplotypiques au sein du génome (voir Introduction section 1.4.3), beaucoup des SNPs génotypés dans ce projet sont corrélés entre eux. La cartographie des haplotypes a alors permis la caractérisation des déséquilibres de liaison pour les paires de SNPs d'une même région. Cela a conduit à l'identification d'un nombre plus restreint de SNPs représentant la même information génétique qu'avec l'ensemble des polymorphismes. De tels SNPs sont appelés tagSNP (Figure 29). On estime ainsi que seulement 600 000 SNPs permettraient de capturer la

diversité des 10 millions de SNPs fréquents du génome. Cette découverte a largement été utilisée pour la conception des puces de génotypage.

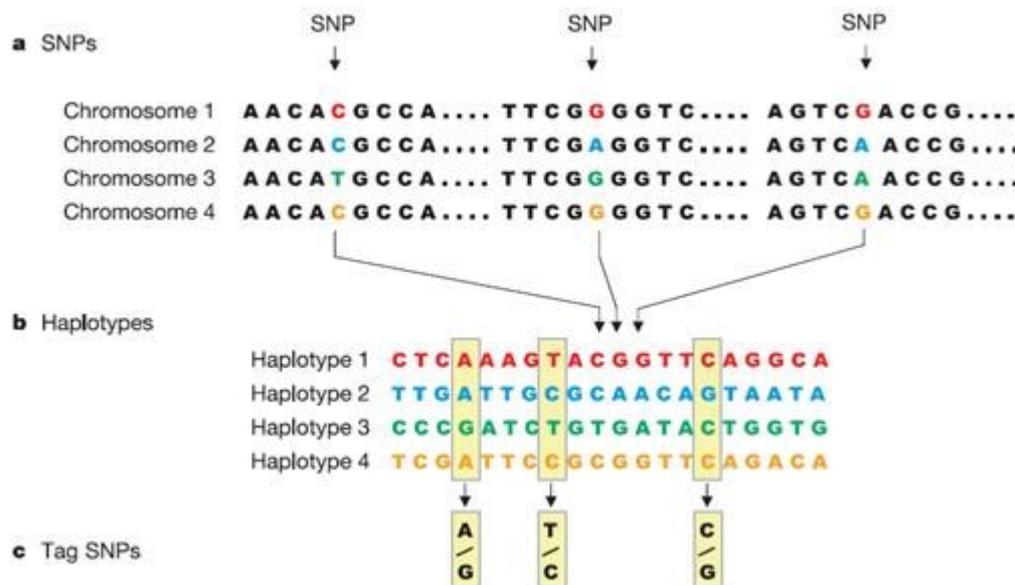


Figure 29. Notion de "tagSNP" (12). (a) Identification de 3 SNPs (en couleur) dans une portion chromosomique. (b) Reconstruction des haplotypes constitués de 20 SNPs dont les trois identifiés précédemment. (c) Détermination de 3 tagSNPs dont la connaissance suffit pour identifier les 4 haplotypes de la population. Par exemple, un profil G-T-C pour ces 3 tagSNPs correspond toujours à l'haplotype 3.

1.1.3 Le projet 1000 Genomes

Plus récemment, le projet 1000 Genomes, dont l'objectif est de séquencer le génome de 2 500 individus issus de 28 populations différentes, a vu le jour (4). Le séquençage intégral de ces génomes vise principalement à identifier l'ensemble des polymorphismes et en particulier les SNPs de faible fréquence qui ne sont pas inclus dans le projet HapMap. A l'instar de ce dernier, le projet 1000 Genomes s'est déroulé en plusieurs phases se différenciant principalement par le nombre de polymorphismes inclus dans l'étude. Actuellement, 2 577 individus ont été séquencés et environ 80 millions de polymorphismes ont été identifiés. Cependant, certaines précautions doivent être prises concernant les données mises à disposition. En effet, en plus des nombreuses difficultés inhérentes au séquençage, la couverture moyenne est de 4x (un locus est « lu » 4 fois en moyenne), ce qui entraîne une certaine incertitude quant aux polymorphismes repérés. Pour la même raison, les génotypes ne sont donnés que sous la forme de probabilité. Enfin, la faible fréquence de certains polymorphismes peut être à l'origine d'incertitude

concernant leur qualité. En conséquence, les données mises à disposition sont régulièrement corrigées ou mises à jour et il est alors préférable d'utiliser des versions moins complètes mais plus fiables du projet.

Cependant, ce projet ouvre de larges perspectives concernant l'implication de polymorphismes peu fréquents dans les maladies multifactorielles. En effet, le rôle de ces mutations est aujourd'hui largement soutenu (332). Enfin, les données de 1000 Genomes servent également de panel de référence pour de nombreuses problématiques telles que le phasage ou l'imputation.

1.2 Le génotypage et les puces à ADN

Le génotypage vise à déterminer les génotypes d'un polymorphisme chez un individu. Des puces à ADN, permettant de génotyper un nombre croissant de SNPs ont été développées depuis 10 ans. Schématiquement, selon le procédé utilisé par Illumina, une puce est une plaque sur laquelle se trouvent des microbilles, chacune spécifique d'un SNP (Figure 30), recouvertes de fragments d'ADN, appelés oligonucléotides. Ces oligonucléotides correspondent à la séquence située en amont du SNP. L'ADN génomique, préalablement amplifié et segmenté, est ensuite mis en contact avec les microbilles et les fragments se fixent sur les oligonucléotides correspondants. La plaque est ensuite lavée afin d'éliminer les fragments d'ADN qui ne se sont pas fixés. Des terminateurs de synthèse (didésoxyribonucléotides) associés à des fluorochromes, rouge ou vert selon la base, sont ensuite déposés sur la plaque et vont s'ajouter à l'oligonucléotide de manière complémentaire à l'ADN génomique grâce à une polymérase. La couleur rouge ou vert permettant d'indiquer l'allèle observé, la plaque est scannée pour analyser la fluorescence et déterminer les génotypes des SNPs de la puce.

Depuis 10 ans, le nombre de SNPs caractérisés sur les puces ADN a été multiplié et il est désormais possible d'analyser jusqu'à 5 millions de polymorphismes (SNPs et quelques CNVs) par puce. De même, une même plaque peut servir au génotypage de plusieurs individus simultanément, réduisant ainsi le coût des procédures expérimentales.

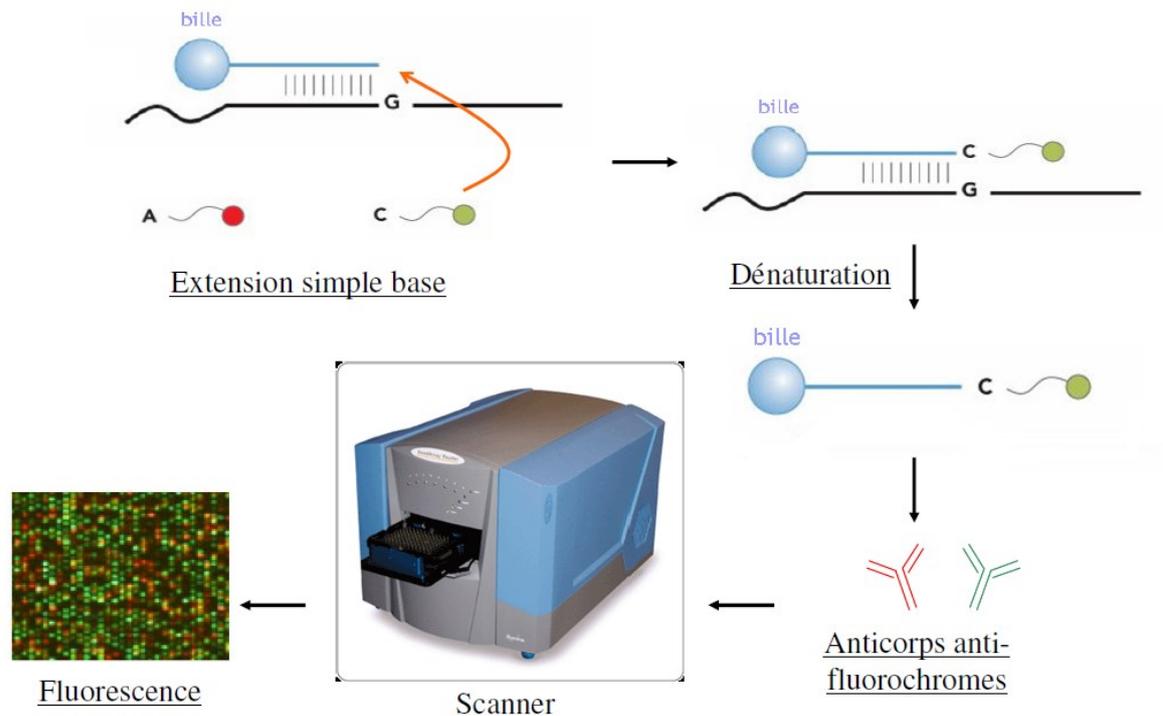


Figure 30. Schématisation des étapes du génotypage selon le procédé Illumina.

Le développement des projets HapMap (12) et 1000 Génomes (4) a largement contribué à l'augmentation de la couverture des puces. En effet, bien qu'utilisant des stratégies différentes, les deux grandes sociétés de génotypage, Affymetrix et Illumina, ont utilisé les SNPs identifiables à l'aide du projet HapMap. Affymetrix a initialement basé le choix des SNPs présents sur les puces en fonction de leur position sur le génome et de leur fréquence dans la population d'intérêt pour essayer d'assurer une couverture homogène du génome alors que la stratégie développée par Illumina utilise les tagSNPs permettant de capturer le plus de variabilité génétique possible.

1.3 Reconstruction des haplotypes

Considérons par exemple deux SNPs S1 et S2 dont les allèles respectifs sont A/a et B/b. Pour chacun de ces deux SNPs, trois génotypes sont possibles (AA, Aa ou aa pour S1) et les techniques de génotypage donnent cette information pour chaque individu. Toutefois, ces données ne permettent pas de connaître la phase de ces allèles, c'est-à-dire les haplotypes (voir Introduction section 1.4.3). En effet, supposons que les génotypes d'un individu pour S1 et S2 soient respectivement Aa et Bb. Dans ce cas, deux paires d'haplotypes (AB/ab ou Ab/aB) peuvent être observés chez cet individu. Les méthodes d'haplotypage visent à attribuer une

paire d'haplotypes à chaque individu d'un groupe à partir des génotypes de ce groupe (Figure 31).

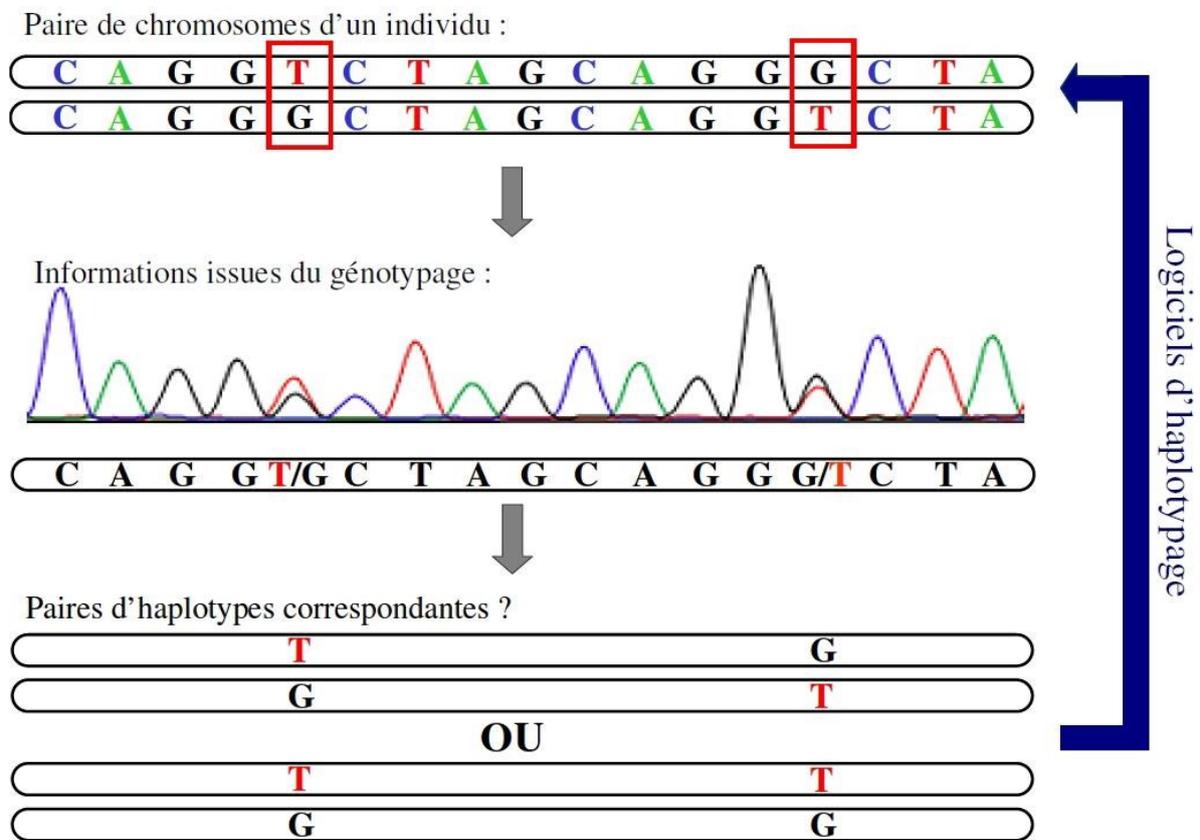


Figure 31. Problématique de l'haplotype

Plusieurs méthodes ont été développées afin de reconstruire les haplotypes. Les méthodes combinatoires étudient tous les haplotypes possibles et n'en conservent qu'une paire pour chaque individu en se basant sur un critère de parcimonie (333) ou de phylogénie (334, 335). Des méthodes d'inférence statistique ont été développées et se basent sur la vraisemblance en recherchant les haplotypes qui sont les plus probables étant donnés les génotypes observés (336, 337). Enfin, plus récemment, des méthodes bayésiennes utilisant des *a priori* biologiques et des modèles de coalescence ont été mises au point (338-340). Ces dernières méthodes semblent être les plus précises (341).

1.4 Imputation

L'imputation vise à déterminer les génotypes de SNPs absents de la puce en se basant sur les informations de déséquilibre de liaison et des taux de recombinaison des haplotypes mesurés à l'aide d'un panel de référence (Figure 32). Ces panels de référence sont constitués d'un grand

nombre de données haplotypées renseignant un nombre de SNPs supérieur à celui de marqueurs présents sur la puce. Généralement, les données de grands projets comme 1000 Genomes ou Hapmap sont utilisées.

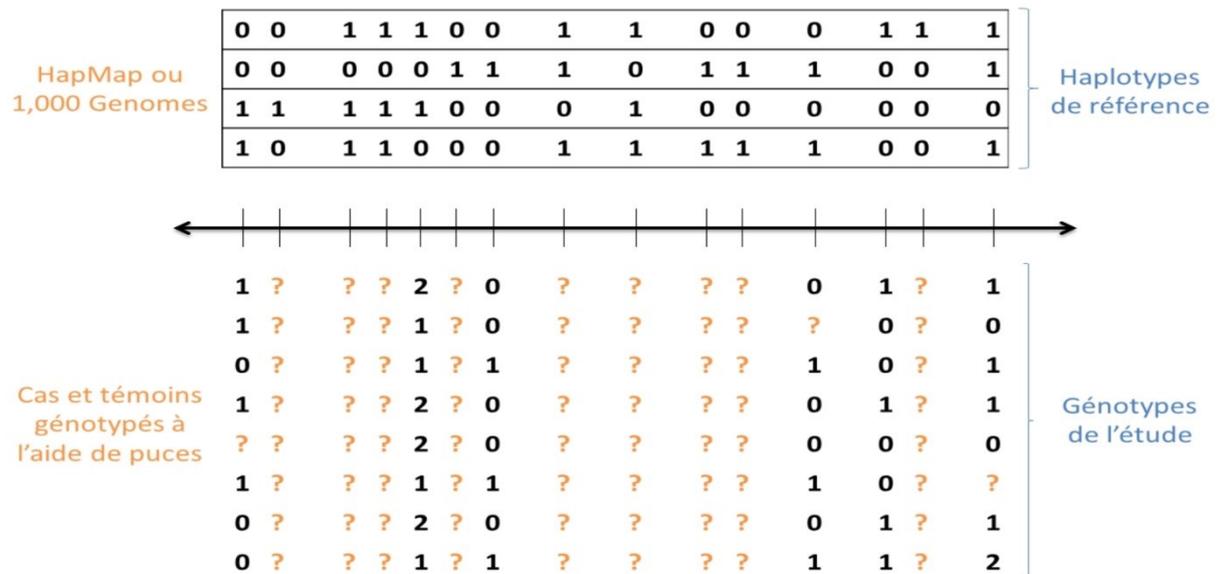


Figure 32. Représentation schématique du problème d'imputation. Les génotypes manquants de la puce seront imputés à l'aide des haplotypes d'un panel de référence. Adapté de (342).

Les puces de génotypage sont conçues à partir d'une liste de polymorphismes déterminée *a priori* permettant de capturer la majorité de l'information génétique. Bien que leur développement permette le génotypage d'un nombre important de SNPs (jusqu'à 5 millions), il reste un grand nombre de polymorphismes non renseignés. De plus, les SNPs génotypés d'une puce à l'autre ne sont pas identiques, ce qui rend souvent difficile la réplication des résultats obtenus d'une GWAS à l'autre. Enfin, certains SNPs peuvent être exclus d'une étude GWAS lors des contrôles de qualité en raison d'un trop grand nombre de données manquantes.

Les méthodes d'imputation développées au cours de ces dernières années peuvent apporter une réponse à ces problématiques en permettant : la connaissance de génotypes de SNPs absents de la puce de génotypage, la caractérisation des données manquantes, la comparaison de jeux de données obtenus par l'intermédiaire de puces différentes, l'augmentation de la puissance statistique de l'étude, la caractérisation plus fine de régions chromosomiques identifiées lors d'une GWAS.

En pratique, la qualité de l'imputation dépend grandement de la taille du panel de référence. Plus ce panel inclut de SNPs et d'individus, plus la caractérisation des déséquilibres de liaison est fine et plus l'inférence des SNPs est fiable. Cependant, la complexité de l'imputation et les temps de calcul augmentent avec la taille du panel. Pour cette raison, il peut être très efficace de réaliser l'imputation à partir de données phasées car cette étape de reconstruction des haplotypes permet un gain de temps considérable tout en assurant une qualité d'imputation quasi-identique (343).

2 ETUDES D'ASSOCIATION GENOME ENTIER

2.1 Population étudiée

2.1.1 La cohorte ANRS CO 13 HEPAVIH

La collaboration ANRS CO13 HEPAVIH est un dispositif de cohorte prospective multicentrique de patients coinfectés par le VIH et le VHC, inclus et suivis actuellement dans 17 services hospitaliers français (CHU Cochin, Pitié-Salpêtrière, Tenon, Avicenne, Bichat - Claude-Bernard, Saint Louis, Saint Antoine, Bicêtre, Paul-Brousse, Necker, à Paris; CHU Sainte-Marguerite à Marseille, CHU Purpan et Hôpital Joseph-Ducuing à Toulouse, CHU l'Archet à Nice, CHU Pellegrin, Saint-André et Haut-Levêque à Bordeaux). Cette collaboration a été mise en place afin de disposer d'un outil de suivi prospectif pour répondre aux diverses questions que pose la prise en charge des patients coinfectés VIH/VHC mais aussi pour décrire l'histoire naturelle de cette coinfection en termes de morbidité et de mortalité et ses déterminants afin de mieux comprendre les mécanismes mis en jeu.

Les critères d'inclusion des patients dans la cohorte ANRS CO13 HEPAVIH sont :

- La majorité (≥ 18 ans) ;
- Une coinfection par le VIH-1 et le VHC, confirmée par un test positif anti-VIH-1 et détection d'ARN VHC par PCR ;
- Etre suivi dans l'un des centres cliniques associés ;
- Avoir donné un consentement écrit.

Il s'agit d'une cohorte prospective de type fermé. La participation des différents centres cliniques est volontaire. Des questionnaires standardisés permettent la collecte confidentielle des données épidémiologiques, socio-comportementales, cliniques, biologiques, thérapeutiques, ainsi que des données relatives à l'évaluation de l'atteinte hépatique (ponction-biopsie hépatique PBH, Fibroscan®, Fibrotest®).

La durée de suivi est de cinq ans renouvelables. Cette durée est justifiée par la durée de l'évolution de la maladie hépatique. Deux types de suivi sont proposés selon le stade histologique de l'hépatite C : un suivi annuel pour les patients non porteurs de cirrhose (suivi standard) et un suivi semestriel pour les patients ayant une cirrhose (suivi cirrhose).

La période des inclusions s'est étalée sur environ trois ans, à partir de fin 2005. Au total, 1166 patients ont été inclus dans la cohorte. Les données de 494 de ces patients ont servi à l'analyse lors de ma thèse.

2.2 Phénotypes analysés

2.2.1 Méthodes d'évaluation de la fibrose hépatique

L'évaluation initiale de la fibrose hépatique, à l'exception du diagnostic évident de cirrhose, peut être fondée sur la réalisation de l'une des trois méthodes actuellement validées: PBH, Fibrotest® ou Fibroscan®. Les résultats de ces méthodes sont généralement convertis en scores reflétant le grade d'activité et le stade d'avancée de la fibrose.

2.2.1.1 La PBH

La PBH est l'examen de référence pour évaluer la fibrose hépatique. L'évaluation de l'atteinte hépatique au cours de l'hépatite C guide et conditionne l'attitude thérapeutique et la surveillance du patient. Elle a un intérêt diagnostique par la recherche d'autres comorbidités ayant une valeur pronostique sur la progression de la fibrose telles que la stéatose. Ainsi, l'importance de la fibrose est décrite en routine de façon semi-quantitative en utilisant des scores de précision variables qui sont parfois peu reproductibles hormis le score METAVIR (344).

La PBH consiste en un prélèvement d'un petit fragment de foie, réalisé sous anesthésie locale ou générale, par voie transpariétale (percutanée intercostale droite) ou par voie transjugulaire (en cas de troubles de l'hémostase). Elle est réalisée sous guidage échographique afin d'améliorer la qualité du prélèvement. Les limites de ce geste sont la nécessité d'un bilan prébiopsie, des effets secondaires, son coût, et les contraintes pratiques qu'il engendre. De plus, ce geste invasif comporte des risques de morbidité de 0,3 % à 0,6 % et un risque de mortalité de 0 à 0,05 % (345).

2.2.1.2 Marqueurs non-invasifs

Le Fibrotest® et l'Actitest® sont deux marqueurs sériques non invasifs de l'activité nécrotico-inflammatoire et de la fibrose hépatique. Le Fibrotest® combine cinq marqueurs (alpha-2-macroglobuline, haptoglobine, apolipoprotéine A1, bilirubinémie totale, gamma-glutamyl transpeptidase) avec un ajustement sur le sexe et l'âge. L'Actitest® requière en plus le dosage des ALAT. Le résultat, combinant les différentes valeurs biologiques obtenues, est compris entre 0 et 1 et ce dernier est converti en score METAVIR.

Les principales limites du Fibrotest® (346) sont constituées par les conditions suivantes susceptible d'induire des faux positifs : hémolyses chroniques, sepsis, maladie de Gilbert (touchant 5 % de la population générale), hépatite aiguë, syndrome inflammatoire, cholestase extra-hépatique, traitements par anti-protéases (atazanavir et indinavir) parfois responsables d'une hyperbilirubinémie, patients porteurs de valve cardiaque, causes associées d'hépatopathie, dont certaines sont plus fréquentes chez les patients coinfectés VIH/VHC.

2.2.1.3 L'élastographie transitoire

L'élastographie transitoire, évaluée par le Fibroscan® (EchoSens, Paris, France) (347), est une méthode non-invasive de diagnostic de la fibrose hépatique. Le principe est de mesurer l'élasticité du foie à partir d'une onde sonore à basse fréquence qui est générée par une sonde placée à la surface de la peau. La vitesse de propagation de l'onde dans le foie, mesurée par ultrason, permet d'estimer l'élasticité du foie. Plus le foie est rigide, plus l'onde se propage rapidement. La mesure obtenue est en kilo Pascal (kPa) et peut être convertie en score METAVIR grâce à une table de correspondance.

Cette méthode est performante pour le diagnostic d'une fibrose avancée ou d'une cirrhose et pour prédire les complications associées à la maladie hépatique. Cependant ses limites sont l'applicabilité, la reproductibilité et l'influence d'autres facteurs que la fibrose sur l'élasticité du foie. Il est ainsi nécessaire d'effectuer 10 mesures au minimum avec un niveau de reproductibilité acceptable (réussite de mesure > 60%, disparité des valeurs : interquartile < 30%). Néanmoins, cette méthode présente une meilleure précision diagnostique que les marqueurs sériques chez les patients coinfectés VIH/VHC.

2.2.1.4 Le score METAVIR

Le score METAVIR, proposé en 1996 par des hépatologues français (344), sert à estimer le grade d'activité et le stade d'avancée de la fibrose. Le grade d'activité (A) traduit l'activité

hépatique et la rapidité de l'évolution alors que le stade de fibrose (F) caractérise les lésions fibreuses hépatiques existantes. C'est une évaluation semi-quantitative qui permet de « normaliser » la plupart des méthodes d'évaluation de la fibrose hépatique. Concernant la PBH, la conversion se fait par observation des coupes histologiques hépatiques. Les caractéristiques principales du score de Fibrose (F) sont résumées dans le Tableau 1.

Tableau 1. Classification METAVIR (344)

Score Activité		Score Fibrose	
A0	sans activité	F0	sans fibrose
A1	activité minimale	F1	fibrose portale sans septa
A2	activité modérée	F2	fibrose portale et quelques septas
A3	activité sévère	F3	fibrose septale sans cirrhose
-	-	F4	cirrhose

Pour la conversion du score quantitatif obtenu par Fibroscan®, il a été nécessaire de définir des limites pour chaque stade de l'échelle METAVIR (Figure 33). A noter que les deux premiers stades METAVIR ne sont pas significativement distinguables par le test Fibroscan®.

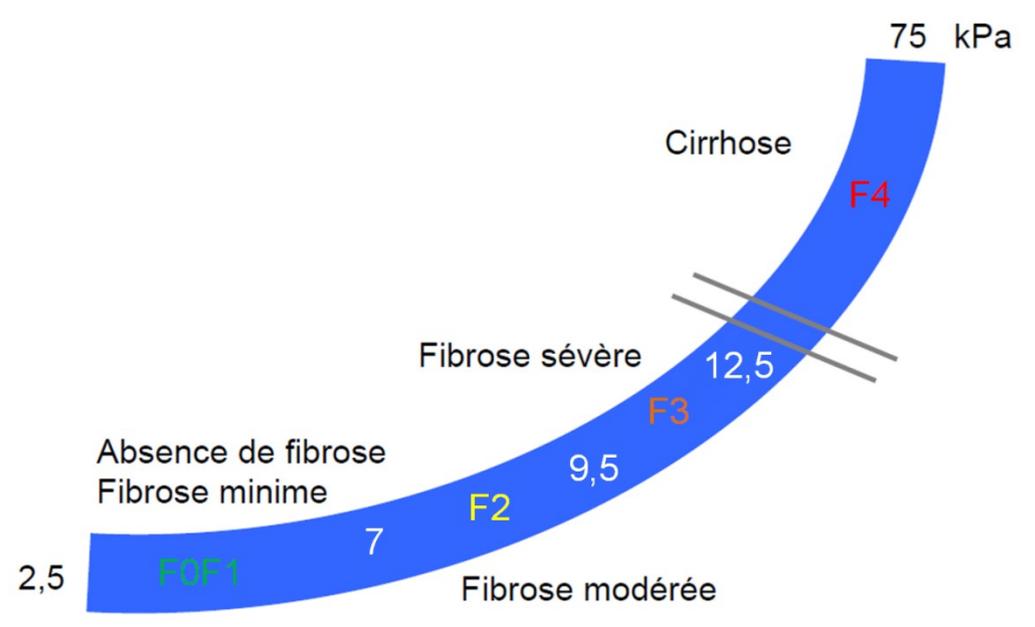


Figure 33. Echelle des valeurs d'élasticité hépatique en kPa (2.5, 7, 9.5, 12.5, 75) selon le stade de fibrose et sa correspondance en score METAVIR (stade F du vert au rouge).

2.2.2 Définition des phénotypes analysés

Pour chacun des individus, il a été estimé un score de fibrose METAVIR à l'inclusion et un suivi de l'élasticité du foie par Fibroscan®.

A l'inclusion, le score METAVIR a été défini par différents examens selon les individus, une PBH (22.5%), un Fibrotest® (23%) ou un Fibroscan® (54.5%). Ainsi, un phénotype dichotomique a été défini : fibrose minimale ou modérée pour un score METAVIR F0, F1 ou F2 et fibrose sévère ou une cirrhose pour des scores F3 ou F4. Ce phénotype sera, dès lors, décrit comme METAVIR F0F1F2 vs F3F4 représentant la comparaison de ces deux groupes.

Lors du suivi des patients après inclusion, l'élasticité hépatique a été évaluée par Fibroscan® au cours du temps. Afin de définir un phénotype quantitatif, il a été décidé de sélectionner la première valeur de Fibroscan® obtenue lors du suivi, en démarrant dès l'inclusion si un examen de Fibroscan® était disponible. La variance et l'asymétrie de la distribution de ce phénotype quantitatif ont été réduites en normalisant l'élasticité hépatique par transformation de la racine carrée. Ce phénotype sera, dès lors, décrit comme sqrtLS (*square root transformation of Liver Stiffness*).

2.3 Covariables utilisées dans l'analyse statistique

Plusieurs facteurs susceptibles d'avoir une influence sur l'évaluation de la fibrose ont été pris en compte lors des analyses statistiques :

- L'âge en années
- Le genre (femme ou homme)
- La consommation d'alcool en trois classes (jamais, passée ou courante)
- Le génotype du VHC (1 versus les autres)
- Le temps supposé en années de coinfection VIH/VHC
- Le temps effectif en années de traitement anti-VIH

2.4 Génotypage

Sur les 494 patients prévus pour le génotypage initial, 120 échantillons d'ADN n'ont pas été sélectionnés pour cause de données manquantes au niveau des covariables, indispensables pour la correction des tests statistiques. Les 374 individus restants ont été génotypés en utilisant des puces de génotypage Illumina Omni2.5 BeadChip renseignant 2 391 739 polymorphismes. L'ADN génomique de chacun des individus a été amplifié, fragmenté, dénaturé puis hybridé sur une puce pendant au moins 16 heures à 48°C. Les fragments hybridés de manière non spécifique ont été éliminés après lavage. Les fragments restants, hybridés de façon spécifique, ont été marqués par fluorescence et détectés en utilisant un scanner Illumina IScan. Les intensités normalisées ont ensuite été interprétées en génotypes en utilisant la version 1.6.3 du logiciel GenomeStudio d'Illumina. Pour la suite des analyses, nous nous sommes focalisés uniquement sur les SNPs.

2.5 Contrôle qualité du génotypage

Les données brutes de génotypage ont été analysées à l'aide de la version 1.6.3 du logiciel GenomeStudio développé par Illumina. Tout d'abord, 30 individus ont été éliminés de l'étude en raison d'un taux de SNPs génotypés par individu (*call rate* en anglais) inférieur à 95%. Ensuite, les SNPs avec une proportion d'individus génotypés (*call frequency* en anglais) inférieure à 99% ont été à nouveau inférés. Suite à cela, les individus dont le *call rate* était inférieur à 98% ont été exclus des analyses. Ce protocole correspond aux recommandations faites par Illumina afin de minimiser les erreurs d'inférence et de les corriger manuellement si nécessaire. Deux paires d'individus ont présenté un niveau élevé de parenté en analyse par le logiciel PLINK (348). Le patient de chaque paire avec le moins de données manquantes a été retenu pour l'analyse.

Ensuite, 210 240 SNPs présentant un taux de données manquantes supérieur à 2% ont été exclus. Afin de limiter les erreurs de génotypage, 1 025 509 SNPs dont la fréquence de l'allèle mineur (MAF, *Minor Allele Frequency* en anglais) était inférieure à 5% n'ont pas été retenus pour l'analyse. Enfin, un écart trop important à l'équilibre d'Hardy-Weinberg pouvant être la conséquence de problèmes de génotypage, 312 SNPs dont la *p*-valeur du test exact d'adéquation à l'équilibre d'Hardy-Weinberg était inférieure à 10^{-6} ont également été exclus. Finalement, 1 155 778 SNPs et 342 individus ont été sélectionnés pour les procédures d'imputation et d'analyse.

2.6 Etude de la stratification de la population

Afin de corriger une possible stratification au sein de la cohorte, les génotypes ont été analysés par l'intermédiaire d'une analyse en composantes principales implémentée dans le logiciel EIGENSTRAT issu de la suite EIGENSOFT (349). Au préalable de cette analyse, 23 individus ont été exclus, puisque lors du questionnaire sociodémographique, ils s'étaient décrits comme « non-européen ». Une première analyse, réalisée en incluant deux populations asiatiques de HapMap (Chinois de Pékin, CHB, et Japonais de Tokyo, JPT), deux populations africaines (Yoruba du Nigéria, YRI, et Masaï du Kenya, MKK), ainsi qu'une population d'origine européenne (Caucasien de l'Utah USA, CEU), a permis d'identifier 27 individus atypiques. Une seconde analyse, réalisée sur la cohorte en écartant les individus éliminés lors de la première analyse, a permis de conclure que les individus restants formaient une population homogène par rapport à la population contrôle CEU de HapMap. Au total, 50 individus ont été écartés de l'analyse lors de l'analyse de la stratification. Finalement, une dernière analyse a été réalisée avec les individus restants afin de déterminer les deux premiers vecteurs propres de la stratification, inclus dans les analyses d'associations afin de tenir compte de la structure de la population.

2.7 Imputation

Tout d'abord, les données ont été phasées à l'aide du logiciel SHAPEIT2 (338). Les données phasées ont ensuite été imputées en utilisant le logiciel IMPUTE2 (342). Les génotypes de la population européenne du projet 1000 Genomes (EUR, 1000 Genomes phase 1 integrated release version 3, <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/>) ont été utilisés comme panel de référence. Après toutes les procédures, 8 426 597 SNPs avec un score d'imputation supérieur à 0,8 et une MAF supérieure à 5% sont retenus pour les analyses statistiques.

2.8 Analyses statistiques

Sur les 494 patients initialement sélectionnés pour l'étude, 170 d'entre eux ont été écartés : 120 présentaient des données manquantes au niveau d'un ou plusieurs facteurs ayant une influence sur la progression de la fibrose (95 avec une date de contamination inconnue, 6 avec un statut manquant de consommation d'alcool, 31 sans renseignement sur le temps de traitement anti-VIH et 13 sans génotype du VHC connu), 23 se sont déclarés non-européens et 27 ont été

identifiés comme étant des individus atypiques lors de l'étude de la stratification. Finalement, 292 patients ont été inclus dans les analyses.

Les caractéristiques démographiques des 292 patients ont été résumées statistiquement, grâce au logiciel R version 3.0.2 (<https://cran.r-project.org/>). Les médianes et les écart-types des variables quantitatives ont été calculés, ainsi que les pourcentages de répartition des variables qualitatives et cela en fonction de la date d'examen du Fibroscan® pour le phénotype sqrtLS et de la date d'inclusion pour la comparaison des scores METAVIR F0F1F2 et F3F4. Les statistiques ont aussi été obtenues selon la sévérité de la fibrose par un phénotype METAVIR dichotomique (<F3 : fibroses faibles ou modérées, >=F3 : fibroses sévères ou cirrhoses). Pour la comparaison des scores METAVIR F0F1F2 vs F3F4, les groupes étaient déjà définis. Pour le phénotype sqrtLS, la conversion en groupes METAVIR a été effectuée selon que les données non-transformées du phénotype soient inférieures à 9,5 kPa (< F3) ou supérieures à 9,5 kPa (>=F3), grâce aux seuils de correspondances présentés Figure 33.

Ensuite, pour chacun des 1 155 778 SNPs, les associations entre les génotypes, imputés ou non, et chacun des phénotypes définis ont été mesurées en utilisant des régressions linéaires multivariées pour le phénotype quantitatif sqrtLS et des régressions logistiques multivariées pour la comparaison des scores METAVIR F0F1F2 vs F3F4, à l'aide du logiciel SNPTEST version 2.5.2 (350). Dans les deux cas, les différentes covariables ainsi que les deux premiers vecteurs propres de la stratification ont été inclus dans le modèle.

La correction des tests multiples a été prise en compte pour chaque phénotype analysé. Les *p*-valeurs d'association ont été ajustées selon la méthode de Bonferroni et comparées au seuil de significativité « génome entier » (5×10^{-8}).

2.9 Approfondissement des associations trouvées

2.9.1 Déséquilibre de liaison

Pour chaque SNP identifié lors des études d'association, tous les SNPs en fort déséquilibre de liaison avec ce polymorphisme ($r^2 > 0,8$) dans la population européenne du projet 1000 Genomes (EUR, 1000 Genomes phase 1 integrated release version 3) ont été identifiés. Seule la liste de SNPs distincts a été conservée.

2.9.2 Exploration des résultats par bioinformatique

Pour chaque SNP d'intérêt, nous avons exploré différentes bases de données afin d'établir d'éventuelles associations avec un mécanisme biologique. Plusieurs bases de données d'expression, disponibles publiquement, permettent d'établir des corrélations entre polymorphisme et variation du niveau d'expression des ARNm, ainsi que l'expression tissu spécifique des gènes:

- Genevar (www.sanger.ac.uk/resources/software/genevar/),
- mRNA by SNP Browser (www.sph.umich.edu/csg/liang/asthma/),
- GHS-Express (<http://genecanvas.ecgene.net/uploads/ForReview/>),
- BioGPS (<http://biogps.org/>)
- et plus récemment GTex Portal (www.gtexportal.org/home/).

Des informations sont également disponibles sur des données de méthylation (Genevar, <http://www.sanger.ac.uk/resources/software/genevar/>); de polyadénylation (PolyApred, <http://www.imtech.res.in/raghava/polyapred/submission.html>); de sites de fixation de facteurs de transcription (regulomedb, <http://regulomedb.org/> dérivée de la base de données TRANSFAC et HaploRegv2, www.broadinstitute.org/mammals/haploreg/haploreg.php); gènes ou cibles de miRNA (miRBAs, <http://www.mirbase.org/>; miRTarBase, <http://mirtarbase.mbc.nctu.edu.tw/>; et MicroCosm Targets, <http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/>); et les données d'ENCODE (regulomedb, <http://regulomedb.org/>).

2.9.3 Réplication dans des cohortes de patients mono-infectés VHC

Une étude de réplication a été effectuée dans une cohorte de patients infectés par le VHC seul. Cette cohorte est la combinaison de deux cohortes de patients adultes européens d'origine française (groupe d'étude ANRS Genoscan) et d'origine Suisse (the Swiss Hepatitis C Cohort Study) avec une infection chronique au VHC. Les patients ont tous un niveau de fibrose déterminé par PBH convertit en score METAVIR.

Une précédente GWAS de cette cohorte Franco-Suisse relative à la progression de la fibrose dans la mono-infection VHC a servi pour tester les résultats d'association génétiques présentant des p -valeurs $< 10^{-6}$ observées avec les phénotypes sqrtLS et la comparaison des scores METAVIR F0F1F2 vs F3F4 définis dans la coinfection. Le phénotype utilisé lors de cette

GWAS a été défini comme étant le taux de progression de fibrose (FPR, *fibrosis progression rate* en anglais) et correspond au ratio du score METAVIR et du temps estimé d'infection en années (unités METAVIR/an). Ce phénotype ne suivant pas une distribution normale, il a subi une transformation logarithmique. Afin de pouvoir comparer le log(FPR) avec le sqrtLS, une corrélation par la méthode Spearman a été effectuée à partir des individus possédant à une même date un score METAVIR, pour générer un FPR, et la valeur de Fibroscan® ayant servi pour le phénotype sqrtLS.

Troisième partie

Résultats

Le travail réalisé durant ma thèse a essentiellement porté sur l'analyse du phénotype de fibrose hépatique chez les patients coinfecteds par le VIH et le VHC. Il a abouti à deux articles en cours de publication qui sont présentés ci-après.

1 Etude « génome entier » de la cohorte ANRS CO13 HEPAVIH de patients coinfectés par le VHC et le VIH : à la recherche d'associations liées à la quantification de la fibrose

A new 3p25 locus is associated with liver fibrosis progression in HIV/HCV co-infected patients

Damien Ulveling*, **Sigrid Le Clerc***, **Aurélie Cobat***, Taoufik Labib, Josselin Noirel, Vincent Laville, Cédric Coulonges, Wassila Carpentier, Bertrand Nalpas, Markus H. Heim, Thierry Poynard, Andreas Cerny, Stanislas Pol, Pierre-Yves Bochud, François Dabis, Ioannis Theodorou, Yves Lévy, Dominique Salmon, Laurent Abel*, Stéphanie Dominguez* and Jean-François Zagury* and the HEPAVIH ANRS CO13 cohort, the Swiss Hepatitis C Cohort Study Group, and the French ANRS HC EP 26 Genoscan Study Group.

Hepatology (en seconde révision)

RESUMÉ :

Afin d'identifier les facteurs génétiques impliqués dans la sévérité de fibrose hépatique lors de la coinfection VIH/VHC, nous avons réalisé une analyse « génome entier » sur 289 patients issus de la cohorte ANRS CO13 HEPAVIH (351), en prenant en covariables les facteurs non-génétiques connus pour influencer la fibrose hépatique tels la surconsommation d'alcool, le sexe ou encore le temps d'infection. L'élasticité hépatique (en kPa) définissant le niveau de fibrose a été évalué par Fibroscan® pour chacun des patients et la première valeur disponible après inclusion dans la cohorte a permis l'établissement d'un phénotype quantitatif après normalisation (sqrtLS, racine carré de l'élasticité du foie en kPa).

Nous avons analysé un jeu de données de 8 426 597 de SNPs génotypés ou ayant un fort score de qualité d'imputation, après un contrôle qualité et des procédures d'imputation utilisant le panel de référence Européen de 1000 Génomes Phase I (N = 379) à partir de données de génotypage d'une puce Illumina Omni2.5 BeadChip (Illumina, San Diego, USA).

Nos résultats ont permis de mettre en évidence deux locus présentant des signaux dont la *p*-valeur d'association en mode dominant était inférieure au seuil de significativité génome-entier (5×10^{-8}). Le premier locus, constitué de cinq SNPs, était localisé sur chromosome 3p25 dans le gène *OXTR*. Le signal était représenté par le SNP génotypé rs61183828, un variant intronique du gène *OXTR* (beta = 1,02, standard error (SE) = 0,17, *p*-valeur = $3,8 \times 10^{-9}$). Ce SNP rs61183828 expliquait 11% de la variance de notre phénotype. Ce locus a été répliqué dans deux cohortes de patients mono-infectés par le VHC renforçant son intérêt biologique. Le second signal, situé sur le chromosome 9p22 dans le pseudogène *MAP1LC3BP1*, correspondait à deux SNPs représenté par le SNP génotypé rs11790131 (beta = 0,60, SE = 0,11, *p*-valeur = $9,3 \times 10^{-9}$). Ce signal n'a pas été répliqué dans les cohortes de mono-infection par le VHC. Finalement, *OXTR* ne présentant pas de lien avec la fibrose hépatique, nous avons étendu l'analyse aux gènes voisins *CAV3* et *RAD18* où le signal était présent grâce au déséquilibre de liaison. Ces deux gènes présentent un intérêt biologique lié à la fibrose hépatique car *CAV3* est important dans les mécanismes de maintien cellulaire et *RAD18* a été relié à la réplication du VHC.

Cette étude a donc permis d'identifier deux locus significativement associés à la sévérité de la fibrose. L'un des résultats majeurs de cette étude est la réplication du locus 3p25 dans la mono-infection par le VHC, présentant deux gènes pertinents sur le plan biologique de par leur fonction dans la cellule et leur potentiel impact sur les mécanismes de fibrogenèse hépatique. Ces nouvelles perspectives dans la compréhension des mécanismes de fibrose dans le contexte de la coinfection VIH/VHC pourraient aider à l'identification de nouvelles cibles pour la création de médicaments ou de tests diagnostiques afin d'améliorer les soins des patients.

A new 3p25 locus is associated with liver fibrosis progression in HIV/HCV co-infected patients

Damien Ulveling^{1*}, Sigrid Le Clerc^{1*}, Aurélie Cobat^{2,3*}, Taoufik Labib¹, Josselin Noirel¹, Vincent Laville¹, Cédric Coulonges¹, Wassila Carpentier⁴, Bertrand Nalpas⁵, Markus H. Heim⁶, Thierry Poynard⁷, Andreas Cerny⁸, Stanislas Pol^{5,9}, Pierre-Yves Bochud¹⁰, François Dabis¹¹, Ioannis Theodorou^{12,13}, Yves Lévy¹⁴, Dominique Salmon¹⁵, Laurent Abel^{2,3,16*}, Stéphanie Dominguez^{14*} and Jean-François Zagury^{1*} and the HEPAVIH ANRS CO13 cohort^{**}, the Swiss Hepatitis C Cohort Study Group^{**}, and the French ANRS HC EP 26 Genoscan Study Group^{**}.

¹Équipe Génomique, Bioinformatique et Applications (EA4627), Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France.

² Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale (INSERM) U1163, Paris, France.

³ Paris Descartes University, Imagine Institute, Paris, France.

⁴ Plateforme Post-Génomique P3S, AP-HP, UPMC Université Paris 6, Faculté de Médecine Pitié Salpêtrière, Paris, France

⁵ Département d'Hépatologie, Hôpital Cochin (AP-HP), Université Paris Descartes, Paris, France.

⁶ Department of Gastroenterology, University Hospital, Basel, Switzerland

⁷ Université Pierre et Marie Curie, Service d'Hépatogastroentérologie, Hôpital Pitié-Salpêtrière (AP-HP), Paris, France

⁸ Epatocentro Ticino, Lugano, Switzerland

⁹ INSERM UMS20, Institut Pasteur, Paris, France.

¹⁰ Infectious Diseases Service, Department of Medicine, University Hospital and University of Lausanne, Switzerland

¹¹ Centre de Recherche INSERM U897, Epidemiologie-Biostatistique, Institut de Santé Publique, Epidémiologie et Développement, Université de Bordeaux, Bordeaux, France

¹² Laboratory of Immunity and Infection, Centre d'Immunologie et des Maladies Infectieuses de Paris (CIMI), INSERM U1135, Hôpital Pitié-Salpêtrière (AP-HP), Paris, France

¹³ Plateforme Génomique INSERM-ANRS, Groupe Hospitalier Pitié Salpêtrière, AP-HP, UPMC Université Paris 6, Paris, France

¹⁴ INSERM U955, AP-HP, Groupe Henri-Mondor Albert-Chenevier, Immunologie Clinique, Créteil, France

¹⁵ Department of Infectious Diseases, Cochin Hospital, Paris, France

¹⁶ St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA.

* These authors contributed equally to this study

** See Supporting Information for the full names of all members of the HEPAVIH ANRS CO13 cohort study group, the Swiss Hepatitis C Cohort Study Group and the French ANRS HC EP 26 Genoscan Study Group

Keywords: HIV/HCV, Liver disease, GWAS, CAV3, RAD18

Corresponding authors:

Stéphanie Dominguez, Département d'Immunologie Clinique, INSERM U955, AP-HP, Groupe Henri-Mondor Albert-Chenevier, 51, avenue du Mal de Lattre de Tassigny 94010 Créteil,

Phone: 33 01 49 81 44 42; Fx: 33 01 49 81 24 69; e-mail =

stephanie.dominguez@hmn.aphp.fr

Jean-François Zagury, Équipe Génomique, Bioinformatique et Applications (EA4627),

Chaire de Bioinformatique, CNAM, 75003, Paris,

Phone: 33 1 58 80 88 20; e-mail = zagury@cnam.fr

List of Abbreviations: HCV: hepatitis C virus; HIV: human immunodeficiency virus; HAART: highly active antiretroviral therapy; GWAS: genome-wide association study; SNP: single-nucleotide polymorphism; HWE: Hardy-Weinberg equilibrium; MAF: minor allele frequency; sqrtLS: square-root of liver stiffness; FPR: fibrosis progression rate; logFPR: logarithm of FPR; METAVIR: meta-analysis of histological data in viral hepatitis; SVR: sustained virological response; IVDU: intravenous drug user; LD: linkage disequilibrium; UTR: untranslated region; lncRNA: long non-coding RNA

Financial Support:

This study was sponsored and funded by the French National Agency for Research on AIDS and Viral Hepatitis (ANRS), with the participation of Glaxo-SmithKline, Roche and Bristol Myers Squibb. The study was also partly funded by a grant from Merck Sharp and Dohme – France. The ANRS Study HC EP 26 Genoscan group is supported by the National Agency for Research on AIDS and Viral Hepatitis (ANRS). The Swiss Hepatitis C Cohort Study (SCCS) is supported by grants from the Swiss National Science Foundation (33CS30_148417). The SCCS genetics project is supported by the Leenaards foundation, the Santos-Suarez foundation and the Swiss National Science Foundation (324730-144054).

Abstract (N=263 words)

Background: There is growing evidence that human genetic variants contribute to liver fibrosis in subjects with hepatitis C virus (HCV) mono-infection, but this aspect has been little investigated in patients co-infected with HCV and human immunodeficiency virus (HIV). We performed the first genome-wide association study (GWAS) of liver fibrosis progression in patients co-infected with HCV and HIV, using the well-characterized French ANRS CO13 HEPAVIH cohort. Liver fibrosis was assessed by elastography (FibroScan®) providing a quantitative fibrosis score. After quality control, GWAS was conducted on 289 Caucasian patients, for a total of 8,426,597 genotyped (Illumina Omni2.5 BeadChip) or reliably imputed SNPs. SNPs with p -values $< 10^{-6}$ were investigated in two independent replication cohorts of European patients infected with HCV alone.

Results: Two signals of genome-wide significance (p -value $< 5 \times 10^{-8}$) were obtained. The first, on chromosome 3p25 and corresponding to rs61183828 (p -value = 3.8×10^{-9}), was replicated in the two independent cohorts of patients with HCV mono-infection. The cluster of SNPs in linkage disequilibrium with rs61183828 was located close to two genes involved in mechanisms affecting both cell signaling and cell structure (*CAV3*) or HCV replication (*RAD18*). The second signal, obtained with rs11790131 (p -value = 9.3×10^{-9}) on chromosome region 9p22, was not replicated.

Conclusion: Our GWAS identified a new locus associated with liver fibrosis severity in patients with HIV/HCV co-infection, on chromosome 3p25. This finding was replicated in patients with HCV mono-infection. These results provide new relevant hypotheses for the pathogenesis of liver fibrosis in patients with HIV/HCV co-infection that may help define new targets for drug development or new prognostic tests, to improve patient care.

Introduction

Hepatitis C virus (HCV) and human immunodeficiency virus (HIV) infections affect 185 million (1) and 37 million (2) individuals, respectively, worldwide. These two viruses are transmitted in similar ways, and 15-30% of individuals with HIV infection are, therefore, also co-infected with HCV (3). With the introduction of highly active antiretroviral therapy (HAART), which has improved the survival of AIDS patients, HCV has emerged as a major comorbid disease causing progressive liver disease, potentially leading to end-stage liver disease, hepatocellular carcinoma, or death (4, 5). Liver disease has been observed to occur more rapidly in patients with HIV/HCV co-infection than in patients infected with HCV alone, but the risk factors for progressive liver disease in HIV/HCV-co-infected patients are not yet fully understood (6-8). Several studies have suggested that the effective treatment of HIV and HCV may be associated with a lower risk of liver disease progression (9-11). Risk factors such as alcohol consumption, age, and obesity have also been described, but account for only a small proportion of the variability in liver fibrosis development (12). There is growing evidence to suggest that host genetic factors are involved, at least in patients infected with HCV alone, but these factors remain largely unknown (13, 14).

A number of candidate gene approaches have been used to identify variants influencing the development of liver disease in patients with HCV infection, generally as a single infection, but most of these studies produced results that were not consistently replicated (15, 16). Genome-wide association studies (GWAS) provide a broader, unbiased approach for the discovery of genetic factors involved in disease susceptibility (17). In HIV-related phenotypes, such studies have mostly identified HLA loci associated with the control of viral load and with non-progression to AIDS (18). Several GWAS of HCV infection-related

phenotypes have identified polymorphisms of the *IL28B* (also denoted as *IFNL3/4*) gene associated with both the spontaneous clearance of HCV and a favorable response to interferon/ribavirin-based treatment, both in patients infected with HCV only (19) and in patients with HIV/HCV co-infection (20). By a candidate gene approach, polymorphisms of *IL28B* gene have also been associated with liver inflammation and hepatic fibrosis in HCV mono-infection (21-23). Only two GWAS have focused on the HCV-related liver fibrosis outcome in patients with mono-infections. The first, in a Caucasian population, identified four susceptibility loci, three of which were linked to genes involved in apoptosis (24). The second, conducted in Japan, detected variants within the HLA region (25). Moreover, GWAS identified *DEPDC5* and *MICA* as susceptibility genes for HCV-induced hepatocarcinoma in Japanese populations (26-28). No GWAS of fibrosis progression in patients with HIV/HCV co-infection has yet been published. We therefore carried out this GWAS to identify new genetic markers associated with fibrosis progression in the prospective French HEPAVIH ANRS CO13 cohort of adults with HIV/HCV co-infection.

Materials and methods

Study design and population

This study was based on a population sample of 374 patients with HIV/HCV co-infection who had provided written informed genetic consent and were enrolled in the French National Agency for Research on AIDS and Viral Hepatitis (ANRS) CO13 HEPAVIH cohort (29). The enrollment criteria were as follows: patients aged 18 years or over, with chronic HIV/HCV co-infection confirmed by a positive test for HIV-1 antibody and by an HCV RNA assay (regardless of clinical stage, sex or transmission group). The dual infection in this cohort was well-characterized because patients had full follow-up medical visits every six months in cases of cirrhosis and annually for patients without cirrhosis (median follow-up for the cohort: ~4 years). A self-administered questionnaire was used to record sociodemographic characteristics and data concerning past and current smoking habits, drug use and alcohol consumption.

Assessment of liver fibrosis and HCV fibrosis phenotype for the GWAS

Liver fibrosis was assessed by transient elastography (FibroScan[®]; Echosens, Paris, France) (30). At least 10 valid elastometric measurements were required for each patient. Liver elasticity was recorded as the median of all single measurements. Only liver stiffness measurements obtained with at least 10 successful acquisitions and a success rate of at least > 60% and interquartile ranges of < 30% were considered reliable. For the GWAS, we used the first liver stiffness measurement, in kPa, obtained by the FibroScan[®] method during patient follow-up as the phenotype of interest (Supplemental Figure 1A). The variance and skewness of the phenotype distribution were reduced by normalizing the liver stiffness data by square root transformation, to yield sqrtLS (Supplemental Figure 1B). We also converted liver stiffness in kPa into a METAVIR score, from F0 to F4 (31). We defined two groups of patients: F0F1F2

(no or mild fibrosis, liver stiffness < 9.5 kPa) and F3F4 (advanced fibrosis, liver stiffness \geq 9.5 kPa) (32).

Genotyping and quality control procedures

The Illumina Omni2.5 BeadChip (Illumina, San Diego, USA), which contains 2,391,739 markers, was used to genotype 374 patients. Raw data were first analyzed with Genome Studio software (version 1.6.3; Illumina) to obtain genotype calls. Thirty samples with a call rate (percentage of SNPs genotyped per sample) < 95% were removed. Two pairs of individuals displayed high levels of relatedness on analysis with PLINK software (33). One patient from each of these pairs was retained for the analysis. At this stage, 342 samples were retained. We subsequently excluded 210,140 SNPs with a low call rate (< 98%), 312 SNPs not in Hardy-Weinberg equilibrium (HWE, p -value < 10^{-6}) and 1,025,509 SNPs with a low minor allele frequency (MAF < 5%). In total, 1,155,778 SNPs were retained for imputation procedures and analysis.

Population stratification analysis

The genotypes were analyzed by principal component analysis (PCA), with the SMARTPCA utility of the EIGENSOFT package version 4.2 (34), to correct for possible population stratification. Before the analysis, we removed 23 individuals who described themselves as “non-European”. We then carried out two rounds of SMARTPCA, which identified 27 outliers. These outliers were excluded from subsequent analyses (Supplemental Figure 2A), which were carried out on a final sample of 292 patients used for the GWAS. A third analysis without outliers was performed to determine the eigenvectors (Supplemental Figure 2B). In the

statistical analysis, we used the two first eigenvectors as covariates to correct for population substructure in the association analyses.

Imputation procedures

The genotype data were phased with SHAPEIT2, and the phased data were then imputed with IMPUTE2 (35, 36). Haplotype data for 1092 European individuals from the phase I integrated variant set of the 1000 Genomes project released in March 2012 and updated in August 2012 (37) were used as a reference panel. After all procedures had been carried out, we retained 8,426,597 SNPs with a high imputation quality score, as measured by the information metric (> 0.8), and with a MAF $> 5\%$.

Covariates used for the association analysis

We focused more specifically on the genetic factors affecting liver fibrosis, by adjusting the regression analyses for possible confounding factors: sex, age (in years), alcohol consumption (never, previous and current), HCV genotype (1 versus other), duration of HIV/HCV infection (in years) and effective time on antiretroviral treatment for HIV infection (in years). The full set of covariates was available for 289 patients. Overall, 200 patients (69%) received a pegylated interferon +/- ribavirin treatment for HCV infection, and 26 of these patients (9% of the total sample) achieved a sustained virological response (SVR) before the first FibroScan® evaluation. Because of their low number and in order to assess the fibrosis evolution of SVR patients, we kept those 26 patients in the initial analyses. However, we also assessed the effect of SVR on our_GWAS results by analyzing the most significant signals after removal of these 26 individuals in the regression model. The values used for all covariates were those on the date

of FibroScan[®] evaluation. The top two eigenvectors were also used to correct for population substructure.

Statistical tests

Genomic associations with sqrtLS were investigated by linear regression analysis in SNPTEST 2.4.1 software (38), with a genetic additive model and genotype dosage for the imputed SNPs, and a complete set of the relevant covariates. The classical genome-wide significance threshold (p -values $\leq 5 \times 10^{-8}$) was used (38, 39).

Replication in patients with HCV mono-infection

Replication studies were performed in two cohorts of patients infected with HCV alone. The 22 SNPs with p -values $< 10^{-6}$ in the primary scan were first tested for replication in the primary cohort studied in a previous GWAS of liver fibrosis progression (24). This first replication sample combined data from two cohorts of adult patients of European descent from France (the ANRS Genoscan study group) and Switzerland (the Swiss Hepatitis C Cohort Study) with chronic HCV infection. In these patients, liver fibrosis stage was determined by examining a liver biopsy specimen obtained before treatment METAVIR scores from F0 to F4 were obtained. Liver fibrosis progression was also considered as a quantitative phenotype, with fibrosis progression rate (FPR), corresponding to METAVIR score divided by the estimated duration of infection in years (METAVIR units per year) (40, 41). Given the distribution of FPR, this phenotype was log-transformed to yield logFPR values. Genotyping was performed with a combination of Illumina arrays (HumanCNV370-Duo and Human1M-Duo), and imputation was performed with the IMPUTE2 software package and the 1000 Genomes Project haplotypes released on 16 June 2014 in the phase 1 integrated variant set release. FPR was

available for 1064 individuals and was used to test for replication of the 22 SNPs with a p -value $< 10^{-6}$.

The three clusters replicated in the Franco-Swiss GWAS for FPR phenotype were further tested in a second independent sample of French patients with HCV mono-infection recruited from several hospitals in Paris (follow-up of the ANRS Genoscan study). Liver fibrosis was assessed by FibroScan[®] and data were available for 309 patients. Thirty-nine patients had a SVR before liver fibrosis evaluation and were excluded from the analysis. In the remaining 270 patients, we used the same phenotypic definition as in the HEPAVIH cohort: the square root transformation of the liver stiffness obtained by the FibroScan[®] method. Genotyping was performed with the *TaqMan*[®] SNP *Genotyping* Assay. The association between logFPR (first cohort) or sqrtLS (second cohort) and the number of alleles (additive model) was assessed by linear regression analysis in R software (<https://cran.r-project.org/>). All the analyses were adjusted for sex, age at infection, alcohol consumption, HCV genotype and mode of contamination.

Correlation between FibroScan[®] values and METAVIR scores converted into FPR

In addition to liver stiffness measurements by FibroScan[®], 70 patients from the HEPAVIH cohort also underwent a quantification of liver fibrosis based on the determination of METAVIR score on liver biopsy. We were therefore able to evaluate the correlation (Spearman's rank correlation coefficient) between sqrtLS and logFPR phenotypes for the 70 patients for which both outcomes and the complete set of confounding covariates were available.

Bioinformatics exploration databases

We explored the polymorphisms detected in our GWAS further, by looking for changes in mRNA levels (GTEx Portal, www.gtexportal.org/home/), splicing (NetGene2, www.cbs.dtu.dk/services/NetGene2/), transcription factor binding sites (RegulomeDB, www.regulomedb.org/, HaploReg v2, www.broadinstitute.org/mammals/haploreg/haploreg.php), and miRNA genes or miRNA targets (miRBAs, www.mirbase.org/, miRTarBase, www.mirtarbase.mbc.nctu.edu.tw/, MicroCosm Targets, www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/).

Results

Sample description

The demographic characteristics of the 289 patients retained for the genetic association study (see Methods) are presented in Supplemental Table 1. The data shown were recorded at the time of the first FibroScan[®] measurement for the patient, generally in the year following inclusion in the cohort. About 75% of the patients were infected via intravenous drug injection and the median age at infection was about 20 years. The median duration of infection was 24.4 years. Men accounted for 70% of the patients in the cohort and most patients were infected with HCV genotype 1 (57.5%). The median liver stiffness value obtained on FibroScan[®] was 6.7 kPa (standard deviation, SD = 9.8, range: 3.2-75) (Supplemental Figure 1A). Median CD4⁺ cell count was 468 cells/mm³ and median HIV viral load was 3.7 log viral copies/ml, with 24% of patients having detectable amounts of virus. The median duration of antiretroviral treatment for HIV infection was about 10 years. We also split the patients in two groups according to liver stiffness converted into a METAVIR score (F0F1F2 versus F3F4). Sex ratio, mean age, and duration of infection did not differ significantly between the two groups. As expected, there was a significant difference in median liver stiffness (in kPa) between the two groups (F0F1F2: 5.9 ± 1.5 versus F3F4: 15.3 ± 13.9). Median CD4⁺ cell counts were significantly lower for the F3F4 group than for the F0F1F2 group, whereas there was no significant difference in HIV viral load between these two groups (Supplemental Table 1).

Associations with sqrtLS

After the quality control and imputation procedures, we checked for genetic associations with sqrtLS in an additive model, using data for 289 patients with HIV/HCV co-infection and 8,426,597 genotyped or reliably imputed SNPs. The genetic association results are shown in a

Manhattan plot (Figure 1) and the distribution of p -values is shown in a QQ plot (Supplemental Figure 3). The QQ plot gave a genomic inflation factor close to one ($\lambda = 0.996$), indicating an absence of stratification after adjustment for the first two principal components. Seven SNP clusters provided p -values $< 10^{-6}$ (Table 1), including two above the significant genome-wide threshold of 5×10^{-8} . The first significant signal, for chromosome 3p25, corresponded to a cluster of five SNPs ($r^2 > 0.8$ in 1000 Genomes European population) tagged by the genotyped SNP rs61183828, an intronic variant of the *OXTR* gene (beta = 1.02, standard error (SE) = 0.17, p -value = 3.8×10^{-9}). The second signal was located on chromosome region 9p22 and corresponded to two SNPs tagged by the genotyped SNP rs11790131 within the *MAP1LC3BP1* pseudogene (beta = 0.60, SE = 0.11, p -value = 9.3×10^{-9}). The thorough analysis of the more recent 1000 Genomes phase 3 European data did not allow to detect any additional variants in strong LD with the SNPs of these two clusters. Multivariate analysis confirmed that the association results within each of the two clusters were supported by a single signal. In addition, haplotype analysis within each of the two clusters did not identify any combination of alleles for which there was substantially stronger evidence for association than for the best single SNP. Information about the other clusters with p -values $< 10^{-6}$ is provided in Table 1.

We then investigated the influence on the results of the 26 patients with a SVR before the evaluation of liver fibrosis, by removing these patients from the analysis. The signal observed for rs61183828 remained significant (p -value = 1.33×10^{-8}), and there was a slight increase in the genetic effect, as measured by the regression coefficient beta (1.065 rather than 1.02). By contrast, the association with rs11790131 no longer displayed GW significance (p -value = 5.5×10^{-6}), and a decrease was observed in the beta coefficient (0.57 rather than 0.60).

Replication in independent cohorts of patients with HCV mono-infection

The 22 SNPs with a p -value $< 10^{-6}$, corresponding to seven independent clusters (Table 1), were first tested for replication in the primary cohort of a previous Franco-Swiss GWAS of liver fibrosis in patients infected with HCV alone. Only one of the 22 SNPs had been genotyped in the Franco-Swiss GWAS. The other 21 SNPs were imputed in at least one cohort, but with a high imputation quality score (imputation information > 0.7 , Table 2). In this first cohort, we used the log-transformed FPR phenotype, logFPR, as a proxy for sqrtLS phenotype. In the HEPAVIH cohort, we found a significant correlation between logFPR and sqrtLS in one subgroup of patients ($r = 0.49$, p -value = 6.5×10^{-5}). Evidence for replication was found for three of the seven clusters, on chromosomes 3p25, 11p13 and 13q12, in the Franco-Swiss GWAS, with the same risk allele as in the HEPAVIH cohort, and with p values less than 0.05 (Table 2). We also investigated whether our seven clusters could be related with variants that were found in reported GWAS dealing with liver fibrosis from various etiologies. We identified 322 SNPs with p -value $< 10^{-5}$ in 31 previously published GWAS (Supplemental Table 3). These SNPs corresponded to 131 genetic regions, and three of these regions, 6q23, 11p13 and 13q12, were common with the 7 regions identified in our study (Table 1). None of the five SNPs of the previous GWAS belonging to these common regions exhibited any LD with the significant SNPs found in our study (distance more than 1Mb), and their p -values obtained in our GWAS were not significant ($p > 5 \times 10^{-2}$).

The three replicated clusters and the initial GW significant signal on chromosome region 9p22 were further tested for replication in an additional French cohort of patients with HCV mono-infection for whom FibroScan[®] data were available. Only the cluster on chromosome region 3p25, with the rs61183828 SNP (beta = 0.625, SE=0.229, one-tailed p -value = 0.004) was replicated in this second cohort. The initial replication of the 3p25 region with the logFPR

phenotype in the Franco-Swiss cohort was observed with another SNP of the cluster, rs73132859 (one-tailed p -value = 0.02, imputation score = 0.92 in the Franco-Swiss cohort) whereas only a trend (p -value = 0.12) was observed for rs61183828. The SNPs rs73132859 and rs61183828 are in strong LD ($r^2 = 1$ in the 1000 Genomes European reference population) but the imputation score was slightly higher for rs73132859 than for rs61183828 (Table 2), potentially accounting for the difference in the results of association analyses. Overall, the replication study in subjects with HCV mono-infection replicated the most significant initial signal for the cluster on chromosome 3p25, in particular with the SNP rs61183828 which was genotyped in both the primary HEPAVIH cohort, and the second Genoscan replication cohort.

The cluster on chromosome 3p25 was significant genome-wide in the HEPAVIH sample, for both the full sample and for the subsample of patients without SVR. This cluster was replicated in two independent cohorts of patients with HCV mono-infection, via rs61183828 in particular. This A/G SNP had a minor allele frequency (MAF) of 0.06 (allele G) in the 289 HEPAVIH patients, with 32 heterozygous AG patients, and two homozygous GG patients. Demographic information according to the SNP rs61183828 genotypes (AA versus AG+GG) is shown in the supplemental Table 2. No significant differences were observed in the distribution of the different covariates between the two genotype groups. A similar MAF was observed in the French/Swiss GWAS ($=0.05$), and the second Genoscan cohort ($=0.06$), and a MAF of 0.04 was obtained for the 1000 Genomes European population. The rare G allele was found to be associated with greater liver stiffness, with a difference of almost 10 kPa between the means of AA and the AG patients (Figure 2A). There were only two homozygous GG patients, but both had high degrees of liver stiffness, with one having an extreme value of 47.2 kPa. Overall, SNP rs61183828 explained 11% of the sqrtLS variance. The distributions of liver stiffness and of the corresponding binary F0F1F2/F3F4 fibrosis phenotypes according to rs61183828 genotype

are shown in Figure 2B. Both the GG patients had a binary fibrosis phenotype (Figure 2B) of F3F4, and more than 50% of the AG heterozygotes were F3F4, versus only 22.8% of the AA patients. According to this binary definition, the odds ratio of developing liver fibrosis (ie having a F3F4 status) for subjects of AG or GG genotypes vs subjects of AA genotype was 4.3 [95% confidence interval: 1.9; 9.7].

Discussion

The major pathological consequences of HCV infection are progression to liver fibrosis and its potential complications. Several factors have clearly been shown to be associated with fibrosis progression. These factors include the duration of infection, age, being male, obesity and alcohol consumption (42). HIV/HCV co-infection also results in liver fibrosis progressing more rapidly, with a viral load for HCV twice that in patients with HCV mono-infection (42). Co-infected patients appear to be at high risk of developing insulin resistance or diabetes (43) and are more likely to develop hepatotoxicity due to HAART than patients infected with HIV alone. The HEPAVIH cohort is slightly particular since a rather high proportion of patients (69%, more than usual) received anti-HCV treatment in addition to anti-HIV treatment, although only 9% of them achieved SVR before their fibrosis evaluation. Indeed, the French HIV guidelines recommend to treat HCV infection in co-infected patients, irrespective to the CD4⁺ count, using the less hepatotoxic drugs. Fibrosis progression was also found to be more likely in patients with low CD4⁺ cell counts and high viral loads for HIV (42), as in the patients of the HEPAVIH cohort, despite the lack of a significant difference in HIV viral load between the F0F1F2 and F3F4 groups (Supplemental Table 1). The identification of genetic variants associated with a risk of developing liver fibrosis in this specific population of co-infected patients is therefore of major interest.

Our GWAS detected, for the first time, two significant signals associated with severe liver fibrosis in patients with HIV/HCV co-infection. One of these signals was weaker if patients with SVR were removed from the analysis and was not replicated in two cohorts of patients with HCV mono-infection. Additional samples of co-infected patients should therefore be investigated to determine the role of this signal in this specific population. By contrast, the signal on chromosome 3p25 clearly remained significant in the sample of co-infected patients

without SVR, and was replicated in two independent cohorts of patients with HCV mono-infection, through the highly correlated SNPs rs61183828 and rs73132859. The association replicated in the mono-infected cohorts was of a smaller magnitude than that in the HIV/HCV cohort, possibly due to the “winner’s curse” effect, sampling variation, and/or an impact of HIV co-infection on the strength of the association. Furthermore, the weaker effect observed in the first replication sample (Franco-Swiss GWAS) may reflect differences in the phenotypes studied, despite the correlation between logFPR and sqrtLS (0.49). More significant replication was observed in the second replication cohort, based on exactly the same phenotype as the GWAS.

Extensive LD analysis of the 1000 Genomes European population showed that the replicated cluster on chromosome 3p25 included five SNPs ($r^2 > 0.8$ with rs61183828) located in a region containing three genes, *OXTR*, *CAV3* and *RADI8* (Supplemental Figure 4). Four of these SNPs were found to be located in *OXTR* introns, including rs61183828, whereas rs73132859 was intergenic. Four of the SNPs of the cluster were found to be located in a potential promoter region enriched in CpG islands, including rs78828026 ($r^2=1$ with rs61183828 according to 1000 Genomes European population), which is located in a validated transcription factor binding site for FOXA1 (forkhead box protein A1) (44). FOXA1 is a liver-specific transcriptional activator (45), consistent with this signal being of biological interest for the phenotype studied. We found no published evidence linking the five SNPs or FOXA1 with the expression of *OXTR*, *CAV3* or *RADI8*. One of these genes, *OXTR*, encodes a G-protein-coupled receptor that acts as a receptor for oxytocin and does not seem to be related to liver or fibrosis outcomes. However, the other two genes, *CAV3* and *RADI8*, appear more interesting as their products are involved in mechanisms likely to affect both cell signaling and the maintenance of cell structure for *CAV3* or HCV replication for *RADI8*.

The *CAV3* gene encodes a member of the caveolin family that functions as a component of the caveolae found in the plasma membranes of most cell types, in which it acts as a scaffold, organizing other proteins involved in cell signaling and the maintenance of cell structure (46). However, this gene also encodes a long non-coding RNA (lncRNA, 680 bases) with exons overlapping the intron sequences of *RAD18* (Supplemental Figure 4). *CAV3* may belong to the growing class of “bifunctional genes” encoding not only proteins but also lncRNA isoforms, which often have a function antagonistic to that of the protein (47). *CAV3* mRNA is also detected in liver tissues ([GeneAtlas U133A, gcrma](#)) and this gene has been reported to be associated with hepatic insulin signaling (48). The *RAD18* gene may also be relevant. Indeed, the potential promoter region mentioned above and including rs78828026, is located 5’ to *RAD18*. This gene encodes a component of the DNA damage repair system. Cells lacking Rad18 are hypersusceptible to infection with murine leukemia virus and HIV (49), and this protein has been implicated in the post-replication DNA repair process occurring during HIV-1 infection (50). Several viruses, including HCV, are known to disturb the cellular DNA repair and cell cycle pathways to maximize their own replication (51).

In conclusion, we have identified, in this GWAS, a significant, replicated genetic association on chromosome 3p25 with the SNP rs61183828 which exhibits a substantial impact on liver stiffness in co-infected patients. These results suggest new relevant hypotheses for the pathogenesis of liver fibrosis in co-infected patients, pointing out novel molecular mechanisms potentially accounting for this genetic effect. These findings may help define new targets for the development of drugs or prognostic tools in the treatment of HCV-infected patients.

References:

1. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, Barnes E. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* 2015;61:77-87.
2. UNAIDS. How AIDS changed everything; 2015.
3. Sherman KE, Rouster SD, Chung RT, Rajicic N. Hepatitis C Virus prevalence among patients infected with Human Immunodeficiency Virus: a cross-sectional analysis of the US adult AIDS Clinical Trials Group. *Clin Infect Dis* 2002;34:831-837.
4. Bica I, McGovern B, Dhar R, Stone D, McGowan K, Scheib R, Snyderman DR. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis* 2001;32:492-497.
5. Rosenthal E, Poiree M, Pradier C, Perronne C, Salmon-Ceron D, Geffray L, Myers RP, et al. Mortality due to hepatitis C-related liver disease in HIV-infected patients in France (Mortavic 2001 study). *AIDS* 2003;17:1803-1809.
6. Benhamou Y, Di Martino V, Bochet M, Colombet G, Thibault V, Liou A, Katlama C, et al. Factors affecting liver fibrosis in human immunodeficiency virus-and hepatitis C virus-coinfected patients: impact of protease inhibitor therapy. *Hepatology* 2001;34:283-287.
7. Konerman MA, Mehta SH, Sutcliffe CG, Vu T, Higgins Y, Torbenson MS, Moore RD, et al. Fibrosis progression in human immunodeficiency virus/hepatitis C virus coinfected adults: prospective analysis of 435 liver biopsy pairs. *Hepatology* 2014;59:767-775.
8. Sulkowski MS, Mehta SH, Torbenson MS, Higgins Y, Brinkley SC, de Oca RM, Moore RD, et al. Rapid fibrosis progression among HIV/hepatitis C virus-co-infected adults. *AIDS* 2007;21:2209-2216.
9. Bani-Sadr F, Lapidus N, Bedossa P, De Boever CM, Perronne C, Halfon P, Pol S, et al. Progression of fibrosis in HIV and hepatitis C virus-coinfected patients treated with interferon plus ribavirin-based therapy: analysis of risk factors. *Clin Infect Dis* 2008;46:768-774.
10. Brau N, Salvatore M, Rios-Bedoya CF, Fernandez-Carbia A, Paronetto F, Rodriguez-Orengo JF, Rodriguez-Torres M. Slower fibrosis progression in HIV/HCV-coinfected patients with successful HIV suppression using antiretroviral therapy. *J Hepatol* 2006;44:47-55.
11. Soriano V, Martin-Carbonero L, Garcia-Samaniego J. Treatment of chronic hepatitis C virus infection: we must target the virus or liver fibrosis? *AIDS* 2003;17:751-753.
12. Rueger S, Bochud PY, Dufour JF, Mullhaupt B, Semela D, Heim MH, Moradpour D, et al. Impact of common risk factors of fibrosis progression in chronic hepatitis C. *Gut* 2015;64:1605-1615.
13. Lutz P, Wasmuth JC, Nischalke HD, Vidovic N, Grunhage F, Lammert F, Oldenburg J, et al. Progression of liver fibrosis in HIV/HCV genotype 1 co-infected patients is related to the T allele of the rs12979860 polymorphism of the IL28B gene. *Eur J Med Res* 2011;16:335-341.
14. Rau M, Baur K, Geier A. Host genetic variants in the pathogenesis of hepatitis C. *Viruses* 2012;4:3281-3302.
15. Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, Elledge SJ. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc Natl Acad Sci U S A* 2009;106:16410-16415.
16. Bataller R, Brenner DA. Liver fibrosis. *J Clin Invest* 2005;115:209-218.
17. Alcais A, Abel L, Casanova JL. Human genetics of infectious diseases: between proof of principle and paradigm. *J Clin Invest* 2009;119:2506-2514.
18. Limou S, Zagury JF. Immunogenetics: Genome-Wide Association of Non-Progressive HIV and Viral Load Control: HLA Genes and Beyond. *Front Immunol* 2013;4:118.
19. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461:399-401.

20. Pineda JA, Caruz A, Di Lello FA, Camacho A, Mesa P, Neukam K, Rivero-juarez A, et al. Low-density lipoprotein receptor genotyping enhances the predictive value of IL28B genotype in HIV/hepatitis C virus-coinfected patients. *AIDS* 2011;25:1415-1420.
21. Rizk NM, Derbala MF. Genetic polymorphisms of ICAM 1 and IL28 as predictors of liver fibrosis severity and viral clearance in hepatitis C genotype 4. *Clin Res Hepatol Gastroenterol* 2013;37:262-268.
22. Bochud PY, Bibert S, Kutalik Z, Patin E, Guergnon J, Nalpas B, Goossens N, et al. IL28B alleles associated with poor hepatitis C virus (HCV) clearance protect against inflammation and fibrosis in patients infected with non-1 HCV genotypes. *Hepatology* 2012;55:384-394.
23. Abe H, Ochi H, Maekawa T, Hayes CN, Tsuge M, Miki D, Mitsui F, et al. Common variation of IL28 affects gamma-GTP levels and inflammation of the liver in chronically infected hepatitis C virus patients. *J Hepatol* 2010;53:439-443.
24. Patin E, Kutalik Z, Guergnon J, Bibert S, Nalpas B, Jouanguy E, Munteanu M, et al. Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. *Gastroenterology* 2012;143:1244-1252 e1241-1212.
25. Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, Hosono N, et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. *J Hepatol* 2013;58:875-882.
26. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, Otsuka M, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011;43:455-458.
27. Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, Aikata H, Ikeda K, et al. Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. *Nat Genet* 2011;43:797-800.
28. Lange CM, Bibert S, Dufour JF, Cellerai C, Cerny A, Heim MH, Kaiser L, et al. Comparative genetic analyses point to HCP5 as susceptibility locus for HCV-associated hepatocellular carcinoma. *J Hepatol* 2013;59:504-509.
29. Loko MA, Salmon D, Carrieri P, Winnock M, Mora M, Merchadou L, Gillet S, et al. The French national prospective cohort of patients co-infected with HIV and HCV (ANRS CO13 HEPAVIH): early findings, 2006-2010. *BMC Infect Dis* 2010;10:303.
30. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343-350.
31. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289-293.
32. Loko MA, Bani-Sadr F, Winnock M, Lacombe K, Carrieri P, Neau D, Morlat P, et al. Impact of HAART exposure and associated lipodystrophy on advanced liver fibrosis in HIV/HCV-coinfected patients. *J Viral Hepat* 2011;18:e307-314.
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.
34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.
35. Delaneau O, Zagury JF. Haplotype inference. *Methods Mol Biol* 2012;888:177-196.
36. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5-6.

37. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073.
38. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906-913.
39. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012;8:e1002822.
40. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* 1997;349:825-832.
41. Sobesky R, Mathurin P, Charlotte F, Moussalli J, Olivi M, Vidaud M, Ratziu V, et al. Modeling the impact of interferon alfa treatment on liver fibrosis progression in chronic hepatitis C: a dynamic view. The Multivirc Group. *Gastroenterology* 1999;116:378-386.
42. Karageorgopoulos DE, Allen J, Bhagani S. Hepatitis C in human immunodeficiency virus co-infected individuals: Is this still a "special population"? *World J Hepatol* 2015;7:1936-1952.
43. Merchante N, Rivero A, de Los Santos-Gil I, Merino D, Marquez M, Lopez-Ruz MA, Rodriguez-Bano J, et al. Insulin resistance is associated with liver stiffness in HIV/HCV co-infected patients. *Gut* 2009;58:1654-1660.
44. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012;489:91-100.
45. Xu CR, Zaret KS. Chromatin "pre-pattern" and epigenetic modulation in the cell fate choice of liver over pancreas in the endoderm. *Nucleus* 2012;3:150-154.
46. Gazzero E, Sotgia F, Bruno C, Lisanti MP, Minetti C. Caveolinopathies: from the biology of caveolin-3 to human diseases. *Eur J Hum Genet* 2010;18:137-145.
47. Ulveling D, Francastel C, Hube F. When one is better than two: RNA with dual functions. *Biochimie* 2011;93:633-644.
48. Dugail I, Postic C. Little caves ameliorate hepatic insulin signaling. Focus on "caveolin gene transfer improves glucose metabolism in diabetic mice". *Am J Physiol Cell Physiol* 2010;298:C442-445.
49. Lloyd AG, Tateishi S, Bieniasz PD, Muesing MA, Yamaizumi M, Mulder LC. Effect of DNA repair protein Rad18 on viral infection. *PLoS Pathog* 2006;2:e40.
50. Mulder LC, Chakrabarti LA, Muesing MA. Interaction of HIV-1 integrase with DNA repair protein hRad18. *J Biol Chem* 2002;277:27489-27493.
51. Chaurushiya MS, Weitzman MD. Viral manipulation of DNA repair and cell cycle checkpoints. *DNA Repair (Amst)* 2009;8:1166-1176.

Table 1 – Associations from with the sqrtLS in additive mode such that p -value $< 10^{-6}$, p -values adjusted with covariates (see Methods)

SNP	CHR	BP(HG19)		MAF	Beta (SE)	Adjusted p -value	MAF 1KG	Genes
rs73132848	3p25	8806424	A	0.06	1.13 (0.18)	1.00×10^{-9}	0.04	<i>OXTR</i>
rs78828026	3p25	8807582	C	0.06	1.01 (0.18)	2.02×10^{-8}	0.04	<i>OXTR</i>
rs61183828 ^G	3p25	8807809	C	0.06	1.01 (0.17)	3.86×10^{-9}	0.04	<i>OXTR</i>
rs79896191	3p25	8808411	T	0.06	1.02 (0.17)	3.24×10^{-9}	0.04	<i>OXTR</i>
rs73132859	3p25	8811394	C	0.06	0.91 (0.18)	2.20×10^{-7}	0.03	intergenic
rs148718180	6q23	131802744	C	0.05	1.01 (0.19)	1.28×10^{-7}	0.04	intergenic
rs117102357	6q23	131823969	T	0.05	1.01 (0.19)	1.19×10^{-7}	0.04	intergenic
rs143633948	6q23	131825594	A	0.05	1.07 (0.2)	9.34×10^{-8}	0.04	intergenic
rs145959532	6q23	131828358	A	0.05	1.01 (0.19)	1.22×10^{-7}	0.04	intergenic
rs78203620	6q23	131830410	A	0.05	1.01 (0.19)	1.21×10^{-7}	0.04	intergenic
rs75544527	6q23	131832335	A	0.05	1.01 (0.19)	1.20×10^{-7}	0.04	intergenic
rs112612132	6q23	131838280	A	0.05	1.01 (0.19)	1.17×10^{-7}	0.04	intergenic
rs3904567	6q23	131840255	T	0.05	1.01 (0.19)	1.17×10^{-7}	0.04	intergenic
rs74819266	6q23	131840790	G	0.05	1.01 (0.19)	1.17×10^{-7}	0.04	intergenic
rs190131975	6q23	131850254	A	0.05	1.01 (0.19)	1.20×10^{-7}	0.04	intergenic
rs150383367 ^G	6q23	131851956	T	0.05	1.01 (0.19)	1.21×10^{-7}	0.04	intergenic
rs11790112	9p22	19469751	A	0.20	0.59 (0.1)	1.79×10^{-8}	0.22	<i>MAP1LC3BP1</i>
rs11790131 ^G	9p22	19469846	T	0.20	0.60 (0.1)	9.26×10^{-9}	0.22	<i>MAP1LC3BP1</i>
rs35897606	11p13	33457857	G	0.08	0.80 (0.15)	5.76×10^{-8}	0.07	intergenic
rs79709413	13q12	25314626	T	0.05	0.95 (0.19)	7.54×10^{-7}	0.06	intergenic
rs35467001 ^G	17q25	71380062	A	0.05	0.87 (0.17)	4.57×10^{-7}	0.06	<i>SDK2</i>
rs73084982	20p13	3411053	A	0.06	0.92 (0.18)	2.05×10^{-7}	0.07	intergenic

^G: indicates that the SNP is genotyped

Table 2 – Replication table in the Franco-Swiss GWAS and the Genoscan cohort.

CHR	SNP	Tested allele HEPAVIH / Franco-Swiss / Genoscan	<i>p</i> -values HEPAVIH	GWAS Franco- Swiss <i>Imputation</i> <i>information</i> ^s	GWAS Franco- Swiss <i>p</i> -value	Genoscan <i>p</i> -value (beta)	MAF HEPAVIH / Franco-Swiss / Genoscan
3p25	rs73132848	A	1.00x10 ⁻⁹	0.92	0.107	-	0.06 / 0.05 / NA
	rs78828026	C	2.02x10 ⁻⁸	0.92	0.081	-	0.06 / 0.05 / NA
	<i>rs61183828</i> ^G	C	<i>3.86x10⁻⁹</i>	0.91	<i>0.123</i>	<i>0.004 (0.625)</i>	<i>0.06 / 0.05 / 0.06</i>
	rs79896191	T	3.24x10 ⁻⁹	0.93	0.120	-	0.06 / 0.05 / NA
	rs73132859	C	2.20x10⁻⁷	0.92	0.020	-	0.06 / 0.05 / NA
6q23	rs148718180	C	1.28x10 ⁻⁷	0.95	0.373	-	0.05 / 0.04 / NA
	rs117102357	T	1.19x10 ⁻⁷	0.99	0.499	-	0.05 / 0.04 / NA
	rs143633948	A	9.34x10 ⁻⁸	0.96	> 0.5	-	0.05 / 0.04 / NA
	rs145959532	A	1.22x10 ⁻⁷	0.98	0.484	-	0.05 / 0.04 / NA
	rs78203620	A	1.21x10 ⁻⁷	0.99	0.496	-	0.05 / 0.04 / NA
	rs75544527	A	1.20x10 ⁻⁷	0.98	0.492	-	0.05 / 0.04 / NA
	rs112612132	A	1.17x10 ⁻⁷	0.98	0.481	-	0.05 / 0.04 / NA
	rs3904567 ^G	T	1.17x10 ⁻⁷	0.98	0.472	-	0.05 / 0.04 / NA
	rs74819266	G	1.17x10 ⁻⁷	0.98	0.478	-	0.05 / 0.04 / NA
	rs190131975	A	1.20x10 ⁻⁷	0.98	0.468	-	0.05 / 0.04 / NA
	rs150383367	T	1.21x10 ⁻⁷	0.98	0.474	-	0.05 / 0.04 / NA
9p22	rs11790112	A	1.79x10 ⁻⁸	0.99	0.468	-	0.20 / 0.20 / NA
	rs11790131 ^G	T	9.26x10 ⁻⁹	1	> 0.5	> 0.5	0.20 / 0.20 / 0.17
11p13	rs35897606	G	5.76x10⁻⁸	0.80	0.006	-*	0.08 / 0.07 / NA
	rs17252741	T	1.05x10⁻⁶	0.82	0.008	> 0.5	0.08 / 0.07 / 0.05
13q12	rs79709413	T	7.54x10⁻⁷	0.78	0.040	> 0.5	0.05 / 0.06 / 0.05
17q25	rs35467001 ^G	A	4.57x10 ⁻⁷	0.86	0.448	-	0.05 / 0.05 / NA
20p13	rs73084982	A	2.05x10 ⁻⁷	0.86	> 0.5	-	0.06 / 0.04 / NA

^s: Imputation information lies in the range (0,1). An information value of 0 indicates that the genotype probabilities are obtained from the allele frequency (i.e. complete uncertainty about the imputed genotype), and an information value of 1 indicates that there is no uncertainty about the imputed genotype (e.g. genotyped SNP as rs11790131).

^G: indicates that the SNP was genotyped in the primary Hepavir cohort

*: rs35897606 could not be genotyped in Genoscan for technical reason and was replaced by rs17252741 ($r^2 = 1$ in 1000 Genomes European populations)

> 0.5: indicates an opposite risk allele

Bold font indicates a replication in the GWAS Franco-Swiss (p -value < 0.05)

Bold and italic font indicates a replication in Genoscan (p -value < 0.05)

Figure legends

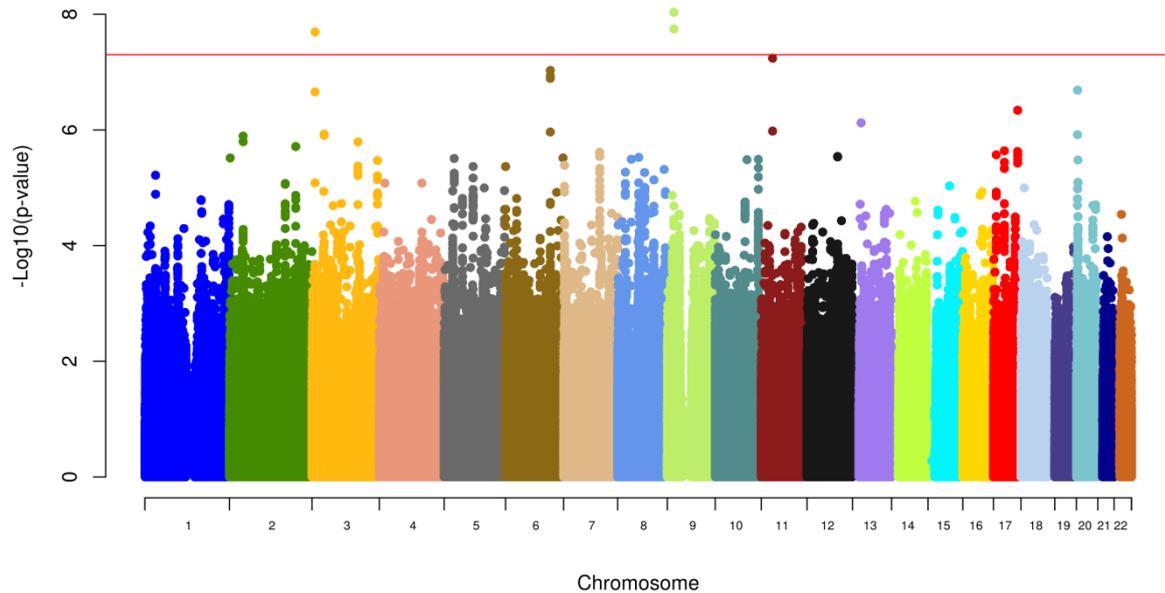


Figure 1 – Manhattan plot of the genome-wide association results with sqrtLS. After genotype imputation, ~8,500,000 common variants were tested for association with sqrtLS in 289 individuals, in linear regression analysis with an additive model. Each dot corresponds to a SNP, with its chromosomal position (x axis) and $-\log_{10}(p\text{-value})$ (y axis) as the coordinates. Association signals of genome-wide significance ($p\text{-value} < 5 \times 10^{-8}$, red line) were observed on chromosomes 3 and 9. The strongest associated SNPs were rs61183828 on chromosome 3p25 ($p\text{-value} = 3.8 \times 10^{-9}$) and rs11790131 on chromosome 9p22 ($p\text{-value} = 9.3 \times 10^{-9}$).

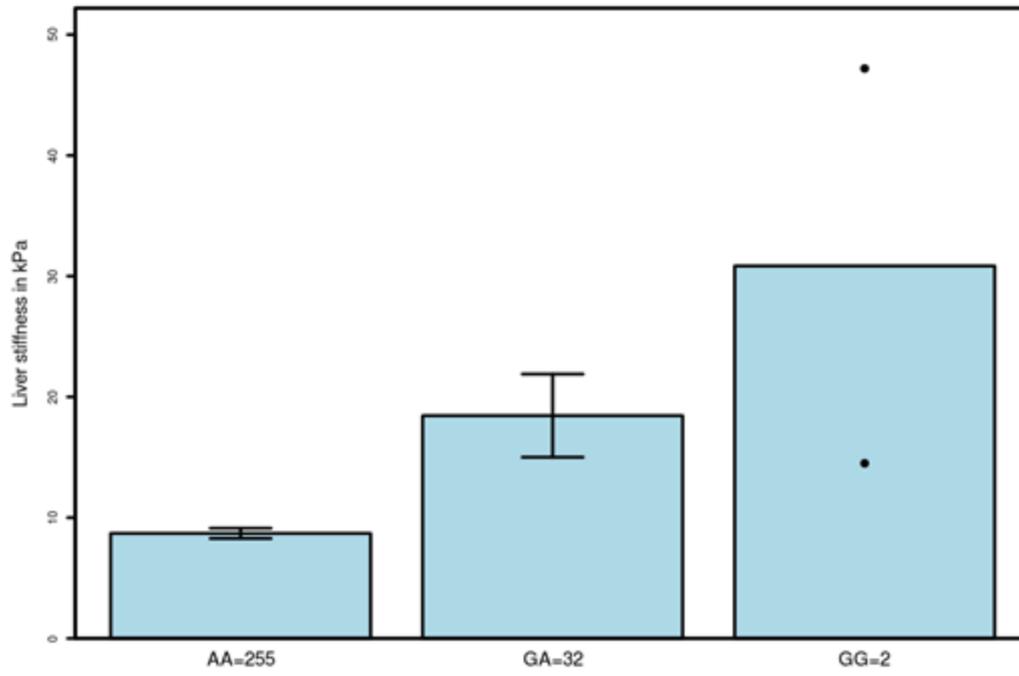
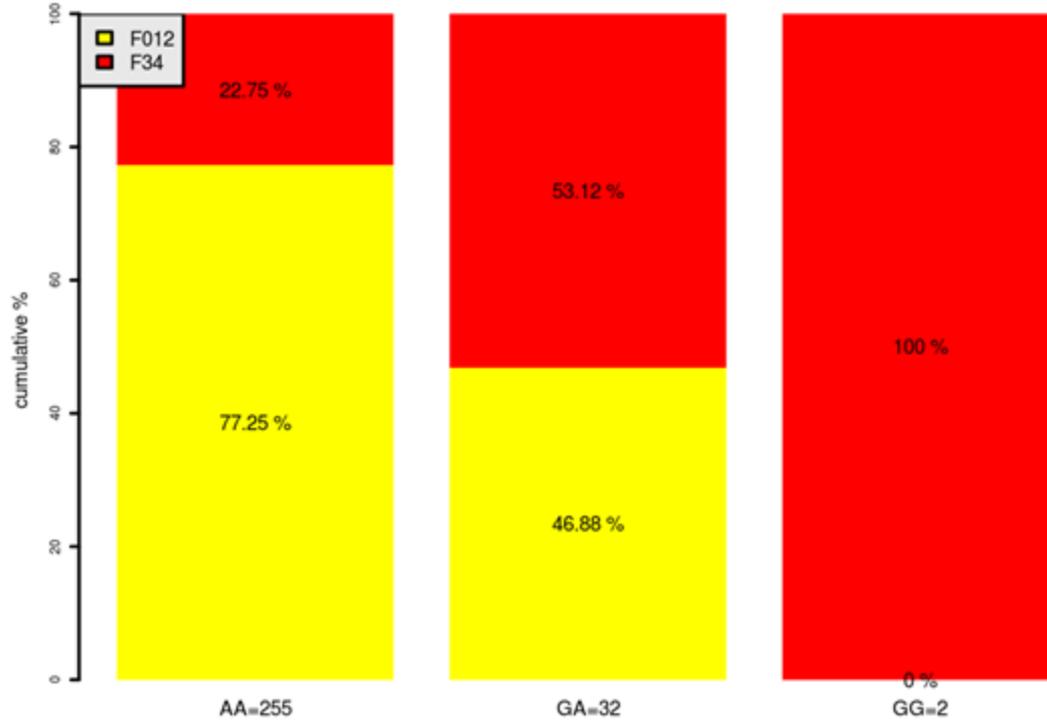
A**B**

Figure 2 – Distribution of liver stiffness according to the genotype at SNP rs61183828 associated with sqrtLS. (A). Mean of non-transformed liver stiffness by genotype. Each genotype, with its sample size, is plotted on the *x*-axis, whereas mean of liver stiffness, in kPa, is plotted on the *y*-axis. Error bars represented the standard error of the mean and the two black points show the GG individuals **(B) Non-transformed liver stiffness converted into a METAVIR score and used as a case/control outcome (F0F1F2 versus F3F4).** For each genotype on the *x*-axis, the sample size is also provided, and METAVIR scores group is indicated on the *y*-axis, together with the percentage of the subjects in the corresponding group.

Supporting Information:

Separated file (Supplementary_file_ulveling_et_al.docx)

Acknowledgments: The authors thank all the members of the ANRS CO13 HEPAVIH Study Group, the Swiss Hepatitis C Cohort Study Group, and the French ANRS HC EP 26 Genoscan Study Group, especially the physicians and nurses involved in cohort follow-up, and all the patients who took part in this study.

Conflict of Interest: The authors declare no conflict of interest

Supplemental Information:

Members of the Scientific Committee of the ANRS CO13 HEPAVIH Study Group: D Salmon (Investigator coordinator), F Dabis (co-investigator coordinator), L Wittkop (scientific and methodological coordinator), MA Loko, M. Winnock, P Sogni, P Trimoulet, J Izopet, V Paradis, B Spire, P Carrieri, C Katlama, G Pialoux, MA Valantin, J Chas, I Poizot-Martin, B Marchou, E Rosenthal, D Garipuy, O Bouchaud, A Gervais, C Lascoux-Combe, C Goujard, K Lacombe, A Bicart-See, C Duvivier, D Vittecoq, D Neau, P Morlat, F Bani-Sadr, L Meyer, F Boufassa, L Piroth, ML Chaix, F Abravanel, B Autran, AM Roque, C Solas, H Fontaine, L Serfaty, M Guiguet, G Chêne, D Costagliola, D Zucman, A Simon, S Dominguez, E Billaud, P Miaillhes, J Polo Devoto, H Aumaître, D Rey, O Segeal, L Marchand (ANRS), S Couffin-Cadiergues (ANRS).

Members of the Swiss Hepatitis C Cohort Study Group: Francesco Negro, Laurent Kaiser (Geneva), Darius Moradpour, Vincent Aubert (Lausanne), Jean-François Dufour, Fabio Giudici, Olivia Keiser, Meri Gorgievski (Berne), Markus Heim, Hans Hirsch, Marielle Rutquist, Pascal Benkert, Thomas Fabbro (Basel), Andreas Cerny (Lugano), Gladys Martinetti-Lucchini (Bellinzona), Raffaele Malinverni (Neuchâtel), Hans H. Siegrist (La Chaux-de-Fonds), David Semela, Pietro Vernazza, Patrick Schmid, Günter Dollenmaier (St. Gallen), Beat Müllhaupt, Philip Bruggmann, Elisabeth Probst-Müller (Zürich).

Members of the French ANRS HC EP 26 Genoscan Study Group: Bertrand Nalpas (Paris), Laurent Abel (Paris), Mona Monteanu (Paris), Laurence Bousquet (Paris), Yen Ngo (Paris), Pascal Lebray (Paris), Joseph Moussalli (Paris), Yves Benhamou (Paris), Dominique Thabut (Paris), Anaïs Vallet-Pichard (Paris), Hélène Fontaine (Paris), Vincent Mallet (Paris), Philippe Sogni (Paris), Jean-Baptiste Trabut (Paris), Marc Bourlière (Marseilles), Ioannis Theodorou (Paris), Jean-François Delfraissy (Paris), Thierry Poynard (Paris), Stanislas Pol (Paris).

Supplemental Tables:

Supplemental Table 1 – Demographic statistics based on the date of the first liver stiffness measure by FibroScan®. Statistical differences (*p*) are determined between F0F1F2 and F3F4 using the t-test for quantitative variables or chi² test for qualitative variables, included in the R software version 3.0.2 (<https://cran.r-project.org/>).

Characteristic	All patients (n = 289)	F0F1F2 (n = 210)	F3F4 (n = 79)	<i>p</i>
Sex (Male/Female) n (%)	200 (69.2) / 89 (30.8)	138 (66) / 72 (34)	61 (77.2) / 18 (22.8)	0.05
Age (years), median +/- SD	45.8 +/- 6.1	45.8 +/- 6.3	45.7 +/- 5.5	0.25
Liver stiffness (kPa) median +/- SD	6.7 +/- 9.8	5.9 +/- 1.5	15.3 +/- 13.8	2x10 ⁻¹⁴
Infection time (years), median +/- SD	24.4 +/- 7.9	24.3 +/- 8.1	24.7 +/- 7.3	0.38
HCV genotype, n (%)				
1	168 (58.1)	120 (57.1)	47 (59.5)	0.66
2	11 (3.8)	8 (3.8)	3 (3.8)	
3	52 (18)	35 (16.7)	17 (21.5)	
4	57 (19.7)	46 (21.9)	12 (15.2)	
5	1 (0.3)	1 (0.5)	0 (0)	
METAVIR score, n (%)				
0	109 (37.4)	109 (52.2)	0 (0)	ND
1	49 (17)	49 (23.4)	0 (0)	
2	52 (18)	52 (24.9)	0 (0)	
3	20 (6.9)	0 (0)	20 (6.9)	
4	59 (20.4)	0 (0)	59 (20.4)	
Alcohol status, n (%) (never/past/current)	49 (17) / 81 (28) / 159 (55)	34 (16.1) / 53 (25.2) / 123 (58.6)	15 (19) / (28 (35.4) / 36 (45.6)	0.13
SVR after treatment (Yes/No) n (%)	26 (9) / 263 (91)	21 (10) / 189 (90)	5 (6.3) / 74 (93.7)	0.49
CD4+ (cell/mm ³), median +/- SD	469 +/- 302.1	505 +/- 305.5	383 +/- 278.8	0.003
HIV viral load *	66 (22.8) / 223 (77.2) 3.7 +/- 1.7	50 (23.8) / 160 (76.2) 3.7 +/- 1.5	16 (20.3) / 60 (75.9) / 3 (3.8) 3.7 +/- 2.1	0.27
ARVT ** (years), median +/- SD	10.5 +/- 4.7	10.4 +/- 4.7	10.8 +/- 4.6	0.81

* HIV viral load: viral load status (detectable / undetectable / NA, n (%)) and logarithm of HIV copy per ml (median +/- SD)

** ARVT: effective time of HIV anti-retroviral treatment

Supplemental Table 2 – Distribution of covariates according to the genotypes AA or AG+GG at SNP rs61183828. Statistical differences (*p*) were determined between genotype groups using the t-test for quantitative variables or chi² test for qualitative variables, included in the R software version 3.0.2.

	Characteristic	AA (n = 255)	AG + GG (n = 32+2)	<i>p</i>
Covariates used in the linear regression model	Sex (Male/Female) n (%)	178 (69.8) / 77 (30.2)	22 (64.7) / 12 (35.3)	0.68
	Age (years), median +/- SD	45.4 +/- 6.1	47.2 +/- 5.8	0.34
	Infection time (years), median +/- SD	24.3 +/- 8	25.2 +/- 6.8	0.62
	HCV genotype 1 versus other	144 (56.5) / 111 (43.5)	22 (64.7) / 12 (35.3)	0.51
	Alcohol status, n (%) (never/past/current)	43 (16.9) / 69 (27.1) / 139 (54.5)	3 (8.8) / 12 (35.3) / 19 (55.9)	0.32
	ARVT * (years), median +/- SD	10.4 +/- 4.7	10.4 +/- 4.7	0.83
	SVR after treatment (Yes/No) n (%)	25 (9.8) / 230 (90.2)	1 (2.9) / 33 (97.1)	0.33
	CD4+ (cell/mm ³), median +/- SD	473.5 +/- 306.4	443 +/- 273.3	0.21
	HIV viral load **	3.7 +/- 1.7 58 (22.7) / 196 (76.9) / 1 (0.4)	3.7 +/- 1.7 9 (26.5) / 23 (67.6) / 2 (5.9)	0.51

* ARVT: effective time of HIV anti-retroviral treatment

** HIV viral load: status (detectable / undetectable / NA, n (%)) and logarithm of HIV copy per ml (median +/- SD)

Supplemental Table 3 – GWAS associations dealing with liver diseases. List of 322 SNPs, with $p < 10^{-5}$ in GWAS catalog (<https://www.ebi.ac.uk/gwas/>), from 31 GWAS dealing with 6 liver diseases. Five SNPs belonging to three common regions of our study (6q23, 11p13, 13q12) are highlighted in green and the p-values of association obtained in our cohort are indicated (p sqrtLS)

- **Summary**

Disease / Trait	Number of GWAS
PSC	6
Autoimmune hepatitis	1
PBC	5
ALD/NAFLD	12
liver related HBV	5
liver related HCV	2

- **PSC = primary sclerosing cholangitis**

First author	journal	pmid	SNPs	P<=10-5	region	Reported genes
Ellinghaus D	Hepatology	22821403	rs4676410	2,00E-009	2q37	<i>GPR35</i>
			rs1452787	3,00E-008	18q21	<i>TCF4</i>
Mellum E	Nat Genet	21151127	rs3197999	1,00E-016	3p21	<i>MST1</i>
			rs6720394	4,00E-008	2q13	<i>BCL2L11</i>
			rs12735793	1,60E-006	-	<i>PTPRU</i>
			rs896403	1,50E-006	-	<i>1q24</i>
			rs4391646	8,80E-007	-	<i>1q24</i>
			rs6678400	9,50E-007	-	<i>1q24</i>
			rs12043426	3,50E-006	-	<i>1q42</i>
			rs6971637	1,60E-007	-	<i>PION</i>
			rs7791854	1,00E-007	-	<i>PION</i>
			rs3807746	2,10E-006	-	<i>MAGI2</i>
			rs6973565	1,40E-006	-	<i>PLXNA4</i>
			rs13231950	4,30E-006	-	<i>PLXNA4</i>
			rs7462577	1,90E-006	-	<i>8q24</i>
			rs7038037	2,70E-006	-	<i>BSPRY</i>
			rs706778	5,40E-006	-	<i>IL2RA</i>
			rs3134883	8,50E-007	-	<i>IL2RA</i>
rs4147359	4,70E-006	-	<i>IL2RA</i>			
rs10905718	2,40E-007	-	<i>IL2RA</i>			
rs10891130	5,00E-006	-	<i>FDX1</i>			

			rs1793660	3,60E-006	-	<i>KIRREL3</i>
			rs12808353	2,40E-006	-	<i>VPS26B</i>
			rs11168249	1,00E-006	-	<i>HDAC7</i>
			rs9520835	9,00E-006	-	<i>TNFSF13B</i>
			rs4321167	5,90E-006	-	<i>15q26</i>
			rs17683107	8,50E-007	-	<i>GAS7</i>
			rs12458015	7,30E-006	-	<i>TCF4</i>
Karlsen TH	Gastroenterology	19944697	rs3099844	2,60E-026	6p21	<i>HLA-B</i>
			rs2844559	4,20E-026	6p21	<i>HLA-B</i>
Ellinghaus D	Nat Genet	26974007	rs2234161	1,53E-006	1	<i>TNFRSF14</i>
			rs7608910	7,21E-006	2	<i>PUS10</i>
			rs7426056	4,36E-022	2	<i>CD28</i>
			rs11306716	2,08E-006	2	<i>NA</i>
			rs11676348	6,48E-007	2	<i>CXCR2</i>
			rs7556897	4,74E-009	2	<i>AC073065.3</i>
			rs3749171	3,67E-009	2	<i>GPR35</i>
			rs3197999	1,84E-028	3	<i>MST1</i>
			rs116046827	4,50E-011	3	<i>BSN</i>
			rs3774937	6,11E-009	4	<i>NFKB1</i>
			rs13132308	2,53E-009	4	<i>IL21</i>
			rs9687958	1,28E-006	5	<i>NA</i>
			rs72928038	8,05E-012	6	<i>BACH2</i>
			rs61839660	1,82E-010	10	<i>IL2RA</i>
			rs3118471	1,70E-011	10	<i>IL2RA</i>
			rs10748781	4,37E-006	10	<i>NKX2-3</i>
			rs11168249	5,49E-009	12	<i>HDAC7</i>
			rs12369214	1,29E-009	12	<i>RIC8B</i>
			rs3184504	1,73E-011	12	<i>SH2B3</i>
			rs11649613	1,28E-011	16	<i>RP11-396B14.2</i>
			rs66504140	1,14E-006	18	<i>CD226</i>
			rs259964	2,78E-006	20	<i>ZNF831</i>
			rs9977672	4,90E-015	21	<i>NA</i>
			rs1893592	5,93E-008	21	<i>UBASH3A</i>
			rs4456788	5,05E-006	21	<i>AP001057.1</i>
Lui JZ	Nat Genet	23603763	rs3748816	7,40E-012	1p36	<i>MMEL1</i>
			rs7426056	1,90E-020	2q33	<i>CD28</i>
			rs317999	2,40E-026	3p21	<i>MST1</i>
			rs13140464	8,90E-013	4q27	<i>IL21</i>
			rs56258221	8,40E-012	6q15	<i>BACH2</i>
			rs4147359	8,20E-017	10p15	<i>IL2RA</i>
			rs7937682	3,20E-009	11q23	<i>SIK2</i>
			rs11168249	5,50E-009	12q13	<i>HDAC7</i>
			rs3118504	5,90E-011	12q24	<i>SH2B3</i>

			rs1788097	3,10E-008	18q22	<i>CD226</i>
			rs60652743	6,50E-010	19q13	<i>PRKD2</i>
			rs2836883	3,20E-017	21q22	<i>PSMG1</i>
Folseraas T	J Hepatol	22521342	rs3748816	2,10E-008	1	<i>TNFRSF14</i>
			rs13132245	1,20E-007	4	<i>IL21</i>
			rs13119723	9,90E-008	4	<i>IL21</i>
			rs2903692	6,40E-006	16	<i>CLEC16A</i>
			rs602662	1,90E-006	19	<i>FUT2</i>
			rs601338	2,70E-006	19	<i>FUT2</i>
			rs11682163	7,60E-006	2	<i>ALLC</i>
			rs281377	2,10E-006	19	<i>FUT2</i>

- **AutoimmuneHepatitis**

First author	journal	pmid	SNPs	P<=10-5	region	Reported genes
deBoer YS	Gastroenterology	24768677	rs2187668	1,50E-078	6p21	<i>(HLA-DQA1)</i>
			rs3184504	7,70E-008	12q24	<i>(SH2B3)</i>
			rs6000782	3,00E-006	22q13	<i>(CARD10)</i>

- **PBC = primary biliary cirrhosis**

First author	journal	pmid	SNPs	P <= 10-5	region	Reported genes
Cordell HJ	Nat Commun.	26394269	rs12712133	5,19E-009	2q12	<i>IL1R1</i>
			rs4973341	2,34E-010	2q36	<i>RNA5SP121</i>
			rs11724804	9,01E-012	4p16	<i>GAK</i>
			rs526231	1,14E-008	5q21	<i>PAM</i>
			rs2546890	1,06E-010	5q33	<i>RNF145</i>
		p sqrtLS = 0.44	rs6933404	1,27E-010	6q23	<i>LOC102723649</i>
			rs1859308	0,000000163	16p11	<i>IL4R</i>
Hirschfield GM	N Engl J Med	19458352	rs4679904	0,000001	3q25	<i>ARF7</i>
			rs6441286	2,00E-014	3q25	<i>IL12A</i>
			rs3790567	3,00E-011	1p31	<i>IL12RB2</i>
			rs907092	0,000008	17q12	<i>IKZF3</i>
			rs10488631	0,0000002	7q32	<i>IRF5-TNPO3</i>
			rs2395148, rs3135363, rs2856683, rs9357152	1,00E-010	6p21	<i>c6orf10</i>

			rs2395148, rs3135363, rs2856683, rs9357152	7,00E-010	6p21	<i>c6orf10</i>
Liu X	Nat Genet	20639880	rs10488631	3,00E-010	7q32	<i>TNPO3</i>
			rs9303277	0,000000002	17q21	<i>GSDMB</i>
			rs3745516	8,00E-011	19q13	<i>SPIB</i>
			rs3790567	8,00E-012	1p31	<i>IL12RB2</i>
			rs6441286	8,00E-012	3q25	<i>IL12A</i>
			rs7774434	3,00E-026	6p21	<i>HLA-DQB1</i>
Mells GF	Nat Genet	21399635	rs12134279	2,00E-014	1q31	<i>DENND1B</i>
			rs10931468	2,00E-019	2q32	<i>STAT4</i>
			rs2293370	3,00E-011	3q13	<i>CD80</i>
			rs7665090	4,00E-012	4q24	<i>NFKB1</i>
			rs860413	1,00E-011	5p13	<i>IL7R</i>
			rs6974491	0,00000004	7p14	intergenic
			rs6421571	3,00E-012	11q23	<i>CXCR5</i>
			rs1800693	0,000000002	12p13	<i>TNFRSF1A</i>
			rs911263	2,00E-011	14q24	<i>RAD51L1</i>
			rs12924729	3,00E-012	16p13	<i>CLEC16A</i>
			rs11117432	5,00E-011	16q24	intergenic
			rs968451	0,000000001	22q13	<i>MAP3K7IP1</i>
			rs17129789	9,00E-020	1p31	<i>IL12RB2</i>
			rs485499	2,00E-016	3q25	<i>IL12A</i>
			rs12531711	9,00E-017	7q32	<i>IRF5</i>
			rs7208487	0,00000008	17q12	<i>ORMDL3</i>
			rs3745516	2,00E-013	19q13	<i>SPIB</i>
			rs1372072	0,00000002	3p24	<i>PLCL2</i>
			rs538147	2,00E-010	11q13	<i>RPS6KA4</i>
			rs8017161	3,00E-013	14q32	<i>TNFAIP2</i>
			rs7774434	4,00E-034	6p21	<i>MHC</i>
Nakamura M	Am J Hum Genet	23000144	rs6890853	0,00000004	5p13	<i>IL7R</i>
			rs9303277	0,000000004	17q21	<i>GSDMB</i>
				0,0000001	2q32	<i>STAT4</i>
			rs2293370	0,000000003	3q13	<i>CD80</i>
			rs7665590	0,00000001	4q23	<i>NFKB1</i>
			rs4979462	3,00E-014	9q32	<i>TNFSF15</i>
			rs4938534	0,00000002	11q23	<i>POU2AF1</i>

- **ALD = alcoholic liver disease / NAFLD = non-alcoholic fatty liver disease**

First author	journal	pmid	SNPs	P <= 10-5	region	Reported genes
Romeo S	Nat Genet	18820647	rs738409	5,90E-010	22q13	<i>PNPLA3</i>

Yuan X	Am J Hum Genet	18940312	rs11597390	1,50E-008	10	<i>CPNI</i>
			rs11597086	1,80E-008	10	<i>CHUK</i>
			rs11591741	3,00E-008	10	<i>CHUK</i>
			rs2281135	8,40E-016	22	<i>PNPLA3</i>
			rs2143571	7,20E-009	22	<i>SAMM50</i>
			rs1169313	1,80E-010	12	<i>HNFA</i>
			rs4820599	4,00E-011	22	<i>GGT1</i>
			rs1780324	7,00E-015	1	<i>NBPF3-ALPL</i>
			rs9461011	1,90E-010	6	<i>GPLD1</i>
			rs9467160	1,20E-011	6	<i>GPLD1</i>
			rs8176720	4,30E-012	9	<i>ABO</i>
			rs641959	2,10E-011	9	<i>ABO</i>
			rs514708	1,90E-011	9	<i>ABO</i>
			rs672316	4,40E-010	9	<i>ABO</i>
			rs657152	1,70E-030	9	<i>ABO</i>
			rs474279	8,30E-009	9	<i>ABO</i>
			rs552148	1,20E-009	9	<i>ABO</i>
			rs12355784	5,00E-010	10	<i>JMJD1C</i>
			rs10761779	6,90E-010	10	<i>REEP3</i>
Chambers JC	Nat Genet	22001757	rs6834314	3,10E-009	4q22	<i>HSD17B13</i>
			rs2954021	5,30E-009	8q24	<i>TRIB1</i>
			rs10883437	4,00E-009	10q24	<i>CPNI</i>
			rs738409	1,20E-045	22q13	<i>PNPLA3</i>
			rs1976403	1,80E-050	1p36	<i>ALPL</i>
			rs16856332	1,60E-009	2q24	<i>ABCB11</i>
			rs1883415	5,60E-026	6p22	<i>ALDH5A1</i>
			rs6984305	2,10E-010	8p23	<i>PPP1R3B</i>
			rs2954021	2,30E-013	8q24	<i>TRIB1</i>
			rs10819937	1,00E-009	9q21	<i>ALDOB</i>
			rs579459	2,60E-123	9q34	<i>ABO</i>
			rs7923609	5,90E-023	10q21	<i>JMJD1C</i>
			rs174601	2,60E-009	11q12	<i>C11orf10</i>
			rs2236653	1,80E-009	11q.24	<i>ST3GAL4</i>
			rs7186908	4,80E-009	16q22	<i>HPR</i>
			rs314253	8,40E-012	17p13	<i>ASGR1</i>
			rs281377	1,10E-015	19q13	<i>FUT2</i>
			rs7267979	7,40E-010	20p11	<i>ABHD12</i>
			rs1497406	2,80E-019	1p36	<i>RSG1</i>
			rs12145922	3,80E-011	1p22	<i>CCBL2</i>
			rs1335645	7,30E-009	1p13	<i>CEPT1</i>
			rs10908458	1,70E-015	1q21	<i>DPM3</i>
			rs1260326	3,90E-013	2p23	<i>C2orf16</i>
			rs13030978	1,10E-011	2q12	<i>MYO1B</i>
			rs2140773	1,10E-009	2q37	<i>EFHD1</i>
			rs10513686	6,10E-011	3q26	<i>SLC2A2</i>
			rs4547811	2,50E-027	4q31	<i>ZNF827</i>

			rs6888304	1,20E-009	5p15	<i>CDH6</i>
			rs4074793	3,40E-010	5q11	<i>ITGA1</i>
			rs9296736	2,60E-009	6p12	<i>MLIP</i>
			rs17145750	2,90E-009	7q11	<i>MLXIPL</i>
			rs754466	6,40E-010	10q23	<i>DLG5</i>
			rs7310409	7,00E-045	12q24	<i>HNFA1A</i>
			rs944002	5,80E-029	14q32	<i>C14orf73</i>
			rs339969	6,60E-020	15q21	<i>RORA</i>
			rs8038465	1,40E-009	15q23	<i>CD276</i>
			rs4581712	3,10E-009	16q23	<i>DYNLRB2</i>
			rs9913711	1,30E-009	17q24	<i>FLJ37644</i>
			rs12968116	8,90E-010	18q21	<i>ATP8B1</i>
			rs4503880	3,00E-012	18q21	<i>NEDD4L</i>
			rs516246	7,60E-010	19q13	<i>FUT2</i>
			rs1076540	9,60E-017	22q11	<i>MICAL3</i>
			rs2739330	1,70E-009	22q11	<i>DDT</i>
			rs2073398	1,10E-109	22q11	<i>GGT1</i>
Feitosa MF	Atherosclerosis	23477746	rs2862954	4,88E-010	10q24	<i>ERLIN1</i>
			rs1408579	4,57E-010	10q24	<i>ERLIN1</i>
			rs10883451	4,29E-010	10q24	<i>ERLIN1</i>
			rs11597086	1,46E-009	10q24	<i>CHUK</i>
			rs11591741	1,84E-009	10q24	<i>CHUK</i>
			rs17729876	1,82E-009	10q24	<i>CWF19L1</i>
			rs17668255	1,82E-009	10q24	<i>CWF19L1</i>
			rs17668357	1,65E-009	10q24	<i>CWF19L1</i>
			rs12784396	2,47E-09	10q24	<i>CWF19L1</i>
Butch S	Nat Genet	26482880	rs738409	1,54E-048	22	<i>PNPLA3</i>
			rs10401969	7,89E-010	19	<i>TM6SF2</i>
			rs58542926	1,33E-009	19	<i>TM6SF2</i>
			rs626283	1,03E-009	19	<i>MBOAT7</i>
Adams LA	Hepatology	23213074	rs12743824	5,00E-006	1p21	<i>LPPR4</i>
			rs222054	1,00E-006	4q13	<i>GC</i>
			rs7324845	3,00E-006	13q14	<i>LCPI</i>
			rs11864146	2,00E-006	16q23	<i>SLC38</i>
Chalasan N	Gastroenterology	20708005	rs343064	3,00E-008	7p14	intergenic
			rs11669592	4,00E-006	19p13	intergenic
			rs4243849	2,00E-006	8q24	intergenic
			rs2358462	4,00E-006	4q31	intergenic
			rs9977253	5,00E-006	21q21	intergenic
			rs6027755	6,00E-006	20q13	intergenic
			rs6079395	7,00E-006	20p12	<i>MACROD2</i>
			rs2800	3,00E-006	3q24	<i>SLC9A9</i>
			rs894177	3,00E-006	3q24	intergenic
			rs1305088	9,00E-006	13q12	<i>SLC46A3</i>
			p sqrtLS =	0.45		

			rs1227756	2,00E-007	10q22	<i>COL13A1</i>
			rs887304	8,00E-007	12p13	<i>EFCAB4B</i>
			rs9584805	2,00E-006	13q32	<i>FARP1</i>
			rs4237591	2,00E-006	11q22	intergenic
			rs2986971	3,00E-006	10p11	intergenic
			rs2216228	4,00E-006	12p12	<i>ST8SIA1</i>
			rs7632299	4,00E-006	3q24	intergenic
			rs10067427	5,00E-006	5q21	intergenic
			rs10954668	7,00E-006	7q21	<i>CACNA2D1</i>
			rs7077164	7,00E-006	10q22	<i>COL13A1</i>
			rs11083271	8,00E-006	18q12	intergenic
			rs2499604	2,00E-006	1q43	<i>ZP4</i>
			rs2710833	6,00E-007	4q32	<i>DDX60L</i>
	p sqrtLS =	0.82	rs2049953	4,00E-006	6q23	intergenic
			rs12344488	5,00E-006	9q21	intergenic
			rs1529093	2,00E-006	2q31	intergenic
			rs959903	7,00E-006	4p15	<i>SEL1L3</i>
			rs2645424	7,00E-007	8p23	<i>FDFT1</i>
			rs1836127	3,00E-006	12q22	intergenic
			rs643608	1,00E-006	21q22	intergenic
			rs6591182	9,00E-007	11q13	<i>LTBP3</i>
			rs2645424	3,00E-006	8p23	<i>FDFT1</i>
			rs6487679	1,00E-006	12p13	<i>PZP</i>
			rs1414896	2,00E-006	1p21	intergenic
			rs10510146	7,00E-006	10q26	<i>C10orf122</i>
Kawaguchi T	PLoS One	22719876	rs738409	1,00E-010	22q13	<i>PNPLA3</i>
			rs11206226	5,00E-006	1p32	<i>YIPF1</i>
			rs1390096	3,00E-006	4p15	<i>HS3ST1</i>
			rs2143571	6,00E-007	22q13	<i>SAMM50</i>
			rs2073080	8,00E-007	22q13	<i>PARVB</i>
Kitamoto T	Hum Genet	23535911	rs6691847	7,00E-006	1p35	<i>PTPRU</i>
			rs7552722	7,00E-006	1p13	<i>NGF</i>
			rs2051090	7,00E-006	13q13	<i>DCLK1</i>
			rs2896019	2,00E-020	22q13	<i>PARVB</i>
Speliotes EK	PLoS Genet	21423719	rs738409	4,00E-034	22q13	<i>PNPLA3</i>
			rs2228603	1,22E-011	19p12	<i>NCAN</i>
			rs12137855	2,59E-008	1	<i>GCKR</i>
Kozlitina	Nat Genet	24531328	rs738409	4,00E-016	22q13	<i>PNPLA3</i>
			rs2281135	6,90E-012	22q13	<i>PNPLA3</i>
			rs58542926	5,70E-008	19p13	<i>TM6SF2</i>
DiStefano JK	Acta Diabetol	25246029	rs10401969	3,10E-007	19p13	<i>SUGP1</i>
			rs4823173	2,80E-007	22q13	<i>PNPLA3</i>
			rs2896019	3,10E-007	22q13	<i>PNPLA3</i>

- Hepatitis B liver disease

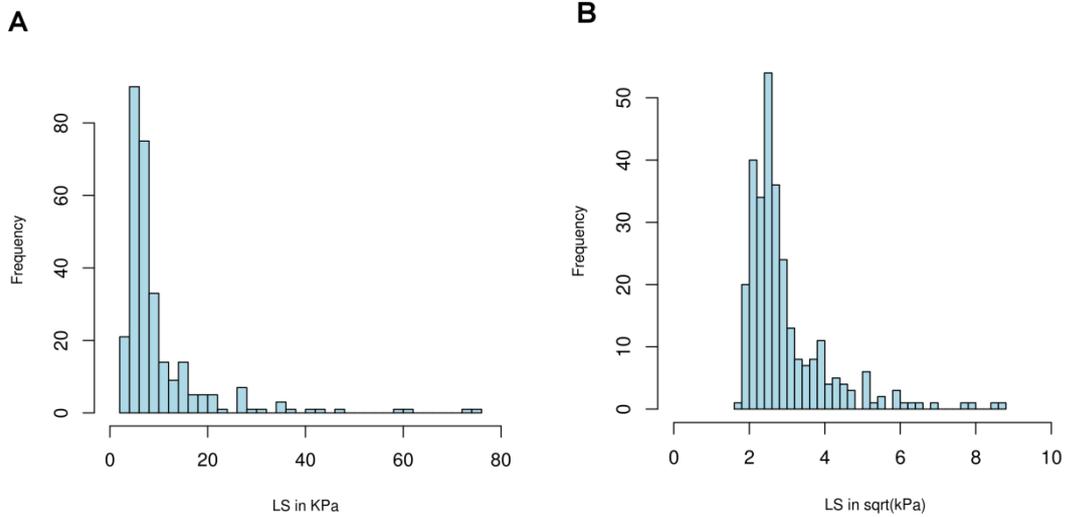
First author	journal	pmid	SNPs	P <= 10-5	region	Reported genes		
Jiang DK	Nat Genet	23242368	rs7574865	2,00E-010	2q32	<i>STAT4</i>		
			rs9275319	3,00E-017	6p21	<i>HLA-DQ</i>		
Zhang H	Nat Genet	20676096	rs17401966	2,00E-018	1p36	<i>UBE4B</i>		
			rs7898005	7,00E-008	10q24	intergenic		
			rs2396470	7,00E-007	2q37	<i>SLC19A3</i>		
Li S	PLoS Genet	22807686	rs455804	5,00E-010	21q21	<i>GRIK1</i>		
			rs9272105	5,00E-022	6p21	<i>HLA-DRB1</i>		
			rs7424161	8,76E-006	2q22	<i>LRP1B</i>		
			rs3825023	3,14E-006	11p15	<i>SAA2</i>		
			rs10896464	6,45E-006	11q13	<i>MYEOV/CCND1</i>		
			rs3092194	4,41E-006	20q12	<i>CHD6</i>		
Al-Qahtani A	J Med Genet	24065354	rs7553612	4,33E-006	1p31	<i>LOC100289178</i>		
			rs6716352	9,82E-006	2q22	<i>ARHGAP15</i>		
			rs34246867	2,54E-006	3q23	<i>MRPS22</i>		
			rs35312459	4,67E-006	4q12	<i>EXOC1</i>		
			rs6867291	7,81E-006	5q23	intergenic		
			rs6867291	8,02E-006	5q23	intergenic		
			rs502071	7,68E-006	6q22	<i>HSF2</i>		
			rs4445240	5,93E-006	8q21	<i>CNBD1</i>		
			rs12784800	3,71E-006	10q22	<i>PPIF</i>		
			rs2724432	1,89E-007	11q22	<i>FDX1</i>		
			rs17033545	5,97E-006	12q23	<i>RP11-328J6.1</i>		
				p sqrtLS = 0.79	rs1778797	6,98E-006	13q12	<i>AL390738.1</i>
					rs7998378	8,20E-006	13q22	<i>KLF5</i>
					rs8054283	8,93E-006	16q23	<i>CDH13</i>
					rs753610	6,45E-006	18q22	<i>LOC643542</i>
					rs17766247	3,78E-006	18q22	<i>LOC643542</i>
		rs5761699	3,98E-006	22q12	<i>CTA-211A9.5</i>			
		rs5761699	7,64E-006	22q12	<i>CTA-211A9.5</i>			
		rs5930539	1,49E-008	X	<i>MST4</i>			
		rs5930539	4,26E-007	X	<i>MST4</i>			
		rs6637934	1,80E-007	X	<i>FRMD7</i>			
		rs6637934	3,70E-006	X	<i>FRMD7</i>			
		rs16989995	7,93E-006	X	<i>DMD</i>			
		rs12014086	4,06E-006	X	<i>EFHC2</i>			
		rs17347621	2,31E-006	X	<i>ACOT9</i>			
Chan KY	PLoS One	22174901	rs12682266	6,69E-006	8	intergenic		
			rs7821974	7,01E-006	8	intergenic		

			rs2275959	6,35E-006	8	intergenic
			rs1573266	7,44E-006	8	intergenic
	p sqrtLS =	0.52	rs2611145	9,31E-006	11p13	<i>ABTD2</i>

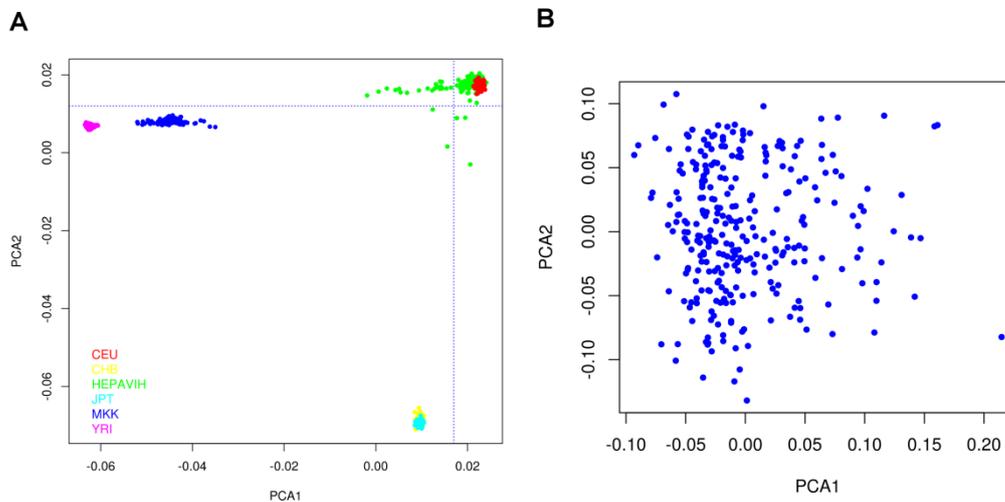
- **Hepatitis C liver disease**

First author	journal	pmid	SNPs	P <= 10-5	region	Reported genes
Patin E	Gastroenterology	22841784	rs16851720	9,00E-009	3q23	<i>RNF7</i>
			rs4374383	1,00E-009	2q13	<i>MERTK</i>
			rs9380516	5,00E-007	6p21	<i>TULP1</i>
			rs2629751	1,00E-007	12q23	<i>GLT8D2</i>
			rs883924	2,00E-006	9q22	<i>LOC340515</i>
			rs7800244	3,00E-006	7p12	<i>PKD1L1</i>
Urabe Y	J Hepatol	23321320	rs910049	9,00E-011	6p21	<i>C6orf10</i>
			rs3135363	1,00E-010	6p21	intergenic
			rs3129860	1,00E-009	6p21	intergenic
			rs3817963	1,00E-008	6p21	<i>BTNL2</i>
			rs9405098	1,00E-010	6p21	intergenic

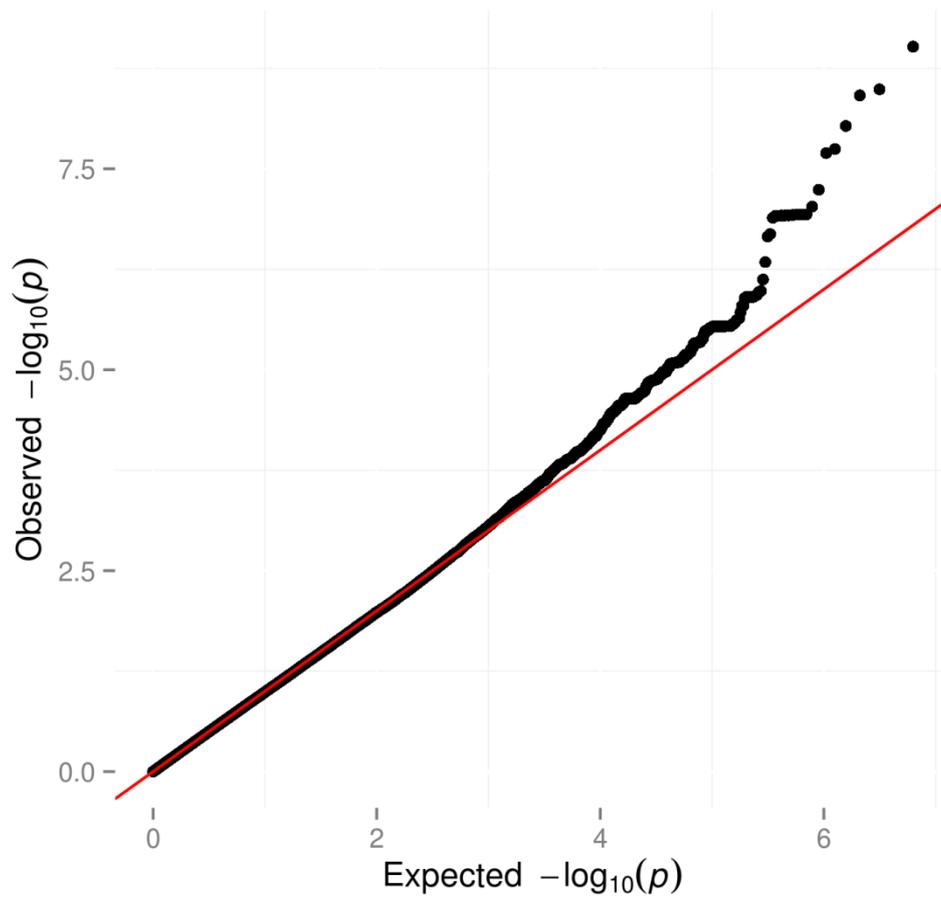
Supplemental Figures:



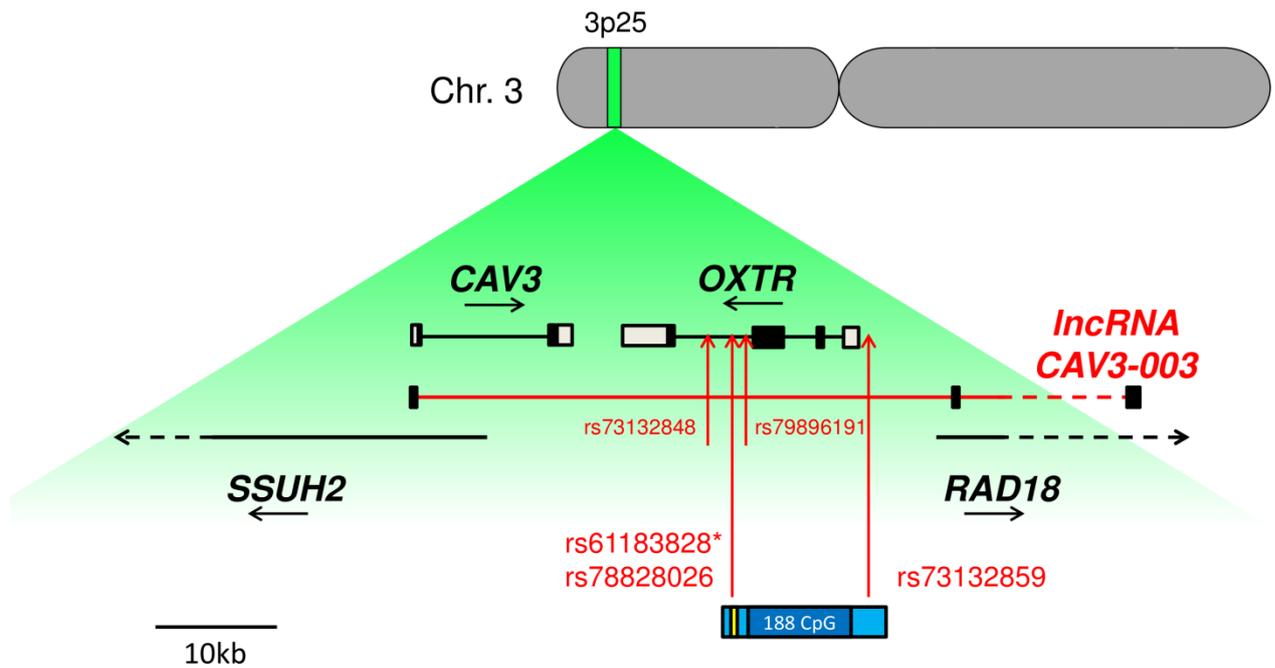
Supplemental Figure 1 – Distribution of liver stiffness in the HEPAVIH sample used in analysis. (A). Liver stiffness (LS) in kPa. (B). Liver stiffness in square-root (sqrtLS) of kPa.



Supplemental Figure 2 – Principal component analysis (PCA) plots. (A) PCA of HEPAVIH and HAPMAP individuals. Each point represents a patient. HEPAVIH individuals are in green. (B) PCA of HEPAVIH individuals after removing of non-European samples. Stratification of the 292 individuals include in the analysis.



Supplemental Figure 3 – Quantile–quantile (QQ) plot. The plot displays the observed (y-axis) versus expected (x-axis) $-\log_{10}$ (p-value). The lambda value is the genomic inflation factor which shows no inflation of the test statistic (lambda = 0.996).



Supplemental Figure 4 – Genomic context of the cluster signal on chromosome region 3p25. Four SNPs are located in *OXTR* in introns and one in the 5' intergenic region of the gene; the genotyped SNP rs61183828 is followed by an asterisk. The rs78828026; in perfect LD with the rs61183828; is within a FOXA1 transcription factor binding site (TFBS in yellow). The TFBS is in a potential transcriptional region (lightblue) which has 188 CpG islands (blue). Gene regions are characterized by white box for UTR regions, black box for exons, and black line for introns. The transcriptional direction is oriented by the arrow below the gene name. The *CAV3* isoform 3 is in red line with two exons in *RAD18* intronic sequences.

2 Etude « génome entier » de la cohorte ANRS CO13 HEPAVIH de patients coinfecteds par le VIH et le VHC : à la recherche d'associations liées à l'impact de la gravité de fibrose par comparaison de groupes de score METAVIR F0F1F2 et F3F4.

A GWAS reveals *CTNND2* gene polymorphisms associations in liver fibrosis by comparing F0F1F2 and F3F4 METAVIR score groups in HIV/HCV coinfection.

Damien Ulveling*, Sigrid Le Clerc, Taoufik Labib, Josselin Noirel, Vincent Laville, Cédric Coulonges, Wassila Carpentier, Dominique Salmon, François Dabis, Yves Lévy, Stéphanie Dominguez* and Jean-François Zagury* and the HEPAVIH ANRS CO13 cohort

Article en préparation

RESUMÉ :

Le score de fibrose METAVIR (F) permet d'estimer le stade de fibrose caractérisant les lésions fibreuses hépatiques existantes. C'est une évaluation semi-quantitative qui permet de « normaliser » la plupart des méthodes d'évaluation de la fibrose hépatique. Dans cette étude les scores METAVIR de 292 patients ont été évalués à partir d'examen de Fibrotest®, Fibroscan® et ponction hépatique, à la date d'inclusion dans la cohorte ANRS CO13 HEPAVIH. Avec un phénotype dichotomique définissant un groupe de 197 patients avec pas ou peu de fibrose (F0F1F2) et un groupe de 95 patients présentant des fibroses sévères voire des cirroses (F3F4), nous avons entrepris de comparer ces deux groupes au niveau de chaque variant génétique dans le cadre d'une analyse « génome entier ».

Nous avons analysé un jeu de données de 8 426 597 de SNPs génotypés ou ayant un fort score de qualité d'imputation, après un contrôle qualité et des procédures d'imputation utilisant le

panel de référence Européen de 1000 Génomes Phase I (N = 379) à partir de données de génotypage d'une puce Illumina Omni2.5 BeadChip (Illumina, San Diego, USA).

La comparaison des groupes de score METAVIR F0F1F2 et F3F4 a permis d'identifier deux signaux « génome entier » significatif. Le premier a été trouvé dans la région chromosomique 1q31 (rs72727113, $\beta = 2,37$, SE = 0,41, p -valeur = $8,2 \times 10^{-9}$) où de nombreux CNVs ont été décrits. Le second signal est situé dans la région chromosomique 5p15 et correspond à trois SNPs et deux indels, tous en LD ($r^2 > 0,9$ selon 1000 Génomes) et introniques du gène *CTNND2*. Le signal est porté par l'association du SNP rs2158099 ($\beta = -1,67$, SE = 0,29, p -valeur = $6,1 \times 10^{-9}$). Le gène *CTNND2* semble important dans les mécanismes de formation de cancer, notamment pour les hépatocarcinomes cellulaires. Ce gène code une protéine, la δ -caténine, impliquée dans les jonctions adhérentes cellulaires et interagit avec des protéines de la voie de signalisation Wnt/Fzd/ β -caténine, qui joue un rôle important dans la physiopathologie hépatique.

Dans cette étude génome-entier, nous avons donc identifié deux signaux, dont un très pertinent pour les maladies hépatique. Ces résultats constituent une avancée dans la compréhension des mécanismes moléculaires associés à la fibrose sévère. Il est important d'essayer de répliquer ces résultats d'autres cohortes de patients coinfectés VIH/VHC.

A GWAS reveals CTNND2 gene polymorphisms associations in liver fibrosis by comparing F0F1F2 and F3F4 METAVIR score groups in HIV/HCV coinfection.

Damien Ulveling^{1*}, Sigrid Le Clerc¹, Taoufik Labib¹, Josselin Noirel¹, Vincent Laville¹, Cédric Coulonges¹, Wassila Carpentier², Dominique Salmon³, François Dabis⁴, Yves Lévy⁵, Stéphanie Dominguez^{5*} and Jean-François Zagury^{1*} and the HEPAVIH ANRS CO13 cohort**

¹Équipe Génomique, Bioinformatique et Applications (EA4627), Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France.

²Plateforme Post-Génomique P3S, AP-HP, UPMC Université Paris 6, Faculté de Médecine Pitié Salpêtrière, Paris, France

³Department of Infectious Diseases, Cochin Hospital, Paris, France

⁴Centre de Recherche INSERM U897, Epidémiologie-Biostatistique, Institut de Santé Publique, Epidémiologie et Développement, Université de Bordeaux, Bordeaux, France

⁵INSERM U955, AP-HP, Groupe Henri-Mondor Albert-Chenevier, Immunologie Clinique, Créteil, France

* These authors contributed equally to this study

** See Supporting Information for the full names of all members of the HEPAVIH ANRS CO13 cohort study group

Keywords: HIV/HCV, Liver Disease, GWAS, METAVIR, CTNND2

Corresponding authors:

Stéphanie Dominguez, Département d'Immunologie Clinique, INSERM U955, AP-HP, Groupe Henri-Mondor Albert-Chenevier, 51, avenue du Mal de Lattre de Tassigny 94010 Créteil,

Phone: 33 01 49 81 44 42; Fx: 33 01 49 81 24 69; e-mail =

stephanie.dominguez@hmn.aphp.fr

Jean-François Zagury, Équipe Génomique, Bioinformatique et Applications (EA4627),

Chaire de Bioinformatique, CNAM, 75003, Paris,

Phone: 33 1 58 80 88 20; e-mail = zagury@cnam.fr

List of Abbreviations:

HCV: hepatitis C virus; HIV: human immunodeficiency virus; HAART: highly active antiretroviral therapy; GWAS: genome-wide association study; SNP: single-nucleotide polymorphism; HWE: Hardy-Weinberg equilibrium; MAF: minor allele frequency; METAVIR: meta-analysis of histological data in viral hepatitis; SVR: sustained virological response; LD: linkage disequilibrium; CNV: copy number variant

Financial Support:

This study was sponsored and funded by the French National Agency for Research on AIDS and Viral Hepatitis (ANRS), with the participation of Glaxo-SmithKline, Roche and Bristol Myers Squibb. The study was also partly funded by a grant from Merck Sharp and Dohme – France.

Abstract

Background: There is growing evidence that human genetic variants contribute to liver fibrosis in subjects with hepatitis C virus (HCV) mono-infection, but this aspect has been little investigated in patients co-infected with HCV and human immunodeficiency virus (HIV). We performed the first genome-wide association study (GWAS) of liver fibrosis of patients co-infected with HCV and HIV, using the well-characterized French ANRS CO13 HEPAVIH cohort. Liver fibrosis was evaluated in METAVIR score from Fibroscan®, Fibrotest® or liver biopsy, providing a qualitative fibrosis score. After quality control, GWAS was conducted on 292 Caucasian patients, by comparing 197 patients with low METAVIR score (< F3) with 95 patients with high METAVIR score (\geq F3), for a total of 8,426,597 genotyped (Illumina Omni2.5 BeadChip) or reliably imputed SNPs.

Results: Two signals of genome-wide significance (p -value $< 5 \times 10^{-8}$) were obtained. The first, on chromosome 1q31 and corresponding to rs72727113 (p -value = 8.18×10^{-9}) a single imputed SNP with no obvious link to a gene. The second signal, obtained through rs2158099 (p -value = 6.05×10^{-9}) on chromosome region 5p15, includes 5 intronic variants of the *CTNND2* gene which seems involved in HCV-induced hepatocellular carcinoma mechanisms.

Conclusion: Our GWAS identified a new locus associated with liver fibrosis severity in patients with HIV/HCV co-infection, on chromosome 5p15. This result was found in *CTNND2*, a gene involved in the Wnt/Fzd/ β -catenin signaling pathway which plays a role in liver physiopathology. These findings provide new insight into the molecular mechanisms of liver fibrosis in patients with HIV/HCV co-infection.

Introduction

There are 37 million individuals living with the Human Immunodeficiency Virus (HIV) infection (1) and it is estimated that approximately 20-30% of HIV patients are also infected with the Hepatitis C Virus (HCV) (2). HIV/HCV co-infection is common due to their shared routes of transmission and the main risk factor among co-infected patient is the intravenous injection of drugs with prevalence rates reported as high as 90% (3). Since the introduction of highly active antiretroviral therapy (HAART), which has improved the survival of AIDS patients, HCV has emerged as a major comorbid disease causing progressive liver disease, potentially leading to end-stage liver disease, hepatocellular carcinoma, or death (4, 5). It appears that the rate of fibrosis progression and liver related diseases are accelerated in patients with HIV/HCV co-infection than in patients infected with HCV alone (6, 7). However, the risk factors for progressive liver disease in HIV/HCV-co-infected patients are not yet fully understood (8). Several studies have suggested that the effective treatment of HIV and HCV may be associated with a lower risk of liver disease progression (9-11). Risk factors such as alcohol consumption, age, and obesity have also been described, but account for only a small proportion of the variability in liver fibrosis development (12). There is growing evidence to suggest that host genetic factors are involved, at least in patients infected with HCV alone, but these factors remain largely unknown (13, 14).

Several candidate gene approaches have been used to identify variants influencing the development or the treatment of liver disease in patients with HCV infection, generally as a single infection, but most of these studies produced results that were not consistently replicated (15, 16). Genome-wide association studies (GWAS) provide a broader, unbiased approach for the discovery of genetic factor involved in disease susceptibility (17). In HIV-

related phenotypes, such studies have mostly identified HLA loci associated with the control of viral load and with non-progression to AIDS (18). Several GWAS of HCV infection-related phenotypes have identified polymorphisms of the *IL28B* (also denoted as *IFNL3/4*) gene associated with both the spontaneous clearance of HCV and a favorable response to interferon/ribavirin-based treatment, both in patients infected with HCV only (19) and in patients with HIV/HCV co-infection (20). By a candidate gene approach, polymorphisms of *IL28B* gene have also been associated with liver inflammation and hepatic fibrosis in HCV mono-infection (21-23). Only two GWAS have focused on the HCV-related liver fibrosis outcome in patients with mono-infections. The first, in a Caucasian population, identified four susceptibility loci, three of which were linked to genes involved in apoptosis (22). The second, conducted in Japan, detected variants within the HLA region (24). Moreover, GWAS identified *DEPDC5* and *MICA* as susceptibility genes for HCV-induced hepatocarcinoma in Japanese populations (25-27). We therefore carried out this GWAS to identify new genetic markers associated with liver fibrosis by comparing subjects with a low METAVIR score (F0F1F2 group) and with subjects with a high METAVIR score (F3F4 group), in the prospective French HEPAVIH ANRS CO13 cohort of adults with HIV/HCV co-infection.

Materials and methods

Study design and population

This study was based on a population sample of 374 patients with HIV/HCV co-infection who had provided written informed genetic consent and were enrolled in the French National Agency for Research on AIDS and Viral Hepatitis (ANRS) CO13 HEPAVIH cohort (28). The enrollment criteria were as follows: patients aged 18 years or over, with chronic HIV/HCV co-infection confirmed by a positive test for HIV-1 antibody and by an HCV RNA assay (regardless of clinical stage, sex or transmission group). The dual infection in this cohort was well-characterized because patients had full follow-up medical visits every six months in cases of cirrhosis and annually for patients without cirrhosis (median follow-up for the cohort: ~4 years). A self-administered questionnaire was used to record socio-demographic characteristics and data concerning past and current smoking habits, drug use and alcohol consumption.

374 patients were genotyped, however, after quality control and validation of the dataset information (availability of covariates); only 292 patients remained included for the genetic analyses (see below). All these patients are treated with HAART and two-thirds received a pegylated interferon +/- ribavirin treatment for HCV infection (n=200). 25 (8.5%) patients have achieved a sustained virological response (SVR) after treatment before the inclusion in the cohort but continued the monitoring. Because of their low number and in order to assess the fibrosis evolution after SVR, we kept those patients in the analysis. We also assessed the effect of SVR on GWAS results by analyzing the most significant signals by removing of these 25 patients.

Liver histology and staging of liver fibrosis for the binary comparison of low versus high METAVIR score groups

The stage of liver fibrosis in patients co-infected was determined either by an examination of a liver biopsy specimen (24%), or by a Fibrotest® (16.8%), or by a FibroScan® (59.2%), leading to the standardized attribution of the METAVIR score in a five-point scale from F0 to F4 (29). The distribution of individuals according to their METAVIR score is presented in Supplemental Table1 and Supplemental Figure 1.

We used a classic case/control approach to look for genetic determinants of liver fibrosis, by comparing the group of patients with score of F0, F1, or F2 with the group of patients with score F3 or F4.

Genotyping and quality control procedures

The Illumina Omni2.5 BeadChip (Illumina, San Diego, USA), which contains 2,391,739 markers, was used to genotype 374 patients. Raw data were first analyzed with Genome Studio software (version 1.6.3; Illumina) to obtain genotype calls. Thirty samples with a call rate (percentage of SNPs genotyped per sample) < 95% were removed. Two pairs of individuals displayed high levels of relatedness on analysis with PLINK software (30). One patient from each of these pairs was retained for the analysis. At this stage, 342 samples were retained. We subsequently excluded 210,140 SNPs with a low call rate (< 98%), 312 SNPs not in Hardy-Weinberg equilibrium (HWE, p -value < 10^{-6}) and 1,025,509 SNPs with a low minor allele frequency (MAF < 5%). In total, 1,155,778 SNPs were retained for imputation procedures and analysis.

Population stratification analysis

The genotypes were analyzed by principal component analysis (PCA), with the SMARTPCA utility of the EIGENSOFT package version 4.2 (31), to correct for possible population stratification. Before the analysis, we removed 23 individuals who described themselves as

“non-European”. We then carried out two rounds of SMARTPCA, which identified 27 outliers. These outliers were excluded from subsequent analyses (Supplemental Figure 2A), which were carried out on a final sample of 292 patients used for the GWAS. A third analysis without outliers was performed to determine the eigenvectors (Supplemental Figure 2B). In the statistical analysis, we used the two first eigenvectors as covariates to correct for population substructure in the association analyses.

Imputation procedures

The genotype data were phased with SHAPEIT2, and the phased data were then imputed with IMPUTE2 (32, 33). Haplotype data for 1094 European individuals from the phase I integrated variant set of the 1000 Genomes project released in March 2012 and updated in August 2012 (34) were used as a reference panel. After all procedures had been carried out, we retained 8,426,597 SNPs with a high imputation quality score, as measured by the information metric (> 0.8), and with a MAF $> 5\%$.

Covariates used for the association analysis

We focused more specifically on the genetic factors affecting liver fibrosis, by adjusting the regression analyses for possible confounding factors defined at the enrollment date: sex, age (in years), alcohol consumption (never, previous and current), HCV genotype (1 versus other), duration of HIV/HCV infection (in years) and effective time on antiretroviral treatment for HIV infection (in years). The full set of covariates was available of 292 patients. The top two eigenvectors were also used to correct for population substructure.

Statistical tests

Genomic associations with the comparison of F0F1F2 and F3F4 groups were investigated by logistic regression analysis in SNPTEST 2.4.1 software (35), with a genetic dominant model and genotype dosage for the imputed variants, and a complete set of the relevant covariates. The classical genome-wide significance threshold (p -values $\leq 5 \times 10^{-8}$) was used (35, 36).

Regional association plots

Regional association plots were created using the locus zoom “plot your own data” function (<https://statgen.sph.umich.edu/locuszoom/genform.php?type=yourdata>). Plots were created utilizing the genome build/LD population hg19/1000 Genomes Nov 2014 EUR.

Search of common signals with GWAS dealing with other liver diseases

From our results, we have performed a thorough search for replication in the GWAS dealing with liver diseases, first in the GWAS catalog (<https://www.ebi.ac.uk/gwas/>), second in the published literature (<http://www.ncbi.nlm.nih.gov/pubmed>). Overall, we have recovered 322 SNPs (p -value $< 10^{-5}$) from 31 GWAS dealing with HCV liver diseases or non-HCV liver diseases (Supplemental Table 3, HBV mono-infection, primary sclerosing cholangitis, autoimmune hepatitis, primary biliary cirrhosis, alcoholic liver disease and non-alcoholic fatty liver disease).

Bioinformatics exploration databases

We explored the polymorphisms detected in our GWAS further, by looking for changes in mRNA levels (GTEx Portal, www.gtexportal.org/home/), splicing (NetGene2, www.cbs.dtu.dk/services/NetGene2/), transcription factor binding sites (RegulomeDB, www.regulomedb.org/), HaploReg www.broadinstitute.org/mammals/haploreg/haploreg.php, v2, www.broadinstitute.org/mammals/haploreg/haploreg.php), and miRNA genes or miRNA

targets (miRBase, www.mirbase.org/, miRTarBase, www.mirtarbase.mbc.nctu.edu.tw/,
MicroCosm Targets, www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/).

Results

Sample description

Baseline demographic description of the 292 patients retained for the genetic association study (see Methods) is described in Supplemental Table 1. The data shown were recorded at the enrollment date of patients. About 75% of patients were co-infected via intravenous drug injection and the median age at infection was about 20 years. The median duration of co-infection was 24.1 years. Men accounted for 70% of the patients in the cohort and most patients were infected with HCV genotype 1 (57.9%). Median CD4⁺ cell count was 467 cells/mm³ and median HIV viral load was 3.7 log viral copies/ml, with 28.1% of patients having detectable amounts of virus. The median duration of antiretroviral treatment for HIV infection was about 10.3 years. We also split the patients in two groups according to METAVIR score F0F1F2 and F3F4 groups. Sex ratio, mean age, and duration of infection did not differ significantly between the two groups. Median CD4⁺ cell counts were significantly lower for the F3F4 group than for the F0F1F2 group, whereas there was no significant difference in HIV viral load between these two groups (Supplemental Table 1).

Associations following the comparison of the F0F1F2 and F3F4 groups

After the quality-control analysis and imputation procedures, 292 patients remained included in the study for a total of 8,426,597 genotyped or reliably imputed variants. The 197 individuals exhibiting a low METAVIR score (< F3) were compared with the 95 individuals exhibiting a high METAVIR score (\geq F3) in the dominant genetic model. The *p*-value distribution is shown in the Manhattan plot of Figure 1 and in the QQ plot of Supplemental Figure 3. The QQ plot gave a genomic inflation factor close to one ($\lambda = 1.078$), indicating an absence of stratification after adjustment for the first two principal components. The GWAS

identified 6 variants in two distinct loci showing a genome-wide significance level (p -values $< 5 \times 10^{-8}$), for which information is provided in Table 1. The first significant signal, for chromosome region 5p15, corresponded to a cluster of five variants (three SNPs and two indels, $r^2 > 0.9$ in 1000 Genomes European populations and HEPAVIH cohort, Supplemental Figure 4) tagged by two genotyped SNPs in total LD, rs2158099 and rs1423493, which are intronic variants of the *CTNND2* gene (betas = -1.67, standard errors (SE) = 0.29, p -values = 6.1×10^{-9}). Multivariate analysis confirmed that these association results within the cluster were supported by a single signal. The second signal was located on chromosome region 1q31 and corresponded to a single imputed SNP rs72727113 within an intergenic region (beta = 2.37, SE = 0.41, p -value = 8.2×10^{-9}). We also investigated the influence of the 25 patients who achieved a SVR after treatment on the results, by removing them from the analysis. Both signals remained significant, according to the genome-wide threshold (rs2158099, beta = -1.69, SE = 0.31, p -value = 2.96×10^{-8} , and rs72727113, beta = 2.37, SE = 0.41, p -value = 1.09×10^{-8}).

We have fixed as a principal association of the chromosome region 5p15 cluster the genotyped SNP rs2158099 for further representations. This C/G SNP had a minor allele frequency (MAF) of 0.49 (allele C) in the 292 HEPAVIH patients. A similar MAF was observed in the 1000 Genomes European population (MAF CEU = 0.49). The distribution of the SNP rs2158099 genotype according to the binary F0F1F2/F3F4 definition is shown in Figure 2. The homozygous CC patients ($n = 85$) in HEPAVIH cohort showed a greater susceptibility to severe fibrosis (METAVIR score $\geq F3$) with 57.65% of the sample size. Supplemental Table 2 provides demographic information according to the SNP rs2158099 genotypes. By performing the Fisher exact test in dominant model (allele C dominant), an Odd-Ratio (OR) of 0.194 [CI 95%: 0.103; 0.353] and a p -value = 3.5×10^{-9} were found. The

second signal carried the G/A SNP rs72727113 which had a MAF of 0.06, as observed in the 1000 Genomes European population (MAF CEU = 0.06). No homozygous AA patients was found in HEPAVIH cohort, and 25 heterozygous AG subjects (25/34) showed susceptibility to severe fibrosis (METAVIR score \geq F3). This is corresponding to an OR of 7.40 [CI 95%: 3.15; 18.94] and a p -value = 2.1×10^{-7} , obtained by performing the Fisher exact test in dominant model (allele G dominant).

We have also investigated 322 SNPs (p -value $< 10^{-5}$) previously published in 31 GWAS dealing with other liver disease, including HCV monoinfection (Supplemental Table 3) These SNPs corresponded to 131 genetic regions. None common regions could provide significance evidence in this study, according to Bonferroni correction (p -value $> 5 \times 10^{-3}$) and taking into account LD information.

Discussion

The major pathological consequence of HCV infection is progression to liver fibrosis and its potential complications. Several factors have clearly been shown to be associated with fibrosis progression. These factors include the duration of infection, age, being male, and alcohol consumption (37), which have been taken as covariates in our study. HIV/HCV co-infection also results in liver fibrosis progressing more rapidly, with a viral load for HCV twice that in patients with HCV mono-infection (37). Co-infected patients appear to be at high risk of developing insulin resistance or diabetes (38) and are more likely to develop hepatotoxicity due to HAART than patients infected with HIV alone. In the HEPAVIH cohort all patients received anti-HIV treatment, and the French HIV guidelines also recommended to treat HCV infection in these patients, irrespective to the CD4⁺ count, using the less hepatotoxic drugs. Fibrosis progression was also found to be more likely in patients with low CD4⁺ cell counts and high viral loads for HIV (37), we indeed observe that patients with a high METAVIR score display lower CD4⁺ cell counts. Interestingly, a recent study (39) has suggested that HIV mono-infection does not impact the fibrosis progression. The identification of genetic variants associated with a risk of developing liver fibrosis in this specific population of co-infected patients is therefore of major interest.

The most significant associations found in our GWAS are located in the *CTNND2* gene and include five intronic variants (two 'indel' and three SNPs). The *CTNND2* gene encodes for a δ -catenin, an adhesive junction associated protein of the armadillo/ β -catenin superfamily. The δ -catenin is important in brain and eye development (40) but also in cancer formation (41). Several splice variants were described, and an isoform and dose feedback of the gene expression is important in cellular morphogenesis, apoptosis, and cancer (42). However, no

influence of the five associations on splicing mechanism was predicted. Indeed, this gene is involved in the mechanisms of maintenance of adherence junctions which are responsible for cell-cell contact (43). The δ -catenin affects the localization and stability of p120-catenin (*CTNND1*) by competitively interacting with E-cadherin that plays an essential role in the formation of adherence junctions and in the function of epithelial cells (44). It has been reported that abnormal expression of E-cadherin and p120-catenin in intra-hepatic cholangiocarcinoma is correlated with tumor differentiation, intra-hepatic metastasis, and survival of patients (45). Additionally, *CTNND2* is expressed in vascular endothelium and the simple lack of one gene copy is sufficient to impair endothelial cell motility and vascular assembly *in vitro* and pathological angiogenesis *in vivo* (46). *CTNND2* has also been reported as deregulated in hepatocellular carcinoma (47, 48). Furthermore, δ -catenin has been reported to interact with β -catenin (49). This interaction is of particular interest in HCV induced hepatocellular carcinoma because β -catenin together with the E-cadherin plays a significant role in liver physiology and pathology through the Wnt/Fzd/ β -catenin signaling pathway (50) and they have been placed on the list of seric markers of liver carcinogenesis (50).

Regarding the second signal, LD investigation of the 1q31 region (Supplemental Figure 5) did not reveal any obvious link with a gene or biological function. Nevertheless, SNP rs72727113 is located within a reported region of copy number variants. Several gains or losses of copies that encompass the 1q31 locus have been described in specialized databases (51) with no obvious functional impact.

The 6 SNPs found in our study were not present in the GWAS catalog, which records all the genetic association *p*-values below 10^{-6} . Reciprocally, we have also looked if identified

regions in our study on liver fibrosis could be replicate in other GWAS dealing with liver fibrosis in HCV mono-infection and also in non-HCV-related liver diseases (see Methods). Unfortunately, none of the regions tested ($p < 10^{-5}$) has emerged replicated in our study taking account into LD information ($p < 5 \cdot 10^{-2}$) (Supplemental Table 3).

To date, no GWAS dealing with fibrosis in HIV/HCV co-infection has been published. Here, we have detected for the first time two significant signals associated with severe fibrosis defined by METAVIR scores in patients with HIV/HCV co-infection. Both signals were persistent with or without patients who achieved a SVR after treatment, suggesting mechanisms independent from a sustained virological response. The *CTNND2* gene seems to be involved in a spectrum of mechanisms that could be relevant to HCV-induced hepatocellular carcinoma but no obvious biological function could be derived from the second signal. As for any GWAS, it will be important to try and replicate these signals in other co-infection cohorts. Several HIV/HCV co-infection cohorts have been described (52) and these interesting results may trigger new genetic studies.

References

1. UNAIDS. How AIDS changed everything; 2015.
2. Sherman KE, Rouster SD, Chung RT, Rajicic N. Hepatitis C Virus prevalence among patients infected with Human Immunodeficiency Virus: a cross-sectional analysis of the US adult AIDS Clinical Trials Group. *Clin Infect Dis* 2002;34:831-837.
3. Cacoub P, Dabis F, Costagliola D, Almeida K, Lert F, Piroth L, Semaille C. Burden of HIV and hepatitis C co-infection: the changing epidemiology of hepatitis C in HIV-infected patients in France. *Liver Int* 2015;35:65-70.
4. Bica I, McGovern B, Dhar R, Stone D, McGowan K, Scheib R, Snyderman DR. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis* 2001;32:492-497.
5. Rosenthal E, Poiree M, Pradier C, Perronne C, Salmon-Ceron D, Geffray L, Myers RP, et al. Mortality due to hepatitis C-related liver disease in HIV-infected patients in France (Mortavic 2001 study). *AIDS* 2003;17:1803-1809.
6. Konerman MA, Mehta SH, Sutcliffe CG, Vu T, Higgins Y, Torbenson MS, Moore RD, et al. Fibrosis progression in human immunodeficiency virus/hepatitis C virus coinfecting adults: prospective analysis of 435 liver biopsy pairs. *Hepatology* 2014;59:767-775.
7. Sulkowski MS, Mehta SH, Torbenson MS, Higgins Y, Brinkley SC, de Oca RM, Moore RD, et al. Rapid fibrosis progression among HIV/hepatitis C virus-co-infected adults. *AIDS* 2007;21:2209-2216.
8. Benhamou Y, Di Martino V, Bochet M, Colombet G, Thibault V, Liou A, Katlama C, et al. Factors affecting liver fibrosis in human immunodeficiency virus-and hepatitis C virus-coinfected patients: impact of protease inhibitor therapy. *Hepatology* 2001;34:283-287.
9. Bani-Sadr F, Lapidus N, Bedossa P, De Boever CM, Perronne C, Halfon P, Pol S, et al. Progression of fibrosis in HIV and hepatitis C virus-coinfected patients treated with interferon plus ribavirin-based therapy: analysis of risk factors. *Clin Infect Dis* 2008;46:768-774.
10. Brau N, Salvatore M, Rios-Bedoya CF, Fernandez-Carbia A, Paronetto F, Rodriguez-Orengo JF, Rodriguez-Torres M. Slower fibrosis progression in HIV/HCV-coinfected patients with successful HIV suppression using antiretroviral therapy. *J Hepatol* 2006;44:47-55.
11. Soriano V, Martin-Carbonero L, Garcia-Samaniego J. Treatment of chronic hepatitis C virus infection: we must target the virus or liver fibrosis? *AIDS* 2003;17:751-753.
12. Rueger S, Bochud PY, Dufour JF, Mullhaupt B, Semela D, Heim MH, Moradpour D, et al. Impact of common risk factors of fibrosis progression in chronic hepatitis C. *Gut* 2015;64:1605-1615.
13. Lutz P, Wasmuth JC, Nischalke HD, Vidovic N, Grunhage F, Lammert F, Oldenburg J, et al. Progression of liver fibrosis in HIV/HCV genotype 1 co-infected patients is related to the T allele of the rs12979860 polymorphism of the IL28B gene. *Eur J Med Res* 2011;16:335-341.
14. Rau M, Baur K, Geier A. Host genetic variants in the pathogenesis of hepatitis C. *Viruses* 2012;4:3281-3302.
15. Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, Elledge SJ. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc Natl Acad Sci U S A* 2009;106:16410-16415.
16. Bataller R, Brenner DA. Liver fibrosis. *J Clin Invest* 2005;115:209-218.
17. Alcais A, Abel L, Casanova JL. Human genetics of infectious diseases: between proof of principle and paradigm. *J Clin Invest* 2009;119:2506-2514.
18. Limou S, Zagury JF. Immunogenetics: Genome-Wide Association of Non-Progressive HIV and Viral Load Control: HLA Genes and Beyond. *Front Immunol* 2013;4:118.

19. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461:399-401.
20. Pineda JA, Caruz A, Di Lello FA, Camacho A, Mesa P, Neukam K, Rivero-juarez A, et al. Low-density lipoprotein receptor genotyping enhances the predictive value of IL28B genotype in HIV/hepatitis C virus-coinfected patients. *AIDS* 2011;25:1415-1420.
21. Rizk NM, Derbala MF. Genetic polymorphisms of ICAM 1 and IL28 as predictors of liver fibrosis severity and viral clearance in hepatitis C genotype 4. *Clin Res Hepatol Gastroenterol* 2013;37:262-268.
22. Bochud PY, Bibert S, Kutalik Z, Patin E, Guergnon J, Nalpas B, Goossens N, et al. IL28B alleles associated with poor hepatitis C virus (HCV) clearance protect against inflammation and fibrosis in patients infected with non-1 HCV genotypes. *Hepatology* 2012;55:384-394.
23. Abe H, Ochi H, Maekawa T, Hayes CN, Tsuge M, Miki D, Mitsui F, et al. Common variation of IL28 affects gamma-GTP levels and inflammation of the liver in chronically infected hepatitis C virus patients. *J Hepatol* 2010;53:439-443.
24. Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, Hosono N, et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. *J Hepatol* 2013;58:875-882.
25. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, Otsuka M, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011;43:455-458.
26. Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, Aikata H, Ikeda K, et al. Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. *Nat Genet* 2011;43:797-800.
27. Lange CM, Bibert S, Dufour JF, Celleraï C, Cerny A, Heim MH, Kaiser L, et al. Comparative genetic analyses point to HCP5 as susceptibility locus for HCV-associated hepatocellular carcinoma. *J Hepatol* 2013;59:504-509.
28. Loko MA, Salmon D, Carrieri P, Winnock M, Mora M, Merchadou L, Gillet S, et al. The French national prospective cohort of patients co-infected with HIV and HCV (ANRS CO13 HEPAVIH): early findings, 2006-2010. *BMC Infect Dis* 2010;10:303.
29. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343-350.
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.
31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.
32. Delaneau O, Zagury JF. Haplotype inference. *Methods Mol Biol* 2012;888:177-196.
33. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5-6.
34. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073.
35. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906-913.
36. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012;8:e1002822.

37. Karageorgopoulos DE, Allen J, Bhagani S. Hepatitis C in human immunodeficiency virus co-infected individuals: Is this still a "special population"? *World J Hepatol* 2015;7:1936-1952.
38. Merchante N, Rivero A, de Los Santos-Gil I, Merino D, Marquez M, Lopez-Ruz MA, Rodriguez-Bano J, et al. Insulin resistance is associated with liver stiffness in HIV/HCV co-infected patients. *Gut* 2009;58:1654-1660.
39. Swanson S, Ma Y, Scherzer R, Huhn G, French AL, Plankey MW, Grunfeld C, et al. Association of HIV, Hepatitis C Virus, and Liver Fibrosis Severity With the Enhanced Liver Fibrosis Score. *J Infect Dis* 2016;213:1079-1086.
40. Israely I, Costa RM, Xie CW, Silva AJ, Kosik KS, Liu X. Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. *Curr Biol* 2004;14:1657-1663.
41. Lu Q. delta-Catenin dysregulation in cancer: interactions with E-cadherin and beyond. *J Pathol* 2010;222:119-123.
42. Zhang J, Lu JP, Suter DM, Krause KH, Fini ME, Chen B, Lu Q. Isoform- and dose-sensitive feedback interactions between paired box 6 gene and delta-catenin in cell differentiation and death. *Exp Cell Res* 2010;316:1070-1081.
43. Reynolds AB. p120-catenin: Past and present. *Biochim Biophys Acta* 2007;1773:2-7.
44. Yang I, Chang O, Lu Q, Kim K. Delta-catenin affects the localization and stability of p120-catenin by competitively interacting with E-cadherin. *Mol Cells* 2010;29:233-237.
45. Zhai B, Yan HX, Liu SQ, Chen L, Wu MC, Wang HY. Reduced expression of P120 catenin in cholangiocarcinoma correlated with tumor clinicopathologic parameters. *World J Gastroenterol* 2008;14:3739-3744.
46. DeBusk LM, Boelte K, Min Y, Lin PC. Heterozygous deficiency of delta-catenin impairs pathological angiogenesis. *J Exp Med* 2010;207:77-84.
47. Chen R, Dong Y, Xie X, Chen J, Gao D, Liu Y, Ren Z, et al. Screening candidate metastasis-associated genes in three-dimensional HCC spheroids with different metastasis potential. *Int J Clin Exp Pathol* 2014;7:2527-2535.
48. Wang HW, Hsieh TH, Huang SY, Chau GY, Tung CY, Su CW, Wu JC. Forfeited hepatogenesis program and increased embryonic stem cell traits in young hepatocellular carcinoma (HCC) comparing to elderly HCC. *BMC Genomics* 2013;14:736.
49. Lu Q, Paredes M, Medina M, Zhou J, Cavallo R, Peifer M, Orecchio L, et al. delta-catenin, an adhesive junction-associated protein which promotes cell scattering. *J Cell Biol* 1999;144:519-532.
50. Rogacki K, Kasprzak A, Stepinski A. Alterations of Wnt/beta-catenin signaling pathway in hepatocellular carcinomas associated with hepatitis C virus. *Pol J Pathol* 2015;66:9-21.
51. Hehir-Kwa JY, Pfundt R, Veltman JA, de Leeuw N. Pathogenic or not? Assessing the clinical relevance of copy number variants. *Clin Genet* 2013;84:415-421.
52. Peters L, Klein MB. Epidemiology of hepatitis C virus in HIV-infected patients. *Curr Opin HIV AIDS* 2015;10:297-302.

Tables

Table 1 – Associations from with the binary METAVIR F0F1F2/F3F4 phenotype in dominant mode such that p -value $< 5 \times 10^{-8}$, p -values adjusted with covariates (see Methods)

SNP	CHR	BP(HG19)	Alleles minor/Major	MAF	Beta (SE)	Adjusted p -value	MAF CEU 1KG	Genes
rs2158099 ^G	5p15	11362268	C/G	0.49	-1.67 (0.29)	6.05x10 ⁻⁹	0.49	<i>CTNND2</i>
rs35268064 ^G indel	5p15	11362322	TG/T	0.48	-1.64 (0.28)	8.29x10 ⁻⁹	0.50	<i>CTNND2</i>
rs11367210 indel	5p15	11362853	GT/G	0.48	-1.69 (0.29)	4.80x10 ⁻⁹	0.49	<i>CTNND2</i>
rs1423493 ^G	5p15	11363645	C/T	0.49	-1.67 (0.29)	6.05x10 ⁻⁹	0.49	<i>CTNND2</i>
rs2074260 ^G	5p15	11365166	C/T	0.48	-1.57 (0.28)	3.00x10 ⁻⁸	0.49	<i>CTNND2</i>
rs72727113	1q31	191555232	G/A	0.06	2.37 (0.41)	8.18x10 ⁻⁹	0.06	intergenic

^G: indicates that the variant is genotyped

Figures legends

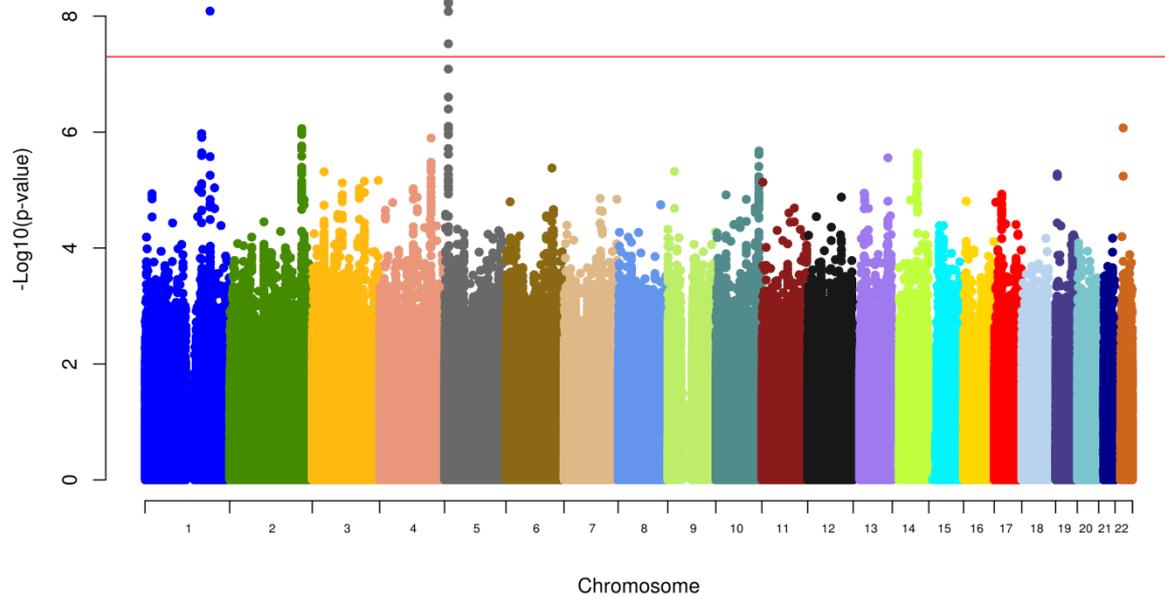


Figure 1 – Manhattan plot of the genome-wide association results by comparing F0F1F2 and F3F4 METAVIR score groups. After genotype imputation, ~8,500,000 common variants were tested for association by comparing 197 subjects with low METAVIR score ($< F3$) with 95 subjects with high METAVIR score ($\geq F3$), in logistic regression analysis with a dominant model. Each dot corresponds to a SNP, with its chromosomal position (x axis) and $-\log_{10}(p\text{-value})$ (y axis) as the coordinates. Association signals of genome-wide significance ($p\text{-value} < 5 \times 10^{-8}$, red line) were observed on chromosomes 1 and 5. The strongest associated locus were on chromosome 5p15 carried by the SNP rs2158099 ($p\text{-value} = 6.05 \times 10^{-9}$) and on chromosome 1q31 with the SNP rs72727113 ($p\text{-value} = 8.18 \times 10^{-9}$).

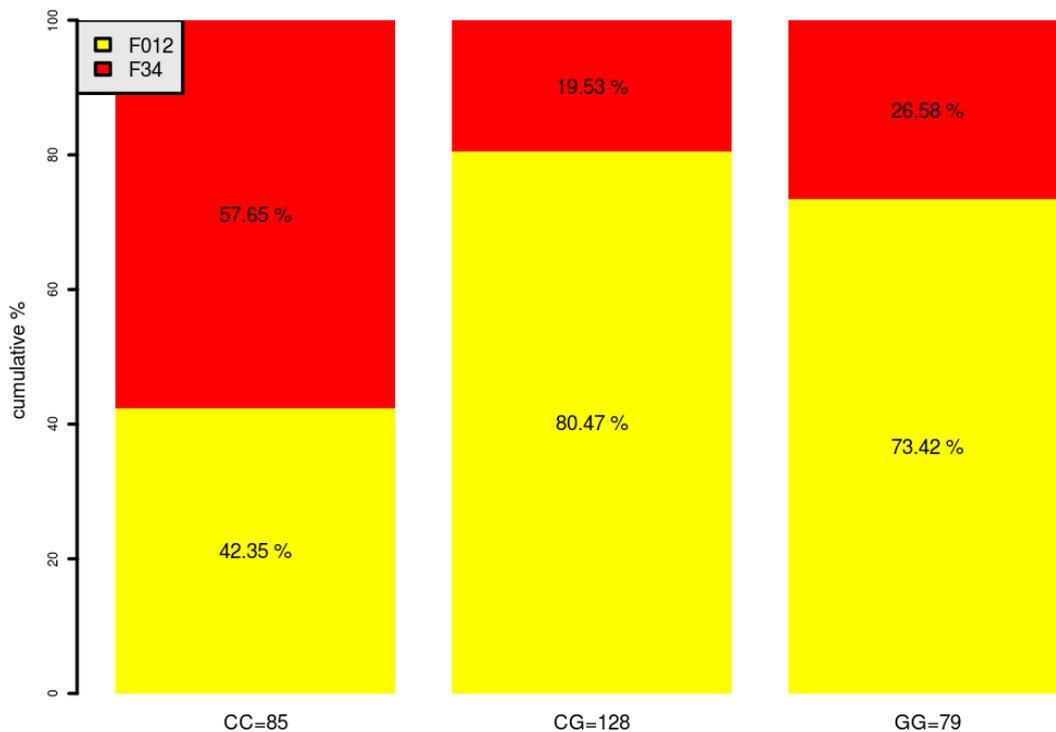


Figure 2 – Distribution by genotype for the SNP rs2158099 according to the F0F1F2 and F3F4 METAVIR score groups. For each genotype on the x-axis, the sample size is also provided, and METAVIR scores group is indicated on the y-axis, together with the percentage of the subjects in the corresponding group.

Supporting Information:

Separated file

Acknowledgments: The authors thank all the members of the ANRS CO13 HEPAVIH Study Group, especially the physicians and nurses involved in cohort follow-up, and all the patients who took part in this study.

Conflict of Interest: The authors declare no conflict of interest

Supplemental Information:

Members of the Scientific Committee of the ANRS CO13 HEPAVIH Study Group: D Salmon (Investigator coordinator), F Dabis (co-investigator coordinator), L Wittkop (scientific and methodological coordinator), MA Loko, M. Winnock, P Sogni, P Trimoulet, J Izopet, V Paradis, B Spire, P Carrieri, C Katlama, G Pialoux, MA Valantin, J Chas, I Poizot-Martin, B Marchou, E Rosenthal, D Garipuy, O Bouchaud, A Gervais, C Lascoux-Combe, C Goujard, K Lacombe, A Bicart-See, C Duvivier, D Vittecoq, D Neau, P Morlat, F Bani-Sadr, L Meyer, F Boufassa, L Piroth, ML Chaix, F Abravanel, B Autran, AM Roque, C Solas, H Fontaine, L Serfaty, M Guiguet, G Chêne, D Costagliola, D Zucman, A Simon, S Dominguez, E Billaud, P Miaillhes, J Polo Devoto, H Aumaître, D Rey, O Segeal, L Marchand (ANRS), S Couffin-Cadiergues (ANRS).

Supplemental Tables:

Supplemental Table 1 – Demographic statistics based on the date of patients inclusions in the cohort. Statistical differences are determined between F0F1F2 and F3F4 using the t-test (quantitatives variables) or the chi² test (qualitatives variables), included in the R software version 3.0.2 (<https://cran.r-project.org/>).

Characteristic	All patients (n = 292)	F0F1F2 (n = 197)	F3F4 (n = 95)	<i>p</i>
Sex (Male/Female) n (%)	202 (69.5) / 90 (30.5)	132 (67) / 65 (33)	71 (74.7) / 24 (25.3)	0.169
Age (years), median +/- SD	45.4 +/- 5.9	45.4 +/- 5.9	45.7 +/- 5.8	0.143
Infection time (years), median +/- SD	24.1 +/- 7.9	23.8 +/- 8.5	24.3 +/- 6.5	0.018
Alcohol status, n (%) (never/past/current)	50 (17.1) / 81 (27.7) / 161 (55.1)	36 (18.3) / 50 (25.4) / 111 (56.3)	14 (14.7) / 31 (32.6) / 50 (52.6)	0.398
HCV genotype 1 vs other, n (%)	169 (57.9) / 123 (42.1)	112 (56.9) / 85 (43.1)	57 (60) / 38 (40)	0.701
METAVIR score, n (%)				ND
0	11 (3.8)	11 (5.6)	0 (0)	
1	134 (45.9)	134 (68)	0 (0)	
2	52 (17.8)	52 (26.4)	0 (0)	
3	22 (7.5)	0 (0)	22 (23.2)	
4	73 (25)	0 (0)	73 (76.8)	
SVR after treatment (Yes/No) n (%)	25 (8.6) / 267 (91.4)	16 (8.1) / 181(91.9)	9 (9.5) / 86 (90.5)	0.870
CD4+ (cell/mm ³), median +/- SD	467 +/- 297.3	468 +/- 307	430 +/- 269.7	0.025
HIV viral load *	82 (28.1) / 210 (71.9) 3.7 +/- 1.9	57 (28.9) / 140 (71.1) 3.7 +/- 1.7	25 (26.3) / 70 (73.7) 3.7 +/- 2.3	0.303
ARVT ** (years), median +/- SD	10.3 +/- 4.6	10 +/- 4.6	10.6 +/- 4.6	0.264

* HIV viral load: HIV viral load status (detectable / undetectable, n (%)) and logarithm of HIV copy per ml (median +/- SD)

** ARVT: effective time of HIV anti-retroviral treatment

Supplemental Table 2 – Demographic statistic based on Supplemental Table 1 according to each genotype (CC/CG/GG) of the SNP rs2158099. *p* indicates the statistical difference from a t-test for quantitative variable or from a chi² test for qualitative variable.

	Characteristic	CC (n = 85)	CG (n = 128)	GG (n = 79)	<i>p</i>
Covariates used in the linear regression model	Sex (Male/Female) n (%)	64 (75.3) / 21 (24.7)	79 (61.7) / 49 (38.3)	60 (75.9) / 19 (24.1)	0.296
	Age (years), median +/- SD	46.4 +/- 6.3	44.6 +/- 5	45.3 +/- 6.6	0.928
	Infection time (years), median +/- SD	24.2 +/- 8.4	24.1 +/- 7.2	24 +/- 8.6	0.036
	HCV genotype 1 versus other	41 (48.2) / 44 (51.8)	81 (63.3) / 47 (36.7)	47 (59.5) / 32 (40.5)	0.137
	Alcohol status, n (%) (never/past/current)	14 (16.5) / 25 (29.4) / 46 (54.1)	20 (15.6) / 33 (25.8) / 75 (58.6)	16 (20.3) / 23 (29.1) / 40 (50.6)	0.976
	ARVT * (years), median +/- SD	10.2 +/- 5.1	9.8 +/- 4.5	10.5 +/- 4	0.270
	SVR after treatment (Yes/No) n (%)	10 (11.8) / 75 (88.2)	8 (6.2) / 120 (93.8)	7 (8.9) / 72 (91.1)	ND
	CD4+ (cell/mm ³), median +/- SD	475 +/- 258.6	468.5 +/- 309	441 +/- 315.4	ND
	HIV viral load **	3.7 +/- 2 24 (28.2) / 61 (71.8)	3.7 +/- 1.9 33 (25.8) / 95 (74.2)	3.7 +/- 1.9 25 (31.6) / 54 (68.4)	ND

* ARVT: effective time of HIV anti-retroviral treatment

** HIV viral load: status (detectable / undetectable, n (%)) and logarithm of HIV copy per ml (median +/- SD)

Supplemental Table 3 – GWAS associations dealing with liver disease

- PSC = primary sclerosing cholangitis

First author	journal	pmid	SNPs	P<=10-5	region	Reported genes
Ellinghaus D	Hepatology	22821403	rs4676410	2,00E-009	2q37	GPR35
			rs1452787	3,00E-008	18q21	TCF4
Mellum E	Nat Genet	21151127	rs3197999	1,00E-016	3p21	MST1
			rs6720394	4,00E-008	2q13	BCL2L11
			rs12735793	1,60E-006	-	PTPRU
			rs896403	1,50E-006	-	1q24
			rs4391646	8,80E-007	-	1q24
			rs6678400	9,50E-007	-	1q24
			rs12043426	3,50E-006	-	1q42
			rs6971637	1,60E-007	-	PION
			rs7791854	1,00E-007	-	PION
			rs3807746	2,10E-006	-	MAGI2
			rs6973565	1,40E-006	-	PLXNA4
			rs13231950	4,30E-006	-	PLXNA4
			rs7462577	1,90E-006	-	8q24
			rs7038037	2,70E-006	-	BSPRY
			rs706778	5,40E-006	-	IL2RA
			rs3134883	8,50E-007	-	IL2RA
			rs4147359	4,70E-006	-	IL2RA
			rs10905718	2,40E-007	-	IL2RA
			rs10891130	5,00E-006	-	FDX1
			rs1793660	3,60E-006	-	KIRREL3
			rs12808353	2,40E-006	-	VPS26B
			rs11168249	1,00E-006	-	HDAC7
			rs9520835	9,00E-006	-	TNFSF13B
rs4321167	5,90E-006	-	15q26			
rs17683107	8,50E-007	-	GAS7			
rs12458015	7,30E-006	-	TCF4			
Karlsen TH	Gastroenterology	19944697	rs3099844	2,60E-026	6p21	HLA-B
			rs2844559	4,20E-026	6p21	HLA-B
Ellinghaus D	Nat Genet	26974007	rs2234161	1,53E-006	1	TNFRSF14
			rs7608910	7,21E-006	2	PUS10
			rs7426056	4,36E-022	2	CD28
			rs11306716	2,08E-006	2	NA
			rs11676348	6,48E-007	2	CXCR2
			rs7556897	4,74E-009	2	AC073065.3
			rs3749171	3,67E-009	2	GPR35
			rs3197999	1,84E-028	3	MST1
			rs116046827	4,50E-011	3	BSN

			rs3774937	6,11E-009	4	NFKB1
			rs13132308	2,53E-009	4	IL21
			rs9687958	1,28E-006	5	NA
			rs72928038	8,05E-012	6	BACH2
			rs61839660	1,82E-010	10	IL2RA
			rs3118471	1,70E-011	10	IL2RA
			rs10748781	4,37E-006	10	NKX2-3
			rs11168249	5,49E-009	12	HDAC7
			rs12369214	1,29E-009	12	RIC8B
			rs3184504	1,73E-011	12	SH2B3
			rs11649613	1,28E-011	16	RP11-396B14.2
			rs66504140	1,14E-006	18	CD226
			rs259964	2,78E-006	20	ZNF831
			rs9977672	4,90E-015	21	NA
			rs1893592	5,93E-008	21	UBASH3A
			rs4456788	5,05E-006	21	AP001057.1
Lui JZ	Nat Genet	23603763	rs3748816	7,40E-012	1p36	MMEL1
			rs7426056	1,90E-020	2q33	CD28
			rs317999	2,40E-026	3p21	MST1
			rs13140464	8,90E-013	4q27	IL21
			rs56258221	8,40E-012	6q15	BACH2
			rs4147359	8,20E-017	10p15	IL2RA
			rs7937682	3,20E-009	11q23	SIK2
			rs11168249	5,50E-009	12q13	HDAC7
			rs3118504	5,90E-011	12q24	SH2B3
			rs1788097	3,10E-008	18q22	CD226
			rs60652743	6,50E-010	19q13	PRKD2
			rs2836883	3,20E-017	21q22	PSMG1
Folseraas T	J Hepatol	22521342	rs3748816	2,10E-008	1	TNFRSF14
			rs13132245	1,20E-007	4	IL21
			rs13119723	9,90E-008	4	IL21
			rs2903692	6,40E-006	16	CLEC16A
			rs602662	1,90E-006	19	FUT2
			rs601338	2,70E-006	19	FUT2
			rs11682163	7,60E-006	2	ALLC
			rs281377	2,10E-006	19	FUT2

- **Autoimmune Hepatitis**

First author	journal	pmid	SNPs	P<=10-5	region	Reported genes
deBoer YS	Gastroenterology	24768677	rs2187668	1,50E-078	6p21	(HLA-DQA1)
			rs3184504	7,70E-008	12q24	(SH2B3)
			rs6000782	3,00E-006	22q13	(CARD10)

- **PBC = primary biliary cirrhosis**

First author	journal	pmid	SNPs	P <= 10 ⁻⁵	region	Reported genes
Cordell HJ	Nat Commun.	26394269	rs12712133	5,19E-009	2q12	IL1R1
			rs4973341	2,34E-010	2q36	RNA5SP121
			rs11724804	9,01E-012	4p16	GAK
			rs526231	1,14E-008	5q21	PAM
			rs2546890	1,06E-010	5q33	RNF145
			rs6933404	1,27E-010	6q23	LOC102723649
			rs1859308	0,000000163	16p11	IL4R
Hirschfield GM	N Engl J Med	19458352	rs4679904	0,000001	3q25	ARF7
			rs6441286	2,00E-014	3q25	IL12A
			rs3790567	3,00E-011	1p31	IL12RB2
			rs907092	0,000008	17q12	IKZF3
			rs10488631	0,0000002	7q32	IRF5-TNPO3
		rs2395148, rs3135363, rs2856683, rs9357152	1,00E-010	6p21	c6orf10	
		rs2395148, rs3135363, rs2856683, rs9357152	7,00E-010	6p21	c6orf10	
Liu X	Nat Genet	20639880	rs10488631	3,00E-010	7q32	TNPO3
			rs9303277	0,000000002	17q21	GSDMB
			rs3745516	8,00E-011	19q13	SPIB
			rs3790567	8,00E-012	1p31	IL12RB2
			rs6441286	8,00E-012	3q25	IL12A
			rs7774434	3,00E-026	6p21	HLA-DQB1
Mells GF	Nat Genet	21399635	rs12134279	2,00E-014	1q31	DENND1B
			rs10931468	2,00E-019	2q32	STAT4
			rs2293370	3,00E-011	3q13	CD80
			rs7665090	4,00E-012	4q24	NFKB1
			rs860413	1,00E-011	5p13	IL7R
			rs6974491	0,00000004	7p14	intergenic
			rs6421571	3,00E-012	11q23	CXCR5
			rs1800693	0,000000002	12p13	TNFRSF1A
			rs911263	2,00E-011	14q24	RAD51L1
			rs12924729	3,00E-012	16p13	CLEC16A
			rs11117432	5,00E-011	16q24	intergenic
			rs968451	0,000000001	22q13	MAP3K7IP1
			rs17129789	9,00E-020	1p31	IL12RB2
			rs485499	2,00E-016	3q25	IL12A
			rs12531711	9,00E-017	7q32	IRF5

			rs7208487	0,0000008	17q12	ORMDL3
			rs3745516	2,00E-013	19q13	SPIB
			rs1372072	0,00000002	3p24	PLCL2
			rs538147	2,00E-010	11q13	RPS6KA4
			rs8017161	3,00E-013	14q32	TNFAIP2
			rs7774434	4,00E-034	6p21	MHC
Nakamura M	Am J Hum Genet	23000144	rs6890853	0,00000004	5p13	IL7R
			rs9303277	0,000000004	17q21	GSDMB
				0,000001	2q32	STAT4
			rs2293370	0,000000003	3q13	CD80
			rs7665590	0,0000001	4q23	NFKB1
			rs4979462	3,00E-014	9q32	TNFSF15
			rs4938534	0,00000002	11q23	POU2AF1

- **ALD = alcoholic liver disease / NAFLD = non-alcoholic fatty liver disease**

First author	journal	pmid	SNPs	P <= 10-5	region	Reported genes
Romeo S	Nat Genet	18820647	rs738409	5,90E-010	22q13	PNPLA3
Yuan X	Am J Hum Genet	18940312	rs11597390	1,50E-008	10	<i>CPNI</i>
			rs11597086	1,80E-008	10	<i>CHUK</i>
			rs11591741	3,00E-008	10	<i>CHUK</i>
			rs2281135	8,40E-016	22	<i>PNPLA3</i>
			rs2143571	7,20E-009	22	<i>SAMM50</i>
			rs1169313	1,80E-010	12	<i>HNF1A</i>
			rs4820599	4,00E-011	22	<i>GGT1</i>
			rs1780324	7,00E-015	1	<i>NBPF3-ALPL</i>
			rs9461011	1,90E-010	6	<i>GPLD1</i>
			rs9467160	1,20E-011	6	<i>GPLD1</i>
			rs8176720	4,30E-012	9	<i>ABO</i>
			rs641959	2,10E-011	9	<i>ABO</i>
			rs514708	1,90E-011	9	<i>ABO</i>
			rs672316	4,40E-010	9	<i>ABO</i>
			rs657152	1,70E-030	9	<i>ABO</i>
			rs474279	8,30E-009	9	<i>ABO</i>
			rs552148	1,20E-009	9	<i>ABO</i>
			rs12355784	5,00E-010	10	<i>JMJD1C</i>
			rs10761779	6,90E-010	10	<i>REEP3</i>
Chambers JC	Nat Genet	22001757	rs6834314	3,10E-009	4q22	<i>HSD17B13</i>
			rs2954021	5,30E-009	8q24	<i>TRIB1</i>
			rs10883437	4,00E-009	10q24	<i>CPNI</i>
			rs738409	1,20E-045	22q13	<i>PNPLA3</i>
			rs1976403	1,80E-050	1p36	<i>ALPL</i>
			rs16856332	1,60E-009	2q24	<i>ABCB11</i>

			rs1883415	5,60E-026	6p22	<i>ALDH5A1</i>
			rs6984305	2,10E-010	8p23	<i>PPP1R3B</i>
			rs2954021	2,30E-013	8q24	<i>TRIB1</i>
			rs10819937	1,00E-009	9q21	<i>ALDOB</i>
			rs579459	2,60E-123	9q34	<i>ABO</i>
			rs7923609	5,90E-023	10q21	<i>JMJD1C</i>
			rs174601	2,60E-009	11q12	<i>C11orf10</i>
			rs2236653	1,80E-009	11q.24	<i>ST3GAL4</i>
			rs7186908	4,80E-009	16q22	<i>HPR</i>
			rs314253	8,40E-012	17p13	<i>ASGR1</i>
			rs281377	1,10E-015	19q13	<i>FUT2</i>
			rs7267979	7,40E-010	20p11	<i>ABHD12</i>
			rs1497406	2,80E-019	1p36	<i>RSG1</i>
			rs12145922	3,80E-011	1p22	<i>CCBL2</i>
			rs1335645	7,30E-009	1p13	<i>CEPT1</i>
			rs10908458	1,70E-015	1q21	<i>DPM3</i>
			rs1260326	3,90E-013	2p23	<i>C2orf16</i>
			rs13030978	1,10E-011	2q12	<i>MYO1B</i>
			rs2140773	1,10E-009	2q37	<i>EFHD1</i>
			rs10513686	6,10E-011	3q26	<i>SLC2A2</i>
			rs4547811	2,50E-027	4q31	<i>ZNF827</i>
			rs6888304	1,20E-009	5p13	<i>CDH6</i>
			rs4074793	3,40E-010	5q11	<i>ITGA1</i>
			rs9296736	2,60E-009	6p12	<i>MLIP</i>
			rs17145750	2,90E-009	7q11	<i>MLXIPL</i>
			rs754466	6,40E-010	10q23	<i>DLG5</i>
			rs7310409	7,00E-045	12q24	<i>HNF1A</i>
			rs944002	5,80E-029	14q32	<i>C14orf73</i>
			rs339969	6,60E-020	15q21	<i>RORA</i>
			rs8038465	1,40E-009	15q23	<i>CD276</i>
			rs4581712	3,10E-009	16q23	<i>DYNLRB2</i>
			rs9913711	1,30E-009	17q24	<i>FLJ37644</i>
			rs12968116	8,90E-010	18q21	<i>ATP8B1</i>
			rs4503880	3,00E-012	18q21	<i>NEDD4L</i>
			rs516246	7,60E-010	19q13	<i>FUT2</i>
			rs1076540	9,60E-017	22q11	<i>MICAL3</i>
			rs2739330	1,70E-009	22q11	<i>DDT</i>
			rs2073398	1,10E-109	22q11	<i>GGTI</i>
Feitosa MF	Atherosclerosis	23477746	rs2862954	4,88E-010	10q24	<i>ERLIN1</i>
			rs1408579	4,57E-010	10q24	<i>ERLIN1</i>
			rs10883451	4,29E-010	10q24	<i>ERLIN1</i>
			rs11597086	1,46E-009	10q24	<i>CHUK</i>
			rs11591741	1,84E-009	10q24	<i>CHUK</i>
			rs17729876	1,82E-009	10q24	<i>CWF19L1</i>
			rs17668255	1,82E-009	10q24	<i>CWF19L1</i>
			rs17668357	1,65E-009	10q24	<i>CWF19L1</i>

			rs12784396	2,47E-09	10q24	CWF19L1
Butch S	Nat Genet	26482880	rs738409	1,54E-048	22	PNPLA3
			rs10401969	7,89E-010	19	TM6SF2
			rs58542926	1,33E-009	19	TM6SF2
			rs626283	1,03E-009	19	MBOAT7
Adams LA	Hepatology	23213074	rs12743824	5,00E-006	1p21	LPPR4
			rs222054	1,00E-006	4q13	GC
			rs7324845	3,00E-006	13q14	LCPI
			rs11864146	2,00E-006	16q23	SLC38
Chalasani N	Gastroenterology	20708005	rs343064	3,00E-008	7p14	intergenic
			rs11669592	4,00E-006	19p13	intergenic
			rs4243849	2,00E-006	8q24	intergenic
			rs2358462	4,00E-006	4q31	intergenic
			rs9977253	5,00E-006	21q21	intergenic
			rs6027755	6,00E-006	20q13	intergenic
			rs6079395	7,00E-006	20p12	MACROD2
			rs2800	3,00E-006	3q24	SLC9A9
			rs894177	3,00E-006	3q24	intergenic
			rs1305088	9,00E-006	13q12	SLC46A3
			rs1227756	2,00E-007	10q22	COL13A1
			rs887304	8,00E-007	12p13	EFCAB4B
			rs9584805	2,00E-006	13q32	FARP1
			rs4237591	2,00E-006	11q22	intergenic
			rs2986971	3,00E-006	10p11	intergenic
			rs2216228	4,00E-006	12p12	ST8SIA1
			rs7632299	4,00E-006	3q24	intergenic
			rs10067427	5,00E-006	5q21	intergenic
			rs10954668	7,00E-006	7q21	CACNA2D1
			rs7077164	7,00E-006	10q22	COL13A1
			rs11083271	8,00E-006	18q12	intergenic
			rs2499604	2,00E-006	1q43	ZP4
			rs2710833	6,00E-007	4q32	DDX60L
			rs2049953	4,00E-006	6q23	intergenic
			rs12344488	5,00E-006	9q21	intergenic
			rs1529093	2,00E-006	2q31	intergenic
			rs959903	7,00E-006	4p15	SEL1L3
			rs2645424	7,00E-007	8p23	FDFT1
			rs1836127	3,00E-006	12q22	intergenic
			rs643608	1,00E-006	21q22	intergenic
			rs6591182	9,00E-007	11q13	LTBP3
			rs2645424	3,00E-006	8p23	FDFT1
			rs6487679	1,00E-006	12p13	PZP
			rs1414896	2,00E-006	1p21	intergenic
			rs10510146	7,00E-006	10q26	C10orf122

Kawaguchi T	PLoS One	22719876	rs738409	1,00E-010	22q13	PNPLA3
			rs11206226	5,00E-006	1p32	YIPF1
			rs1390096	3,00E-006	4p15	HS3ST1
			rs2143571	6,00E-007	22q13	SAMM50
			rs2073080	8,00E-007	22q13	PARVB
Kitamoto T	Hum Genet	23535911	rs6691847	7,00E-006	1p35	PTPRU
			rs7552722	7,00E-006	1p13	NGF
			rs2051090	7,00E-006	13q13	DCLK1
			rs2896019	2,00E-020	22q13	PARVB
Speliotes EK	PLoS Genet	21423719	rs738409	4,00E-034	22q13	PNPLA3
			rs2228603	1,22E-011	19p12	NCAN
			rs12137855	2,59E-008	1	GCKR
Kozlitina	Nat Genet	24531328	rs738409	4,00E-016	22q13	PNPLA3
			rs2281135	6,90E-012	22q13	PNPLA3
			rs58542926	5,70E-008	19p13	TM6SF2
DiStefano JK	Acta Diabetol	25246029	rs10401969	3,10E-007	19p13	SUGP1
			rs4823173	2,80E-007	22q13	PNPLA3
			rs2896019	3,10E-007	22q13	PNPLA3
			rs2281135	6,00E-007	22q13	PNPLA3

- **Hepatitis B liver disease**

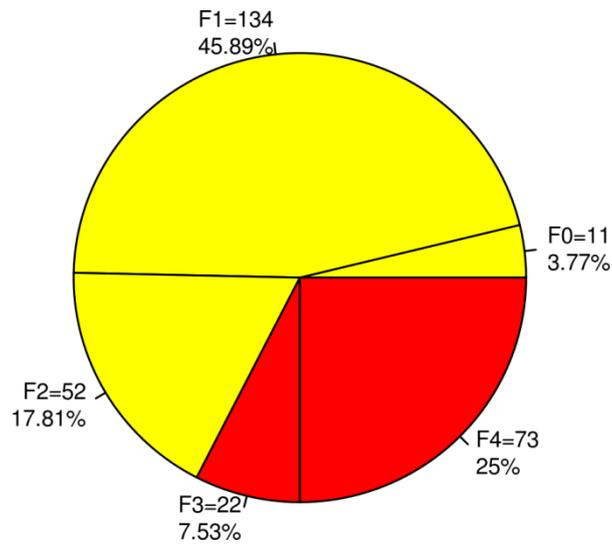
First author	journal	pmid	SNPs	P <= 10 ⁻⁵	region	Reported genes
Jiang DK	Nat Genet	23242368	rs7574865	2,00E-010	2q32	STAT4
			rs9275319	3,00E-017	6p21	HLA-DQ
Zhang H	Nat Genet	20676096	rs17401966	2,00E-018	1p36	UBE4B
			rs7898005	7,00E-008	10q24	intergenic
			rs2396470	7,00E-007	2q37	SLC19A3
Li S	PLoS Genet	22807686	rs455804	5,00E-010	21q21	GRIK1
			rs9272105	5,00E-022	6p21	HLA-DRB1
			rs7424161	8,76E-006	2q22	LRP1B
			rs3825023	3,14E-006	11p15	SAA2
			rs10896464	6,45E-006	11q13	MYEOV/CCND1
			rs3092194	4,41E-006	20q12	CHD6
Al-Qahtani A	J Med Genet	24065354	rs7553612	4,33E-006	1p31	LOC100289178
			rs6716352	9,82E-006	2q22	ARHGAP15
			rs34246867	2,54E-006	3q23	MRPS22
			rs35312459	4,67E-006	4q12	EXOC1
			rs6867291	7,81E-006	5q23	intergenic
			rs6867291	8,02E-006	5q23	intergenic

			rs502071	7,68E-006	6q22	HSF2
			rs4445240	5,93E-006	8q21	CNBD1
			rs12784800	3,71E-006	10q22	PPIF
			rs2724432	1,89E-007	11q22	FDX1
			rs17033545	5,97E-006	12q23	RP11-328J6.1
			rs1778797	6,98E-006	13q12	AL390738.1
			rs7998378	8,20E-006	13q22	KLF5
			rs8054283	8,93E-006	16q23	CDH13
			rs753610	6,45E-006	18q22	LOC643542
			rs17766247	3,78E-006	18q22	LOC643542
			rs5761699	3,98E-006	22q12	CTA-211A9.5
			rs5761699	7,64E-006	22q12	CTA-211A9.5
			rs5930539	1,49E-008	X	MST4
			rs5930539	4,26E-007	X	MST4
			rs6637934	1,80E-007	X	FRMD7
			rs6637934	3,70E-006	X	FRMD7
			rs16989995	7,93E-006	X	DMD
			rs12014086	4,06E-006	X	EFHC2
			rs17347621	2,31E-006	X	ACOT9
Chan KY	PLoS One	22174901	rs12682266	6,69E-006	8	intergenic
			rs7821974	7,01E-006	8	intergenic
			rs2275959	6,35E-006	8	intergenic
			rs1573266	7,44E-006	8	intergenic
			rs2611145	9,31E-006	11p13	ABTD2

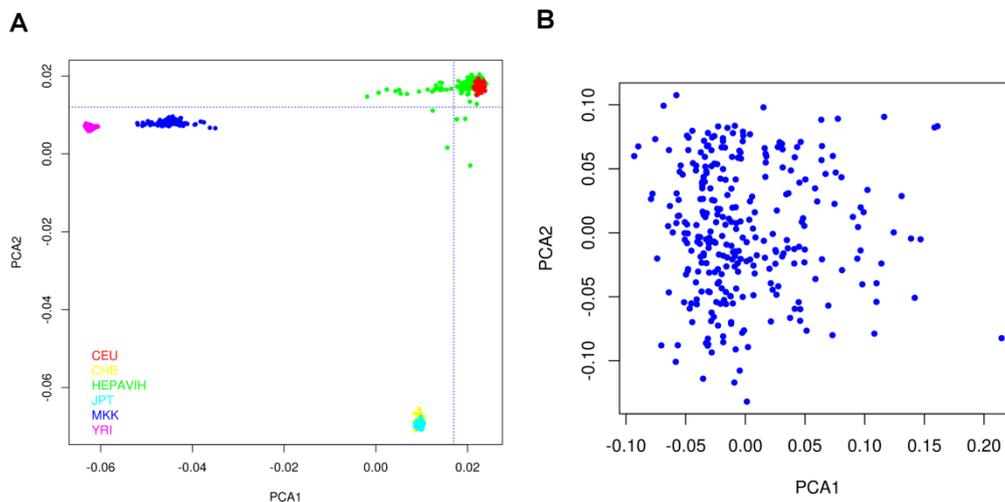
- **Hepatitis C liver disease**

First author	journal	pmid	SNPs	P <= 10 ⁻⁵	region	Reported genes
Patin E	Gastroenterology	22841784	rs16851720	9,00E-009	3q23	RNF7
			rs4374383	1,00E-009	2q13	MERTK
			rs9380516	5,00E-007	6p21	TULP1
			rs2629751	1,00E-007	12q23	GLT8D2
			rs883924	2,00E-006	9q22	LOC340515
			rs7800244	3,00E-006	7p12	PKD1L1
Urabe Y	J Hepatol	23321320	rs910049	9,00E-011	6p21	C6orf10
			rs3135363	1,00E-010	6p21	intergenic
			rs3129860	1,00E-009	6p21	intergenic
			rs3817963	1,00E-008	6p21	BTNL2
			rs9405098	1,00E-010	6p21	intergenic

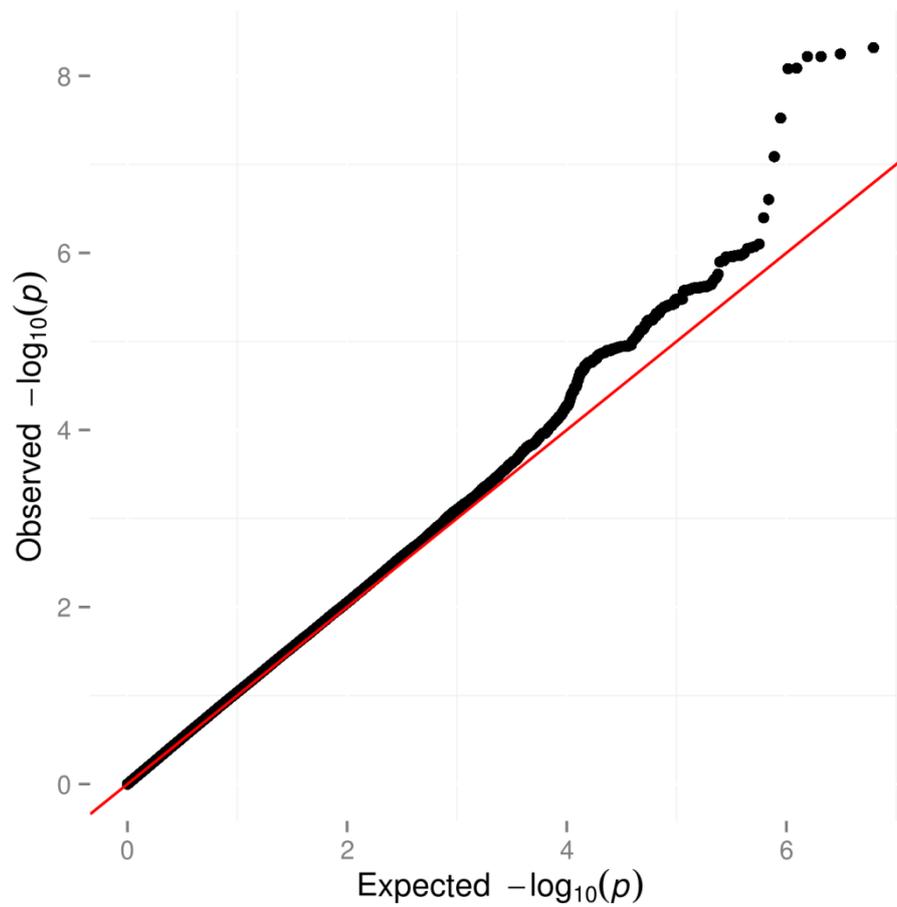
Supplemental Figures:



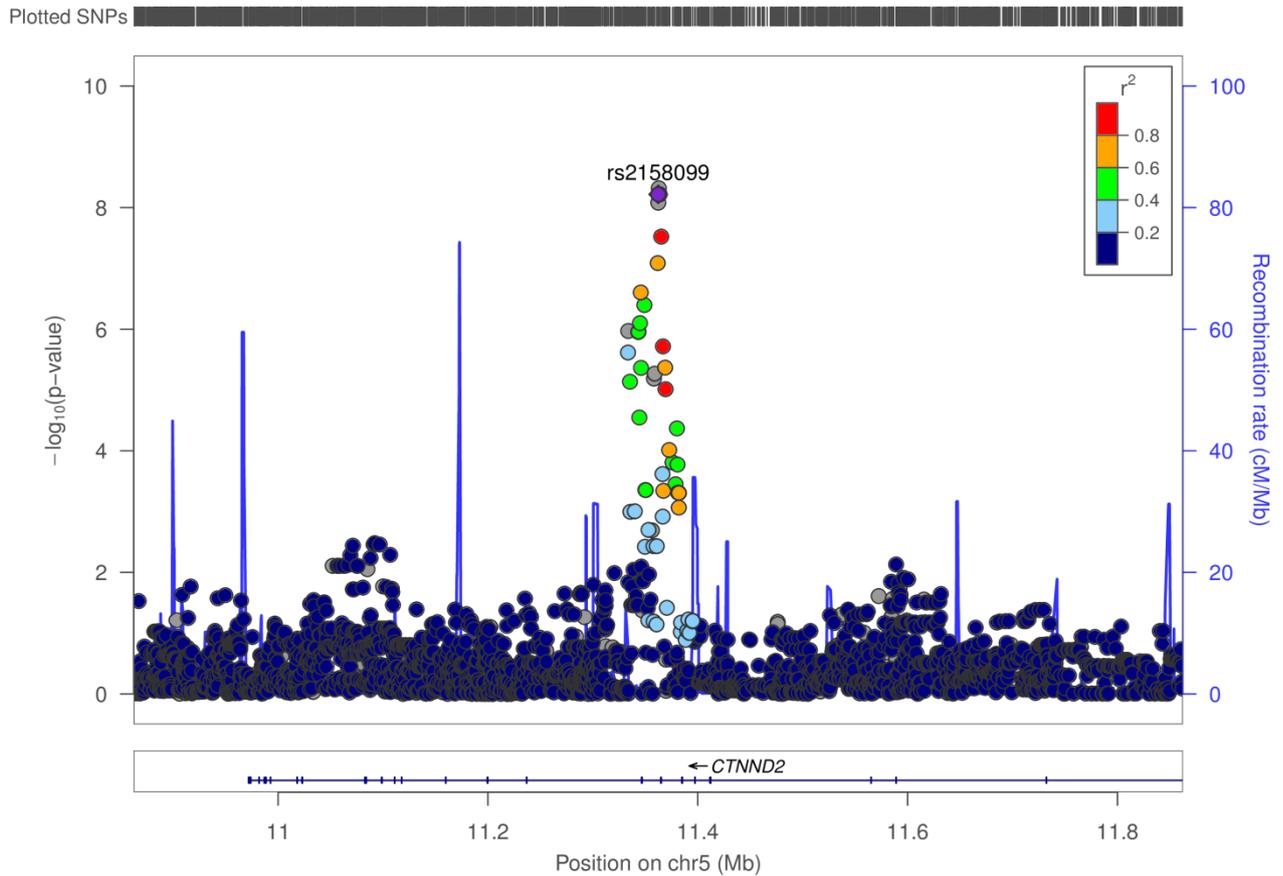
Supplemental Figure 1 – Distribution of METAVIR scores in the HEPAVIH sample used in analysis. The binary definition according to the breakpoint METAVIR F3 score is represented by the yellow and red colors (yellow : $<F3$, $n=197$, red : $\geq F3$, $n=95$)



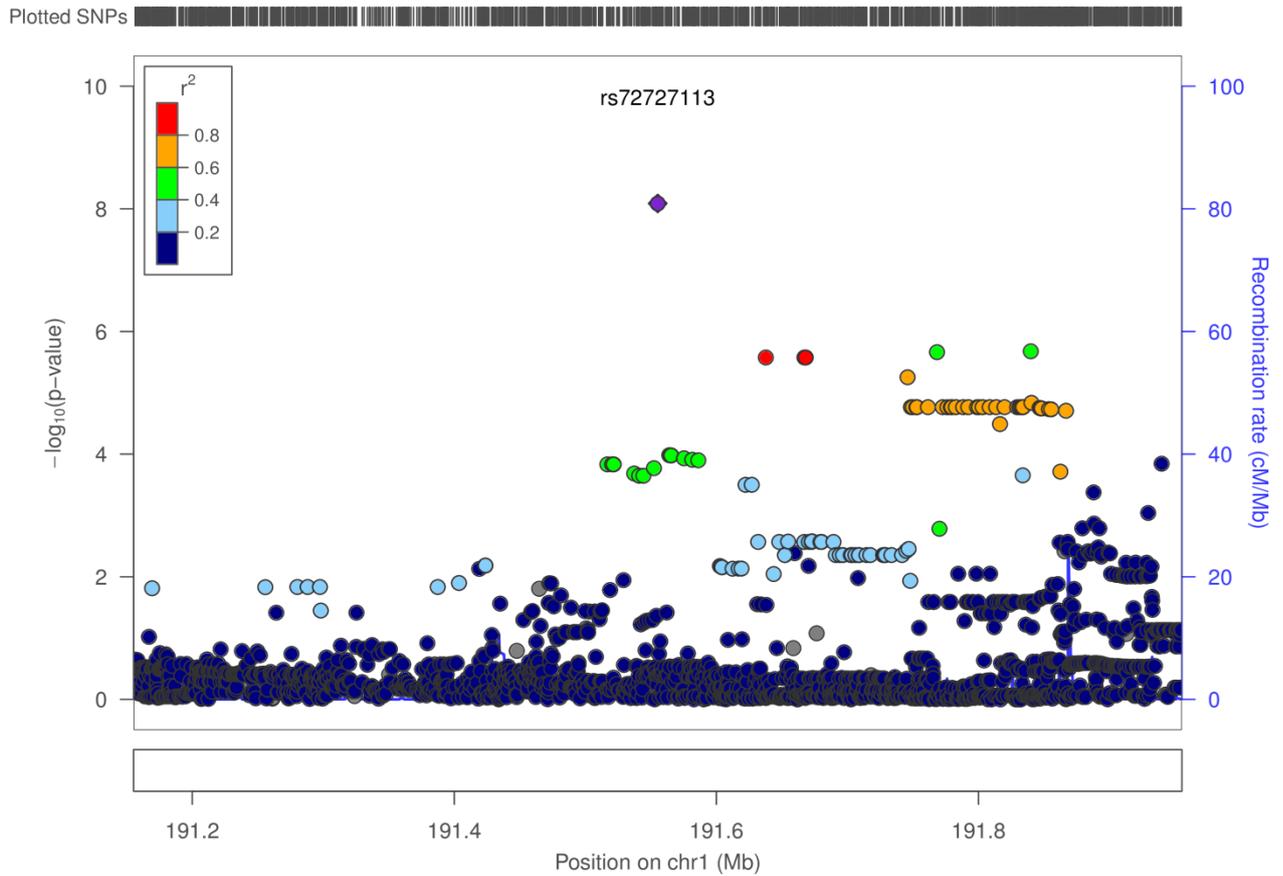
Supplemental Figure 2 – Principal component analysis (PCA) plots. (A) PCA of HEPAVIH and HAPMAP individuals. Each point represents a patient. HEPAVIH individuals are in green. **(B) PCA of HEPAVIH individuals after removing of non-European samples.** Stratification of the 292 individuals include in the analysis.



Supplemental Figure 3 – Quantile–quantile (QQ) plot. The plot displays the observed (y-axis) versus expected (x-axis) $-\log_{10}(p\text{-value})$. The lambda value is the genomic inflation factor which shows no inflation of the test statistic (lambda = 1.078).



Supplemental Figure 4 – 5p15 associations regional plot for binary F0F1F2/F3F4. Locus Zoom plots for the *CTNND2* gene region, with the most significant genotyped SNP rs2158099 (in purple). Vertical axis is $-\log_{10}$ of the p -value, the horizontal axis is the chromosomal position. Each dot represents a SNP tested for association by comparing F0F1F2 and F3F4 METAVIR score groups, in dominant model. Approximate linkage disequilibrium, between the most significant SNP and the other SNPs in the plot, is shown by the r^2 legend in top left of the plot. SNPs with missing LD information are shown in *grey*.



Supplemental Figure 5 – 1q31 associations regional plot for binary F0F1F2/F3F4. Locus Zoom plots for the surrounding region of the most significant SNP rs72727113 (in purple). Vertical axis is $-\log_{10}$ of the p -value, the horizontal axis is the chromosomal position. Each dot represents a SNP tested for association by comparing F0F1F2 and F3F4 METAVIR score groups, in dominant model. Approximate linkage disequilibrium, between the most significant SNP and the other SNPs in the plot, is shown by the r^2 legend in top left of the plot. SNPs with missing LD information are shown in *grey*.

3 RESULTATS ADDITIONNELS

3.1 Etude d'épistasie

Suite à l'association répliquée pour le SNP rs61183828 lié à la sévérité de la fibrose dans la coinfection VIH/VHC, j'ai réalisé une étude d'épistasie entre ce SNP et des variants connus pour être associés avec la charge virale lors de l'infection VIH ou la progression vers le SIDA. Cette analyse avait pour but de déterminer si des interactions entre deux SNPs pouvaient être impliquées dans le phénomène d'accélération de la fibrose observé lors de la coinfection VIH/VHC par comparaison avec la monoinfection VHC. Les variants liés au VIH ont été choisis à partir de la revue récente faite par O'Brien *et al.* datant de 2013 (81) (Tableau 2). L'analyse d'épistasie a été effectuée grâce au logiciel interSNP (www.intersnp.meb.uni-bonn.de).

Tableau 2. Listes des gènes et des SNPs associés dans le SIDA

Gènes liés au VIH	SNPs associés	SNPs en LD*
CCR5	rs333	84
CCR2	rs1799987	2
CCL5	rs2280789	101
HLA-C	rs9264942	8**
HCP5	rs2395029	49
ZNRD1	rs9261174	58
CXCR6	rs2234358	12
NOTCH4	rs8192591	16
MICA	rs4418214	53**

*SNPs en LD > 0,8 avec les SNPs associés au VIH testés dans nos études

**Recherche avec LD > 0,5 car aucun SNP n'était récupéré avec LD > 0,8

Deux effets épistatiques ont été observés entre le SNP d'intérêt rs61183828 et des SNPs des gènes *CCR5* et *CCL5*, dans un modèle additif d'interaction prenant en compte les covariables des 289 patients analysés dans l'étude « génome entier » portant sur le phénotype sqrtLS. Ces effets passaient le seuil de significativité après correction de Bonferroni sur le nombre de SNP

indépendants testés ($p < 1,8 \times 10^{-3}$), pour *CCR5* le SNP rs3092960, $p = 7,2 \times 10^{-4}$ et pour *CCL5* le SNP rs11653282 $p = 6,1 \times 10^{-4}$, les plus significatifs. Afin de voir si ces effets étaient liés au rôle supposé de ces gènes dans l'infection VIH, nous avons ajouté la charge virale et le compte de $CD4^+$ dans le modèle d'interaction en s'attendant à observer une diminution du signal. L'interaction avec le SNP du gène *CCR5* a effectivement diminué à une p -valeur de $6,9 \times 10^{-3}$, alors que l'effet avec le rs11653282 du gène *CCL5* se renforçait à une p -valeur de $6,8 \times 10^{-5}$. Les MAF observées pour ces deux SNPs (rs3092960 G/A et rs11653282 C/T respectivement) sont de 0,12 et 0,10 dans la population de CEU de 1000 génomes ainsi que chez les individus de la cohorte ANRS CO13 HEPAVIH.

Du fait des faibles effectifs, nous avons décidé de réunir les génotypes portant l'allèle mineur et de tester l'interaction en modèle dominant. Nous avons observé des p -valeurs de $8,6 \times 10^{-4}$ pour le SNP rs3092960 (*CCR5*), et $p = 1,3 \times 10^{-3}$ pour le SNP rs11653282 (*CCL5*). Nous avons calculé les moyennes de fibrose hépatique, à partir des mesure de Fibroscan® en kPa des 289 patients de l'étude sqrtLS, en fonction des combinaisons de génotypes entre le SNP rs61183828 et les deux SNPs décrits dans le SIDA (Tableau 3 :rs3092960 et Tableau 4 :rs11653282).

Tableau 3. Moyennes de fibrose en kPa, en fonction des génotypes du SNP rs61183828 et du SNP rs3092960 (*CCR5*)

rs61183828	rs3092960	Moyenne fibrose kPa +/- sd (n)
AA	GG	8,92 +/- 7,02 (194)
	GA+GG	8,39 +/- 6,42 (62)
AG+GG	GG	15,82 +/- 17,01 (26)
	GA+GG	30,08 +/- 24,19 (8)

Tableau 4. Moyennes de fibrose en kPa, en fonction des génotypes du SNP rs61183828 et du SNP rs3092960 (*CCL5*)

rs61183828	rs11653282	Moyenne fibrose kPa +/- sd (n)
AA	CC	9,02 +/- 7,37 (205)
	CT+TT	7,83 +/- 4,15 (53)
AG+GG	CC	15,34 +/- 14,63 (25)
	CT+TT	29,84 +/- 27,49 (9)

L'effet identifié étant assez faible au vue des moyennes de fibrose lors des interactions, nous avons effectué des permutations des génotypes pour chacun des SNPs associés au SIDA pour garder la structure phénotypique et génotypique de l'effet du rs61183828 sur la fibrose et évalué

le réel impact des SNPs associés au SIDA. Ainsi, l'épistasie a été évaluée avec interSNP en mode dominant en réalisant 10000 permutations et nous avons obtenue des p -valeurs corrigées de $3,5 \times 10^{-2}$ qui ne passait plus le seuil de Bonferroni.

3.2 Réplication de signaux obtenus par d'autres GWAS sur des maladies du foie

Nous n'avons pas trouvé de réplication des meilleurs signaux identifiés dans nos deux GWAS (ni des régions génétiques concernées) dans d'autres GWAS portant sur des maladies du foie qui ont été publiées dans la littérature scientifique. Comme indiqué, dans les deux publications, nous avons répertorié 31 GWAS étudiées pour 6 maladies du foie différentes, qui aboutissaient à 322 SNP avec $p < 10^{-5}$ tous différents et sans LD avec les nôtres.

Réciproquement, à partir des résultats de l'étude d'association avec le phénotype sqrtLS en modèle additif, nous avons observé des réplifications d'associations décrites dans plusieurs maladies liées au foie ($p < 0.05$, Tableau 5). Les réplifications les plus intéressantes sont issues de l'étude GWAS « cas/témoins » publiée par Urabe *et al.* où les cas étaient proches car patients atteints d'hépatite C chronique présentant des cirrhoses hépatiques.

Tableau 5. Listes des SNPs associés à des maladies hépatiques, répliqués dans notre étude d'association génétique à sqrtLS

SNP	CHR	POS	A1	A2	p sqrtLS	Auteur (PMID)	Maladie	p GWAS d'origine	Gène rapporté
rs3817963	6	32368087	T	C	0,0022	Urabe (23321320)	HCV	1,00E-08	<i>BTNL2</i>
rs7774434	6	32657578	T	C	0,0049	Liu (20639880)	PBC	3,00E-26	<i>HLA</i>
rs9275319	6	32666295	A	G	0,0106	Jiang (23242368)	HBV	3,00E-17	<i>HLA</i>
rs672316	9	136138125	G	T	0,0254	Yuan (18940312)	ALD/NAFLD	4,40E-10	<i>ABO</i>
rs11864146	16	84046715	A	G	0,0255	Adams (23213074)	ALD/NAFLD	2,00E-06	<i>SLC38</i>
rs641959	9	136133699	C	A	0,0291	Yuan (18940312)	ALD/NAFLD	2,10E-11	<i>ABO</i>
rs514708	9	136133743	T	C	0,0291	Yuan (18940312)	ALD/NAFLD	1,90E-11	<i>ABO</i>
rs3129860	6	32401079	A	G	0,0292	Urabe (23321320)	HCV	1,00E-09	intergenic

rs1976403	1	21766453	A	C	0,0297	Chambers (22001757)	ALD/NAFLD	1,80E-50	<i>ALPL</i>
rs2143571	22	44391686	G	A	0,0435	Yuan (18940312)	ALD/NAFLD	7,20E-09	<i>SAMM50</i>
rs3135363	6	32389648	A	G	0,0477	Urabe (23321320)	HCV	1,00E-10	<i>BTNL2</i>
rs10896464	11	69078912	T	C	0,0480	Li (22807686)	HBV	6,45E-06	<i>CCND1</i>
rs474279	9	136139617	T	C	0,0489	Yuan (18940312)	ALD/NAFLD	8,30E-09	<i>ABO</i>

Pour l'étude par comparaison des groupes METAVIR (F0F1F2 et F3F4) plusieurs réplifications ont aussi été mises en évidence ($p < 0.05$) à partir de nos résultats obtenus en modèle dominant (Tableau 6). De manière très intéressante, sur ce phénotype légèrement différent (score METAVIR), on retrouve une même association, le SNP rs3817963, issue de l'étude d'Urabe *et al.*.

Tableau 6. Listes des SNPs associés à des maladies hépatiques, répliqués dans notre étude d'association génétique de comparaison des groupes METAVIR F0F1F2 et F3F4.

SNP	CHR	POS	A1	A2	p METAVIR F0F1F2 vs F3F4	Auteur (PMID)	Maladie	p GWAS d'origine	Gène rapporté
rs9357152	6	32664960	A	G	0.0081	Hirschfield (19458352)	PBC	1,00E-10	<i>c6orf10</i>
rs3817963	6	32368087	T	C	0.0207	Urabe (23321320)	HCV	1,00E-08	<i>BTNL2</i>
rs11676348	2	219010146	C	T	0.0217	Ellinghaus (26974007)	PSC	6,48E-07	<i>CXCR2</i>
rs4973341	2	228660362	C	T	0.0256	Cordell (26394269)	PBC	2,34E-10	<i>RNA5SP121</i>
rs11864146	16	84046715	A	G	0.0357	Adams (23213074)	ALD/NAFLD	2,00E-06	<i>SLC38</i>
rs10488631	7	128594183	T	C	0.0454	Liu (20639880)	PBC	3,00E-10	<i>TNPO3</i>
rs944002	14	103572815	A	G	0.0473	Chambers (22001757)	ALD/NAFLD	5,80E-29	<i>C14orf73</i>

Quatrième partie
Discussion et perspectives

1 BILAN DES DEUX ETUDES GENETIQUES REALISEES SUR LA COHORTE DE COINFECTION ANRS CO13 HEPAVIH

1.1 Travaux portant sur l'élasticité hépatique mesurée par Fibroscan®

1.1.1 Rappel des résultats

Nous avons mené une étude d'association « génome entier » afin d'identifier des gènes impliqués dans des mécanismes menant à une fibrose hépatique sévère.

A partir de données quantitatives normalisées (sqrtLS) définissant l'élasticité hépatique mesurée par Fibroscan® (en kPa), deux locus se sont révélés significativement associés à notre phénotype. Le premier locus 3p25 est porté par le SNP génotypé rs61183828 ($p < 5 \times 10^{-8}$) et les SNPs imputés, rs73132848, rs78828026, rs79896191 ($p < 5 \times 10^{-8}$) et rs73132859 ($p < 10^{-7}$). Ces SNPs sont situés dans des régions introniques du gène *OXTR*. Le second locus, 9p22, est porté par le SNP génotypé rs11790131 ($p < 5 \times 10^{-8}$) et le SNP imputé rs11790112 ($p < 5 \times 10^{-8}$) tous deux localisés dans le pseudogène *MAP1LC3BP1*. L'analyse initiale comprenant les patients ayant achevés une SVR après traitement, nous avons vérifié leur influence sur nos résultats en les retirant de l'analyse. Le locus 3p25 reste significatif alors que le locus 9p22 ne passe plus le seuil de « génome entier ».

Parmi les 7 locus identifiés ($p < 10^{-6}$), trois locus ont été répliqués, dont le locus 3p25 via le SNP rs73132859 ($p < 0,05$), dans une étude génétique Franco-Suisse de la mono-infection VHC par analyse « génome entier » portant sur la vitesse de progression de la fibrose (log de FPR, *logarithme of fibrosis progression rate*). Dans la cohorte ANRS CO13 HEPAVIH, nous avons observé une corrélation entre les phénotypes log de FPR et sqrtLS dans un sous-groupe de patients co-infectés ($r = 0,49$, p -valeur = $6,5 \times 10^{-5}$). Nous avons ensuite entrepris une seconde étape de répliquation chez des patients français mono-infectés par le VHC issus de la cohorte Genoscan, où des mesures de Fibroscan® étaient disponibles. Les données génétiques de Genoscan n'existant pas, nous avons génotypé un SNP de chacun des trois locus répliqués avec log de FPR. Nous avons également génotypé un SNP du locus 9p22 initialement observé

comme passant le seuil de significativité « génome entier » lors de notre étude. Finalement, seul le locus 3p25 *via* le SNP rs61183828 ($p < 5 \times 10^{-3}$) a été considéré comme répliqué.

1.1.2 Interprétation biologique

Le pseudogène *MAP1LC3BP1* ne semble pas avoir de lien direct avec la fibrose hépatique. Ce pseudogène de *microtubule associated protein 1 light chain 3 beta* est exprimé dans le foie (GTex et BioGPS). Néanmoins, aucune information supplémentaire n'a été décrite à ce jour.

Une analyse extensive du déséquilibre de liaison du locus 3p25 répliqué dans la monoinfection par le VHC a montré une région comportant trois gènes : *OXTR*, *CAV3* et *RAD18*. Quatre des cinq SNPs du locus sont introniques d'*OXTR* mais sont aussi localisés dans une région promotrice potentielle de *RAD18* enrichie en « îlots CpG ». En effet, le SNP rs61183828 est compris dans cette région promotrice, et le SNP rs78822806 ($r^2=1$ avec rs61183828 selon 1000 Génomes) est trouvé dans un site de fixation de facteurs de transcription FOXA1 (352) (*forkhead box protein A1*) de cette même région. FOXA1 est un activateur transcriptionnel spécifique du foie (353), ce qui est consistant avec un intérêt biologique lié au phénotype étudié. Néanmoins, aucun impact de ces SNPs n'a été observé sur l'expression des gènes *OXTR*, *CAV3* et *RAD18*.

Le gène *OXTR* code pour un récepteur de l'ocytocine couplé à une protéine G et ne semble pas être lié à la fibrose hépatique. Au contraire, *CAV3* et *RAD18* sont très pertinents car leurs produits sont impliqués, respectivement, dans des mécanismes de maintien de la structure cellulaire et de la réplication du VHC.

Le gène *CAV3* code pour une cavéoline qui est un composant des cavéoles retrouvées dans la membrane plasmique de la plupart des cellules. C'est une protéine d'échafaudage organisant d'autres protéines impliquées dans la signalisation cellulaire et dans le maintien de la structure cellulaire. L'ARNm de *CAV3* est détecté dans les tissus hépatiques et le gène a été rapporté comme étant associé avec la signalisation hépatique de l'insuline (354) et le syndrome dépressif avec dépendance à l'alcool (355). De plus, un long ARN non-codant est transcrit (lncRNA, 680 bases) chevauchant les gènes *OXTR* et *RAD18*. Ce gène pourrait appartenir à la nouvelle classe émergente des gènes bifonctionnels, codant pour des protéines mais aussi pour des isoformes lncRNA fonctionnels qui ont souvent une fonction antagoniste à leurs protéines (356).

Le gène *RAD18* pourrait aussi être pertinent car le signal est localisé dans une région promotrice potentielle du gène et dans un site de liaison à FOXA1. Ce gène code pour une protéine participant au système de réparation des dommages à l'ADN. Or, une déficience cellulaire en Rad18 rend les cellules plus permissives à l'infection virale (notamment le VIH) (357) et la protéine est impliquée dans les processus de réparation post-réplication survenant lors de l'infection par le VIH-1 (358). Plusieurs virus, dont le VHC, sont connus pour perturber les mécanismes de réparation de l'ADN et du cycle cellulaire pour maximiser leur réplication (359).

1.1.3 Comparaison avec la littérature

L'étude d'association « génome entier » avec la sévérité de la fibrose hépatique réalisée dans le contexte de la coinfection VIH/VHC est une première du genre, et aucune étude de ce type n'avait été précédemment publiée dans la littérature.

A ce jour, seules trois études d'analyse génétique « génome entier » ont traité d'associations avec des phénotypes liés à l'état hépatique dans le cadre de la monoinfection par le VHC. En 2011, une première étude publiée dans *Nature Genetics* par Kumar *et al.* a identifié le gène *MICA* (279) chez des patients japonais en analyse « cas/témoins », chez des patients chroniquement infectés par le VHC où les cas étaient atteints de carcinomes hépatocellulaires induits par le VHC. S'en est suivi une étude « cas/témoins » du même groupe, par Urabe *et al.* dans le *Journal of Hepatology* mettant en avant la région du CMH chez des patients japonais chroniquement infectés par le VHC où les cas présentaient des cirrhoses hépatiques (360). En 2012 dans le journal *Gastroenterology*, Patin *et al.* (274) ont révélé de nouvelles associations dans une population Franco-Suisse de patients monoinfectés par le VHC par analyse du taux de progression de fibrose (log de FPR). Ainsi, les gènes *RNF7*, *MERTK* et *TULP1*, impliqués dans des mécanismes de l'apoptose, ont été identifiés. Les résultats d'associations génétiques de l'étude « génome entier » dans cette population Franco-Suisse ont servi pour la première réplication des associations trouvées avec le phénotype sqrtLS de ma première étude (voir Résultats section 1).

Nous avons également entrepris une recherche de réplication dans des études « génome entier » liées à des maladies hépatiques : PSC (*primary sclerosing cholangitis*), l'hépatite auto-immune, PBC (*primary biliary cirrhosis*), VHC, VHB ou encore ALD (*alcoholic liver disease*) et NAFLD (*non-alcoholic fatty liver disease*). Finalement, nous avons récupéré sur 131 régions

chromosomiques 322 SNPs présentant des p -valeurs $< 10^{-5}$ dans 31 GWAS des 6 maladies mentionnées précédemment. Nous avons trouvé trois régions en commun mais les SNPs de notre étude n'étaient pas en LD avec ceux de la littérature et aucune réplique n'a pu être mise en évidence sur ces régions (6q23, 11p13 et 13q12).

1.2 Travaux portant sur l'impact de la gravité de fibrose entre les groupes METAVIR F0F1F2 et F3F4

1.2.1 Rappel des résultats

Lors de cette étude, nous avons défini deux groupes de patients coïnfectés VIH/VHC afin de les comparer par analyse génétique « génome entier ». Le premier groupe était constitué de 197 patients avec des fibroses minimales ou modérées (score METAVIR $< F3$). Le second présentait 95 patients montrant des fibroses sévères voire des cirrhoses (score METAVIR $\geq F3$).

Ainsi, la comparaison des groupes de score METAVIR F0F1F2 et F3F4 a permis d'identifier deux signaux passant le seuil de significativité « génome entier » par régression logistique en modèle dominant. Le premier est localisé dans la région chromosomique 1q31 et correspond à l'unique SNP imputé rs72727113 (beta = 2,37, SE = 0,41, p -valeur = $8,2 \times 10^{-9}$), où de nombreux CNVs ont été rapportés. Le second signal est situé dans la région chromosomique 5p15 et inclut trois SNPs et deux indels, tous en LD ($r^2 > 0,9$ selon 1000 Génomes) et introniques du gène *CTNND2*. Le signal est porté par l'association du SNP rs2158099 (beta = -1,67, SE = 0,29, p -valeur = $6,1 \times 10^{-9}$).

Nous avons observés un OR de 7,40 [CI 95%: 3,15; 18,94] pour le SNP rs72727113 du locus 1q31 et un OR de 0,194 [CI 95%: 0,103; 0,353] pour le SNP rs2158099 du locus 5p15.

1.2.2 Interprétation biologique

L'association du SNP rs72727113 (1q31) ne permet pas de mettre en évidence une fonction biologique ou un gène précis. Ce SNP est situé dans une région rapportée comme présentant de nombreux CNVs. Plusieurs gains et pertes de copies, incluant le locus 1q31, y ont été décrits dans des bases de données spécialisées sans impact fonctionnel clair.

Les 5 variants du second signal 5p15 ont été localisés dans l'intron 1 et 2 du gène (3 SNPs et 2 indels). Le gène *CTNND2* code une protéine, la δ -caténine, qui appartient à la superfamille des protéines armadillo/ β -caténine associées aux jonctions adhérentes cellulaires. Cette protéine

est très importante dans le développement du cerveau et de l'œil (361), mais aussi dans la formation de cancer (362). Plusieurs isoformes ont été décrits, et l'expression de *CTNND2* via ses isoformes permet un signal retour important pour la morphogenèse cellulaire, l'apoptose et les mécanismes de cancer (363). Cependant, les 5 SNPs n'ont pas été prédits comme impactant le phénomène d'épissage alternatif. La δ -caténine est impliquée dans les mécanismes de jonctions adhérentes permettant les contacts cellulaires en interagissant avec l'E-cadhérine, déstabilisant le complexe E-cadhérine/p120 (*CTNND1*) très important pour la formation des jonctions adhérentes et la fonction des cellules épithéliales (364). Il a été rapporté qu'une expression anormale de l'E-cadhérine et de p120 était corrélée avec des cholangio-carcinomes, des métastases intra-hépatiques et la survie de patient (365). De plus, la perte d'une copie du gène *CTNND2*, exprimé dans l'endothélium vasculaire, empêche *in vitro* la motilité des cellules endothéliales et l'assemblage vasculaire ainsi qu'une angiogenèse pathologique *in vivo* (366). L'expression de *CTNND2* a aussi été rapportée comme dérégulée dans de nombreux cas de HCC (367, 368). De plus, la δ -caténine interagit avec la β -caténine (369), et ensemble avec l'E-cadhérine, ils jouent un rôle significatif dans la physiopathologie hépatique à travers la voie de signalisation Wnt/Fzd/ β -caténine (370).

1.2.3 Comparaison avec la littérature

Les 6 associations des deux régions observées lors de notre étude n'ont pas été rapportées dans GWAS catalog (<https://www.ebi.ac.uk/gwas/home>) ($p < 10^{-5}$). Nous avons également recherché une réplique de SNPs dans notre étude au sein de ces deux régions ($p < 0,005$), potentiellement décrites dans d'autres maladies impliquant une fibrose hépatique liée ou non au VHC (VHC, PSC, l'hépatite auto-immune, PBC, VHB ou encore ALD et NAFLD). Sur les 6 maladies prospectées (31 GWAS pour 131 régions chromosomiques incluant 322 SNPs rapportés), aucune région commune n'a pu être mise en évidence en prenant en compte l'information du déséquilibre de liaison.

Néanmoins, en se focalisant sur le gène *CTNND2* lui-même et non plus aux SNPs, nous avons trouvés 7 GWAS présentant des associations avec $p < 10^{-6}$, sans rapport à l'infection par le VHC. En effet, une étude présentait une association avec une p -valeur inférieure au seuil de significativité « génome entier » ($p < 10^{-8}$) lié à la sclérose latérale amyotrophique. Les autres études ne présentaient également aucun lien avec la fonction hépatique ou le cancer, mais avec des désordres bipolaires, la schizophrénie, la myopie ou un déclin de la fonction pulmonaire. Aucun de ces SNPs n'a été répliqué ($p < 0,05$).

1.3 Discussion sur les résultats additionnels

1.3.1 Etude d'épistasie

En marge de l'étude génétique impliquant le phénotype sqrtLS, une étude d'épistasie a été réalisée pour identifier des interactions entre le SNP rs61183828, appartenant au locus 3p25 répliqué dans deux cohortes indépendantes de monoinfection par le VHC, et 9 SNPs connus pour être associés à la charge virale du VIH dans le SIDA. Par test de régression, nous avons observé deux effets épistatiques impliquant des SNPs des gènes *CCR5* ou *CCL5* et si on ajoutait la charge virale et le compte de $CD4^+$ en covariables, l'effet d'interaction avec *CCR5* ($p = 6,9 \times 10^{-3}$) diminuait alors que l'effet avec *CCL5* se renforçait ($p = 6,8 \times 10^{-5}$).

Nous avons aussi réalisé une approche de permutations sur les génotypes des SNPs associés au SIDA, et les deux signaux ne passaient plus le seuil de significativité établi à $1,8 \times 10^{-3}$.

L'interaction observée entre le SNP rs61183828 et les SNPs des gènes *CCR5* ou *CCL5* est pertinente dans le contexte des mécanismes de progression de la fibrose hépatique. Les chimiokines sont régulées positivement dans le foie et le sang périphérique et elles y jouent un rôle important lors de l'infection par le VHC (371). Le ligand *CCL5* est une chimiokine attractrice de cellules immunitaires (lymphocytes T actifs et monocytes). Ces cellules expriment le récepteur à motif C-C 5 (*CCR5*) de *CCL5*. Dans le foie, *CCR5* est exprimé dans les HSC (cellules étoilées du foie) qui sont la cause majeure de production de MEC et de cellules profibrogenique dans le foie endommagé (371). Des réponses précoces à l'infection par le VHC induisent l'expression de *CCL5* dans le sang périphérique et peuvent se traduire par une inflammation hépatique moins sévère (372). De plus, il a été décrit un mécanisme cellulaire plausible dans lequel un antagoniste du récepteur de *CCL5* accélère la régression d'une fibrose hépatique dans des modèles *in vivo* murins présentant des dommages chroniques au foie non-induit par le VHC (373). Enfin, l'axe *CCL5/CCR5* a été décrit comme médiateur de l'apoptose des lymphocytes T (374).

De façon intéressante, nous avons observé que l'effet épistatique diminuait avec *CCR5* alors qu'il augmentait avec *CCL5* en ajoutant des covariables spécifiques du VIH dans le test d'interaction. L'implication du gène *CCR5* dans le SIDA est bien connue y compris sur le plan génétique (34, 83-85), et la diminution de l'effet observé pour ce SNP avec ces covariables est donc logique. Le renforcement de l'effet épistatique observé pour les SNPs de *CCL5* lors de

l'ajout en covariables des paramètres clés du SIDA suggère plutôt que ce ligand a un rôle dans l'évolution de la fibrose, indépendamment du VIH, allant dans le sens de la littérature citée précédemment.

Cependant, ces effets épistatiques restent modestes et ne passent plus le seuil de Bonferroni après correction statistique par permutations. Aussi intéressants qu'ils soient, ces deux interactions avec notre SNP d'intérêt rs61183828 nécessitent donc d'être validées et si l'hypothèse de l'implication de *CCL5* dans la fibrose est exacte, l'épistasie avec notre SNP rs61183828 devrait même pouvoir être évaluée sur des cohortes de patients mono-infectés VHC déjà disponibles, et pas seulement par réplication sur de futures cohortes de coinfection VIH/VHC à venir.

1.3.2 Réplication des signaux obtenus par d'autres GWAS sur des maladies du foie.

Les réplifications sont essentielles pour confirmer la réalité d'une association génétique dans des populations indépendantes. De la liste de 322 SNPs décrits dans la littérature comme liés à des maladies du foie, aucun ne passerait le seuil de Bonferroni pour une réplication *stricto sensu* dans nos deux études ($p < 1,5 \times 10^{-4}$ pour 322 SNPs).

Parmi ces 322 SNPs de GWAS de maladies liées au foie, nous n'avons identifié que 13 SNPs dans l'étude génétique quantitative (sqrLS) et que 7 SNPs dans l'étude génétique comparant les groupes METAVIR F0F1F2 vs F3F4 montrant une valeur $p < 0,05$. Aucun de ces SNPs ne faisait partie des signaux majeurs trouvés dans de nos deux études. Parmi eux, deux SNPs présentaient un intérêt particulier car ils sont chacun retrouvés dans nos deux études. Il s'agit du SNP rs3817963 (*BTNL2*) associé ($p = 10^{-8}$) à la cirrhose hépatique dans l'hépatite C chronique et le SNP rs1186146 (*SLC38*) observé (2×10^{-6}) dans une étude génétique traitant la NAFLD.

L'étude « cas/témoins » d'Urabe *et al.* portait sur des cas de cirrhose hépatique dans l'hépatite C chronique et c'est, en fin de compte, la seule qui présentait un phénotype vraiment similaire à ceux de nos deux études sur la fibrose dans la coinfection VIH/VHC. Urabe *et al.* ont comparé 682 patients japonais présentant des cirrhoses induites par l'infection VHC avec 1045 patients japonais chroniquement infectés par le VHC sans cirrhose. Le phénotype dichotomique de cette

étude est très extrême par rapport aux deux études réalisées sur la cohorte ANRS CO13 HEPAVIH. En effet, la cirrhose correspond à un score F4 sur l'échelle METAVIR et nos deux études ne présentaient pas autant de patients cirrhotiques. Le SNP rs3817963 est situé dans le gène *BTNL2* dans la région du HLA. Nous observons une réplique de ce SNP plus forte dans notre étude du phénotype quantitatif. La distribution quantitative permet de rendre plus extrême la fibrose chez les patients par rapport à l'étude de comparaison de groupes METAVIR qui nivelle la fibrose en classes. On peut remarquer que c'est assez rare d'arriver à répliquer un SNP dans le locus HLA sur des populations d'origine très différente comme des caucasiens et des asiatiques.

Cette réplique souligne l'intérêt de travailler sur des cohortes de coinfection par le VIH et par le VHC, qui permettent dans notre cas de confirmer des associations génétiques observées en monoinfection par le VHC et de découvrir en plus de nouveaux signaux génétiques potentiellement impliqués dans la coinfection par le VIH et par le VHC.

1.4 Comparaison des deux études génétiques « génome entier » réalisées sur la cohorte ANRS CO13 HEPAVIH

1.4.1 Comparaison entre les deux études

Les deux études ont permis d'identifier des associations avec la sévérité de la fibrose dans la coinfection VIH/VHC. Dans la première, la fibrose était définie quantitativement par Fibroscan® à différentes dates (sqrtLS), dans la deuxième, il s'agissait de comparer deux groupes « cas/témoins » de scores METAVIR F0F1F2 et F3F4 à l'inclusion des patients issus de la cohorte ANRS CO13 HEPAVIH.

Les résultats de la première étude du trait quantitatif ont été recherchés dans la seconde étude « cas/témoins » et réciproquement. L'ensemble des *p*-valeurs des SNPs obtenus pour les deux études est résumé dans le Tableau 7.

Tableau 7. Ensemble des SNPs et des p-valeurs associés à la sévérité de la fibrose dans chacune des deux études génétiques. Par trait quantitatif (sqrtLS) ou par « cas/témoins » (F0F1F2 vs F3F4)

SNPs	locus	<i>p</i> -valeur sqrtLS	<i>p</i> -valeur F0F1F2 vs F3F4
rs73132848	3p25	1,00x10 ⁻⁹	0,0052
rs78828026	3p25	2,02x10 ⁻⁸	0,0118
rs61183828	3p25	3,86x10 ⁻⁹	0,0038
rs79896191	3p25	3,24x10 ⁻⁹	0,0039
rs73132859	3p25	2,20x10 ⁻⁷	0,0350
rs11790112	9p22	1,79x10 ⁻⁸	2,07x10 ⁻⁵
rs11790131	9p22	9,26x10 ⁻⁹	4,78x10 ⁻⁶
rs2158099	5p15	0,0151	6,05x10 ⁻⁹
rs35268064	5p15	0,0099	8,29x10 ⁻⁹
rs11367210	5p15	0,0169	4,80x10 ⁻⁹
rs1423493	5p15	0,0151	6,05x10 ⁻⁹
rs2074260	5p15	0,0141	3,00x10 ⁻⁸
rs72727113	1q31	0,1144	8,18x10 ⁻⁹

On observe que les SNPs identifiés par association à la fibrose quantitative montrent des *p*-valeurs en étude « cas/témoins » passant un seuil de réplification simple de 0,05. En particulier le signal du locus 9p22 qui présente des *p*-valeurs plus fortement répliquées en étude « cas/témoins » ($p < 10^{-5}$). Pour le locus 3p25, répliqué en monoinfection par le VHC, les *p*-valeurs sont aussi répliqué en étude « cas/témoins » et atteignent même $3,8 \times 10^{-3}$ pour le SNP rs61183828. Inversement, les associations observées en étude de comparaison de groupes de score METAVIR montrent des *p*-valeurs autour de 10^{-3} pour le locus 5p15 et une simple tendance ne passant pas le seuil de 0,05 pour le locus 1q31. Cette différence s'explique dans la définition des phénotypes. La définition de chacun des deux phénotypes a été faite à des dates différentes et par des méthodes d'évaluation de fibrose partiellement différentes.

Dans l'étude « cas/témoins », l'évaluation de fibrose en scores METAVIR était issue de trois méthodes (pbh, 24%, Fibrotest®, 16.8%, FibroScan® 59.2%) alors que l'étude quantitative (sqrtLS) présentait des valeurs de fibrose uniquement évaluées par Fibroscan®. Dans l'optique d'une comparaison de phénotypes basés sur des scores METAVIR ou sur des valeurs d'élasticité hépatique obtenues par Fibroscan®, nous avons réalisé une analyse de corrélation. Ainsi, nous obtenons une corrélation significative mais que partielle entre le phénotype basé

sur le FPR (score METAVIR par années d'infection) et le phénotype quantitatif (sqrtLS ~ log de FPR, $r = 0,49$, p -valeur = $6,5 \times 10^{-5}$).

1.4.2 Interprétation des différences observées

Comme nous l'avons mentionné plus haut, la définition des phénotypes est une étape critique du « design » d'une étude « génome entier ». Il a été démontré qu'un trait quantitatif permet d'obtenir plus de puissance statistique que la dichotomisation de ce dernier en « cas/témoins ». En effet, la transformation de données quantitatives en classes fait perdre de l'information (375).

Dans cette thèse, le phénotype quantitatif d'évaluation de fibrose a été obtenu par Fibroscan® et n'a pas été dichotomisé. Néanmoins, lors de l'étude « cas/témoins », les groupes ont été définis à partir de mesures par Fibroscan® mais aussi à partir de pbh ou par Fibrotest®. En plus d'une perte d'information par classification de la fibrose en groupe, un problème de définition des scores METAVIR peut exister. En effet, l'évaluation d'un score METAVIR F2 par des méthodes non-invasives (Fibroscan® et Fibrotest® dans notre cas) génère souvent des erreurs de classification. Ainsi, des scores F2 définissant une fibrose modérée et le début des stades ultérieurs, peuvent en réalité être de niveau F1 (376). De meilleures méthodes de prédiction de ce stade pourraient apporter plus d'information et donc plus de puissance de détection par analyse « génome entier ».

Les études « cas/témoins » définissant deux groupes extrêmes basés sur une distribution de valeurs quantitatives peuvent souvent révéler une augmentation de puissance statistique (377). Dans notre étude « cas/témoins », la définition dichotomique ne favorise pas les extrêmes par la présence du groupe intermédiaire F2 qui peut induire de l'ambiguïté. Ainsi, dans l'étude « cas/témoins », nous avons entrepris de retirer les individus avec des fibroses intermédiaires de score METAVIR F2 (N individus = 55) afin de rendre les porteurs de phénotypes « plus extrêmes » et éliminer une source de biais potentielle. L'analyse a été faite entre les groupes F0F1 et F3F4 (N individus = 237) et nos signaux initiaux « flirtent » avec le seuil de significativité « génome entier » (p -valeurs $< 7 \times 10^{-8}$). En enlevant en plus le groupe F3, on définit une comparaison de 215 individus (F0F1 vs F4) et nos principaux signaux sont observés avec des p -valeurs $> 5 \times 10^{-7}$, le gain de puissance statistique étant probablement impacté par le retrait total de 77 individus ayant des scores METAVIR de fibroses F2 ou F3.

Cette analyse permet de montrer que le retrait des scores METAVIR F2 de notre étude « cas/témoins » influe légèrement sur nos résultats et permet de garder une puissance statistique malgré un faible nombre d'individus analysés. Cependant, le retrait des scores METAVIR F3 a tendance à faire disparaître nos signaux. A trop vouloir rendre extrême un phénotype dichotomique, on peut perdre de l'information et passer à côté d'associations pertinentes. Enlever les scores METAVIR F3 n'est pas forcément nécessaire car, bien qu'intermédiaire avec le stade F2, ce score définit une réelle fibrose clinique et apporte de l'information au phénotype dichotomique.

De façon surprenante, nous avons observé que le signal 3p25, identifié en étude quantitative (sqrtLS), se renforçait en rendant le phénotype dichotomique plus extrême (rs61183828, F0F1F2 vs F3F4 $p = 3,8 \times 10^{-3}$, F01 vs F3F4 $p = 2,1 \times 10^{-3}$, F0F1 vs F4 $p = 1,0 \times 10^{-4}$). Ce signal a été expliqué par des individus ayant une fibrose très extrême lors de l'étude quantitative, notamment deux individus au génotype rs61183828_GG.

1.4.3 Intérêt de la cohorte de patients coinfectés ANRS CO13 HEPAVIH

La coinfection par le VIH et le VHC conduit à des réactions complexes qui à ce jour ne sont que partiellement comprises et décrites. Les études dans le domaine se sont surtout focalisées, à juste titre, sur la compréhension des mécanismes liés à la guérison spontanée ou après traitement, à l'image de la découverte de l'implication du gène *IL28B* (224, 251-253). La fibrose hépatique induite par le VHC est aggravée par des facteurs environnementaux, et la coinfection par le VIH accélère ce phénomène. Cependant, les mécanismes génétiques sous-jacents impliqués lors de la coinfection avec le VIH ne sont pas connus.

La cohorte ANRS CO13 HEPAVIH a permis de centraliser des données cliniques relatives à la coinfection par le VIH et le VHC (voir Matériels et méthodes section 2.1.1). Cette cohorte est donc très intéressante pour des études génétiques puisqu'il est connu que l'évolution naturelle et le pronostic de l'hépatite C sont plus sévères en cas de coinfection par le VIH. En particulier, une vitesse de progression de la fibrose hépatique accélérée et une évolution rapide vers la cirrhose -et ses complications- ont été observées dans la coinfection VIH/VHC. Plusieurs investigations ont été menées afin de comprendre ces phénomènes, notamment en étudiant la dérégulation immunitaire ainsi que la déplétion des CD4⁺ intestinaux (75). Un faible taux lymphocytaire CD4⁺ et une charge virale du VIH élevée ont été observés chez des patients coinfectés VIH/VHC présentant une progression de fibrose hépatique sévère et rapide (378).

Une suggestion serait que le VIH accélère la fibrose hépatique par interaction de sa protéine gp120 avec les corécepteurs CCR5 et CXCR4 des hépatocytes et d'autres cellules hépatiques par l'intermédiaire du TGF- β (307).

Une des hypothèses sous-jacente à notre travail était justement que des signaux génétiques moins visibles en monoinfection VHC pourraient ressortir plus fortement dans le cas d'une coinfection par le VIH.

Dans le cadre de notre étude génétique sur 292 patients, nous avons identifié un signal très fortement associé à la fibrose hépatique au niveau du locus 3p25. Ce locus présente un intérêt biologique certain et a été répliqué, bien plus faiblement, dans deux cohortes indépendantes de patients monoinfectés par le VHC. Plusieurs hypothèses peuvent être avancées pour cette réplification plus faible :

- L'hypothèse initiale d'un réel impact de la coinfection par les deux virus
- L'effet connu du « winner's curse » qui surestime les associations génétiques identifiées en étude « génome entier » (379)
- La différence des phénotypes pour la réplification dans l'étude basée sur la population Franco-Suisse monoinfectée VHC. Néanmoins, la cohorte Genoscan permettant une modeste réplification présentait des individus monoinfectés VHC avec la même définition de fibrose quantitative

Afin de pouvoir valider ces hypothèses, il sera nécessaire de pouvoir reproduire notre étude dans une autre cohorte de patients coinfectés VIH/VHC, bien que de telles cohortes ne soient pas pour le moment exploitées en études génétiques.

Dans notre étude, nous avons aussi recherché des effets épistatiques entre le SNP rs61183828 que nous avons identifié et des variants génétiques connus pour influencer l'évolution du VIH. De manière intéressante, une tendance significative (sans correction de Bonferroni) a été observée entre le SNP rs61183828 et deux marqueurs connus de l'infection par VIH-1 : *CCL5* et *CCR5*. Comme cela a été discuté précédemment (voir Discussion section 1.3.1), si ces deux marqueurs ont aussi été suspectés pour leur implication dans les mécanismes de fibrose hépatique, il se pourrait que l'effet épistatique de *CCL5* soit aussi observable chez des monoinfectés VHC.

Tous ces résultats génétiques permettent de mieux comprendre les mécanismes moléculaires sous-jacents à la fibrose hépatique lors de l'infection par le VHC. Il faudra bien-sûr les valider par l'étude génomique d'autres cohortes, mais ils soulignent l'intérêt de travailler sur ce phénotype de coinfection.

2 CRITIQUE DES METHODES UTILISEES

2.1 Etudes « génome entier »

Le succès des analyses d'associations « génome entier » dans l'identification des facteurs de risques associés aux maladies a été massif durant la dernière décennie. Cependant, en dépit des investissements financiers croissants, beaucoup de variants génétiques associés à ces maladies complexes – multifactorielles - comme le SIDA ou l'hépatite sont encore inconnus.

2.1.1 Intérêt

L'avantage indéniable apporté par ces études « génomes entier » a été la possibilité d'une recherche extensive de variants communs associés à un caractère sur le génome entier, en excluant tout *a priori* biologique. Il a été ainsi possible d'identifier des régions chromosomiques associées à des caractères/maladies alors que ces liens étaient jusque-là insoupçonnés.

L'explosion des publications portant sur des GWAS a aussi été facilitée par la relative simplicité de mise en œuvre et d'exécution de cette nouvelle technologie des puces de génotypage. Les méthodologies utilisées lors d'une GWAS sont désormais bien établies et le protocole à suivre pour réaliser une telle étude est bien défini. De plus, beaucoup de publications détaillent précisément les différentes étapes à suivre lors d'une GWAS, aussi bien au niveau du contrôle qualité qu'au niveau de l'analyse statistique (256, 332, 380, 381).

Enfin, l'évolution des techniques expérimentales et de la bioinformatique a aussi contribué à l'essor de ces études « génome-entier ». En effet, les puces de génotypage permettent aujourd'hui de génotyper jusqu'à 5 millions de SNPs chez plusieurs milliers d'individus à faible coût. Par conséquent, les GWAS sont désormais des analyses relativement peu coûteuses permettant de recueillir une grande quantité d'information. En parallèle, les progrès de la bioinformatique, et notamment les méthodes d'imputation, ont facilité la comparaison des résultats obtenus sur différentes cohortes ainsi que la mise en place de méta-analyses. Cependant, ces études présentent aussi des inconvénients, notamment statistiques.

2.1.2 Limites

L'inconvénient principal des études « génome-entier » apparaît lors de la correction des tests multiples permettant de repérer les polymorphismes significativement associés au phénotype analysé. En effet, la méthode la plus couramment utilisée pour pallier au problème de comparaison multiple est la méthode de Bonferroni. Bien qu'elle soit simple et rapide à mettre en place, cette méthode est très conservatrice. Seuls les signaux très associés seront détectés, indiquant des résultats avec un risque d'erreur statistique de 5% sur l'ensemble des SNPs testés en « génome entier ». Mais, des signaux réellement associés au caractère étudié dont les effets sont plus modestes pourront ne pas être considérés comme significativement associés, avec pour cause la stringence de cette correction. L'approche alternative s'appuyant sur le taux de fausse découverte (FDR, *false discovery rate* en anglais) permet d'atténuer ce phénomène, mais seulement de manière partielle. Avoir accès à des outils statistiques permettant d'abaisser la stringence des tests multiples permettrait de ne pas laisser de côté des signaux réels mais plus modestes.

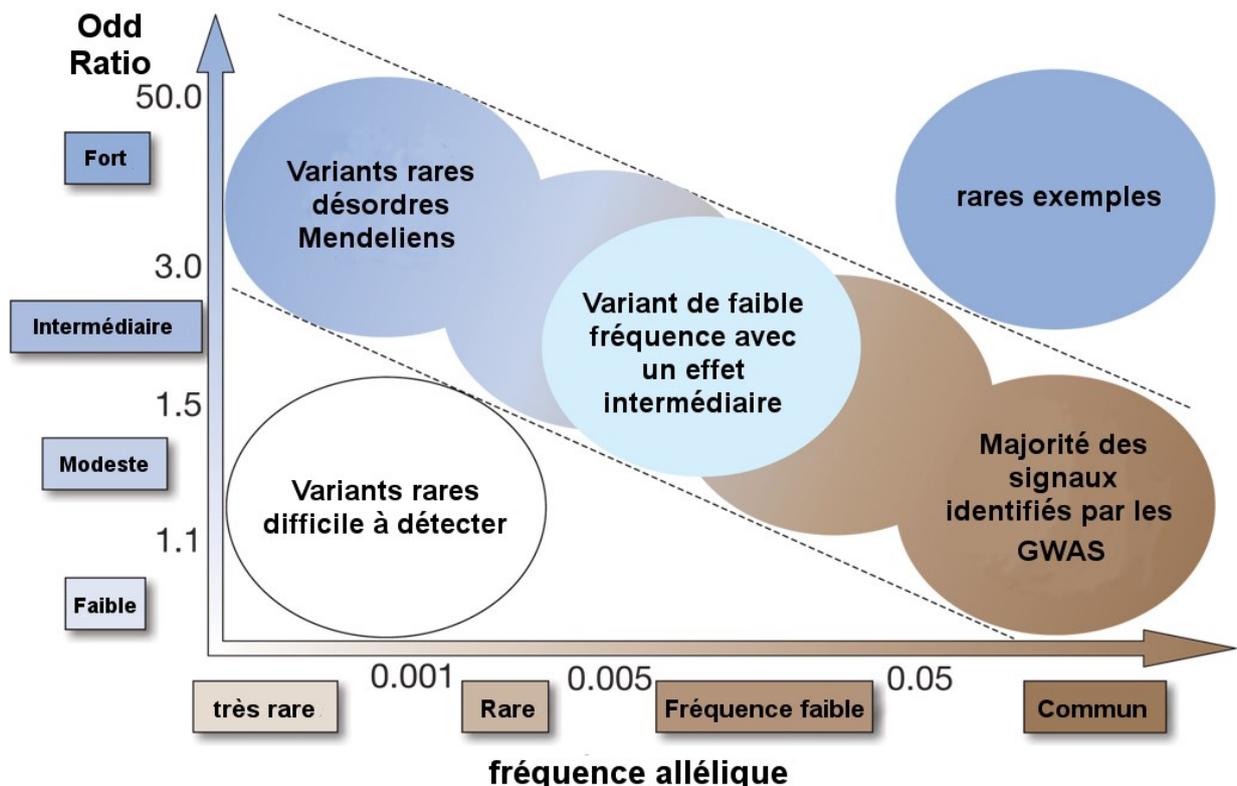


Figure 34. Classification des marqueurs génétiques associés aux maladies en fonction de leur fréquence et de leur pénétrance dans la maladie

Une autre limite de ces méthodes résulte directement du paradigme selon lequel les maladies fréquentes sont les conséquences de polymorphismes fréquents. Ainsi, les puces de génotypage ont été élaborées de façon à privilégier les SNPs dont la MAF est supérieure à 5%. Les SNPs moins fréquents sont de fait exclus de ces analyses. Or, depuis quelques années, ce paradigme est remis en question car les variants communs ne suffisent pas à déterminer complètement l'architecture génétique des pathologies complexes (382). L'imputation a partiellement rendu possible l'analyse des SNPs de fréquence faible, en permettant l'inférence des génotypes de ces SNPs mais la qualité de leur imputation peut s'avérer problématique. Se pose également le problème de la puissance statistique lors de l'analyse de tels polymorphismes. En effet, un grand nombre de patients est nécessaire afin d'observer la significativité statistique d'une association entre un SNP de fréquence faible et un phénotype. Une solution simple est la mise en commun des données issues de différentes cohortes dans le but d'entreprendre des méta-analyses, avec cependant la difficulté éventuelle de normalisation de toutes ces données indépendantes.

3 PERSPECTIVES

3.1 Réplifications et méta-analyse dans la coinfection

La réplification des résultats obtenus lors d'une GWAS est essentielle. Elle consiste à comparer les p -valeurs des signaux obtenus à ceux trouvés dans des cohortes indépendantes étudiant un phénotype identique ou similaire. Nous avons répliqué un de nos résultats dans la monoinfection VHC (premier article) mais nous n'avons pas pu répliquer notre seconde étude sur le phénotype METAVIR car il n'y a pas d'autre cohorte de coinfection génotypée pour le moment.

Afin de trouver de nouvelles associations avec la sévérité de la fibrose, nous pourrions combiner des données génomiques de différentes études traitant de la fibrose hépatique dans le contexte de coinfection VIH/VHC. Ce type d'étude est appelée méta-analyse mais pour le moment, nous avons réalisé la seule étude génétique traitant cette thématique dans la coinfection VIH/VHC.

Dans le futur, une collaboration avec une cohorte allemande de malades coinfectés par le VIH/VHC (PROBE-C) est en discussion.

3.2 Approches multi-marqueurs

Les études d'association « génome-entier » analysent des marqueurs simples. Or, pour des maladies complexes, la combinaison de plusieurs polymorphismes, et non un seul, peut être à l'origine de la maladie. Il semble donc pertinent de mettre en place des stratégies permettant d'évaluer ces possibilités.

3.2.1 Haplotypes

Les haplotypes sont définis comme la combinaison de n polymorphismes situés sur un même chromosome (voir Introduction section 1.4.3). Ces haplotypes peuvent être à l'origine de dysfonctionnement et ainsi expliquer certaines maladies (383-386). Ces exemples illustrent la pertinence de la recherche d'haplotypes comme facteur de risque de certaines pathologies. Toutefois, de nombreuses questions sont sous-jacentes à ces analyses. En effet, comment définir les haplotypes étudiés ? Combien et quels marqueurs inclure dans les haplotypes considérés ?

Enfin, étant donné le nombre d'haplotypes qu'il est possible d'analyser, comment contourner les limites des corrections multiples afin de ne pas être trop conservatif ?

3.2.2 Epistasie

L'épistasie désigne l'interaction entre deux loci : l'impact d'un génotype à un locus particulier dépend du génotype à un autre locus. Les GWAS ont pour vocation de rechercher les polymorphismes ayant individuellement un impact sur le caractère étudié mais ceux-ci ne représentent qu'une fraction des facteurs génétiques responsables du phénotype (382). Les interactions entre SNPs peuvent expliquer une autre portion de ces facteurs génétiques.

En effet, plusieurs exemples d'interaction génétique ont été observés. Des interactions entre le gène *PTPN2* et des gènes du métabolisme de la vitamine D contribuent au risque d'arthrite chronique juvénile (387). De même, une interaction entre le gène *ERAPI* et le gène *HLA-C* influence la susceptibilité au psoriasis (388). Ces deux exemples illustrent la pertinence de la recherche d'épistasies pour déterminer les facteurs génétiques impliqués dans des maladies complexes.

Plusieurs méthodes ont été implémentées afin de rechercher les épistasies (348, 389-393). Cependant, la recherche d'interactions est encore confrontée à des problèmes statistiques tels que la puissance ou la correction des tests multiples. Une publication récente détaille les principales méthodes existantes et les problèmes inhérents à ces techniques (394).

Lors de mes travaux de thèse, nous avons réalisé une étude d'épistasie. Nous avons ainsi observé une tendance d'interaction entre le SNP rs61183828 favorisant la fibrose hépatique induite par le VHC et des SNPs candidats liés au VIH-1. Cette étude a été confrontée aux limites statistiques décrites plus haut du fait de faibles effectifs (voir Discussion section 1.3.1).

Une investigation d'épistasie plus large sur tous les SNPs génotypés pourrait révéler de nouvelles interactions dans la coinfection VIH/VHC. Il serait alors possible de tester ces effets dans des cohortes de monoinfection par le VHC, mais l'idéal serait bien-sûr une réplification dans la coinfection VIH/VHC.

3.3 Analyse des CNVs

Les avancées dans la connaissance du génome ont mis en lumière l'importance d'autres marqueurs de variabilité comme les CNVs (Copy Number Variation) qui sont des grandes

insertions ou délétions (voir Introduction section 1.3.5). Ces marqueurs se retrouvent dans toutes les populations sans affecter nécessairement la santé des patients.

Ces polymorphismes ont été omis volontairement des analyses présentées dans ce travail puisque leur inférence à l'aide des puces de génotypage est complexe (395, 396). Cependant, leur analyse ne doit pas être négligée dans la mesure où des variants de cette nature ont déjà été associés à plusieurs maladies comme le cancer (397), l'autisme (398) ou l'obésité (399, 400), et que leur influence sur la variabilité observée face aux maladies est reconnue. Une étude se focalisant sur la région d'*IL28B* a été réalisée dans la mono-infection par le VHC sans révéler le moindre rôle de CNVs dans la forte influence d'*IL28B* dans la réponse virologique soutenue spontanée (401). Bien que l'étude des CNVs soit encore problématique à partir des données obtenues avec les puces, une analyse avec des phénotypes extrêmes liés à la fibrose dans la cohorte ANRS CO13 HEPAVIH est en cours de développement.

3.4 Analyses d'eQTLs

Les eQTLs (*expression Quantitative Trait Locus* en anglais) représentent les traits (principalement des SNPs) dont l'allèle change l'expression d'un gène cible. Nous distinguons deux types d'eQTLs :

- les cis-eQTLs représentent les marqueurs situés relativement près des gènes affectés sur le chromosome et qui modifieraient par exemple la fixation des facteurs de transcription.
- Les trans-eQTLs sont des marqueurs éloignés de leurs gènes cibles, qui peuvent ne pas être sur le même chromosome, et qui influencent de manière indirecte l'expression de ceux-ci, par des éléments régulateurs positionnés près du variant tels que les miRNA ou des facteurs de transcription.

Un eQTL peut être détecté en calculant le facteur de corrélation entre un allèle (trait qualitatif) et l'expression d'un autre gène (trait quantitatif). Lors de ma thèse, une recherche systématique d'eQTLs a été faite afin de définir si les polymorphismes trouvés étaient impliqués dans la régulation d'expression de gènes. Néanmoins, une recherche plus large des résultats ($p < 10^{-3}$) de GWAS pourrait faire émerger de nouveaux signaux impliqués dans la fibrose hépatique liée à l'infection VHC par décomposition de la variance génomique (402) comme cela a été fait dans le SIDA (403).

3.5 Nouvelles technologies de séquençage ou NGS

En biologie, l'évolution technique et technologique est constante et continue, mais dans le cas du séquençage de l'ADN/ARN, elle se fait avec une rapidité exponentielle, modifiant en permanence les concepts et les conditions de l'exercice médical, notamment en oncologie et en génétique. Ainsi, plusieurs technologies ont émergées dont les NGS (*Next Generation Sequencing*) qui ont permis le séquençage massif et très rapide de l'ADN et de l'ARN mais aussi d'évaluer les modifications épigénétique de l'ADN :

Le séquençage intégral du génome ou de ses parties codantes (nommées exome) chez un grand nombre de personnes comparativement aux technologies plus anciennes a été rendu possible. L'analyse des données générées par de telles techniques permet de caractériser des polymorphismes rares et a ouvert de nouvelles perspectives pour de nombreuses maladies (404).

Au cours des dernières années, les approches par NGS ont non seulement permis de générer rapidement des séquences du VHC, mais sont aussi devenues un outil clinique pour le suivi des patients. Par exemple, le séquençage profond a été utilisé pour identifier de nouveaux sous-types du VHC et étudier l'histoire de l'épidémie du VHC dans des populations éloignées (405). Récemment, des approches par NGS ont permis de comparer les « quasi-espèces » du VHC chez des patients hémophiles coinfectés VIH/VHC et des patients mono-infectés VHC dans une population japonaise. Cette étude a montré un lien entre des « quasi-espèces » du VHC associé à la résistance aux DAAs (protéase NS3) et le succès d'une réponse virologique soutenue chez des patients hémophiles coinfectés VIH/HC (406).

Au-delà de l'approche génétique stricte, les NGS peuvent permettre d'étudier l'expression des gènes par RNAseq. Cette méthode sans *a priori* biologique est hautement reproductible des résultats obtenus par « puce d'expression », mais apporte une plus haute sensibilité de détection des ARN (*i.e* ARNm rare) et peut permettre l'identification de différents variants d'épissage. En effet, le transcriptome constitue l'ensemble de tous les ARNs issus de la transcription du génome. Ils comprennent les ARNs messagers, ribosomiques, de transfert et autres espèces d'ARN comme les micro-ARN (miRNA). Plusieurs études ont été réalisées dans le contexte de la mono-infection VHC, identifiant des gènes dérégulés lors de la phase aiguë de la maladie (407), lors de la phase chronique après traitement basé sur l'IFN- α (408), mais aussi la dérégulation par des miRNA consécutif de l'infection (409). L'analyse du transcriptome est une

approche complémentaire aux études génétiques pour expliquer la physiopathologie de maladies multifactorielles, et se base sur l'identification de gènes qui sont différentiellement exprimés entre les patients et les témoins.

L'expression d'un gène dépend très faiblement de sa séquence d'ADN, mais plus de facteurs de transcription ou des mécanismes épigénétiques, comme la méthylation des histones, qui sont difficilement détectables par l'étude du génome (410). Les NGS permettent également d'étudier les modifications/altérations épigénétiques dans le contexte d'une maladie et ainsi appréhender à un niveau supérieur la compréhension des mécanismes sous-jacent, conduisant au suivi clinique des patients et la découverte de marqueurs pronostiques (411). Le rôle joué par l'épigénétique dans la plasticité cellulaire est non-négligeable et semble important dans la formation des cancers (412). Lors de l'infection par le VHC chez des patients alcooliques présentant des HCC, plusieurs signatures de la méthylation de l'ADN se sont révélées dérégulées alors que d'autres semblaient conservées (413).

Toutes ces approches sont complémentaires aux études génétiques classiques, et peu d'entre elles ont été réalisées dans la coinfection VIH/VHC. Dès lors, l'utilisation technologique des NGS pourra certainement être utile pour une meilleure compréhension des mécanismes découlant de l'interaction entre le VIH et VHC et des mécanismes associés.

3.6 Biologie des systèmes

Les études d'association « génome-entier » permettent la détection de polymorphismes génétiques associés à un phénotype. Toutefois, l'intégration de données biologiques est aujourd'hui une étape possible grâce à la disponibilité de grandes bases de données biologiques et cela devrait permettre d'aller plus loin que les tests statistiques utilisés. Ainsi, la biologie des systèmes est l'étude intégrative par modélisation des gènes, des protéines, des réactions biochimiques et autres processus physico-chimiques considérés en tant que système dynamique d'éléments en interaction.

Plusieurs disciplines ont pour vocation de générer des données, notamment issues des NGS, et permettent la mise en place d'une telle intégration :

- Le transcriptome qui étudie les ARNs messagers issus de l'expression d'une partie du génome. La comparaison des données transcriptomiques et des données génomiques a

notamment permis l'élaboration de bases de données renseignant l'impact de polymorphismes sur l'expression des gènes (414-416) ;

- L'épigénomique qui étudie les mécanismes de régulation de l'expression des gènes pouvant être influencés par l'environnement ;
- Le protéome qui étudie l'ensemble des protéines présentes dans la cellule ;
- Le métabolome qui étudie de petites molécules comme les hormones par exemple dans différents tissus.

Le projet ENCODE (417, 418), qui a débuté en 2003, a pour ambition de décrire les éléments fonctionnels du génome humain en combinant ces différentes disciplines. Les données d'ores et déjà générées sont publiques et ont été exploitées dans différentes bases de données comme RegulomeDB (419). L'émergence de la biologie systémique ouvre de nombreuses perspectives pour une meilleure compréhension des mécanismes biologiques.

L'intégration des données de GWAS avec les informations des voies de signalisation biologiques peut permettre d'identifier des voies spécifiques associées au phénotype étudié. Cette analyse présente l'avantage d'utiliser les données de GWAS dans leur ensemble. Les signaux n'atteignant pas le seuil de significativité « génome entier » sont utilisés pour identifier l'enrichissement dans des gènes de voies biologiques, révélant des pistes biologiques potentielles non soupçonnées. Plusieurs méthodes ont été développées comme l'algorithme GSEA (*Gene Set Enrichment Analysis* en anglais) (420), mais ces méthodes restent relativement récentes et ne font pas consensus au niveau des protocoles d'analyse. La définition des voies de signalisations biologiques est complexe du fait des redondances de fonction des gènes et de leur implication dans de multiples mécanismes biologiques. Cependant, l'intégration d'un *a priori* biologique doit permettre d'identifier plus facilement les mécanismes impliqués dans un phénotype, et d'améliorer l'interprétation de l'impact biologique d'un ou plusieurs polymorphismes. Plusieurs revues énumèrent les principales limites des techniques développées pour identifier des voies de signalisation associées à un phénotype (421-425).

Aucune étude à partir de résultats GWAS n'a été menée dans l'infection par le VHC. Une étude a analysé les différences de régulation des voies de cytotoxicité des CMSPs par profilage transcriptomique de 12 patients mono-infectés VIH ou VHC et de 10 patients co-infectés VIH/VHC. Ces données de « puce d'expression » ont permis d'identifier des interactions

cytokines/récepteurs régulées négativement chez les coinfectés par rapport aux monoinfectés (426).

3.7 Autres phénotypes à explorer

L'intérêt de la cohorte ANRS CO13 HEPAVIH réside dans son caractère extrêmement bien documenté donnant de la robustesse et de la puissance aux résultats d'association. Dans cette thèse, nous nous sommes focalisés sur le phénotype le plus évident de la coinfection par le VIH et le VHC, l'évolution de la fibrose hépatique. Néanmoins, d'autres aspects restent à explorer liés à l'infection par le VIH et le VHC :

- Etude des risques d'apparition de lipodystrophie. Tous les patients de la cohorte étaient traités par HAART et la présence ou l'absence de lipodystrophie y est très bien documentée. Nous avons réalisé des tests préliminaires avec une définition « cas/témoins » à un temps t. Aucune association génétique n'a pu être clairement mise en évidence. Des études plus approfondies seront nécessaires, peut-être en distinguant plus les traitements utilisés par les patients.
- Etude de la charge virale VIH et des réponses CD4 après traitements. Le taux de lymphocytes CD4 et la charge virale VIH avant traitement étaient connus pour certains des patients. Il serait donc intéressant de rechercher les facteurs génétiques qui peuvent déterminer ces paramètres ainsi que ceux discriminant les mauvais répondeurs aux traitements par HAART.
- Etude des patients ayant achevé une SVR spontanée ou après traitement. La cohorte présentait des patients ayant achevés une SVR après traitement. Néanmoins, le faible nombre d'individus en SVR génotypés n'a pas permis de réaliser une étude statistiquement solide. Le génotypage de nouveaux individus ayant guéri spontanément ou après traitement pourrait permettre de confirmer l'impact du gène *IL28B* et d'identifier de nouveaux facteurs génétiques dans la cohorte ANRS CO13 HEPAVIH.
- Etude des charges virales VIH et VHC. Une recherche chez des patients présentant des charges virales élevées ou basses pour les deux virus pourrait dévoiler des polymorphismes influençant simultanément le contrôle des réponses aux deux virus. On retrouve là encore la possibilité d'identifier des effets interactifs liés à la coinfection.

Cinquième partie
Conclusion

1 Conclusion

Mon travail de thèse a porté sur l'étude génomique d'une cohorte unique de patients coinfectés par le virus d'immunodéficience humaine (VIH) et le virus de l'hépatite C (VHC). Cette cohorte de patients, extrêmement bien documentée, devait permettre de réaliser des analyses d'association génétique avec plusieurs paramètres biologiques et cliniques liés à ces infections.

Avant mon travail, aucune publication d'étude d'association génétique n'avait été publiée sur des patients coinfectés VIH/VHC. Compte-tenu de la complexité des données, j'ai focalisé mon travail sur un paramètre majeur de l'hépatite C, qu'est la fibrose hépatique. J'ai identifié de nouvelles associations génétiques en utilisant deux méthodologies différentes (trait quantitatif et analyse « cas/témoins »), dont une qui a été répliquée dans deux cohortes indépendantes de patients monoïnfectés VHC. Réciproquement, trois études génomiques avaient été réalisées sur la fibrose chez des patients monoïnfectés par le VHC et mon étude a permis de répliquer un des signaux identifiés.

Ces résultats montrent l'intérêt de travailler sur des cohortes ayant des phénotypes très particuliers comme ici, la coinfection par deux virus, le VIH et le VHC. Les résultats que j'ai obtenus suggèrent que ces virus ont des interactions, vues par la génétique, pour leurs effets pathogéniques. De plus, mes travaux de thèse permettent de poser de nouvelles hypothèses, améliorant la compréhension de l'étiologie moléculaire de la fibrose hépatique dans la coinfection VIH/VHC. En approfondissant ces pistes biologiques, les gènes identifiés pourraient se révéler important pour l'élaboration d'outils d'aide au diagnostic et au suivi clinique des patients coinfectés ainsi que pour suggérer de nouvelles pistes thérapeutiques. De nombreux autres phénotypes sont à explorer afin de mieux comprendre les mécanismes de la coinfection et ainsi permettre, à terme, un suivi clinique amélioré pour les patients.

Sur un plan personnel, mon travail de thèse m'a permis d'évoluer dans un domaine à l'interface de la biologie, la génétique, la statistique et la bioinformatique. Le cadre du laboratoire m'a aidé à comprendre et mettre en œuvre des méthodologies dans un domaine en perpétuelle évolution et à l'aube de sa création. Ces années de thèse très enrichissantes m'auront données l'opportunité de compléter ma formation initiale au sein d'un laboratoire dynamique à l'image de la bioinformatique.

Bibliographie

1. Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol* 2005;278:274-288.
2. Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in *Nature*, number 4356 April 25, 1953. *Nature* 1974;248:765.
3. Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* 2013;9:e1003569.
4. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073.
5. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308-311.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444-454.
7. Hardy GH. Mendelian Proportions in a Mixed Population. *Science* 1908;28:49-50.
8. Weinberg W. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 1908;64:368-382.
9. Robbins RB. Applications of Mathematics to Breeding Problems II. *Genetics* 1918;3:73-92.
10. Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 1964;49:49-67.
11. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968;38:226-231.
12. International HapMap C. The International HapMap Project. *Nature* 2003;426:789-796.
13. Myers S, Spencer CC, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G. The distribution and causes of meiotic recombination in the human genome. *Biochem Soc Trans* 2006;34:526-530.
14. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 2005;310:321-324.
15. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-2229.
16. Zhang W, Hui KY, Gusev A, Warner N, Ng SM, Ferguson J, Choi M, et al. Extended haplotype association study in Crohn's disease identifies a novel, Ashkenazi Jewish-specific missense mutation in the NF-kappaB pathway gene, HEATR3. *Genes Immun* 2013;14:310-316.
17. Clark AG. The role of haplotypes in candidate gene studies. *Genet Epidemiol* 2004;27:321-333.
18. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, et al. Genes mirror geography within Europe. *Nature* 2008;456:98-101.

19. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073-1080.
20. Luzzatto L. Sickle cell anaemia and malaria. *Mediterr J Hematol Infect Dis* 2012;4:e2012065.
21. Macdonald M, Ambrose C, Duyao M, Myers R, Lin C, Srinidhi L, Barnes G, et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* 1993;72:971-983.
22. Chawla J. Stepwise approach to myopathy in systemic disease. *Front Neurol* 2011;2:49.
23. Sadikovic B, Al-Romaih K, Squire JA, Zielenska M. Cause and consequences of genetic and epigenetic alterations in human cancer. *Curr Genomics* 2008;9:394-408.
24. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;447:433-440.
25. Dye C. After 2015: infectious diseases in a new era of health and development. *Philos Trans R Soc Lond B Biol Sci* 2014;369:20130426.
26. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;7:781-791.
27. Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989;245:1059-1065.
28. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989;245:1066-1073.
29. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. *J Clin Invest* 2005;115:1449-1457.
30. Nakamura S. [Huntington's disease--advances in gene mapping]. *Nihon Rinsho* 1993;51:2481-2487.
31. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983;306:234-238.
32. Mira MT, Alcais A, di Pietrantonio T, Thuc NV, Phuong MC, Abel L, Schurr E. Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes. *Genes Immun* 2003;4:67-73.
33. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC Genet* 2013;14:39.
34. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 1996;273:1856-1862.

35. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001-1006.
36. Gottlieb MS. Pneumocystis pneumonia--Los Angeles. 1981. *Am J Public Health* 2006;96:980-981; discussion 982-983.
37. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983;220:868-871.
38. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 1984;224:500-503.
39. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986;233:343-346.
40. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, Traore I, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* 1994;265:1587-1590.
41. Adjorlolo-Johnson G, De Cock KM, Ekpini E, Vetter KM, Sibailly T, Brattegaard K, Yavo D, et al. Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. *JAMA* 1994;272:462-466.
42. van Manen D, Kootstra NA, Boeser-Nunnink B, Handulle MA, van't Wout AB, Schuitemaker H. Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS* 2009;23:19-28.
43. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 1995;267:483-489.
44. Bister K, Jansen HW. Oncogenes in retroviruses and cells: biochemistry and molecular genetics. *Adv Cancer Res* 1986;47:99-188.
45. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet* 2008;42:709-732.
46. Swain A, Coffin JM. Mechanism of transduction by retroviruses. *Science* 1992;255:841-845.
47. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 1998;67:1-25.
48. Greene WC. The molecular biology of human immunodeficiency virus type 1 infection. *N Engl J Med* 1991;324:308-317.
49. Jones KA, Peterlin BM. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem* 1994;63:717-743.
50. Levy JA. Virus-host interactions in HIV pathogenesis: directions for therapy. *Adv Dent Res* 2011;23:13-18.
51. Levy JA. HIV pathogenesis: 25 years of progress and persistent challenges. *AIDS* 2009;23:147-160.

52. Munier S, Borjabad A, Lemaire M, Mariot V, Hazan U. In vitro infection of human primary adipose cells with HIV-1: a reassessment. *AIDS* 2003;17:2537-2539.
53. Hazan U, Romero IA, Canello R, Valente S, Perrin V, Mariot V, Dumonceaux J, et al. Human adipose cells express CD4, CXCR4, and CCR5 [corrected] receptors: a new target cell type for the immunodeficiency virus-1? *FASEB J* 2002;16:1254-1256.
54. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 1996;381:661-666.
55. Homsy J, Meyer M, Tateno M, Clarkson S, Levy JA. The Fc and not CD4 receptor mediates antibody enhancement of HIV infection in human cells. *Science* 1989;244:1357-1360.
56. Brinchmann JE, Albert J, Vartdal F. Few infected CD4+ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. *J Virol* 1991;65:2019-2023.
57. Fisher AG, Collalti E, Ratner L, Gallo RC, Wong-Staal F. A molecular clone of HTLV-III with biological activity. *Nature* 1985;316:262-265.
58. Lifson JD, Feinberg MB, Reyes GR, Rabin L, Banapour B, Chakrabarti S, Moss B, et al. Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. *Nature* 1986;323:725-728.
59. Sodroski J, Goh WC, Rosen C, Campbell K, Haseltine WA. Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity. *Nature* 1986;322:470-474.
60. Garg H, Blumenthal R. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. *Cell Mol Life Sci* 2008;65:3134-3144.
61. Plata F, Autran B, Martins LP, Wain-Hobson S, Raphael M, Mayaud C, Denis M, et al. AIDS virus-specific cytotoxic T lymphocytes in lung disorders. *Nature* 1987;328:348-351.
62. Walker BD, Chakrabarti S, Moss B, Paradis TJ, Flynn T, Durno AG, Blumberg RS, et al. HIV-specific cytotoxic T lymphocytes in seropositive individuals. *Nature* 1987;328:345-348.
63. Kion TA, Hoffmann GW. Anti-HIV and anti-anti-MHC antibodies in alloimmune and autoimmune mice. *Science* 1991;253:1138-1140.
64. Laurence J, Hodtsev AS, Posnett DN. Superantigen implicated in dependence of HIV-1 replication in T cells on TCR V beta expression. *Nature* 1992;358:255-259.
65. Dalgleish AG, Wilson S, Gompels M, Ludlam C, Gazzard B, Coates AM, Habeshaw J. T-cell receptor variable gene products and early HIV-1 infection. *Lancet* 1992;339:824-828.
66. Chowdhury MI, Munakata T, Koyanagi Y, Arai S, Yamamoto N. Mycoplasma stimulates HIV-1 expression from acutely- and dormant-lymphocyte/monoblastoid cell lines. *Arch Virol* 1994;139:431-438.
67. Chowdhury IH, Munakata T, Koyanagi Y, Kobayashi S, Arai S, Yamamoto N. Mycoplasma can enhance HIV replication in vitro: a possible cofactor responsible for the progression of AIDS. *Biochem Biophys Res Commun* 1990;170:1365-1370.
68. Gougeon ML, Montagnier L. Programmed cell death as a mechanism of CD4 and CD8 T cell deletion in AIDS. Molecular control and effect of highly active anti-retroviral therapy. *Ann N Y Acad Sci* 1999;887:199-212.

69. Estaquier J, Idziorek T, de Bels F, Barre-Sinoussi F, Hurtrel B, Aubertin AM, Venet A, et al. Programmed cell death and AIDS: significance of T-cell apoptosis in pathogenic and nonpathogenic primate lentiviral infections. *Proc Natl Acad Sci U S A* 1994;91:9431-9435.
70. Greenberg ME, Iafrate AJ, Skowronski J. The SH3 domain-binding surface and an acidic motif in HIV-1 Nef regulate trafficking of class I MHC complexes. *EMBO J* 1998;17:2777-2789.
71. Guatelli JC. Interactions of viral protein U (Vpu) with cellular factors. *Curr Top Microbiol Immunol* 2009;339:27-45.
72. Romani B, Engelbrecht S, Glashoff RH. Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *J Gen Virol* 2010;91:1-12.
73. Romagnani S, Del Prete G, Manetti R, Ravina A, Annunziato F, De Carli M, Mazzetti M, et al. Role of TH1/TH2 cytokines in HIV infection. *Immunol Rev* 1994;140:73-92.
74. Clerici M, Shearer GM. A TH1-->TH2 switch is a critical step in the etiology of HIV infection. *Immunol Today* 1993;14:107-111.
75. Marchetti G, Tincati C, Silvestri G. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clin Microbiol Rev* 2013;26:2-18.
76. Caradonna L, Amati L, Magrone T, Pellegrino NM, Jirillo E, Caccavo D. Enteric bacteria, lipopolysaccharides and related cytokines in inflammatory bowel disease: biological and clinical significance. *J Endotoxin Res* 2000;6:205-214.
77. Anderson KV. Toll signaling pathways in the innate immune response. *Curr Opin Immunol* 2000;12:13-19.
78. Weissman JB. Availability of AZT for treatment of AIDS. *N Engl J Med* 1987;316:1158.
79. Peytavin G, Calvez V, Katlama C. [CCR5 antagonists: a new class of antiretrovirals]. *Therapie* 2009;64:9-16.
80. De Clercq E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* 2009;33:307-320.
81. O'Brien SJ, Hendrickson SL. Host genomic influences on HIV/AIDS. *Genome Biol* 2013;14:201.
82. An P, Winkler CA. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet* 2010;26:119-131.
83. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 1996;382:722-725.
84. Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, Lee B, et al. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* 1998;282:1907-1911.
85. McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, Murphy PM. CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). *Lancet* 1998;352:866-870.
86. Carrington M, Martin MP, van Bergen J. KIR-HLA intercourse in HIV disease. *Trends Microbiol* 2008;16:620-627.

87. Kaslow RA, Dorak T, Tang JJ. Influence of host genetic variation on susceptibility to HIV type 1 infection. *J Infect Dis* 2005;191 Suppl 1:S68-77.
88. Magierowska M, Theodorou I, Debre P, Sanson F, Autran B, Riviere Y, Charron D, et al. Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 1999;93:936-941.
89. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317:944-947.
90. Le Clerc S, Limou S, Coulonges C, Carpentier W, Dina C, Taing L, Delaneau O, et al. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *J Infect Dis* 2009;200:1194-1201.
91. Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O, Labib T, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 2009;199:419-426.
92. Feinstone SM, Kapikian AZ, Purcell RH, Alter HJ, Holland PV. Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N Engl J Med* 1975;292:767-770.
93. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 1989;244:359-362.
94. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2095-2128.
95. Lavanchy D. Evolving epidemiology of hepatitis C virus. *Clin Microbiol Infect* 2011;17:107-115.
96. Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* 2013;57:1333-1342.
97. Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* 2003;20:381-387.
98. Fauvelle C, Lepiller Q, Felmlee DJ, Fofana I, Habersetzer F, Stoll-Keller F, Baumert TF, et al. Hepatitis C virus vaccines--progress and perspectives. *Microb Pathog* 2013;58:66-72.
99. Donahue JG, Munoz A, Ness PM, Brown DE, Jr., Yawn DH, McAllister HA, Jr., Reitz BA, et al. The declining risk of post-transfusion hepatitis C virus infection. *N Engl J Med* 1992;327:369-373.
100. van den Hoek JA, van Haastrecht HJ, Goudsmit J, de Wolf F, Coutinho RA. Prevalence, incidence, and risk factors of hepatitis C virus infection among drug users in Amsterdam. *J Infect Dis* 1990;162:823-826.
101. Ciesek S, Friesland M, Steinmann J, Becker B, Wedemeyer H, Manns MP, Steinmann J, et al. How stable is the hepatitis C virus (HCV)? Environmental stability of HCV and its susceptibility to chemical biocides. *J Infect Dis* 2010;201:1859-1866.

102. Humar A, Morris M, Blumberg E, Freeman R, Preiksaitis J, Kiberd B, Schweitzer E, et al. Nucleic acid testing (NAT) of organ donors: is the 'best' test the right test? A consensus conference report. *Am J Transplant* 2010;10:889-899.
103. Pozzetto B, Memmi M, Garraud O, Roblin X, Berthelot P. Health care-associated hepatitis C virus infection. *World J Gastroenterol* 2014;20:17265-17278.
104. Terrault NA, Dodge JL, Murphy EL, Tavis JE, Kiss A, Levin TR, Gish RG, et al. Sexual transmission of hepatitis C virus among monogamous heterosexual couples: the HCV partners study. *Hepatology* 2013;57:881-889.
105. Vandelli C, Renzo F, Romano L, Tisminetzky S, De Palma M, Stroffolini T, Ventura E, et al. Lack of evidence of sexual transmission of hepatitis C among monogamous couples: results of a 10-year prospective follow-up study. *Am J Gastroenterol* 2004;99:855-859.
106. Vellozzi C, Buchacz K, Baker R, Spradling PR, Richardson J, Moorman A, Tedaldi E, et al. Treatment of hepatitis C virus (HCV) infection in patients coinfecting with HIV in the HIV Outpatient Study (HOPS), 1999-2007. *J Viral Hepat* 2011;18:316-324.
107. Danta M, Rodger AJ. Transmission of HCV in HIV-positive populations. *Curr Opin HIV AIDS* 2011;6:451-458.
108. Roberts EA, Yeung L. Maternal-infant transmission of hepatitis C virus infection. *Hepatology* 2002;36:S106-113.
109. Penin F, Dubuisson J, Rey FA, Moradpour D, Pawlotsky JM. Structural biology of hepatitis C virus. *Hepatology* 2004;39:5-19.
110. Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. *Nat Rev Microbiol* 2007;5:453-463.
111. Kaito M, Watanabe S, Tsukiyama-Kohara K, Yamaguchi K, Kobayashi Y, Konishi M, Yokoi M, et al. Hepatitis C virus particle detected by immunoelectron microscopic study. *J Gen Virol* 1994;75 (Pt 7):1755-1760.
112. Catanese MT, Uryu K, Kopp M, Edwards TJ, Andrus L, Rice WJ, Silvestry M, et al. Ultrastructural analysis of hepatitis C virus particles. *Proc Natl Acad Sci U S A* 2013;110:9505-9510.
113. Lindenbach BD. Virion assembly and release. *Curr Top Microbiol Immunol* 2013;369:199-218.
114. Mina MM, Luciani F, Cameron B, Bull RA, Beard MR, Booth D, Lloyd AR. Resistance to hepatitis C virus: potential genetic and immunological determinants. *Lancet Infect Dis* 2015;15:451-460.
115. Da Costa D, Turek M, Felmlee DJ, Girardi E, Pfeffer S, Long G, Bartenschlager R, et al. Reconstitution of the entire hepatitis C virus life cycle in nonhepatic cells. *J Virol* 2012;86:11919-11925.
116. Cocquerel L, Voisset C, Dubuisson J. Hepatitis C virus entry: potential receptors and their biological functions. *J Gen Virol* 2006;87:1075-1084.
117. Cai L, Ji A, de Beer FC, Tannock LR, van der Westhuyzen DR. SR-BI protects against endotoxemia in mice through its roles in glucocorticoid production and hepatic clearance. *J Clin Invest* 2008;118:364-375.

118. Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, Filocamo G, Traboni C, et al. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J* 2002;21:5017-5025.
119. Wustner D, Mondal M, Huang A, Maxfield FR. Different transport routes for high density lipoprotein and its associated free sterol in polarized hepatic cells. *J Lipid Res* 2004;45:427-437.
120. Martin F, Roth DM, Jans DA, Pouton CW, Partridge LJ, Monk PN, Moseley GW. Tetraspanins in viral infections: a fundamental role in viral biology? *J Virol* 2005;79:10839-10851.
121. Zhang YY, Zhang BH, Ishii K, Liang TJ. Novel function of CD81 in controlling hepatitis C virus replication. *J Virol* 2010;84:3396-3407.
122. Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, Weiner AJ, et al. Binding of hepatitis C virus to CD81. *Science* 1998;282:938-941.
123. Mee CJ, Harris HJ, Farquhar MJ, Wilson G, Reynolds G, Davis C, van ISC, et al. Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells. *J Virol* 2009;83:6211-6221.
124. Meuleman P, Hesselgesser J, Paulson M, Vanwolleghem T, Desombere I, Reiser H, Leroux-Roels G. Anti-CD81 antibodies can prevent a hepatitis C virus infection in vivo. *Hepatology* 2008;48:1761-1768.
125. Paris L, Tonutti L, Vannini C, Bazzoni G. Structural organization of the tight junctions. *Biochim Biophys Acta* 2008;1778:646-659.
126. Yang W, Qiu C, Biswas N, Jin J, Watkins SC, Montelaro RC, Coyne CB, et al. Correlation of the tight junction-like distribution of Claudin-1 to the cellular tropism of hepatitis C virus. *J Biol Chem* 2008;283:8643-8653.
127. Fofana I, Krieger SE, Grunert F, Glaubien S, Xiao F, Fafi-Kremer S, Soulier E, et al. Monoclonal anti-claudin 1 antibodies prevent hepatitis C virus infection of primary human hepatocytes. *Gastroenterology* 2010;139:953-964, 964 e951-954.
128. Evans MJ, von Hahn T, Tscherne DM, Syder AJ, Panis M, Wolk B, Hatzioannou T, et al. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* 2007;446:801-805.
129. Meertens L, Bertaux C, Cukierman L, Cormier E, Lavillette D, Cosset FL, Dragic T. The tight junction proteins claudin-1, -6, and -9 are entry cofactors for hepatitis C virus. *J Virol* 2008;82:3555-3560.
130. Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, Rice CM. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* 2009;457:882-886.
131. Mee CJ, Grove J, Harris HJ, Hu K, Balfe P, McKeating JA. Effect of cell polarization on hepatitis C virus entry. *J Virol* 2008;82:461-470.
132. Rao R. Occludin phosphorylation in regulation of epithelial tight junctions. *Ann N Y Acad Sci* 2009;1165:62-68.
133. Butt AM, Feng D, Nasrullah I, Tahir S, Idrees M, Tong Y, Lu J. Computational identification of interplay between phosphorylation and O-beta-glycosylation of human

occludin as potential mechanism to impair hepatitis C virus entry. *Infect Genet Evol* 2012;12:1235-1245.

134. Lupberger J, Zeisel MB, Xiao F, Thumann C, Fofana I, Zona L, Davis C, et al. EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat Med* 2011;17:589-595.

135. Lindenbach BD, Rice CM. Unravelling hepatitis C virus replication from genome to function. *Nature* 2005;436:933-938.

136. Simmonds P. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* 2004;85:3173-3188.

137. Sievert W, Altraif I, Razavi HA, Abdo A, Ahmed EA, Alomair A, Amarapurkar D, et al. A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt. *Liver Int* 2011;31 Suppl 2:61-80.

138. Cornberg M, Razavi HA, Alberti A, Bernasconi E, Buti M, Cooper C, Dalgard O, et al. A systematic review of hepatitis C virus epidemiology in Europe, Canada and Israel. *Liver Int* 2011;31 Suppl 2:30-60.

139. Khatrab MA, Ferenci P, Hadziyannis SJ, Colombo M, Manns MP, Almasio PL, Esteban R, et al. Management of hepatitis C virus genotype 4: recommendations of an international expert panel. *J Hepatol* 2011;54:1250-1262.

140. Murphy DG, Sablon E, Chamberland J, Fournier E, Dandavino R, Tremblay CL. Hepatitis C virus genotype 7, a new genotype originating from central Africa. *J Clin Microbiol* 2015;53:967-972.

141. van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F, Rezza G, et al. Spread of hepatitis C virus among European injection drug users infected with HIV: a phylogenetic analysis. *J Infect Dis* 2004;189:292-302.

142. Martell M, Esteban JI, Quer J, Genesca J, Weiner A, Esteban R, Guardia J, et al. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* 1992;66:3225-3229.

143. Gill K, Ghazinian H, Manch R, Gish R. Hepatitis C virus as a systemic disease: reaching beyond the liver. *Hepatol Int* 2015.

144. Schwarz AK, Grove J, Hu K, Mee CJ, Balfe P, McKeating JA. Hepatoma cell density promotes claudin-1 and scavenger receptor BI expression and hepatitis C virus internalization. *J Virol* 2009;83:12407-12414.

145. Akazawa D, Date T, Morikawa K, Murayama A, Miyamoto M, Kaga M, Barth H, et al. CD81 expression is important for the permissiveness of Huh7 cell clones for heterogeneous hepatitis C virus infection. *J Virol* 2007;81:5036-5045.

146. Grove J, Huby T, Stamataki Z, Vanwolleghem T, Meuleman P, Farquhar M, Schwarz A, et al. Scavenger receptor BI and BII expression levels modulate hepatitis C virus infectivity. *J Virol* 2007;81:3162-3169.

147. Koutsoudakis G, Herrmann E, Kallis S, Bartenschlager R, Pietschmann T. The level of CD81 cell surface expression is a key determinant for productive entry of hepatitis C virus into host cells. *J Virol* 2007;81:588-598.

148. Cacoub P, Gragnani L, Comarmond C, Zignego AL. Extrahepatic manifestations of chronic hepatitis C virus infection. *Dig Liver Dis* 2014;46 Suppl 5:S165-173.

149. Vargas HE, Laskus T, Radkowski M, Wilkinson J, Balan V, Douglas DD, Harrison ME, et al. Detection of hepatitis C virus sequences in brain tissue obtained in recurrent hepatitis C after liver transplantation. *Liver Transpl* 2002;8:1014-1019.
150. Maggi F, Giorgi M, Fornai C, Morrica A, Vatteroni ML, Pistello M, Siciliano G, et al. Detection and quasispecies analysis of hepatitis C virus in the cerebrospinal fluid of infected patients. *J Neurovirol* 1999;5:319-323.
151. Bolay H, Soylemezoglu F, Nurlu G, Tuncer S, Varli K. PCR detected hepatitis C virus genome in the brain of a case with progressive encephalomyelitis with rigidity. *Clin Neurol Neurosurg* 1996;98:305-308.
152. Burgel B, Friesland M, Koch A, Manns MP, Wedemeyer H, Weissenborn K, Schulz-Schaeffer WJ, et al. Hepatitis C virus enters human peripheral neuroblastoma cells - evidence for extra-hepatic cells sustaining hepatitis C virus penetration. *J Viral Hepat* 2011;18:562-570.
153. Fletcher NF, Yang JP, Farquhar MJ, Hu K, Davis C, He Q, Dowd K, et al. Hepatitis C virus infection of neuroepithelioma cell lines. *Gastroenterology* 2010;139:1365-1374.
154. Rodrigue-Gervais IG, Jouan L, Beaulieu G, Sauve D, Bruneau J, Willems B, Sekaly RP, et al. Poly(I:C) and lipopolysaccharide innate sensing functions of circulating human myeloid dendritic cells are affected in vivo in hepatitis C virus-infected patients. *J Virol* 2007;81:5537-5546.
155. Fournillier A, Freida D, Defrance T, Merle P, Trepo C, Inchauspe G. Analysis of B-lymphocyte differentiation in patients infected with hepatitis C virus. *J Med Virol* 2004;72:566-574.
156. Goutagny N, Fatmi A, De Ledinghen V, Penin F, Couzigou P, Inchauspe G, Bain C. Evidence of viral replication in circulating dendritic cells during hepatitis C virus infection. *J Infect Dis* 2003;187:1951-1958.
157. Laporte J, Bain C, Maurel P, Inchauspe G, Agut H, Cahour A. Differential distribution and internal translation efficiency of hepatitis C virus quasispecies present in dendritic and liver cells. *Blood* 2003;101:52-57.
158. Navas MC, Fuchs A, Schvoerer E, Bohbot A, Aubertin AM, Stoll-Keller F. Dendritic cell susceptibility to hepatitis C virus genotype 1 infection. *J Med Virol* 2002;67:152-161.
159. Caussin-Schwemling C, Schmitt C, Stoll-Keller F. Study of the infection of human blood derived monocyte/macrophages with hepatitis C virus in vitro. *J Med Virol* 2001;65:14-22.
160. Marukian S, Jones CT, Andrus L, Evans MJ, Ritola KD, Charles ED, Rice CM, et al. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* 2008;48:1843-1850.
161. Boisvert J, He XS, Cheung R, Keefe EB, Wright T, Greenberg HB. Quantitative analysis of hepatitis C virus in peripheral blood and liver: replication detected only in liver. *J Infect Dis* 2001;184:827-835.
162. Zehender G, Meroni L, De Maddalena C, Varchetta S, Monti G, Galli M. Detection of hepatitis C virus RNA in CD19 peripheral blood mononuclear cells of chronically infected patients. *J Infect Dis* 1997;176:1209-1214.
163. Fujiwara K, Allison RD, Wang RY, Bare P, Matsuura K, Schechterly C, Murthy K, et al. Investigation of residual hepatitis C virus in presumed recovered subjects. *Hepatology* 2013;57:483-491.

164. Di Lello FA, Culasso AC, Parodi C, Bare P, Campos RH, Garcia G. New evidence of replication of hepatitis C virus in short-term peripheral blood mononuclear cell cultures. *Virus Res* 2014;191:1-9.
165. Gong GZ, Lai LY, Jiang YF, He Y, Su XS. HCV replication in PBMC and its influence on interferon therapy. *World J Gastroenterol* 2003;9:291-294.
166. Welker MW, Zeuzem S. Occult hepatitis C: how convincing are the current data? *Hepatology* 2009;49:665-675.
167. European Association for the Study of the L. EASL Clinical Practice Guidelines: management of hepatitis C virus infection. *J Hepatol* 2011;55:245-264.
168. Santantonio T, Wiegand J, Gerlach JT. Acute hepatitis C: current status and remaining challenges. *J Hepatol* 2008;49:625-633.
169. Orland JR, Wright TL, Cooper S. Acute hepatitis C. *Hepatology* 2001;33:321-327.
170. Marcellin P, Pequignot F, Delarocque-Astagneau E, Zarski JP, Ganne N, Hillon P, Antona D, et al. Mortality related to chronic hepatitis B and chronic hepatitis C in France: evidence for the role of HIV coinfection and alcohol consumption. *J Hepatol* 2008;48:200-207.
171. Craxi A, Laffi G, Zignego AL. Hepatitis C virus (HCV) infection: a systemic disease. *Mol Aspects Med* 2008;29:85-95.
172. Seeff LB. Natural history of chronic hepatitis C. *Hepatology* 2002;36:S35-46.
173. Machado MV, Cortez-Pinto H. Insulin resistance and steatosis in chronic hepatitis C. *Ann Hepatol* 2009;8 Suppl 1:S67-75.
174. Catanese MT, Dorner M. Advances in experimental systems to study hepatitis C virus in vitro and in vivo. *Virology* 2015;479-480:221-233.
175. Hoshida Y, Fuchs BC, Bardeesy N, Baumert TF, Chung RT. Pathogenesis and prevention of hepatitis C virus-induced hepatocellular carcinoma. *J Hepatol* 2014;61:S79-90.
176. Gokhale NS, Vazquez C, Horner SM. Hepatitis C Virus. Strategies to Evade Antiviral Responses. *Future Virol* 2014;9:1061-1075.
177. Bantel H, Schulze-Osthoff K. Apoptosis in hepatitis C virus infection. *Cell Death Differ* 2003;10 Suppl 1:S48-58.
178. Horner SM, Gale M, Jr. Regulation of hepatic innate immunity by hepatitis C virus. *Nat Med* 2013;19:879-888.
179. Yoon JC, Shiina M, Ahlenstiel G, Rehmann B. Natural killer cell function is intact after direct exposure to infectious hepatitis C virions. *Hepatology* 2009;49:12-21.
180. Tsai SL, Liaw YF, Chen MH, Huang CY, Kuo GC. Detection of type 2-like T-helper cells in hepatitis C virus infection: implications for hepatitis C virus chronicity. *Hepatology* 1997;25:449-458.
181. Neumann-Haefelin C, Blum HE, Chisari FV, Thimme R. T cell response in hepatitis C virus infection. *J Clin Virol* 2005;32:75-85.
182. Schnuriger A, Dominguez S, Guiguet M, Harfouch S, Samri A, Ouazene Z, Slama L, et al. Acute hepatitis C in HIV-infected patients: rare spontaneous clearance correlates with weak memory CD4 T-cell responses to hepatitis C virus. *AIDS* 2009;23:2079-2089.
183. Zeisel MB, Cosset FL, Baumert TF. Host neutralizing responses and pathogenesis of hepatitis C virus infection. *Hepatology* 2008;48:299-307.

184. Zhang P, Wu CG, Mihalik K, Virata-Theimer ML, Yu MY, Alter HJ, Feinstone SM. Hepatitis C virus epitope-specific neutralizing antibodies in Igs prepared from human plasma. *Proc Natl Acad Sci U S A* 2007;104:8449-8454.
185. von Hahn T, Yoon JC, Alter H, Rice CM, Rehermann B, Balfe P, McKeating JA. Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. *Gastroenterology* 2007;132:667-678.
186. Pestka JM, Zeisel MB, Blaser E, Schurmann P, Bartosch B, Cosset FL, Patel AH, et al. Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. *Proc Natl Acad Sci U S A* 2007;104:6025-6030.
187. Lavillette D, Morice Y, Germanidis G, Donot P, Soulier A, Pagkalos E, Sakellariou G, et al. Human serum facilitates hepatitis C virus infection, and neutralizing responses inversely correlate with viral replication kinetics at the acute phase of hepatitis C virus infection. *J Virol* 2005;79:6023-6034.
188. Fiel MI. Pathology of chronic hepatitis B and chronic hepatitis C. *Clin Liver Dis* 2010;14:555-575.
189. Sobue S, Nomura T, Ishikawa T, Ito S, Saso K, Ohara H, Joh T, et al. Th1/Th2 cytokine profiles and their relationship to clinical features in patients with chronic hepatitis C virus infection. *J Gastroenterol* 2001;36:544-551.
190. Lechmann M, Woitas RP, Langhans B, Kaiser R, Ihlenfeldt HG, Jung G, Sauerbruch T, et al. Decreased frequency of HCV core-specific peripheral blood mononuclear cells with type 1 cytokine secretion in chronic hepatitis C. *J Hepatol* 1999;31:971-978.
191. Koziel MJ, Dudley D, Afdhal N, Grakoui A, Rice CM, Choo QL, Houghton M, et al. HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. *J Clin Invest* 1995;96:2311-2321.
192. Fallahi P, Ferri C, Ferrari SM, Corrado A, Sansonno D, Antonelli A. Cytokines and HCV-related disorders. *Clin Dev Immunol* 2012;2012:468107.
193. Tanaka S, Takenaka K, Matsumata T, Mori R, Sugimachi K. Hepatitis C virus replication is associated with expression of transforming growth factor-alpha and insulin-like growth factor-II in cirrhotic livers. *Dig Dis Sci* 1996;41:208-215.
194. Bertolotti A, D'Elios MM, Boni C, De Carli M, Zignego AL, Durazzo M, Missale G, et al. Different cytokine profiles of intraphepatic T cells in chronic hepatitis B and hepatitis C virus infections. *Gastroenterology* 1997;112:193-199.
195. Khan MH, Farrell GC, Byth K, Lin R, Weltman M, George J, Samarasinghe D, et al. Which patients with hepatitis C develop liver complications? *Hepatology* 2000;31:513-520.
196. Schuppan D. Structure of the extracellular matrix in normal and fibrotic liver: collagens and glycoproteins. *Semin Liver Dis* 1990;10:1-10.
197. Friedman SL. Mechanisms of hepatic fibrogenesis. *Gastroenterology* 2008;134:1655-1669.
198. Parsons CJ, Takashima M, Rippe RA. Molecular mechanisms of hepatic fibrogenesis. *J Gastroenterol Hepatol* 2007;22 Suppl 1:S79-84.
199. Pilette C, Fort J, Calès P. Traitement pharmacologique de la fibrose hépatique : bientôt disponible ? . *Hépatogastro & Oncologie Digestive* 1998;5:223-230.

200. Clement B, Grimaud JA, Campion JP, Deugnier Y, Guillouzo A. Cell types involved in collagen and fibronectin production in normal and fibrotic human liver. *Hepatology* 1986;6:225-234.
201. Geerts A, Greenwel P, Cunningham M, De Bleser P, Rogiers V, Wisse E, Rojkind M. Identification of connective tissue gene transcripts in freshly isolated parenchymal, endothelial, Kupffer and fat-storing cells by northern hybridization analysis. *J Hepatol* 1993;19:148-158.
202. Rescan PY, Loreal O, Hassell JR, Yamada Y, Guillouzo A, Clement B. Distribution and origin of the basement membrane component perlecan in rat liver and primary hepatocyte culture. *Am J Pathol* 1993;142:199-208.
203. Milani S, Herbst H, Schuppan D, Kim KY, Riecken EO, Stein H. Procollagen expression by nonparenchymal rat liver cells in experimental biliary fibrosis. *Gastroenterology* 1990;98:175-184.
204. Mazzocca A, Carloni V, Sciammetta S, Cordella C, Pantaleo P, Caldini A, Gentilini P, et al. Expression of transmembrane 4 superfamily (TM4SF) proteins and their role in hepatic stellate cell motility and wound healing migration. *J Hepatol* 2002;37:322-330.
205. Florimond A, Chouteau P, Bruscella P, Le Seyec J, Merour E, Ahnou N, Mallat A, et al. Human hepatic stellate cells are not permissive for hepatitis C virus entry and replication. *Gut* 2015;64:957-965.
206. Seki E, Schwabe RF. Hepatic inflammation and fibrosis: functional links and key pathways. *Hepatology* 2015;61:1066-1079.
207. Pinzani M, Knauss TC, Pierce GF, Hsieh P, Kenney W, Dubyak GR, Abboud HE. Mitogenic signals for platelet-derived growth factor isoforms in liver fat-storing cells. *Am J Physiol* 1991;260:C485-491.
208. Ma L. Pathology Features and Molecular Genetic Mechanisms of Hepatocellular Carcinoma Development in Patients with Hepatitis C Associated Liver Cirrhosis. *Hereditary Genet* 2014;3:e109.
209. Block TM, Mehta AS, Fimmel CJ, Jordan R. Molecular viral oncology of hepatocellular carcinoma. *Oncogene* 2003;22:5093-5107.
210. Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, Matsuura Y, et al. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat Med* 1998;4:1065-1067.
211. Lai MM, Ware CF. Hepatitis C virus core protein: possible roles in viral pathogenesis. *Curr Top Microbiol Immunol* 2000;242:117-134.
212. Zhao LJ, Wang L, Ren H, Cao J, Li L, Ke JS, Qi ZT. Hepatitis C virus E2 protein promotes human hepatoma cell proliferation through the MAPK/ERK signaling pathway via cellular receptors. *Exp Cell Res* 2005;305:23-32.
213. Tseng CT, Klimpel GR. Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. *J Exp Med* 2002;195:43-49.
214. Plentz RR, Park YN, Lechel A, Kim H, Nellessen F, Langkopf BH, Wilkens L, et al. Telomere shortening and inactivation of cell cycle checkpoints characterize human hepatocarcinogenesis. *Hepatology* 2007;45:968-976.
215. Lu W, Lo SY, Chen M, Wu K, Fung YK, Ou JH. Activation of p53 tumor suppressor by hepatitis C virus core protein. *Virology* 1999;264:134-141.

216. Biron-Shental T, Amiel A, Anchidin R, Sharony R, Hadary R, Kitay-Cohen Y. Telomere length and telomerase reverse transcriptase mRNA expression in patients with hepatitis C. *Hepatology* 2013;60:1713-1716.
217. Wiemann SU, Satyanarayana A, Tsahuridu M, Tillmann HL, Zender L, Klempnauer J, Flemming P, et al. Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis. *FASEB J* 2002;16:935-942.
218. Jang HS, Kang KM, Choi BO, Chai GY, Hong SC, Ha WS, Jirtle RL. Clinical significance of loss of heterozygosity for M6P/IGF2R in patients with primary hepatocellular carcinoma. *World J Gastroenterol* 2008;14:1394-1398.
219. Dore MP, Realdi G, Mura D, Onida A, Massarelli G, Dettori G, Graham DY, et al. Genomic instability in chronic viral hepatitis and hepatocellular carcinoma. *Hum Pathol* 2001;32:698-703.
220. Kondoh N, Wakatsuki T, Hada A, Shuda M, Tanaka K, Arai M, Yamamoto M. Genetic and epigenetic events in human hepatocarcinogenesis. *Int J Oncol* 2001;18:1271-1278.
221. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* 2012;44:694-698.
222. Tornesello ML, Buonaguro L, Tatangelo F, Botti G, Izzo F, Buonaguro FM. Mutations in TP53, CTNNB1 and PIK3CA genes in hepatocellular carcinoma associated with hepatitis B and hepatitis C virus infections. *Genomics* 2013;102:74-83.
223. Luxon BA, Grace M, Brassard D, Bordens R. Pegylated interferons for the treatment of chronic hepatitis C infection. *Clin Ther* 2002;24:1363-1383.
224. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461:399-401.
225. Le Page C, Genin P, Baines MG, Hiscott J. Interferon activation and innate immunity. *Rev Immunogenet* 2000;2:374-386.
226. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol* 2011;1:519-525.
227. Sadler AJ, Williams BR. Interferon-inducible antiviral effectors. *Nat Rev Immunol* 2008;8:559-568.
228. Gibbert K, Schlaak JF, Yang D, Dittmer U. IFN-alpha subtypes: distinct biological activities in anti-viral therapy. *Br J Pharmacol* 2013;168:1048-1058.
229. Gerlach N, Gibbert K, Alter C, Nair S, Zelinskyy G, James CM, Dittmer U. Anti-retroviral effects of type I IFN subtypes in vivo. *Eur J Immunol* 2009;39:136-146.
230. Witkowski JT, Robins RK, Sidwell RW, Simon LN. Design, synthesis, and broad spectrum antiviral activity of 1-*D*-ribofuranosyl-1,2,4-triazole-3-carboxamide and related nucleosides. *J Med Chem* 1972;15:1150-1154.
231. Sidwell RW, Huffman JH, Khare GP, Allen LB, Witkowski JT, Robins RK. Broad-spectrum antiviral activity of Virazole: 1- β -*D*-ribofuranosyl-1,2,4-triazole-3-carboxamide. *Science* 1972;177:705-706.

232. Carvalho OV, Saraiva GL, Ferreira CG, Felix DM, Fietto JL, Bressan GC, Almeida MR, et al. In-vitro antiviral efficacy of ribavirin and interferon-alpha against canine distemper virus. *Can J Vet Res* 2014;78:283-289.
233. Kamar N, Rostaing L, Abravanel F, Garrouste C, Lhomme S, Esposito L, Basse G, et al. Ribavirin therapy inhibits viral replication on patients with chronic hepatitis e virus infection. *Gastroenterology* 2010;139:1612-1618.
234. Fried MW, Fong TL, Swain MG, Park Y, Beames MP, Banks SM, Hoofnagle JH, et al. Therapy of chronic hepatitis B with a 6-month course of ribavirin. *J Hepatol* 1994;21:145-150.
235. Kakumu S, Yoshioka K, Wakita T, Ishikawa T, Takayanagi M, Higashi Y. Pilot study of ribavirin and interferon-beta for chronic hepatitis B. *Hepatology* 1993;18:258-263.
236. Parker WB. Metabolism and antiviral activity of ribavirin. *Virus Res* 2005;107:165-171.
237. Pawlotsky JM, Dahari H, Neumann AU, Hezode C, Germanidis G, Lonjon I, Castera L, et al. Antiviral action of ribavirin in chronic hepatitis C. *Gastroenterology* 2004;126:703-714.
238. Ibarra KD, Pfeiffer JK. Reduced ribavirin antiviral efficacy via nucleoside transporter-mediated drug resistance. *J Virol* 2009;83:4538-4547.
239. Dixit NM, Perelson AS. The metabolism, pharmacokinetics and mechanisms of antiviral activity of ribavirin against hepatitis C virus. *Cell Mol Life Sci* 2006;63:832-842.
240. Lau JY, Tam RC, Liang TJ, Hong Z. Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. *Hepatology* 2002;35:1002-1009.
241. Feld JJ, Hoofnagle JH. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* 2005;436:967-972.
242. Casey LC, Lee WM. Hepatitis C therapy update. *Curr Opin Gastroenterol* 2012;28:188-192.
243. Jacobson IM, McHutchison JG, Dusheiko G, Di Bisceglie AM, Reddy KR, Bzowej NH, Marcellin P, et al. Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med* 2011;364:2405-2416.
244. Poordad F, McCone J, Jr., Bacon BR, Bruno S, Manns MP, Sulkowski MS, Jacobson IM, et al. Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med* 2011;364:1195-1206.
245. Zhang X. Direct anti-HCV agents. *Acta Pharm Sin B* 2016;6:26-31.
246. Degasperi E, Aghemo A. Sofosbuvir for the treatment of chronic hepatitis C: between current evidence and future perspectives. *Hepat Med* 2014;6:25-33.
247. Sulkowski MS, Kang M, Matining R, Wyles D, Johnson VA, Morse GD, Amorosa V, et al. Safety and antiviral activity of the HCV entry inhibitor ITX5061 in treatment-naive HCV-infected adults: a randomized, double-blind, phase 1b study. *J Infect Dis* 2014;209:658-667.
248. Fofana I, Xiao F, Thumann C, Turek M, Zona L, Tawar RG, Grunert F, et al. A novel monoclonal anti-CD81 antibody produced by genetic immunization efficiently inhibits Hepatitis C virus cell-cell transmission. *PLoS One* 2013;8:e64221.
249. Lacey K, Vercauteren K, Grzyb K, Naddeo M, Verhoye L, Slowikowski MP, Fafi-Kremer S, et al. Novel human SR-BI antibodies prevent infection and dissemination of HCV in vitro and in humanized mice. *J Hepatol* 2012;57:17-23.

250. Matsuura K, Tanaka Y. Host genetic variants influencing the clinical course of hepatitis C virus infection. *J Med Virol* 2016;88:185-195.
251. Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, Bochud M, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* 2010;138:1338-1345, 1345 e1331-1337.
252. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, Bassendine M, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 2009;41:1100-1104.
253. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 2009;41:1105-1109.
254. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, Kidd J, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 2009;461:798-801.
255. Grebely J, Raffa JD, Lai C, Krajden M, Conway B, Tyndall MW. Factors associated with spontaneous clearance of hepatitis C virus among illicit drug users. *Can J Gastroenterol* 2007;21:447-451.
256. Busch MP, Glynn SA, Stramer SL, Orland J, Murphy EL, Wright DJ, Kleinman S, et al. Correlates of hepatitis C virus (HCV) RNA negativity among HCV-seropositive blood donors. *Transfusion* 2006;46:469-475.
257. Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N, Nolt K, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA* 2000;284:450-456.
258. Prokunina-Olsson L, Muchmore B, Tang W, Pfeiffer RM, Park H, Dickensheets H, Hergott D, et al. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. *Nat Genet* 2013;45:164-171.
259. Duggal P, Thio CL, Wojcik GL, Goedert JJ, Mangia A, Latanich R, Kim AY, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. *Ann Intern Med* 2013;158:235-245.
260. Mangia A, Gentile R, Cascavilla I, Margaglione M, Villani MR, Stella F, Modola G, et al. HLA class II favors clearance of HCV infection and progression of the chronic liver damage. *J Hepatol* 1999;30:984-989.
261. Cramp ME, Carucci P, Underhill J, Naoumov NV, Williams R, Donaldson PT. Association between HLA class II genotype and spontaneous clearance of hepatitis C viraemia. *J Hepatol* 1998;29:207-213.
262. Falletti E, Fabris C, Vandelli C, Colletta C, Cussigh A, Smirne C, Fontanini E, et al. Genetic polymorphisms of interleukin-6 modulate fibrosis progression in mild chronic hepatitis C. *Hum Immunol* 2010;71:999-1004.
263. Hellier S, Frodsham AJ, Hennig BJ, Klenerman P, Knapp S, Ramaley P, Satsangi J, et al. Association of genetic variants of the chemokine receptor CCR5 and its ligands, RANTES and MCP-2, with outcome of HCV infection. *Hepatology* 2003;38:1468-1476.
264. Muhlbauer M, Bosserhoff AK, Hartmann A, Thasler WE, Weiss TS, Herfarth H, Lock G, et al. A novel MCP-1 gene polymorphism is associated with hepatic MCP-1 expression and severity of HCV-related liver disease. *Gastroenterology* 2003;125:1085-1093.

265. Yee LJ, Tang J, Herrera J, Kaslow RA, van Leeuwen DJ. Tumor necrosis factor gene polymorphisms in patients with cirrhosis from chronic hepatitis C virus infection. *Genes Immun* 2000;1:386-390.
266. Powell EE, Edwards-Smith CJ, Hay JL, Clouston AD, Crawford DH, Shorthouse C, Purdie DM, et al. Host genetic factors influence disease progression in chronic hepatitis C. *Hepatology* 2000;31:828-833.
267. Yu RB, Hong X, Ding WL, Tan YF, Zhang YX, Sun NX, Wu GL, et al. The association between the genetic polymorphism of HLA-DQA1, DQB1, and DRB1 and serum alanine aminotransferase levels in chronic hepatitis C in the Chinese population. *J Gastroenterol Hepatol* 2008;23:1394-1402.
268. Yoshizawa K, Ota M, Saito S, Maruyama A, Yamaura T, Rokuhara A, Orii K, et al. Long-term follow-up of hepatitis C virus infection: HLA class II loci influences the natural history of the disease. *Tissue Antigens* 2003;61:159-165.
269. Fabris C, Falletti E, Cussigh A, Bitetto D, Fontanini E, Bignulin S, Cmet S, et al. IL-28B rs12979860 C/T allele distribution in patients with liver cirrhosis: role in the course of chronic viral hepatitis and the development of HCC. *J Hepatol* 2011;54:716-722.
270. Abe H, Ochi H, Maekawa T, Hayes CN, Tsuge M, Miki D, Mitsui F, et al. Common variation of IL28 affects gamma-GTP levels and inflammation of the liver in chronically infected hepatitis C virus patients. *J Hepatol* 2010;53:439-443.
271. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle E, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2008;40:1461-1465.
272. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat Genet* 2010;42:21-23.
273. Trepo E, Pradat P, Potthoff A, Momozawa Y, Quertinmont E, Gustot T, Lemmers A, et al. Impact of patatin-like phospholipase-3 (rs738409 C>G) polymorphism on fibrosis progression and steatosis in chronic hepatitis C. *Hepatology* 2011;54:60-69.
274. Patin E, Kutalik Z, Guergnon J, Bibert S, Nalpas B, Jouanguy E, Munteanu M, et al. Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. *Gastroenterology* 2012;143:1244-1252 e1241-1212.
275. Peng Q, Lao X, Chen Z, Lai H, Deng Y, Wang J, Mo C, et al. TP53 and MDM2 gene polymorphisms, gene-gene interaction, and hepatocellular carcinoma risk: evidence from an updated meta-analysis. *PLoS One* 2013;8:e82773.
276. Abu Dayyeh BK, Yang M, Fuchs BC, Karl DL, Yamada S, Sninsky JJ, O'Brien TR, et al. A functional polymorphism in the epidermal growth factor gene is associated with risk for hepatocellular carcinoma. *Gastroenterology* 2011;141:141-149.
277. Tanabe KK, Lemoine A, Finkelstein DM, Kawasaki H, Fujii T, Chung RT, Lauwers GY, et al. Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. *JAMA* 2008;299:53-60.
278. Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, Aikata H, Ikeda K, et al. Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. *Nat Genet* 2011;43:797-800.

279. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, Otsuka M, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011;43:455-458.
280. Sagiv A, Burton DG, Moshayev Z, Vadai E, Wensveen F, Ben-Dor S, Golani O, et al. NKG2D ligands mediate immunosurveillance of senescent cells. *Aging (Albany NY)* 2016.
281. Lange CM, Bibert S, Dufour JF, Cellerai C, Cerny A, Heim MH, Kaiser L, et al. Comparative genetic analyses point to HCP5 as susceptibility locus for HCV-associated hepatocellular carcinoma. *J Hepatol* 2013;59:504-509.
282. Taylor LE, Swan T, Mayer KH. HIV coinfection with hepatitis C virus: evolving epidemiology and treatment paradigms. *Clin Infect Dis* 2012;55 Suppl 1:S33-42.
283. Bica I, McGovern B, Dhar R, Stone D, McGowan K, Scheib R, Snyderman DR. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis* 2001;32:492-497.
284. Di Martino V, Rufat P, Boyer N, Renard P, Degos F, Martinot-Peignoux M, Matheron S, et al. The influence of human immunodeficiency virus coinfection on chronic hepatitis C in injection drug users: a long-term retrospective cohort study. *Hepatology* 2001;34:1193-1199.
285. Yee TT, Griffioen A, Sabin CA, Dusheiko G, Lee CA. The natural history of HCV in a cohort of haemophilic patients infected between 1961 and 1985. *Gut* 2000;47:845-851.
286. Tedaldi EM, Hullsiek KH, Malvestutto CD, Arduino RC, Fisher EJ, Gaglio PJ, Jenny-Avital ER, et al. Prevalence and characteristics of hepatitis C virus coinfection in a human immunodeficiency virus clinical trials group: the Terry Bein Community Programs for Clinical Research on AIDS. *Clin Infect Dis* 2003;36:1313-1317.
287. Thomson EC, Smith JA, Klenerman P. The natural history of early hepatitis C virus evolution; lessons from a global outbreak in human immunodeficiency virus-1-infected individuals. *J Gen Virol* 2011;92:2227-2236.
288. Ghosn J, Pierre-Francois S, Thibault V, Duvivier C, Tubiana R, Simon A, Valantin MA, et al. Acute hepatitis C in HIV-infected men who have sex with men. *HIV Med* 2004;5:303-306.
289. Beld M, Penning M, Lukashov V, McMorrow M, Roos M, Pakker N, van den Hoek A, et al. Evidence that both HIV and HIV-induced immunodeficiency enhance HCV replication among HCV seroconverters. *Virology* 1998;244:504-512.
290. Eyster ME, Fried MW, Di Bisceglie AM, Goedert JJ. Increasing hepatitis C virus RNA levels in hemophiliacs: relationship to human immunodeficiency virus infection and liver disease. Multicenter Hemophilia Cohort Study. *Blood* 1994;84:1020-1023.
291. Daar ES, Lynn H, Donfield S, Gomperts E, Hilgartner MW, Hoots WK, Chernoff D, et al. Relation between HIV-1 and hepatitis C viral load in patients with hemophilia. *J Acquir Immune Defic Syndr* 2001;26:466-472.
292. Bonacini M, Govindarajan S, Blatt LM, Schmid P, Conrad A, Lindsay KL. Patients co-infected with human immunodeficiency virus and hepatitis C virus demonstrate higher levels of hepatic HCV RNA. *J Viral Hepat* 1999;6:203-208.
293. Bower WA, Culver DH, Castor D, Wu Y, James VN, Zheng H, Hammer S, et al. Changes in hepatitis C virus (HCV) viral load and interferon-alpha levels in HIV/HCV-coinfected patients treated with highly active antiretroviral therapy. *J Acquir Immune Defic Syndr* 2006;42:293-297.

294. Chung RT, Evans SR, Yang Y, Theodore D, Valdez H, Clark R, Shikuma C, et al. Immune recovery is associated with persistent rise in hepatitis C virus RNA, infrequent liver test flares, and is not impaired by hepatitis C virus in co-infected subjects. *AIDS* 2002;16:1915-1923.
295. Ananthkrishnan AN, McGinley EL, Fangman J, Saeian K. Hepatitis C/HIV co-infection is associated with higher mortality in hospitalized patients with hepatitis C or HIV. *J Viral Hepat* 2010;17:720-729.
296. Rotman Y, Liang TJ. Coinfection with hepatitis C virus and human immunodeficiency virus: virological, immunological, and clinical outcomes. *J Virol* 2009;83:7366-7374.
297. Salmon-Ceron D, Rosenthal E, Lewden C, Bouteloup V, May T, Burty C, Bonnet F, et al. Emerging role of hepatocellular carcinoma among liver-related causes of deaths in HIV-infected patients: The French national Mortalite 2005 study. *J Hepatol* 2009;50:736-745.
298. Fierer DS, Uriel AJ, Carriero DC, Klepper A, Dieterich DT, Mullen MP, Thung SN, et al. Liver fibrosis during an outbreak of acute hepatitis C virus infection in HIV-infected men: a prospective cohort study. *J Infect Dis* 2008;198:683-686.
299. Thein HH, Yi Q, Dore GJ, Krahn MD. Natural history of hepatitis C virus infection in HIV-infected individuals and the impact of HIV in the era of highly active antiretroviral therapy: a meta-analysis. *AIDS* 2008;22:1979-1991.
300. Puoti M, Prestini K, Putzolu V, Zanini B, Baiguera C, Antonini MG, Pagani P, et al. HIV/HCV co-infection: natural history. *J Biol Regul Homeost Agents* 2003;17:144-146.
301. Hernandez MD, Sherman KE. HIV/hepatitis C coinfection natural history and disease progression. *Curr Opin HIV AIDS* 2011;6:478-482.
302. Greub G, Ledergerber B, Battegay M, Grob P, Perrin L, Furrer H, Burgisser P, et al. Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV Cohort Study. *Lancet* 2000;356:1800-1805.
303. Sulkowski MS, Moore RD, Mehta SH, Chaisson RE, Thomas DL. Hepatitis C and progression of HIV disease. *JAMA* 2002;288:199-206.
304. Brookmeyer R, Gail MH, Polk BF. The prevalent cohort study and the acquired immunodeficiency syndrome. *Am J Epidemiol* 1987;126:14-24.
305. Miller MF, Haley C, Koziel MJ, Rowley CF. Impact of hepatitis C virus on immune restoration in HIV-infected patients who start highly active antiretroviral therapy: a meta-analysis. *Clin Infect Dis* 2005;41:713-720.
306. Chen JY, Feeney ER, Chung RT. HCV and HIV co-infection: mechanisms and management. *Nat Rev Gastroenterol Hepatol* 2014;11:362-371.
307. Lin W, Weinberg EM, Tai AW, Peng LF, Brockman MA, Kim KA, Kim SS, et al. HIV increases HCV replication in a TGF-beta1-dependent manner. *Gastroenterology* 2008;134:803-811.
308. Babu CK, Suwansrinon K, Bren GD, Badley AD, Rizza SA. HIV induces TRAIL sensitivity in hepatocytes. *PLoS One* 2009;4:e4623.
309. Tuyama AC, Hong F, Saiman Y, Wang C, Ozkok D, Mosoian A, Chen P, et al. Human immunodeficiency virus (HIV)-1 infects human hepatic stellate cells and promotes collagen I

and monocyte chemoattractant protein-1 expression: implications for the pathogenesis of HIV/hepatitis C virus-induced liver fibrosis. *Hepatology* 2010;52:612-622.

310. Glassner A, Eisenhardt M, Kokordelis P, Kramer B, Wolter F, Nischalke HD, Boesecke C, et al. Impaired CD4(+) T cell stimulation of NK cell anti-fibrotic activity may contribute to accelerated liver fibrosis progression in HIV/HCV patients. *J Hepatol* 2013;59:427-433.

311. Armitage AE, Stacey AR, Giannoulatou E, Marshall E, Sturges P, Chatha K, Smith NM, et al. Distinct patterns of hepcidin and iron regulation during HIV-1, HBV, and HCV infections. *Proc Natl Acad Sci U S A* 2014;111:12187-12192.

312. Liu Y, Lv Q, Gao J, Long L, Duan Z, Liang H, Shen T, et al. Coinfection with HIV-1 alleviates iron accumulation in patients with chronic hepatitis C virus infection. *PLoS One* 2014;9:e98039.

313. Martinelli AL, Ramalho LN, Zucoloto S. Hepatic stellate cells in hepatitis C patients: relationship with liver iron deposits and severity of liver disease. *J Gastroenterol Hepatol* 2004;19:91-98.

314. Danta M, Semmo N, Fabris P, Brown D, Pybus OG, Sabin CA, Bhagani S, et al. Impact of HIV on host-virus interactions during early hepatitis C virus infection. *J Infect Dis* 2008;197:1558-1566.

315. Gonzalez-Peralta RP, Qian K, She JY, Davis GL, Ohno T, Mizokami M, Lau JY. Clinical implications of viral quasispecies heterogeneity in chronic hepatitis C. *J Med Virol* 1996;49:242-247.

316. Balagopal A, Philp FH, Astemborski J, Block TM, Mehta A, Long R, Kirk GD, et al. Human immunodeficiency virus-related microbial translocation and progression of hepatitis C. *Gastroenterology* 2008;135:226-233.

317. Laguno M, Murillas J, Blanco JL, Martinez E, Miquel R, Sanchez-Tapias JM, Bargallo X, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for treatment of HIV/HCV co-infected patients. *AIDS* 2004;18:F27-36.

318. Lissen E, Clumeck N, Sola R, Mendes-Correa M, Montaner J, Nelson M, DePamphilis J, et al. Histological response to pegIFNalpha-2a (40KD) plus ribavirin in HIV-hepatitis C virus co-infection. *AIDS* 2006;20:2175-2181.

319. Soriano V, Maida I, Nunez M, Garcia-Samaniego J, Barreiro P, Martin-Carbonero L, Gonzalez-Lahoz J. Long-term follow-up of HIV-infected patients with chronic hepatitis C virus infection treated with interferon-based therapies. *Antivir Ther* 2004;9:987-992.

320. Mandorfer M, Schwabl P, Steiner S, Reiberger T, Peck-Radosavljevic M. Advances in the management of HIV/HCV coinfection. *Hepatology* 2016;62:424-435.

321. Monforte Ade A, Bugarini R, Pezzotti P, De Luca A, Antinori A, Mussini C, Vigevani GM, et al. Low frequency of severe hepatotoxicity and association with HCV coinfection in HIV-positive patients treated with HAART. *J Acquir Immune Defic Syndr* 2001;28:114-123.

322. den Brinker M, Wit FW, Wertheim-van Dillen PM, Jurriaans S, Weel J, van Leeuwen R, Pakker NG, et al. Hepatitis B and C virus co-infection and the risk for hepatotoxicity of highly active antiretroviral therapy in HIV-1 infection. *AIDS* 2000;14:2895-2902.

323. Puoti M, Torti C, Ripamonti D, Castelli F, Zaltron S, Zanini B, Spinetti A, et al. Severe hepatotoxicity during combination antiretroviral treatment: incidence, liver histology, and outcome. *J Acquir Immune Defic Syndr* 2003;32:259-267.

324. Nunez M. Hepatotoxicity of antiretrovirals: incidence, mechanisms and management. *J Hepatol* 2006;44:S132-139.
325. Stone SF, Lee S, Keane NM, Price P, French MA. Association of increased hepatitis C virus (HCV)-specific IgG and soluble CD26 dipeptidyl peptidase IV enzyme activity with hepatotoxicity after highly active antiretroviral therapy in human immunodeficiency virus-HCV-coinfected patients. *J Infect Dis* 2002;186:1498-1502.
326. Labarga P, Soriano V, Vispo ME, Pinilla J, Martin-Carbonero L, Castellares C, Casado R, et al. Hepatotoxicity of antiretroviral drugs is reduced after successful treatment of chronic hepatitis C in HIV-infected patients. *J Infect Dis* 2007;196:670-676.
327. Eckenfelder A, Restrepo N, Pinzon, Ulveling D, Zagury J-F, Seitz H, Emiliani S, Gallois-Montbrun S. Identification of HIV-1 sequences targeted by the miRNA Induced silencing complex (miRISC) using Argonaute 2 cross-linking and immunoprecipitation (HITS-CLIP). *Retrovirology* 2013;10:P33.
328. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385-389.
329. International HapMap C. A haplotype map of the human genome. *Nature* 2005;437:1299-1320.
330. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-861.
331. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52-58.
332. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9:356-369.
333. Wang L, Xu Y. Haplotype inference by maximum parsimony. *Bioinformatics* 2003;19:1773-1780.
334. Bafna V, Gusfield D, Lancia G, Yooseph S. Haplotyping as perfect phylogeny: a direct approach. *J Comput Biol* 2003;10:323-340.
335. Halperin E, Eskin E. Haplotype reconstruction from genotype data using Imperfect Phylogeny. *Bioinformatics* 2004;20:1842-1849.
336. Niu T, Qin ZS, Xu X, Liu JS. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet* 2002;70:157-169.
337. Zhao JH, Sham PC. Faster haplotype frequency estimation using unrelated subjects. *Hum Hered* 2002;53:36-41.
338. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5-6.
339. Delaneau O, Zagury JF. Haplotype inference. *Methods Mol Biol* 2012;888:177-196.
340. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-989.

341. Adkins RM. Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. *BMC Genet* 2004;5:22.
342. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
343. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* 2011;1:457-470.
344. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289-293.
345. Cadranet JF, Rufat P, Degos F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). *Hepatology* 2000;32:477-481.
346. Blanc JF, Bioulac-Sage P, Balabaud C, Desmouliere A. Investigation of liver fibrosis in clinical practice. *Hepatol Res* 2005;32:1-8.
347. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343-350.
348. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.
349. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.
350. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906-913.
351. Loko MA, Salmon D, Carrieri P, Winnock M, Mora M, Merchadou L, Gillet S, et al. The French national prospective cohort of patients co-infected with HIV and HCV (ANRS CO13 HEPAVIH): early findings, 2006-2010. *BMC Infect Dis* 2010;10:303.
352. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012;489:91-100.
353. Xu CR, Zaret KS. Chromatin "pre-pattern" and epigenetic modulation in the cell fate choice of liver over pancreas in the endoderm. *Nucleus* 2012;3:150-154.
354. Dugail I, Postic C. Little caves ameliorate hepatic insulin signaling. Focus on "caveolin gene transfer improves glucose metabolism in diabetic mice". *Am J Physiol Cell Physiol* 2010;298:C442-445.
355. Edwards AC, Aliev F, Bierut LJ, Bucholz KK, Edenberg H, Hesselbrock V, Kramer J, et al. Genome-wide association study of comorbid depressive syndrome and alcohol dependence. *Psychiatr Genet* 2012;22:31-41.
356. Ulveling D, Francastel C, Hube F. When one is better than two: RNA with dual functions. *Biochimie* 2011;93:633-644.
357. Lloyd AG, Tateishi S, Bieniasz PD, Muesing MA, Yamaizumi M, Mulder LC. Effect of DNA repair protein Rad18 on viral infection. *PLoS Pathog* 2006;2:e40.

358. Mulder LC, Chakrabarti LA, Muesing MA. Interaction of HIV-1 integrase with DNA repair protein hRad18. *J Biol Chem* 2002;277:27489-27493.
359. Chaurushiya MS, Weitzman MD. Viral manipulation of DNA repair and cell cycle checkpoints. *DNA Repair (Amst)* 2009;8:1166-1176.
360. Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, Hosono N, et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. *J Hepatol* 2013;58:875-882.
361. Israely I, Costa RM, Xie CW, Silva AJ, Kosik KS, Liu X. Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. *Curr Biol* 2004;14:1657-1663.
362. Lu Q. delta-Catenin dysregulation in cancer: interactions with E-cadherin and beyond. *J Pathol* 2010;222:119-123.
363. Zhang J, Lu JP, Suter DM, Krause KH, Fini ME, Chen B, Lu Q. Isoform- and dose-sensitive feedback interactions between paired box 6 gene and delta-catenin in cell differentiation and death. *Exp Cell Res* 2010;316:1070-1081.
364. Yang I, Chang O, Lu Q, Kim K. Delta-catenin affects the localization and stability of p120-catenin by competitively interacting with E-cadherin. *Mol Cells* 2010;29:233-237.
365. Zhai B, Yan HX, Liu SQ, Chen L, Wu MC, Wang HY. Reduced expression of P120 catenin in cholangiocarcinoma correlated with tumor clinicopathologic parameters. *World J Gastroenterol* 2008;14:3739-3744.
366. DeBusk LM, Boelte K, Min Y, Lin PC. Heterozygous deficiency of delta-catenin impairs pathological angiogenesis. *J Exp Med* 2010;207:77-84.
367. Chen R, Dong Y, Xie X, Chen J, Gao D, Liu Y, Ren Z, et al. Screening candidate metastasis-associated genes in three-dimensional HCC spheroids with different metastasis potential. *Int J Clin Exp Pathol* 2014;7:2527-2535.
368. Wang HW, Hsieh TH, Huang SY, Chau GY, Tung CY, Su CW, Wu JC. Forfeited hepatogenesis program and increased embryonic stem cell traits in young hepatocellular carcinoma (HCC) comparing to elderly HCC. *BMC Genomics* 2013;14:736.
369. Lu Q, Paredes M, Medina M, Zhou J, Cavallo R, Peifer M, Orecchio L, et al. delta-catenin, an adhesive junction-associated protein which promotes cell scattering. *J Cell Biol* 1999;144:519-532.
370. Rogacki K, Kasprzak A, Stepinski A. Alterations of Wnt/beta-catenin signaling pathway in hepatocellular carcinomas associated with hepatitis C virus. *Pol J Pathol* 2015;66:9-21.
371. Katsounas A, Schlaak JF, Lempicki RA. CCL5: a double-edged sword in host defense against the hepatitis C virus. *Int Rev Immunol* 2011;30:366-378.
372. Fahey S, Dempsey E, Long A. The role of chemokines in acute and chronic hepatitis C infection. *Cell Mol Immunol* 2014;11:25-40.
373. Stock MK, Hammerich L, do ON, Berres ML, Alsamman M, Heinrichs D, Nellen A, et al. Met-CCL5 modifies monocyte subpopulations during liver fibrosis regression. *Int J Clin Exp Pathol* 2013;6:678-685.
374. Murooka TT, Wong MM, Rahbar R, Majchrzak-Kita B, Proudfoot AE, Fish EN. CCL5-CCR5-mediated apoptosis in T cells: Requirement for glycosaminoglycan binding and CCL5 aggregation. *J Biol Chem* 2006;281:25184-25194.

375. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genet Epidemiol* 2010;34:254-257.
376. Chen SH, Li YF, Lai HC, Kao JT, Peng CY, Chuang PH, Su WP, et al. Effects of patient factors on noninvasive liver stiffness measurement using acoustic radiation force impulse elastography in patients with chronic hepatitis C. *BMC Gastroenterol* 2012;12:105.
377. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol* 2011;35:790-799.
378. Karageorgopoulos DE, Allen J, Bhagani S. Hepatitis C in human immunodeficiency virus co-infected individuals: Is this still a "special population"? *World J Hepatol* 2015;7:1936-1952.
379. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology* 2008;19:649-651; discussion 657-648.
380. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010;5:1564-1573.
381. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011;6:121-133.
382. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008;456:18-21.
383. Wang Q, Lv H, Lv W, Shi M, Zhang M, Luan M, Zhu H, et al. Genome-wide haplotype association study identifies BLM as a risk gene for prostate cancer in Chinese population. *Tumour Biol* 2015;36:2703-2707.
384. Lambert JC, Grenier-Boley B, Harold D, Zelenika D, Chouraki V, Kamatani Y, Sleegers K, et al. Genome-wide haplotype association study identifies the FRMD4A gene as a risk locus for Alzheimer's disease. *Mol Psychiatry* 2013;18:461-470.
385. Tregouet DA, König IR, Erdmann J, Munteanu A, Braund PS, Hall AS, Grosshennig A, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* 2009;41:283-285.
386. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870-874.
387. Ellis JA, Scurrah KJ, Li YR, Ponsonby AL, Chavez RA, Pezic A, Dwyer T, et al. Epistasis amongst PTPN2 and genes of the vitamin D pathway contributes to risk of juvenile idiopathic arthritis. *J Steroid Biochem Mol Biol* 2015;145:113-120.
388. Genetic Analysis of Psoriasis C, the Wellcome Trust Case Control C, Strange A, Capon F, Spencer CC, Knight J, Weale ME, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet* 2010;42:985-990.
389. Prabhu S, Pe'er I. Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. *Genome Res* 2012;22:2230-2240.
390. Gyenesei A, Moody J, Semple CA, Haley CS, Wei WH. High-throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics* 2012;28:1957-1964.

391. Zhang X, Huang S, Zou F, Wang W. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics* 2010;26:i217-227.
392. Herold C, Steffens M, Brockschmidt FF, Baur MP, Becker T. INTERSNP: genome-wide interaction analysis guided by a priori information. *Bioinformatics* 2009;25:3275-3281.
393. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. *Nat Genet* 2007;39:1167-1173.
394. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet* 2014;15:722-733.
395. Yau C, Holmes CC. CNV discovery using SNP genotyping arrays. *Cytogenet Genome Res* 2008;123:307-312.
396. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 2007;39:S16-21.
397. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008;40:722-729.
398. Jimenez-Barron LT, O'Rawe JA, Wu Y, Yoon M, Fang H, Iossifov I, Lyon GJ. Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline. *Cold Spring Harb Mol Case Stud* 2015;1:a000422.
399. Marcovecchio ML, Florio R, Verginelli F, De Lellis L, Capelli C, Verzilli D, Chiarelli F, et al. Low AMY1 Gene Copy Number Is Associated with Increased Body Mass Index in Prepubertal Boys. *PLoS One* 2016;11:e0154961.
400. Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, Sudmant PH, et al. Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet* 2014;46:492-497.
401. di Iulio J, Ciuffi A, Fitzmaurice K, Kelleher D, Rotger M, Fellay J, Martinez R, et al. Estimating the net contribution of interleukin-28B variation to spontaneous hepatitis C virus clearance. *Hepatology* 2011;53:1446-1454.
402. Ehsani A, Janss L, Pomp D, Sorensen P. Decomposing genomic variance using information from GWA, GWE and eQTL analysis. *Anim Genet* 2016;47:165-173.
403. Spadoni JL, Rucart P, Le Clerc S, van Manen D, Coulonges C, Ulveling D, Laville V, et al. Identification of Genes Whose Expression Profile Is Associated with Non-Progression towards AIDS Using eQTLs. *PLoS One* 2015;10:e0136989.
404. Matullo G, Di Gaetano C, Guarrera S. Next generation sequencing and rare genetic variants: from human population studies to medical genetics. *Environ Mol Mutagen* 2013;54:518-532.
405. Quinones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. Deep sequencing: becoming a critical tool in clinical virology. *J Clin Virol* 2014;61:9-19.
406. Ogishi M, Yotsuyanagi H, Tsutsumi T, Gatanaga H, Ode H, Sugiura W, Moriya K, et al. Deconvoluting the composition of low-frequency hepatitis C viral quasispecies: comparison of genotypes and NS3 resistance-associated variants between HCV/HIV coinfecting hemophiliacs and HCV mono-infected patients in Japan. *PLoS One* 2015;10:e0119145.

407. Papić N, Maxwell CI, Delker DA, Liu S, Heale BS, Hagedorn CH. RNA-sequencing analysis of 5' capped RNAs identifies many new differentially expressed genes in acute hepatitis C virus infection. *Viruses* 2012;4:581-612.
408. Hou J, Groothuisink ZM, Koning L, Roomer R, van IWF, Kreeft K, Liu BS, et al. Analysis of the transcriptome and immune function of monocytes during IFN α -based therapy in chronic HCV revealed induction of TLR7 responsiveness. *Antiviral Res* 2014;109:116-124.
409. Song R, Liu Q, Liu T, Li J. Connecting rules from paired miRNA and mRNA expression data sets of HCV patients to detect both inverse and positive regulatory relationships. *BMC Genomics* 2015;16 Suppl 2:S11.
410. Bird A. Perceptions of epigenetics. *Nature* 2007;447:396-398.
411. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* 2012;13:679-692.
412. Feinberg AP, Koldobskiy MA, Gondor A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* 2016;17:284-299.
413. Hlady RA, Tiedemann RL, Puszyk W, Zendejas I, Roberts LR, Choi JH, Liu C, et al. Epigenetic signatures of alcohol abuse and hepatitis infection during human hepatocarcinogenesis. *Oncotarget* 2014;5:9425-9443.
414. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010;5:e10693.
415. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009;325:1246-1250.
416. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, et al. A genome-wide association study of global gene expression. *Nat Genet* 2007;39:1202-1207.
417. Qu H, Fang X. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics* 2013;11:135-141.
418. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004;306:636-640.
419. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790-1797.
420. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278-1283.
421. Evangelou M, Rendon A, Ouwehand WH, Wernisch L, Dudbridge F. Comparison of methods for competitive tests of pathway analysis. *PLoS One* 2012;7:e41018.
422. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* 2012;28:323-332.
423. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012;8:e1002375.

424. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* 2011;19:837-843.
425. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010;11:843-854.
426. Wu JQ, Saksena MM, Soriano V, Vispo E, Saksena NK. Differential regulation of cytotoxicity pathway discriminating between HIV, HCV mono- and co-infection identified by transcriptome profiling of PBMCs. *Virology* 2015;12:4.

Liste des publications

Ulveling D.*, Sigrid Le Clerc S.*, Labib T., Noirel J., Laville V., Coulonges C., Carpentier W., Salmon D., Dabis F., Lévy Y., Dominguez S. and Zagury J.F. and the HEPAVIH ANRS CO13 cohort. « A GWAS reveals CTNND2 gene polymorphisms associations in liver fibrosis by comparing F0F1F2 and F3F4 METAVIR score groups in HIV/HCV coinfection. » - Article en préparation.

Laville V.*, Le Clerc S.*, Ezzedine K., Jdid R., Taing L., Labib T., Coulonges T., **Ulveling D.**, Carpentier W., Galan P., Hercberg S., Morizot F., Latreille J., Malvy D., Tschachler E., Guinot C., Zagury J.F. « A Genome-Wide Association Study identifies new genetic associations in upper eyelid sagging. » - Article en préparation.

Ulveling D.*, Le Clerc S.*, Cobat A.*, Labib T., Noirel J., Laville V., Coulonges C., Carpentier W., Nalpas B., Heim M.H., Poynard T., Cerny A., Pol A., Bochud P.Y., Dabis F., Theodorou I., Lévy Y., Salmon D., Abel L., Dominguez L. and Zagury J.F. and the HEPAVIH ANRS CO13 cohort, the Swiss Hepatitis C Cohort Study Group, and the French ANRS HC EP 26 Genoscan Study Group. « A new 3p25 locus is associated with liver fibrosis progression in HIV/HCV co-infected patients », **Hepatology**. - Article en cours de seconde révision.

Laville V.*, Le Clerc S.*, Ezzedine K., Jdid R., Taing L., Labib T., Coulonges T., **Ulveling D.**, Carpentier W., Galan P., Hercberg S., Morizot F., Latreille J., Malvy D., Tschachler E., Guinot C., Zagury J.F. « A Genome-Wide Association Study in Caucasian Women reveals the involvement of HLA genes in the severity of facial solar lentigines », **Pigment Cell and Melanoma Research**. - Article en cours de révision.

Spadoni J.L.*, Rucart P.*, Le Clerc S., van Manen D., Coulonges C., **Ulveling D.**, Laville V., Labib T., Taing L., Delaneau O., Montes M., Schuitemaker H., Noirel J., Zagury J.F. « Identification of Genes Whose Expression Profile Is Associated with Non-Progression towards AIDS Using eQTLs. », **PLoS ONE**, 2015, 10(9): e0136989.

Ulveling D., Dinger M.E., Francastel C. and Hubé F. « Identification of a dinucleotide signature that discriminates coding from non-coding long RNAs. », **Frontiers in Genetics**, 2014, doi: 10.3389/fgene.2014.00316.

Le Clerc S., Delaneau O., Coulonges C., Spadoni J.L., Labib T., Laville V., **Ulveling D.**, Noirel J., Montes M., Schächter F., Caillat-Zucman S., Zagury J.F., « Evidence After Imputation for a Role of MICA Variants in Nonprogression and Elite Control of HIV Type 1 Infection. », **Journal Of Infectious Disease**, 2014, 210(12), 1946-50.

Eckenfelder A., Restrepo N., **Ulveling D.**, Zagury J.F., Seitz H., Emiliani S., Gallois-Montbrun S. « Identification of HIV-1 sequences targeted by the miRNA Induced silencing complex (miRISC) using Argonaute 2 cross-linking and immunoprecipitation (HITS-CLIP) », **Retrovirology**, 2013, BioMed Central, 2013, 10 (Suppl 1), pp.P33.

Liste des communications orales

Ulveling D. « Genome-wide association studies for liver fibrosis in HIV/HCV coinfection and replication in HCV mono-infection ». Genomics and disease pathogenesis. Chissay, 17-18 Juin 2015.

Ulveling D. « ANRS CO13 HEPAVIH: exploration and perspectives. ». Genomics and disease pathogenesis. Ermenonville, 16-17 Décembre 2014.

Posters

Le Clerc S., **Ulveling D.**, Labib T., Winnock M., Theodorou I., Abel L., Lévi Y., Salmon D., Zagury J.F., Dominguez S. « GWAS reveals new genetic associations with liver fibrosis progression in HIV/HCV. **Conference on Retroviruses and Opportunistic Infection (CROI)** » Boston, **Février 2014**.

Analyses génomiques de la coinfection par les virus VIH et VHC

Résumé

Un tiers des infectés par le VIH à travers le monde sont coinfectés par le VHC. Au cours des dernières années, l'hépatite C est devenue une cause majeure de mortalité lors de la coinfection et la progression de la fibrose hépatique y est accélérée par rapport à la monoinfection. En plus de facteurs environnementaux bien connus, la génétique semble jouer un rôle important dans l'évolution de l'hépatite C.

Durant ma thèse, j'ai réalisé des GWAS avec les données génomiques de 292 patients coinfectés VIH/VHC issus d'une cohorte où l'histoire naturelle du VIH et du VHC y est renseignée de manière détaillée. Deux études portant sur la sévérité de la fibrose (respectivement quantitative et qualitative) ont permis d'abord d'identifier un locus répliqué dans la monoinfection VHC, impliquant des gènes liés à la structure cellulaire (*CAV3*) et à la réplication VHC (*RAD18*) puis le gène *CTNND2* lié indirectement à des maladies hépatiques.

Ces résultats permettent de nouvelles perspectives dans la compréhension de mécanismes liés à la fibrose hépatique lors d'infection par le VHC.

Mots clés : GWAS, SNP, coinfection VIH/VHC, fibrose hépatique, maladie du foie

Résumé en anglais

A third of HIV worldwide people are coinfecting with HCV. In recent years, hepatitis C has become a major cause of death during coinfection and the progression of hepatic fibrosis is accelerated relative to the monoinfection. In addition to well-known environmental factors, genetics appears to play an important role in the evolution of hepatitis C.

During my PhD, I performed GWAS with genomic data of 292 patients HIV/HCV coinfecting from a cohort where the natural history of HIV and HCV is informed in detail. Two studies on the severity of fibrosis (respectively quantitative and qualitative) allowed to first identify a locus replicated in monoinfection HCV, involving genes linked to cell structure (*CAV3*) and HCV replication (*RAD18*) then *CTNND2* gene indirectly related to liver disease.

Overall, these findings provide new insights in the understanding of mechanisms related to hepatic fibrosis during HCV infection.

Keywords : GWAS, SNP, HIV/HCV coinfection, liver fibrosis, liver disease