



**HAL**  
open science

# Identification de régulateurs clés de la carcinogénèse hépatique humaine : Études clinico-pathologiques, moléculaires et fonctionnelles

Alexandre dos Santos dos Santos

► **To cite this version:**

Alexandre dos Santos dos Santos. Identification de régulateurs clés de la carcinogénèse hépatique humaine : Études clinico-pathologiques, moléculaires et fonctionnelles. Cancer. Université Paris Saclay (COmUE), 2019. Français. NNT : 2019SACLS382 . tel-04265543

**HAL Id: tel-04265543**

**<https://theses.hal.science/tel-04265543>**

Submitted on 31 Oct 2023

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

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

# Identification de régulateurs clés de la carcinogenèse hépatique humaine : Etudes clinico- pathologiques, moléculaires et fonctionnelles

Thèse de doctorat de l'Université Paris-Saclay  
préparée à L'INSERM UMR-S 1193

École doctorale n° n°582 : cancérologie : biologie - médecine - santé  
(CBMS)

Spécialité : Aspects moléculaires et cellulaires de la biologie

Thèse présentée et soutenue à Villejuif, le 30 octobre 2019, par

**Alexandre Dos Santos**

Composition du Jury :

Didier Samuel PUPH, Université Paris-Sud (INSERM U1193)	Président
Anne Corlu DR2, CNRS (INSERM U1241)	Rapporteur
Pascal Pineau DR, Institut Pasteur (Organisation nucléaire et oncogénèse)	Rapporteur
Sabine Colnot DR2, INSERM (INSERM U1138)	Examineur
Olivier Rosmorduc PUPH, Sorbonne-Université (INSERM U1193)	Examineur
Antonio Sa Cunha PUPH, Université Paris-Sud (INSERM U1193)	Examineur
Jamila Faivre PUPH, Université Paris-Sud (INSERM U1193)	Directeur de thèse

*Aucune carte du monde n'est digne d'un regard si le  
pays de l'utopie n'y figure pas ... ..*

*Oscar Wilde*

*« Dis donc Kiki ! c'est quand que tu soutiens ? »*

## Remerciements

Je remercie les membres du jury pour m'avoir fait l'honneur d'accepter de juger ce travail de thèse.

Je remercie sincèrement les Docteurs **Anne Corlu** et **Pascal Pineau**, pour avoir accepté de juger ce travail en qualité de rapporteurs, pour le temps qu'ils ont consacré à la lecture du manuscrit et à leurs patiences.

Je remercie le Docteur **Sabine Colnot**, le Professeur **Olivier Rosmorduc** et le Professeur **Antonio Sa Cunha** d'avoir accepté d'être examinateurs de mon travail.

Je tiens à adresser ma profonde reconnaissance et gratitude au Professeur **Jamila Faivre**. Jamila ! Nous nous connaissons depuis bientôt 17 ans. J'étais un tout jeune blanc bec arrivant dans le laboratoire de Christian Bréchet à Necker à la 370. Tu étais « étudiante en thèse » et dirigeais déjà un groupe dans l'unité. Nous avons commencé l'aventure de la 785 à Paul Brousse et très vite j'ai commencé à travailler avec toi d'abord de manière officieuse après ma réussite au concours INSERM puis j'ai rejoint ton équipe de manière officielle il y a plus de 8 ans. Nous avons essuyé de nombreux échecs et remporté de très belles victoires au cours de ces années. Nous avons été souvent en désaccord toi et moi, mais nos vifs échanges ont toujours fait avancer les projets du groupe. Je me suis toujours plu à penser que j'étais en quelque sorte « ton bras armé » pour désamorcer des conflits ou pour trouver des solutions rapides « en dehors des clous ». Tu as été et est aujourd'hui encore une « école de la vie » et je t'admire pour cela. J'apprécie ta ténacité, ta résistance, ton opiniâtreté sans faille. Je te remercie de m'avoir fait confiance quand j'ai exprimé il y a quelques années, ma volonté de me spécialiser dans la bio-informatique. Avec ton plein soutien, j'ai pu développer de nouveaux outils de travail dans le groupe, outils dont l'utilité a été reconnue par mes pairs, l'année dernière, lors de ma promotion aux corps des ingénieurs d'études de l'INSERM. Enfin je te remercie pour ta patience et ta disponibilité (« si, si ta disponibilité, malgré tes nombreuses fonctions à l'hôpital et à l'université, tu as toujours été là quand j'en avais besoin »).

Je suis ravi d'avoir travaillé sous ta direction et j'espère encore passer de nombreuses années au sein de l'équipe que tu animes et diriges dans l'unité. Je te souhaite et nous souhaite encore de belles conquêtes.

Je tiens à adresser mes plus sincères remerciements au Professeur **Didier Samuel** de m'avoir accueilli dans son unité, au tout début de cette aventure villejuifoise.

Je remercie le Professeur **Christian Bréchet**, pour m'avoir accueilli dans son laboratoire en 2002, et pour m'avoir toujours témoigné sa confiance.

Je tiens à remercier les membres de l'équipe « Microbiote, inflammation et cancer du foie ». le Docteur **Nicolas Moniaux**. Nicolas, merci pour tes conseils précieux, nos discussions à l'aube, nos échanges et ton soutien, dès la première heure, à ma reconversion professionnelle. **Alice**, je te remercie pour ton honnêteté, tes « crises de colère et de pleurs » qui cachent un profond respect de notre institution et une volonté inébranlable d'améliorer les conditions de travail du laboratoire. Merci pour ces tranches de vie, à la cantine, à mon bureau, que tu m'as fait partager. Petit bémol..... le foot, mais dans ma grande mansuétude, je ne t'en veux pas de supporter l'ASNL. Je te souhaite une grande réussite dans ce métier. **Emma**, merci pour les moments partagés lors de la révision du papier dans Gastro (difficiles, mais on s'est bien marré !!) et pour nos discussions culturelles. Il est grand temps que les établissements publics comprennent l'importance de ton métier et je te souhaite de tout cœur de réussir. **Julie**, ma « compagne OMICS », je te remercie pour ces derniers mois où tu as travaillé pour m'aider à finir ce projet. Je te remercie pour ta rigueur, pour nos discussions frénétiques au sous-sol du CHB et pour ta profonde gentillesse. Je te souhaite de réussir et de t'épanouir dans ce métier. **Marie-Annick**, je te remercie pour ta douceur, ton humour et pour tes conseils toujours adéquats. Aux étudiants en thèse du groupe, **Nicolas** et **Hala** « Halette », je vous souhaite tout le succès que vous méritez. Enfin je remercie tous les autres membres qui m'ont supporté toutes ces années **Michael**, **Patrick**, **Marion**, **Delphine**, **Myriam**, **Stéphanie** et les M2.

Je tiens à exprimer ma gratitude au Docteur **Christophe Desterke** qui m'a « infecté » avec le virus de la bio-informatique. J'ai beaucoup appris avec toi ! Je t'en serais toujours reconnaissant.

-----

J'exprime ma profonde gratitude et reconnaissance au Docteur et amie **Claire Lacoste**. Claire, je t'ai connu en deux temps : Thésarde rigoureuse, drôle, généreuse et humaniste puis ingénieure rigoureuse, drôle, généreuse et humaniste. Tu es une personne inoubliable dans tous les sens du terme. Ta détermination et ton énergie à faire bouger les lignes, ton sens du sacrifice et ta profonde estime pour l'Homme sont pour moi des idéaux à suivre en exemple dans ma carrière

professionnelle comme dans ma vie personnelle. Rares sont les gens qui ont su déceler cette richesse en toi ! Claire, j'ai la chance (« oui je sais la chance n'existe pas, je suis le premier à le dire mais je vais faire une exception... ») de faire partie de tes proches et je savoure chaque instant (« que Dieu fait ? Non, je veux bien faire des exceptions mais faut pas exagérer... ») à tes côtés.

Je tiens à exprimer ma profonde reconnaissance et gratitude au Docteur **Franck Chiappini**. Voisin de bureau (« Ouhla ! Il a pas l'air rigolo le nouveau ! ») et dans Villejuif (« Y a beaucoup de personnes qui rentrent du sport à 4h du mat' »), tu es devenu au fil du temps un ami sincère. Tu m'as fait découvrir le langage R alors que j'y étais farouchement opposé et je suis devenu aujourd'hui, grâce à toi un « Ayatollah » de RStudio. J'ai la chance de partager avec toi encore maintenant des fous rires, des discussions futuristes et ces moments de complicité qui ont forgé notre relation. L'INSERM a perdu une perle rare en te laissant partir. Pour tous ces instants et pour les suivants, merci de tout mon cœur mon Francky !

Je tiens à remercier chaleureusement le Docteur **Serena Battaglia** (Battaglagla pour les intimes). Je te connais depuis des lustres (« qu'est-ce qu'on était jeune au début à Necker ! ») et tu es devenue une amie véritable. J'estime ta franchise, ton authenticité, ta bonne humeur, ta mauvaise humeur, ton humanisme, ta misogynie (ceux qui savent, savent...), ton énergie. Il me tarde de reprendre nos soirées bobos du jeudi.

Je tiens à remercier le Docteur **Isabel Serra de Matamala**. Isabel, tu fais partie de ces personnes qui ont traversé ma vie et ont laissé une empreinte indélébile dans mon cœur. Ton dynamisme, ta fraîcheur, ton humour ont égayé ce laboratoire et j'ai la chance aujourd'hui de goûter encore à ces instants chaleureux avec toi. *Sempre seràs al meu cor.*

-----

Je remercie le Docteur **Nassima Benzoubir**. Ma Nass, merci pour ta joie de vivre contagieuse, ton dynamisme, tes fous rires, ta disponibilité. Merci aussi de m'avoir montré ces derniers temps, qu'on pouvait faire de belles choses avec la fluo. J'espère que l'Université Paris-Saclay comprendra vite ta vraie valeur et te promouvra au poste que tu mérites.

Je remercie **Christophe Masselon** et **Magali Court** de m'avoir appris le principe de l'AMT. Nos échanges ont toujours été fructueux.

Je remercie le Professeur **Catherine Guettier** et le Docteur **Mylène Sebah** pour m'avoir enseigné les bases de la pathologie hépatique.

Je remercie **Katia** pour m'avoir appris les bases de l'expérimentation en histologie. Sans toi, ce projet, et d'autres, n'auraient pu être réalisés. Merci pour ta bonne humeur et j'attends avec impatience ton prochain message du lundi matin « Coucou de PB ! ». Je remercie aussi toutes les techniciennes d'Anapath, **Martine**, les **deux Véro**, **Marie-Jo**, **Violette**, **Françoise** et les autres qui m'ont donné un coup de main dans mes travaux.

Je remercie nos gestionnaires de choc et de charmes, **Marina**, **Françoise**, **Christine**, **Vincent** (Vincenzo), pour leurs bonnes humeurs, leurs efficacités et leurs disponibilités. On ne le dit que trop rarement, les laboratoires tournent grâce à vous !!

Je tiens à remercier le Docteur **Rodolphe Sobesky**, pour son humour, sa gentillesse, son humour, sa gentillesse, son humour, sa gentillesse, son humour et son humour. À très bientôt pour nos discussions politicardes.

J'exprime ma gratitude envers le Docteur **Valérie Thiers**, qui fut mon chef d'équipe pendant très longtemps, qui m'a initié à la microdissection laser et à la biologie moléculaire.

Je tiens à exprimer ma profonde reconnaissance au Docteur **France Demaugre**. Françonnette, tu as été comme une mère pour moi dans le laboratoire. Mère attentionnée, mère sévère, mère « chiante », mère juste. Ce que je suis devenu aujourd'hui, je te le dois en grande partie.

Je remercie les membres de l'unité 785 et 1193 que j'ai côtoyé et avec qui j'ai passé des moments inoubliables. **Juan**, « mon frère » **Abdallah**, **Ola**, **Mohyeddine**, **Jean-Charles**, **Slavka**, **Julie**, **Yannick**, **Francesca**, **Elvire**, **François R**, **Michelle**, **Amel**, **Guillaume**, **Aline**, **Nazha**.  
MERCIIII

Je remercie les personnels APHP que j'ai eu plaisir à croiser, **Laurent**, **Pascal**, les infirmières de l'hôpital de jour et celles de réa, les services techniques, les services de sécurité, **Carima**, **Rashmi** et **Guadalupe**, les ARCs (**Magali**, **Isabelle**, **Meriem**, **Frédérique**, **Sylvie**, **Maryam**). Je remercie **Françoise Laigle** pour sa bonne humeur et son énergie communicative.

-----

Je remercie le Docteur **Jeanine Quillard**, pour m'avoir initié à la télé-pathologie. J'ai passé d'excellents moments avec vous pendant les séances de scan à Bicêtre.



Je remercie **Olivier Trassard**, pour son expertise en analyse immunohistochimique et pour sa disponibilité et sa rapidité (« Euh Olivier, tu peux me scanner des lames pour avant-hier ? »)

Je remercie le Docteur **Adrienne Anginot**, pour m'avoir permis de faire des manip préliminaires en hypoxie.

Je remercie le Docteur **Julia Delahousse**, pour le dosage de la 2-hydroxyglutarate et sa disponibilité.

-----

Je remercie également mes amis Mickaël, Kevin et Setareh. **Mickaël**, tu m'as honoré en m'accordant le privilège d'être ton témoin de mariage. Je garde de merveilleux souvenirs de ton passage au labo, de cette fameuse matinée où tu as appris que le poste d'IE t'était attribué et de nombreuses soirées jeux. Merci pour ton soutien ! Et à très bientôt pour des nuits blanches. **Kevin**, tu m'as honoré en m'accordant le privilège d'être ton chauffeur de mariage ;-). Kevin, tu es comme un petit frère pour moi qui suis fils unique. Je garde d'excellents souvenirs de ta présence au labo (« Alex ! Tu peux vérifier mon calcul ? Nico ne me croit pas quand je lui dis que c'est bon ! »). Je te souhaite de trouver un poste qui soit vraiment en adéquation avec tes compétences. Sorbonne Université risque de perdre une pièce centrale de sa structure. À très bientôt aussi pour des soirées jeux. **Setareh**, je me souviens pendant ton M1, avoir passé de longues heures à t'enseigner le principe de la 2D-PAGE. Ces moments furent durs pour toi et m'ont révélé ta force intérieure. Setareh et Kevin, je vous souhaite tout le bonheur du monde. J'ai une chance inouïe de vous avoir tous les 3 pour amis.

Je remercie chaleureusement **Camille** et **François** de leur soutien et de « m'avoir botté le cul » pour que je termine cette thèse.

Je tiens à remercier **Joanna**. Merci de m'avoir encouragé et soutenu au début de cette aventure.

-----

J'ai une pensée pour **Maurice Fontaine**, qui nous a quitté trop tôt. Je garde des souvenirs mémorables avec toi.

J'ai une pensée affectueuse pour le Docteur **Sokhavuth Sar**. Ma Sokha, je ne suis pas religieux tu le sais bien, mais il m'arrive parfois de sentir ta présence à mes côtés. J'ai été profondément affecté par ton décès, toi qui te faisais une joie d'être une grand-mère poule pour tes petits

enfants. Je t'ai rencontré en 2002, tout jeune, tout juste sorti de mes études et j'arrivais dans un très grand labo. Les premiers instants, j'étais perdu et au milieu de cette foule, je t'ai trouvé. Une force tranquille, débordante de gentillesse et d'affection. Toujours souriante et positive. Toujours à l'écoute et disponible. Tout semblait nous séparer : nos âges, nos cultures, nos histoires et pourtant, tu fus l'une des personnes avec qui j'ai été le plus complice dans ma carrière. Il suffisait de s'échanger un regard (notamment pendant les réunions interminables...) pour se comprendre, sourire et rire.

Sokha, je te dédie cette thèse.

Je tiens à remercier **mes parents**. Vous êtes mes modèles. Vous avez, malgré l'adversité, toujours réussi à accomplir ce qui vous paraissait le plus juste. Vous vous êtes sacrifiés pour moi et j'espère de tout cœur que vous êtes fier de moi et de ce que je suis devenu. Merci de m'avoir élevé avec ces valeurs de persévérance et de partage, Merci de m'avoir toujours encouragé, épaulé, et merci pour votre confiance.

Je tiens à remercier ma chère **Clara**. Ma fille ! merci pour ta bonne humeur (c'est normal tu tiens de ton père !), pour ton dynamisme (c'est normal tu tiens de ton père !), pour ton affection (c'est normal tu tiens de ton père !), et même pour ces moments où tu m'agaces (oui ça aussi tu tiens de ton père...) ! Mon lapin, je te souhaite de t'épanouir à l'école et de trouver ta voie. Par contre, le jour où tu voudras faire une thèse, viens me voir pour en discuter...

Enfin je n'y serais jamais arrivé sans **toi** qui partages ma vie, qui m'aime d'un amour tendre, ardent, idéal. Je voudrais d'abord m'excuser pour tous les tracas du quotidien que je t'ai fait subir. Mon stress, mes indécisions, mes paniques ... tu as été d'une patience sans faille avec moi et tu m'as offert tout ce dont j'avais besoin : une écoute, une absence de jugement et un formidable optimisme. Merci d'être présente à mes côtés, merci d'être toi. Je t'aime... ,, @@@

## Résumé

Le carcinome hépatocellulaire (CHC) est la forme la plus fréquente de cancer du foie et l'une des principales causes de mortalité par cancer dans le monde. Il s'agit d'une maladie de mauvais pronostic, aux ressources thérapeutiques limitées, hétérogène sur le plan immunophénotypique et génomique, qui se développe très souvent sur un foie remanié cirrhotique. Les études moléculaires ont révélé plusieurs sous-classes de CHC caractérisés par des signatures génomiques et protéomiques distinctes. Au cours de mon travail de thèse, nous avons contribué à améliorer notre compréhension de la biologie des CHC et des classifications moléculaires en cartographiant le génome non-codant de tumeurs de CHC induites par des virus hépatotropes (VHB, VHC) et en étudiant la sous-classe moléculaire de CHC la plus agressive KRT19-positif. Nous avons établi la première carte de transcriptome à ARN non codants du CHC et révélé une forte activation intra-tumorale des rétrotransposons à LTR, qui sont principalement inhibés dans les cellules hépatiques normales, dans des CHC induits par les VHB et VHC. Certains des transcrits dérivés de LTR se sont révélés être des régulateurs clés de l'expression génique et donc activer la croissance cellulaire. Dans la deuxième étude, nous identifions une nouvelle voie de régulation des CHC KRT19-positif affectant le métabolisme énergétique de ces tumeurs. Les CHC KRT19-positif sont des tumeurs fortement glycolytiques liée à une activation de la réponse à l'hypoxie. L'excès de production par les CHC KRT19-positif de l'oncométabolite 2-hydroxyglutarate en absence de mutation des gènes IDH1 et IDH2 était associé à un profil aberrant hyperméthylé sur la lysine 9 de l'histone H3 (H3K9me3) suggérant une répression de la transcription notamment des gènes impliqués dans la différenciation cellulaire.

## Abstract

Hepatocellular carcinoma (HCC) is the main primary liver cancer and one of the most leading cause of cancer-related death worldwide. This heterogeneous disease with a worse prognosis has been subjected of numerous studies aimed to establish global phenotypic profiles. During my thesis, I dedicated my work to improve these classifications by identifying signatures on the non-coding genome and working on a very aggressive form of HCC expressing progenitors markers. With help of a Japanese team, we demonstrated that LTR-derived ncRNAs were active in HCC and that correlation correlates with expression of common cancer markers (GPC3) and TP53 mutations. This signature can also be used to discriminate HCCs at high risk of recurrence. Finally, we have showed that these LTRs are detectable on preneoplastic stages in the *Mdr2* KO mouse model. In parallel, I worked on HCC that expresses progenitor markers such as cytokeratin 19. Using proteomic and transcriptional approaches and in silico analyses, we propose that the occurrence of this type of cancer is due to an hypoxic event likely related to trans-arterial chemoembolization. These tumors have a highly glycolytic phenotype with production of an oncometabolite (2-hydroxyglutarate) that has been generally found in IDH1/2 mutated cholangiocarcinomas. Finally we suggest the use of metformin, type 2 diabetes drug, to reverse metabolic reprogramming and restore sensitivity to chemotherapy.

<b>I INTRODUCTION.....</b>	<b>19</b>
1. ÉPIDÉMIOLOGIE DU CARCINOME HÉPATOCELLULAIRE.....	19
2. ÉTIOLOGIES DU CARCINOME HÉPATOCELLULAIRE.....	19
2.1 Etiologies virales.....	19
2.2 Etiologies non virales.....	21
2.3 Facteurs de risque génétiques.....	22
3. HISTOIRE NATURELLE DE LA CARCINOGENÈSE HÉPATIQUE.....	22
3.1 De l'hépatite chronique au CHC.....	23
3.2 Fibrogenèse hépatique.....	25
3.3 Complications de la cirrhose.....	26
3.4 Du foie sain au CHC.....	29
4. DIAGNOSTIC DU CARCINOME HÉPATOCELLULAIRE.....	29
4.1 Imagerie.....	30
4.2 Diagnostic biologique (marqueurs).....	30
4.3 Diagnostic histologique.....	32
5. TRAITEMENTS DU CARCINOME HÉPATOCELLULAIRE.....	34
5.1 Radiofréquence.....	34
5.2 Chimio-embolisation trans-Artérielle.....	34
5.3 Traitements chirurgicaux.....	35
5.4 Thérapies ciblées.....	35
5.5 Immunothérapie.....	36
6. PRINCIPAUX FACTEURS DE CARCINOGENÈSE HÉPATIQUE.....	36
6.1 Mutagenèse insertionnelle.....	36
6.2 Action directe des protéines virales.....	40
6.3 Maladies métaboliques.....	41
6.4 Microbiote intestinal.....	41
7. EFFECTEURS MOLÉCULAIRES.....	42
7.1 Anomalies du génome codant.....	42
7.2 Anomalies du génome non-codant.....	48
8. LE MÉTABOLISME DU CARCINOME HÉPATOCELLULAIRE.....	51
8.1 Généralités sur le métabolisme des cellules tumorales.....	51
8.2 Voies de signalisation impliquées dans la reprogrammation métabolique du CHC.....	52
8.3 Hétérogénéité métabolique des cancers.....	55
8.4 Le métabolisme comme cible pour la thérapie anti-cancéreuse.....	55

9. CLASSIFICATIONS MOLÉCULAIRES.....	57
9.1 Classifications transcriptionnelles.....	57
9.2 Signature immunitaire.....	59
9.3 Signature protéomique.....	60
9.4 Signature métabolomique.....	61
<b>II ÉTUDE 1 : ACTIVATION DE PROMOTEURS RÉTROVIRAUX LTR DANS LES CARCINOMES HÉPATOCELLULAIRES VIRUS-INDUITS.....</b>	<b>64</b>
1. PRÉSENTATION DE L'ÉTUDE.....	64
2. ARTICLE.....	67
<b>III ÉTUDE 2 : CARACTÉRISATION DU CARCINOME HÉPATOCELLULAIRE À DOUBLE DIFFÉRENCIATION.....</b>	<b>125</b>
1. PRÉSENTATION DE L'ÉTUDE.....	125
2. MATÉRIEL ET MÉTHODES.....	128
2.1 Sélection des échantillons.....	128
2.2 Microdissection Laser.....	128
2.3 Spectrométrie de masse.....	129
2.4 Dosage tissulaire de la 2-Hydroxyglutarate.....	131
2.5 Détection des protéines par immunohistochimie.....	132
2.6 Analyse des transcrits.....	133
2.7 Enrichissement fonctionnel.....	133
2.8 Modèle in vitro.....	134
2.9 Analyses statistiques.....	135
3. RÉSULTATS.....	136
3.1 La protéomique quantitative lie les voies métaboliques aux CHC KRT19-positif.....	136
3.2 La réponse à l'hypoxie et la glycolyse dont des marques distinctives des CHC KRT19-positif.....	138
3.3 Forte association entre réponse à l'hypoxie et expression de KRT19 dans la cohorte de CHC du TCGA.....	141
3.4 La réponse à l'hypoxie provoque l'activation de l'expression de KRT19.....	143
3.5 Les CHC hypoxiques produisent du 2-hydroxyglutarate.....	144
4. DISCUSSION.....	146
5. CONCLUSIONS ET PERSPECTIVES.....	151
6. FIGURES ET TABLEAUX SUPPLÉMENTAIRES.....	152
<b>IV ETUDE 3 : RÔLE DE LA LECTINE REG3A DANS L'HOMEOSTASIE DU MICROBIOTE INTESTINAL : CONTRÔLE DE L'INFLAMMATION DANS UN MODÈLE DE COLITE.....</b>	<b>162</b>
1. PRÉSENTATION DE L'ÉTUDE.....	162

2. ARTICLE.....	165
V DISCUSSION GÉNÉRALE.....	194
VI RÉFÉRENCES BIBLIOGRAPHIQUES.....	200

## Publications

Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, Nicoli P, Kindlova M, Ghisletti S, **Dos Santos A**, Rapoud D, Samuel D, Faivre J, Ewing AD, Richardson SR, Faulkner GJ. *L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis*. Genome Res. 2018 May;28(5):639-653

Sebagh M, Bosselut N, **Dos Santos A**, Allard MA, Ruiz A, Saffroy R, Cherqui D, Vibert E, Castaing D, Adam R, Cunha AS, Lemoine A. *Rare genetic heterogeneity within single tumor discovered for the first time in colorectal liver metastases after liver resection*. Oncotarget. 2018 Apr 24;9(31):21921-21929

Altinel K, Hashimoto K, Wei Y, Neuveut C, Gupta I, Suzuki AM, **Dos Santos A**, Moreau P, Xia T, Kojima S, Kato S, Takikawa Y, Hidaka I, Shimizu M, Matsuura T, Tsubota A, Ikeda H, Nagoshi S, Suzuki H, Michel ML, Samuel D, Buendia MA, Faivre J, Carninci P. *Single-Nucleotide Resolution Mapping of Hepatitis B Virus Promoters in Infected Human Livers and Hepatocellular Carcinoma*. J Virol. 2016 Nov 14;90(23):10811-10822.

Benzoubir N, Mussini C, Lejamtel C, **Dos Santos A**, Guillaume C, Desterke C, Samuel D, Br  chot C, Bourgeade MF, Guettier C. Gamma-smooth muscle actin expression is associated with epithelial-mesenchymal transition and stem-like properties in hepatocellular carcinoma. PLoS One. 2015 Jun 25;10(6):e0130559

Moniaux N, Darnaud M, Garbin K, **Dos Santos A**, Guettier C, Samuel D, Amouyal G, Amouyal P, Br  chot C, Faivre J. The Reg3   (HIP/PAP) Lectin Suppresses Extracellular Oxidative Stress in a Murine Model of Acute Liver Failure. PloS One. 2015 May 4;10(5):e0125584.

Sebagh M, Castillo-Rama M, Azoulay D, Coilly A, Delvart V, Allard MA, **Dos Santos A**, Johanet C, Roque-Afonso AM, Saliba F, Duclos-Vall  e JC, Samuel D, Demetris AJ. Histologic findings predictive of a diagnosis of de novo autoimmune hepatitis after liver transplantation in adults. Transplantation. 2013 Oct 15;96(7):670-8.

Shukla R, Upton KR, Mu  oz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, Sinha S, Iannelli F, Radaelli E, **Dos Santos A**, Rapoud D, Guettier C, Samuel D, Natoli G, Carninci P, Ciccarelli FD, Garcia-Perez JL, Faivre J, Faulkner GJ. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 2013 Mar 28;153(1):101-11.



## Abbréviations

### 0-9 :

2-HG : 2-Hydroxyglutarate

4-HNE : 4-hydroxynonanal

### A :

AbI : Index d'abondance

ADN: Acide désoxyribonucléique

AFP : Alpha foeto-protéine

AMT : *Accurate Mass and Time*

ARN: Acide ribonucléique

ARNm : ARN messenger

ATP : Adenosine tri-phosphate

AUC : Aire sous la courbe (*Area under curve*)

### B

BCLC : *Barcelona Clinic Liver Cancer*

### C

cfDNA : ADN nu (*circulating free DNA*)

CHC : Carcinome Hépatocellulaire

CoCl<sub>2</sub> : Chlorure de cobalt

CpG : Cytosine phosphate guanine

CTC : Cellules tumorales circulantes

### D

DAG : Diacylglycérol

DEN : Diethylnitrosamine

DMEM : *Dulbecco's Modified Eagle Medium*

DMOG :dimethyloxaloylglycin

DNMT : Méthyltransférases de l'ADN

DTT : Dithiothréitol

### E

EMT : Transition épithélio-mésenchymateuse (*Epithelial to mesenchymal transition*)

EPCAM : *EPithelial Cell Adhesion Molecule*

## **G**

GALAD : *Gender, Age, AFP-L3, AFP, and Des-carboxy-prothrombin*

GWAS : Etude d'association pan-génomique (*Genome Wide Association Study*)

## **H**

HIF1- $\alpha$  : *Hypoxia inducible factor 1 alpha*

HK : Hexokinase

HRE : Élément de réponse à l'hypoxie

HSC : Cellule étoilée hépatique (*Hepatic stellate cell*)

## **I**

IDH : isocitrate déshydrogénase

IFN $\alpha$ : Interféron alpha

## **K**

KEGG : *Kyoto Encyclopedia of Genes and Genomes*

*KRT19* ou *CK 19*: Cytokératine 19

## **L**

LINE: Long élément nucléaire intercalé (*Long interspersed nuclear elements*)

LPS : Lipopolysaccharide

LTR: Séquence terminale longue répétée (*Long Terminal Repeats*)

## **M-N**

miR : microARN

MMP : Métalloprotéinase

mTOR : *mammalian Target of rapamycin*

NAFLD : Stéatose hépatique non liée à l'alcool (*Non alcoholic fatty liver disease*)

NASH : Stéatohépatite non liée à l'alcool (*Non alcoholic steato-hepatitis*)

## **O**

OMICS : néologisme anglais décrivant la caractérisation et la quantification globales de molécules biologiques

OMS ou WHO : Organisation mondiale de la santé

ORF : Cadre de lecture ouvert (*Open reading frame*)

## **P-R**

PCR : Amplification en chaîne par polymérase (*Polymerase chain reaction*)

PK : Pyruvate kinase

pVHL : *Von Hippel-Lindau tumor suppressor*

ROS : Espèces réactives à l'oxygène (*Reactive oxygen species*)

RT : transcription inverse (*reverse transcription*)

## T

TACE : Chimio-embolisation trans-artérielle (*Trans-arterial chemoembolization*)

TCGA : *The Cancer Genome Atlas*

TERT : Sous-unité catalytique de la télomérase (*Telomerase reverse transcriptase*)

TG : Triglycérides

TGF- $\beta$  : *Transforming Growth Factor beta*

TLR4 : *Toll like receptor 4*

TUNEL : *Terminal Transferase dUTP Nick End Labeling*

## V

VHB : Virus de l'hépatite B

VHC : Virus de l'hépatite C

VIH : Virus de l'immunodéficience humaine

## Index des figures

Figure 1 : Épidémiologie du carcinome hépatocellulaire dans le monde

Figure 2 : La stéatose hépatique

Figure 3 : Algorithme décisionnel pour la prise en charge d'un patient cirrhotique

Figure 4 : Grades de différenciation selon la classification d'Edmondson et Steiner

Figure 5 : Classification BCLC (Barcelona Clinic Liver Cancer) pour le traitement du CHC

Figure 6 : Structure du rétrotransposon L1 et rétrotransposition

Figure 7 : Voie Wnt/ $\beta$ -caténine

Figure 8 : Transport et métabolisme du glucose des cellules tumorales et normales

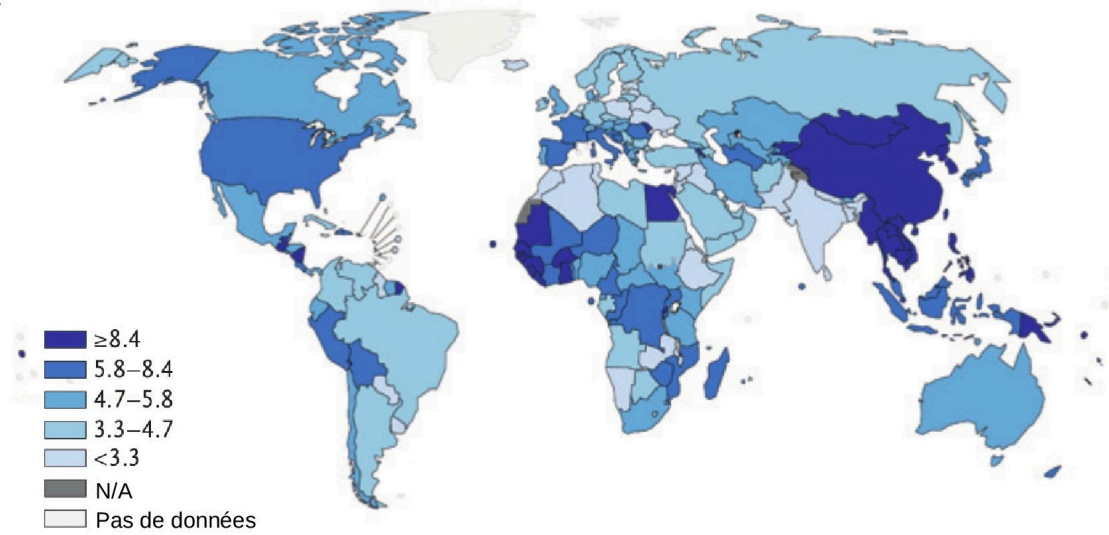
Figure 9 : La protéine HIF1- $\alpha$  dans des conditions normoxiques et hypoxiques

Figure 10 : Classification moléculaire des CHC à 6 groupes selon Boyault *et al.*

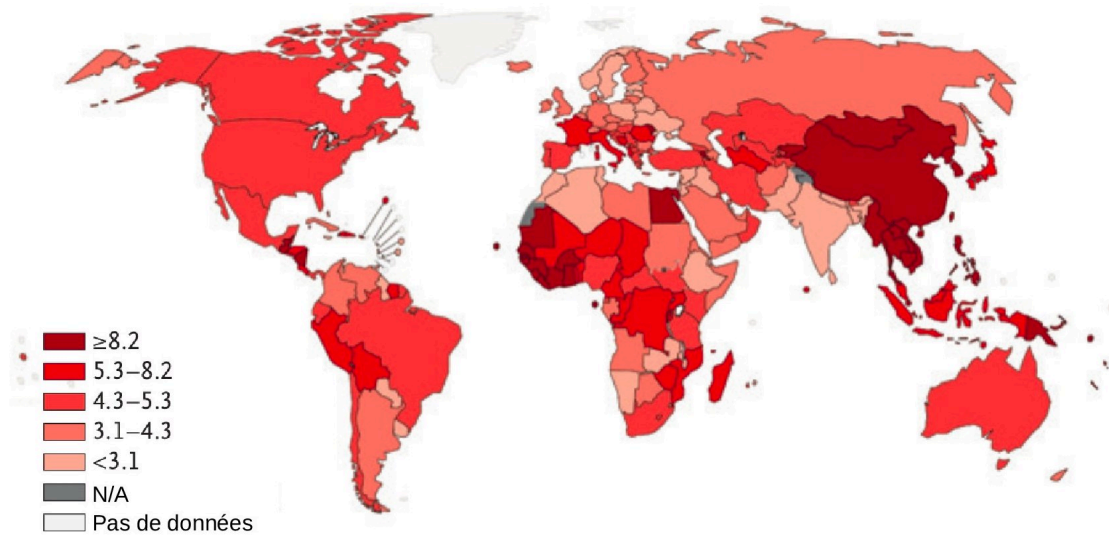
Figure 11 : Classification moléculaire des CHC à 3 groupes selon Hoshida *et al.*

Figure 12 : Classification moléculaire des CHC basée sur le profil immunitaire des tumeurs et de leur microenvironnement

A



B



**Figure 1: Epidémiologie du carcinome hépatocellulaire dans le monde** Les données (octobre 2018) sont tirées de l'Organisation Mondiale de la Santé. Les données pour l'incidence (A) et la mortalité (B) sont exprimées pour 100 000 personnes.

---

# I INTRODUCTION

## 1. Épidémiologie du carcinome hépatocellulaire

Le carcinome hépatocellulaire (CHC) est la tumeur maligne primitive du foie la plus fréquente et l'une des plus meurtrières dans le monde, tous cancers confondus (deuxième en terme de mortalité).

En moyenne 840 000 nouveaux cas par an surviennent dans le monde avec un rapport incidence/mortalité proche de 1 (780 000 décès par an). Les hommes sont davantage affectés par la maladie que les femmes avec un sexe ratio de 3:1 des cas diagnostiqués (Ferlay *et al.*, 2019).

Il existe de fortes disparités géographiques. La Chine, le sud-est asiatique, le Japon et l'Afrique sub-saharienne sont des zones de forte incidence avec 20 nouveaux-cas pour 100 000 habitants, alors qu'en Scandinavie et au Canada, l'incidence est plus faible (moins de 5/100 000). En France et aux USA, l'incidence du CHC est moyenne (entre 5 et 10 par 100 000 habitants). Ces variabilités géographiques sont liées à des expositions variables aux facteurs de risque de CHC. En Europe centrale et de l'est, l'alcool est le principal facteur de risque alors qu'en Asie et en Afrique, l'infection par le virus de l'hépatite B est prédominante (Akinyemiju *et al.*, 2017). La Figure 1 montre l'incidence du CHC dans le monde d'après l'organisation mondiale de la santé.

En France, selon les dernières estimations de l'Institut National de Veille Sanitaire, l'incidence en 2017 était de 8 273 nouveaux cas par an chez les hommes et 2 432 cas chez les femmes (Jéhannin-Ligier *et al.*, 2017)

## 2. Étiologies du carcinome hépatocellulaire

### 2.1 Étiologies virales

Le CHC se développe principalement sur un foie inflammatoire chroniquement endommagé, présentant des cycles répétés de nécrose/régénération, une hépatite chronique, une fibrose et finalement une cirrhose. Dans un contexte d'infection par des virus hépatotropes, notamment les virus de l'hépatite B et de l'hépatite C, qui sont des facteurs de risque majeurs du CHC, le foie est le site de production de nombreux médiateurs inflammatoires et oxydatifs en réponse à

l'infection. De nombreuses cellules immunes participent également à la maintenance de l'homéostasie hépatique. En cas de régénération hépatique, les lymphocytes T *via* notamment la cytokine lymphotoxine régulent ce processus indispensable à la vie du foie (Tumanov *et al.*, 2009). En réponse à l'infection virale, et donc à une inflammation prolongée, la création d'un environnement prolifératif peut conduire à un remodelage du parenchyme hépatique, une fibrose et éventuellement à long terme un cancer (Pikarsky *et al.*, 2004; Haybaeck *et al.*, 2009). Les virus des hépatites B et C sont également capables d'induire une carcinogenèse hépatique par des mécanismes directs.

### 2.1.1 Le virus de l'hépatite B

L'infection par le virus de l'hépatite B (VHB) est le facteur de risque du CHC le plus important au monde (la moitié des cas) (Parkin, 2006; Venook *et al.*, 2010).

Il existe dans le monde plus de 350 millions de porteurs chroniques du VHB et cette infection cause plus d'un million de décès chaque année. C'est un virus responsable d'une infection aiguë susceptible d'évoluer vers la chronicité. L'infection aiguë est le plus souvent asymptomatique. Elle évolue vers la guérison dans plus de 90 % des cas. L'hépatite fulminante est rare, observée en moyenne dans un cas sur 1 000. Le risque de passage vers la chronicité est estimé à 5-10 %. La cirrhose est habituellement observée chez 20 % des sujets, 15 à 20 ans en moyenne après la survenue de l'infection. Le CHC est diagnostiqué 10 à 15 ans après celui de la cirrhose (Tang *et al.*, 2018). Le risque de passage à la chronicité est, par contre, beaucoup plus élevé chez les enfants car il est de 25 à 50 % quand l'infection survient entre 1 et 5 ans et peut atteindre 90 % en cas d'infection néo-natale (Abdel-Hady and Kelly, 2013).

Le virus du VHB est un virus enveloppé possédant un génome viral à ADN partiellement double brin de 3,2 kb. Le génome code une protéine de capsid, une ADN polymérase, des protéines d'enveloppe (S, Pré-S1 et pré-S2), l'antigène e et la protéine HBx qui est une protéine transactivatrice. Le risque de développer un CHC est lié, au niveau de la réplication virale, au génotype viral et à la présence de mutations génomiques. Une équipe taïwanaise a montré que la survenue du CHC est corrélée au nombre de copies d'ADN du virus retrouvé dans le sérum des patients infectés (Chen *et al.*, 2006). Il a également été montré que le génotype C, par comparaison avec le génotype B, est un facteur de risque plus important dans la survenue du cancer, d'autant plus s'il est associé à une cirrhose (Chan *et al.*, 2004) Pour autant les auteurs ne concluent pas que ce génotype a un effet plus important sur la carcinogenèse hépatique.

## 2.1.2 Le virus de l'hépatite C

Le VHC est un virus à ARN de polarité positive appartenant à la famille des *flaviviridae* (virus de la dengue, fièvre jaune, etc.). On estime que 170 millions de personnes sont infectés par le VHC dans le monde. Il n'existe pas de vaccins prévenant la survenue de l'infection par le VHC. Le traitement curatif de choix était jusqu'à peu l'association de l'interféron alpha (IFN- $\alpha$ ) pégylé avec la ribavirine (analogue nucléosidique de synthèse). La réponse au traitement varie en fonction des génotypes (Flamm, 2003). La bithérapie est plus efficace pour les génotypes 2 et 3 (80 % de bons répondeurs) que pour le génotype 1 (autour de 40 %) (Boyer and Marcellin, 2002). La résistance au traitement peut s'expliquer par une forte réplication virale, un taux important de mutations génétiques favorisant l'émergence de mutants résistants, une demi-vie très courte des virions et par l'action de certaines protéines virales comme la protéine NS5A dans le mécanisme de résistance à l'interféron (Enomoto *et al.*, 1996). Depuis peu, l'émergence de nouveaux principes actifs, appelés agents antiviraux directs (AAD) permettent d'obtenir des taux de guérison supérieurs à 95% chez des patients jamais traités mais aussi chez des patients en échec de traitements antérieurs. Les différentes classes thérapeutiques disponibles sont les inhibiteurs de la protéase NS3/4A (siméprévir, paritaprévir...), les inhibiteurs de NS5A (daclatasvir, ledipasvir...) et les inhibiteurs de la polymérase NS5B (sofosbuvir, dasabuvir...) (Gane *et al.*, 2013). En raison du coût extrêmement élevé de ces médicaments et d'une moindre réponse clinique des génotypes viraux 2 et 3, la recherche de nouvelles stratégies thérapeutiques reste une priorité pour les chercheurs. Récemment, une équipe américaine a proposé l'utilisation d'un anti-allergique peu coûteux dans le traitement de l'infection (He *et al.*, 2015), mais il reste encore de nombreux tests à effectuer avant que celui-ci ne soit autorisé à être administré aux patients infectés.

## 2.2 Étiologies non virales

### 2.2.1 Alcool

La consommation abusive d'alcool est connue pour augmenter le risque de développer un CHC en favorisant le développement d'une stéatohépatite et d'une cirrhose (Mittal and El-Serag, 2013). De plus le risque de développer une cirrhose est une relation linéaire à la consommation d'alcool (Scoccianti *et al.*, 2016).

L'alcool est aussi connu pour agir en synergie avec d'autres facteurs de risque et en particulier avec les virus des hépatites. Le risque de développer un CHC augmente de 2 à 3 fois chez des



patients consommateurs de plus de 60 g/jour d'alcool lorsqu'ils sont respectivement infectés par le VHC ou le VHB (Ikeda *et al.*, 1998; Hutchinson *et al.*, 2005).

### 2.2.2 Facteurs environnementaux toxiques

L'exposition à des produits chimiques en particulier chez certaines catégories d'ouvriers est responsable d'un nombre substantiel de cas de cancers professionnels même s'il est difficile d'estimer le nombre de cas, en raison du manque d'informations concernant la durée d'exposition et l'interaction avec d'autres carcinogènes (Jiang *et al.*, 2017). Cependant, les rares études s'intéressant au sujet ont permis d'identifier des substances comme le chlorure de vinyle (Fred *et al.*, 2008) ou le trichloroéthylène (Uccello *et al.*, 2012) comme ayant un effet carcinogène sur le foie et pour lesquelles il y a suffisamment de preuves chez l'homme et chez l'animal pour les incriminer directement dans la pathogenèse des cancers hépatiques.

## 2.3 Facteurs de risque génétiques

La plupart des prédispositions génétiques à la survenue d'un CHC sont des maladies métaboliques. On retrouve l'hémochromatose (mutation C282Y du gène HFE1) se caractérisant par une surcharge ferrique dans le foie, la maladie de Wilson (mutations dans le gène ATP7B) conduisant à un défaut d'excrétion hépatique du cuivre, le déficit en alpha-1 antitrypsine (mutations du gène SERPINA1), la tyrosinémie de type 1 (mutations du gène FAH) et la porphyrie aiguë intermittente (mutations du gène HMBS).

## 3. Histoire naturelle de la carcinogenèse hépatique

Le carcinome hépatocellulaire se développe principalement sur une hépatite chronique. La fibrose hépatique est une pathologie résultant d'une prolifération anormale de tissus cicatriciels composés de collagènes en réponse à une hépatite chronique. La cirrhose est le stade ultime de la fibrose hépatique et se définit par une déstructuration de l'architecture du foie et le développement de nodules de régénération. Cliniquement, la cirrhose peut être associée à une hypertension portale et à une insuffisance hépatique de sévérité variable. Au sein des nodules cirrhotiques, les hépatocytes font l'objet de cycles répétés de mort cellulaire et de prolifération compensatrice pouvant fournir un environnement mutagène propice à la transformation cellulaire et au développement tumoral. Les causes de la cirrhose sont multifactorielles et sont principalement dues aux infections par les virus hépatotropes B et C, et une consommation abusive d'alcool. D'autres facteurs étiologiques comme une agression hépatique par des

aflatoxines, l'existence d'une maladie métabolique héréditaire ou l'épidémie d'obésité et le développement d'une stéatohépatite sont également des causes de cirrhose. Les mécanismes conduisant à la fibrose hépatique semblent cependant indépendants de l'étiologie.

### 3.1 De l'hépatite chronique au CHC

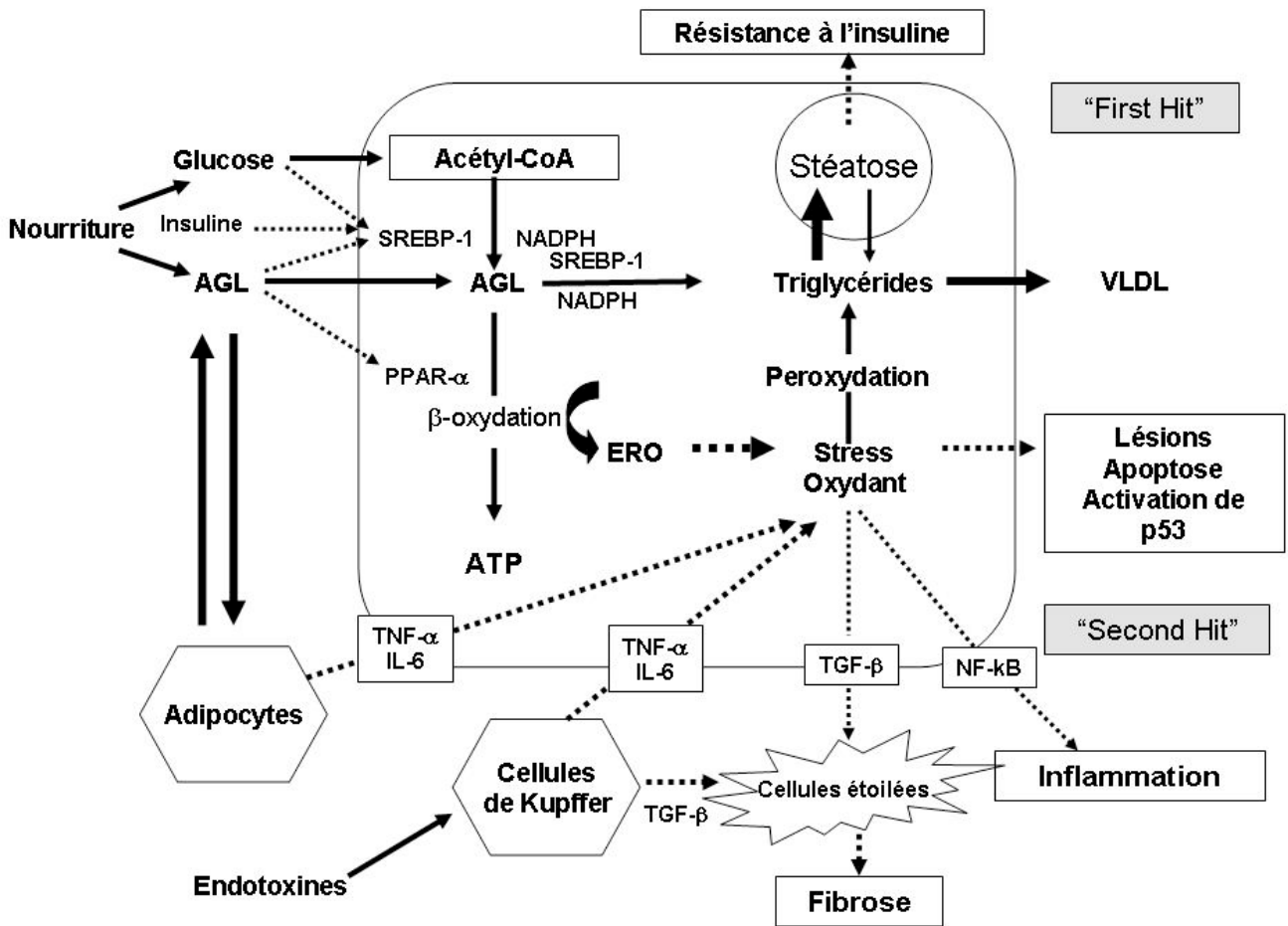
#### 3.1.1 L'hépatite chronique viro-induite

L'infection par le VHC est une cause majeure d'hépatite chronique. La physiopathologie de l'hépatite C chronique est liée à l'inflammation induite par la réponse immune dirigée contre les antigènes du virus. De plus, les patients atteints d'hépatite chronique virale C développent, dans 40 à 60 % des cas, une stéatose. Particulièrement, chez les patients infectés par le génotype 3, le degré de stéatose est significativement corrélé à la concentration intra-hépatique en ARN du VHC, suggérant l'implication directe du VHC dans la genèse de la stéatose (Rubbia-Brandt *et al.*, 2000, 2001).

Comme pour le VHC, l'infection par le virus de l'hépatite B provoque des cycles de nécrose-régénération du tissu hépatique qui progresse en fibrose, et dans le contexte dû à la persistance de l'inflammation, participe à l'établissement d'un terrain favorable à l'émergence d'altérations génétiques et épigénétiques conduisant à la tumorigenèse (Neuveut *et al.*, 2010). Le virus de l'hépatite B a, de plus, des effets directs sur la dynamique cellulaire qui seront décrits dans les paragraphes (Intégration par le virus de l'Hépatite B p36 et HBx p40)

#### 3.1.2 L'hépatite chronique alcoolique

L'action de l'alcool sur la carcinogenèse hépatique est multiple. L'acétaldéhyde, produit du métabolisme de l'éthanol par l'aldéhyde déshydrogénase (protéine cytosolique) et le cytochrome P450 2E1 (protéine du réticulum endoplasmique), est connu pour former des adduits avec l'ADN génomique (Brooks and Theruvathu, 2005) et inhiber l'enzyme responsable des réparations de l'ADN suite à l'action d'agents alkylants (Espina *et al.*, 1988). Ces adduits peuvent conduire à des mutations et à des erreurs de réplication dans des oncogènes et des gènes suppresseurs de tumeur (Purohit *et al.*, 2013). L'acétaldéhyde est capable également de faciliter l'invasion cellulaire en augmentant la synthèse des métalloprotéinases (MMPs) comme la MMP 9 (Hsiang *et al.*, 2007). La consommation d'alcool active fortement l'expression du cytochrome P450 2E1. Cette enzyme, qui catalyse de nombreuses réactions d'oxydation de toxiques dans le foie, génère un grand nombre d'espèces réactives de l'oxygène (ROS), augmentant ainsi la peroxydation des lipides et la production d'un métabolite particulier, le 4-



**Figure 2: La stéatose hépatique** Mécanismes de formation et d'évolution de la stéatose hépatique (« first-hit ») vers un état inflammatoire (« second-hit ») aboutissant à la NASH, la fibrose et à plus long terme la cirrhose. D'après Jansen *et al.* 2004.

hydroxynonenal (4-HNE) (Hu *et al.*, 2002). Ce dernier forme des adduits dans l'ADN préférentiellement à la troisième base du codon 249 du gène TP53 (Hu *et al.*, 2002).

D'autres mécanismes, comme la perte chromosomique, le stress oxydatif, la méthylation de l'ADN et la réduction du taux de l'acide rétinolique dans le foie, contribuent au développement d'un CHC sur une pathologie alcoolique (Purohit *et al.*, 2013).

### 3.1.3 NAFLD et NASH

Le syndrome métabolique est un ensemble de plusieurs troubles physiologiques et biochimiques. Il est défini par une obésité abdominale et au moins deux critères de la liste ci-après (taux élevé de triglycérides sérique, taux inférieur à la normale de HDL, une hypertension artérielle et une hyperglycémie à jeun) (Parikh and Mohan, 2012). Dans la majorité des cas, le syndrome métabolique conduit à une insulino-résistance et à une stéatose hépatique.

La stéatopathie hépatique non alcoolique (Non Alcoholic Fatty Liver Diseases, NAFLD) inclut un large spectre d'anomalies hépatiques de la stéatose à la stéatohépatite (Non Alcoholic Steato-Hepatitis, NASH), à la fibrose, la cirrhose et le CHC (cf. Figure 2). La stéatose simple NAFLD est connue pour être associée à l'obésité, la résistance à l'insuline, le diabète et l'hypertriglycéridémie. La prévalence de la NAFLD est en constante augmentation affectant environ un quart de la population mondiale (Møller *et al.*, 1994; Bergström *et al.*, 2001; Wolk *et al.*, 2001; Calle *et al.*, 2003; Younossi *et al.*, 2016). La NAFLD est à la fois la cause et la conséquence du syndrome métabolique (Yki-Järvinen, 2014).

La stéatose hépatique se caractérise par l'accumulation intracellulaire de triglycérides (TG) et de diacylglycérols (DAG) contenus dans des vésicules lipidiques de différentes tailles (stéatose macro- et micro-vésiculaire). L'accumulation de ces lipides est le résultat d'un déséquilibre dans le transport des lipides, l'oxydation et la synthèse *de novo* des lipides (Ginsberg, 2006; Goldberg and Ginsberg, 2006). C'est un phénomène qui peut être réversible mais peut également progresser en stéatohépatite (NASH) (Wong *et al.*, 2010) voire en CHC sur un foie non-cirrhotique (Paradis *et al.*, 2009).

Les facteurs responsables du passage de la stéatose à la NASH restent encore à déterminer ; de nombreuses études font état de la toxicité cellulaire directe des lipides accumulés (Lee *et al.*, 1994; Cusi, 2012), et de la dérégulation du métabolisme lipidique (Puri *et al.*, 2007; Caballero *et al.*, 2009; Min *et al.*, 2012) (voir également le chapitre 9.3 en page 60), générateurs d'inflammation. Récemment, une étude de notre laboratoire a, d'une part, identifié une

signature lipidique spécifique présente dans les foies de patients souffrant de NASH et d'autre part, montré la forte lipotoxicité de ce set de lipides sur des hépatocytes en culture primaire (Chiappini *et al.*, 2017). En effet, il a été observé une augmentation importante de longs acides gras saturés et insaturés (jusqu'à C18) et une baisse de très longs acides gras poly-insaturés dus à une réduction d'activité de l'enzyme FADS1 nécessaire à la synthèse des phospholipides.

Le syndrome métabolique, l'obésité viscérale et la NAFLD sont des maladies inflammatoires. L'obésité viscérale est caractérisée par des niveaux importants de sécrétion de cytokines pro-inflammatoires (TNF- $\alpha$ , IL-1 $\beta$ , IL-6, MCP-1) ce qui altère la sensibilité à l'insuline notamment des adipocytes et induit la polarisation des macrophages en phénotype M1 conduisant à l'initiation et à l'entretien de l'inflammation dans le foie et la progression vers des états pathologiques plus sévères incluant le CHC (Xu *et al.*, 2003; Park *et al.*, 2010).

La progression vers un CHC n'est évidemment pas uniquement liée à ce processus inflammatoire. La dérégulation de voies oncogéniques comme la voie PI3K-Akt-mTOR-PTEN (Leavens *et al.*, 2009; Kudo *et al.*, 2011) et des anomalies mitochondriales (Bournat and Brown, 2010; Montgomery and Turner, 2015) sont également des facteurs importants de progression vers la carcinogenèse.

### 3.2 Fibrogenèse hépatique

Les cellules étoilées du foie (HSCs, hepatic stellate cells) jouent un rôle clé dans la fibrogenèse hépatique. Ces cellules sont localisées dans l'espace de Disse entre la face sinusoidale des hépatocytes et les cellules endothéliales, et ont différents rôles incluant le stockage de la vitamine A (Hendriks *et al.*, 1987), le contrôle du tonus vasomoteur (Kawada *et al.*, 1993) et surtout la sécrétion et la dégradation de la matrice extracellulaire (Clement *et al.*, 1986).

L'agression des hépatocytes par des virus ou des toxines de diverses natures est à l'origine d'une inflammation associée à une infiltration cellulaire composée notamment de cellules de Küpfer, de macrophages et de plaquettes. Les HSC sont par la suite activées par des cytokines pro-inflammatoires et des facteurs de croissance qui sont sécrétés par des cellules immunes infiltrantes. À l'état activé, les HSC prolifèrent et se transforment en myofibroblastes sécrétant des molécules pro-fibrogéniques (TGF- $\beta$ , CTGF, PDGF) (Paradis *et al.*, 2002; Borkham-Kamphorst *et al.*, 2007; Inagaki and Okazaki, 2007), du collagène (Tsukada *et al.*, 2006) et des enzymes de dégradation de la matrice extracellulaire (métalloprotéases MMP2 et MMP9) (Arthur *et al.*, 1992; Han *et al.*, 2007).

### 3.2.1 La cirrhose

La cirrhose est le stade ultime de la fibrose hépatique conduisant à de profondes modifications de l'architecture hépatique. Ce phénomène est souvent asymptomatique jusqu'au stade symptomatique de la décompensation. Elle est caractérisée sur le plan histologique par l'existence de barrières fibreuses mutilantes et vascularisées qui relient les voies portales entre elles et à la veine centrolobulaire. Les hépatocytes sont enchâssés dans des nodules fibrotiques appelés nodules de régénération. L'évolution de la cirrhose est la suivante : au cours d'une première phase, la cirrhose est non compliquée (cirrhose compensée) pouvant se développer de manière asymptomatique. Elle peut être découverte de manière fortuite lors d'un examen clinique « de routine » ou à l'occasion d'une chirurgie abdominale. Trente à 40 % des cas de cirrhose sont asymptomatiques. Un second temps est marqué par la survenue de complications (cirrhose décompensée) avec détérioration de l'état général, insuffisance hépatique et hypertension portale responsables de multiples complications comme l'apparition d'une ascite, une péritonite bactérienne spontanée, une hémorragie digestive ou une encéphalopathie hépatique (Schuppan and Afdhal, 2008).

Au cours de la cirrhose, la synthèse d'ADN est accélérée au sein des nodules de régénération et par conséquent les réarrangements chromosomiques sont plus fréquents, favorisant le développement de nodules hyperplasiques. Ces derniers peuvent évoluer en nodules dysplasiques qui sont des lésions pré-néoplasiques. Cette évolution s'accompagne d'une accumulation d'altérations moléculaires « gains de fonction » conduisant à l'acquisition d'activités biologiques et de capacités fonctionnelles nouvelles essentielles à l'initiation de la carcinogenèse hépatique et à sa progression.

## 3.3 Complications de la cirrhose

### 3.3.1 Ascite

L'ascite, définie comme la présence d'un excès de liquide dans la cavité péritonéale, est la principale complication de la cirrhose (Runyon and AASLD, 2013). L'ascite représente un stade avancé de progression vers une phase décompensée de la cirrhose et est associée à un pronostic défavorable avec une mortalité de 50 % dans les 2 ans qui suivent son apparition (Moore and Aithal, 2006). Elle résulte principalement de défaillances liées à l'insuffisance hépatique et à l'hypertension portale. L'infection du liquide d'ascite est également une complication grave chez 10 à 30 % des patients cirrhotiques (Ascione *et al.*, 2017).

### 3.3.2 Encéphalopathie

L'encéphalopathie hépatique est une complication neuropsychique sévère liée notamment à l'accumulation de substances neurotoxiques comme celles issues du métabolisme de la glutamine pris en charge par des bactéries du microbiote intestinal (Desjardins *et al.*, 2012). L'accumulation, dans le sang des patients, d'ammoniaque, qui n'est pas dégradé par un foie insuffisant, est à l'origine d'anomalies cérébrales structurales et fonctionnelles. L'existence d'une encéphalopathie hépatique est une indication à la transplantation hépatique.

### 3.3.3 Hypertension portale et translocation bactérienne

L'hypertension portale est une conséquence fréquente de la cirrhose et est reliée à des conséquences morbides comme des encéphalopathies et des saignements dans le tube digestif. De nombreuses voies de signalisation sont aujourd'hui connues comme ayant un rôle dans la pathogénicité de l'hypertension portale et dans la progression vers le cancer. L'hypertension portale provoque des changements structuraux majeurs dans l'intestin comme une congestion, un épaissement de la musculature muqueuse et une réduction drastique de l'intégrité des jonctions serrées (Iwakiri, 2014) ce qui a pour conséquence d'augmenter la perméabilité intestinale et la translocation bactérienne épithéliale en direction du foie. De plus, l'anabolisme des acides biliaires est profondément modifié chez les patients cirrhotiques alors que les acides biliaires sont connus pour exercer diverses fonctions visant à atténuer la translocation bactérienne et la prolifération bactérienne excessive par l'intermédiaire du récepteur nucléaire FXR (Kakiyama *et al.*, 2014; Ridlon *et al.*, 2015). L'hypertension portale perturbe donc fortement la perméabilité intestinale et facilite l'inflammation par le biais de la translocation bactérienne (Wiest *et al.*, 1999, 2014). Cette dernière est fréquente chez les patients avec cirrhose avancée (Cirera *et al.*, 2001) et des taux sériques importants d'endotoxines peuvent être retrouvés chez ces patients (Zhang *et al.*, 2012). Dans un modèle de carcinogenèse induite par le Diéthylnitrosamine (DEN), il a été montré que l'action combinée de la translocation bactérienne et de la dysbiose intestinale induite par une antibiothérapie conduit à une inflammation soutenue dans le foie et à l'apparition de lésions fibrotiques et néoplasiques (Zhang *et al.*, 2012).

La voie TLR4 (Toll like receptor 4) impliquée dans la reconnaissance des LPS bactériens présents dans des microbiotes intestinaux dysbiotiques semble avoir un impact pathogénique très important dans la carcinogenèse hépatique. La voie LPS/TLR4 favoriserait la prolifération hépatocytaire via la production de TNF $\alpha$  et d'IL-6 par les macrophages conduisant à une

progression accélérée de tumeurs chimiquement induites (Yu *et al.*, 2010). L'influence de cette voie a été confirmée dans un autre modèle de carcinogénèse induite (Dapito *et al.*, 2012). Les cellules cibles des LPS, à savoir les hépatocytes et les cellules étoilées du foie, entretiennent un dialogue moléculaire pour activer la voie NF- $\kappa$ B et favoriser la progression tumorale. De plus, la production hépatocytaire du mitogène, l'épiréguline, accentue le développement des nodules tumoraux. Le lien entre microbiote et cancers primitifs du foie a été formellement démontré dans des modèles animaux cancéreux par des expériences d'administration d'antibiothérapie chez des animaux immunocompétents et des expérimentations réalisées chez des souris axéniques. L'ensemble de ces animaux présentait un développement tumoral fortement réduit.

L'emploi d'antibiotiques, et surtout de probiotiques ou de prébiotiques pour corriger une dysbiose et, par là l'état pathologique associé, est reconnu comme étant une approche thérapeutique d'intérêt par de nombreuses équipes dans le monde. En effet, la dysbiose intestinale, secondaire en partie aux modifications qualitatives et quantitatives du pool des acides biliaires, chez les patients atteints d'hépatite chronique, contribuerait à l'inflammation chronique hépatique. Les espèces bactériennes anaérobies strictes productrices de butyrate, un acide gras à chaîne courte aux propriétés anti-inflammatoires et anti-tumorales (Fung *et al.*, 2012; Louis *et al.*, 2014; Wilson *et al.*, 2017) sont moins abondantes dans le microbiote fécal de patients avec hépatopathie chronique que chez des individus sains, au profit d'une surreprésentation d'espèces bactériennes aérotolérantes et anaérobies facultatives telles que des *Enterobacteriaceae* (Milosevic *et al.*, 2019). Cette dysbiose conduit à un changement de métabolisme des acides biliaires secondaires fermentés par les communautés bactériennes et entretient le déséquilibre bactérien intestinal. Une étude pionnière a montré que l'obésité induite chez des souris altère la composition du microbiote intestinal et augmente la production d'acide déoxycholique (DCA) qui provoque des dommages à l'ADN dans les hépatocytes (Yoshimoto *et al.*, 2013). Récemment, un probiotique (Prohep®) a montré des effets anti-carcinogènes dans un modèle de souris par modulation du microbiote intestinal, inhibition de l'angiogénèse et modulation de l'expression de cytokines anti-inflammatoires (IL-17) (Li *et al.*, 2016a).

La recherche de traitements ciblant l'axe microbiome intestinal - foie doit évidemment se poursuivre étant donné le rôle majeur de la translocation bactérienne épithéliale et à distance, et de la dysbiose dans la pathogénèse des hépatopathies chroniques et la carcinogénèse hépatique.



### 3.3.4 Varices œsophagiennes

Autre conséquence de l'hypertension portale est formation de varices dans la paroi de l'estomac ou de l'œsophage. La rupture de ces varices œsophagiennes peut entraîner une hémorragie digestive. Elle est l'une des principales causes de décès chez les patients cirrhotiques en absence de traitement (Garcia-Tsao and Bosch, 2015).

## 3.4 Du foie sain au CHC

Dans une proportion non négligeable de cas (20 % environ), le CHC peut se développer dans un foie non cirrhotique voire dans un parenchyme hépatique non remanié. Dans ce contexte, les tumeurs sont généralement plus volumineuses que celles développées sur foie cirrhotique (Gaddikeri *et al.*, 2014). Les actions directes des virus des hépatites, du syndrome métabolique et de l'exposition à des génotoxiques sont des facteurs de risque reconnus. L'adénome hépatocellulaire est une tumeur bénigne du foie qui peut se compliquer par un CHC (Farges *et al.*, 2011), d'autant plus lorsqu'il existe des anomalies de stockage du glycogène ou la prise de pilules contraceptives par voie orale (Farges and Dokmak, 2010; Shanbhogue *et al.*, 2011).

Enfin, un CHC développé sur foie sain est fréquent dans des maladies rares comme l'hémochromatose, la porphyrie, la maladie de Wilson ou encore la déficience en alpha-1 antitrypsine (Trevisani *et al.*, 2010).

## 4. Diagnostic du carcinome hépatocellulaire

Le pronostic sombre du CHC est souvent dû à une absence de symptomatologie clinique dans les stades précoces de la maladie. Plus de 60 % des patients sont diagnostiqués à des stades avancés (Altekruse *et al.*, 2009) et ont en conséquence une survie réduite comparée à celle des patients diagnostiqués à un stade précoce. Le dépistage des patients à plus haut risque de développer un CHC est primordial et est actuellement basé sur le dosage sérique de l'alpha foeto-protéine et de l'échographie abdominale. Depuis de nombreuses années, des efforts importants ont été menés dans le dépistage précoce des populations cirrhotiques à fort risque de développer un CHC au moyen de marqueurs non-invasifs biologiques et d'imagerie. Compte tenu des risques d'erreurs diagnostiques, de la médiocrité de performance en termes de sensibilité et spécificité des marqueurs disponibles, certains cliniciens privilégient des approches conventionnelles d'exploration d'un nodule suspect par ponction biopsique hépatique dirigée sous échographie ou scanner pour poser un diagnostic histologique (Blanc *et al.*, 2019).

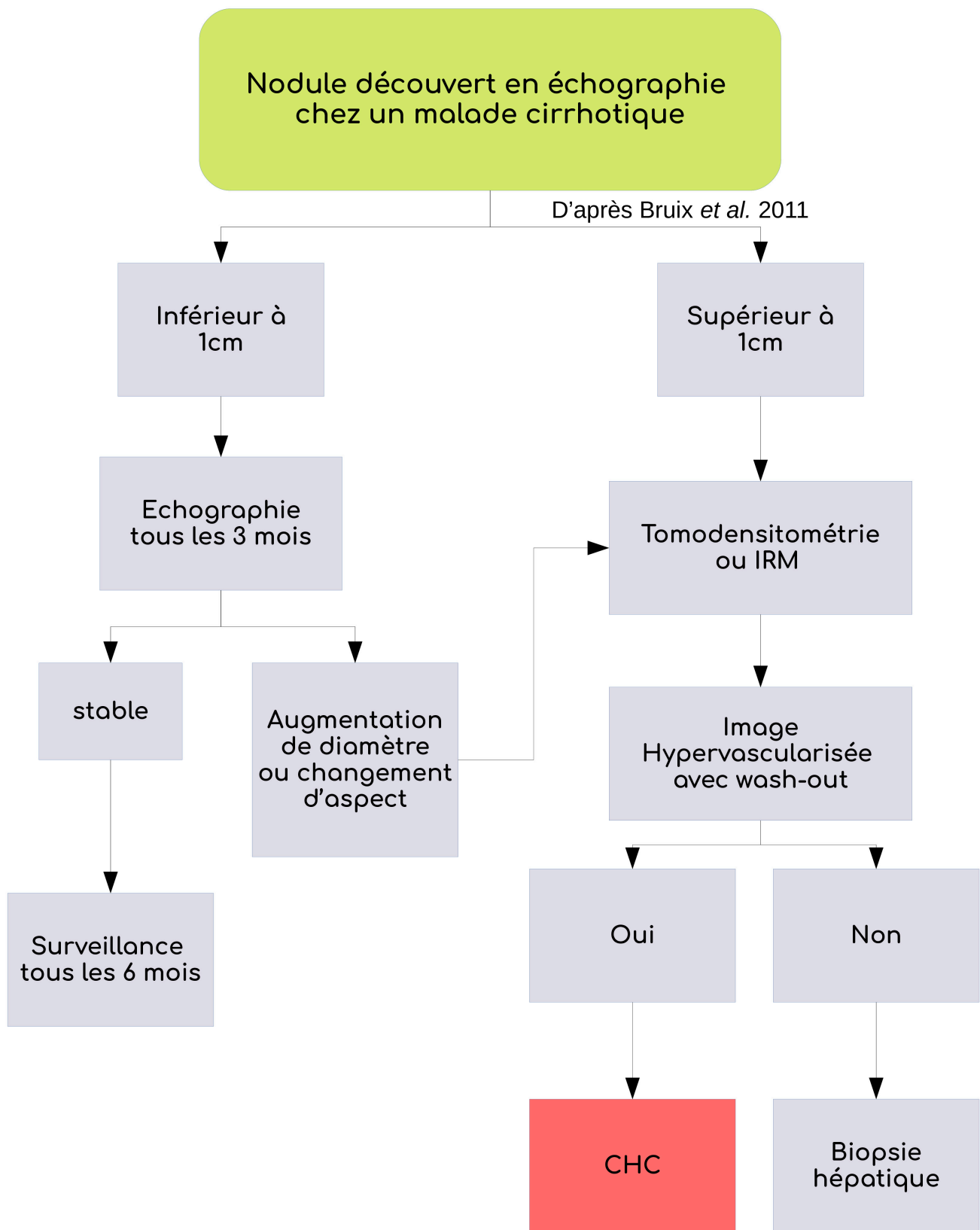


Figure 3: Algorithme décisionnel pour la prise en charge d'un patient cirrhotique

Le risque de dissémination tumorale sur le trajet de l'aiguille est variable selon les travaux. Certaines études ne rapportent aucune dissémination iatrogène (Bret *et al.*, 1988; Buscarini *et al.*, 1990; Nolsøe *et al.*, 1990) alors que d'autres estiment que ce risque est compris entre 2 et 5 % (Chapoutot *et al.*, 1999; Dumortier *et al.*, 2000; Takamori *et al.*, 2000).

## 4.1 Imagerie

La caractérisation des nodules tumoraux de CHC repose sur l'examen précis de la vascularisation par IRM ou échographie de contraste avec injection de produit de contraste. Le diagnostic de CHC peut reposer sur l'imagerie seule interprétée par un expert en associant deux méthodes radiologiques dès lors qu'il existe une hyper-vascularisation artérielle (Ronot and Vilgrain, 2014). En effet, en cas de découverte d'un nodule de taille supérieure à 1 cm hypervascularisé à la phase artérielle avec diminution de l'intensité du signal aux phases portale ou tardive, le diagnostic de CHC peut être retenu. Si le nodule est de taille inférieure à 1 cm, des imageries régulières (échographie) tous les 3 mois environ sont réalisées pour surveiller l'évolution du nodule (cf. Figure 3).

## 4.2 Diagnostic biologique (marqueurs)

### 4.2.1 Protéines sériques

L'alpha foeto-protéine (AFP) est une protéine onco-foetale qui fût pendant très longtemps le seul marqueur non invasif disponible pour le diagnostic du CHC et le plus largement utilisé. Malheureusement les performances de ce marqueur sont limitées. La plupart des études rapporte une sensibilité d'environ 60 % et une faible spécificité (Sato *et al.*, 1994; Adachi *et al.*, 2003; Marrero *et al.*, 2009; Bruix *et al.*, 2016). D'autres biomarqueurs diagnostiques potentiels sont en développement, parmi lesquels l'ostéopontine, une phosphoprotéine de la matrice extracellulaire qui se lie aux intégrines. Cette dernière est retrouvée abondante dans les tissus tumoraux de patients atteints de CHC (Gotoh *et al.*, 2002). Les performances diagnostiques de l'ostéopontine sont comparables à celles de l'AFP et la combinaison des deux biomarqueurs améliorerait le diagnostic avec une sensibilité de 77 % et une spécificité de 82 % dans une étude par méta-analyse (Sun *et al.*, 2018). D'autres études prennent en compte l'hétérogénéité tumorale, en combinant des marqueurs cliniques et protéiques pour l'établissement de scores composites. L'exemple le plus avancé, actuellement en phase III de développement clinique, est le score GALAD qui combine le sexe, l'âge, le taux de l'AFP, de l'AFP-L3, une isoforme de l'AFP, et de la des-carboxy-prothrombine (DCP). Ce score, évalué sur une cohorte de 670

patients (331 avec CHC, 339 avec maladie chronique du foie dont la cirrhose) et validé sur des cohortes externes européennes et asiatiques a une performance diagnostique supérieure à celle de l'échographie hépatique (Johnson *et al.*, 2014; Yang *et al.*, 2019). La validation du score GALAD sur des cohortes prospectives est en cours d'étude (Sengupta and Parikh, 2017).

#### 4.2.2 Autres marqueurs

L'amélioration des technologies d'analyse haut débit a permis d'isoler de nombreux types de molécules dans le sérum des patients avec hépatopathie chronique avec ou sans CHC. C'est le cas des micro-ARNs circulants miR21 et miR199, qui, dans le cadre d'une étude clinique de phase II, sont proposés comme marqueurs de discrimination entre les patients cirrhotiques sans ou avec CHC (Bai *et al.*, 2019; Peng *et al.*, 2019). Les performances diagnostiques de ces micro-ARNs associés au dosage de l'AFP sont très bonnes (AUC=0,95).

Les petites molécules peuvent être identifiées par spectrométrie de masse et chromatographie haute résolution. Nous l'avons mentionné, de nombreuses voies métaboliques et de signalisation sont perturbées dans les cancers, notamment le CHC. Il devient donc indispensable d'identifier les métabolites issus de ces voies. L'un des métabolites prometteurs en terme de développement clinique, est le 1-méthyladénosine dont la mesure sérique sépare clairement, dans une étude chinoise, les patients atteints de CHC des individus sains (Chen *et al.*, 2011). Les performances diagnostiques de ce métabolite associé au dosage de l'AFP sont très bonnes (AUC=0,95).

Malgré des avancées significatives notamment technologiques dans le domaine des biomarqueurs issus du métabolome, ces méthodes souffrent d'un important écueil : de nombreux facteurs, comme le régime alimentaire ou des facteurs environnementaux peuvent modifier considérablement le métabolome indépendamment de la maladie sous-jacente (Monteiro *et al.*, 2013).

Des efforts d'identification de nouveaux outils diagnostiques ont porté sur la recherche de cellules tumorales circulantes (CTC) et de la quantification de l'ADN nu tumoral (cfDNA, circulating free DNA) circulant. Pour la détection des CTC, deux systèmes existent : la méthode ISET qui utilise un système de filtration des cellules en fonction de leur taille et le système CellSearch® qui, au moyen de billes magnétiques, capture des CTC *via* des marqueurs présents à leur membrane plasmique (EPCAM, KRT18, KRT19, CD45).

La première étude de détection de CTC dans le sang de patients atteints de CHC fut réalisée par une équipe française. Plus de la moitié des patients testés présentaient des CTC dans leur sang

circulant et la présence des CTC était corrélée à l'invasion tumorale et à la thrombose portale (Vona *et al.*, 2000). Une autre équipe a détecté des CTC exprimant le récepteur membranaire ASGR1 (asialoglycoprotein receptor 1), récepteur facilitateur de l'infection par le virus de l'hépatite B (Witzigmann *et al.*, 2016). Les CTC ont été retrouvées chez plus de 80 % des patients testés ainsi que chez des patients présentant un petit CHC (Xu *et al.*, 2011).

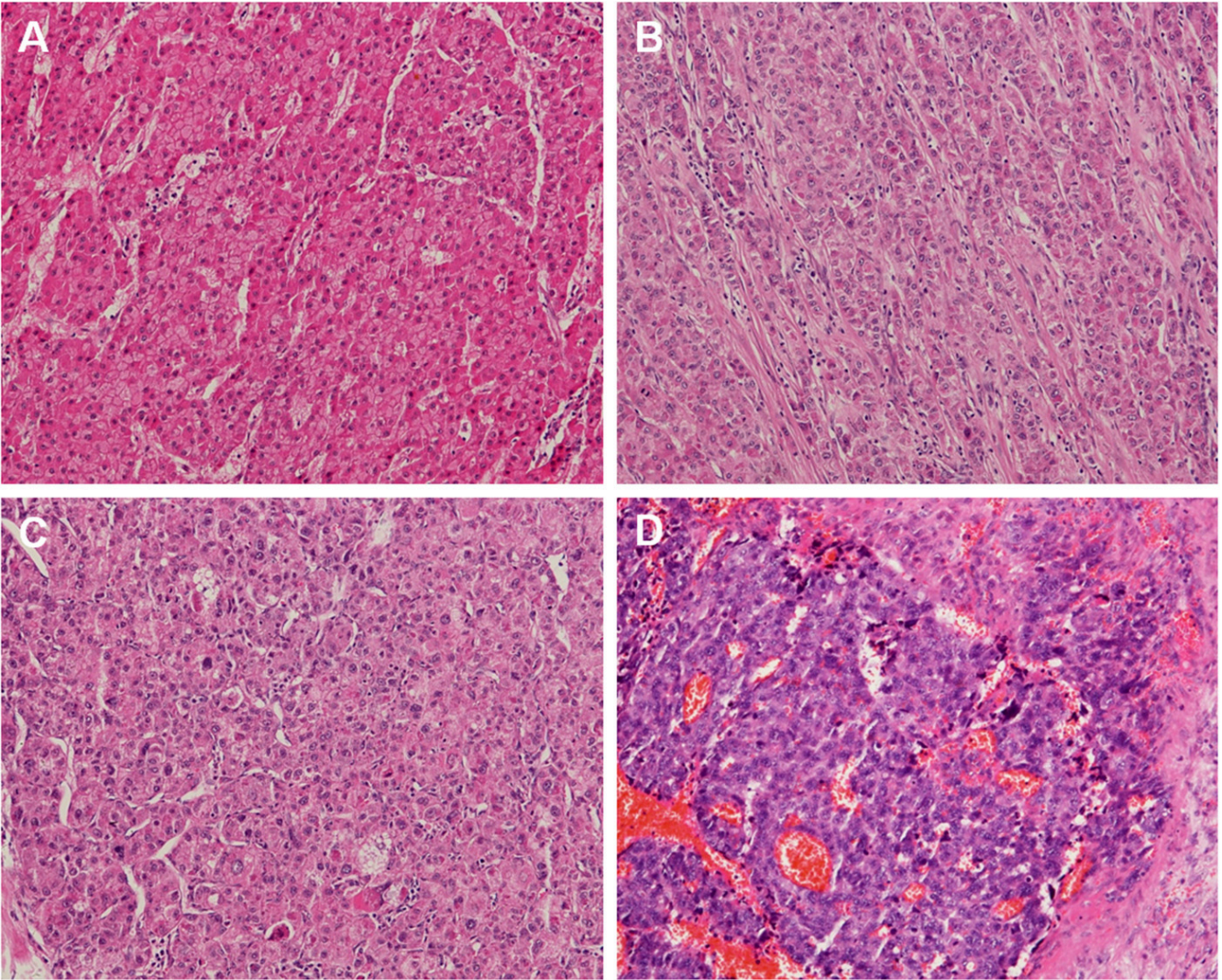
L'ADN tumoral circulant comporte des petits fragments d'ADN de 160 à 180 paires de bases (taille d'un nucléosome) qui circulent dans le sang de patients. Leur présence serait fortement liée à leur relargage par des cellules apoptotiques (Jung *et al.*, 2010; Jiang *et al.*, 2015). La concentration sanguine de l'ADN circulant serait associée à la taille des tumeurs (Iizuka *et al.*, 2006; Ren *et al.*, 2006), au grade tumoral (Iizuka *et al.*, 2006; Ng *et al.*, 2018) et pourrait servir de marqueur pronostique après une chirurgie de résection (Tokuhisa *et al.*, 2007). Toutefois, d'un point de vue diagnostique, l'intérêt de la détection des CTC est importante combinée au dosage de l'AFP (Liao *et al.*, 2015).

La possibilité qui est aujourd'hui offerte par le séquençage de nouvelle génération d'améliorer la caractérisation des CTC et la détection des mutations présentes dans le cfDNA peut dans un avenir proche conduire à l'établissement de tests de détection sensibles et très spécifiques pour ce type de cancer (Aravanis *et al.*, 2017).

### 4.3 Diagnostic histologique

Le CHC peut se présenter sous la forme d'une tumeur épithéliale nodulaire unique avec parfois des nodules satellites, d'une lésion diffuse et/ou multifocale, souvent polychrome, blanche, beige, jaune, ou verdâtre avec des remaniements nécrotiques et hémorragiques. La plupart des CHC expriment le marqueur antigénique hépatocytaire (HepPar1) et les cytokératines 8 et 18. Le pôle canaliculaire des hépatocytes est très souvent marqué par l'anticorps polyclonal anti-ACE (antigène carcinoembryonnaire). L'alpha foetoprotéine peut être exprimée dans environ 30 % des CHC, et les cytokératines 7 et 19, spécifiques des cellules biliaires, sont généralement absentes. Enfin la néo-angiogenèse tumorale est observée par le biais de marqueurs endothéliaux comme CD31, CD34 et BNH9.

L'apparence histologique varie fortement d'un patient à l'autre mais aussi à l'intérieur d'une tumeur d'une zone à l'autre. Il existe également différents grades de différenciation tumorale.



**Figure 4: Grades de différenciation selon la classification d'Edmondson et Steiner** Grades de différenciation selon Edmonson-Steiner, Grade I (A), Grade II (B), Grade III (C) et Grade IV (D). Images tirées de Zhou *et al.* 2017.

### 4.3.1 Profils architecturaux

Sur le plan microscopique, l'architecture des CHC varie avec des cellules qui peuvent être disposées en travées (trabéculaire) plus ou moins épaisses selon l'état de différenciation (plus la tumeur est dédifférenciée, plus les travées sont épaisses). Ces travées sont séparées par des sinusoides. Ce type architectural est souvent accompagné de structures pseudo-glandulaires (ou acineuse) ressemblant à des canaux avec parfois une production de bile. Les CHC peuvent présenter également des formes compactes où les espaces sinusoidaux sont très discrets. Il existe également des formes squirreuses où le stroma fibreux est très abondant et des CHC fibro-lamellaires qui sont des formes rares de CHC généralement développées chez des patients jeunes non-cirrhotiques avec des travées hépatocytaires en forme de feuilles mais aussi des formes de tumeurs ressemblant à un cholangiocarcinome.

### 4.3.2 Grades de différenciation

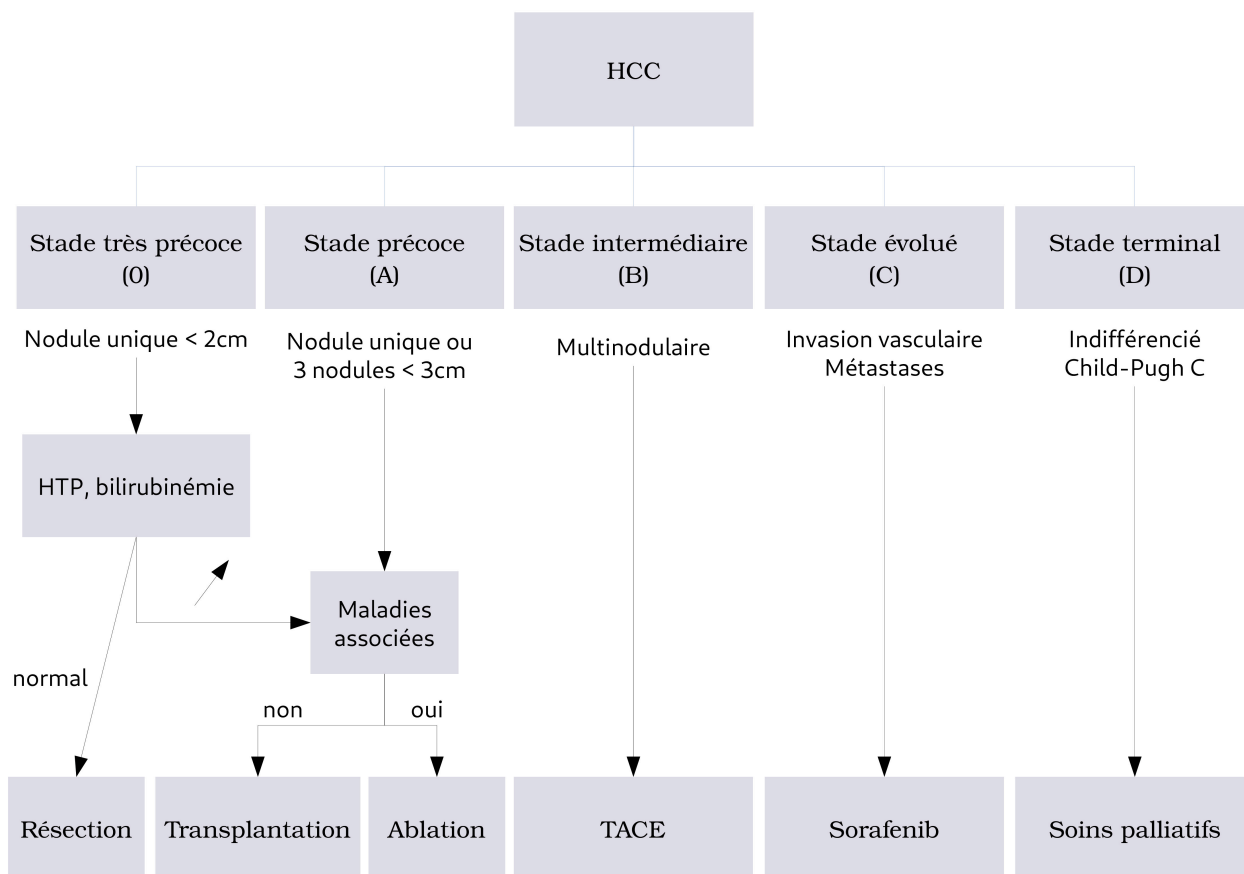
Les CHC peuvent présenter différents grades de différenciation estimés selon deux types d'échelle : le grade histologique d'Edmondson et Steiner développé par deux médecins américains dans les années 50, aujourd'hui encore utilisé (Edmondson and Steiner, 1954). L'Organisation Mondiale de la Santé (OMS) a développé un grade qui est très proche de celui d'Edmondson et Steiner. Ces grades tiennent compte de l'architecture tumorale et des atypies cyto-nucléaires (cf. Figure 4).

#### 4.3.2.1 Le CHC bien différencié

Il correspond au grade I et II d'Edmondson et Steiner, et est généralement retrouvé dans les petites tumeurs (de taille inférieure à 2 cm) à un stade précoce de la maladie. Il a une architecture trabéculée et pseudo-acineuse. Les cellules ont peu d'atypies avec un rapport noyau-cytoplasme augmenté. De plus, les cellules tumorales sont plus petites que les hépatocytes normaux. Dans le cas des tumeurs de taille supérieure à 3 cm, ce grade de différenciation peut être retrouvé à la périphérie du nodule tumoral.

#### 4.3.2.2 Le CHC moyennement différencié

Il correspond au grade III d'Edmondson et Steiner. Les travées sont plus épaisses, le cytoplasme est très éosinophile, le rapport noyau-cytoplasme est proche de celui observé dans les hépatocytes normaux. C'est le grade le plus fréquemment observé dans les CHC.



**Figure 5: Classification BCLC (Barcelona Clinic Liver Cancer) pour le traitement du CHC**

TACE : Chimioembolisation trans-artérielle ; HTP : Hypertension portale



### 4.3.2.3 Le CHC peu différencié

Il correspond au grade IV d'Edmondson et Steiner. La tumeur est compacte avec un rapport noyau-cytoplasme supérieur à celui des hépatocytes normaux. Les cellules sont pléomorphiques avec présence de cellules géantes mono- ou multinucléées et une forte activité mitotique.

## 5. Traitements du carcinome hépatocellulaire

La transplantation hépatique est à ce jour la meilleure option thérapeutique pour traiter les patients atteints de CHC. Malheureusement, la plupart des patients sont diagnostiqués à un stade avancé de la maladie ce qui contre-indique la transplantation hépatique. La surveillance systématique des patients à haut risque de développer un cancer permet de diagnostiquer précocement les patients et ainsi de leur offrir une relative large gamme de traitements (Forner *et al.*, 2012, 2019). L'évaluation du stade de la maladie est donc essentielle pour une prise en charge optimisée des patients. Parmi l'ensemble des classifications utilisées en pratique clinique, la classification BCLC (*Barcelona Clinic Liver Cancer*) est la plus consensuelle. Cet algorithme décisionnel, présenté en Figure 5, représente une aide importante pour une sélection personnalisée des options thérapeutiques.

### 5.1 Radiofréquence

L'ablation de la tumeur par radiofréquence est le traitement de choix pour un nodule unique de diamètre inférieur à 3 cm localisé dans un foie ayant conservé une bonne fonction hépatique. Les taux de survie à 5 ans sont larges entre 40 et 90 %. Il y a de fortes disparités dans l'estimation des taux de récurrence en fonction des études et des équipes cliniques. Ce taux peut être très faible (pour les nodules uniques inférieurs à 2 cm) (Livraghi *et al.*, 2008) pouvant atteindre 70 % (pour les nodules compris entre 3 et 4 cm) (Zytoon *et al.*, 2007; Montasser *et al.*, 2014).

### 5.2 Chimio-embolisation trans-Artérielle

La chimio-embolisation trans-artérielle (TACE) est un traitement adapté aux patients atteints de CHC de stade intermédiaire (BCLC B) et non éligibles à un traitement curatif. Associée à une chimiothérapie (de type doxorubicine ou cisplatine), la TACE entraîne une nécrose ischémique de la tumeur qui notamment déclenche l'activation fonctionnelle de protéines pro-angiogéniques, comme le VEGF, (Petrillo *et al.*, 2018a) et l'expression de marqueurs de cellules

progénitrices (Nahm *et al.*, 2017b), pouvant contribuer à un regain d'agressivité du tissu tumoral non détruit.

## 5.3 Traitements chirurgicaux

### 5.3.1 Résection

La résection chirurgicale est, comme la radiofréquence, adaptée au nodule unique développé sur foie non-cirrhotique. Les patients ne présentant pas d'hypertension portale avec une fonction hépatique préservée ont un taux de survie à 5 ans de 70 % mais un risque important de récurrence tumorale, dû notamment à la présence de foyers d'invasion micro-vasculaire (Wong and Lo, 2013).

### 5.3.2 Transplantation

La transplantation hépatique est indiquée chez des patients atteints de CHC développé sur foie cirrhotique, car elle traite à la fois la tumeur et la cause sous-jacente. Il a été établi par de nombreuses équipes médico-chirurgicales que les conditions idéales de transplantation devaient respecter les critères dits de Milan (Mazzaferro *et al.*, 1996), qui sont essentiellement des critères morphologiques (taille et nombre de nodules). Des critères supplémentaires comme le degré de différenciation tumorale et le taux sérique d'AFP sont également utilisés dans la décision thérapeutique, même s'il n'existe aucun consensus. Il est cependant de pratique courante de ne pas proposer de transplantation hépatique à des patients présentant une tumeur peu différenciée (Cillo *et al.*, 2004) ou ayant un taux sérique d'AFP élevé supérieur à 1000 ng/mL (Duvoux *et al.*, 2012; Toso *et al.*, 2015).

## 5.4 Thérapies ciblées

Comme nous l'avons vu précédemment, le CHC est le résultat d'un processus complexe multi-étapes comprenant des altérations génétiques et génomiques nombreuses et variées, qui affectent de nombreuses cascades de signalisation. Les efforts de classification moléculaire de ces tumeurs ont permis d'identifier quelques cibles thérapeutiques potentielles adaptées à chaque patient.

Le sorafenib est un inhibiteur de diverses kinases impliquées dans la progression tumorale. Deux essais cliniques de phase III (Llovet *et al.*, 2008; Cheng *et al.*, 2009) ont démontré son efficacité en améliorant la survie des patients et sa tolérance chez des patients à un stade tumoral avancé. D'autres molécules en phase II de développement ont été testées. Le sunitinib

(anti-VEGFR et anti-PDGFR), le brivanib (anti-VEGFR et anti-FGFR) ou l'everolimus (inhibiteur de mTOR) n'ont pas montré d'efficacité supérieure à celle du sorafenib (Park *et al.*, 2011; Zhu *et al.*, 2011; Finn *et al.*, 2012; Cheng *et al.*, 2013).

## 5.5 Immunothérapie

Des immunothérapies basées sur l'administration d'anticorps monoclonaux dirigés contre des points de contrôles immunitaires sont en cours de développement dans le CHC et font l'objet de multiples essais cliniques de phase II et III. Les traitements en cours d'évaluation sont les anticorps anti-PD1 (nivolumab, pembrolizumab) et les anticorps anti-CTLA4 (tremlimumab) qui semblent montrer des résultats encourageants chez certains patients non répondeurs au sorafenib (El-Khoueiry *et al.* 2017; Duffy *et al.* 2017; Zhu *et al.* 2018)

## 6. Principaux facteurs de carcinogenèse hépatique

### 6.1 Mutagenèse insertionnelle

#### 6.1.1 Intégration par le virus de l'Hépatite B

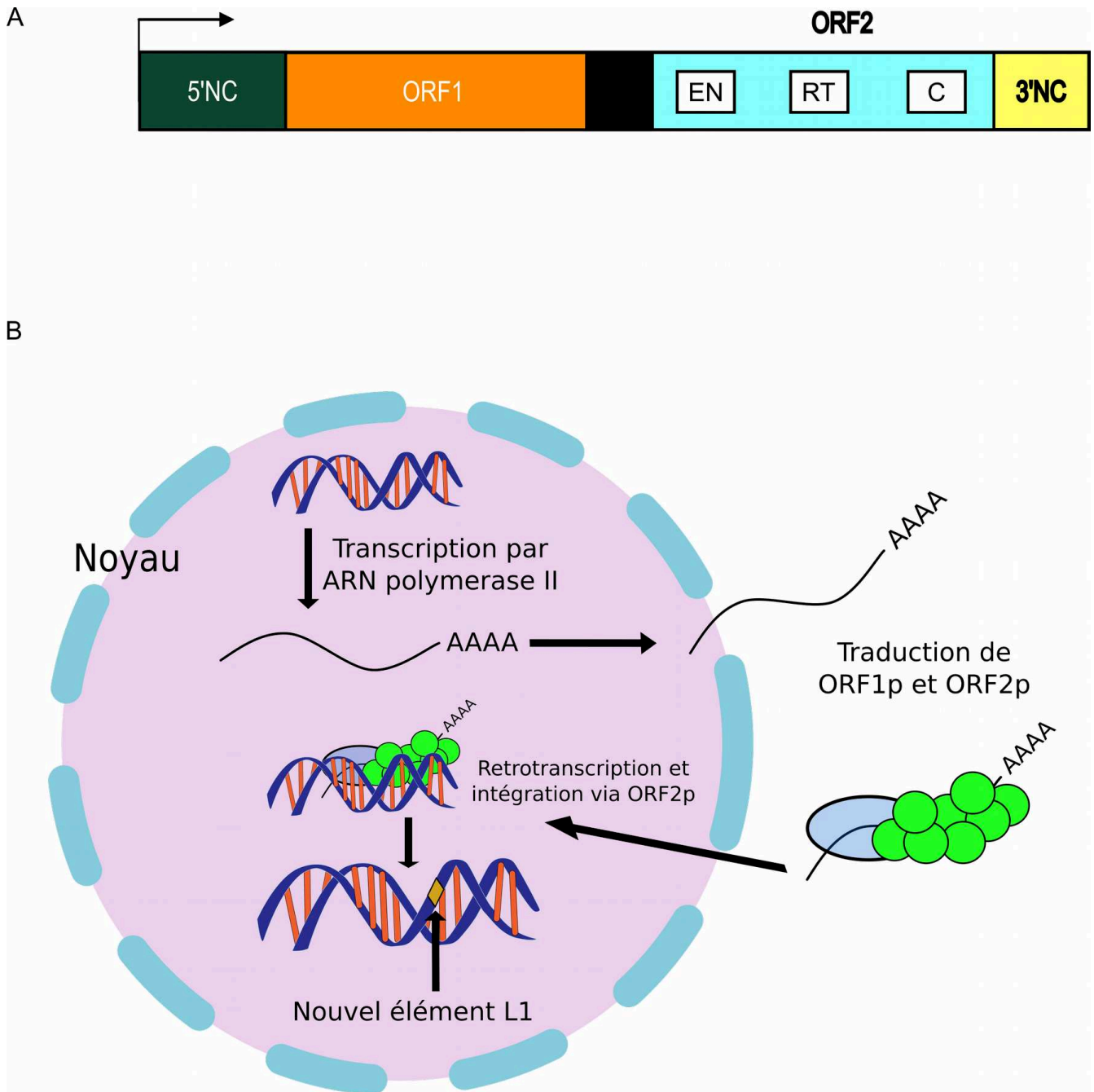
L'intégration de l'ADN du virus de l'hépatite B dans le génome eucaryote est un événement fréquent, détectable chez 75-90% des cas, bien que non essentiel au cycle viral. Les premiers gènes cellulaires affectés par l'intégration du VHB ont été découverts dans les années 90. Il s'agit des gènes codant la cycline A (Wang *et al.*, 1992) et le récepteur  $\beta$  de l'acide rétinoïque (Garcia *et al.*, 1993). Depuis ces travaux, de nombreux sites récurrents d'intégration ont été identifiés grâce notamment à l'essor des technologies de séquençage à haut débit. Une étude majeure publiée en 2012 a analysé le génome des tissus hépatiques provenant de 81 patients atteints de CHC et infectés par le VHB (Sung *et al.*, 2012). Des intégrations virales ont été identifiées chez 75 patients de cette cohorte. Plus précisément, les intégrations ont été retrouvées préférentiellement dans le tissu tumoral plutôt que non tumoral (91 % versus 33 %) suggérant un avantage de croissance conféré par les intégrations au cours de la tumorigenèse. Un seul site d'intégration sur les 399 sites séquencés était partagé entre la tumeur et la non tumeur d'un même patient. Le gène TERT est très fréquemment le siège d'une intégration, principalement au niveau du promoteur et dans des régions introniques du gène, confirmant les données d'une étude précédente (Paterlini-Bréchet *et al.*, 2003). Les autres gènes cibles sont des gènes impliqués dans la méthylation de l'ADN (MLL4, 10 sites d'intégration identifiés), le

cycle cellulaire (CCNE1, 4 sites) et la sumoylation (SENP5, 3 sites). L'intégration virale dans ces régions génomiques conduit à une surexpression des ARNm TERT, MLL4 et CCNE1. Enfin, les auteurs ont montré que ces intégrations sont associées à un âge jeune des patients cancéreux suggérant une possible accélération du développement tumoral par le biais de ce mécanisme et surtout l'association à un pronostic sombre par la présence d'au moins 3 intégrations virales dans le génome tumoral. Deux autres études ont depuis été rapportées, identifiant le set de cibles récurrentes d'intégration (Ding *et al.*, 2012; Fujimoto *et al.*, 2012). Seul le gène TERT est une cible commune d'intégration dans ces deux articles. De plus, Ding et ses collaborateurs n'ont pas retrouvé de différences d'intégration entre les parties tumorales et non tumorales analysées. Les différences observées entre ces études peuvent évidemment être la conséquence de différence dans les cohortes de patients étudiées, dans la méthodologie employée mais aussi des différences dans la notion de « tissu adjacent à la tumeur » pour le choix de la zone non-tumorale à séquencer.

Le phénomène de mutagenèse insertionnelle par intégration d'éléments génétiques dans l'ADN génomique de l'hôte peut également concerner des régions non-codantes de l'ADN, en particulier des séquences répétées comme les éléments transposables LINE-1 (Long interspersed nuclear elements ) et LTR (long terminal repeat sequence), ce qui a été d'ailleurs observé dans les trois articles cités précédemment. Le fait notable concernant l'intégration dans ces régions répétées fut la découverte d'un transcrite chimérique HBx-LINE1 dans la lignée cellulaire HKCl-4 (lignée dérivée d'un CHC-HBV positif). Ce transcrite est capable d'activer la voie Wnt/ $\beta$ -caténine *in vitro*, et de favoriser l'invasion cellulaire et la transition épithélio-mésenchymateuse (EMT). Il fut identifié dans des tumeurs de CHC chez 21 patients infectés par le VHB sur un total de 90 CHC analysés. Ce résultat fut par la suite confirmé dans une cohorte de patients cancéreux chinois (17 CHC exprimaient le transcrite HBx-LINE1 sur 40 cas analysés) (Liang *et al.*, 2016), mais non confirmé dans une cohorte de 50 patients caucasiens présentant un CHC (Cao *et al.*, 2015). D'autres études méritent d'être effectuées afin de clarifier cette différence.

### 6.1.2 Insertion d'éléments LINE-1

Le génome humain contient environ 500 000 copies de LINE-1 (L1), ce qui représente 17% de l'ADN humain. La majorité des L1 perd la capacité de rétrotransposition en raison de tronçures en 5', de réarrangements inversés ou de mutations ponctuelles survenant au cours de la transcription inverse ou de la réplication chromosomique ultérieure. On estime que le génome



**Figure 6: Structure du rétrotransposon L1 et rétrotransposition (A).** L1 est un élément de 6kb qui possède un domaine 5' non codant riche en CpG avec le promoteur pour l'ARN polymérase II, deux cadres de lectures ORF1 et ORF2 (ORF2 codant pour une endonucléase EN, une transcriptase inverse RT) une région C riche en cystéine et une séquence polyadénylée dans la région 3' non codante. (B). L'élément L1 est d'abord transcrit dans le noyau par l'ARN polymérase II qui est ensuite traduite dans le cytoplasme. Les deux protéines ORF1p (ronds verts) et ORF2p (ovale bleu) s'assemblent avec l'ARN L1 et transloquent dans le noyau. La transcription inverse de l'ARN est couplée à l'intégration et est réalisée directement sur le brin d'ADN coupé.

humain moyen contient environ 50 à 120 L1 actifs, avec un sous-ensemble hautement actif (~ 5 à 10% d'éléments actifs) appelé « hot L1 » constituant la majorité de cette activité. Ces L1 actifs utilisent un mécanisme de 'copier-coller' pour s'insérer et se propager dans tout le génome, avec des effets potentiellement perturbateurs sur les gènes voisins ou les séquences régulatrices. De cette manière, les L1 actifs continuent de remodeler le génome humain et deviennent une source de mutagenèse endogène qui provoque la variation du génome individuel et peut participer à la pathogenèse de nombreuses maladies génétiques (cf. Figure 6). Les L1 sont actifs dans la lignée germinale et au cours de l'embryogenèse, et sont inhibés dans les cellules somatiques par voie épigénétique. Dans les cellules cancéreuses, les L1 sont activés de manière aberrante et peuvent jouer un rôle dans l'instabilité du génome, une des caractéristiques de la pathogenèse du cancer. Leurs états de méthylation et leurs activités de rétrotransposition sont associés et fluctuent pendant l'initiation et la progression du cancer. De plus, les protéines ORF1p et ORF2p codées par les L1 peuvent avoir un potentiel pro-oncogène, par exemple en activant des facteurs de transcription oncogéniques ou en séquestrant des oncosuppresseurs (Burns, 2017).

#### **6.1.2.1      *Activation de la voie Wnt/béta-caténine : insertion dans le gène MCC.***

Durant ma thèse, j'ai participé à une étude visant à établir que la mobilisation de L1 dans les cellules germinales et somatiques est un facteur étiologique important dans le CHC (Shukla *et al.*, 2013).

Trois insertions germinales d'éléments L1 ont été identifiées dans les tumeurs CHC viro-induites de 4 patients sur 19 CHC analysés par séquençage haut débit ; celles-ci altéraient fortement l'expression du gène MCC (Mutated in Colorectal Cancer) connu pour réguler de multiples processus cellulaires dans les cellules épithéliales, tels que la prolifération, la réponse aux dommages de l'ADN, la formation de lamellipodes et la migration cellulaire. Le gène MCC peut également réprimer les voies de signalisation Wnt et NFkB (Bouwmeester *et al.*, 2004; Fukuyama *et al.*, 2008; Sigglekow *et al.*, 2012). De nombreuses preuves en particulier dans le colon suggèrent que le gène MCC est un puissant suppresseur de tumeur dans les tissus murins et humains et que le gène MCC réprime la voie  $\beta$ -caténine (Arnaud *et al.*, 2009). La surexpression du gène MCC est associée dans notre étude à une inhibition de la prolifération cellulaire. Une analyse des produits du gène MCC montre une diminution importante de l'expression de la protéine MCC dans les tumeurs de CHC et dans les régions non-tumorales adjacentes chez ces 4 patients comparée à celle observée dans des foies normaux. Cette

inhibition de MCC est associée à une augmentation de l'expression de  $\beta$ -caténine en accord avec les données de la littérature. Bien que le rôle pathogénique de l'insertion germinale d'un élément L1 dans le gène MCC n'a pas été formellement démontré dans notre étude, nos résultats suggèrent l'existence d'une prédisposition génétique au CHC chez ces 4 patients en lien avec une insertion d'éléments L1 et un rôle important de la protéine MCC dans le cancer du foie.

#### **6.1.2.2 Hypométhylation du promoteur L1 dans les CHC.**

Comme décrit précédemment, la plupart des rétrotransposons L1 sont rendus inactifs par la méthylation de leur promoteur. Un séquençage des îlots CpG dans le foie tumoral de 8 patients a montré une hypométhylation du promoteur L1 dans les cellules tumorales comparée à celle de cellules non tumorales adjacentes et de cellules normales. L'expression des transcrits ORF2 et de la région 5' non codante était augmentée dans les parties tumorales et non-tumorales par rapport au foie normal contrôle montrant une forte activation et une transcription de L1. D'autres études citant notre travail ont utilisé l'hypométhylation des éléments L1 comme marqueur pronostic dans le CHC (Gao *et al.*, 2014, Zhu *et al.*, 2014b; Harada *et al.*, 2015). Dans ces travaux, l'hypométhylation des éléments L1 est associée à un pronostic plus sombre des patients indépendamment de tout critère clinico-pathologique, à l'exception du sexe.

#### **6.1.2.3 Activation spécifique du facteur de transcription ST18.**

Dans notre étude, deux autres insertions somatiques dans le gène ST18 ont également été retrouvées dans l'ADN tumoral d'un patient atteint de CHC conduisant à une augmentation d'expression de ST18, un répresseur de transcription à doigt de zinc. Des analyses par RT-PCR et par immunohistochimie ont montré une forte expression de ST18 dans la tumeur comparée à la non-tumeur. Cette forte expression a été retrouvée dans des lignées cellulaires transformées dérivées de CHC humain ainsi que dans des nodules de CHC développés dans le modèle de souris invalidées pour le gène *Mdr (multidrug resistance) 2*. Le rôle de la protéine ST18 dans la carcinogénèse reste controversée ; des études montrent un effet suppresseur de tumeur quand d'autres soulignent le caractère pro-oncogène de ce facteur de transcription suivant le type de cancer étudié. Nos données expérimentales ainsi que l'absence d'expression de la protéine ST18 dans le foie normal suggèrent un rôle oncogénique de la protéine ST18 au moins dans la tumeur de CHC analysée. Au total, notre travail a permis d'identifier un nouveau mécanisme de carcinogénèse hépatique par mutagenèse insertionnelle germinale et somatique de rétrotransposons de type LINE-1. Bien que ce mécanisme soit activé de manière hétérogène

avec des insertions identifiées dans des sous-groupes de patients, les L1 représentent des éléments génétiques importants de mutagenèse insertionnelle (comme le virus de l'hépatite B) pouvant altérer des voies oncogéniques ou suppressives de tumeur et conduire à l'initiation ou la progression tumorale du CHC.

## 6.2 Action directe des protéines virales

### 6.2.1 HBx

La protéine HBx présente de nombreuses activités biologiques pouvant affecter la transcription de nombreux gènes cellulaires, la transduction du signal intracellulaire, la prolifération cellulaire, l'apoptose, la réparation des dommages à l'ADN, l'inflammation et la réponse immunitaire. Elle est capable de stimuler la réplication du VHB en augmentant l'activité de la polymérase virale via notamment la régulation du calcium intracellulaire (Bouchard *et al.*, 2001). Elle stimule la phosphorylation de la protéine core, indispensable à la réplication virale (Melegari *et al.*, 2005). HBx peut également induire des voies d'apoptose et de production d'espèces réactives de l'oxygène (ROS) (Chaturvedi *et al.*, 2019). Enfin, HBx est capable d'interagir avec de nombreuses voies de signalisation qui sont reconnues comme étant majeures dans la pathogenèse du CHC comme les voies Wnt/ $\beta$ -caténine (Cho *et al.*, 2012) et TGF- $\beta$  (Cheng *et al.*, 2018).

### 6.2.2 Capside du VHC

La carcinogenèse liée à l'infection par le VHC est complexe, résultant de la combinaison de facteurs environnementaux, de facteurs liés à l'hôte et de l'action des protéines virales. Le virus a la capacité d'induire directement la transformation cellulaire et la tumorigenèse en altérant différentes voies de signalisation oncogéniques. Notre laboratoire, a par exemple, étudié de manière extensive le rôle de la protéine de capsid du VHC dans la carcinogenèse hépatique et montré que celle-ci interagissait avec la voie du TGF- $\beta$  favorisant ainsi la fonction de promotion tumorale dépendante de cette voie. De plus, l'analyse de différents variants de protéine de capsid isolés de tumeurs de CHC humain a montré que seules certaines séquences de capsid étaient capables d'induire une transition épithélio-mésenchymateuse (EMT) qui contribuait à l'invasion tumorale et au processus métastatique (Pavio *et al.*, 2005; Battaglia *et al.*, 2009).



### 6.3 Maladies métaboliques

L'épidémie d'obésité et de diabète de type 2 est aujourd'hui la cause principale de stéatose hépatique non liée à l'alcool (NAFLD, Non Alcoholic Fatty Liver Disease). L'augmentation importante de l'incidence de ces pathologies qui concerne plus d'un milliard de personnes dans le monde (principalement dans les pays développés) fait de celles-ci, la première cause de maladies chroniques du foie et un problème de santé publique majeur (Browning *et al.*, 2004; Ramachandran and Snehalatha, 2010; NCD Risk Factor Collaboration (NCD-RisC), 2016). La stéatose et la résistance à l'insuline favorisent l'inflammation dans les organes périphériques métaboliquement actifs comme le tissu adipeux, entraînent d'importants changements hormonaux, des changements du statut oxydatif et une lipotoxicité (Baffy *et al.*, 2012). En 2010, une équipe américaine a montré que l'obésité, dans un modèle murin génétique ou après administration d'un régime alimentaire riche en lipides, favorise le développement du cancer du foie *via* la contribution des signalisations IL-6 et TNF $\alpha$  (Park *et al.*, 2010).

### 6.4 Microbiote intestinal

Le foie est constamment exposé à des produits microbiens provenant des flux mésentérique et portal. Ces produits, comme les endotoxines et en particulier les lipopolysaccharides (LPS), activent les voies pro-inflammatoires et oxydatives et peuvent contribuer au développement d'un CHC (Yu *et al.*, 2010). Une équipe française a retrouvé la présence d'ADN bactérien d'*Helicobacter pylori*, bactérie responsable de certains cancers de l'estomac, dans les foies de patients atteints de cirrhose (Rocha *et al.*, 2005). Des études expérimentales chez le rongeur ont montré l'implication de composants moléculaires du microbiote intestinal dysbiotique dans la promotion et la progression tumorale hépatique. Chez des rongeurs sauvages ou axéniques (dépourvus de microbiote intestinal) traités par des carcinogènes chimiques, la progression du CHC est largement freinée par l'application d'une antibiothérapie ou par le statut axénique lui-même *via* une réduction de l'activation de la voie LPS/TLR4 (Dapito *et al.*, 2012; Zhang *et al.*, 2012).

Comme nous l'avons mentionné précédemment, l'obésité et les régimes alimentaires gras sont des facteurs de risque importants du CHC. Ils provoquent des altérations de composition, et donc de fonctionnalité du microbiote intestinal (Ley *et al.*, 2005, 2006; Turnbaugh *et al.*, 2006). Il a été montré que l'administration d'antibiotiques chez des souris obèses retarde l'apparition de nodules tumoraux suggérant fortement l'implication des microbiomes dysbiotiques dans cette carcinogenèse. De plus, l'acide déoxycholique (DCA), métabolite issu de la

transformation des acides biliaires primaires en secondaires par des genres bactériens de type *Clostridium* cluster XI et XIVa, et connu pour son activité génotoxique, est particulièrement surreprésenté dans des souris obèses exposées à des carcinogènes chimiques. Le blocage de la production de DCA ou la réduction des bactéries productrices de DCA empêchent le développement du CHC chez des souris obèses (Yoshimoto *et al.*, 2013).

L'utilisation de prébiotiques et de probiotiques qu'il s'agit d'identifier et qui viseraient à moduler les flores intestinales dysbiotiques représente un axe important de recherche pour la prévention et le traitement du CHC.

## 7. Effecteurs moléculaires

### 7.1 Anomalies du génome codant

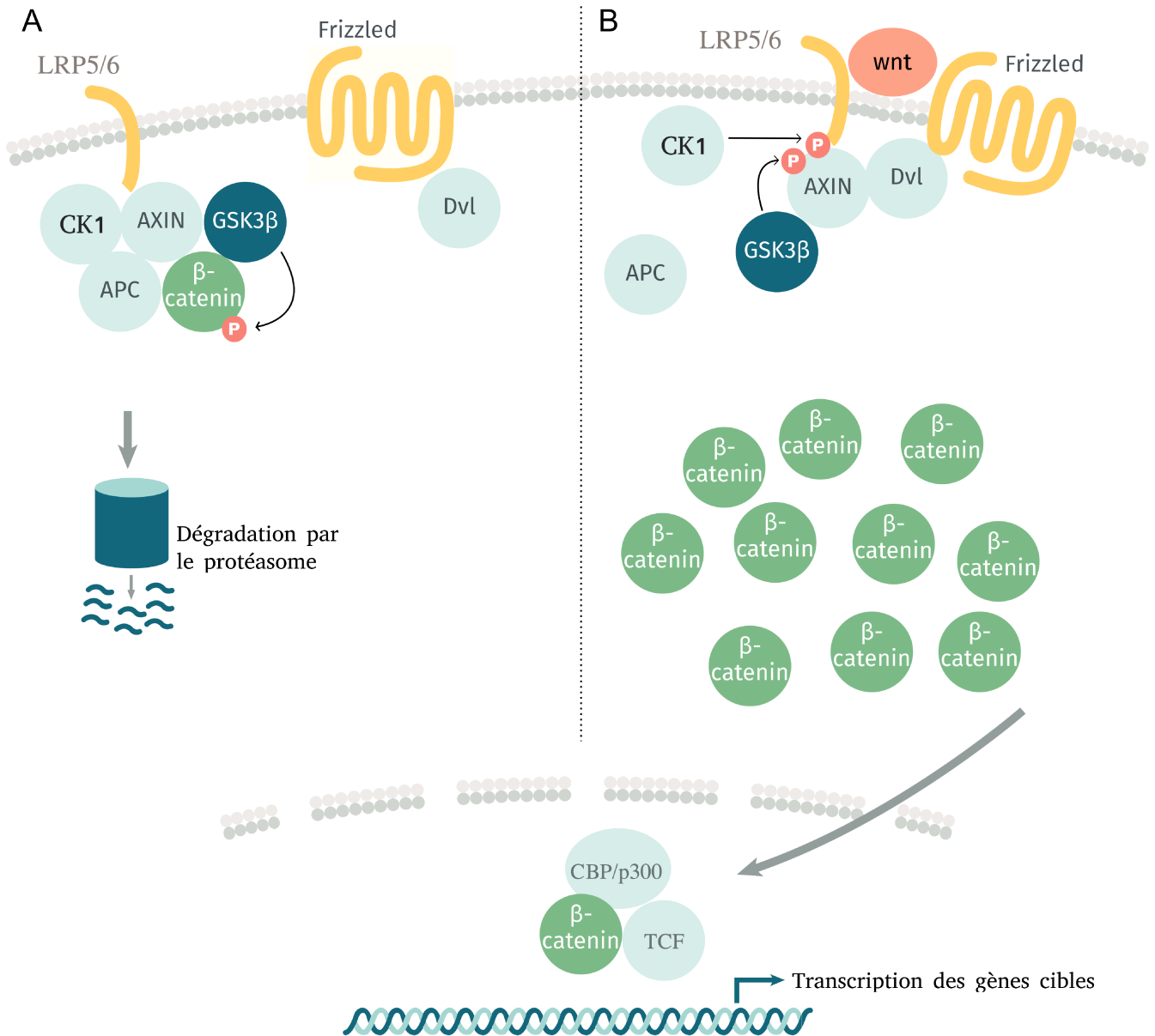
Les progrès des technologies de ces 20 dernières années ont permis des avancées spectaculaires dans l'identification et l'étude des fonctionnalités des mutations somatiques observées dans le CHC. Ces données récentes permettent de mieux comprendre le développement tumoral et sa progression.

#### 7.1.1 Gène p53

Le gène TP53 est un gène suppresseur de tumeur fréquemment muté dans les cancers solides. Nous avons noté précédemment que l'intoxication par l'aflatoxine B1 entraîne une mutation spécifique dans le codon 249 (R249S) de TP53. De nombreuses mutations ont été identifiées dans le gène TP53 mais aucune mutation récurrente n'a pu être observée (Hsu *et al.*, 1991; Guichard *et al.*, 2012; Fujimoto *et al.*, 2015; Schulze *et al.*, 2015). À la différence du gène TERT, les mutations de TP53 sont majoritairement observées dans les stades avancés de CHC (Schulze *et al.*, 2015).

#### 7.1.2 Gène CTNNB1

La voie de signalisation WNT- $\beta$ -caténine joue un rôle majeur dans l'embryogenèse, la zonation du foie ou encore le métabolisme hépatique. Des mutations activatrices du gène CTNNB1 codant la  $\beta$ -caténine ont été retrouvées dans 26 % (de La Coste *et al.*, 1998) ou 37 % des cas (Schulze *et al.*, 2015) étudiés. Les mutations dans le gène CTNNB1 ont pour conséquence une accumulation et une translocation de la protéine  $\beta$ -caténine dans le noyau où elle s'associe aux facteurs de transcription de la famille LEF/TCF pour stimuler l'activité transcriptionnelle de



**Figure 7: Voie Wnt/ $\beta$ caténine**

En l'absence du ligand Wnt, la  $\beta$ -caténine se situe dans un complexe de dégradation comprenant les protéines Axin et APC et les kinases CK1 et GSK3. Ce complexe phosphoryle la  $\beta$ -caténine qui est ensuite ubiquitinylée et dégradée. En présence du ligand Wnt, les récepteurs membranaires Frizzled et LRP5/6 sont activés et phosphorylés par le recrutement de Dvl. Après recrutement de l'Axin, le complexe de dégradation de la  $\beta$ -caténine est rompu, conduisant à l'accumulation de la  $\beta$ -caténine et sa translocation dans le noyau où elle agit comme facteur de transcription dans un complexe formé par TCF et CBP/p300.

gènes cibles impliqués notamment dans la prolifération cellulaire, la migration et l'échappement métastatique. D'autres mutations seraient capables d'activer la voie WNT- $\beta$ -caténine comme les mutations inactivatrices des gènes AXIN1 (11 %), AXIN2 (1 %) et APC (1 %). De manière intéressante les tumeurs mutées pour AXIN1 n'ont pas le même profil transcriptionnel que les tumeurs mutées CTNNB1 ou APC (Laurent-Puig *et al.*, 2001; Hoshida *et al.*, 2009). Des équipes françaises ont montré qu'aucun gène cible de la  $\beta$ -caténine n'était exprimé dans des modèles cellulaires mutés pour AXIN1 suggérant un rôle de ce gène indépendant de cette voie (Zucman-Rossi *et al.*, 2007).

La voie WNT- $\beta$ -caténine est une voie complexe impliquée dans le développement du foie embryonnaire et dans l'homéostasie hépatique adulte contrôlant divers mécanismes clés comme la prolifération cellulaire, la différenciation et la polarité cellulaires.

En absence de ligand de la famille WNT, la  $\beta$ -caténine forme un complexe avec les protéines APC, Axin, CK1 et GSK-3 $\beta$  qui phosphoryle la  $\beta$ -caténine. Cette dernière est ubiquitinylée et dégradée par le protéasome. Lorsqu'un ligand de la famille WNT se fixe à son récepteur membranaire (Frizzled), le complexe ligand-récepteur induit une cascade d'événements moléculaires qui conduit à l'accumulation et la translocation nucléaire de  $\beta$ -caténine où elle jouera un rôle de co-facteur de transcription en présence de TCF/LEF (cf. Figure 7).

Une première étude a montré que l'activation aberrante de la voie WNT- $\beta$ -caténine par le biais d'une inactivation du gène APC conduit au développement d'un CHC chez la souris (Colnot *et al.*, 2004). L'inhibition de  $\beta$ -caténine est associée à une augmentation de l'apoptose et une réduction de la prolifération dans des modèles cellulaires (Behari *et al.*, 2007; Monga, 2015). Dans des modèles animaux, il a été montré que l'infiltration macrophagique est responsable d'une activation forte de la voie WNT- $\beta$ -caténine par la production de ligands de la famille WNT (Debebe *et al.*, 2017) et que l'inhibition pharmacologique de ces ligands réduit la prolifération cellulaire (Wang *et al.*, 2016). Enfin, l'expression de marqueurs de dédifférenciation cellulaire comme EPCAM peut être activée par la voie WNT- $\beta$ -caténine (Yamashita *et al.*, 2007). L'ensemble de ces données démontre l'importance de cette voie dans la carcinogenèse hépatique.

### 7.1.3 PI3-K

Les phosphatidylinositol 3-kinases (PI3K) appartiennent à une famille conservée de lipides kinases intracellulaires qui phosphorylent le groupe 3'-hydroxyle des phosphatidylinositols.

Cette réaction conduit à l'activation de nombreuses voies de signalisation qui régulent des fonctions aussi diverses que le métabolisme cellulaire, la survie des cellules, la polarité cellulaire ainsi que le trafic vésiculaire. La voie PI3-K-AKT-mTOR est fréquemment dérégulée dans les cancers solides et notamment dans le CHC. mTOR contrôle de nombreux processus cellulaires incluant la régulation de la biosynthèse des protéines. Lorsque celle-ci est activée de manière aberrante, cela conduit à une prolifération cellulaire anormale (Mamane *et al.*, 2006; Petroulakis *et al.*, 2006). Même si des mutations ont été retrouvés dans les gènes de cette voie (PI3KCA : 3 % et PTEN : 1 % des tumeurs), il semblerait que cela n'affecte pas la voie PI3-K-AKT-mTOR (Villanueva *et al.*, 2008).

Un taux élevé de protéine AKT phosphorylée, retrouvé dans un CHC sur cinq analysés (Boyault *et al.*, 2007), est caractéristique des classes moléculaires G2 (Boyault *et al.*, 2007) et S2 (Hoshida *et al.*, 2009) et est corrélé à une récurrence précoce des CHC traités. Phospho-mTOR et ses cibles, les S6 kinases, sont fréquemment dérégulés dans les CHC (Sahin *et al.*, 2004). De plus, l'expression du régulateur négatif PTEN est souvent réduite dans le CHC. L'inactivation du gène PTEN dans les hépatocytes conduit au développement d'un CHC dans un modèle animal (Horie *et al.*, 2004) et son expression dans les cancers humains est fortement corrélée à un mauvais pronostic (Hu *et al.*, 2003).

Comme l'indique ces données moléculaires évidemment non exhaustives, la voie PI3-K-AKT-mTOR est fortement impliquée dans la carcinogenèse hépatique et fait l'objet de recherche de composés pharmacologiques pour le traitement du CHC. Des études *in vitro* et *in vivo* évaluant le sirolimus et l'everolimus (inhibiteurs mTOR) ont produit des résultats encourageants d'inhibition de la croissance tumorale (Semela *et al.*, 2007; Pigué *et al.*, 2008; Villanueva *et al.*, 2008). Actuellement des essais cliniques sont en développement, évaluant l'efficacité de ces inhibiteurs en combinaison avec des drogues de chimiothérapie conventionnelle. Malheureusement la plupart des essais déjà réalisés n'a pas montré d'effet positif sur la survie des patients (Kneteman *et al.*, 2004; Zhu *et al.*, 2011, 2014a), ce qui est probablement dû à un rétro-contrôle déclenché par mTORC2. L'inhibition des deux complexes mTORC1 et mTORC2 a donné des résultats prometteurs *in vivo* (Xie *et al.*, 2017), et est actuellement testée dans le cadre d'études cliniques (essai NCT03591965; clinicaltrials.gov)

#### 7.1.4 Stress oxydatif

NRF2 (Nuclear factor erythroid 2-related factor 2) codé par le gène NFE2L2 est un facteur de transcription qui induit l'expression de nombreux gènes de réponse au stress oxydatif. KEAP1

est une molécule inhibitrice de NRF2 entraînant sa dégradation par le protéasome. Il a été montré que l'activation de NRF2 dans des cellules murines cancéreuses protège ces dernières de la toxicité vis-à-vis des espèces réactives de l'oxygène (ROS) et de la mort cellulaire (DeNicola *et al.*, 2011). Dans le CHC, des mutations activatrices du gène NFE2L2 (3 à 6 %) et des mutations inactivatrices du gène KEAP1 (2 à 8 %) ont été identifiées dans différentes études (Guichard *et al.*, 2012; Cleary *et al.*, 2013).

### 7.1.5 Remodelage de la chromatine

Le remodelage de la chromatine est un mécanisme important de contrôle du cycle cellulaire, de la différenciation cellulaire et de la réparation des dommages à l'ADN. Les gènes impliqués dans ce processus sont fréquemment mutés dans les cancers et représentent la troisième famille de gènes les plus fréquemment mutés dans le CHC (89). Des mutations dans les gènes régulateurs de l'épigénétique ARID1A, ARID1B et ARID2 (ARID1A : 16 %, ARID2 : 6%) ont été détectées dans des CHC associés à l'alcool et au VHC (Li *et al.*, 2011a; Guichard *et al.*, 2012; Nault *et al.*, 2019). Dans l'étude sino-américaine de Li et collaborateurs, portant sur une centaine de patients, les auteurs ont montré que les mutations de ARID2 étaient corrélées aux mutations de la bêta-caténine. Le rôle des mutations de ARID2 semble être lié à une réduction de la voie Jak/STAT conduisant à la persistance de l'infection par le VHC. D'autres gènes impliqués dans le remodelage de la chromatine sont également mutés dans les CHC comme le gène CHD1L qui favorise la progression tumorale (Liu *et al.*, 2016).

### 7.1.6 Modifications des histones

Les histones sont la cible d'un grand nombre de modifications post-traductionnelles, telles que l'acétylation, la méthylation, la phosphorylation et l'ubiquitinylation, constituant un code très élaboré dont l'impact sur les fonctions nucléaires comme la régulation transcriptionnelle est majeur. De nombreuses études indiquent que la méthylation des histones est fortement liée à la dérégulation de gènes associés à l'invasion tumorale et à la prolifération cellulaire. Par exemple, la méthylation de l'histone H3 sur la lysine en position 4 et 36 (K4, K36) provoquent des changements conformationnels de la chromatine entraînant l'expression de gènes impliqués dans la transition épithélio-mésenchymateuse (Jenuwein, 2001).

#### 7.1.6.1 La méthylation des histones

EZH2 (enhancer of zeste homologue 2), une protéine du complexe épigénétique Polycomb, est responsable de la mono-, di- ou triméthylation de la lysine située en position 27 de l'histone H3

(H3K27me1, 2 ou 3). EZH2 est une protéine clé du complexe PRC2 (polycomb repressive complex 2) : elle porte l'activité catalytique du complexe et joue un rôle majeur dans la répression transcriptionnelle. La protéine EZH2 est surexprimée dans près de 50 % des CHC. Son expression est associée à une prolifération cellulaire tumorale accrue et à un mauvais pronostic des patients. L'inactivation ou l'inhibition de EZH2 dans des lignées cellulaires dérivées de CHC ralentit la prolifération cellulaire et la formation de tumeurs dans des modèles de souris xéno greffées (Chiba *et al.*, 2012). La phosphorylation de la protéine EZH2 a également une influence sur les modifications des histones en rompant les interactions de EZH2 avec le complexe Polycomb, le rendant ainsi inactif et réduisant, par voie de conséquence, le degré de méthylation de H3K27 et la croissance tumorale (Wan *et al.*, 2018).

La méthylation de la lysine 9 de l'histone H3 (H3K9) est catalysée par deux enzymes principales, les lysines méthyl-transférases KMT1A et SETDB1, contribuant à la répression de gènes de l'euchromatine. L'expression de ces deux enzymes est retrouvée plus importante dans la partie tumorale du CHC que non tumoral. La surexpression de KMT1A est corrélée à une récurrence tumorale plus fréquente et son inhibition réduit la croissance tumorale (Chiba *et al.*, 2015). De plus, il a été retrouvé des mutations inactivatrices de KMT2C (5%) dans les CHC. Son inactivation dans des lignées cellulaires conduit à une activation de la prolifération cellulaire (Greenman *et al.*, 2007).

#### 7.1.6.2 L'acétylation des histones

L'acétylation des histones est régulée par deux groupes d'enzymes aux fonctions opposées : les histones acétyltransférases (HAT) et les histones désacétylases (HDAC) qui ajoutent ou enlèvent un groupement acétyle sur les lysines conservées des queues amino-terminales des histones modifiant ainsi la charge portée par la lysine, l'accessibilité de l'ADN et la transcription (Roth *et al.*, 2001; Ropero and Esteller, 2007).

L'expression des HDAC est fortement dérégulée dans les CHC (Haberland *et al.*, 2009). HDAC3 dont le rôle dans la prolifération tumorale et l'invasion a été montré *in vitro*, est surexprimée dans les CHC et est associée à la récurrence tumorale (Wu *et al.*, 2010).

HDAC1 et 2 sont également associées à la prolifération tumorale et au risque de récurrence tumorale. L'inhibition de HDAC1 et 2 empêche la répression de l'enzyme fructose 1,6 biphosphatase et inhibe la croissance tumorale *in vitro* dans des lignées cellulaires et *in vivo* dans des modèles de xéno greffes (Yang *et al.*, 2017).

De plus, HDAC2 est impliquée dans la modulation de gènes de l'apoptose, du cycle cellulaire et du métabolisme lipidique dans une étude transcriptomique réalisée sur des lignées d'hépatocarcinome (Lee *et al.*, 2014).

Tout comme la méthylation, l'acétylation est une cible thérapeutique majeure et a fait l'objet de développement de candidats médicaments visant à inhiber les HDAC. Le Vorinostat est un inhibiteur de HDAC utilisé dans le traitement des lymphomes cutanés à cellules T (Duvic *et al.*, 2007). Dans des modèles *in vitro*, le Vorinostat a montré des effets notables anti-cancéreux en combinaison avec le sorafenib ou l'oxaliplatine (Hsu *et al.*, 2014; Liao *et al.*, 2017). Des études cliniques de phase I/II d'évaluation du Vorinostat dans le CHC sont en cours (clinicaltrials.gov – juin 2019).

### 7.1.7 Méthylation de l'ADN

La méthylation de l'ADN se produit principalement sur les résidus cytosines de dinucléotides palindromiques 5'-CG-3', appelés dinucléotides CpG, au niveau des promoteurs et à proximité du site d'initiation de la transcription. Les enzymes responsables de l'addition covalente d'un groupement méthyl sur une cytosine sont les méthyltransférases de l'ADN (DNMTs). La méthylation de l'ADN est un mécanisme fondamental de régulation épigénétique de l'expression des gènes. Elle joue un rôle primordial dans le développement, la formation et la maintenance de la chromatine ainsi que d'autres processus fondamentaux. Les enzymes DNMTs sont fréquemment surexprimées dans de nombreux cancers (Subramaniam *et al.*, 2014).

Des gènes suppresseurs de tumeurs, comme la E-Cadherine, p16 ou encore SOCS1 sont souvent réprimés par la méthylation de l'ADN (Wei *et al.*, 2002; Okochi *et al.*, 2003). L'hyperméthylation des îlots CpG observée dans le CHC est associée à une instabilité chromosomique, une forte activité télomérase et un mauvais pronostic clinique (Zhang *et al.*, 2007a, 2008; Wang *et al.*, 2018).

Par ailleurs, les mutations somatiques retrouvées dans le foie non tumoral sont préférentiellement situées dans des régions hautement méthylées dans une importante étude réalisée sur 373 échantillons, suggérant un rôle majeur du contexte chromatinien dans la régulation de la fréquence des mutations somatiques (Hama *et al.*, 2018).

Des candidats inhibiteurs de la méthylation de l'ADN sont développés depuis plusieurs années pour aborder une nouvelle approche de thérapie du CHC. L'azacytidine (5-AZA), analogue de la cytosine et inhibiteur des DNMT, stimule la différenciation des cellules tumorales en



convertissant des cellules résistantes au Sorafenib en cellules très bien différenciées dans un modèle murin de xenogreffes sous-cutanées de cellules dérivées de CHC (Gailhouste *et al.*, 2018). Récemment, une étude de phase I/II chez des patients réfractaires aux traitements classiques a montré une réponse positive en termes de survie des patients traités par de faibles doses de décitabine (5-aza-2'-deoxycytidine) qui présente une fonction similaire à celle du 5-AZA (Fan *et al.*, 2014).

En raison de leur demi-vie courte, du fait de l'inactivation métabolique des cytidines déaminases (Sato *et al.*, 2017), des molécules de seconde génération sont en cours de développement. Il faut noter que de nombreuses études pré-cliniques ont montré une certaine efficacité de ces molécules (comme la guadecitabine) dans des modèles *in vitro* et *in vivo* de xénogreffes (Kuang *et al.*, 2015; Jueliger *et al.*, 2016).

## 7.2 Anomalies du génome non-codant

Les projets de séquençage à grande échelle comme le TCGA (The Cancer Genome Atlas) et l'ICGC (International Cancer Genome Consortium) se sont principalement concentrés dans les variations génomiques existants dans les séquences codantes pour des protéines dans les génomes tumoraux. Ces projets (comme le projet ENCODE (Derrien *et al.*, 2012; ENCODE Project Consortium, 2012)) ont grandement facilité la compréhension et l'identification de nouvelles altérations dans les régions codantes. Cependant, seulement 2 % du génome humain est constitué de régions codantes pour des protéines. Pendant très longtemps, les 98 % restants « non codants » ont été très peu étudiés. Aujourd'hui, de nombreuses études ont montré que des variations, mutations dans cet ADN « poubelle » pouvaient avoir des répercussions importantes dans des maladies comme le cancer.

### 7.2.1 Promoteur de TERT

De nombreuses études ont montré l'importance du système télomères/téломérase dans la pathogenèse des maladies hépatiques et l'initiation tumorale. La sénescence répllicative est un processus physiologique qui contrôle la capacité proliférative des hépatocytes par l'intermédiaire du raccourcissement des télomères (Paradis *et al.*, 2001). Les télomères sont constitués de régions TTAGGG hautement répétitives et de protéines associées situées à l'extrémité des chromosomes assurant leur protection vis-à-vis d'éventuelles recombinaisons non homologues avec fusions chromosomiques et fusions de chromatides sœurs. Au cours de la vie cellulaire, les télomères raccourcissent et conduisent à un arrêt de prolifération, une

sénescence et une apoptose. La télomérase est un complexe enzymatique composé d'une sous-unité catalytique protéique portant l'activité transcriptase inverse (hTERT) et un ARN matriciel (hTR, encore appelé TERC). La télomérase maintient la longueur des télomères en synthétisant et en ajoutant des fragments d'ADN au chromosome. L'expression de cet enzyme est réprimée dans la plupart des cellules somatiques.

La réactivation de la télomérase dans 90 % des cas de CHC est principalement due (i) à des mutations du promoteur du gène TERT (60 % des cas) (Nault *et al.*, 2013*b*), (ii) à une amplification du gène TERT (6 % des cas) (Totoki *et al.*, 2014) et (iii) à une intégration du virus de l'hépatite B dans le promoteur du gène TERT (10-15 % des cas) (Paterlini-Bréchet *et al.*, 2003; Sung *et al.*, 2012). Le taux de mutations du promoteur du gène TERT est faible dans les lésions pré-néoplasiques et augmente fortement dans le CHC suggérant que cette mutation est un événement précoce et récurrent dans la carcinogenèse hépatique (Schulze *et al.*, 2015).

Récemment, un essai clinique visant à inhiber l'activité télomérase par l'Imetelstat, un oligonucléotide 13-mer conjugué à un lipide, a été réalisé montrant une certaine efficacité chez certains patients avec cancers solides (poumons, sein) et hémopathies malignes (myélome multiple, leucémie lymphoïde chronique) ainsi que des enfants atteints de tumeur au cerveau (Salloum *et al.*, 2016).

## 7.2.2 ARNs non codants

### 7.2.2.1 Micro-ARNs

Les micro-ARNs (miARNs) sont des régulateurs traductionnels qui, après appariement à un ARN messenger cible, conduisent à la répression traductionnelle ou à la dégradation de cet ARNm. Ils interviennent dans de multiples processus biologiques comme la différenciation cellulaire, la maintenance des cellules souches ou la transition épithélio-mésenchymateuse. De nombreuses études ont décrit le profil d'expression des miARNs et leur rôle dans la régulation d'expression de gènes dans les cancers et notamment dans les CHC.

miR-122 est l'un des plus importants micro-ARN exprimé dans le foie. Il joue un rôle majeur dans l'homéostasie et les fonctions normales du foie (Li *et al.*, 2011*b*; Szabo and Bala, 2013). miR-122 est spécifique du foie et de nombreux facteurs de transcriptions hépato-spécifiques, comme HNF4 $\alpha$  ou HNF1 $\alpha$ , se lient au promoteur de miR-122 et activent son expression (Li *et al.*, 2011*b*; Hu *et al.*, 2012). miR-122 est également associé à la résistance anti-tumorale des drogues de chimiothérapie, en particulier le sorafenib. Des lignées cellulaires résistantes aux

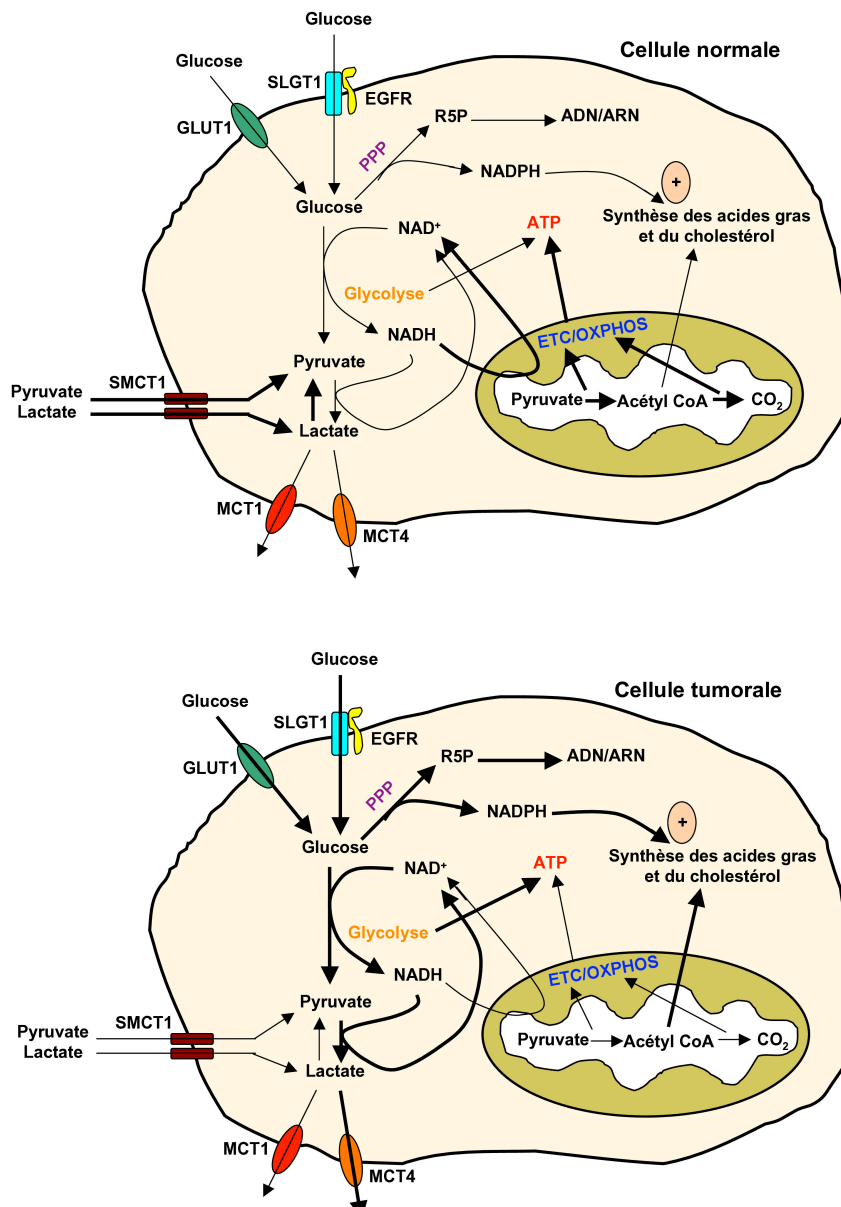
thérapies ont une expression moins importante de miR-122. La réactivation de miR-122 confère aux cellules tumorales une plus grande sensibilité au sorafenib (Xu *et al.*, 2016).

D'autres micro-ARNs sont impliqués dans les fonctions hépatiques et dans la carcinogenèse. On peut citer miR-19a et miR-223 qui sont liés à l'infection par le VHB et ont pour cibles MYC et PTEN (Yu *et al.*, 2016), ou encore miR-222 qui est impliqué dans la fibrogenèse hépatique (Hayes and Chayama, 2016).

De nombreuses analyses globales d'expression des micro-ARNs ont permis de stratifier les tumeurs de CHC. Par exemple, une équipe espagnole a analysé l'expression de 358 miARNs dans 89 CHC chez des patients infectés par le virus de l'hépatite C (Toffanin *et al.*, 2011). Trois sous-classes sensiblement égales en nombre ont été déterminées suivant une classification moléculaire non hiérarchique. L'analyse fonctionnelle de ces sous-classes a permis de dégager de nombreuses voies de signalisation et aberrations génomiques, par ailleurs déjà connues dans les CHC. En effet, la sous-classe A est fortement associée aux mutations du gène CNNTB1 et semble très proche des données antérieures de classification transcriptomique. Le sous-groupe B est caractérisé par une signature de gènes de réponse à l'interféron et est associé à des tumeurs de faible volume. Le sous-groupe C de mauvais pronostic est divisé en 3 catégories, dont l'une d'entre elles présente une expression très marquée des miARNs, en particulier dans une région chromosomique (chr19q13.41) riche en miARNs.

### 7.2.2.2 Longs ARNs non codants

A côté de la multitude d'ARNs non-codants ayant des profils d'expression et de régulation complexes et dont la taille est inférieure à 200 nucléotides comme les micro-ARNs, une large proportion du génome humain est composée de longs ARNs non-codants (lncRNA). Bien que leur rôle soit peu exploré, il est clair que certains d'entre eux jouent un rôle important dans la régulation de l'expression génique et sont impliqués dans diverses pathologies, notamment le cancer. Parmi les nombreux lncRNAs différentiellement exprimés dans le CHC comparé à la cirrhose ou au tissu sain, seuls quelques-uns sont particulièrement bien caractérisés sur le plan structural et fonctionnel. Par exemple, HULC (highly upregulated in liver cancer) agit en séquestrant des micro-ARNs comme miR-107 conduisant à une activation d'expression du facteur de transcription E2F1 et de la sphingosine kinase I, cette dernière ayant un rôle pro-angiogénique dans des modèles *in vivo* (Lu *et al.*, 2016). HULC est capable également d'activer la transition épithélio-mésenchymateuse via miR-200a-3p (Li *et al.*, 2016b), de déréguler le



**Figure 8: Transport et métabolisme du glucose dans des cellules normales et tumorales**

Dans les cellules normales, le glucose est transporté par GLUT1 et SGLT1 puis métabolisé par la glycolyse pour générer majoritairement du pyruvate qui subit l'oxydation mitochondriale. Le transporteur SMCT1 permet l'entrée de lactate et de pyruvate dans les cellules. Le métabolisme du glucose par la voie des pentoses phosphates (PPP) est très faible tout comme la synthèse des acides gras et du cholestérol. Les niveaux d'ATP sont maintenus par phosphorylation oxydative (chaîne respiratoire ETC/OXPHOS). Dans les cellules tumorales, GLUT1 et SGLT1 sont surexprimés. Le pyruvate généré par la glycolyse est majoritairement transformé en lactate, car la fonction mitochondriale est supprimée. MCT4 surexprimé exporte le lactate et SMCT1 est inhibé. Le métabolisme du glucose par la voie des pentoses phosphates est stimulé générant du ribose-5-phosphate pour la synthèse d'ADN/ARN et du NADPH pour la synthèse des acides gras et du cholestérol. Les niveaux d'ATP sont maintenus par la glycolyse. D'après Ganapathy *et al.* 2009

métabolisme lipidique par la voie PPAR $\alpha$ -acyl-Coa synthetase I (Cui *et al.*, 2015) et d'activer la traduction d'ARNs messagers oncogéniques (Li *et al.*, 2016b).

HOTAIR (Hox antisense intergenic RNA) est un lncRNA intergénique impliqué dans la régulation épigénétique des gènes. Fréquemment surexprimé dans les CHC et les lignées cellulaires dérivées de CHC, HOTAIR augmente le recrutement de PRC2 (Polycomb Group Complex 2) sur le génome et influence la répression de gènes cibles en modifiant des marques chromatinienne comme la triméthylation de la lysine 27 de l'histone H3 et la déméthylation de la lysine 4 de l'histone H3 (Tsai *et al.*, 2010).

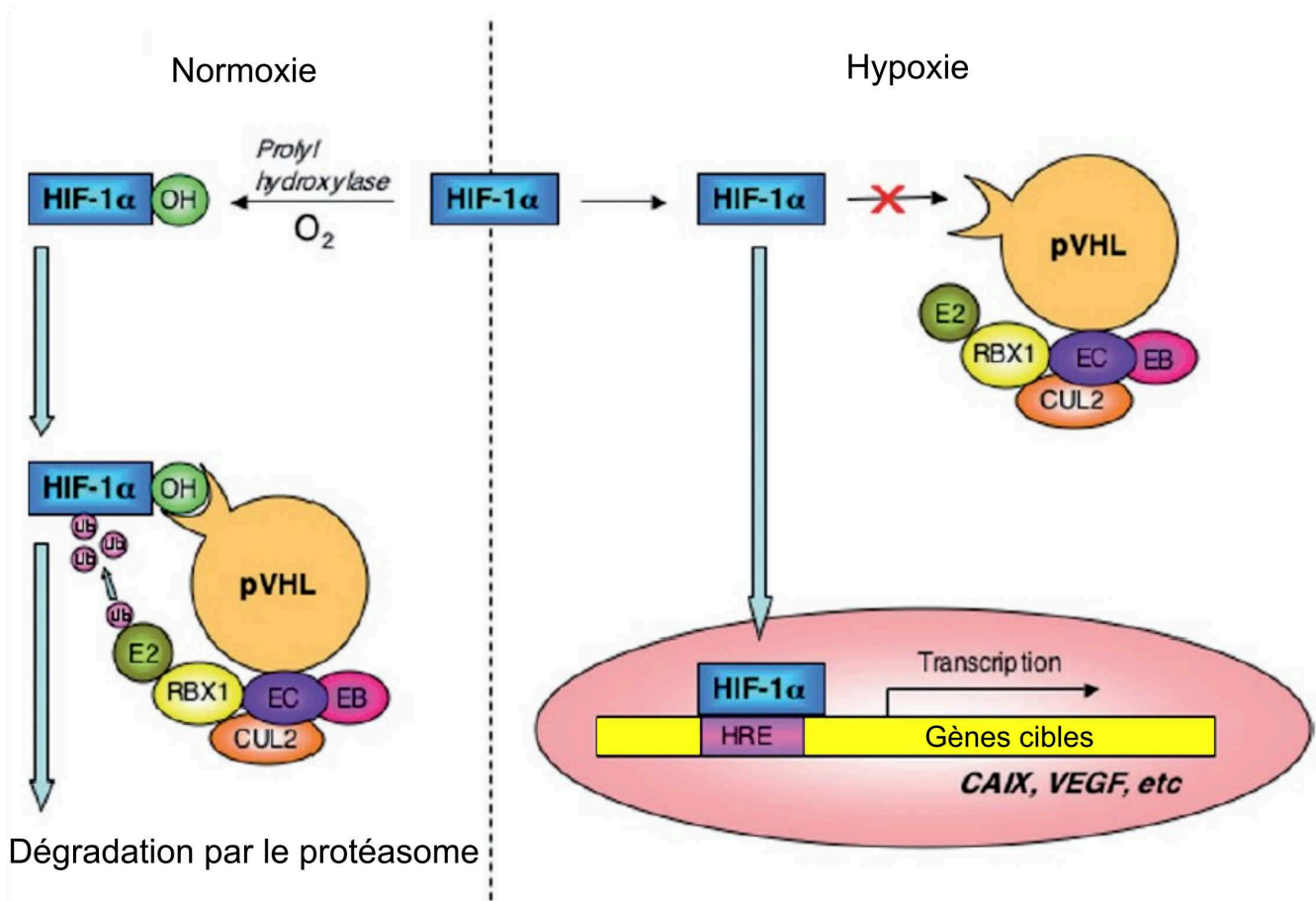
## 8. Le métabolisme du carcinome hépatocellulaire

### 8.1 Généralités sur le métabolisme des cellules tumorales

Les cellules tumorales présentent un métabolisme différent de celui des cellules normales. Il est fréquent d'observer une surexpression des voies de la glycolyse, du glycogène et de la néoglucogenèse dans le but d'ajuster le contrôle métabolique et d'assurer la croissance et les divisions cellulaires. Ce phénomène de reprogrammation métabolique des cellules cancéreuses avec hyper-consommation de glucose par la voie de la glycolyse au détriment de la voie oxydative fût pour la première fois décrit par le biochimiste allemand Otto Warburg.

Dans des conditions normales en présence d'oxygène, les cellules dégradent le glucose en pyruvate via la glycolyse et ce dernier est métabolisé via le cycle des acides tricarboxyliques (cycle de Krebs) et la phosphorylation oxydative dans le compartiment mitochondrial pour la production d'ATP. À travers cette voie mitochondriale, une molécule de glucose génère 36 à 38 molécules d'ATP. Dans des conditions d'environnement anaérobie, la glycolyse est la voie de dégradation préférentielle avec relargage de lactate (cf. Figure 8).

L'orientation métabolique des cellules cancéreuses qualifiée d'effet Warburg a pour objectif de faire face aux importants besoins spécifiques des cellules notamment en matière de production de précurseurs métaboliques nécessaires aux synthèses des protéines et des acides nucléiques. Elle est rendue possible par une augmentation de la captation et du transport intracellulaire du glucose par les transporteurs membranaires GLUT (DeBerardinis *et al.*, 2008; Hsu and Sabatini, 2008; Jones and Thompson, 2009) ainsi que par de très nombreuses modifications enzymatiques. Les cellules tumorales vont, même en présence d'oxygène, limiter leur production d'ATP (2 à 4



**Figure 9: La protéine HIF1-α dans des conditions normoxiques et hypoxiques**

Dans des conditions normoxiques (20 % d'oxygène), les prolyl hydroxylases (PHD) hydroxyle HIF1-α qui se fixe à VHL. VHL forme un complexe avec l'elongine B (EB), l'elongine C (EC), CUL2, RBX1 et l'enzyme d'ubiquitinylation E2. HIF1-α est ensuite ubiquitinylé et dégradé par le protéasome. A l'inverse, en absence d'oxygène, HIF1-α n'est pas hydroxylé et entre dans le noyau pour activer l'expression de gènes cibles comme le facteur d'angiogenèse VEGF et l'anhydrase carbonique 9 (CAIX). D'après Kitamura et Tsukamoto 2008

moles par mole de glucose) en utilisant préférentiellement la voie de la glycolyse (Warburg, 1927).

La glutamine est également utilisée comme autre substrat énergétique pour le cycle de Krebs et la production d'ATP et son catabolisme est fortement augmenté dans la glycolyse aérobie.

## 8.2 Voies de signalisation impliquées dans la reprogrammation métabolique du CHC

L'utilisation préférentielle du glucose est dirigée par diverses voies de régulation qui agissent sur la stabilisation de la protéine HIF1- $\alpha$  (Hypoxia-inducible factor 1 $\alpha$ ), et cela même en présence d'un environnement aérobie. Dans des conditions normoxiques (20 % d'oxygène), HIF1- $\alpha$  est rapidement hydroxylé par des prolyls hydroxylases (réaction dépendante d'oxygène et de fer). HIF1- $\alpha$  hydroxylé se fixe à un complexe protéique comprenant pVHL, les élongines B et C et la culline 2, cette dernière présentant une activité ubiquitine ligase. Après ubiquitinylation, HIF1- $\alpha$  est dégradé par le protéasome. En absence d'oxygène, la première réaction d'hydroxylation n'est pas déclenchée et la fixation de HIF1- $\alpha$  au complexe n'aboutit pas à sa dégradation. La protéine HIF1- $\alpha$  est alors transloquée dans le noyau où elle joue un rôle de facteur de transcription d'activation de gènes sous la dépendance d'éléments de réponse à l'hypoxie (HRE) (cf. Figure 9).

L'hypoxie est donc le principal facteur d'induction de HIF1- $\alpha$  mais la stabilisation de ce dernier peut également avoir lieu en présence d'oxygène.

### 8.2.1 Akt

Nous l'avons vu précédemment (cf. PI3-K p 43), la voie PI3-K-Akt-mTOR est fortement dérégulée dans les CHC et l'activation d'Akt par phosphorylation est observée dans ce cancer. L'expression et la translocation à la membrane plasmique du transporteur GLUT1 sont augmentées par Akt (Barthel *et al.*, 1999; von der Crone *et al.*, 2000, p. 4). De plus, Akt phosphoryle l'hexokinase HK2 et augmente son activité intracellulaire (Gottlob *et al.*, 2001; Miyamoto *et al.*, 2008). Enfin les cellules surexprimant Akt deviennent dépendantes du glucose pour proliférer (Elstrom *et al.*, 2004).

### 8.2.2 TP53

La protéine p53 a un effet négatif sur la glycolyse en inhibant l'import du glucose dans la cellule par inhibition de l'expression des transporteurs GLUT1 et GLUT4 (Kawauchi *et al.*,

2008). La protéine TIGAR, une des cibles de p53, du fait de sa similarité avec le domaine biphosphatase de l'enzyme 6-phosphofructo-2-kinase, réduit le niveau du fructose 2,6-biphosphate et donc inhibe la voie de la glycolyse. Enfin, des mutations inactivatrices de TP53 augmentent l'utilisation de la voie de la glycolyse aérobie par les cellules tumorales en induisant l'activation de NF- $\kappa$ B et de GLUT3 (Kawauchi *et al.*, 2008).

### 8.2.3 AMPK

Le rôle de l'activation de l'AMPK dans la reprogrammation métabolique des cellules tumorales est controversé. Certains travaux montrent une inhibition de la voie de la glycolyse aérobie au profit de la phosphorylation oxydative, conduisant à un arrêt de synthèse des acides nucléiques, des protéines et des lipides et donc à un arrêt de croissance cellulaire (Semenza *et al.*, 1994; Wouters and Koritzinsky, 2008). D'autres études, au contraire, montrent une activation d'expression de GLUT4 et donc d'une importation intracellulaire plus importante de glucose suite à une activation de l'AMPK (Hu *et al.*, 2008).

### 8.2.4 Mutations de la $\beta$ -caténine

Les mutations du gène CTNNB1 codant la  $\beta$ -caténine concernent plus d'un quart des CHC analysés (cf. Gène CTNNB1 p42) et provoquent des changements métaboliques profonds. Une étude très récente a montré dans des modèles animaux que la reprogrammation métabolique de type Warburg n'a pas lieu dans des CHC mutés pour le gène CTNNB1 et que la voie de la glycolyse n'est donc pas la voie métabolique préférentielle pour ces tumeurs. Les auteurs montrent que, dans ces CHC, les voies d'oxydation des acides gras ainsi que l'anabolisme des phospholipides sont privilégiés (Senni *et al.*, 2019). En accord avec ces données expérimentales, la même équipe montre que des CHC murins et humains mutés pour le gène CTNNB1 accumulent le  $^{18}\text{F}$ -flurorocholine, et non pas le fluorodésoxyglucose (Gougelet *et al.*, 2019).

### 8.2.5 Oncogènes

Certains oncogènes peuvent être responsables de la dérégulation métabolique observée dans les cellules cancéreuses. Le proto-oncogène Ras, lorsqu'il est muté, induit l'expression de la lactate déshydrogénase (LDHA) (Chiaradonna *et al.*, 2006) et conduit à l'augmentation de la captation et de la dégradation de glucose par la voie de la glycolyse (Yun *et al.*, 2009).



Des gènes de la glycolyse et du transport de glucose (enolase 1, hexokinase 2, GLUT1) sont des cibles de l'oncogène Myc (DeBerardinis *et al.*, 2008). De plus, Myc induit la dégradation de la glutamine par l'activation de l'expression de la glutaminase (Yuneva *et al.*, 2007).

## 8.2.6 Conséquences de la reprogrammation métabolique

### 8.2.6.1 Fonction mitochondriale et production de ROS

Un organisme aérobie produit naturellement des espèces réactives à l'oxygène (ROS). L'équilibre redox d'une cellule est essentiel à sa survie et est dépendante de la quantité de ROS produits par le métabolisme ainsi que leur recyclage par différents systèmes enzymatiques (catalase, SOD, peroxydases, etc.) et non-enzymatiques (glutathion, vitamines E et C, caroténoïdes, etc.) .

Une dérégulation du potentiel redox est fortement impliquée dans la progression tumorale et la résistance aux chimiothérapies (Hanahan and Weinberg, 2011). D'importantes quantités de ROS produits peuvent être cytotoxiques via les dommages causés aux macromolécules (Collins *et al.*, 2005). L'utilisation préférentielle de la glycolyse réduit *de facto* l'activité mitochondriale et la production mitochondriale de ROS, en particulier via HIF1- $\alpha$  (Yoshida *et al.*, 2000, Zhang *et al.*, 2007b). Dans certains cas pourtant, l'hypoxie conduit paradoxalement à une production importante de ROS (Chandel *et al.*, 1998; Guzy *et al.*, 2005). Ces différences entre études peuvent s'expliquer en partie par la nature des cellules étudiées et par les méthodes de mesure des taux intracellulaires de ROS (Denko, 2008).

Enfin, la synthèse du glutathion réduit, puissant anti-oxydant cellulaire, est dépendante des voies métaboliques du glucose et de la glutamine. En effet, le glucose via la voie des pentoses phosphates fournit le cofacteur NADPH et la glutamine sert de substrat à la synthèse du glutathion (Yuneva, 2008).

### 8.2.6.2 Biosynthèse de macromolécules

Outre la production d'ATP, le métabolisme du glucose et de la glutamine contribuent à la formation de macromolécules indispensables à la survie et à la croissance tumorale. Le glucose fournit les précurseurs de la synthèse du ribose via la voie des pentoses phosphates (PPP) (Wice *et al.*, 1981). Le métabolisme de la glutamine conduit à la synthèse des purines et pyrimidines (Zetterberg and Engström, 1981).

### 8.3 Hétérogénéité métabolique des cancers

Les tumeurs ont des profils métaboliques différents qui sont fonction de paramètres du micro-environnement comme la vascularisation, le taux d'oxygène, la mécano-transduction ou encore l'impact de drogues de chimiothérapie. Il a été montré, dans le cancer du sein, l'existence de sous-populations cellulaires tumorales sensibles, et d'autres résistantes, à la chimiothérapie (Gonzalez-Angulo *et al.*, 2007). Ces différents facteurs de pression de sélection contribuent largement à la survie et à la propagation loco-régionale et à distance du cancer. Par ailleurs, les différents sous-clones peuvent se comporter de manière « symbiotique ». En effet, il a été montré, dans des cellules MCF7 dérivées de cancer du sein humain et dans des modèles animaux, que des ensembles de cellules tumorales développaient un métabolisme glycolytique de type Warburg et sécrétaient du lactate qui était utilisé par d'autres sous-populations tumorales pour former du pyruvate et entrer dans un métabolisme oxydatif donnant ainsi un avantage de croissance aux deux types de populations tumorales (Sonveaux *et al.*, 2008).

Enfin, le concept de cellules souches cancéreuses (CSC) peut contribuer à l'hétérogénéité métabolique intra-tumorale. Un faible nombre de cellules souches à fort taux d'auto-renouvellement et résistantes aux thérapies anticancéreuses classiques alimenterait le pool de cellules tumorales et contribuerait au phénomène de la récurrence tumorale (Nabi and Le, 2018). Leur état « d'indifférenciation » ou de « dédifférenciation » serait associé à une diversité de phénotypes métaboliques (Prasetyanti and Medema, 2017) pouvant conduire à la coexistence de divers sous-clones au sein d'un même nodule tumoral.

### 8.4 Le métabolisme comme cible pour la thérapie anti-cancéreuse

Les enzymes des voies métaboliques dérégulées que nous venons de décrire sont des cibles thérapeutiques potentielles dont l'inhibition aurait pour objectif d'empêcher la tumeur de « s'alimenter » et donc de croître.

#### 8.4.1 Les transporteurs

Les transporteurs de glucose sont en amont des voies métaboliques majeures utilisées par les cellules cancéreuses et sont donc des cibles thérapeutiques potentielles. Les flavonoïdes, par exemple, ont la capacité d'inhiber les transporteurs de glucose. La phloretine et la sylibin sont capables d'inhiber l'import intracellulaire de glucose (Kobori *et al.*, 1997; Cui *et al.*, 2009; Wu *et al.*, 2009; Zhan *et al.*, 2011). À l'heure actuelle, aucune de ces molécules n'a fait l'objet d'étude clinique.

### 8.4.2 L'hexokinase 2

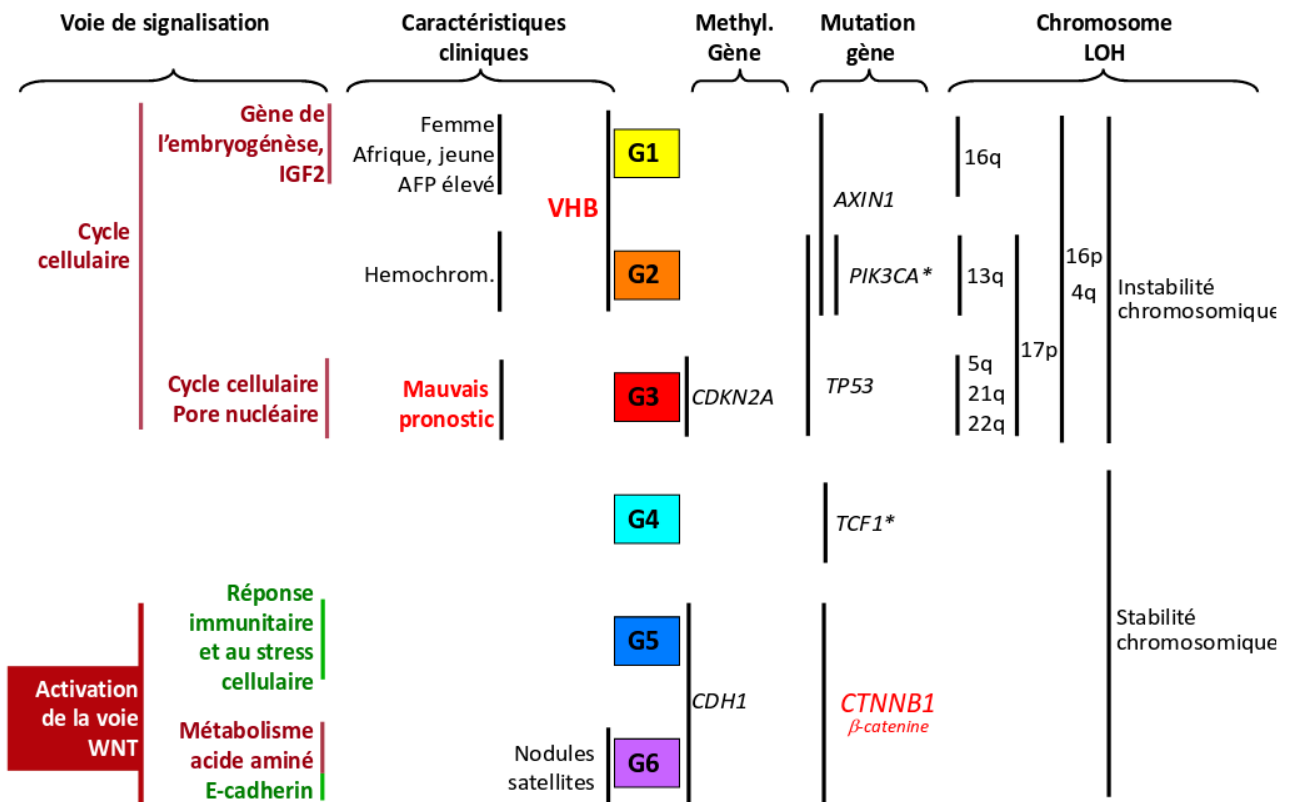
L'hexokinase 2 est la première enzyme de la voie de la glycolyse. En utilisant un analogue du glucose, le 2-déoxy-D-glucose, il a été montré une inactivation de l'hexokinase après phosphorylation (Aft *et al.*, 2002; Zhang *et al.*, 2015) dans des lignées cellulaires de CHC. La combinaison du 2-deoxy-D-glucose et de la doxorubicine ou du paclitaxel donne des résultats anti-tumoraux encourageants dans des modèles xéno greffés (Maschek *et al.*, 2004). Un autre analogue du glucose, le 3-bromopyruvate, a présenté des effets plus importants d'inhibition de la voie de l'hexokinase ; il a été testé chez un enfant atteint de CHC avec des résultats encourageants (Ko *et al.*, 2012) mais actuellement aucune étude clinique n'a été réalisée évaluant ce type de traitement.

### 8.4.3 La metformine

La metformine est un anti-hyperglycémiant oral utilisé pour le traitement du diabète de type 2. Elle est connue pour activer l'AMPK qui inhibe mTOR (Long and Zierath, 2006). De nombreuses études montrent que les patients prenant cette médication ont un risque réduit de développer un CHC (Donadon *et al.*, 2009, 2010; Hassan *et al.*, 2010; Tseng, 2018) ou encore qu'une survie prolongée a été enregistrée chez des patients atteints de cancer pancréatique ou de lymphome bénéficiant de cette médication (5,89 années contre 1,29 dans le lymphome ; 0,68 contre 0,22 dans le cancer du pancréas)(Wynn *et al.*, 2019). Dans des modèles animaux de cancer, la metformine réduit substantiellement le volume tumoral. Parmi toutes les molécules présentées dans ce chapitre, la metformine bénéficie d'une large littérature dans le traitement des cancers et de nombreux essais cliniques sont en cours pour démontrer son effet bénéfique en association avec des chimiothérapies classiques. Chez des patientes atteintes de cancer du sein, la metformine administrée en pré-opératoire a permis de diminuer significativement l'expression intra-tumorale de marqueurs de prolifération cellulaire et d'augmenter celle de marqueurs apoptotiques (Niraula *et al.*, 2012; Hadad *et al.*, 2015). À l'heure actuelle, quelques essais cliniques sont en cours dans la prévention ou le traitement du CHC en association avec d'autres molécules. Les résultats ne sont pas encore disponibles.

## 9. Classifications moléculaires

Les CHC présentent une pathogenèse et une évolution différentes suivant les patients ainsi qu'un pronostic variable reflétant une hétérogénéité moléculaire et des combinaisons spécifiques de multiples altérations génétiques et épigénétiques modifiant les voies



**Figure 10 Classification moléculaire des CHC à 6 groupes selon Boyault et al.**

Classification transcriptomique de 57 CHC et validée sur une cohorte indépendante de 63 autres CHC. Cette classification à 6 groupes regroupe des données d'altérations chromosomiques et mutationnelles. Cette classification est enrichie par des annotations clinico-pathologiques et de voies de signalisation dérégulées. D'après Boyault et al. 2007.

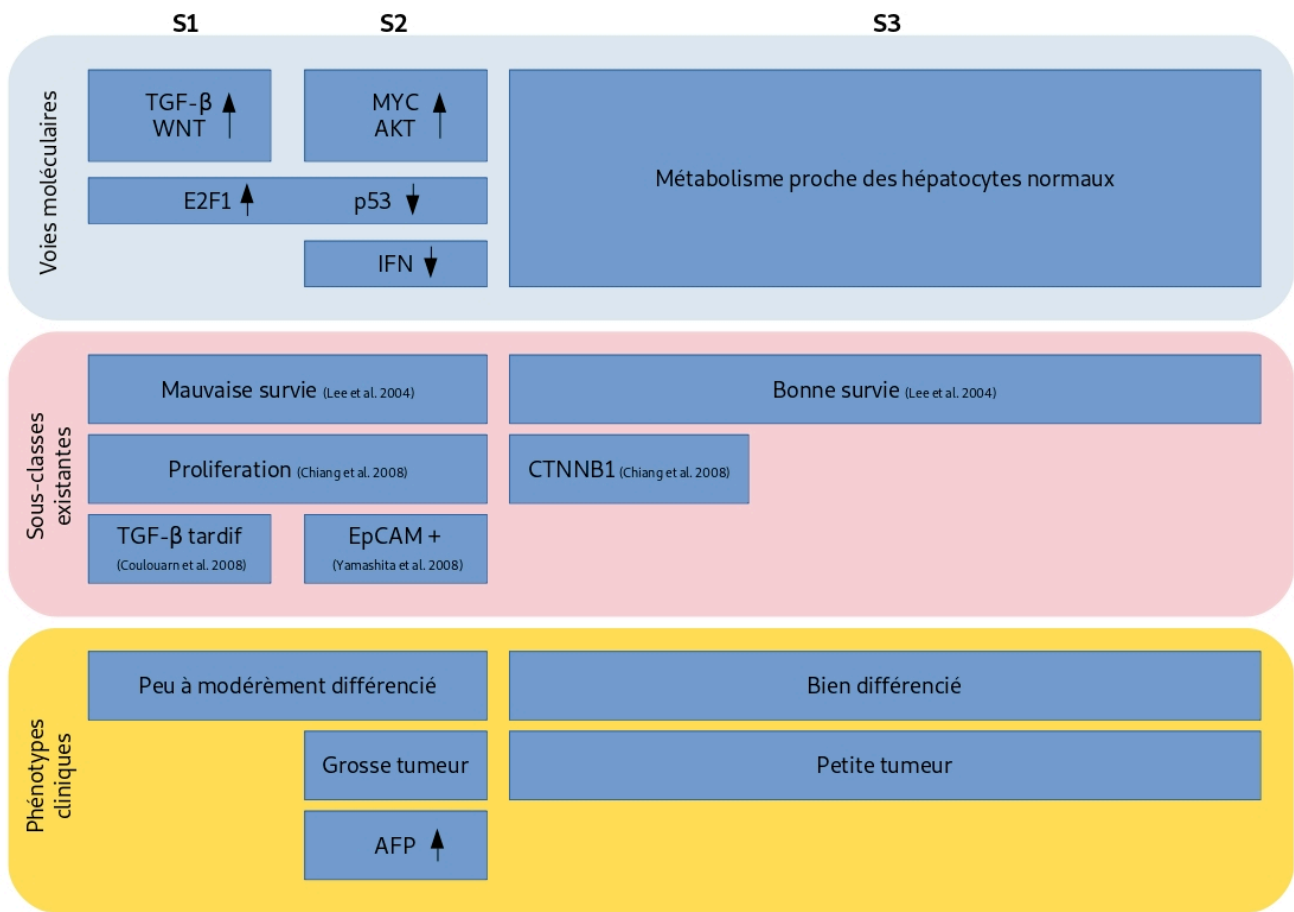
oncogéniques et suppressives de tumeur. L'existence de sous-classes de CHC biologiquement distinctes a été suggérée sur la base d'associations d'études pan-génomiques. Depuis environ une décennie, de nombreuses équipes à travers le monde ont proposé d'ajouter, aux classiques classifications anatomo-cliniques des cancers, une stratification moléculaire permettant de mieux caractériser la biologie de ces tumeurs et offrir aux patients des thérapies personnalisées.

## 9.1 Classifications transcriptionnelles

Les altérations génétiques identifiées dans le CHC ont en effet été associées à des critères clinico-pathologiques. Par exemple, P. Laurent-Puig *et al.* ont décrit l'existence de sous-classes de CHC humains en étudiant une cohorte de 137 tumeurs (Laurent-Puig *et al.*, 2001). Un premier sous-groupe de cancers corrélé à un mauvais pronostic est associé à une forte instabilité chromosomique et des pertes d'hétérozygotie fréquentes. Les gènes *TP53* et *AXIN* y sont fréquemment mutés. Le deuxième sous-groupe est caractérisé par une stabilité chromosomique et des mutations dans le gène *CTNNB1* codant la  $\beta$ -caténine. Ce même groupe de chercheurs réalisa plus tard, sur une autre cohorte de 120 CHC, une analyse génétique ainsi qu'une étude du profil d'expression des transcrits ARNm (transcriptome) sur 57 CHC. Six sous-groupes moléculaires nommés G1 à G6 ont été déterminés selon le profil transcriptomique des tumeurs, correspondant à des altérations différentielles de molécules et voies de signalisation. Les sous-groupes G1 à G3 sont caractérisés par une forte instabilité chromosomique comparés aux sous-groupes G4 à G6. Les sous-groupes G1 et G2 sont caractérisés par davantage d'infections par le virus de l'hépatite B, des mutations dans le gène *AXIN1* et une activation de la voie AKT. Dans le sous-groupe G3, les gènes du cycle cellulaire et de la prolifération sont fortement surexprimés, et ces altérations sont associées à un pronostic sombre des patients appartenant à ce sous-groupe. Les mutations dans le gène *CTNNB1* sont fréquemment retrouvées dans les sous-groupes G5 et G6 et sont associées à une signature moléculaire d'activation de la voie WNT.

Un résumé de cette classification moléculaire, tirée de la publication principale, est présenté dans la Figure 10.

Une importante méta-analyse réalisée à partir de 8 jeux de données publiques correspondant à 603 CHC fut par la suite conduite (Hoshida *et al.*, 2009). Trois sous-groupes transcriptomiques nommés S1 à S3 furent révélés (cf. Figure 11). Les sous-groupes S1 et S2 partagent des caractéristiques phénotypiques communes telles qu'une faible différenciation tumorale, une



**Figure 11: Classification moléculaire des CHC à 3 groupes selon Hoshida *et al.*** Une méta-analyse incluant 603 CHC provenant de 9 jeux de données transcriptomiques a défini trois sous-classes moléculaires caractérisées par un profil d'expression génomique et des annotations clinico-pathologiques spécifiques. Ce schéma illustre également les correspondances de ces 3 sous-classes moléculaires avec les signatures et sous-classes précédemment rapportées. D'après Hoshida *et al.* 2009.

importante prolifération cellulaire, un volume tumoral élevé et un mauvais pronostic clinique. De plus, d'un point de vue moléculaire, les sous-groupes S1 et S2 présentent une activation du facteur de transcription E2F1 et de ses gènes cibles ainsi qu'une répression de TP53 et des gènes cibles associés. Spécifiquement, le sous-groupe S1 est caractérisé par une signature d'activation des voies WNT (en absence de mutation du gène *CTNNB1*) et TGF- $\beta$ . Le sous-groupe S2 présente des taux sériques élevés d'AFP et une signature d'activation de l'oncogène *MYC*. Cette sous-classe a aussi la particularité de présenter un profil moléculaire dit « progéniteur » avec notamment une expression élevée du marqueur EPCAM. Nous reviendrons sur ce point dans le chapitre suivant. Enfin le sous-groupe S3 regroupe des tumeurs de petite taille bien différenciées, associées à une bonne survie des patients. Le gène *CTNNB1* est davantage muté dans ce dernier sous-groupe.

À la différence de l'étude réalisée par Boyault et collaborateurs (Boyault *et al.*, 2007), les infections virales B et C ne semblent pas appartenir à un sous-groupe moléculaire particulier. Cette méta-analyse montre également que la proportion de CHC appartenant aux différents sous-groupes moléculaires est sensiblement identique que les tumeurs proviennent de patients caucasiens ou asiatiques.

Enfin, un travail récent réalisé par le consortium international TCGA a produit une analyse moléculaire détaillée de 363 CHC au moyen de différentes plateformes d'exploration globale de l'épigénome et du génome codant ainsi qu'une exploration ciblée du microenvironnement tumoral et du protéome. Les caractéristiques ethniques des patients, les étiologies et la pathogenèse de ce panel de CHC sont variées. Le traitement des grands ensembles de données produites a conduit les auteurs à proposer une nouvelle classification moléculaire intégrant une grande diversité d'altérations suivant 3 clusters nommés iC1 à iC3. Ces clusters recourent en partie les sous-groupes S1 à S3 décrits par l'équipe du Prof. Hoshida (Hoshida *et al.*, 2009). Il faut noter qu'un grand nombre d'altérations identifiées par le consortium TCGA avait auparavant été décrit et qu'aucune nouvelle altération moléculaire parmi le petit nombre identifié (comme les mutations dans les gènes *LZTR1* et *EEF1A1*) ne pouvait malheureusement être la cible d'une thérapie.

Par la suite d'autres études n'ont fait que confirmer l'existence de ces sous-groupes moléculaires dans le cadre d'études de cohortes de CHC ou de méta-analyses indiquant la quasi-exhaustivité des altérations moléculaires décrites par la communauté scientifique, et, par là, la relevance biologique et clinique de ces défauts (Wang *et al.*, 2017).

Certaines études se sont concentrées sur des sous-ensembles de CHC aux caractéristiques clinico-pathologiques particulières comme, par exemple, des CHC bien différenciés associés à un très bon pronostic clinique (Désert *et al.*, 2017). Deux sous-groupes de CHC distincts par la présence ou non de mutations dans le gène *CTNNB1* ont été décrits : un sous-groupe dit ‘périportal’ (sans mutation du gène *CTNNB1*) et un sous-groupe dit ‘périveineux’ (avec mutations du gène *CTNNB1*) liés à l’expression de gènes impliqués dans le métabolisme et la zonation du foie. Dans le foie normal, le facteur de transcription hépatocytaire HNF4 $\alpha$  et la  $\beta$ -caténine coopèrent pour l’activation de la transcription de gènes liés à la zonation fonctionnelle et métabolique du foie (Colletti *et al.*, 2009; Torre *et al.*, 2011). Il a été montré que ce dialogue était maintenu dans les CHC bien différenciés. Il est à noter que les patients présentant une tumeur bien différenciée et non mutée pour le gène *CTNNB1* présentent un pronostic plus favorable que ceux ayant un CHC muté pour le gène *CTNNB1*. Cet intéressant travail a permis d’approfondir notre connaissance de la biologie de cette sous-classe moléculaire.

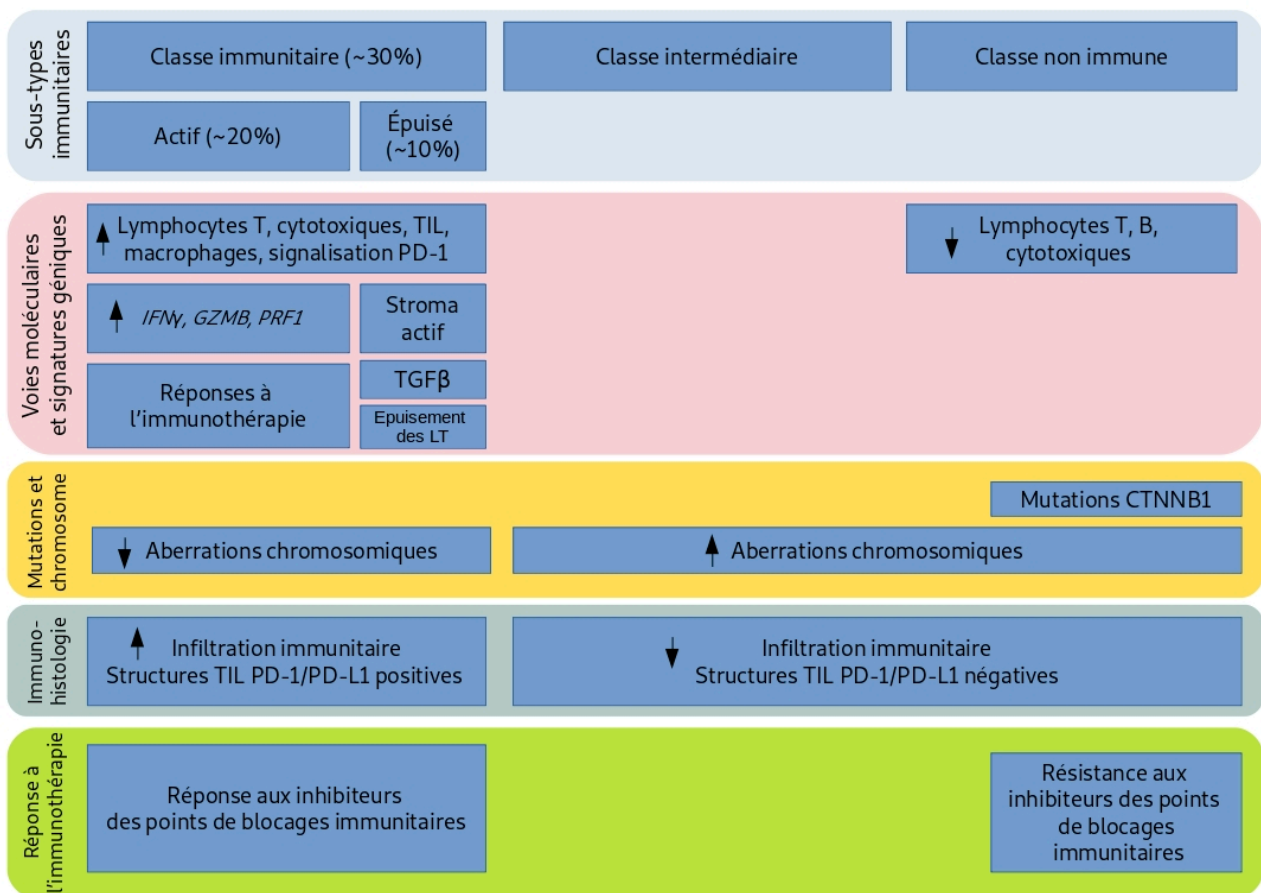
La grande majorité des travaux publiés a permis de stratifier les CHC selon leur profil transcriptomique avec l’objectif de définir des groupes homogènes de tumeurs. Il est également utile de définir des signatures transcriptionnelles partagées par tous les CHC analysés. Ce type d’étude qui a été dirigé par le Dr. Coulouarn (Allain *et al.*, 2016) sur 15 jeux de données publiques correspondant à 784 CHC a permis de dégager une signature de 935 gènes communs dérégulés par rapport au foie non tumoral. Cette signature était principalement associée à des altérations chromosomiques connues et d’apparition précoce dans le CHC. Cette étude a également permis de proposer des cibles thérapeutiques potentielles en complément des thérapies ciblées spécifiques de sous-classes moléculaires.

## 9.2 Signature immunitaire

Il est établi que la réponse immunitaire anti-tumorale est un facteur de survie notable chez les patients. Il est donc important d’identifier les groupes de patients qui développent cette réponse immune efficace et ceux dont la réponse immunitaire est inefficace ou absente afin de personnaliser le traitement du cancer. Il existe cependant à ce jour peu d’études de classification immunitaire des CHC.

Une étude pionnière (Sia *et al.*, 2017) fut conduite dans ce but avec l’analyse des profils d’expression des cellules immunitaires du microenvironnement tumoral chez environ 1000 CHC. Une sous-classe immunitaire particulière fut découverte et concernait environ un quart des cas de CHC de la cohorte. Cette sous-classe est caractérisée par de forts niveaux d’abondance de lymphocytes T cytotoxiques CD8+, de macrophages et de cellules Natural





D'après Pinyol et al. 2019

### Figure 12: Classification moléculaire des CHC basée sur le profil immunitaire des tumeurs et de leur microenvironnement

Environ 20 % des CHC appartiennent à la sous-classe dite « immunitaire » caractérisée par de forts niveaux d'infiltrations lymphocytaires, l'expression des points de contrôle immunitaire PD-1 et PD-L1, et l'activation de la voie de signalisation de l'interféron gamma. Ces tumeurs seraient potentiellement sensibles aux traitements par des inhibiteurs des points de contrôle immunitaire. À l'opposé, des CHC présentent un profil immunologiquement silencieux, parmi lesquels les CHC mutés pour le gène codant pour la  $\beta$ -caténine, et qui seraient résistantes à l'immunothérapie anti-cancéreuse. Enfin, il existe une classe dite intermédiaire en termes de degré d'infiltration lymphocytaire tumorale. D'après Pinyol *et al.* 2019

Killer (NK) constituant une infiltration cellulaire majeure du microenvironnement tumoral et représentant potentiellement un groupe candidat idéal pour une immunothérapie anti-cancéreuse. Il a également été observé que les patients appartenant à cette sous-classe ont une meilleure survie. Enfin, les tumeurs caractérisées par une signature WNT/ $\beta$ -caténine sont totalement absentes de ce sous-groupe immunitaire. Ce fait a été confirmé par une autre équipe qui a étudié une cohorte de patients asiatiques (Shimada *et al.*, 2018).

La Figure 12 montre une synthèse des différentes classifications moléculaires et immunitaires abordées dans ce chapitre.

### 9.3 Signature protéomique

Malgré les récents progrès technologiques, l'exploration protéique globale des cancers est sous-représentée dans les études OMICS. Quelques équipes ont cependant étudié le protéome altéré des CHC et établi des signatures protéiques spécifiques dans un nombre important de tumeurs. Par exemple, des sets de protéines dérégulées liés à un phénotype « progéniteur » (Orimo *et al.*, 2008), à la récurrence précoce du cancer (Morofuji *et al.*, 2016) ou à la résistance au Sorafenib (Kim *et al.*, 2017) ont été décrits. Certaines protéines comme ERK1 sont actuellement cibles d'inhibiteurs en cours d'essai clinique. Récemment, un consortium chinois a réalisé une importante analyse protéomique et phospho-protéomique dans les CHC de stade précoce de 100 patients (BCLC-0 et A) (Chinese Human Proteome Project (CNHPP) Consortium *et al.*, 2019). Une classification hiérarchique non supervisée a révélé l'existence de 3 sous-classes protéomiques. La première sous-classe est caractérisée par une surreprésentation de protéines impliquées dans le métabolisme et la différenciation hépatocytaire alors que les deux autres sous-classes sont associées à des protéomes impliqués dans la prolifération cellulaire. La stratification des CHC en trois sous-classes protéomiques est assez proche de la classification transcriptomique proposée par Hoshida *et al.* (Hoshida *et al.*, 2009). La classification protéomique détecte cependant des caractéristiques supplémentaires, par exemple, le degré d'infiltration immunitaire ou encore des altérations du métabolisme glucidique dans le cas de la sous-classe III de mauvais pronostic. Ces auteurs ont également décrit l'importance du métabolisme du cholestérol dans la carcinogenèse hépatique par la découverte d'un marqueur tissulaire SOAT1, qui lorsqu'il est bloqué par l'avasimibe (traitement de l'hypercholestérolémie familiale) inhibe la prolifération cellulaire et la progression tumorale dans des modèles cellulaires et animaux.

## 9.4 Signature métabolomique

L'exploration du métabolome connaît un essor considérable du fait de l'amélioration des techniques d'analyse, l'existence d'importantes bases de données de spectrométrie de masse et l'accès aisé aux fluides biologiques (sang, urine, liquide céphalo-rachidien, salive, etc.). La métabolomique constitue donc un outil de choix pour détecter des modifications cellulaires et tissulaires globales et ainsi identifier les changements physiologiques ou pathologiques associés à la santé et aux maladies. Dans le contexte hépatique, l'étude du métabolome est particulièrement informative car le foie est le siège d'une importante activité métabolique et enzymatique pour le contrôle de l'homéostasie de l'organisme tout entier.

Plusieurs études (rapportées dans la revue (Safaei *et al.*, 2016)) ont mis en évidence des voies métaboliques hépatiques différenciellement dérégulées dans l'état cirrhotique et cancéreux.

Le métabolisme lipidique est profondément perturbé avec une augmentation massive de nombreux phospholipides (Yang *et al.*, 2007) et une perte importante d'acides gras libres (Patterson *et al.*, 2011) probablement liée à une augmentation de la  $\beta$ -oxydation peroxysomale via le facteur PPAR- $\alpha$ , phénomène connu dans le CHC (Lalwani *et al.*, 1981). Cette diminution des lipides peut être également le témoin d'une utilisation préférentielle du glucose comme source d'énergie par les cellules cancéreuses. De plus, le niveau de saturation des lipides est modifié suivant le degré de différenciation de la tumeur. Dans les cancers peu différenciés de mauvais pronostic, le taux des acides gras insaturés est nettement plus important que celui des acides gras saturés (Budhu *et al.*, 2013) révélant une signature lipidique particulière associée à l'agressivité tumorale.

Le foie est l'organe producteur d'acides biliaires favorisant la digestion et l'absorption des nutriments. L'élévation des acides biliaires est caractéristique de la cirrhose (Raedsch *et al.*, 1981) et peut l'être dans le CHC en cas d'invasion tumorale et d'obstruction de la voie biliaire (El-Mir *et al.*, 2001). En règle générale, la production d'acides biliaires est diminuée dans la tumeur comparée à celle du foie cirrhotique.

Le stress oxydatif est une voie importante de carcinogenèse hépatique et de développement du CHC ; une signature de dommages oxydatifs est retrouvée dans le sérum de patients atteints de CHC. Une élévation de peptides  $\gamma$ -glutamyl, précurseurs du glutathion, est retrouvée chez ces patients (Soga *et al.*, 2011) ainsi que celle de la xanthine, un intermédiaire pour la production d'ion superoxyde (Fitian *et al.*, 2014).

Pour finir, une équipe espagnole a récemment réalisé une étude métabolomique avec l'objectif d'identifier des biomarqueurs sériques diagnostiques (Banales *et al.*, 2019). Parmi les 13 métabolites différentiellement abondants dans le sang circulant de patients avec CHC et d'individus sains, seul l'acide glycocholique a été retrouvé plus abondant dans le cancer. En combinaison avec l'AFP, leur sensibilité et spécificité se trouvaient être grandement améliorées pour le diagnostic du CHC.

## Préambule

Fort initialement d'une expertise en spectrométrie de masse appliquée aux cancers primitifs du foie, j'ai rejoint l'équipe « Microbiote, Inflammation et Cancer » dirigée par le Professeur Jamila Faivre avec l'objectif de développer des analyses intégrées de données de (méta)protéomique, (méta)génomique et (méta)transcriptomique ainsi que des outils de biostatistique de données « BIG DATA » appliquées aux cancers du foie et à son environnement (microbiote intestinal, système immunitaire).

Trois études moléculaires auxquelles j'ai largement participé vous sont présentées dans ce mémoire de thèse. Elles mettent en évidence des régulateurs potentiels jusque-là non décrits des maladies chroniques du foie et de la carcinogenèse hépatique. Dans le cadre d'un consortium européen (MODHEP), nous avons caractérisé en collaboration avec RIKEN institute (Japon) le génome non codant de CHC humains viro-induits et identifié en particulier une activation forte intra-tumorale des rétrotransposons à LTR, ces derniers représentant 8 % du génome humain, ouvrant ainsi un nouveau champ d'investigation du rôle des séquences répétées dans la régulation de l'expression génique dans le cancer. (Hashimoto *et al.* Genome Res 2015)

Le second travail avait pour objectif d'approfondir notre connaissance de la biologie d'un sous-type moléculaire de CHC exprimant des marqueurs progéniteurs. Nous avons caractérisé le protéome et le transcriptome altérés de ces tumeurs particulières, et révélons le rôle de la voie de réponse à l'hypoxie dans les altérations métaboliques présentes dans ce sous-type agressif de tumeurs, ouvrant ainsi la voie à de nouvelles perspectives de prise en charge des patients. (Dos Santos *et al.* Manuscrit en préparation)

Le troisième travail a porté sur la réponse à l'inflammation par les protéines antimicrobiennes (lectine REG3A) via la modification du microbiote intestinal, offrant un nouveau paradigme pour la régulation de l'interaction hôte-microbiote basée sur l'activité anti-oxydante d'une molécule de l'immunité innée. (Darnaud\*, Dos Santos\*, Gonzalez\* *et al.* Gastroenterology, 2018)

---

# II ÉTUDE 1 : ACTIVATION DE PROMOTEURS RÉTROVIRAUX LTR DANS LES CARCINOMES HÉPATOCELLULAIRES VIRUS-INDUITS

## 1. Présentation de l'étude

Le projet Human Genome, initié en 1990 et terminé dans les années 2003, avait pour but de séquencer le génome humain. Bien qu'il fût considéré comme une avancée majeure dans la compréhension du vivant, un grand nombre de questions émergèrent.

En 2003, le génome humain était finalement séquencé. Cet indéniable succès scientifique, vu à l'époque comme l'étape indispensable à la compréhension des processus biologiques, a rapidement fait place à d'autres questions. En effet, connaître la séquence de tous les gènes ne donne aucune indication sur les mécanismes d'activation et sur l'orchestration nécessaire qui contrôlent l'expression temporelle et spatiale des gènes et de leurs produits. Ensuite, il est rapidement apparu que les gènes codant pour les protéines ne représentaient qu'une petite fraction du génome, soit environ 2 %. La fonction de la majorité des trois milliards de bases composant le génome nous est en fait encore largement inconnue. Cette simple observation suggère que la complexité biologique tire son origine dans d'autres portions du génome. Dans cet univers de séquences non codantes, on trouve notamment les séquences régulatrices, qui permettent l'expression coordonnée et complexe des gènes, de manière spécifique dans un organe ou un type cellulaire donné, et qui participent à la complexité biologique. La création d'un catalogue des régulateurs de l'expression des gènes est nécessaire, non seulement pour mieux comprendre les phénomènes biologiques, mais également appréhender les mécanismes à l'oeuvre dans les états pathologiques.

Dans ce contexte, le projet ENCODE (Encyclopedia of DNA elements) s'est attaché depuis une dizaine d'années à annoter systématiquement le génome (voir le site <http://www.nature.com/encode/>); l'objectif principal est d'identifier les éléments fonctionnels. Un élément fonctionnel peut être défini comme une séquence déterminée sur le génome portant une information spécifique. Par exemple, l'information peut être celle qui permet la production d'un ARN, codant ou non codant. Mais l'élément fonctionnel peut également être défini par la présence d'une « signature protéique » sur l'ADN, tel que la liaison d'un facteur de transcription ou d'autres régulateurs d'expression. Il peut s'agir aussi d'une modification

biochimique de l'ADN ou des histones. Par exemple, la méthylation des cytosines dans les promoteurs des gènes est généralement associée à une répression de l'expression. D'autre part, la méthylation de la lysine 4 de l'histone 3 (H3K4Me3) marque un promoteur, alors que l'acétylation de l'histone 3 sur la lysine 27 (H3K27Ac) marque une séquence amplificatrice (enhancer) activée. L'ensemble de ces modifications complexes contribuent à la régulation épigénétique de l'expression des gènes. Comme on le voit, on s'affranchit d'un modèle linéaire qui stipule que l'une des quelque 20 000 séquences composant le génome codant contient l'information permettant la synthèse d'un ARN messager qui, lui-même, permettra la synthèse d'une protéine responsable de l'expression d'un caractère. Au contraire, l'un des messages importants du projet ENCODE est que plus de 80 % du génome contient des éléments fonctionnels pouvant être associés à des fonctions biochimiques dans au moins un type de cellules composant notre organisme. L'organisation du génome, telle que l'envisage ENCODE, implique une régulation en réseaux où de multiples régions sont appelées à moduler l'expression finale d'un ou plusieurs gènes, de façon à assurer une réponse coordonnée.

Les rétrotransposons à LTR constituent environ 8 % du génome humain. Ils sont divisés en deux groupes suivant leur capacité de rétrotransposition. On distingue les rétrotransposons LTR autonomes dérivés de rétrovirus (HERV). Bien que la plupart des gènes issus de HERVs ne sont pas intacts, les HERV à LTR comprennent des caractéristiques fonctionnelles comme des promoteurs, des sites de fixation aux facteurs de transcription et des séquences régulatrices.

En plus des HERVs, les rétrotransposons à LTR contiennent une deuxième grande classe que sont les MaLR (Mammalian apparent LTR Retrotransposons). Ces MaLRs représentent 3,65 % du génome et se caractérisent par une structure ressemblant aux rétrovirus à l'exception d'une perte des gènes *gag* et *pol* et donc une absence d'autonomie à rétrotransposer par manque de transcriptase inverse (Smit, 1993). L'intégration de ces éléments dans le génome est plus ancienne que celle des HERVs (80-100 millions d'années vs. 70 millions d'années pour HERV).

Depuis plusieurs années, de nombreuses équipes ont montré que les LTR pouvaient réguler l'expression de gènes voisins (cellulaires ou viraux) dans le développement placentaire et la cancérogenèse. La mutation insertionnelle, la recombinaison et le polymorphisme sont les trois principaux mécanismes de cette instabilité. De plus, l'hypométhylation, la transactivation et les transcrits anti-sens peuvent réguler l'activité des LTRs et *de facto* l'expression de gènes voisins.

Les éléments LTR et les rétrovirus endogènes contribuent à la carcinogenèse en affectant l'expression fonctionnelle de gènes adjacents. C'est le cas, par exemple, observé par une équipe allemande (Lamprecht *et al.*, 2010), dans le lymphome Hodgkinien. Le récepteur CSF1R joue un rôle essentiel dans la survie, la prolifération et la différenciation des précurseurs hématopoïétiques (Bourette *et al.*, 1993; Wrobel *et al.*, 2004). Son expression aberrante peut conduire à une prolifération cellulaire maligne, une invasion cellulaire et la formation de métastases dans de nombreux cancers tels que celui du sein (Hiraga and Nakamura, 2009; Patsialou *et al.*, 2009), de l'ovaire (Horiguchi *et al.*, 1988; Bast *et al.*, 1993) et de la prostate (Ide *et al.*, 2002). La transcription de ce récepteur dans les lymphomes est initiée par une séquence LTR de la famille des MaLR (THE1B) située en amont du promoteur classique. Ce mécanisme, dû à la perte du contrôle épigénétique, est la première démonstration que l'activation d'un élément LTR pouvait conduire à un cancer (Lamprecht *et al.*, 2010). Depuis, de nombreuses études ont montré l'activation de transcrits HERV LTR dans différents cancers (Boese *et al.*, 2000; Armbruster *et al.*, 2002; Goering *et al.*, 2011, Chen *et al.*, 2013b, p. 9; Singh *et al.*, 2013). Par exemple, dans la Leucémie myéloïde, HERV-K est fréquemment réactivé et le gène viral Np9 a été récemment identifié comme un oncogène (Chen *et al.*, 2013b). La protéine Np9 est retrouvée préférentiellement chez les patients atteints de leucémie. De plus, des expériences *in vitro* ont montré que l'expression de Np9 est essentielle à la survie et à la croissance cellulaires et activait de nombreuses voies de signalisation relatives à la dédifférenciation cellulaire (Wnt/ $\beta$ catenin, c-myc/AKT, ras-ERK, Notch1). Cette étude ouvre de nouvelles voies de recherche ainsi qu'une nouvelle définition de cibles thérapeutiques dans le traitement de la leucémie.

Enfin, récemment, il a été découvert que les LTRs étaient fortement exprimés dans les cellules souches embryonnaires et les iPS (Fort *et al.*, 2014) et que l'activation de certains d'entre eux était essentielle à la reprogrammation cellulaire (Lu *et al.*, 2014; Ohnuki *et al.*, 2014).

Prises dans leur ensemble, toutes ces études suggèrent un rôle capital de ce type de séquences génomique dans la tumorigenèse.



## 2. Article

Le séquençage du génome a permis d'élucider les voies moléculaires fréquemment perturbées dans le CHC, comme pour d'autres cancers, permettant potentiellement une intervention thérapeutique reposant sur la signature mutationnelle de la tumeur. La capacité de répertorier tout le spectre des aberrations génétiques survenant dans le CHC revêt donc une importance cruciale. Le but de notre étude a été de cartographier le génome non codant de tumeurs de CHC induites par des virus hépatotropes, avec l'objectif de découvrir de nouvelles altérations génétiques du CHC affectant le génome non codant et pouvant favoriser la prolifération et la transformation cellulaires. Nous avons établi la première carte de transcriptome à ARN non codants du CHC et révélé une forte activation des promoteurs LTR, qui sont principalement inhibés dans les cellules hépatiques normales, dans des CHC induits par les VHB et VHC. Certains des transcrits dérivés de LTR se sont révélés être des régulateurs clés de l'expression génique et donc activer la croissance cellulaire.

# CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors

Kosuke Hashimoto,<sup>1</sup> Ana Maria Suzuki,<sup>1</sup> Alexandre Dos Santos,<sup>2,3</sup> Christophe Desterke,<sup>2,3</sup> Agnese Collino,<sup>4</sup> Serena Ghisletti,<sup>4</sup> Emilie Braun,<sup>2,3</sup> Alessandro Bonetti,<sup>1</sup> Alexandre Fort,<sup>1</sup> Xian-Yang Qin,<sup>5</sup> Enrico Radaelli,<sup>6</sup> Bogumil Kaczkowski,<sup>1</sup> Alistair R.R. Forrest,<sup>1</sup> Soichi Kojima,<sup>5</sup> Didier Samuel,<sup>2,3</sup> Giocchino Natoli,<sup>4</sup> Marie Annick Buendia,<sup>2,3</sup> Jamila Faivre,<sup>2,3,7</sup> and Piero Carninci<sup>1</sup>

<sup>1</sup>RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, 230-0045 Japan; <sup>2</sup>INSERM, U1193, Paul-Brousse Hospital, Hepatobiliary Centre, 94800 Villejuif, France; <sup>3</sup>Université Paris Saclay, Faculté de Médecine Le Kremlin Bicêtre, 94800 Villejuif, France; <sup>4</sup>European Institute of Oncology (IEO), Department of Experimental Oncology, IFOM-IEO Campus, 20139 Milan, Italy; <sup>5</sup>RIKEN Center for Life Science Technologies, Division of Bio-function Dynamics Imaging, Wako, Saitama, 351-0198, Japan; <sup>6</sup>VIB Center for the Biology of Disease, KU Leuven Center for Human Genetics, B-3000 Leuven, Belgium; <sup>7</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Pôle de Biologie Médicale, Paul-Brousse Hospital, 94800 Villejuif, France

An increasing number of noncoding RNAs (ncRNAs) have been implicated in various human diseases including cancer; however, the ncRNA transcriptome of hepatocellular carcinoma (HCC) is largely unexplored. We used CAGE to map transcription start sites across various types of human and mouse HCCs with emphasis on ncRNAs distant from protein-coding genes. Here, we report that retroviral LTR promoters, expressed in healthy tissues such as testis and placenta but not liver, are widely activated in liver tumors. Despite HCC heterogeneity, a subset of LTR-derived ncRNAs were more than 10-fold up-regulated in the vast majority of samples. HCCs with a high LTR activity mostly had a viral etiology, were less differentiated, and showed higher risk of recurrence. ChIP-seq data show that MYC and MAX are associated with ncRNA deregulation. Globally, CAGE enabled us to build a mammalian promoter map for HCC, which uncovers a new layer of complexity in HCC genomics.

[Supplemental material is available for this article.]

Hepatocellular carcinoma (HCC) accounts for 70%–85% of total liver cancers. The major risk factors of HCC are chronic HBV and HCV infections and alcohol (Jemal et al. 2011). The development of HCC is a heterogeneous multistep process associated with genetic alteration and dysregulation of gene expression. Recent whole-genome studies in HCC identified tens of thousands of somatic mutations, several of which occur in chromatin regulators, suggesting that the transcriptional network might have been disrupted through reorganization of chromatin structure (Totoki et al. 2011; Fujimoto et al. 2012). Genome-wide analyses of gene expression in human HCC have identified overexpressed genes (Ladeiro et al. 2008), activated pathways (Huang et al. 2011), and subtypes of HCC (Boyault et al. 2007). However, most of the studies focused on protein-coding genes or microRNAs, and the repertoire of long ncRNAs in HCC tumor tissues remains largely unexplored.

Comprehensive transcriptome studies have revealed that a large proportion of mammalian genomes, including transposable elements (TEs), are transcribed (Kapranov et al. 2002, 2007; Okazaki et al. 2002; Carninci et al. 2005; Cheng et al. 2005;

Faulkner et al. 2009; Djebali et al. 2012). LTR retroposons are a major class of TEs, accounting for 8% of the human genome (Lander et al. 2001). The vast majority of long terminal repeat (LTR) retroposons have lost their internal-domain encoding genes and reside as solitary LTRs lacking the ability to retrotranspose (Kovalskaya et al. 2006). Nevertheless, regulatory sequences including promoters and transcription factor binding sites are widely observed within LTR elements (Bourque et al. 2008). We have recently shown that LTRs are massively transcribed in embryonic stem cells and iPS cells, some of which are involved in the maintenance of pluripotency (Fort et al. 2014). Other studies demonstrated a high activity of distinct LTR subfamilies in stem cells using several different methods, including RNA-seq (Kelley and Rinn 2012; St Laurent et al. 2013), DNase-seq (Jacques et al. 2013), and MeDIP-seq (Xie et al. 2013). Furthermore, an appropriate activation of LTRs is essential for iPS reprogramming (Lu et al. 2014; Ohnuki et al. 2014), suggesting that the expression of LTRs might be associated with cancerous features, such as poor differentiation and high proliferation potency.

The Cap Analysis of Gene Expression (CAGE) method has been widely used to identify transcription start sites (TSSs) of ncRNAs and messenger RNAs by capturing the capped 5' ends of

**Corresponding authors:** [carninci@riken.jp](mailto:carninci@riken.jp), [jamila.faivre@inserm.fr](mailto:jamila.faivre@inserm.fr)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.191031.115>. Freely available online through the *Genome Research* Open Access option.

© 2015 Hashimoto et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

the RNAs. The reproducibility of CAGE for expression measurement has been demonstrated through many studies, including FANTOM (The FANTOM Consortium et al. 2014) and ENCODE (Djebali et al. 2012), discovering a large number of novel ncRNAs and active enhancers. Here, we report the ncRNA transcriptome of human and *Mdr2* knockout (KO) mouse HCC using CAGE, with special emphasis on ncRNAs distant from protein-coding genes. We show that a large proportion of the distal ncRNAs are LTR-derived in human and mouse HCC genomes. The CAGE data revealed three well-defined subclasses of human HCCs corresponding to high, intermediate, and low expression levels of a selected set of LTR ncRNAs, respectively. The LTR-high subclass was correlated with definite clinical features (viral etiology, less differentiated tumors, high risk of recurrence) and MYC pathway activation. ChIP-seq data indicate an active role for transcription factors such as MYC-based complexes in the deregulation of LTR-ncRNAs in HCC.

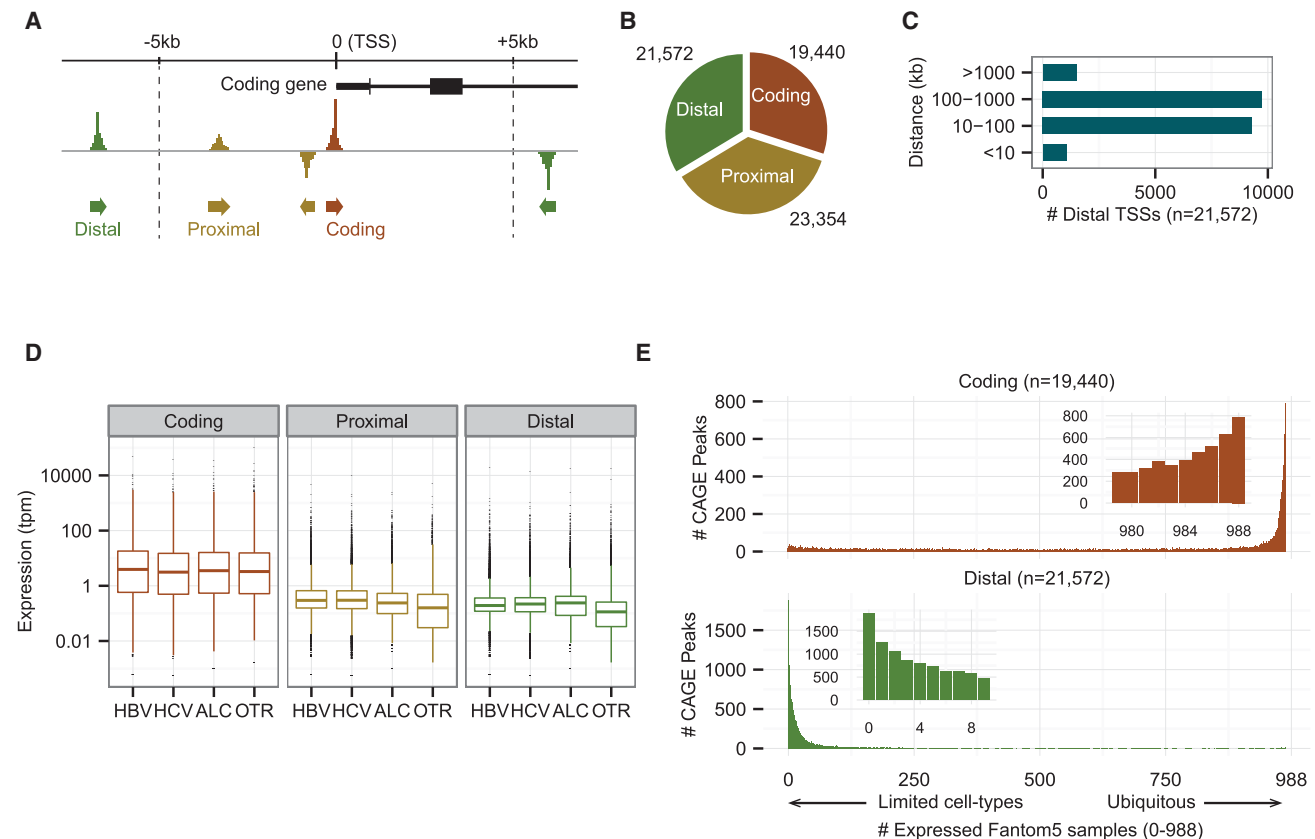
## Results

### Specific features of distal ncRNAs in HCC transcriptome

We sequenced CAGE libraries for 50 HCC tumor tissues and 50 matched nontumor (NT) tissues from patients with various etiologies, mostly HBV, HCV, and alcohol abuse. We also prepared

samples from morphologically normal liver tissues collected at a distance from a liver metastasis of colon cancer in five patients. These last five samples are referred to as “normal” as opposed to the “nontumor” and control for gene signatures affected in morphologically normal liver tissue by nearby inflammation and virus activities. The total number of uniquely mapped reads is 1.7 billion—an average “mapped read count” of 16.0 million for tumors, 15.9 million for nontumors, and 19.5 million for normal tissues (Supplemental Table S1). CAGE peaks (corresponding to TSSs) were determined based on the 5′ position of sequenced reads using Paraclu (Frith et al. 2008).

We identified 64,366 distinct CAGE peaks as the transcriptome of human HCC tumor tissues. All the peaks fulfill the following two criteria, as used previously (Fort et al. 2014): The expression level should exceed one tag per million (tpm) in at least one tumor sample; and the peak should be expressed in at least two tumor samples. Each individual peak represents a transcription initiation region at a median length of 45 bp covering 0.17% of the human genome. To distinguish potential ncRNAs from protein-coding genes, we classified the 64,366 peaks identified into three groups with the following distinctive features according to the GENCODE12 annotation (Harrow et al. 2012): (1) a coding peak, which overlaps with a protein-coding TSS ( $\pm 100$  bp) in sense orientation; (2) a proximal peak, which is located



**Figure 1.** The transcriptome of HCC tumors contains a large number of distal noncoding transcripts. (A) Schematic representation of the coding, proximal, and distal CAGE peaks. (B) Numbers of peaks in the different groups. (C) Distribution of distances between distal peaks and their closest coding peaks. (D) Expression levels in the different groups for the indicated etiologies. The data are means over tumors belonging to a given etiology group (tpm: tags per million + 0.001). (E) Expression patterns of coding and distal peaks across human body (988 human expression data from the FANTOM5 project). We counted the numbers of tags located within each TSS region in each sample, and then the number of samples that expressed tumor transcripts with the criterion of at least one tag. Coding peaks are ubiquitously expressed, whereas distal peaks are only expressed in limited cell types. Proximal and distal peaks show a similar expression pattern (Supplemental Fig. S1D).

within 5 kb of a coding peak or overlaps with a coding exon; and (3) a distal peak, which could not be assigned as either a coding nor a proximal peak (Fig. 1A). It can be assumed that distal peaks are mostly ncRNAs, whereas proximal peaks are mixtures of ncRNAs and unannotated alternative promoters of protein-coding genes. Distal peaks account for 33.5% (21,572/64,366) of the tumor transcriptome (Fig. 1B); half of them are located far (at >100 kb) from expressed coding peaks (Fig. 1C). On the other hand, proximal peaks are enriched in the upstream of coding peaks, which are likely to be products of bidirectional transcription (Supplemental Fig. S1A; Katayama et al. 2005). Expression levels of proximal peaks and their closest coding peaks tend to be positively correlated, whereas those of distal peaks and their closest coding peaks are not correlated, suggesting that distal peaks are not necessarily associated with their neighboring protein-coding genes (Supplemental Fig. S1B). One of the characteristics of ncRNAs is a lower expression level compared to coding RNAs, except for some massively transcribed ncRNAs such as *MALAT1*. Indeed, median expression of proximal and distal transcripts are ~20–30 times lower than coding transcripts in all the HCCs tumors, irrespective of the etiology (Fig. 1D). Despite their lower level of expression, 90% of the distal peaks are represented by a total of at least 50 reads, identified cumulatively across the collection of samples studied here, and are expressed in more than 10 tumor samples, indicating that the transcripts were not sample-dependent but common to most HCC tumors (Supplemental Fig. S1C).

To estimate how widely the HCC transcripts are expressed in the human body, we used the FANTOM5 atlas, which provides expression data measured by CAGE across a broad panel of primary cells, cell lines, and tissues, including normal adult and fetal liver tissues (The FANTOM Consortium et al. 2014). The distributions of expressed sample counts are strikingly different between coding and distal transcripts (Fig. 1E; Supplemental Fig. S1D). About 80% of coding transcripts are expressed in more than half of the FANTOM5 samples, with the highest peak (mode) at 988 samples, whereas ~94% of distal transcripts are detected in less than half of the FANTOM5 samples, with the highest peak at 0 samples, indicating that the distal transcripts are expressed in a limited number of tissues and cell lineages.

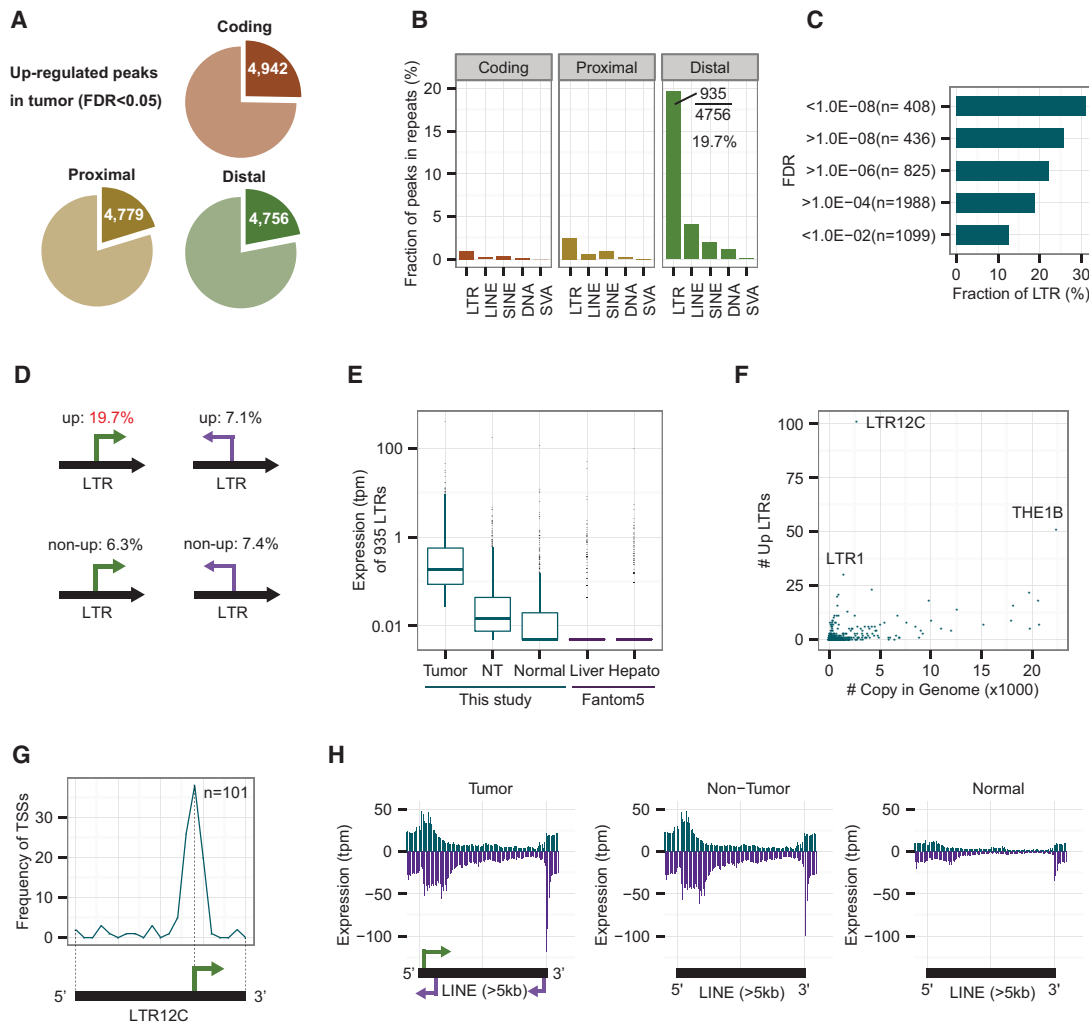
### Aberrant activation of LTR retroviral promoters in HCC tumors

We compared expression levels between a case group of 50 tumors and a control group of 50 matched nontumors using edgeR (Robinson et al. 2010) to explore misactivated transcripts in tumors (see Methods). We identified 14,477 significantly up-regulated peaks in tumors with a low FDR threshold ( $FDR < 0.05$ ), among which 4942, 4779, and 4756 were coding, proximal, and distal peaks, respectively (Fig. 2A). The most significantly up-regulated protein-coding gene is *GPC3* (Supplemental Fig. S2A), which is a known biomarker and a possible target for the treatment of HCC (Capurro et al. 2003; Gao et al. 2014). About 42% of the top 100 distal peaks overlapped with repetitive elements in the sense direction. To estimate the trend toward association of up-regulated peaks and repetitive elements, we examined which fraction of the up-regulated peaks overlapped with major repetitive elements (LINE, LTR, and SINE). The family names and genomic positions of repetitive elements we used are those defined by RepeatMasker (<http://www.repeatmasker.org>) and RepBase (Jurka et al. 2005) based on their classification and nomenclature of eukaryotic transposable elements (Kapitonov and Jurka 2008). Interestingly, ~20% (935/4756) of the up-regulated distal peaks overlapped with LTR el-

ements in the sense direction (Fig. 2B). The fraction of LTRs increases to 30% (127/408) when limited to the most significantly up-regulated peaks with FDR below  $1.0 \times 10^{-8}$  (Fig. 2C). These percentages are significantly higher ( $P < 1.0 \times 10^{-5}$  in all cases; one-sided Fisher's exact test) than those found for (1) non-up-regulated peaks (6.3%) (Supplemental Fig. S2B), (2) up-regulated peaks in the antisense direction (7.1%) (Supplemental Fig. S2C), (3) non-up-regulated peaks in the anti-sense direction (7.4%) (Supplemental Fig. S2D), (4) LINE and SINE elements in sense and antisense (1%–7%) (Fig. 2B; Supplemental Fig. S2C), and (5) randomized distal peaks (5%–6%) (Supplemental Fig. S2E). A summary of the main four overlap patterns is shown in Figure 2D. It was also true (1) when we excluded “-int” elements from the LTR category, which are internal elements of ERV (endogenous retrovirus) (Supplemental Fig. S2F), (2) when we used tags with mapping quality 30 or more (Supplemental Fig. S2G), and (3) when we used GENCODE v.19 annotations for our peak classification instead of GENCODE v.12 (Supplemental Fig. S2H). This clearly indicates that the promoters located in LTR elements are widely activated in HCCs.

Next, we studied the expression levels of the 935 activated LTR promoters in 10 normal livers using three data sets: normal liver tissues from this study ( $n = 5$ ) (Supplemental Table S1); adult and fetus liver tissues from FANTOM5 ( $n = 2$ ) (Supplemental Table S2); and primary hepatocytes from FANTOM5 ( $n = 3$ ) (Supplemental Table S2). As expected, the expression levels of LTRs in normal livers are much lower than in nontumor and tumor HCC tissues. More than half of the LTRs were not expressed at all in any of the 10 normal samples (Fig. 2E). We thus conclude that the up-regulation of LTR promoters is a hallmark feature of HCCs. Note that we have confirmed that the non-LTR distal transcripts are also weakly expressed in normal livers (Supplemental Fig. S2I).

About 500 LTR subfamilies with various copy numbers have been defined by RepeatMasker and RepBase for the hg19 human genome. If LTR promoters were randomly activated in unusual conditions such as long-term inflammation, the numbers of up-regulated LTRs should correlate with copy numbers in the genome. If, on the contrary, they were activated through a selective process, one should observe the activation of specific subfamilies, such as LTR7 and HERVH-int, which are markers for pluripotent cells (Santoni et al. 2012; Fort et al. 2014). Figure 2F shows that some subfamilies with relatively small numbers of copies per genome were preferentially activated in HCC tumors compared to the random distribution (Supplemental Fig. S2J), especially the LTR12C subfamily of which 101 elements (104 CAGE peaks) were significantly up-regulated in HCC tumors (Supplemental Table S3). Moreover, the peaks were consistently located at a specific position on the LTR12C elements (Fig. 2G), implying that promoter architecture of LTR elements is conserved. In many cases, the transcription start sites of LTR12C were located within a specific 5-nt (GTGGC) sequence (Supplemental Fig. S2K), whereas nonexpressed LTR12C elements often lost this promoter sequence. Nevertheless, the median size (1411 bp) of 101 up-regulated LTR12C members was similar to the median (1417 bp) of all LTR12C members (Supplemental Fig. S2L), and a subset of nonexpressed LTR12C retained a promoter sequence similar to that of the active elements, indicating that the active and inactive members cannot be distinguished solely by the presence or absence of promoter sequences. Other subfamilies than LTR12C showed enrichment of TSSs around the middle of the elements, which were slightly upstream compared to the position they had in LTR12C (Supplemental Fig. S2M). Conversely, only five elements were



**Figure 2.** LTR retroviral promoters are up-regulated in human HCC tumors. (A) Fractions of significantly up-regulated coding, proximal, and distal peaks in HCC tumors. The threshold is set to  $P < 0.05$  (FDR, calculated with edgeR). (B) Fractions of peaks overlapping with repetitive elements in the sense direction. Nearly 20% (935/4756) of the up-regulated distal peaks overlap with LTR elements. (C) Fractions of peaks overlapping with LTRs at different levels of significance. The most significant bin (FDR  $< 1.0 \times 10^{-8}$ ) shows the highest fraction (30%) of LTRs. (D) A summary of peaks in LTR elements (sense versus antisense and up versus non-up). Fractions of peaks in other repeats do not show such differences (Fig. 2B; Supplemental S2B–D). For example, SINE elements are 2.0% for up and 2.6%–4.2% for other cases, while SVA elements are 0.1% for up and 0.0%–0.1% for other cases. (E) Expression levels of 935 LTRs in tumor, nontumor (NT), and normal liver tissues. “Liver” (normal adult and fetus liver tissues) and “Hepato” (three biological replicates of primary hepatocytes) data from the FANTOM5 Project are shown for comparison. To avoid log-of-zero errors, 0.005 was added to the tpm values for this box plot. The median values of “Normal,” “Liver,” and “Hepato” are 0.005, indicating 0 tpm. (F) Activated LTR subfamilies versus copy numbers in the human genome. (G) Relative position of CAGE peaks on LTR12C. The majority of CAGE peaks are located at a relative position of 70% from the 5' end. Black bar and green arrow: presumed promoter architecture conserved among LTR12C elements. (H) Distribution of CAGE tags on long (>5 kb) LINE elements. Blue (purple) bars: sense (antisense) direction with respect to the LINE elements. Arrows: potential promoters embedded in full-length L1 elements.

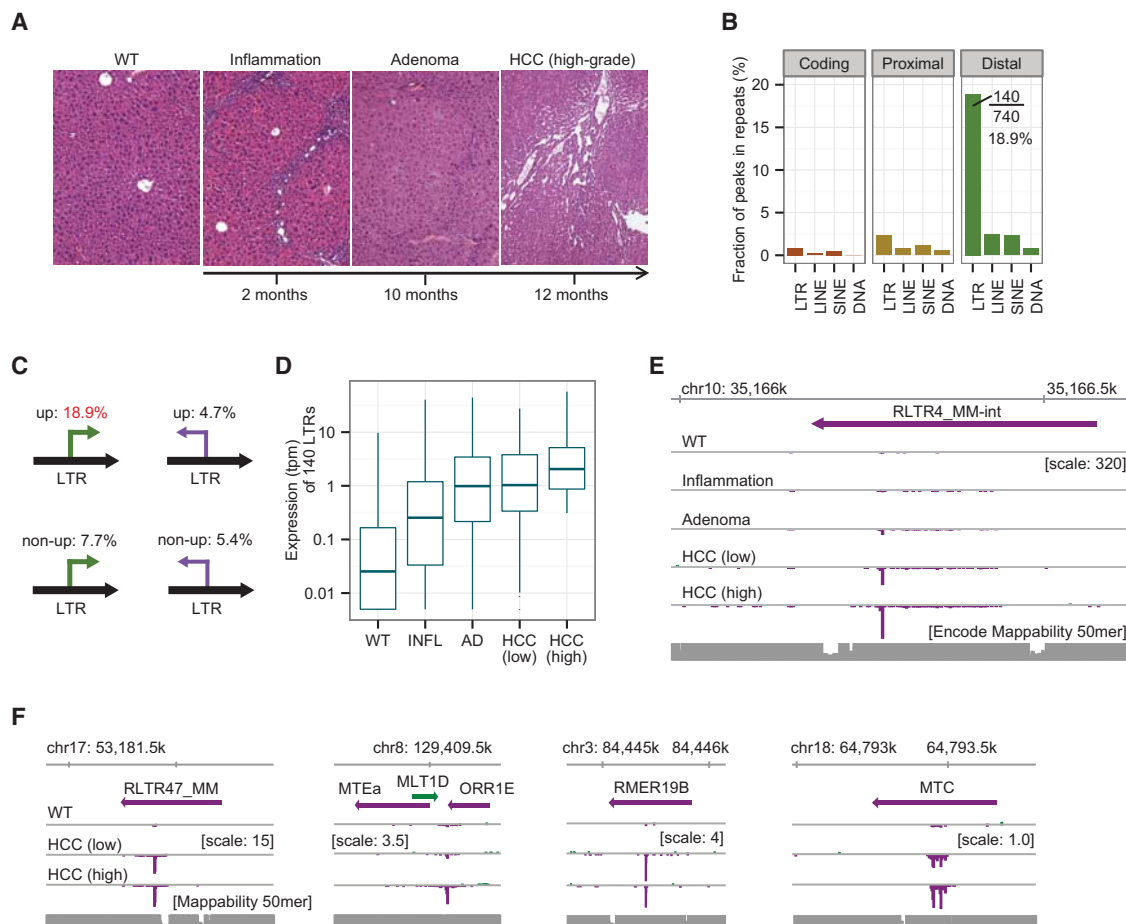
up-regulated in the LTR7 and HERVH-int subfamilies, which, it is worth noting, have at least hundreds of elements transcribed in stem cells. In summary, nearly 1000 LTRs were up-regulated in HCC, only a few of which were stem cell-specific LTRs.

The LTRs identified are mostly solitary ones, with a median size of 472 bp, which probably lost their ability to replicate but kept their promoter activity. On the other hand, full-length LINE elements are known to be activated with copy number increase in multiple cancers (Lee et al. 2012). To investigate the activity of full-length LINE elements, we mapped CAGE tags to L1 elements longer than 5 kb. Interestingly, tumor and nontumor HCC samples showed very similar expression patterns, irrespective of the etiology, whereas normal livers had very low expression lev-

els of L1 elements (Fig. 2H; Supplemental Fig. S2N). This is consistent with our previous report that increases in copy number were to be found both in tumor and nontumor areas of HCCs (Shukla et al. 2013), although the multiple steps ranging from expression of full-length LINE elements to their retrotranspositions will need to be validated by further experiments. In addition, the antisense promoters that the L1 elements are known to have were clearly observed around 5' and 3' regions (Fig. 2H).

#### LTR activation increases as HCC progresses in *Mdr2* knockout mice

We previously reported that LTR activation is frequently observed in human and mouse stem cells (Fort et al. 2014). Given that LTR



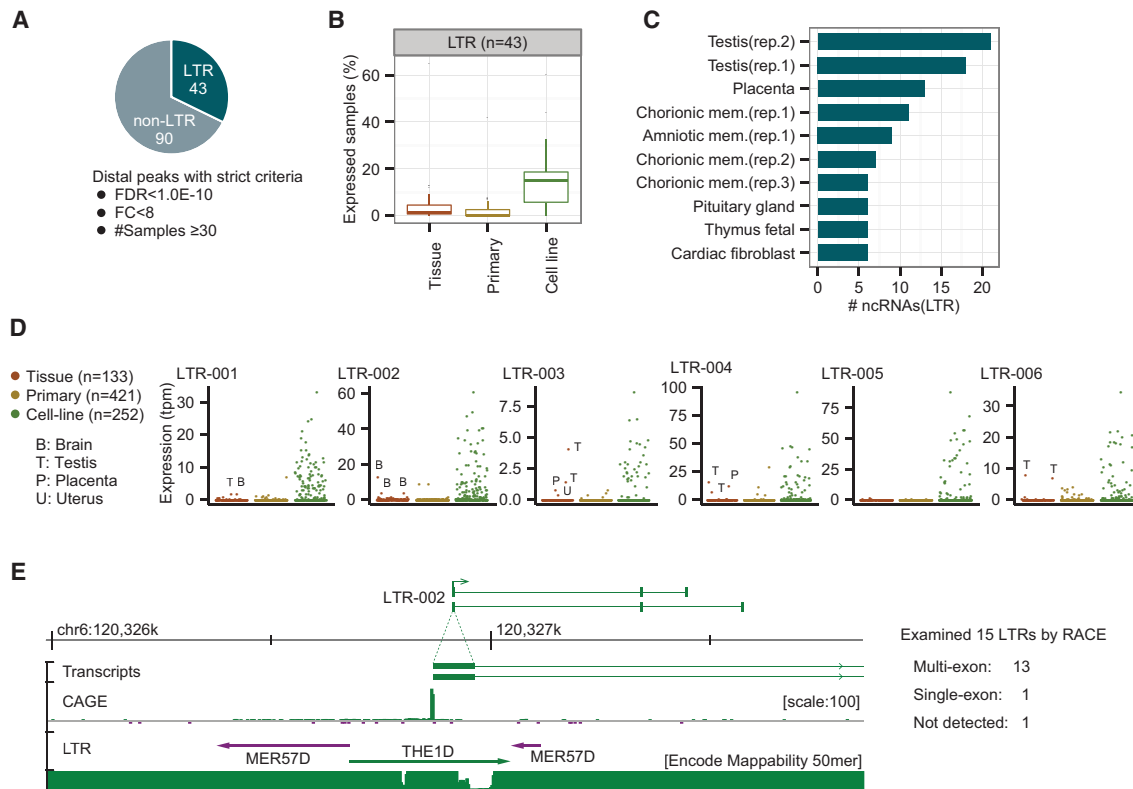
**Figure 3.** LTR promoters are activated in mouse HCCs. (A) Representative liver histological sections at the indicated stages of carcinogenesis in *Mdr2* KO mice. (B) Fraction of peaks overlapping with repetitive elements in the sense direction. Nearly 19% (140/740) of up-regulated distal peaks overlap with LTR elements. (C) Fractions of peaks overlapping with LTRs according to the (sense versus antisense) direction of transcription and the degree of up-regulation. (D) Expression levels of the 140 LTRs in normal (WT), inflammatory (INFL), adenoma (AD), low-grade, and high-grade HCC tissues. The data are means over tumors belonging to a given histological group (tpm + 0.005). (E) CAGE signal of the most significantly up-regulated LTR promoter at the indicated stages of the liver disease. The signal increases as the disease develops. CAGE tags are pooled for each group and visualized using the ZENBU Browser. (F) Expression patterns of LTRs selected from four different subfamilies.

activation occurs in human HCC, as shown above, we assumed that it also occurs in mouse HCC, giving us the opportunity to study LTR activation at different stages of liver carcinogenesis. To this aim, we used the *Mdr2* KO mouse, a well-known model of inflammation-driven HCC (Pikarsky et al. 2004; Iannelli et al. 2014). We examined 37 liver samples from either wild-type (WT) or *Mdr2* KO mice. Liver tissues from *Mdr2* KO mice were categorized into the following four histological groups: inflammation; adenoma; low-grade HCC; or high-grade HCC (>50%) (Fig. 3A; Supplemental Table S4; Katzenellenbogen et al. 2007). We sequenced CAGE libraries for six WT, three inflammation, four adenoma, 16 low-grade HCC, and eight high-grade HCC samples. We determined 49,096 CAGE peaks classified into coding, proximal, and distal groups using the Ensembl annotation (Supplemental Fig. S3A). By comparing the expression levels of WT ( $n = 3$ ) and high-grade HCC ( $n = 4$ ) tissues, we identified 740 significantly up-regulated distal peaks characteristic of high-grade HCC (Supplemental Fig. S3B). As in the case of human HCC, ~19% (140/740) of the up-regulated distal peaks were located within LTR elements with the same strand (Fig. 3B; Supplemental Table S5), which is a much higher proportion than for the other patterns (Fig. 3C; Supplemental Fig. S3C).

It was also true (1) when we excluded “-int” elements from the LTR category and (2) when we used tags with a mapping quality of 30 or more (Supplemental Fig. S3D). These same 140 LTRs were also overexpressed in adenoma and HCC tissues of *Mdr2* KO mice starting from the preneoplastic inflammation stage (Fig. 3D). Figure 3E shows the most significantly up-regulated distal peaks derived from LTRs, the expression levels of which increased as carcinogenesis progressed. The expression patterns of the four different elements selected from noninternal LTR families show consistent promoter activation within LTR elements in low- and high-grade HCC (Fig. 3F). Taken together, these results indicate that activation of LTR promoters is a shared feature of human HCC and the mouse HCC model studied.

#### Molecular signature of HCC comprising top 43 up-regulated LTR-derived transcripts

Having found significantly up-regulated distal peaks with a low (<0.05) FDR threshold in human and mouse HCCs, we sought to select human ncRNA candidates appropriate for determining a



**Figure 4.** LTR-derived ncRNAs found in HCCs are also expressed in reproductive tissues and cell lines. (A) Number of LTR-derived (43) and non-LTR-derived (90) ncRNAs among the up-regulated distal ncRNAs. The selection criteria are as follows: FDR threshold <math>1.0 \times 10^{-10}</math>; fold change > 8; expression in at least 30 tumor samples. (B) Proportion of tissues, primary cells, and cell lines that expressed the 43 LTR-derived ncRNAs according to FANTOM5. The LTR expression threshold value is the median expression level of the working set of 50 HCC tumors. (C) Top 10 normal tissues and primary cells that express the LTRs according to FANTOM5. (Rep) Biological replicates prepared in FANTOM5. (D) Expression patterns of the six most strongly expressed LTRs in the human body. LTR-003, 004, and 006 are only expressed in reproductive tissues (testis, placenta, and/or uterus). LTR-001 and 002 are expressed in brain. (E) Full-length ncRNAs with LTR promoters determined by 3' RACE. Thirteen out of 15 LTRs are confirmed as promoters of multiexon ncRNAs. LTR-002 is shown as an example of identified transcripts.

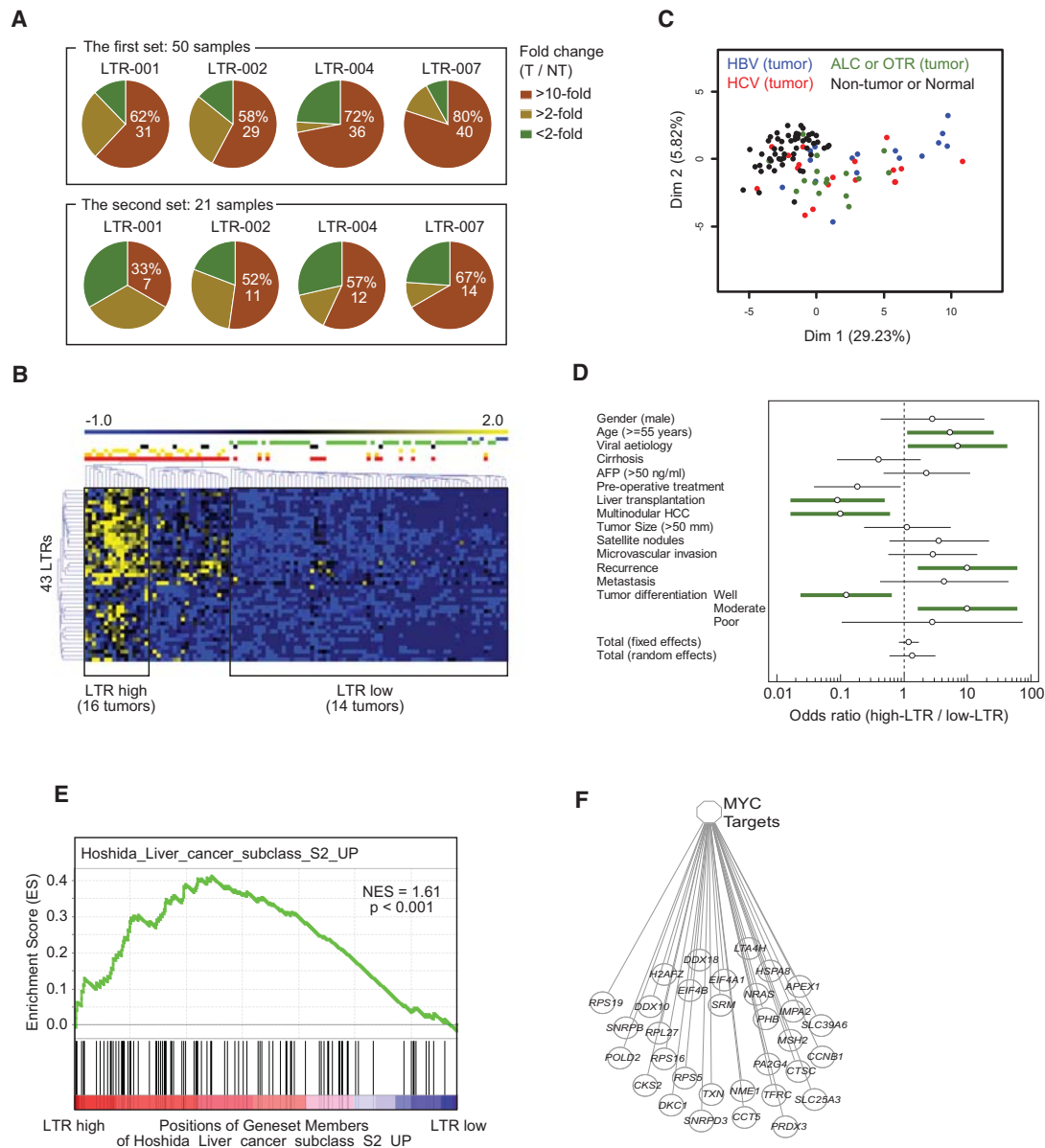
molecular signature of HCC using the following stringent criteria: (1) FDR threshold below  $1.0 \times 10^{-10}$ ; (2) fold change above 8.0; and (3) expression in at least 30 tumor samples. About one-third of the distal peaks (43/133) that passed the criteria coincided with LTRs (Fig. 4A). These 43 signature-LTRs were given sequential names from LTR-001 to LTR-043 in ascending order of FDR values. For a full list and the expression patterns, see Supplemental Table S6 and Supplemental Figure S4A. The fact that the signature-LTRs are scarcely transcribed in normal liver tissues raises the question of the nature of the tissues in which they are intrinsically programmed for expression. According to FANTOM5, their expression is limited to 1%–2% of the normal tissues and primary cells (Fig. 4B; Supplemental Fig. S4B). Intriguingly, reproductive tissues and primary cells occupy the top seven positions. Testis, which expresses about half of the 43 signature-LTRs, is in the first position followed by placenta, chorionic, and amniotic membranes (Fig. 4C). Some of the highest-ranking signature-LTRs, such as LTR-003, LTR-004, and LTR-006, are exclusively expressed in reproduction-related tissues (Fig. 4D). In contrast, the top six signature-LTRs are widely detected in all hepatic and nonhepatic cancer cell lines (median: 38 samples) (Fig. 4B,D; Supplemental Fig. S4B), emphasizing the close link between these LTRs and carcinogenesis.

We submitted 15 of the 43 signature-LTRs to 3' RACE validation (Supplemental Table S7), among which LTR-001 and 002 were further investigated by nested PCRs using 10 more primer

pairs designed on different exons (Supplemental Fig. S4C). Deep sequencing of 3' RACE and PCR products using MiSeq successfully determined the transcript structures of 14 of them, among which 13 had at least one splice site (Supplemental Table S8). The 33 transcripts detected, including splice variants, were all predicted to be ncRNAs with low coding potential by the Coding-Potential Assessment Tool (CPAT) (Supplemental Table S9; Wang et al. 2013). The median exon size is 154 bp, which is comparable to the average coding gene size (122 bp), whereas the median intron size is ~37 kb, i.e., much longer than the general protein-coding gene size (1 kb) (Lander et al. 2001). Figure 4E shows an example of novel transcripts, LTR-002, determined by the 3' RACE. The TSS is located at the middle of an LTR element, and the second and third exons are located far downstream from the first exon without encoding any viral proteins, suggesting that LTR-002 has become fixed as a promoter provider to the ncRNA. Note that we also confirmed that seven selected non-LTR candidates are all multiexon ncRNAs by 3' RACE and CPAT (Supplemental Tables S8, S9).

#### Stratification of HCCs based on the activation level of LTR promoters

We validated the CAGE data by measuring the expression levels of LTR-001, 002, 004, and 007, which are all multiexon ncRNAs, as determined by 3' RACE, using quantitative real-time PCR in the



**Figure 5.** LTR promoters are frequently activated in human HCCs, correlating with viral etiology and aggressiveness. (A) Fold changes of LTR-001, 002, 004, and 007 in tumor compared to nontumor samples measured by qPCR in the working (first) and validation (second) set of HCC samples. (B) Unsupervised classification of HCC samples according to the expression profile of the 43 signature-LTRs using the MeV (MultiExperiment Viewer) software. Three (LTR-high, LTR-intermediate, LTR-low) distinct clusters are clearly visible. The LTR-high cluster included 16 tumor samples. The LTR-low cluster comprised 14 tumor, 50 NT, and five normal samples. (Blue) Normal liver, (green) NT, (black) alcohol, (yellow) HBV, (orange) HCV, (red) all tumors. (C) PCA based on the expression profile of 43 signature-LTRs. (D) Correlation between the indicated clinical parameters and the expression level of the signature-LTRs. Thick green lines: clinical features significantly associated with either the LTR-high or the LTR-low subclass according to the criterion  $FDR < 0.05$ . (E) Gene set enrichment analysis (GSEA) plot of HCCs related to LTR activity. Genes were rank-ordered according to differential expression between LTR-high and LTR-low HCC subclasses. The LTR-high HCC subclass revealed an enrichment of proliferative genes comparable to that of the S2 subclass of Hoshida et al. (2009). The black bars on the x-axis represent positions of gene set members on the rank ordered list of gene set Hoshida\_liver\_cancer\_subclass\_S2\_up. The y-axis represents the enrichment score (ES) of the gene set Hoshida\_liver\_cancer\_subclass\_S2\_up, in the subclass of samples LTR-high HCC. (NES) Normalized Enrichment Score; NES = ES/mean (all ES of other gene sets in MSigDB 4.0). (F) Enrichment plot showing an up-regulation of MYC target genes in the LTR-high subclass.

previous working set of 50 tumor and matched nontumor (T/NT) HCC samples and in a validation set of 21 new T/NT HCC samples. The primer sequences are listed in Supplemental Table S10. Consistent with the CAGE data, the expression of these LTRs was more than 10-fold higher in ~60%–80% of the tumor samples compared to the matched NT in the working set (Fig. 5A;

Supplemental Fig. S5). An up-regulation of the selected LTRs was also observed in a large majority of the tumors in the validation set. Globally, LTR-007 was the most frequently (76%, 54/71) expressed LTR in HCC tumors. These data show that LTRs are recurrently activated in HCC, with different activation levels for different LTRs and different HCCs, indicating that the LTR



activation is not the consequence of genome-wide random events during tumorigenesis. Figure 5B shows an unsupervised classification of the 105 CAGE samples according to the ncRNA expression level of the 43 signature-LTRs. This revealed three (high, moderate, low/inexistent) well-defined clusters of samples. The LTR-high cluster contained only tumor HCC samples (16 in number), while the LTR-low subclass contained all the nontumor and normal-liver samples and 14 tumor samples. Principal component analysis (PCA) showed that LTR-high ncRNA expressions clustered almost exclusively (90%) with viral etiology, the highest expressions occurring mostly in HBV-positive tumors (Fig. 5C). The remaining 10% were related to alcoholism. The tumors of the LTR-low subclass were equally divided between the different etiologies.

A multivariate-analysis based on the set of clinical parameters listed in Figure 5D shows that the HCCs of the LTR-high subclass were globally more severe than those of the LTR-low subclass (Fig. 5D; Supplemental Table S11). The LTR-high HCCs were significantly correlated with age  $\geq 55$  yr, viral etiology (as already mentioned), and a high risk of recurrence. Their differentiation grade ranged from moderate to poor. They, moreover, had a clear tendency to be associated with more satellite nodules, more microvascular invasion, and more distant metastasis. LTR-low HCCs were mostly well-differentiated, and more importantly, they had a much lower risk of recurrence (2/14) than the LTR-high ones (10/16;  $P = 0.011$ ). A correlation with patient outcome was irrelevant to our purpose due to the fact that the majority of the patients with LTR-low HCC received a hepatic graft, improving their outcome compared to patients with LTR-high HCC who benefited from a partial hepatectomy.

Finally, to gain insight on the molecular pathways activated in the different HCC subclasses, we performed a gene set enrichment analysis (GSEA) from the CAGE transcriptome. We found a great enrichment of genes related to cell cycle and replication in the LTR-high subclass compared to the LTR-low one (Fig. 5E,F). Interestingly, the gene enrichment pattern of the LTR-high HCCs was closely similar to that of the S2 molecular subclass defined by Hoshida et al. (2009), which is characterized by proliferation as well as *MYC* and *AKT1* activation. These results suggest that the LTR expression level be taken as a new characteristic in the definition of HCC molecular subclasses, corresponding to an as-yet unsuspected layer of complexity in the genomics of HCC.

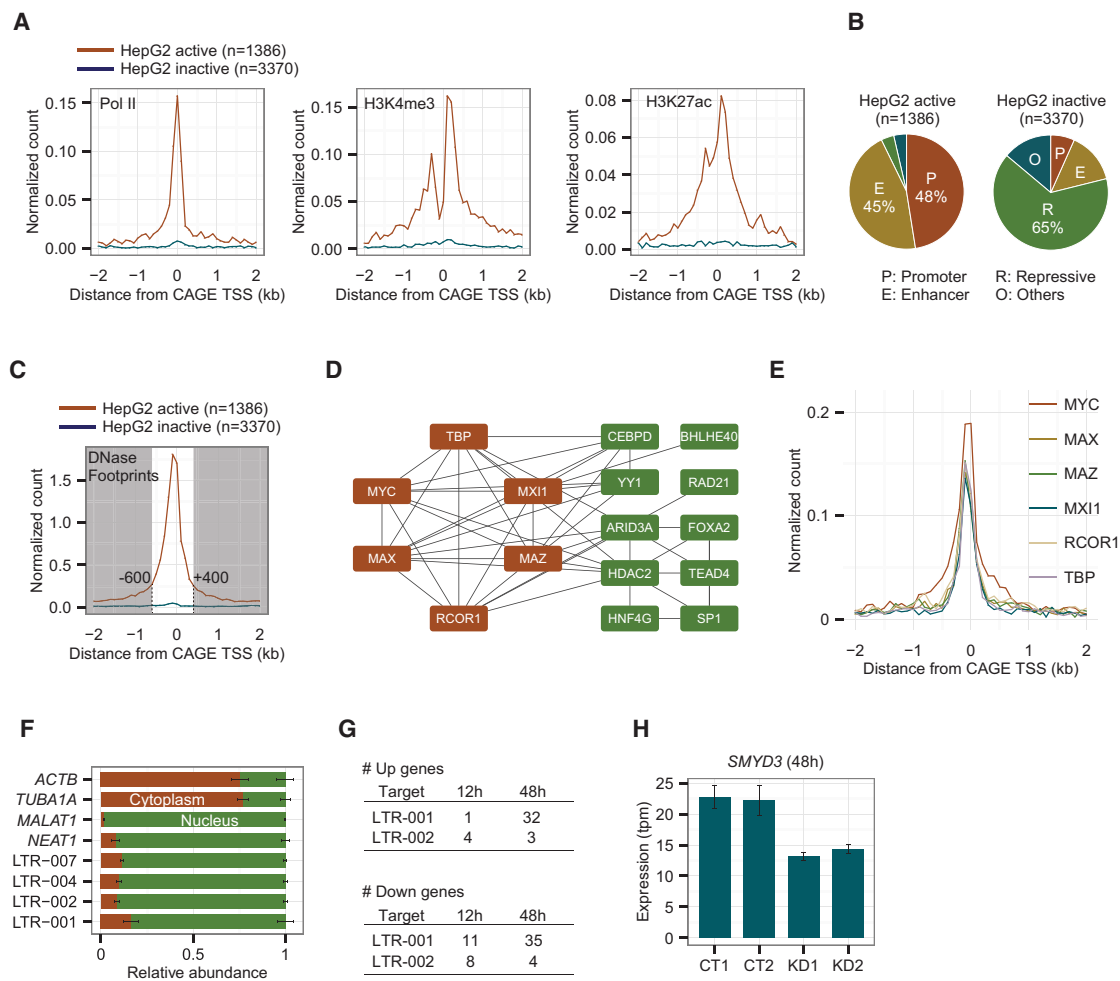
### Distal ncRNAs activated by cobinding transcription factors

To cast light on the mechanism of up-regulation of ncRNAs, we combined our data for human HCCs with the large set of data for HepG2 cells provided by ENCODE. We first examined whether the up-regulated distal peaks that we found in human HCCs (see Fig. 2A) were also active in HepG2 cells. Of the 4756 up-regulated distal peaks in human HCCs, 1386 had both transcripts and open chromatin signals in HepG2 cells. This observation is further validated by independent ChIP-seq data on RNA polymerase II (Pol II) and active histone marks (H3K4me3 and H3K27ac). Figure 6A shows that these active signals were strongly enriched around HepG2 active peaks, whereas none was enriched around HepG2 inactive peaks. Furthermore, the former loci were mostly marked as promoters or enhancers, while the latter were marked as repressive states according to chromatin state annotation based on combinatorial patterns of chromatin marks (Fig. 6B; Ernst et al. 2011). We thus conclude that 1386 distal peaks were commonly active both in HCC tumor tissues and the HepG2 cell line. Subsequently, we used only these active peaks for further analyses. The list of

ENCODE data (CAGE, DNase-seq, FAIRE-seq, ChIP-seq, and chromatin states) used to assess the activities of the distal peaks in HepG2 cells is available in Supplemental Table S12.

The analysis of the promoter region around the 1386 distal peaks active in HepG2 cells revealed that distal ncRNAs carry multiple binding sites of various transcription factors in distinct positions. The DNase I footprint represents direct binding loci of regulatory elements including transcription factors. DNase I footprint data showed that the region spanning from  $-600$  to  $+400$  bp (local regulatory region) around the active peaks was highly enriched in transcription factor binding sites, with five or more footprints at more than half of the peaks (Fig. 6C). We then searched for the transcription factors that contributed to up-regulation of ncRNAs using the ENCODE ChIP-seq data of HepG2 for 51 transcription factors (Supplemental Table S13). Interestingly, *MYC*, an oncogene which is widely activated in many cancers, bound to  $\sim 45\%$  (635/1386) of the active peaks, while it bound to a significantly smaller proportion of the non-up-regulated peaks (one-sided Fisher's exact test with Bonferroni correction  $< 0.05$ ). In total, 19 transcription factors were significantly associated with up-regulated peaks, including TBP (TATA box binding protein), which had the lowest *P*-value, and ARID3A (AT-rich interactive domain-containing protein 3), which had the highest binding counts (Supplemental Fig. S6A). We then investigated which pairs of TFs shared common targets using the Jaccard index. More specifically, we generated a cobinding network, in which each node represents a transcription factor and each edge represents a cobinding pair, whose Jaccard Index is above 0.5 (Fig. 6D). Six transcription factors in the network are fully connected to one another, and four of them include *MYC*, the *MYC*-associated protein *MAZ*, the *MYC* dimerization partner *MAX*, and the alternative *MAX* partner *MXI1*. Moreover, all six transcription factors show binding immediately upstream of TSSs (between  $-100$  and  $0$  bp), suggesting an active role for these transcription factor complexes in the deregulation of ncRNAs in HCC (Fig. 6E). In addition, we performed similar analyses for LTR-derived distal peaks, which is a subset of the 1386 peaks analyzed above, and observed similar enrichment patterns of RNA Pol II, active histone marks, DNase footprints, and six transcription factors (Supplemental Fig. S6B,C).

We finally performed knockdown experiments to investigate whether LTR-derived ncRNAs have any functional role in the regulatory network of HepG2 cells. Because recurrently activated LTR-derived ncRNAs are enriched in the nuclear fraction of HepG2 cells (Fig. 6F), we designed locked nucleic acid (LNA) oligonucleotides (Petersen and Wengel 2003) for the top two LTR-derived ncRNAs. LNAs were designed on the common exon of all isoforms. LTR-001 and LTR-002 were stably suppressed by two different LNAs per target (Supplemental Table S14) at 12, 24, 48, and 72 h after transfection into HepG2 cells (Supplemental Fig. S6D). We selected the 12- and 48-h time points to study early responses to the depletion of LTR-001 and LTR-002 by CAGE analysis (Supplemental Fig. S6E). We confirmed that the transcriptome of cells treated by two different LNAs were highly correlated (Supplemental Fig. S6F). A total of 98 differentially expressed genes (FDR  $< 0.05$ ) were identified by comparison between knockdown and control samples using edgeR (Fig. 6G). Supplemental Tables S15 and S16 display the full list of the dysregulated genes for LTR-001 and LTR-002 at 48 h. Among the down-regulated genes with LTR-001 stands *SMYD3*, which encodes a lysine methyltransferase involved in the methylation of MAP3K2, increasing MAP kinase signaling and promoting the formation of Ras-driven carcinomas (Hamamoto et al. 2004; Mazur et al. 2014). We found



**Figure 6.** A subset of up-regulated distal ncRNAs including LTRs and non-LTRs has active histone marks and TF binding sites in the HepG2 cell line. (A) Enrichment of ChIP-seq peaks for RNA Pol II, H3K4me3, and H3K27ac within 2 kb from the center of the CAGE peaks in 100-bp bins. The red and blue lines represent active and inactive peaks in the HepG2 cell line, respectively. (B) Chromatin states annotated by ENCODE based on combinatorial patterns of chromatin marks for the HepG2 active and inactive loci. (C) Distribution of DNase I footprints within 2 kb from the center of the CAGE peaks in 100 bp bins. (D) The co-occurrence of transcription factors at up-regulated distal ncRNA loci. Nodes are connected by a solid line if their Jaccard Index is above 0.5. Nodes shown in red are fully connected to one another. (E) Enrichment of ChIP-seq peaks for the fully connected six TFs within 2 kb from the center of the CAGE peaks in 100-bp bins. (F) Relative abundance of four LTR ncRNAs in nucleus and cytoplasm with *ACTB*, *TUBA1A*, *MALAT1*, and *NEAT1* as cytoplasmic and nuclear controls. (G) Numbers of significantly up-regulated genes after 12 and 48 h of knocking down LTR-001 and 002. (H) Expression levels of *SMYD3* measured by CAGE in CT (control) and KD (knockdown) samples at 48 h of the ASO treatment.

that the expression level of *SMYD3* was unchanged at 12 h but significantly lowered at 48 h (Fig. 6H). However, the protein level of *SMYD3*, expression levels of MAP kinase target genes, and cell proliferation were not significantly changed, suggesting that a knockdown of LTR-001 is not enough to disturb the downstream pathways of *SMYD3*. Among the up-regulated genes, *IGFBP1* (insulin-like growth factor binding protein 1) has important inhibitory role in the development and/or growth of HCC, and *RBBP6* (retinoblastoma binding protein) acts as a cell-cycle regulator.

## Discussion

A growing number of ncRNAs are being considered as regulators of various cellular processes (Mercer and Mattick 2013), whereas relatively few ncRNAs are reported to be associated with hepatocarcinogenesis (Braconi et al. 2011; Spizzo et al. 2012). We have estab-

lished a comprehensive map of promoters for ncRNAs actively transcribed in a large set of human HCCs as well as in the *Mdr2* KO mouse model of HCC using CAGE. This unique resource provides a mammalian promoter map for liver cancer at single-nucleotide resolution, which enables us to connect promoter activities with surrounding DNA regulatory elements, including LTR retroviral promoters and transcription factor binding sites. The comparison of the expression levels between tumors and matched nontumor tissues revealed that more than 14,000 promoters are significantly up-regulated in HCC tumors, one-third of which are at least 5 kb away from the promoters of known protein coding genes. Remarkably, 20% (935/4756) of the up-regulated distal peaks are LTR retroviral promoters, which are mostly silenced in normal hepatocytes and liver tissues. Nevertheless, a subset of the LTRs is strongly expressed in reproduction tissues, such as testis and placenta, indicating a highly tissue-specific regulation of the LTRs. The mechanisms of regulation of the LTRs was previously

investigated in iPS and ES cells, which express high levels of specific LTR families, namely, LTR7 and HERVH. An appropriate activation of these LTRs was shown to be essential for iPS reprogramming (Lu et al. 2014; Ohnuki et al. 2014). Depleting LTRs in ES cells led to cell differentiation (Fort et al. 2014; Lu et al. 2014), whereas augmenting LTRs impaired cell differentiation (Ohnuki et al. 2014). Concordantly, we found that HCCs can be stratified according to LTR expression level and that LTR-high HCCs are less well-differentiated than LTR-low ones, confirming that an excess of LTR activity impairs cell differentiation. It is interesting that nonorthologous LTR subfamilies are activated in both human and mouse HCCs. The same type of unexpected concordance was previously observed in stem cell-associated LTRs (Fort et al. 2014). Whether or not this should be attributed to convergent evolution of functionally similar features will remain uncertain as long as LTR-related functional pathways are not determined.

In this study, we used tags uniquely mapped to the genomes. This strategy might miss transcripts derived from young ERV elements that accumulate few mutations. To assess the potential impact of multimap tags that mapped to multiple genomic loci (mapping quality < 10), we have performed analyses of repetitive elements at the family level. We measured expression values for each repeat family defined by RepeatMasker in the human and mouse genomes based on all tags, including multimap tags, and compared tumors with nontumors using edgeR. We confirmed that various LTR families are significantly up-regulated in tumors, whereas most young ERV elements, in particular, human ERVK (LTR5\_Hs), SVA, *Alu*, and mouse IAP elements, are not up-regulated or are even down-regulated in tumors (Supplemental Figs. S2O, S3E). This shows that young repetitive elements are not largely activated in liver cancer.

This study provides evidence supporting the idea that activation of LTR promoters might contribute to liver carcinogenesis. First, a subset of 43 LTRs was up-regulated (more than eightfold) in the vast majority of the HCCs studied. For example, LTR-007, which was totally suppressed in normal liver and most other healthy tissues, was more than 10-fold up-regulated in ~80% of the HCC tumors. This high frequency of LTR-007 activation is comparable to the frequencies of the *GPC3* activation (Jia et al. 2007) and *TP53* mutations in HCC (Fujimoto et al. 2012). In addition, LTR-007 is up-regulated more than 1000-fold in 15 patients from the initial data set and seven patients from the second data set. Up-regulation rates of this magnitude are likely due to promoter activations rather than copy-number increase. Second, LTR-high tumors showed a higher risk of recurrence and, more generally, were more aggressive than LTR-low ones. The LTR-high subclass was closely similar to the well-recognized S2 molecular subclass of HCCs (Hoshida et al. 2009), suggesting taking into account the LTR promoter activity in the taxonomy of HCCs. Third, the LTRs found in HCC tumors were expressed in several cancer cell lines. Fourth, LTR up-regulation appeared from the preneoplastic stage in the *Mdr2* KO mouse model of HCC. This study uncovers a new layer of complexity in the transcriptome of HCC, which highlights the importance of understanding LTR-derived ncRNAs in cancer. The next stage would be the functional characterization of the ncRNAs, which requires a series of experiments including overexpression, knockdown, and/or knockout. The transcriptional responses of cell lines and HCC tissues are different, and the best ncRNA candidates found in HCC tissues might not be testable using cell lines, suggesting that a much larger number would be needed to find ncRNAs that are equally significant in function.

## Methods

### CAGE libraries for human liver tissues

Liver tissues were collected from 50 patients resected for HCC and five patients resected for metastatic liver colon cancer. The latter were used as control samples. The ethics evaluation committee of the Inserm (IRB00003888, FWA00005831) and the ethical review committee of RIKEN (H24-4) approved the use of human liver samples in this study. Informed consent was obtained from each subject. Total RNAs were extracted from the tissues using a mirVana kit (Ambion). The quality of the RNA samples was assessed using an Agilent RNA 6000 Nano kit (RIN scores provided in Supplemental Table S1). We prepared CAGE libraries following the protocol described in detail in Takahashi et al. (2012). We used 5 µg of total RNA to synthesize cDNA with random primers. Full-length cDNAs were biotinylated and captured by streptavidin-coated magnetic beads. The cDNAs were released from cap-trapped RNAs, ligated to 5' linkers including barcode sequences, and digested with EcoP15I. CAGE libraries were sequenced with single-end reads of 50 bp on the Illumina HiSeq 2000 platform.

### Determination of CAGE TSSs

Multiplexed sequencing reads were split by barcode sequences. Reads with ambiguous bases "N" were removed, and linker sequences were trimmed from the 3' end. Artificial sequences and ribosomal RNAs were identified and removed using TagDust (Lassmann et al. 2009) and RNAdust (<http://compbio.gsc.riken.jp/rmadust.html>). We then aligned the extracted CAGE tags to the human genome (hg19/GRCh37 assembly) using BWA v0.5.9 (Li and Durbin 2009) with default parameters on the MOIRAI pipeline platform (Hasegawa et al. 2014). The uniquely mapped CAGE tags with a minimal mapping quality 10 were used in this study. Because the most 5' position of a CAGE tag represents a TSS, the 5' coordinates were extracted from the tags and generated a genome-wide TSS map at single-nucleotide resolution. We then clustered the tags to define distinct CAGE peaks using Paraclu with the following parameters: (1) a minimum of 10 tags per cluster; (2) a minimum density increase of 2; and (3) a maximum cluster length of 500 bp. Raw tag counts for each peak were divided by a total tag count of the library to calculate normalized expression values. The unit of the expression value is TPM, tags per uniquely mapped million tags. CAGE tags and peaks were visualized using ZENBU (Severin et al. 2014).

### Identification of differentially expressed peaks

The differentially expressed peaks were identified using edgeR version 2.6.3 (Robinson et al. 2010) in a Bioconductor package in R. Briefly, the dispersion of expression values was estimated by the quantile-adjusted conditional maximum likelihood (qCML) method using estimateCommonDisp and estimateTagwiseDisp functions with default parameters. Exact *P*-values and false discovery rates were calculated using exactTest and topTags functions.

### *Mdr2* KO mouse and CAGE libraries

Animal experiments have been performed in agreement with the Italian Laws (D.L.vo 116/92) and the guidelines of the European Commission Recommendation 2007/526/EC-June 18, 2007. The project has been reported to the Italian Ministry of Health (project n. 106/11). Founders of the FVB.129P2-Abcb4tm1Bor/J (*Mdr2* KO, stock number: 002539) and FVB/NJ (*Mdr2* WT, stock number: 001800) mice were purchased from the Jackson Laboratory. Mouse colonies were maintained in a specific pathogen-free animal facility. Tumor growth in *Mdr2* KO livers is multinodular,

and HCC tends to develop within an adenoma (as a focus of tumor progression). The histological composition of grossly detectable hepatic nodules was semiquantitatively determined based on reported classification criteria (Thoolen et al. 2010). In order to analyze sample sets based on histological content, we classified samples as “adenoma” when scoring a variable content in adenomatous tissue without any detectable presence of carcinoma, “low-grade HCC” with a histologic content of carcinoma  $\leq 50\%$ , and “high-grade HCC” when the carcinoma content was higher than 50%. The histological composition of the liver tissues was determined based on the criteria reported in the literature (Katzenellenbogen et al. 2007).

We prepared and sequenced CAGE libraries from control and tumor tissues as described above. The sequences were aligned to the mouse genome (mm9/NCBI37 assembly) using BWA v0.5.9 (Li and Durbin 2009) with default parameters. The uniquely mapped CAGE tags with a minimal mapping quality 10 were used for the downstream analysis. CAGE peaks were detected using the Paraclu code.

### Comparative analysis with FANTOM5 data

We calculated expression values of the 43 LTRs in various cell types using FANTOM5 expression data (The FANTOM Consortium et al. 2014). More specifically, the number of tags within a CAGE peak was counted and normalized by the library size with mapping quality  $\geq 20$  using the files in the *ctss.bed* format provided in the FANTOM5 web site (<http://fantom.gsc.riken.jp/5/datafiles/latest/basic/>). This process was done for the CAGE libraries of 135 tissues, 432 primary cells, and 241 cell lines (the full sample list is available in Supplemental Table 1 of the FANTOM5 paper [The FANTOM Consortium et al. 2014]). To see which cell types express the LTRs, expression values of the LTRs in the FANTOM5 samples were compared to the expression values of the tumors. The number of samples that express the LTRs with more than the median values of 50 liver tumors was counted for tissue, primary cell, and cell line groups. We also counted the number of LTRs that express in each FANTOM5 sample.

### Sample clustering based on the LTR expression

The graphical representation of CAGE LTR expression profile and unsupervised classification of HCC samples were performed with the MeV (MultiExperiment Viewer) software, version 4.9.0 using the Euclidean distance metric with average linkage algorithm during classification (Saeed et al. 2003). The obtained cluster allowed discrimination of two classes of tumor samples: tumors with low LTR expression and tumors with high LTR expression.

### Multivariate analysis on clinical biological parameters

The detection of LTR level expression allowed us to distinguish two groups of tumor samples by CAGE analysis (high level and low level of LTR discriminated by unsupervised clustering). The odds ratio of each clinical and biological criterion was calculated for each class of patients (high-LTR and low-LTR). Meta-analysis performed with the totality of clinical and biological criteria taken into account during the study allowed testing of the heterogeneity of the calculated odds ratios. A forest plot was drawn with odds ratios and interval of confidence at 95% for each criterion. Significance of the meta-analysis was retained if the *P*-value of the test for heterogeneity  $< 0.05$  and if the random effects on the forest plot are well centered.

### Comparative analysis with ENCODE HepG2 data

We used CAGE, DNase-seq, and FAIRE-seq data of the HepG2 cell line, downloaded from the ENCODE website (<http://genome.ucsc.edu/ENCODE/>) to determine which tumor TSSs are active in HepG2. First, all CAGE tags produced for different compartments (Supplemental Table S12, No.1–9) were combined. Secondly, narrow peaks, defined by the ENCODE pipeline, of DNase-seq rep1, rep2, and FAIRE-seq (Supplemental Table S12, No.10–12) were combined. Finally, tumor TSSs were determined to be active in HepG2 if at least one CAGE tag was found in the TSS and at least one narrow peak of the open chromatin signals was found within 1 kb of the TSS.

We used chromatin states systematically annotated based on combinatorial patterns of chromatin marks (Supplemental Table S12, No.16; Ernst et al. 2011) to examine epigenetic differences in active and inactive TSS loci. The original 15 categories were simplified into four categories, in which active, weak, and poised promoters were merged as “Promoter,” strong and weak enhancers were merged as “Enhancer,” repressed and heterochromatin were merged as “Repressive,” and all the others were merged as “Others.” Each TSS was annotated to one of these categories. When a TSS is overlapped with multiple segments, the annotation was prioritized in the order of Promoter, Enhancer, and Repressive.

We used ChIP-seq peaks, defined by the ENCODE pipeline, for 51 transcription factors (Supplemental Table S13) to test associations between TF bindings and up-regulations. For each TF, a  $2 \times 2$  contingency table was generated in which one factor is whether or not distal TSSs are significantly up-regulated and the other factor is whether the TSSs have a binding site of the TF (mid-point of the peak) in a range from  $-600$  to  $+400$  bp. A *P*-value was calculated for each TF using a one-sided Fisher’s exact test and adjusted for multiple testing using Bonferroni’s correction. A Jaccard index was calculated for all pairs of significantly associated TFs.

We used ChIP-seq peaks, defined by the ENCODE pipeline, for 51 transcription factors (Supplemental Table S13) to test associations between TF bindings and up-regulations. For each TF, a  $2 \times 2$  contingency table was generated in which one factor is whether or not distal TSSs are significantly up-regulated and the other factor is whether the TSSs have a binding site of the TF (mid-point of the peak) in a range from  $-600$  to  $+400$  bp. A *P*-value was calculated for each TF using a one-sided Fisher’s exact test and adjusted for multiple testing using Bonferroni’s correction. A Jaccard index was calculated for all pairs of significantly associated TFs.

### LNA knockdown

HepG2 cells were maintained in Dulbecco’s modified Eagle’s medium, 10% FBS, and 100 units/mL penicillin streptomycin (Gibco) at 37°C, 5% CO<sub>2</sub>. For the qPCR experiment, 80,000 cells were seeded in 12-well plates, and for the CAGE experiment, 160,000 cells were used in 6-well plates. Transfection was done according to the Lipofectamine RNAiMAX protocol. We used LNA longRNA GapmeRs by Exiqon, which are antisense oligonucleotides (ASOs) containing a central stretch of DNA monomers flanked by LNA blocks. Target RNAs are cleaved by RNase H, activated by ASOs. The ASOs were transfected into HepG2 cells at a concentration of 20 nM. Lipofectamine alone and scramble ASOs were used as controls. Knockdown efficiency was measured by qPCR with specific primer sets (Supplemental Table S10) for biological triplicates at 12, 24, 48, and 72 h after the transfection. *GAPDH* was used for normalization.

### Data access

CAGE data from this study have been submitted to the NCBI database of Genotypes and Phenotypes (dbGaP; <http://www.ncbi.nlm.nih.gov/gap/>) under accession number phs000885.v1.p1 for human HCC and the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE60982 for mouse HCC.

### Acknowledgments

This work was supported by the European Union’s Seventh Framework Programme (FP7) under grant agreement No. 259743 (MODHEP consortium) to the INSERM (D.S., M.A.B., J.F.), the RIKEN Center (K.H., A.M.S., P.C.), and IEO (A.C., S.G., G.N.), as well as a Research Grant from the Japanese Ministry of

Education, Culture, Sports, Science and Technology (MEXT) to the RIKEN Center for Life Science Technologies. The authors thank RIKEN GenAS for the sequencing of the libraries. We also thank Professor René Adam, Professor Daniel Azoulay, Professor Daniel Cherqui, Professor Antonio Sa Cunha, Professor Catherine Guettier, and Dr. Eric Vibert (Centre Hépatobiliaire, Villejuif), and the Biological Resources Centres (AP-HP, Université Paris-Sud, National Liver cancer biobanks network) for providing HCC specimens. J.F. acknowledges the support of the French National Institute of Cancer (INCa PAIR-HCC).

## References

- Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**: 1752–1762.
- Boyault S, Rickman DS, de Reynies A, Balabaud C, Rebouissou S, Jeannot E, Hérault A, Saric J, Belghiti J, Franco D, et al. 2007. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**: 42–52.
- Braconi C, Valeri N, Kogure T, Gasparini P, Huang N, Nuovo GJ, Terracciano L, Croce CM, Patel T. 2011. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci* **108**: 786–791.
- Capurro M, Wanless IR, Sherman M, Deboer G, Shi W, Miyoshi E, Filmus J. 2003. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* **125**: 89–97.
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309**: 1559–1563.
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149–1154.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101–108.
- Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–49.
- The FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Lassmann T, Itoh M, Summers KM, Suzuki H, et al. 2014. A promoter-level mammalian expression atlas. *Nature* **507**: 462–470.
- Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41**: 563–571.
- Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, Bonetti A, Voineagu I, Bertin N, Kratz A, et al. 2014. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* **46**: 558–566.
- Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. 2008. A code for transcription initiation in mammalian genomes. *Genome Res* **18**: 1–12.
- Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, et al. 2012. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* **44**: 760–764.
- Gao W, Kim H, Feng M, Phung Y, Xavier CP, Rubin JS, Ho M. 2014. Inactivation of Wnt signaling by a human antibody that recognizes the heparan sulfate chains of glypican-3 for liver cancer therapy. *Hepatology* **60**: 576–587.
- Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. 2004. *SMYD3* encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* **6**: 731–740.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760–1774.
- Hasegawa A, Daub C, Carninci P, Hayashizaki Y, Lassmann T. 2014. MOIRAI: a compact workflow system for CAGE analysis. *BMC Bioinformatics* **15**: 144.
- Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, et al. 2009. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* **69**: 7385–7392.
- Huang Q, Lin B, Liu H, Ma X, Mo F, Yu W, Li L, Li H, Tian T, Wu D, et al. 2011. RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. *PLoS One* **6**: e26168.
- Iannelli F, Collino A, Sinha S, Radaelli E, Nicoli P, D'Antiga L, Sonzogni A, Faivre J, Buendia MA, Sturm E, et al. 2014. Massive gene amplification drives paediatric hepatocellular carcinoma caused by bile salt export pump deficiency. *Nat Commun* **5**: 3850.
- Jacques PE, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9**: e1003504.
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics. *CA Cancer J Clin* **61**: 69–90.
- Jia HL, Ye QH, Qin LX, Budhu A, Forgues M, Chen Y, Liu YK, Sun HC, Wang L, Lu HZ, et al. 2007. Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma. *Clin Cancer Res* **13**: 1133–1139.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**: 462–467.
- Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* **9**: 411–412.
- Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916–919.
- Kapranov P, Cheng J, Dike S, Nix DA, Duttgupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**: 1484–1488.
- Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564–1566.
- Katzenellenbogen M, Mizrahi L, Pappo O, Klopstock N, Olam D, Jacob-Hirsch J, Amariglio N, Rechavi G, Domany E, Galun E, et al. 2007. Molecular mechanisms of liver carcinogenesis in the *mdr2*-knockout mice. *Mol Cancer Res* **5**: 1159–1170.
- Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* **13**: R107.
- Kovalskaya E, Buzdin A, Gogvadze E, Vinogradova T, Sverdlov E. 2006. Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions. *Virology* **346**: 373–378.
- Ladeiro Y, Couchy G, Balabaud C, Bioulac-Sage P, Pelletier L, Rebouissou S, Zucman-Rossi J. 2008. MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations. *Hepatology* **47**: 1955–1963.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Lassmann T, Hayashizaki Y, Daub CO. 2009. TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics* **25**: 2839–2840.
- Lee E, Iskov R, Yang L, Gokcumen O, Haseley P, Luquette LJ III, Lohr JG, Harris CC, Ding L, Wilson RK, et al. 2012. Landscape of somatic retrotransposition in human cancers. *Science* **337**: 967–971.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21**: 423–425.
- Mazur PK, Reynoird N, Khatri P, Jansen PW, Wilkinson AW, Liu S, Barbash O, Van Aller GS, Huddleston M, Dhanak D, et al. 2014. *SMYD3* links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**: 283–287.
- Mercer TR, Mattick JS. 2013. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* **20**: 300–307.
- Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, Nakamura M, Tokunaga Y, Nakamura M, Watanabe A, et al. 2014. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci* **111**: 12426–12431.
- Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Petersen M, Wengel J. 2003. LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol* **21**: 74–81.
- Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, Urieli-Shoval S, Galun E, Ben-Neriah Y. 2004. NF- $\kappa$ B functions

- as a tumour promoter in inflammation-associated cancer. *Nature* **431**: 461–466.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, et al. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374–378.
- Santoni FA, Guerra J, Luban J. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9**: 111.
- Severin J, Lizio M, Harshbarger J, Kawaji H, Daub CO, Hayashizaki Y, The FANTOM Consortium, Bertin N, Forrest AR. 2014. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat Biotechnol* **32**: 217–219.
- Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, et al. 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**: 101–111.
- Spizzo R, Almeida MI, Colombatti A, Calin GA. 2012. Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* **31**: 4577–4587.
- St Laurent G, Shtokalo D, Dong B, Tackett MR, Fan X, Lazorthes S, Nicolas E, Sang N, Triche TJ, McCaffrey TA, et al. 2013. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol* **14**: R73.
- Takahashi H, Lassmann T, Murata M, Carninci P. 2012. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc* **7**: 542–561.
- Thoolen B, Maronpot RR, Harada T, Nyska A, Rousseaux C, Nolte T, Malarkey DE, Kaufmann W, Kuttler K, Deschl U, et al. 2010. Proliferative and nonproliferative lesions of the rat and mouse hepatobiliary system. *Toxicol Pathol* **38**: 5S–81S.
- Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, et al. 2011. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43**: 464–469.
- Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**: e74.
- Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**: 836–841.

Received February 14, 2015; accepted in revised form September 30, 2015.

## Supplementary materials

### Supplementary Figures

Fig. S1: Characterization of coding, proximal, and distal CAGE peaks

Fig. S2: Activation of LTR retroviral promoters in human HCC tumors

Fig. S3: Activation of LTR retroviral promoters in mouse HCC tumors

Fig. S4: Expression profiles of 43 LTR-derived ncRNAs

Fig. S5: qPCR validation of LTR-derived ncRNAs

Fig. S6: Activity of the ncRNAs in HepG2 cells

### Supplementary Tables

Table S1: HCC samples with etiology and map counts

Table S2: Fantom5 data used to measure expression in normal livers

Table S3: A list of up-regulated LTR12C in human HCC

Table S4: mouse liver samples with histology and map counts

Table S5: A list of up-regulated LTR in mouse HCC

Table S6: The most significantly up-regulated distal peaks (43 LTRs and 90 non-LTRs)

Table S7: Targets of 3' RACE validation (15 LTRs and 7 non-LTRs)

Table S8: Full-length ncRNAs determined by 3' RACE

Table S9: Prediction of coding potential using CPAT web system

Table S10: qPCR primers for LTR-derived ncRNAs

Table S11: Clinical features associated with the high LTR class

Table S12: ENCODE data used to measure activities of TSSs in HepG2

Table S13: ENCODE ChIP-Seq data used to measure TF activities in HepG2

Table S14: Targets and sequences of LNAs

Table S15: Significantly dysregulated genes with LTR-001 KD at 48h

Table S16: Significantly dysregulated genes with LTR-002 KD at 48h

## Supplementary Data Sets

### Human Data

[http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000885.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000885.v1.p1)

#### Expression table

File name	: GSE60982_Expression.hg19.HCC.64366peaks.txt.gz
Columns 1-4	: Genomic position of TSS (hg19)
Columns 5	: total read counts of all samples
Columns 6	: strand (+ or -)
Columns 7	: classification (Coding, Proximal, or Distal)
Columns 8-9	: Log2 fold change and FDR (50 tumors vs 50 NTs)
Columns 10-11	: Gene names and IDs
Columns 12-61	: Normalized expression values (tpm) for 50 tumors
Columns 62-111	: Normalized expression values (tpm) for 50 NTs
Columns 112-116	: Normalized expression values (tpm) for 5 normal livers

### Mouse Data

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60982>

#### Expression table

File name	: GSE60982_Expression.mm9.HCC.49096peaks.txt.gz
Columns 1-4	: Genomic position of TSS (mm9)
Columns 5	: total read counts of all samples
Columns 6	: strand (+ or -)
Columns 7	: classification (Coding, Proximal, or Distal)
Columns 8-9	: Log2 fold change and FDR (3 WTs vs 4 high-grade HCC)
Columns 10-11	: Gene names and IDs
Columns 12-17	: Normalized expression values (tpm) for 6 WT
Columns 18-20	: Normalized expression values (tpm) for 3 inflammation
Columns 21-24	: Normalized expression values (tpm) for 4 adenoma
Columns 25-40	: Normalized expression values (tpm) for 16 low-grade HCC
Columns 41-48	: Normalized expression values (tpm) for 8 high-grade HCC



### Example of data access and file conversion

```
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX%2FSRX690%2FSRX690748/SRR1562475/SRR1562475.sra
```

```
sam-dump SRR1562475.sra | samtools view -bhS - > SRR1562475.bam
```

```
samtools index SRR1562475.bam
```

## Supplementary Methods

### *Generation of random dataset to analyze human LTR activities*

The significantly up-regulated peaks (genomic positions) were randomized using the shuffleBed function of the Bedtools v2.17.0 with the '-seed' option 1, 2, and 3 for dataset 1, 2, and 3 (Quinlan and Hall 2010). Gapped regions including centromere and telomere filled with 'N' in the human genome were excluded using the '-excl' option. The generated random peaks were classified into coding, proximal, and distal groups in the same way described in the Results section. The coordinates of gapped regions were downloaded from the UCSC table browser. We then counted the number of distal peaks that overlap with LTRs in sense and antisense directions (Fig. S2E) and that overlap with LTR subfamilies in the sense direction (Fig. S2J).

### *Rapid amplification of 3' cDNA ends (3' RACE)*

We manually selected CAGE peaks derived from LTRs for 3' RACE. RNA samples from tumor and non-tumor samples (149T and 149NT respectively) were fractionated into polyA+ and polyA- pools with MicroPurist kit (Ambion). PolyA+ RNA fractions were reverse transcribed with PrimeScript (Takara) and random primers. For each target two subsequent

RACE PCRs were performed using Ex Taq (Takara), gene-specific forward primers and oligo dT reverse primers. Libraries of the RACE products were prepared using Nextera XT DNA Sample Preparation Kit (Illumina) and paired-end sequenced on Miseq platform (Illumina). The sequencing reads were assembled by Trinity with default parameters (Grabherr et al. 2011), and the assembled sequences were mapped to the human genome (hg19/GRCh37) using blat (version 34) (Kent 2002). A missing 5' sequence from the CAGE peak to a primer position was added to the assembled sequence. Coding potential was predicted for the obtained full-length sequences using the CPAT web server with default parameters (Wang et al. 2013).

#### *Quantitative PCR for LTRs*

Total RNA was extracted from tumor and non-tumor samples of 71 human HCCs. cDNA was synthesized using SuperScript II reverse transcriptase with random primers. Expressions levels of LTR-001, 002, 004, and 007 were measured by qPCR using target-specific primer sets and normalized to HMBS (Table S10).

#### *Gene Set Enrichment Analysis*

Gene set enrichment analysis was performed with GSEA 2.013 from broad institute and MsigDb 4.0 (Subramanian et al. 2005). Molecules known to be target of MYC were used to build a custom gene set compatible with GSEA software from Broad Institute: MYC targets were curated from publications by Cairo and al (Cairo et al. 2008). Genes targets of MYC enriched by GSEA analysis were implemented into Cytoscape version 3.0.1 to build a network (Shannon et al. 2003).

#### *CAGE for LNA Knockdown samples*

We prepared CAGE libraries for biological duplicates of Lipofectamine alone, scramble LNAs, two different LNAs targeting LTR-001, and two LNAs targeting LTR-002 at 12h and 48h after the transfection. A total of 24 CAGE libraries were sequenced with single-end reads of 50 bp on the Illumina HiSeq 2000 platform. Extraction of CAGE tags, mapping to the human genome, and TSS peak detections were done as described above. We compared expression levels between 4 knockdown samples using two different LNAs as a case group and 4 Lipofectamine with/without scramble LNAs as a control group with edgeR version 2.6.3. CAGE peaks that had bias to a specific barcode sequence were detected and removed by edgeR (FDR<1.0E-5).

## References

- Cairo S, Armengol C, De Reynies A, Wei Y, Thomas E, Renard CA, Goga A, Balakrishnan A, Semeraro M, Gresh L et al. 2008. Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer cell* **14**: 471-484.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**: 644-652.
- Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome research* **12**: 656-664.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**: 2498-2504.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15545-15550.
- Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic acids research* **41**: e74.

## Supplementary Figure legends

**Figure S1. Characterization of coding, proximal, and distal CAGE peaks.** **A.** Distribution of distances between proximal peaks to their closest coding peaks. Sense indicates that coding and proximal peaks are the same strands and antisense indicates that they are the opposite strands. **B.**

Spearman's correlations for proximal vs. coding peaks and distal vs. coding peaks. **C.** Robustness of coding, proximal, and distal CAGE peaks. Graphs show how many tumor samples support CAGE peaks. A great majority of peaks in all groups are supported by more than half samples. **D.** Expression pattern of proximal peaks across human body. (see also Fig. 1E for coding and distal peaks)

**Figure S2. Activation of LTR retroviral promoters in human HCC tumors.** **A.** Expression levels of *GPC3*, the most significantly up-regulated coding gene in tumors. As reported by previous papers, *GPC3* is largely up-regulated in HCC. **B.** Fraction of peaks in repeats in the sense direction. **C.** Fraction of peaks in repeats in the anti-sense direction. **D.** Fraction of non-up-regulated peaks in repeats repetitive elements in the anti-sense direction. **E.** Fraction of randomized peaks in repeats in the sense direction. R1, 2, and 3 represent three sets of randomized peaks. **F.** Fraction of peaks in repeats excluding ERV internal domains (“-int” elements). **G.** Fraction of peaks in repeats using tags with mapping quality 30 or above. **H.** Fraction of peaks in repeats using the Gencode ver.19 (the latest for the hg19 genome) for the promoter classification instead of Gencode ver.12. **I.** Expression levels of non-LTR distal peaks in tumor, non-tumor (NT), and normal liver tissues. 'Liver' (normal adult and fetus liver tissues)

and 'Hepato' (three biological replicates of primary hepatocytes) data from the Fantom 5 Project are shown for comparison. To avoid log of zero, 0.005 is added to the tpm values. The median values of 'Normal', 'Liver', and 'Hepato' are 0.005, indicating no expression. **J.** Random overlap counts versus copy numbers for LTR subfamilies in the human genome. One dot represents one LTR subfamily. **K.** Sequence patterns of promoters in LTR12C elements. The upper panel shows a sequence logo generated by WebLogo based on the multiple alignment of significantly up-regulated LTR12C sequences. The lower panel shows a sequence logo for non-expressed LTR12C elements that retain the promoter region. **L.** Size distribution of all and up-regulated LTR12C and MER52A elements. **M.** Relative position of CAGE peaks on the LTRs excluding LTR12C subfamily members **N.** Distribution of CAGE tags on long (>5kb) LINE elements for different etiology groups. Blue (purple) bars: sense (antisense) direction with respect to the LINE elements. **O.** Significantly up-regulated repeat families (FDR<0.05) in tumor compared to non-tumor. The analysis was done using single-map as well as multi-map tags (mapping quality below 10). The vast majority of up-regulated families are LTR, whereas no SINE families are significantly changed.

**Figure S3. Activation of LTR retroviral promoters in mouse HCC tumors.** **A.** Numbers of peaks for coding, proximal, and distal peaks **B.** Fractions of significantly upregulated coding, proximal, and distal peaks in high-grade HCC tumors compared to WT. The threshold is set to  $p < 0.05$  (FDR, calculated with edgeR). **C.** Fractions of up-regulated or non-upregulated peaks in repeats in the sense or anti-sense (AS) directions **D.** Fractions of up-regulated peaks excluding “-int” elements or using tags with mapping quality 30 or above. **E.** Significantly up-regulated repeat families (FDR<0.05) in HCC (high) compared to WT. The analysis was done using single-map as well as multi-map tags (mapping quality below 10). The vast majority of up-regulated families are LTR, whereas no SINE families are significantly changed.

**Figure S4. Expression profiles of 43 LTR-derived ncRNAs.** **A.** Expression values of 43 LTRs for tumors (red), non-tumors (yellow), and normal (green) samples. **B.** Expression values of LTRs (LTR-007 to LTR-043) in Fantom 5 samples: 133 human normal tissues (red), 421 primary cells (yellow), and 252 cell lines (green). **C.** Determination of all transcripts in LTR-001 and 002 loci by PCRs using 10 primer pairs designed on different exons. F1 and F1.1 were used as forward primers with reverse primers, R1-4 and R5-6 respectively. F2 and F2.1 were used as forward primers with reverse primers, R8-9 and R9-10 respectively. The obtained



bands marked by arrows were subjected to sequence by MiSeq 250bp paired-end and determined full-length transcripts.

**Figure S5. qPCR validation of LTR-derived ncRNAs.** Fold changes of LTR-001, 002, 004, and 007 in tumor compared to non-tumor samples were measured by qPCR in the first 50 samples and the second 21 samples.

**Figure S6. Activity of the distal ncRNAs in HepG2 cells.** **A.** Transcription factors significantly associated with up-regulated ncRNAs. TBP (TATA box-binding protein) had the lowest p-value, and ARID3A (AT-rich interactive domain-containing protein 3) had the highest binding counts around TSSs of the ncRNAs. **B.** Enrichment of ChIP-Seq peaks of RNA-Polymerase II, H3K4me3, H3K27ac, and DNase Footprints within 2 kb from CAGE peaks in LTR elements. The red and blue lines represent active and inactive peaks in the HepG2 cell line, respectively. **C.** Enrichment of ChIP-seq peaks of 6 transcription factors within 2kb from CAGE peaks in LTR elements. **D.** Knockdown efficiency of LTR-001 and 002 measured by qPCR at 12h, 24h, 48h, and 72h after transfection of LNAs into HepG2 cells. CT1: Lipofectamine alone, CT2: Lipofectamine with scramble LNAs, KD1: Lipofectamine with

LNA designed on the LTR. KD2: Lipofectamine with another LNA designed on the LTR. **E.**

Knockdown efficiency of LTR-001 and 002 at 12h and 48h measured by CAGE. **F.**

Representative correlation plots between two different LNAs targeting LTR-001 and LTR-002.

The average spearman's correlation coefficient between two different LNAs is 0.935 whereas

the correlation between biological replicates of the same LNAs is 0.928.

Figure S1

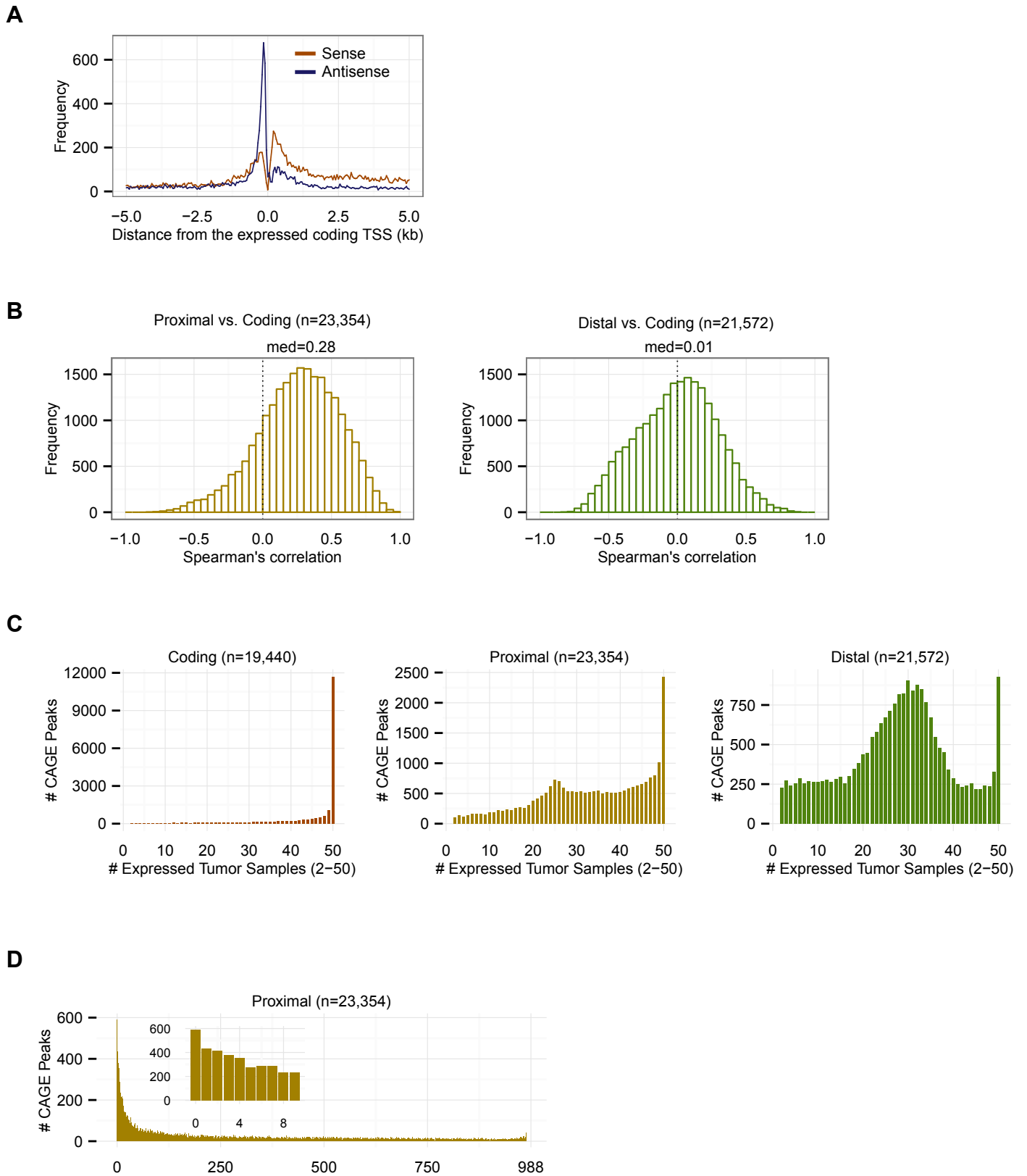
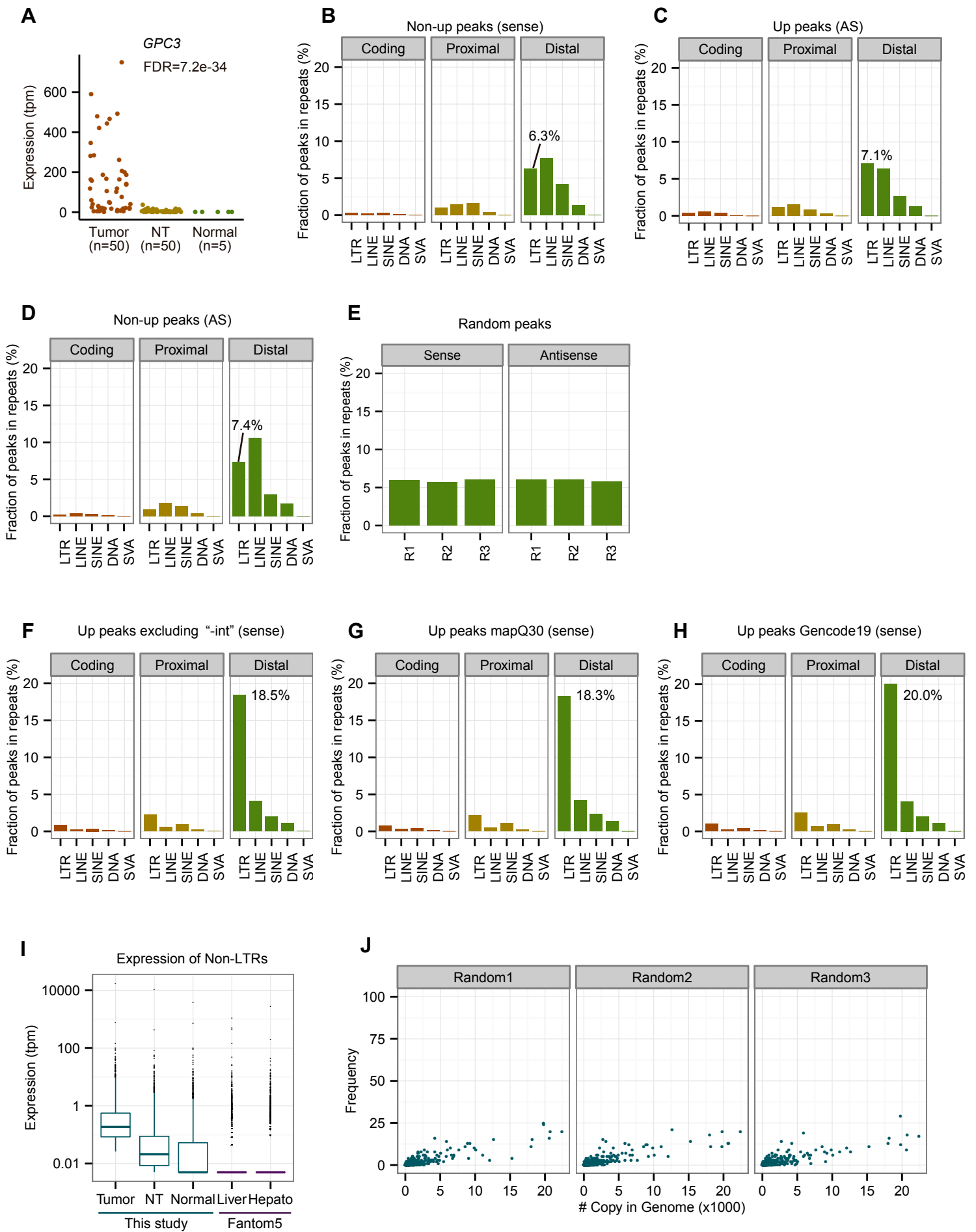


Figure S2



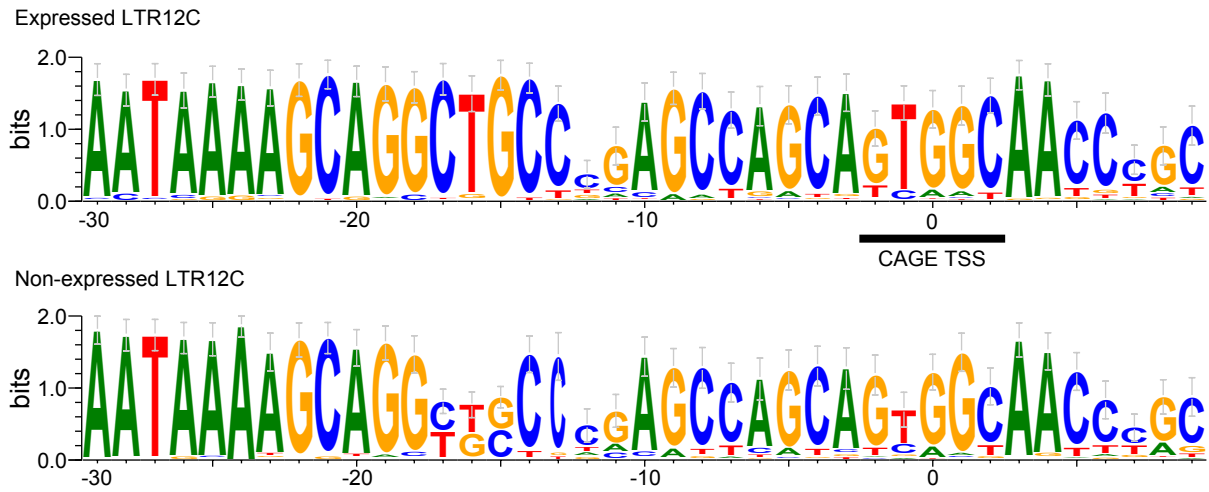
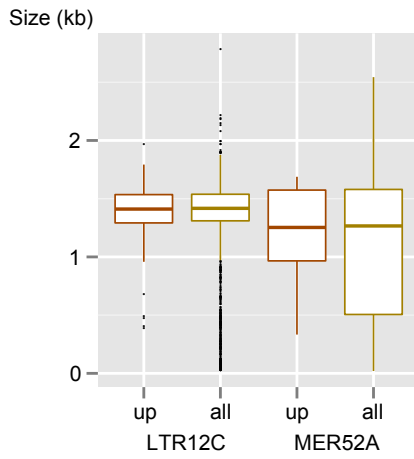
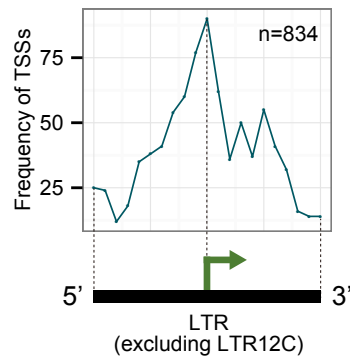
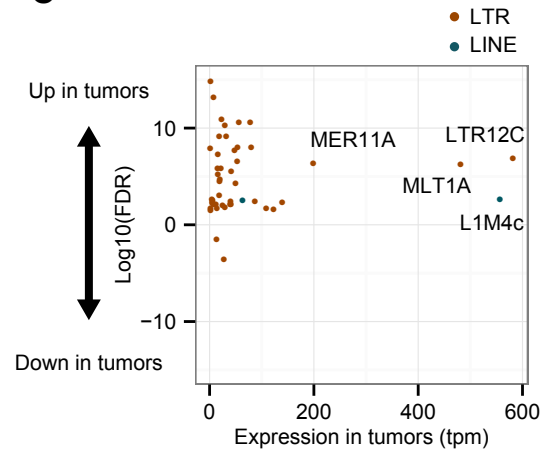
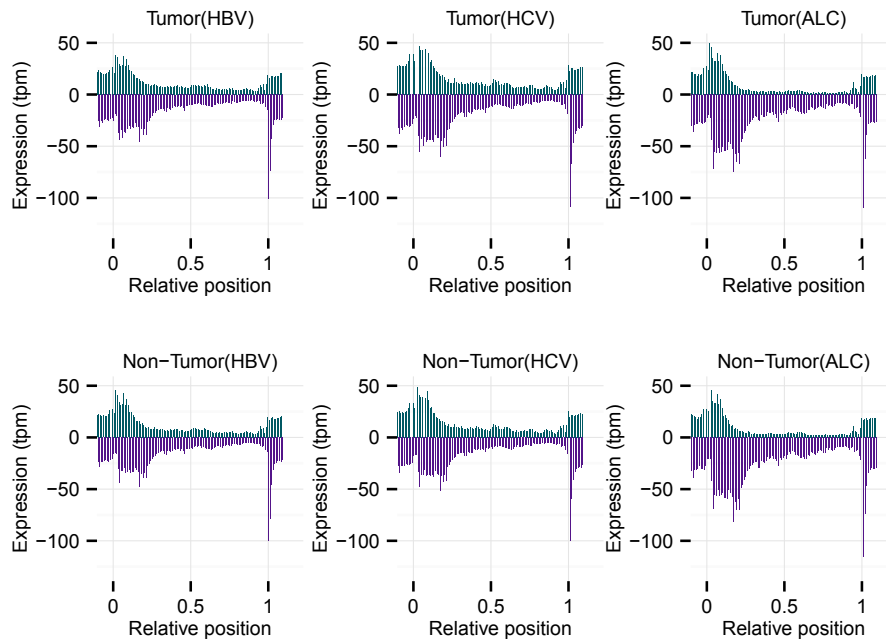
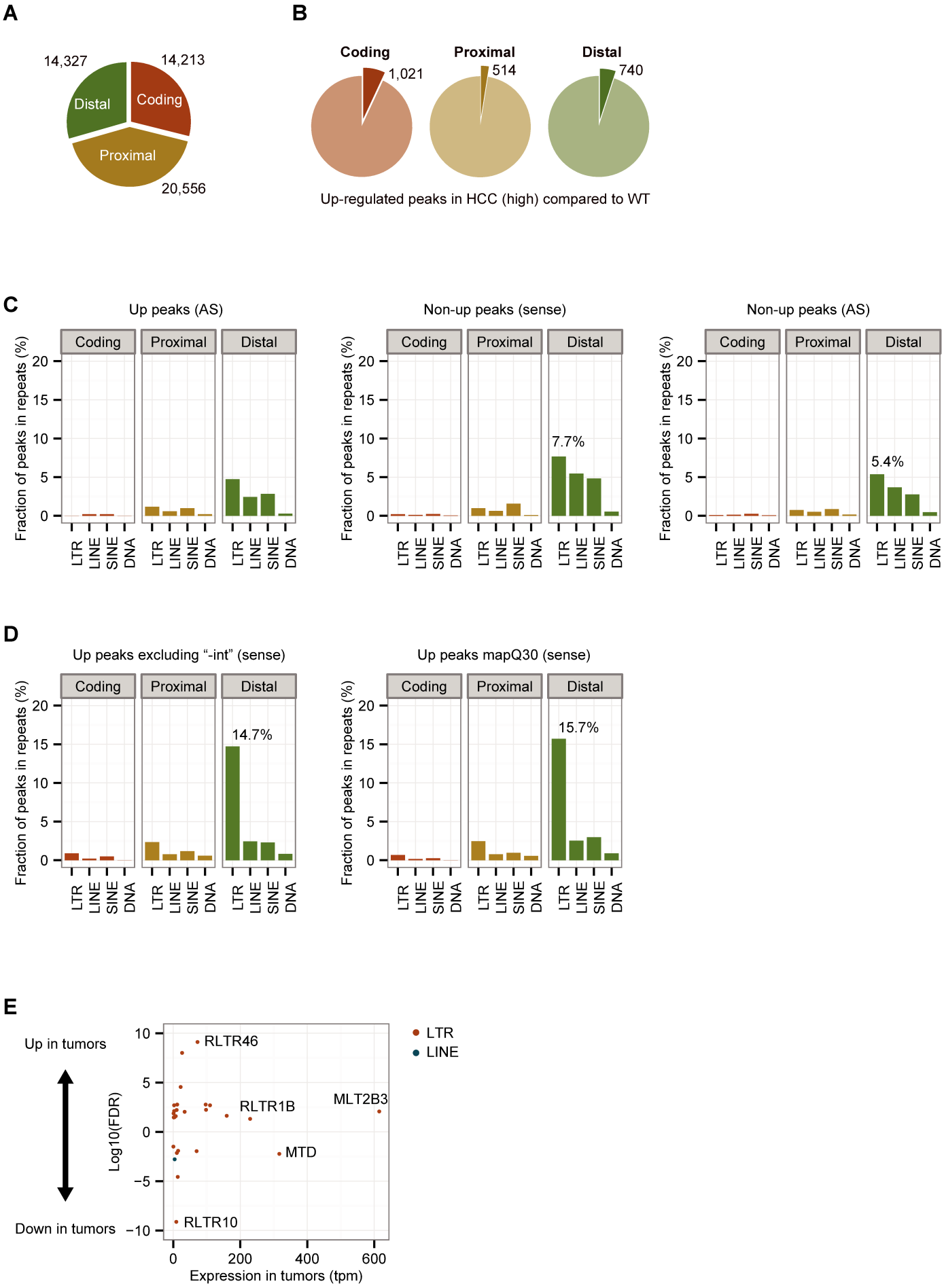
**K****L****M****O****N**

Figure S3



# Figure S4A

Expression profiles of  
43 LTRs in liver tissues

- Tumor (n=50)
- Non-tumor (n=50)
- Normal (n=5)

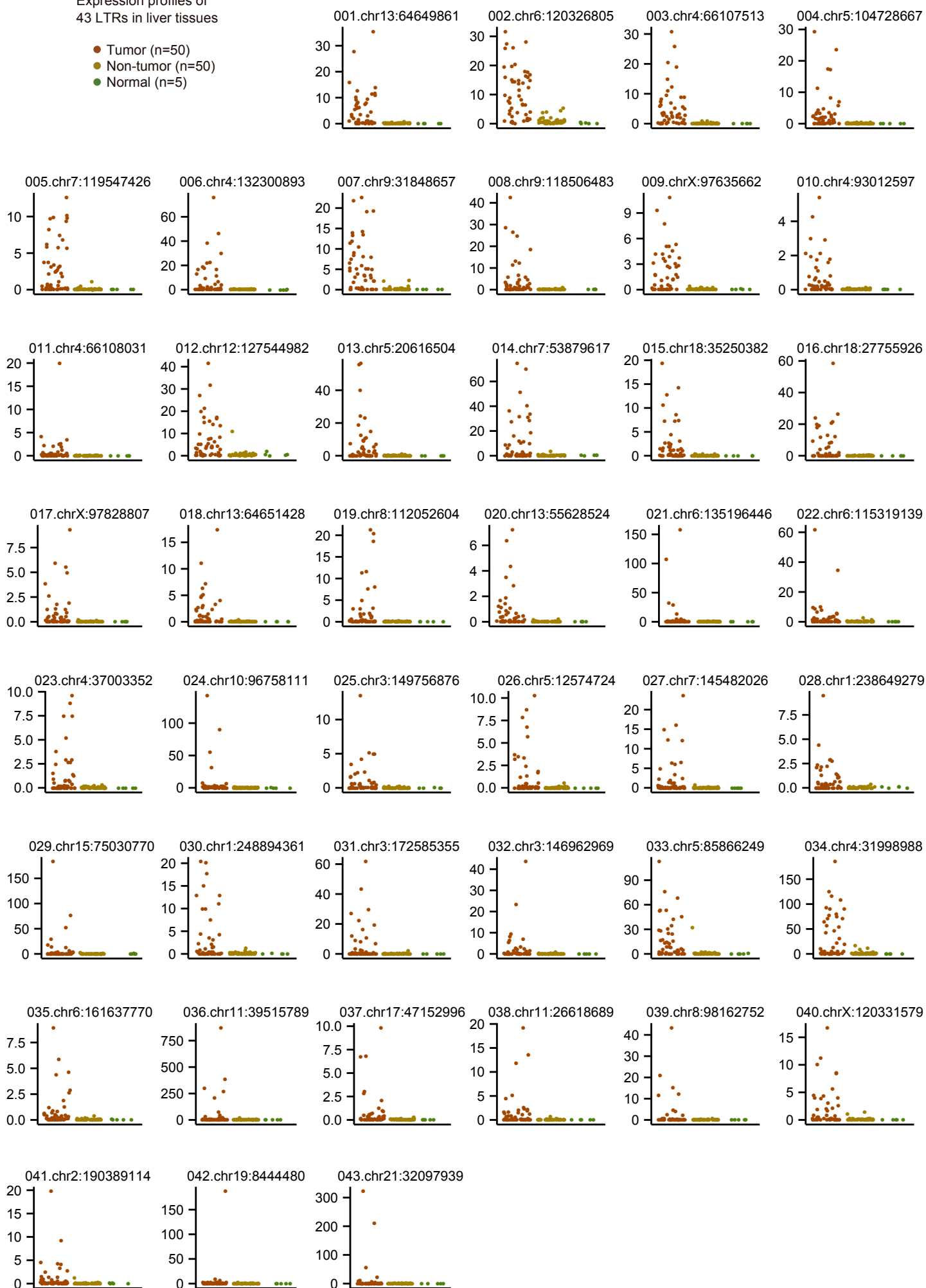
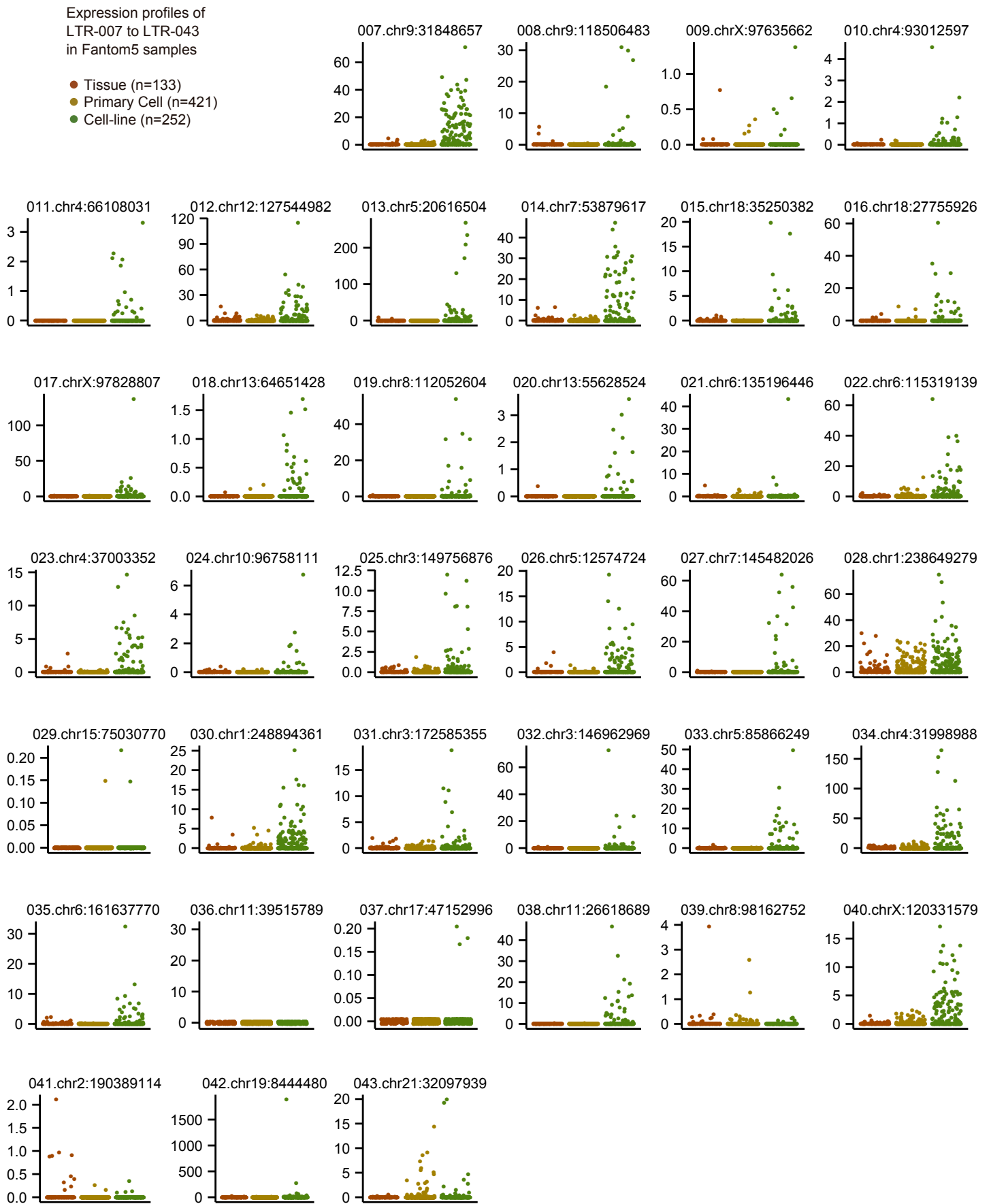


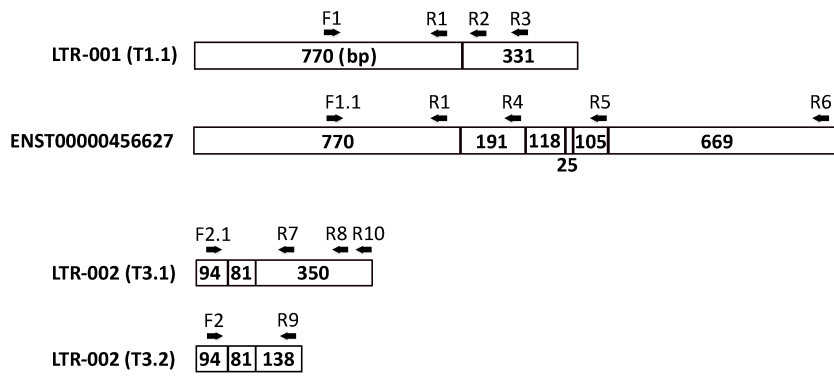
Figure S4B





# Figure S4C

## Primer design



## Forward Primers

F1 GACGCCCTGCCTATGGAAC  
 F1.1 ACCAGCTGTTCACTCTGGGTCTC  
 F2 TCGCTTTCTGCCATGATTGC  
 F2.1 TGTTTCCTACCGCGTTGTGAAG

## Reverse Primers

R1 CCCAGAAAGGGTGTACAGT  
 R2 CTGTGTCCATCCCTCCTCTG  
 R3 AAATCCGACTTGCCTTTGAA  
 R4 GCACATTGGCGTGATACAAC  
 R5 GGGCAGCGTTTCTTATCTGATGG  
 R6 GAGGGAAGGATGGCAACTCATGG  
 R7 AGTGTTATTTAATGTTACTATGGT  
 R8 GGGTCTTTGTGCCACATTTT  
 R9 ATGAACCTGCAGATCCAGTG  
 R10 AATTTTACTTCACITTTATTGTGC

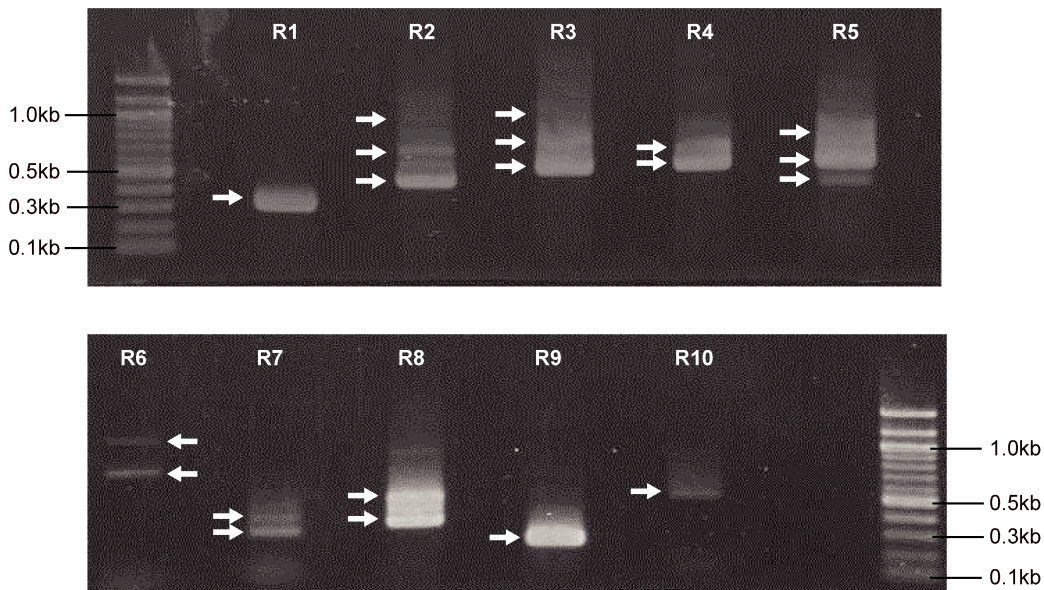
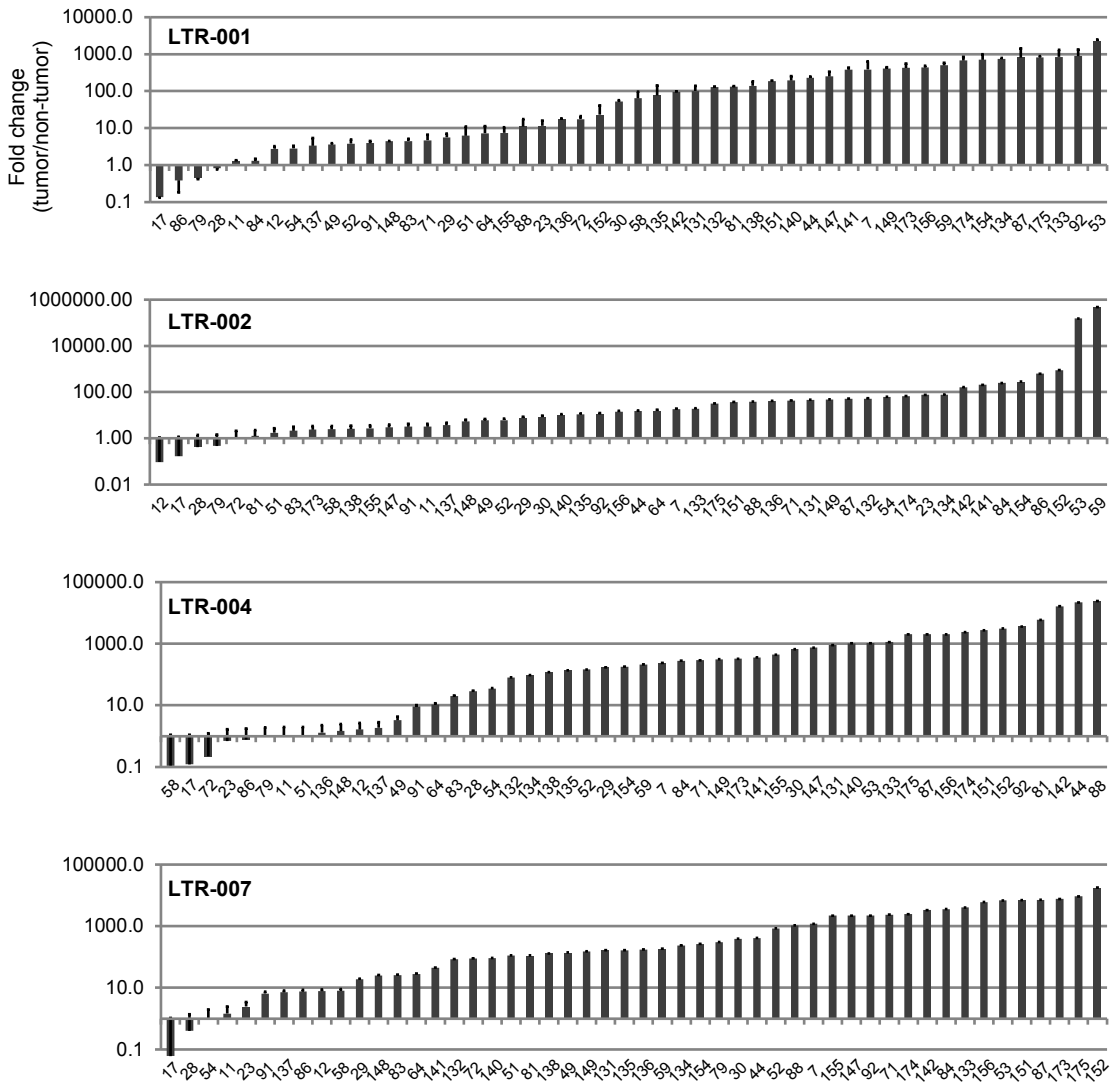
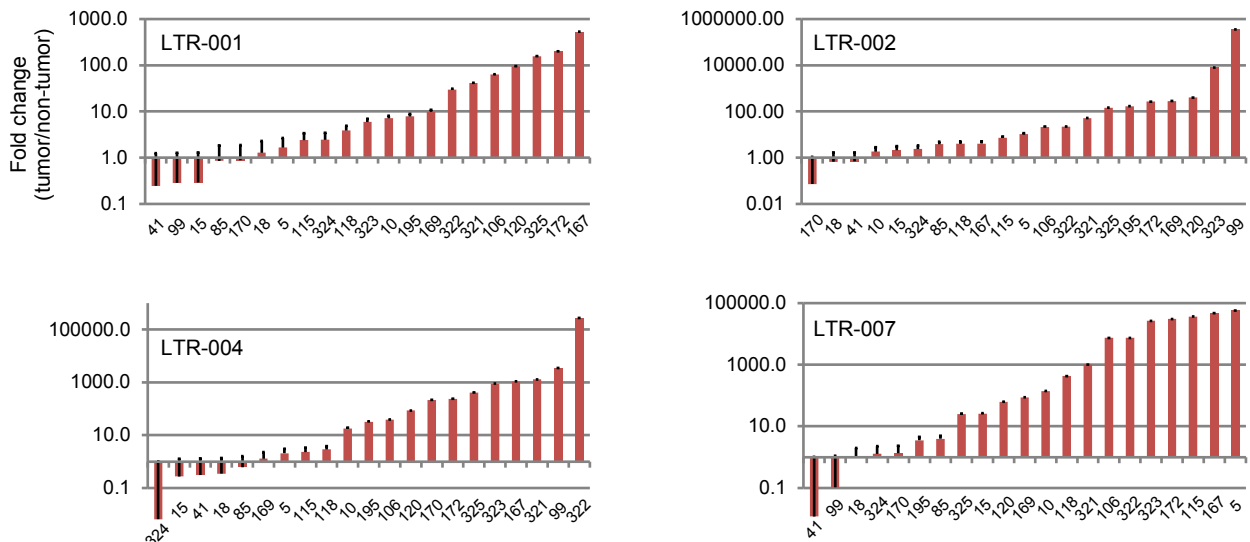


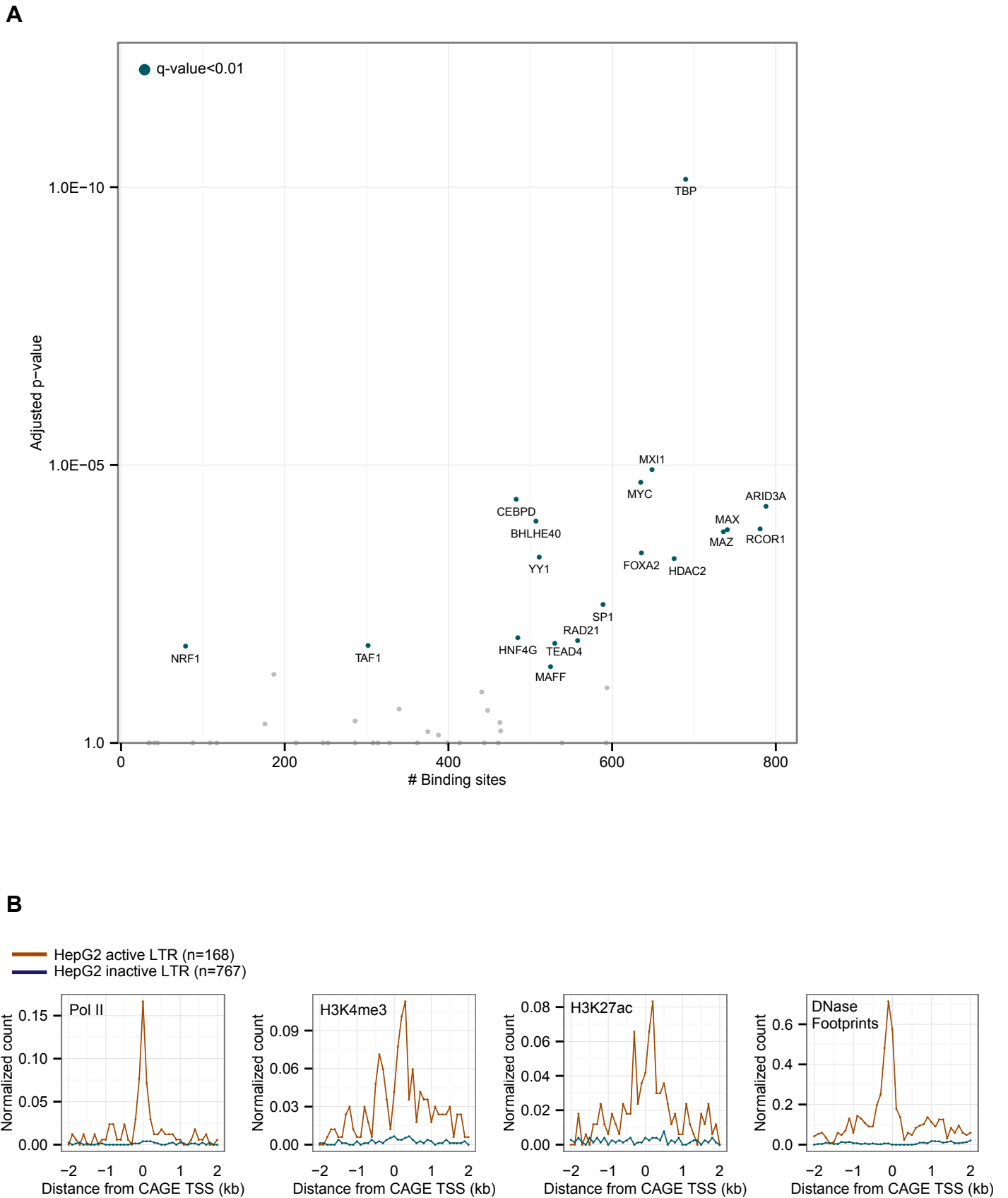
Figure S5

The first set: 50 samples



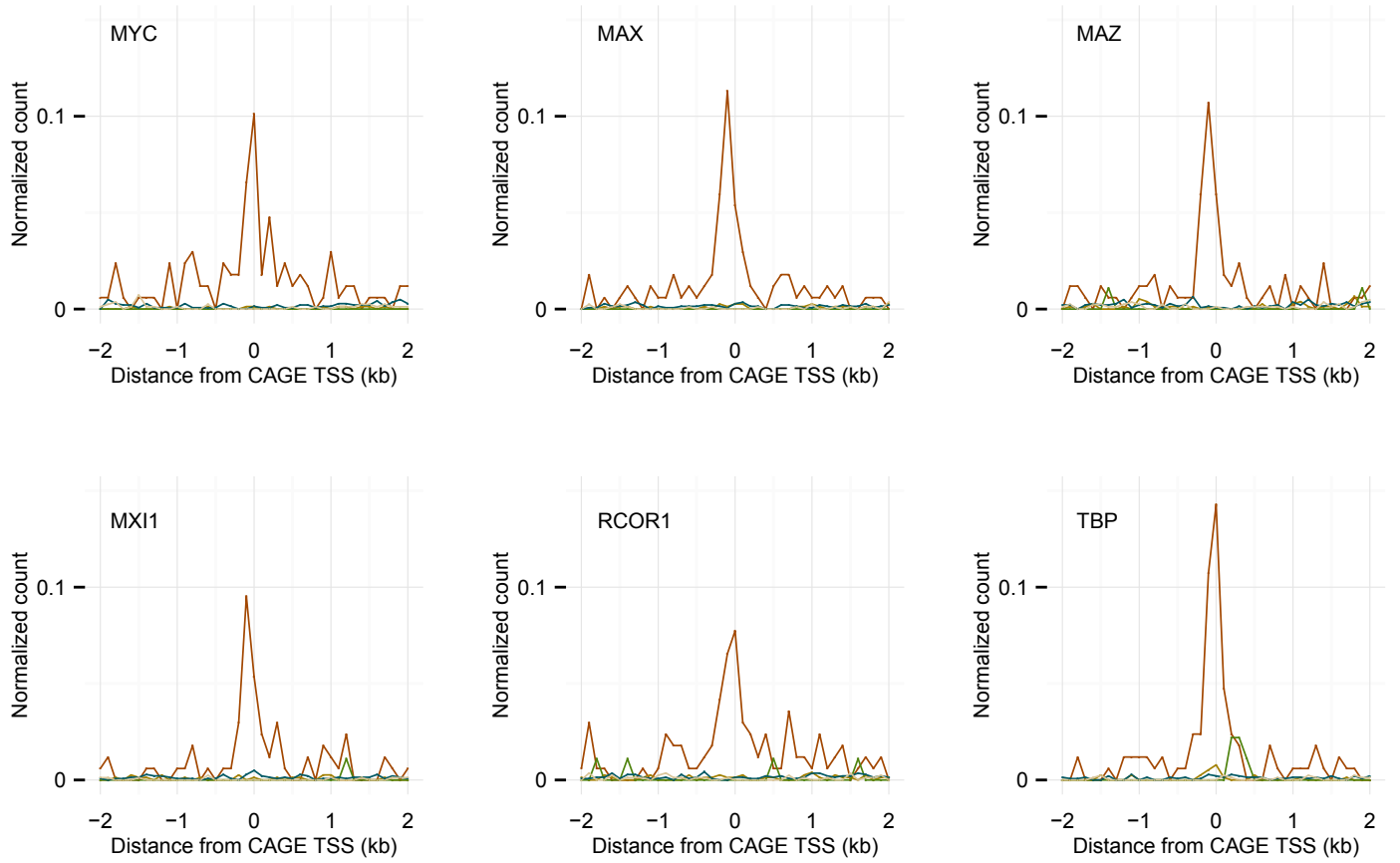
The second set: 21 samples





**C**

- HepG2 active LTR (n=168)
- HepG2 inactive LTR (n=767)
- HepG2 inactive LTR12C (n=91)
- LTR12B,C,D,E, or F (n=1418)
- LTR17 (n=831)



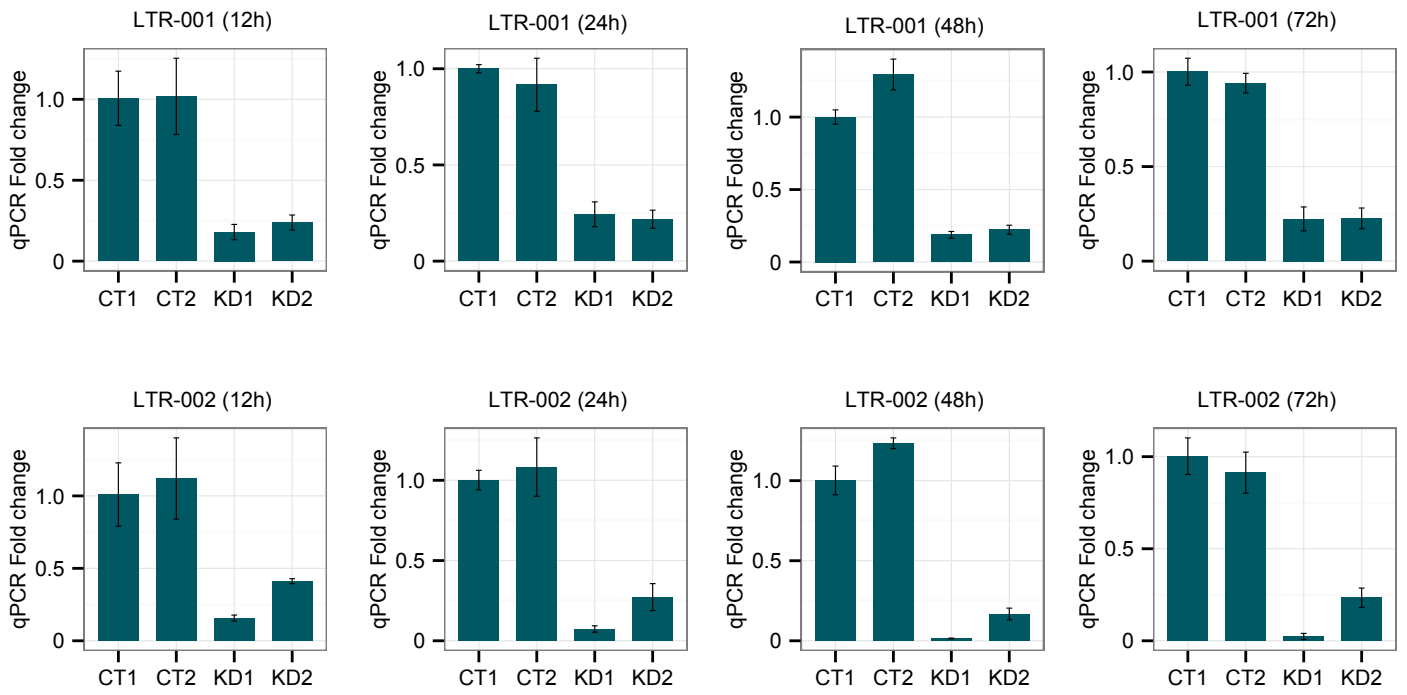
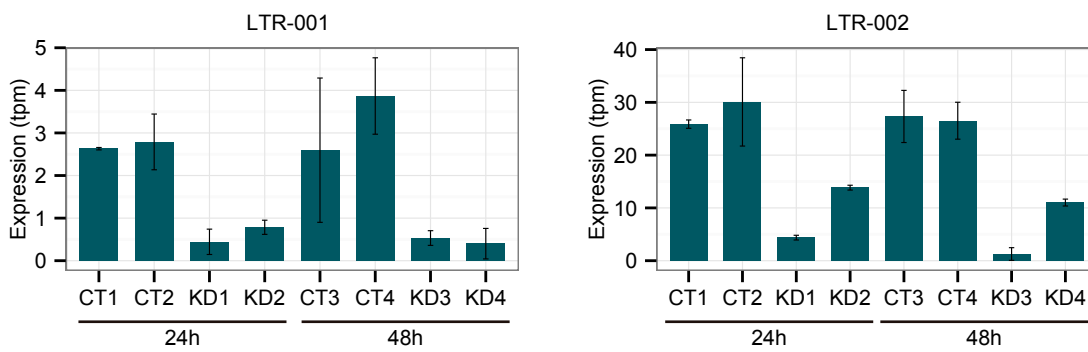
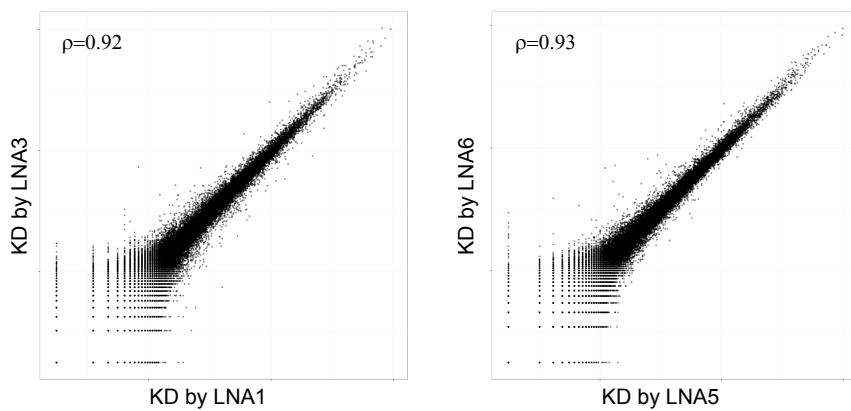
**D****E****F**

Table S1. HCC samples with etiology and map counts

Number	Sample ID	Type	Etiology	Gender	RIN	Read count	Mapped	Single map	Map rate
1	17T	Tumor	HBV	male	8.6	13,704,114	11,131,494	10,057,951	81%
2	23T	Tumor	HBV	male	10	34,526,410	32,495,689	28,344,846	94%
3	52T	Tumor	HBV	male	8.7	9,393,758	8,529,410	7,986,553	91%
4	71T	Tumor	HBV	male	9.5	9,448,052	8,794,498	7,902,752	93%
5	81T	Tumor	HBV	male	7.9	29,757,451	27,673,138	24,261,335	93%
6	86T	Tumor	HBV	female	9	13,982,080	12,941,944	11,148,718	93%
7	133T	Tumor	HBV	male	7.9	13,554,448	12,241,611	10,706,465	90%
8	137T	Tumor	HBV	male	8.2	18,361,250	16,760,712	13,689,986	91%
9	140T	Tumor	HBV	male	7.1	25,775,615	23,950,279	19,768,074	93%
10	141T	Tumor	HBV	male	8	19,346,705	17,575,608	15,395,218	91%
11	154T	Tumor	HBV	male	7.7	18,581,260	16,944,580	15,187,136	91%
12	155T	Tumor	HBV	male	7.5	19,044,662	17,100,978	15,549,557	90%
13	156T	Tumor	HBV	male	7.6	11,163,135	10,290,388	9,269,347	92%
14	173T	Tumor	HBV	male	9.2	33,628,289	31,429,432	27,814,369	93%
15	174T	Tumor	HBV	male	8.4	13,286,218	12,318,911	10,806,132	93%
16	175T	Tumor	HBV	male	8.6	32,000,160	29,904,230	25,970,240	93%
17	11T	Tumor	HCV	male	8.8	31,347,758	29,386,132	25,245,047	94%
18	12T	Tumor	HCV	male	8.9	12,274,044	11,399,462	9,891,139	93%
19	29T	Tumor	HCV	male	8.8	17,434,011	16,246,397	14,100,265	93%
20	49T	Tumor	HCV	male	8.4	26,189,069	24,588,064	21,310,469	94%
21	72T	Tumor	HCV	male	8.5	26,938,828	25,199,079	22,030,051	94%
22	91T	Tumor	HCV	male	8.8	8,618,017	7,586,177	6,504,774	88%
23	131T	Tumor	HCV	female	8.2	24,397,405	22,003,796	18,310,187	90%
24	132T	Tumor	HCV	male	7.9	33,973,535	30,725,852	26,996,995	90%
25	134T	Tumor	HCV	male	8.1	30,285,017	26,693,466	22,398,364	88%
26	135T	Tumor	HCV	male	7.2	23,799,961	21,659,035	18,219,428	91%
27	136T	Tumor	HCV	male	8	19,114,206	17,332,811	14,978,015	91%
28	138T	Tumor	HCV	female	8.3	22,710,452	20,889,761	17,315,752	92%
29	142T	Tumor	HCV	male	6.9	21,295,859	19,295,625	16,544,160	91%
30	147T	Tumor	HCV	female	7.7	24,177,262	22,069,033	19,554,254	91%
31	148T	Tumor	HCV	female	8.2	12,114,410	10,897,676	9,860,845	90%
32	149T	Tumor	HCV	male	9	39,997,354	36,590,187	32,581,254	91%
33	151T	Tumor	HCV	female	8.4	15,890,142	14,356,824	12,764,379	90%
34	152T	Tumor	HCV	female	9	13,669,240	12,590,006	11,505,888	92%
35	28T	Tumor	Alcohol abuse	male	8.4	29,346,572	26,563,450	22,997,478	91%
36	30T	Tumor	Alcohol abuse	male	8.2	18,963,155	17,913,140	15,327,709	94%
37	44T	Tumor	Alcohol abuse	male	7.8	14,571,524	13,685,675	11,835,762	94%
38	7T	Tumor	Alcohol abuse	male	8.4	27,360,816	25,390,730	22,321,213	93%
39	83T	Tumor	Alcohol abuse	male	7.6	10,892,088	9,896,427	8,679,861	91%
40	88T	Tumor	Alcohol abuse	male	8.9	37,410,176	34,755,389	31,355,638	93%
41	92T	Tumor	Alcohol abuse	male	7.6	13,889,818	12,667,987	11,214,084	91%
42	59T	Tumor	Alcohol abuse	male	5.6	10,969,469	9,630,605	8,741,640	88%
43	58T	Tumor	Alcohol abuse	male	8	20,262,303	19,185,690	16,985,353	95%
44	64T	Tumor	Alcohol abuse	male	7	8,239,372	7,499,221	6,606,183	91%
45	51T	Tumor	HBV+HCV	male	9.5	22,472,479	20,667,558	19,046,624	92%
46	53T	Tumor	Autoimmune	female	8.6	13,127,042	12,008,177	10,593,579	91%
47	87T	Tumor	NASH	male	7.9	14,370,235	12,816,899	11,632,369	89%
48	54T	Tumor	WIL	male	7.7	9,866,988	8,832,359	8,123,513	90%
49	84T	Tumor	Unknown	male	8.1	16,839,679	15,081,376	13,596,219	90%
50	79T	Tumor	Unknown	male	8.8	11,366,739	10,671,947	9,450,468	94%
51	17NT	Non-tumor	HBV	male	8.2	4,952,149	4,089,802	3,657,964	83%
52	23NT	Non-tumor	HBV	male	9.3	31,859,986	29,345,240	25,672,297	92%
53	52NT	Non-tumor	HBV	male	9.2	15,459,585	14,346,903	12,737,372	93%
54	71NT	Non-tumor	HBV	male	7.8	24,525,111	22,857,897	19,659,022	93%
55	81NT	Non-tumor	HBV	male	8.6	36,083,947	33,644,047	29,861,955	93%
56	86NT	Non-tumor	HBV	female	6.3	13,267,499	11,990,398	10,084,745	90%

57	133NT	Non-tumor	HBV	male	8.3	15,809,573	13,918,837	12,073,071	88%
58	137NT	Non-tumor	HBV	male	8.2	14,928,380	13,576,095	11,586,526	91%
59	140NT	Non-tumor	HBV	male	7.9	18,848,537	16,907,315	14,932,883	90%
60	141NT	Non-tumor	HBV	male	7.4	20,099,768	18,015,093	15,331,814	90%
61	154NT	Non-tumor	HBV	male	7.3	22,278,862	20,214,152	18,398,102	91%
62	155NT	Non-tumor	HBV	male	6.2	10,352,268	9,422,626	8,656,389	91%
63	156NT	Non-tumor	HBV	male	6.9	8,100,277	7,449,314	6,746,743	92%
64	173NT	Non-tumor	HBV	male	9	31,890,433	30,102,764	26,417,530	94%
65	174NT	Non-tumor	HBV	male	8.6	19,465,347	18,041,909	15,775,145	93%
66	175NT	Non-tumor	HBV	male	6.5	26,287,447	25,091,661	21,805,115	95%
67	11NT	Non-tumor	HCV	male	8	16,738,091	15,788,083	13,572,802	94%
68	12NT	Non-tumor	HCV	male	8.4	22,067,957	20,430,729	17,695,368	93%
69	29NT	Non-tumor	HCV	male	8.7	8,921,719	8,238,391	7,187,997	92%
70	49NT	Non-tumor	HCV	male	8.5	25,164,907	23,470,166	20,504,277	93%
71	72NT	Non-tumor	HCV	male	7.9	28,129,758	25,898,262	22,881,340	92%
72	91NT	Non-tumor	HCV	male	9	16,859,719	15,062,680	13,230,514	89%
73	131NT	Non-tumor	HCV	female	7.8	11,434,712	10,215,653	8,671,702	89%
74	132NT	Non-tumor	HCV	male	7.8	24,134,839	21,738,928	18,785,473	90%
75	134NT	Non-tumor	HCV	male	7.8	18,354,742	16,665,006	14,186,414	91%
76	135NT	Non-tumor	HCV	male	7.4	21,285,078	19,269,560	16,079,945	91%
77	136NT	Non-tumor	HCV	male	7.8	10,372,005	9,213,917	8,114,142	89%
78	138NT	Non-tumor	HCV	female	7.8	25,810,748	23,327,876	19,663,074	90%
79	142NT	Non-tumor	HCV	male	8.8	30,947,952	28,213,004	24,865,650	91%
80	147NT	Non-tumor	HCV	female	7	20,920,849	18,814,767	16,594,177	90%
81	148NT	Non-tumor	HCV	female	8.1	16,256,292	14,614,583	13,176,739	90%
82	149NT	Non-tumor	HCV	male	8	25,049,915	22,501,418	20,198,693	90%
83	151NT	Non-tumor	HCV	female	9	18,103,115	16,333,599	14,602,482	90%
84	152NT	Non-tumor	HCV	female	8.2	15,227,304	13,819,750	12,623,047	91%
85	28NT	Non-tumor	Alcohol abuse	male	9.4	18,415,109	17,427,849	15,212,504	95%
86	30NT	Non-tumor	Alcohol abuse	male	8	36,677,681	34,582,453	29,705,982	94%
87	44NT	Non-tumor	Alcohol abuse	male	7.1	27,601,475	25,721,122	22,322,225	93%
88	7NT	Non-tumor	Alcohol abuse	male	7.6	15,400,894	14,500,578	12,649,686	94%
89	83NT	Non-tumor	Alcohol abuse	male	7.9	12,302,344	11,156,455	9,982,624	91%
90	88NT	Non-tumor	Alcohol abuse	male	7.1	23,386,207	21,203,361	19,163,055	91%
91	92NT	Non-tumor	Alcohol abuse	male	10	19,870,942	18,158,308	16,594,194	91%
92	59NT	Non-tumor	Alcohol abuse	male	6.8	14,761,194	13,053,058	11,926,456	88%
93	58NT	Non-tumor	Alcohol abuse	male	6.5	7,582,225	7,014,493	6,356,770	93%
94	64NT	Non-tumor	Alcohol abuse	male	7.3	9,943,414	9,286,654	8,327,211	93%
95	51NT	Non-tumor	HBV+HCV	male	7.6	11,586,463	10,097,726	8,866,281	87%
96	53NT	Non-tumor	Autoimmune	female	8.9	16,597,955	15,458,580	13,510,315	93%
97	87NT	Non-tumor	NASH	male	9.2	32,470,921	29,524,474	26,649,603	91%
98	54NT	Non-tumor	WIL	male	8.3	26,703,622	24,828,555	21,077,564	93%
99	84NT	Non-tumor	Unknown	male	7.9	25,205,759	22,195,075	19,733,852	88%
100	79NT	Non-tumor	Unknown	male	8.6	18,129,437	17,250,865	14,572,380	95%
101	NL2	Normal	Normal	undefined	7.8	16,883,878	15,142,360	13,574,375	90%
102	NL3	Normal	Normal	undefined	7.2	33,157,781	26,822,976	23,340,295	81%
103	NL4	Normal	Normal	undefined	7.4	24,782,874	22,339,733	19,748,343	90%
104	NL5	Normal	Normal	undefined	7.8	24,054,644	21,695,420	19,453,471	90%
105	NL6	Normal	Normal	undefined	7.3	14,627,021	13,160,229	11,571,324	90%

Table S2. Fantom5 data used to measure expression in normal livers

Number	Technology	Definition	Sample ID
1	CAGE	adult liver tissue	CNhs10624
2	CAGE	fetul liver tissue	CNhs11798
3	CAGE	primary hepatocyte donor1	CNhs12340
4	CAGE	primary hepatocyte donor2	CNhs12349
5	CAGE	primary hepatocyte donor3	CNhs12626



Table S3. A list of up-regulated LTR12C in human HCC

	Peak	Log2FC	FDR
1	chr13_64651428_64651436_-	6.1	2.8E-16
2	chr8_112052604_112052671_-	6.2	7.3E-16
3	chr2_108363389_108363407_-	6.9	1.0E-13
4	chr2_140168479_140168533_+	5.7	2.9E-13
5	chr5_12574724_12574910_-	5.5	3.3E-13
6	chr3_172585355_172585370_+	5.5	5.7E-13
7	chr3_146962969_146962989_-	6.3	6.8E-13
8	chr4_31998988_31999006_+	4.7	1.2E-12
9	chr2_88803503_88803509_+	7.5	1.3E-12
10	chr3_145051902_145051917_+	6.4	7.2E-12
11	chr11_21774394_21774408_-	5.6	2.3E-11
12	chr7_148141630_148141638_-	6.6	8.8E-11
13	chr8_83354646_83354651_+	5.9	1.2E-10
14	chr1_79735629_79735638_+	6.0	2.4E-10
15	chr3_83267071_83267112_-	5.1	8.7E-10
16	chr4_187812082_187812090_+	5.8	1.8E-09
17	chr4_163661593_163661651_+	5.5	2.0E-09
18	chr1_194327154_194327213_-	5.3	2.0E-09
19	chr11_29735181_29735221_+	5.7	9.8E-09
20	chr14_44854772_44854777_+	7.2	1.1E-08
21	chr1_154644064_154644071_+	4.0	1.3E-08
22	chr8_80289026_80289041_-	6.4	2.0E-08
23	chr22_26149438_26149489_+	5.6	5.4E-08
24	chr1_81643879_81643889_-	5.6	9.5E-08
25	chr11_42896529_42896534_+	6.3	1.1E-07
26	chr18_63273306_63273317_+	6.8	1.7E-07
27	chr4_138864174_138864179_+	6.1	2.0E-07
28	chr5_13579412_13579422_+	7.7	2.0E-07
29	chr1_187042495_187042523_-	4.7	6.3E-07
30	chr2_181556732_181556740_+	5.7	6.5E-07
31	chr3_163217645_163217657_+	5.8	1.6E-06
32	chr2_52007008_52007022_-	3.4	2.4E-06
33	chr6_28606947_28606954_-	7.4	2.6E-06
34	chr14_88200640_88200666_-	4.9	4.1E-06
35	chr8_84315924_84315972_+	4.7	4.2E-06
36	chr4_90412098_90412103_+	3.8	5.1E-06
37	chr8_117544072_117544076_+	8.7	5.7E-06
38	chr3_165635822_165635825_-	6.1	6.2E-06
39	chr14_41099157_41099160_-	4.3	7.8E-06
40	chr2_175412747_175412762_-	5.7	8.5E-06
41	chr3_145052033_145052038_+	5.8	9.3E-06
42	chr4_70441336_70441356_-	4.4	1.2E-05
43	chr9_87024408_87024421_+	6.3	1.8E-05
44	chr16_64377812_64377899_+	3.2	1.9E-05
45	chr1_210061899_210061905_-	8.3	2.0E-05
46	chr5_150689584_150689588_+	5.5	2.2E-05
47	chr6_30989607_30989611_+	7.7	3.4E-05
48	chr1_171350384_171350391_-	4.4	3.8E-05
49	chr3_79552183_79552188_+	4.1	4.0E-05
50	chr12_20936266_20936299_-	3.8	5.9E-05
51	chr7_83303632_83303637_+	8.2	7.3E-05
52	chr3_85554118_85554157_-	2.5	8.0E-05
53	chrX_118343588_118343600_-	3.5	8.3E-05
54	chr11_21774185_21774220_-	5.0	1.0E-04
55	chr5_19288676_19288683_+	7.7	1.1E-04
56	chr6_28926720_28926729_+	4.7	1.4E-04

57	chr14_70722346_70722352_+	3.5	1.5E-04
58	chr12_92028652_92028706_+	2.2	1.7E-04
59	chr10_62801186_62801190_+	8.3	1.8E-04
60	chr1_154644441_154644443_+	3.8	1.9E-04
61	chr4_14361000_14361038_+	3.9	1.9E-04
62	chr4_34336965_34336974_-	4.2	2.4E-04
63	chr6_122601794_122601878_+	2.6	3.0E-04
64	chr10_15467620_15467650_-	6.2	3.8E-04
65	chr3_137432930_137432963_-	5.3	5.0E-04
66	chr9_105248812_105248816_-	4.7	5.0E-04
67	chr13_91147974_91147983_+	4.7	5.7E-04
68	chr9_91300401_91300406_-	5.7	6.7E-04
69	chr15_62478283_62478287_-	4.3	9.0E-04
70	chr6_113010124_113010129_-	5.4	9.2E-04
71	chr4_142253763_142253785_-	1.2	9.3E-04
72	chr6_52652225_52652242_+	1.6	1.1E-03
73	chr13_55311031_55311041_-	4.0	1.1E-03
74	chr2_52234734_52234741_+	3.6	1.6E-03
75	chr8_127726265_127726269_-	3.9	2.6E-03
76	chr5_105582640_105582705_+	3.8	2.6E-03
77	chr10_67188563_67188589_-	4.2	3.1E-03
78	chr2_192787301_192787319_+	4.0	3.1E-03
79	chr12_6826148_6826150_+	4.5	3.5E-03
80	chr7_79369447_79369452_+	5.1	3.6E-03
81	chr12_72579863_72579867_+	6.3	4.0E-03
82	chr18_56127649_56127652_+	5.6	4.3E-03
83	chr13_81913922_81913954_-	6.7	4.6E-03
84	chr22_39754393_39754398_+	3.7	5.4E-03
85	chr9_21959870_21959874_+	4.2	5.8E-03
86	chr1_115904381_115904395_-	2.5	6.1E-03
87	chr11_84019132_84019170_-	2.8	8.8E-03
88	chrX_118341749_118341754_-	3.3	9.4E-03
89	chr6_36331397_36331401_-	2.0	9.7E-03
90	chr6_28833662_28833666_-	3.2	1.1E-02
91	chr6_24765752_24765756_-	2.9	1.6E-02
92	chr6_30774806_30774810_-	2.9	1.8E-02
93	chr6_64152249_64152258_-	2.7	1.8E-02
94	chr4_6310364_6310369_-	3.1	2.1E-02
95	chr4_88784227_88784231_-	3.2	2.5E-02
96	chrX_118055535_118055549_-	2.6	2.9E-02
97	chr1_191701634_191701668_+	2.4	2.9E-02
98	chr12_6818234_6818238_+	2.6	3.5E-02
99	chr11_124960335_124960376_+	2.7	4.1E-02
100	chr6_28436674_28436680_-	3.7	4.3E-02
101	chr7_152235561_152235568_-	1.5	4.4E-02
102	chr2_178701305_178701309_+	2.4	4.6E-02
103	chr12_71771003_71771009_+	3.0	4.6E-02
104	chr19_39313478_39313484_+	2.2	4.7E-02

---

Table S4. mouse liver samples with histology and map counts

Number	ID	Category	Histology	Gender	RIN	Read count	Mapped	Single map	Map rate	Used for statistics
1	71	WT	WT tissues	male	7.9	24,948,298	23,275,609	17,015,995	93%	Y
2	72	WT	WT tissues	male	7.5	12,925,948	11,701,886	8,936,593	91%	Y
3	73	WT	WT tissues	male	9.2	51,083,095	47,115,296	34,089,432	92%	Y
4	66	WT	WT cells	male	8.5	24,667,464	21,967,015	17,166,449	89%	N
5	67	WT	WT cells	male	7.3	7,594,431	6,955,074	5,124,318	92%	N
6	68	WT	WT cells	male	9.1	46,641,665	43,485,158	29,644,387	93%	N
7	58	INFL	Inflammation	male	7.5	40,776,157	37,669,539	32,020,059	92%	N
8	60	INFL	Inflammation	male	7.4	12,189,140	10,323,740	8,750,051	85%	N
9	59	INFL	Inflammation	male	5.9	4,396,831	3,423,206	3,074,718	78%	N
10	50	AD	40% AD, 0% HCC	female	8.2	5,747,335	5,110,778	4,625,606	89%	N
11	25	AD	30% AD, 0% HCC	male	6.4	7,122,524	5,031,920	4,542,201	71%	N
12	27	AD	20% AD, 0% HCC	male	6.9	4,613,660	3,909,734	3,520,529	85%	N
13	43	AD	60% AD, 0% HCC	male	6.1	11,085,453	9,407,793	8,580,968	85%	N
14	39	HCC (low)	70% AD, 20% HCC	male	7.5	12,784,580	11,687,696	10,565,343	91%	N
15	35	HCC (low)	70% AD, 10% HCC	male	7.6	22,446,560	20,513,738	18,559,098	91%	N
16	51	HCC (low)	10% AD, 20% HCC	female	7.1	15,155,881	12,725,069	11,857,857	84%	N
17	32	HCC (low)	60% AD, 10% HCC	male	7.3	21,738,865	19,081,480	17,519,961	88%	N
18	48	HCC (low)	50% AD, 20% HCC	male	7.6	3,163,218	2,354,036	2,192,109	74%	N
19	45	HCC (low)	30% AD, 40% HCC	female	7.2	11,416,625	10,515,674	9,597,571	92%	N
20	36	HCC (low)	80% AD, 10% HCC	male	7.6	6,674,328	5,274,349	4,833,128	79%	N
21	40	HCC (low)	80% AD, 20% HCC	male	7	3,634,863	3,115,385	2,770,311	86%	N
22	41	HCC (low)	70% AD, 30% HCC	male	8	9,433,303	8,432,116	7,712,509	89%	N
23	31	HCC (low)	70% AD, 10% HCC	male	6.5	6,082,739	4,954,134	4,493,532	81%	N
24	46	HCC (low)	80% AD, 10% HCC	female	6	2,007,420	1,780,884	1,630,205	89%	N
25	47	HCC (low)	80% AD, 10% HCC	female	6.2	4,055,905	3,670,159	3,354,484	90%	N
26	42	HCC (low)	70% AD, 20% HCC	male	5.2	5,703,096	4,828,486	4,319,918	85%	N
27	34	HCC (low)	70% AD, 20% HCC	male	5.9	6,373,320	5,603,402	5,062,103	88%	N
28	49	HCC (low)	0% AD, 50% HCC	female	7.5	2,798,209	2,271,901	2,110,452	81%	N
29	28	HCC (low)	20% AD, 50% HCC	male	6.1	2,856,689	2,266,860	2,042,043	79%	N
30	29	HCC (high)	0% AD, 90% HCC	male	7.4	12,809,505	11,879,489	10,653,463	93%	Y
31	37	HCC (high)	20% AD, 70% HCC	male	7.9	26,022,427	22,719,787	19,571,009	87%	Y
32	38	HCC (high)	20% AD, 70% HCC	male	9	25,638,306	21,871,072	19,641,767	85%	Y
33	30	HCC (high)	10% AD, 70% HCC	male	7.2	29,985,584	25,454,293	23,145,872	85%	Y
34	24	HCC (high)	20% AD, 80% HCC	male	4	3,332,081	2,318,807	2,169,238	70%	N
35	44	HCC (high)	10% AD, 60% HCC	female	5.4	3,364,254	2,679,636	2,415,675	80%	N
36	26	HCC (high)	10% AD, 60% HCC	male	5.3	2,859,731	2,122,221	1,936,007	74%	N
37	33	HCC (high)	0% AD, 60% HCC	male	4.8	5,682,756	4,096,510	3,692,358	72%	N

Table S5. A list of up-regulated LTR in mouse HCC

Num	Chr	Peak Start	Peak End	STR	FDR	LTR	Num	Chr	Start	End	STR	FDR	LTR
1	chr10	35166197	35166483	-	2.32E-09	RLTR4_MM-int	71	chr3	6375406	6375657	+	5.55E-03	RLTR27
2	chr2	80144374	80144811	-	2.35E-09	RLTR47_MM	72	chr8	25522802	25522807	+	5.65E-03	RLTR12B
3	chr6	141824559	141824573	+	3.81E-08	RLTR41	73	chr10	30354461	30354541	+	5.71E-03	MuRRS4-int
4	chr17	53181751	53181960	-	4.83E-08	RLTR47_MM	74	chr17	21611849	21612028	+	5.75E-03	MMVL30-int
5	chr7	5685012	5685128	+	8.19E-08	RLTR10A	75	chr13	33233239	33233486	-	5.75E-03	MMVL30-int
6	chr16	56116366	56116452	-	3.49E-07	RLTR17	76	chr9	1.22E+08	1.22E+08	-	6.21E-03	MLT1F1
7	chr4	107806906	107806917	+	5.71E-07	RLTR41	77	chr4	75046957	75047195	+	6.30E-03	RLTR44A
8	chr8	129409521	129409604	-	6.34E-07	ORR1E	78	chr12	49549721	49549825	-	6.34E-03	RLTR44E
9	chr3	84445398	84445413	-	7.00E-07	RMER19B	79	chr4	88626393	88626504	+	6.55E-03	RLTR20A2
10	chr11	31445733	31445973	+	1.45E-06	RLTR10D	80	chr10	1.05E+08	1.05E+08	+	8.16E-03	IAPLTR4_I
11	chr3	45018416	45018449	+	1.45E-06	RLTR10D	81	chr13	43444677	43445033	+	8.41E-03	RMER19B
12	chr1	123242725	123242859	-	6.02E-06	ORR1B1	82	chr12	41770170	41770262	+	8.41E-03	MTD
13	chr10	84525235	84525412	-	1.10E-05	RLTR47_MM	83	chr4	55355611	55355680	-	8.64E-03	ORR1A4
14	chr18	74508653	74508811	+	1.15E-05	RLTR47_MM	84	chr14	14766737	14766763	+	9.42E-03	RLTR1B-int
15	chr3	21857346	21857736	-	1.17E-05	MT2C_Mm	85	chr18	38388181	38388256	+	9.78E-03	RMER19B
16	chr4	118616054	118616058	-	2.57E-05	MTA_Mm	86	chr15	57416215	57416222	+	1.01E-02	RLTR10D
17	chr6	43466689	43466920	+	3.27E-05	RMER6A	87	chr4	1.08E+08	1.08E+08	+	1.05E-02	MMVL30-int
18	chr18	64793437	64793497	-	4.16E-05	MTC	88	chrX	12999375	12999753	+	1.07E-02	RMER19B
19	chr15	10762808	10762885	+	4.50E-05	RLTR15	89	chr3	63045386	63045393	+	1.11E-02	RLTR41
20	chr6	141825999	141826123	+	5.57E-05	MMVL30-int	90	chr6	1.4E+08	1.4E+08	+	1.19E-02	RLTR20D
21	chr9	76274130	76274268	-	5.64E-05	MMVL30-int	91	chr12	86392448	86392577	+	1.23E-02	MYSERV6-int
22	chr17	11097002	11097167	+	6.27E-05	MTD	92	chr18	74452655	74452872	-	1.24E-02	MMVL30-int
23	chr5	58968860	58969002	+	7.26E-05	RLTR10A	93	chr12	19125205	19125336	+	1.36E-02	MMVL30-int
24	chr8	25522830	25522945	+	8.36E-05	RLTR12B	94	chr4	85706515	85706551	-	1.37E-02	MLT1A1
25	chr6	54359293	54359449	-	9.01E-05	LTR81C	95	chr10	17724993	17725182	-	1.49E-02	RLTR15
26	chr2	107805033	107805355	+	9.15E-05	MTEa	96	chr10	40537776	40537847	+	1.58E-02	RMER16-int
27	chr13	104285139	104285262	+	1.10E-04	MTD	97	chr17	63653523	63653555	-	1.63E-02	RMER19B
28	chr6	11138865	11138951	+	1.79E-04	RLTR1B-int	98	chrX	1.47E+08	1.47E+08	+	1.64E-02	MMVL30-int
29	chr7	93987704	93987785	+	2.40E-04	RLTR10A	99	chr2	16813973	16814215	-	1.74E-02	MuRRS4-int
30	chr2	7531849	7531853	+	2.93E-04	RLTR6_Mm	100	chr3	40732912	40733009	-	1.78E-02	RLTR44B
31	chr17	17641623	17641633	+	3.07E-04	RLTR1B	101	chrX	1.13E+08	1.13E+08	+	1.86E-02	MMVL30-int
32	chr2	34689323	34689327	-	3.19E-04	RLTR27	102	chr2	1.72E+08	1.72E+08	+	1.91E-02	RLTR47_MM
33	chr14	14767874	14767925	+	3.28E-04	RLTR1B-int	103	chrX	93450865	93451028	+	2.01E-02	MTEa
34	chr15	36656474	36656694	+	3.31E-04	RMER17C-int	104	chr13	52118311	52118373	+	2.18E-02	RLTR31B_Mm
35	chr5	87007787	87008202	+	3.88E-04	MMVL30-int	105	chr1	54379933	54379982	-	2.20E-02	MMVL30-int
36	chr17	49103178	49103236	-	4.22E-04	ORR1E	106	chr1	84846537	84846631	-	2.30E-02	ORR1D1
37	chr17	50560909	50561112	+	4.65E-04	MuRRS4-int	107	chrX	11901036	11901150	+	2.32E-02	RLTR40
38	chr14	14765074	14765109	+	4.81E-04	RLTR1B	108	chr14	36501981	36501987	+	2.33E-02	RLTR44B
39	chr6	81971172	81971284	-	5.23E-04	MTD	109	chr3	28753069	28753183	-	2.53E-02	MuRRS4-int
40	chr15	41747509	41747753	+	6.10E-04	RMER6C	110	chr17	18695410	18695693	+	2.53E-02	IAPLTR4
41	chr16	9176527	9176567	-	6.39E-04	RLTR47_MM	111	chr4	7622683	7622686	-	2.64E-02	RLTR27
42	chr1	15062947	15062972	+	6.64E-04	RLTR10A	112	chr12	49546462	49546535	-	2.72E-02	RLTR44E
43	chr6	130348617	130348728	+	9.81E-04	RLTR27	113	chr6	37489670	37489792	+	2.78E-02	RLTR18B
44	chr15	47199928	47200124	+	9.99E-04	RLTR1B-int	114	chr9	1.23E+08	1.23E+08	+	2.78E-02	ORR1A2
45	chr12	105615772	105615840	+	1.23E-03	RLTR9E	115	chr1	23170919	23171084	+	2.81E-02	ORR1A4
46	chr19	23360512	23360831	+	1.24E-03	RMER19B	116	chr15	32743923	32743925	-	2.82E-02	RLTR10D
47	chr13	35689821	35690130	-	1.27E-03	RMER19B	117	chr2	23893179	23893221	+	2.84E-02	MuRRS-int
48	chr10	102969968	102969996	+	1.40E-03	RLTR10D	118	chr5	23762526	23762531	+	2.86E-02	RLTR13D3
49	chr19	58975066	58975072	-	1.55E-03	RLTR10D	119	chr2	1.04E+08	1.04E+08	+	2.95E-02	RLTR17
50	chr16	31927072	31927135	+	1.57E-03	RLTR47_MM	120	chr1	1.67E+08	1.67E+08	-	2.96E-02	MTD
51	chr1	31217413	31217422	+	1.59E-03	RMER19B	121	chr17	17644928	17644945	+	2.97E-02	RLTR1B-int
52	chr2	122771825	122771892	+	1.77E-03	LTR88c	122	chr3	1.37E+08	1.37E+08	-	3.05E-02	ORR1E
53	chr5	45686513	45686641	-	1.86E-03	MLT2B2	123	chr5	1.09E+08	1.09E+08	+	3.24E-02	RLTR31A_Mm
54	chrX	83781606	83781690	-	1.93E-03	RLTR10A	124	chr18	11215228	11215313	+	3.27E-02	RLTR20B3
55	chr16	78460645	78460697	+	2.07E-03	MuRRS4-int	125	chr15	74762666	74762737	+	3.42E-02	RLTR6-int
56	chr3	11036473	11036567	+	2.18E-03	RLTR41	126	chr11	66795121	66795128	+	3.45E-02	RLTR31A_Mm
57	chr7	10774537	10774707	+	2.42E-03	RLTR10A	127	chr1	58672684	58672752	-	3.45E-02	ORR1E-int
58	chr11	12566633	12566804	+	2.68E-03	RLTR16	128	chr17	18692758	18692892	+	3.52E-02	RLTR9E
59	chr8	3309684	3309830	-	2.86E-03	RLTR44B	129	chr3	17450800	17450803	+	3.53E-02	RLTR10A
60	chr2	84371467	84371639	-	2.90E-03	MTEa	130	chr15	41057268	41057300	+	3.81E-02	LTRIS_Mus
61	chr3	39689498	39689558	-	3.24E-03	RLTR10A	131	chr5	1.48E+08	1.48E+08	-	3.90E-02	MTEa
62	chr10	31023041	31023233	-	3.36E-03	RLTR47_MM	132	chrX	23411012	23411093	+	3.90E-02	RLTR10A
63	chr9	70082940	70083016	+	3.44E-03	MTD	133	chr1	1.8E+08	1.8E+08	-	4.02E-02	RLTR16
64	chr2	59503383	59503529	+	3.83E-03	LTR33A_	134	chr8	42237597	42237645	-	4.16E-02	RLTR20D
65	chr3	85150410	85150547	+	3.90E-03	RMER19B	135	chr19	10042810	10042861	+	4.16E-02	RLTR25B
66	chr14	14765421	14765438	+	4.03E-03	RLTR1B-int	136	chr18	19012082	19012118	+	4.21E-02	MMERGLN-int
67	chr17	10888438	10888508	-	4.46E-03	MTE2b	137	chrX	1.1E+08	1.1E+08	-	4.30E-02	ORR1A4
68	chr4	118669814	118669906	-	4.56E-03	RLTR44B	138	chr3	1.11E+08	1.11E+08	-	4.85E-02	RLTR10D
69	chr17	75219028	75219367	-	5.46E-03	RMER19C	139	chr12	19846033	19846036	+	4.92E-02	MMERGLN-int
70	chr16	31557436	31557768	-	5.50E-03	RLTR47_MM	140	chr7	10724062	10724232	+	4.98E-02	IAPLTR4

Table S6. The most significantly up-regulated distal peaks (43 LTRs and 90 non-LTRs, Y: tested, N: not tested)

	TYPE	Name	Peak	LTR	Log2FC	FDR	RACE	qPCR	KD
1	LTR	LTR-001	chr13_64649861_64650181_-	MER52A	5.2	2.4E-27	Y	Y	Y
2	LTR	LTR-002	chr6_120326805_120326870_+	THE1D	3.5	1.5E-24	Y	Y	Y
3	LTR	LTR-003	chr4_66107513_66107548_-	MER52A	6.1	2.4E-24	Y	N	N
4	LTR	LTR-004	chr5_104728667_104728706_-	MER52A	6.1	2.9E-21	Y	Y	N
5	LTR	LTR-005	chr7_119547426_119547433_-	THE1A	5.8	1.9E-20	N	N	N
6	LTR	LTR-006	chr4_132300893_132300970_+	LTR12D	6.0	3.5E-20	N	N	N
7	LTR	LTR-007	chr9_31848657_31848713_+	MER61A	5.0	2.6E-19	Y	Y	N
8	LTR	LTR-008	chr9_118506483_118506625_-	MER50B	5.5	1.4E-18	Y	N	N
9	LTR	LTR-009	chrX_97635662_97635698_-	MER61B	5.6	4.2E-18	Y	N	N
10	LTR	LTR-010	chr4_93012597_93012646_-	MER41A	4.9	4.2E-18	N	N	N
11	LTR	LTR-011	chr4_66108031_66108092_-	MER52A	6.1	1.6E-17	N	N	N
12	LTR	LTR-012	chr12_127544982_127545366_+	LTR1	3.6	1.6E-17	Y	N	N
13	LTR	LTR-013	chr5_20616504_20616510_+	THE1B	6.7	3.4E-17	Y	N	N
14	LTR	LTR-014	chr7_53879617_53879660_-	MLT2B3	4.9	7.7E-17	Y	N	N
15	LTR	LTR-015	chr18_35250382_35250447_+	MLT1K	6.0	1.6E-16	N	N	N
16	LTR	LTR-016	chr18_27755926_27756006_-	LTR1	6.3	1.8E-16	Y	N	N
17	LTR	LTR-017	chrX_97828807_97828900_+	LTR49	6.6	2.6E-16	N	N	N
18	LTR	LTR-018	chr13_64651428_64651436_-	LTR12C	6.1	2.8E-16	N	N	N
19	LTR	LTR-019	chr8_112052604_112052671_-	LTR12C	6.2	7.3E-16	Y	N	N
20	LTR	LTR-020	chr13_55628524_55628539_+	LTR17	6.0	4.8E-15	N	N	N
21	LTR	LTR-021	chr6_135196446_135196589_-	LTR18A	5.9	1.4E-14	N	N	N
22	LTR	LTR-022	chr6_115319139_115319436_+	LTR12F	4.1	3.5E-14	N	N	N
23	LTR	LTR-023	chr4_37003352_37003605_+	MER48	4.7	1.2E-13	N	N	N
24	LTR	LTR-024	chr10_96758111_96758263_-	LTR10A	6.1	2.6E-13	N	N	N
25	LTR	LTR-025	chr3_149756876_149757003_+	MER51A	5.5	2.9E-13	N	N	N
26	LTR	LTR-026	chr5_12574724_12574910_-	LTR12C	5.5	3.3E-13	N	N	N
27	LTR	LTR-027	chr7_145482026_145482125_+	MER50B	6.0	3.8E-13	Y	N	N
28	LTR	LTR-028	chr1_238649279_238649353_-	MER61A	4.3	3.9E-13	N	N	N
29	LTR	LTR-029	chr15_75030770_75030854_-	THE1B	6.4	5.1E-13	N	N	N
30	LTR	LTR-030	chr1_248894361_248894428_-	MER50B	4.9	5.3E-13	N	N	N
31	LTR	LTR-031	chr3_172585355_172585370_+	LTR12C	5.5	5.7E-13	N	N	N
32	LTR	LTR-032	chr3_146962969_146962989_-	LTR12C	6.3	6.8E-13	N	N	N
33	LTR	LTR-033	chr5_85866249_85866735_-	LTR2C	4.0	9.5E-13	N	N	N
34	LTR	LTR-034	chr4_31998988_31999006_+	LTR12C	4.7	1.2E-12	Y	N	N
35	LTR	LTR-035	chr6_161637770_161637904_-	THE1A	4.7	1.5E-12	N	N	N
36	LTR	LTR-036	chr11_39515789_39515794_+	MER11A	4.4	6.3E-12	N	N	N
37	LTR	LTR-037	chr17_47152996_47153073_-	MER41-int	4.9	8.3E-12	N	N	N
38	LTR	LTR-038	chr11_26618689_26618774_-	LTR17	5.5	8.7E-12	N	N	N
39	LTR	LTR-039	chr8_98162752_98162840_+	THE1B	5.6	1.0E-11	N	N	N
40	LTR	LTR-040	chrX_120331579_120331716_+	LTR25	3.9	1.5E-11	Y	N	N
41	LTR	LTR-041	chr2_190389114_190389189_-	THE1D	4.1	1.9E-11	N	N	N
42	LTR	LTR-042	chr19_8444480_8444487_+	LTR12D	4.9	1.9E-11	N	N	N
43	LTR	LTR-043	chr21_32097939_32098065_-	MLT2D	4.9	5.0E-11	N	N	N
44	non-LTR	nLTR-044	chr4_19456991_19457094_+		6.6	2.3E-29	N	N	N
45	non-LTR	nLTR-045	chr2_145425502_145425854_+		5.7	8.3E-24	Y	N	N
46	non-LTR	nLTR-046	chr22_48027406_48027559_+		5.7	5.2E-23	N	N	N
47	non-LTR	nLTR-047	chr14_25501764_25501806_-		4.8	3.1E-22	N	N	N
48	non-LTR	nLTR-048	chr10_54515147_54515384_-		4.4	1.4E-21	N	N	N
49	non-LTR	nLTR-049	chr9_11065016_11065169_-		6.1	3.5E-21	N	N	N
50	non-LTR	nLTR-050	chr1_143467553_143467742_-		4.8	9.1E-21	N	N	N
51	non-LTR	nLTR-051	chr4_66864638_66864906_+		4.3	2.0E-20	N	N	N
52	non-LTR	nLTR-052	chr6_23177196_23177334_-		5.5	8.0E-20	Y	N	N
53	non-LTR	nLTR-053	chr13_89815398_89815500_-		6.3	1.2E-19	N	N	N
54	non-LTR	nLTR-054	chr17_70396473_70396561_-		4.8	2.3E-19	Y	N	N
55	non-LTR	nLTR-055	chr4_135248598_135248657_-		5.2	2.3E-19	N	N	N
56	non-LTR	nLTR-056	chr1_192708603_192708714_-		4.6	7.3E-19	Y	N	N
57	non-LTR	nLTR-057	chr10_102880778_102880935_+		3.1	7.7E-19	Y	N	N
58	non-LTR	nLTR-058	chr8_47529222_47529328_-		7.2	8.3E-18	N	N	N
59	non-LTR	nLTR-059	chr8_107072668_107072827_-		3.3	1.4E-17	Y	N	N
60	non-LTR	nLTR-060	chr4_66107613_66107620_+		6.7	1.9E-17	N	N	N
61	non-LTR	nLTR-061	chrX_92951158_92951236_+		6.5	3.4E-17	N	N	N
62	non-LTR	nLTR-062	chr21_11098925_11098996_-		5.8	5.9E-17	N	N	N
63	non-LTR	nLTR-063	chr8_124697229_124697426_-		6.8	1.0E-16	N	N	N
64	non-LTR	nLTR-064	chr19_55845276_55845520_-		3.4	2.0E-16	N	N	N
65	non-LTR	nLTR-065	chr1_156426321_156426375_-		3.7	2.1E-16	N	N	N

66	non-LTR	nLTR-066	chr2_133094884_133094955_+	5.6	3.9E-16	N	N	N
67	non-LTR	nLTR-067	chr12_111497811_111497862_-	6.5	7.8E-16	N	N	N
68	non-LTR	nLTR-068	chr14_54080422_54080745_-	3.5	9.6E-16	Y	N	N
69	non-LTR	nLTR-069	chr2_78768982_78769323_-	5.4	1.3E-15	N	N	N
70	non-LTR	nLTR-070	chr8_49088774_49088921_+	3.7	1.6E-15	N	N	N
71	non-LTR	nLTR-071	chr16_76268983_76269339_+	5.4	1.8E-15	N	N	N
72	non-LTR	nLTR-072	chr5_35237985_35238069_-	4.2	2.3E-15	N	N	N
73	non-LTR	nLTR-073	chr1_202830852_202830914_+	3.5	4.3E-15	N	N	N
74	non-LTR	nLTR-074	chr12_127544890_127544992_-	3.5	6.4E-15	N	N	N
75	non-LTR	nLTR-075	chr7_34797830_34797945_-	4.7	9.6E-15	N	N	N
76	non-LTR	nLTR-076	chr4_188593870_188594125_-	4.8	2.5E-14	N	N	N
77	non-LTR	nLTR-077	chr1_147727218_147727318_-	5.7	3.5E-14	N	N	N
78	non-LTR	nLTR-078	chr3_147088170_147088274_-	6.3	6.6E-14	N	N	N
79	non-LTR	nLTR-079	chr22_44242951_44243047_-	5.2	1.8E-13	N	N	N
80	non-LTR	nLTR-080	chr12_70944080_70944102_+	6.4	2.0E-13	N	N	N
81	non-LTR	nLTR-081	chrX_149115424_149115504_+	5.5	2.6E-13	N	N	N
82	non-LTR	nLTR-082	chr6_124669729_124669754_+	6.1	2.9E-13	N	N	N
83	non-LTR	nLTR-083	chr15_66694188_66694236_+	3.7	3.1E-13	N	N	N
84	non-LTR	nLTR-084	chr5_3536011_3536189_-	6.0	3.3E-13	N	N	N
85	non-LTR	nLTR-085	chr1_170143736_170143807_+	5.5	3.8E-13	N	N	N
86	non-LTR	nLTR-086	chr1_246339776_246339914_-	5.1	4.6E-13	N	N	N
87	non-LTR	nLTR-087	chr6_105384844_105384950_+	5.2	5.0E-13	N	N	N
88	non-LTR	nLTR-088	chr7_64734221_64734308_+	4.3	5.6E-13	N	N	N
89	non-LTR	nLTR-089	chr13_110242600_110242753_+	4.0	5.7E-13	N	N	N
90	non-LTR	nLTR-090	chr3_4941365_4941578_-	6.1	5.9E-13	N	N	N
91	non-LTR	nLTR-091	chr4_115068285_115068322_+	4.6	5.9E-13	N	N	N
92	non-LTR	nLTR-092	chr5_173155159_173155203_-	4.4	6.5E-13	N	N	N
93	non-LTR	nLTR-093	chr4_44018712_44018944_+	4.9	7.4E-13	N	N	N
94	non-LTR	nLTR-094	chr10_10836846_10836976_-	3.2	9.6E-13	N	N	N
95	non-LTR	nLTR-095	chr13_68831761_68831830_-	5.9	9.8E-13	N	N	N
96	non-LTR	nLTR-096	chr16_14580591_14580873_+	3.1	1.1E-12	N	N	N
97	non-LTR	nLTR-097	chrX_7511587_7511814_+	5.8	1.3E-12	N	N	N
98	non-LTR	nLTR-098	chr1_145390399_145390478_-	4.3	1.3E-12	N	N	N
99	non-LTR	nLTR-099	chr2_210359776_210360229_+	3.4	1.4E-12	N	N	N
100	non-LTR	nLTR-100	chr5_59837919_59838042_-	4.4	1.9E-12	N	N	N
101	non-LTR	nLTR-101	chr14_28142441_28142491_-	5.3	3.1E-12	N	N	N
102	non-LTR	nLTR-102	chr12_74686399_74686456_-	5.3	4.7E-12	N	N	N
103	non-LTR	nLTR-103	chr3_145829119_145829280_-	4.0	5.6E-12	N	N	N
104	non-LTR	nLTR-104	chr8_36636633_36636738_-	5.3	5.6E-12	N	N	N
105	non-LTR	nLTR-105	chr16_79018417_79018492_+	3.2	7.8E-12	N	N	N
106	non-LTR	nLTR-106	chr1_155618329_155618374_-	4.0	8.6E-12	N	N	N
107	non-LTR	nLTR-107	chr1_201874446_201874513_-	4.9	1.1E-11	N	N	N
108	non-LTR	nLTR-108	chr11_35425907_35426227_-	3.6	1.1E-11	N	N	N
109	non-LTR	nLTR-109	chr12_103292981_103293117_+	3.6	1.1E-11	N	N	N
110	non-LTR	nLTR-110	chr5_557071_557102_-	3.5	1.6E-11	N	N	N
111	non-LTR	nLTR-111	chr16_12713528_12713697_-	4.2	1.8E-11	N	N	N
112	non-LTR	nLTR-112	chr12_34200857_34200951_-	5.1	1.9E-11	N	N	N
113	non-LTR	nLTR-113	chr1_244256324_244256377_-	4.6	2.0E-11	N	N	N
114	non-LTR	nLTR-114	chr12_113644190_113644252_-	4.4	2.1E-11	N	N	N
115	non-LTR	nLTR-115	chr1_237634114_237634543_+	4.3	2.3E-11	N	N	N
116	non-LTR	nLTR-116	chr12_97856596_97856613_+	3.5	2.3E-11	N	N	N
117	non-LTR	nLTR-117	chr19_22966865_22966986_-	3.1	2.5E-11	N	N	N
118	non-LTR	nLTR-118	chr3_150454151_150454451_-	3.6	2.9E-11	N	N	N
119	non-LTR	nLTR-119	chr13_94958159_94958162_+	3.8	3.1E-11	N	N	N
120	non-LTR	nLTR-120	chrX_21788738_21788791_-	3.4	3.2E-11	N	N	N
121	non-LTR	nLTR-121	chr12_128108501_128108573_+	4.7	3.5E-11	N	N	N
122	non-LTR	nLTR-122	chrX_49558230_49558478_+	4.3	3.6E-11	N	N	N
123	non-LTR	nLTR-123	chrX_27417076_27417151_-	4.2	3.9E-11	N	N	N
124	non-LTR	nLTR-124	chr19_29492135_29492199_+	5.3	4.3E-11	N	N	N
125	non-LTR	nLTR-125	chr7_25471788_25471824_-	4.9	5.0E-11	N	N	N
126	non-LTR	nLTR-126	chr1_117008795_117008884_+	3.3	5.1E-11	N	N	N
127	non-LTR	nLTR-127	chr8_144490327_144490432_+	5.1	5.2E-11	N	N	N
128	non-LTR	nLTR-128	chr6_25920064_25920202_+	4.5	5.9E-11	N	N	N
129	non-LTR	nLTR-129	chr10_71850872_71851017_+	3.9	6.1E-11	N	N	N
130	non-LTR	nLTR-130	chr1_187929553_187929791_+	4.2	6.8E-11	N	N	N
131	non-LTR	nLTR-131	chr5_55354832_55354890_+	4.9	6.8E-11	N	N	N
132	non-LTR	nLTR-132	chr3_39839148_39839168_-	4.6	6.9E-11	N	N	N
133	non-LTR	nLTR-133	chr4_69922667_69922730_-	4.3	7.6E-11	N	N	N

Table S7. Targets of 3' RACE validation (15 LTRs and 7 non-LTRs)

Number	Name	TSS	Gencode 19	Primer1	Primer2	# Transcripts
1	LTR-001	chr13_64649861_64650181_-	ENST00000456627.1	CACTGTGGGCTTCGTCTCTG	GACGCCCTGCCTATGGA AAC	7
2	LTR-002	chr6_120326805_120326870_+	Not annotated	CTTGCTGCTTCTTCGCTTTC	TCGCTTTCTGCCATGATTGC	4
3	LTR-003	chr4_66107513_66107548_-	Not annotated	CCGTGTACCTCATTCTTGTC	ACCCTGGGACCCAGTGAATG	2
4	LTR-004	chr5_104728667_104728706_-	ENST00000503650.1	CGTTAGGAGCTACCGTCTTC	TCTTCTCCACCGTGCTCACC	1
5	LTR-007	chr9_31848657_31848713_+	Not annotated	CCGGGTCTTACTCATCATT	TTTGTGGCCGTGGGACAAAG	2
6	LTR-008	chr9_118506483_118506625_-	ENST00000433546.2	CTTGCTTCCCAAAGCTAAGG	AAGCTAAGGAGGGCTCTGTG	1
7	LTR-009	chrX_97635662_97635698_-	Not annotated	GAAACCTGCCAGCAGGACTC	TTCTGTGGGTTCTGTTTCC	3
8	LTR-012	chr12_127544982_127545366_+	Not annotated	CCCAGCCTGGGCCTATAAAC	ACGACCCTAGCAAGGCAGAG	0
9	LTR-013	chr5_20616504_20616510_+	ENST00000512688.1	TTTGCTCCTGCTTCGTCTTC	CGTCTTCTGCCGTGATTGTG	3
10	LTR-014	chr7_53879617_53879660_-	ENST00000380970.2	TCTTACTGTGCTGGGCGGTT	GAGTCAGACTCCAGGTTCTTC	4
11	LTR-016	chr18_27755926_27756006_-	Not annotated	ATCCGAGACCCTAGCAAGAC	AGAAACGGAAGCTGCTGGAC	1
12	LTR-019	chr8_112052604_112052671_-	Not annotated	TCATTCCCGAAGTCAGTGAG	CCAATTCCGGGCACGGTATG	1
13	LTR-027	chr7_145482026_145482125_+	Not annotated	TGTGCCTAGGTCTTGTTC	AGATTTAAAGTTGCGGCCGAC	1
14	LTR-034	chr4_31998988_31999006_+	ENST00000513211.1	GGAGGAAGGAACA ACTACAG	TTCACTCTTGAGCCAGAGAC	2
15	LTR-040	chrX_120331579_120331716_+	Not annotated	CCCACAGTGAGTATTCCTGTT	CTCTACTGACAGAGAGCTTTC	1
16	nLTR-045	chr2_145425502_145425854_+	ENST00000597655.1	GGTCTGAAGGGGAAGAAAAGC	TGTTGAGATCTGGCGACTTG	2
17	nLTR-052	chr6_23177196_23177334_-	Not annotated	GACACTTCGAGGCTCCATTG	CATGCGGCAAATAGGAAACC	2
18	nLTR-054	chr17_70396473_70396561_-	ENST00000453722.2	ATGCCAGGCTGGTGGAGAAC	GCCTGAAGACAGCCCTAAAC	2
19	nLTR-056	chr1_192708603_192708714_-	Not annotated	TTTCTGCCTCCAGACTCTAC	CTACTCTCCTGCCTCAATTC	1
20	nLTR-057	chr10_102880778_102880935_+	Not annotated	CCTCCCAGCAGTGTGGATT	GAGGTAAGGGTGCCGATGAG	1
21	nLTR-059	chr8_107072668_107072827_-	ENST00000520433.1	CGCATTCTACTTTTCTAC	CCAGCAGAAATTCCTGTACAC	4
22	nLTR-068	chr14_54080422_54080745_-	ENST00000436530.1	TGATGAAAGTCAGCCTAGGG	CAGAGACCAAAGGGCCATAC	1

Table S8. Full-length ncRNAs determined by 3' RACE

Number	Name	chr	start	end	Exon size	strand	# Exon	Exon length	Exon start position	id
1	LTR-001	chr13	64425460	64650144	1101	-	2	331,770	0,223914	T1.1
2	LTR-001	chr13	64425460	64650144	1176	-	3	331,75,770	0,2357,223914	T1.2
3	LTR-001	chr13	64560503	64650144	1878	-	6	669,105,25,118,191,770	0,3725,16014,16977,47976,88871	T1.3
4	LTR-001	chr13	64560503	64650144	1687	-	5	669,105,25,118,770	0,3725,16014,16977,88871	T1.4
5	LTR-001	chr13	64560503	64650144	1735	-	4	669,105,191,770	0,3725,47976,88871	T1.5
6	LTR-001	chr13	64560503	64650144	1569	-	4	669,105,25,770	0,3725,16014,88871	T1.6
7	LTR-001	chr13	64560503	64650144	1544	-	3	669,105,770	0,3725,88871	T1.7
8	LTR-002	chr6	120326868	121027885	525	+	3	94,81,350	0,456788,700667	T3.1
9	LTR-002	chr6	120326868	120892061	313	+	3	94,80,139	0,456788,565054	T3.2
10	LTR-002	chr6	120326868	121027885	600	+	4	94,75,81,350	0,122844,456788,700667	T3.3
11	LTR-002	chr6	120326868	121027885	650	+	5	94,75,50,81,350	0,122844,345860,456788,700667	T3.4
12	LTR-003	chr4	66072125	66107533	1242	-	4	376,151,81,634	0,19452,31976,34774	T2.1
13	LTR-003	chr4	66103988	66107533	828	-	2	194,634	0,2911	T2.2
14	LTR-004	chr5	104253029	104728668	971	-	4	570,143,51,207	0,38246,91532,475432	T5.1
15	LTR-007	chr9	31848669	32112835	1002	+	3	236,98,668	0,124343,263498	T10.1
16	LTR-007	chr9	31848669	31977997	839	+	3	236,98,505	0,124343,128823	T10.2
17	LTR-008	chr9	118403718	118506541	749	-	3	542,101,106	0,100022,102717	T15.1
18	LTR-009	chrX	97009741	97635670	581	-	3	273,98,210	0,597260,625719	T18.1
19	LTR-009	chrX	97288808	97635670	616	-	4	149,159,98,210	0,154631,318193,346652	T18.2
20	LTR-009	chrX	97418996	97635670	771	-	5	201,103,159,98,210	0,834,24443,188005,216464	T18.3
21	LTR-013	chr5	20616506	20936877	1257	+	8	83,57,147,61,85,98,108,618	0,44314,120557,191455,279085,316485,318534,319753	T13.1
22	LTR-013	chr5	20616506	20899593	585	+	6	83,57,147,61,85,152	0,44314,120557,191455,279085,282935	T13.2
23	LTR-013	chr5	20616506	20737349	426	+	3	83,57,286	0,44314,120557	T13.3
24	LTR-014	chr7	53724161	53879642	1991	-	6	980,163,174,141,91,442	0,35458,114965,120019,123477,155039	T16.1
25	LTR-014	chr7	53724161	53879642	1676	-	4	980,163,91,442	0,35458,123477,155039	T16.2
26	LTR-014	chr7	53758873	53879642	1757	-	5	909,174,141,91,442	0,80253,85307,88765,120327	T16.3
27	LTR-014	chr7	53838091	53879642	1883	-	4	1209,141,91,442	0,6089,9547,41109	T16.4
28	LTR-016	chr18	27737296	27755981	546	-	2	270,276	0,18409	T17.1
29	LTR-019	chr8	111995211	112052619	992	-	7	158,37,109,120,128,154,286	0,20431,21678,22547,44323,56106,57122	T21.1
30	LTR-027	chr7	145482073	145545509	668	+	3	453,94,121	0,57510,63315	T19.1
31	LTR-034	chr4	31998999	32157445	1170	+	3	445,215,510	0,148675,157936	T22.1
32	LTR-034	chr4	31998999	32157445	955	+	2	445,510	0,157936	T22.2
33	LTR-040	chrX	120331632	120332713	1081	+	1	1081	0	T12.1
34	nLTR-045	chr2	145425550	145940216	888	+	4	125,183,134,446	0,100114,504014,514220	T4.1
35	nLTR-045	chr2	145425550	145612387	532	+	4	125,183,105,119	0,100114,185299,186718	T4.2
36	nLTR-052	chr6	23056524	23177316	1052	-	3	771,86,195	0,118957,120597	T14.1
37	nLTR-052	chr6	23175151	23177316	611	-	2	416,195	0,1970	T14.2
38	nLTR-054	chr17	70319276	70396542	1012	-	3	185,590,237	0,59907,77029	T7.1
39	nLTR-054	chr17	70319276	70396542	576	-	3	185,154,237	0,2657,77029	T7.2
40	nLTR-056	chr1	192460585	192708626	830	-	5	514,70,59,81,106	0,75691,102143,135885,247935	T9.1
41	nLTR-057	chr10	102880846	102897510	1035	+	2	329,706	0,15958	T6.1
42	nLTR-059	chr8	106797034	107072753	1044	-	6	411,155,148,118,67,145	0,2260,2642,13518,142844,275574	T8.1
43	nLTR-059	chr8	106797034	107072753	889	-	5	411,148,118,67,145	0,2642,13518,142844,275574	T8.2
44	nLTR-059	chr8	106673911	107072753	853	-	6	285,169,69,118,67,145	0,1089,69410,136641,265967,398697	T8.3
45	nLTR-059	chr8	106697116	107072753	418	-	5	38,50,118,67,145	0,856,113436,242762,375492	T8.4
46	nLTR-068	chr14	54074027	54080669	614	-	2	315,299	0,6343	T11.1



Table S9. Prediction of coding potential using CPAT web system

Number	Name	Size (bp)	Exon	Genomic Size	Coding Probability	Coding Label
1	T1.1	1101	2	224684	0.07	no
2	T1.2	1176	3	224684	0.02	no
3	T1.3	1878	6	89641	0.04	no
4	T1.4	1687	5	89641	0.05	no
5	T1.5	1735	4	89641	0.03	no
6	T1.6	1569	4	89641	0.03	no
7	T1.7	1544	3	89641	0.03	no
8	T2.1	1242	4	35408	0.02	no
9	T2.2	828	2	3545	0.02	no
10	T3.1	525	3	701017	0.02	no
11	T3.2	313	3	565193	0.02	no
12	T3.3	600	3	701017	0.01	no
13	T3.4	650	3	701017	0.12	no
14	T4.1	888	4	514666	0.08	no
15	T4.2	532	4	186837	0.05	no
16	T5.1	971	4	475639	0.02	no
17	T6.1	1035	2	16664	0.03	no
18	T7.1	1012	3	77266	0.00	no
19	T7.2	576	3	77266	0.00	no
20	T8.1	1044	6	275719	0.02	no
21	T8.2	889	5	275719	0.02	no
22	T8.3	853	6	398842	0.02	no
23	T8.4	418	5	375637	0.02	no
24	T9.1	830	5	248041	0.01	no
25	T10.1	1002	3	264166	0.02	no
26	T10.2	839	3	129328	0.01	no
27	T11.1	614	2	6642	0.03	no
28	T12.1	1081	1	1081	0.00	no
29	T13.1	1257	8	320371	0.00	no
30	T13.2	585	6	283087	0.00	no
31	T13.3	426	3	120843	0.00	no
32	T14.1	1052	3	120792	0.00	no
33	T14.2	611	2	2165	0.00	no
34	T15.1	749	3	102823	0.02	no
35	T16.1	1991	6	155481	0.17	no
36	T16.2	1676	4	155481	0.17	no
37	T16.3	1757	5	120769	0.17	no
38	T16.4	1883	4	41551	0.17	no
39	T17.1	546	2	18685	0.06	no
40	T18.1	581	3	625929	0.02	no
41	T18.2	616	4	346862	0.00	no
42	T18.3	771	5	216674	0.00	no
43	T19.1	668	3	63436	0.05	no
44	T21.1	992	7	57408	0.00	no
45	T22.1	1170	3	158446	0.01	no
46	T22.2	955	2	158446	0.00	no

Table S10. qPCR primers for LTR-derived ncRNAs

Number	Target	Experiments	Forward	Reverse
1	LTR-001	qPCR / LNA KD	TTGAAACATGCCCTTGCTC	AGAAAGGGTGTACAGTCTTGG
2	LTR-002	LNA KD	TCGCTTTCTGCCATGATTGC	TTCCTGGCCCAGACTTCAATC
3	LTR-002	qPCR	GCAGTCCATGGAGCAGTTACA	CCCTGCATCGAGCAAGTCTA
4	LTR-004	qPCR	AGCTTTTTCTCGCTGCTGAG	AGCGTTACAGCCCTTTTTGC
5	LTR-007	qPCR	AGCCCTGCAACAATTTTGGC	TTTCAAAGGTCCCCAACTCG
6	HMBS	qPCR	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC
7	GAPDH	LNA KD	GAAATCCCATCACCATCTTCCAGG	GAGCCCCAGCCTTCTCCATG

Table S11. Clinical features associated with the high LTR class

Study	high-LTR	low-LTR	Odds	95% CI
Gender (Male)	14/16	10/14	2.8	0.427 to 18.376
Age (≥55 years)	12/16	5/14	5.4	1.120 to 26.045
Viral etiology	14/16	7/14	7	1.140 to 42.971
Cirrhosis	8/16	10/14	0.4	0.0876 to 1.826
AFP (≥50 ng/ml)	8/15	4/12	2.286	0.475 to 11.003
Pre-operative treatment	4/16	9/14	0.185	0.0384 to 0.893
Liver transplantation	4/16	11/14	0.0909	0.0165 to 0.501
Multinodular HCC	6/16	12/14	0.1	0.0164 to 0.609
Tumor size (≥50 mm)	5/16	4/14	1.136	0.237 to 5.457
Satellite nodules	6/16	2/14	3.6	0.591 to 21.932
Microvascular invasion	7/16	3/14	2.852	0.568 to 14.326
Recurrence	10/16	2/14	10	1.641 to 60.923
Metastasis	4/16	1/14	4.333	0.423 to 44.430
Differentiation - Well	5/16	11/14	0.124	0.0236 to 0.650
Differentiation - Moderate	10/16	2/14	10	1.641 to 60.923
Differentiation - Poor	1/16	1/14	0.867	0.0492 to 15.280
Total (fixed effects)	118/255	93/222	1.199	0.830 to 1.732
Total (random effects)	118/255	93/222	1.355	0.586 to 3.134

Table S12. ENCODE data used to measure activities of TSSs in HepG2

Number	Technology	Definition	UCSC Accession	FileName
1	CAGE	cell PolyA+ rep1	wgEncodeEH001149	wgEncodeRikenCageHepg2CellPapAlnRep1.bam
2	CAGE	cell PolyA+ rep2	wgEncodeEH001149	wgEncodeRikenCageHepg2CellPapAlnRep2.bam
3	CAGE	nucleus PolyA+ rep1	wgEncodeEH001148	wgEncodeRikenCageHepg2NucleusPapAlnRep1.bam
4	CAGE	nucleus PolyA+ rep2	wgEncodeEH001148	wgEncodeRikenCageHepg2NucleusPapAlnRep2.bam
5	CAGE	nucleus PolyA-	wgEncodeEH000345	wgEncodeRikenCageHepg2NucleusPamAln.bam
6	CAGE	cytosol PolyA+ rep1	wgEncodeEH001147	wgEncodeRikenCageHepg2CytosolPapAlnRep1.bam
7	CAGE	cytosol PolyA+ rep2	wgEncodeEH001147	wgEncodeRikenCageHepg2CytosolPapAlnRep2.bam
8	CAGE	cytosol PolyA-	wgEncodeEH000343	wgEncodeRikenCageHepg2CytosolPamAln.bam
9	CAGE	nucleolus	wgEncodeEH000344	wgEncodeRikenCageHepg2NucleolusTotalAln.bam
10	DNase-Seq	narrowPeak rep1	wgEncodeEH000482	wgEncodeUwDnaseHepg2PkRep1.narrowPeak.gz
11	DNase-Seq	narrowPeak rep2	wgEncodeEH000482	wgEncodeUwTfbsHepg2CtcfStdPkRep2.narrowPeak.gz
12	FAIR-Seq	narrowPeak	wgEncodeEH000546	wgEncodeOpenChromFaireHepg2Pk.narrowPeak.gz
13	ChIP-Seq	RNA pol II	wgEncodeEH002278	wgEncodeHaibTfbsHepg2Pol24h8V0416102PkRep1.broadPeak.gz
14	ChIP-Seq	H3K4me3	wgEncodeEH000413	wgEncodeUwHistoneHepg2H3k4me3StdPkRep1.narrowPeak.gz
15	ChIP-Seq	H3K27ac	wgEncodeEH000094	wgEncodeBroadHistoneHepg2H3k27acStdPk.broadPeak.gz
16	ChIP-Seq	chromatin state	wgEncodeEH000789	wgEncodeBroadHmmHepg2HMM.bed.gz

Table S13. ENCODE ChIP-Seq data used to measure TF activities in HepG2

Number	Name	GEO accession	UCSC Accession	FileName
1	ARID3A	GSM935275	wgEncodeEH002858	wgEncodeSydhTfbsHepg2Arid3anb100279lggrabPk.narrowPeak.gz
2	ATF3	GSM803503	wgEncodeEH001568	wgEncodeHaibTfbsHepg2Atf3V0416101PkRep1.broadPeak.gz
3	BHLHE40	GSM935566	wgEncodeEH002032	wgEncodeSydhTfbsHepg2Bhlhe40clggrabPk.narrowPeak.gz
4	BRCA1	GSM935609	wgEncodeEH001859	wgEncodeSydhTfbsHepg2Brca1a300lggrabPk.narrowPeak.gz
5	CEBPB	GSM1010778	wgEncodeEH002304	wgEncodeHaibTfbsHepg2Cebpbsc150V0416101PkRep2.broadPeak.gz
6	CEBPD	GSM1010777	wgEncodeEH002305	wgEncodeHaibTfbsHepg2Cebpdsc636V0416101PkRep1.broadPeak.gz
7	CHD2	GSM935307	wgEncodeEH001840	wgEncodeSydhTfbsHepg2Chd2ab68301lggrabPk.narrowPeak.gz
8	cJUN	GSM935364	wgEncodeEH001794	wgEncodeSydhTfbsHepg2CjunlggrabPk.narrowPeak.gz
9	cMYC	GSM822291	wgEncodeEH000545	wgEncodeOpenChromChipHepg2CmycPk.narrowPeak.gz
10	CREB1	GSM1010808	wgEncodeEH003312	wgEncodeHaibTfbsHepg2Creb1sc240V0422111PkRep1.broadPeak.gz
11	CTCF	GSM733645	wgEncodeEH000080	wgEncodeBroadHistoneHepg2CtcfStdPk.broadPeak.gz
12	ELF1	GSM803415	wgEncodeEH001641	wgEncodeHaibTfbsHepg2Elf1sc631V0416101PkRep2.broadPeak.gz
13	EZH2	GSM1003487	wgEncodeEH002416	wgEncodeBroadHistoneHepg2Ezh239875Pk.broadPeak.gz
14	FOSL2	GSM803451	wgEncodeEH001501	wgEncodeHaibTfbsHepg2Fosl2V0416101PkRep1.broadPeak.gz
15	FOXA1	GSM803432	wgEncodeEH001630	wgEncodeHaibTfbsHepg2Foxa1sc101058V0416101PkRep1.broadPeak.gz
16	FOXA2	GSM803403	wgEncodeEH001650	wgEncodeHaibTfbsHepg2Foxa2sc6554V0416101PkRep2.broadPeak.gz
17	GABPA	GSM803343	wgEncodeEH001548	wgEncodeHaibTfbsHepg2GabpPcr2xPkRep1.broadPeak.gz
18	HDAC2	GSM803493	wgEncodeEH001618	wgEncodeHaibTfbsHepg2Hdac2sc6296V0416101PkRep1.broadPeak.gz
19	HNF4A	GSM803460	wgEncodeEH001595	wgEncodeHaibTfbsHepg2Hnf4asc8987V0416101PkRep1.broadPeak.gz
20	HNF4G	GSM803404	wgEncodeEH001651	wgEncodeHaibTfbsHepg2Hnf4gsc6558V0416101PkRep1.broadPeak.gz
21	JUND	GSM803500	wgEncodeEH001470	wgEncodeHaibTfbsHepg2JundPcr1xPkRep2.broadPeak.gz
22	MAFF	GSM935306	wgEncodeEH001841	wgEncodeSydhTfbsHepg2Maffm8194lggrabPk.narrowPeak.gz
23	MAFK	GSM935305	wgEncodeEH001842	wgEncodeSydhTfbsHepg2Mafkab50322lggrabPk.narrowPeak.gz
24	MAX	GSM935406	wgEncodeEH002796	wgEncodeSydhTfbsHepg2MaxlggrabPk.narrowPeak.gz
25	MAZ	GSM935335	wgEncodeEH002860	wgEncodeSydhTfbsHepg2Mazab85725lggrabPk.narrowPeak.gz
26	MBD4	GSM1010740	wgEncodeEH002318	wgEncodeHaibTfbsHepg2Mbd4sc271530V0422111PkRep1.broadPeak.gz
27	MXI1	GSM935437	wgEncodeEH002029	wgEncodeSydhTfbsHepg2Mxi1StdPk.narrowPeak.gz
28	NR2F2	GSM1010810	wgEncodeEH003314	wgEncodeHaibTfbsHepg2Nr2f2sc271940V0422111PkRep2.broadPeak.gz
29	NRF1	GSM935648	wgEncodeEH001802	wgEncodeSydhTfbsHepg2Nrf1lggrabPk.narrowPeak.gz
30	NRSF	GSM1010784	wgEncodeEH002263	wgEncodeHaibTfbsHepg2NrsfV0416101PkRep2.broadPeak.gz
31	P300	GSM803499	wgEncodeEH001612	wgEncodeHaibTfbsHepg2P300V0416101PkRep1.broadPeak.gz
32	RAD21	GSM935647	wgEncodeEH001803	wgEncodeSydhTfbsHepg2Rad21lggrabPk.narrowPeak.gz
33	RCOR1	GSM935579	wgEncodeEH002845	wgEncodeSydhTfbsHepg2Corestsc30189lggrabPk.narrowPeak.gz
34	RXRA	GSM803452	wgEncodeEH001506	wgEncodeHaibTfbsHepg2RxraPcr1xPkRep2.broadPeak.gz
35	SIN3AK20	GSM803530	wgEncodeEH001471	wgEncodeHaibTfbsHepg2Sin3ak20Pcr1xPkRep2.broadPeak.gz
36	SMC3	GSM935542	wgEncodeEH001861	wgEncodeSydhTfbsHepg2Smc3ab9263lggrabPk.narrowPeak.gz

37	SP1	GSM803507	wgEncodeEH001561	wgEncodeHaibTfbsHepg2Sp1Pcr1xPkRep1.broadPeak.gz
38	SP2	GSM1010787	wgEncodeEH002264	wgEncodeHaibTfbsHepg2Sp2V0422111PkRep1.broadPeak.gz
39	SRF	GSM803502	wgEncodeEH001611	wgEncodeHaibTfbsHepg2SrfV0416101PkRep2.broadPeak.gz
40	TAF1	GSM803367	wgEncodeEH001551	wgEncodeHaibTfbsHepg2Taf1Pcr2xPkRep1.broadPeak.gz
41	TBP	GSM935280	wgEncodeEH001793	wgEncodeSydhTfbsHepg2TbplggrabPk.narrowPeak.gz
42	TCF12	GSM803336	wgEncodeEH001544	wgEncodeHaibTfbsHepg2Tcf12Pcr1xPkRep1.broadPeak.gz
43	TCF7L2	GSM782122	wgEncodeEH001780	wgEncodeSydhTfbsHepg2Tcf7l2UcdPk.narrowPeak.gz
44	TEAD4	GSM1010875	wgEncodeEH002345	wgEncodeHaibTfbsHepg2Tead4sc101184V0422111PkRep1.broadPeak.gz
45	TR4	GSM935596	wgEncodeEH000675	wgEncodeSydhTfbsHepg2Tr4UcdPk.narrowPeak.gz
46	USF1	GSM803527	wgEncodeEH001472	wgEncodeHaibTfbsHepg2Usf1Pcr1xPkRep2.broadPeak.gz
47	USF2	GSM935646	wgEncodeEH001804	wgEncodeSydhTfbsHepg2Usf2lggrabPk.narrowPeak.gz
48	YY1	GSM803381	wgEncodeEH001661	wgEncodeHaibTfbsHepg2Yy1sc281V0416101PkRep1.broadPeak.gz
49	ZBTB33	GSM803449	wgEncodeEH001503	wgEncodeHaibTfbsHepg2Zbtb33Pcr1xPkRep1.broadPeak.gz
50	ZBTB7A	GSM1010831	wgEncodeEH002296	wgEncodeHaibTfbsHepg2Zbtb7aV0416101PkRep2.broadPeak.gz
51	ZEB1	GSM1010809	wgEncodeEH003315	wgEncodeHaibTfbsHepg2Zeb1V0422111PkRep1.broadPeak.gz

---

Table S14. Targets and sequences of LNAs

Number	Name	Target	LNA sequence 5'-3'
1	L1	LTR-001	GCGAGTAAGCAGAAGA
2	L3	LTR-001	AAATGGAGGGCGAGTA
3	L4	LTR-001	TGTCCAGGATGAATGA
4	L5	LTR-002	CTTCACAACGCGGTA
5	L6	LTR-002	CCCACAATTACCGTTG

Table S15. Significantly dysregulated genes with LTR-001 KD at 48h

Number	TYPE	TSS	Log2(FC)	FDR	Name
1	DOWN	chr10_77191031_77191479_+	-1.2	1.8E-09	C10ORF11
2	DOWN	chr13_64650089_64650181_-	-2.9	2.2E-08	LTR-001
3	DOWN	chr4_19456965_19457245_+	-1.5	7.1E-07	RP11-608O21.1
4	DOWN	chr6_56819437_56819862_-	-1.3	2.5E-05	DST
5	DOWN	chr3_61547207_61547322_+	-1.4	5.2E-05	PTPRG
6	DOWN	chr5_60240918_60241225_+	-0.7	4.3E-04	MIMITIN
7	DOWN	chr16_15528316_15528473_+	-1.4	2.7E-03	C16ORF45
8	DOWN	chr17_43971780_43972028_+	-0.8	3.2E-03	MAPT
9	DOWN	chr1_246670461_246670642_-	-0.8	4.2E-03	SMYD3
10	DOWN	chr8_9911725_9911878_+	-1.2	4.4E-03	MSRA
11	DOWN	chr6_52668605_52668722_-	-0.8	6.6E-03	GSTA1
12	DOWN	chr13_93879056_93879322_+	-0.8	7.5E-03	GPC6
13	DOWN	chr6_5261619_5261844_+	-0.8	8.9E-03	FARS2
14	DOWN	chr8_58906933_58907302_+	-1.1	1.4E-02	FAM110B
15	DOWN	chrX_33760071_33760534_+	-1.3	1.6E-02	RP11-305F18.1
16	DOWN	chr17_64298632_64298764_+	-0.6	1.7E-02	PRKCA
17	DOWN	chr6_144329324_144329589_-	-1.1	1.7E-02	HYMAI
18	DOWN	chr12_4713988_4714206_+	-0.6	1.9E-02	DYRK4
19	DOWN	chr6_120326805_120326902_+	-0.6	2.3E-02	-
20	DOWN	chr7_101458933_101459327_+	-0.8	2.7E-02	CUX1
21	DOWN	chr11_8615481_8615854_-	-0.8	3.1E-02	STK33
22	DOWN	chr13_112025173_112025276_+	-1.1	3.1E-02	-
23	DOWN	chr4_74281767_74281993_+	-0.6	3.1E-02	ALB
24	DOWN	chr13_43597508_43597846_+	-0.5	3.3E-02	DNAJC15
25	DOWN	chr20_18568521_18568735_+	-1.0	3.9E-02	DTD1
26	DOWN	chr5_78281472_78281476_-	-2.5	4.1E-02	ARSB
27	DOWN	chr13_93879552_93879586_+	-1.1	4.2E-02	GPC6
28	DOWN	chr8_81786428_81786911_-	-0.7	4.3E-02	ZNF704
29	DOWN	chr5_43313227_43313616_-	-0.6	4.3E-02	HMGCS1
30	DOWN	chr3_126423023_126423364_+	-0.5	4.6E-02	CHCHD6
31	DOWN	chr19_13106214_13106424_+	-0.8	4.6E-02	NFIX
32	DOWN	chr18_32073222_32073328_+	-1.1	4.6E-02	DTNA
33	DOWN	chr3_99536634_99536855_+	-0.6	4.6E-02	C3ORF26
34	DOWN	chr3_13324361_13324814_-	-0.7	4.6E-02	-
35	DOWN	chr20_36322307_36322533_+	-0.5	4.7E-02	CTNBL1
36	UP	chr7_45927983_45928256_+	2.1	1.1E-04	IGFBP1
37	UP	chr16_24548998_24549082_+	1.2	2.2E-03	RBBP6
38	UP	chr11_67120745_67121095_-	0.8	5.8E-03	POLD4
39	UP	chr10_114154436_114154612_+	1.2	5.8E-03	ACSL5
40	UP	chr8_117658735_117658838_-	4.1	6.6E-03	EIF3H
41	UP	chr12_7052070_7052116_+	1.6	7.5E-03	AK310266
42	UP	chr2_192542809_192542896_+	0.9	8.9E-03	DKFZP313C2412
43	UP	chr10_102046091_102046118_-	1.3	1.5E-02	BLOC1S2
44	UP	chr22_35817451_35817454_+	1.7	2.1E-02	MCM5
45	UP	chr1_151804257_151804322_-	2.5	2.1E-02	RORC
46	UP	chr1_110879128_110879228_-	3.0	2.1E-02	-
47	UP	chr12_53343796_53343854_+	1.0	2.4E-02	KRT18
48	UP	chr11_18258262_18258666_-	0.9	2.4E-02	SAA4
49	UP	chr9_140119596_140119649_+	1.5	2.4E-02	C9ORF169
50	UP	chr1_53556536_53556699_-	1.2	2.7E-02	SLC1A7
51	UP	chrX_70331389_70331494_-	1.6	2.7E-02	IL2RG
52	UP	chr17_74547140_74547349_-	0.7	2.7E-02	CYGB
53	UP	chr19_39826638_39826764_-	0.8	2.8E-02	GMFG
54	UP	chr7_45933004_45933231_+	1.5	2.8E-02	-
55	UP	chr17_48351783_48351864_+	1.7	2.8E-02	TMEM92



56	UP	chr5_38468352_38468731_-	1.8	2.9E-02	CTD-2263F21.1
57	UP	chr17_40715271_40715276_+	1.4	3.1E-02	COASY
58	UP	chr12_96390065_96390144_-	1.9	3.1E-02	HAL
59	UP	chr17_74261257_74261463_+	0.7	3.2E-02	FAM100B
60	UP	chr12_50132698_50132740_+	1.3	3.3E-02	TMBIM6
61	UP	chr1_198126084_198126296_+	0.7	3.4E-02	NEK7
62	UP	chrX_102884608_102884683_+	1.4	3.7E-02	TCEAL1
63	UP	chr6_43595017_43595132_-	1.6	4.1E-02	GTPBP2
64	UP	chr2_208734583_208734798_-	0.9	4.3E-02	AC083900.1
65	UP	chr13_76209878_76209953_-	1.5	4.4E-02	LMO7
66	UP	chr9_33447490_33447641_-	1.7	4.6E-02	AQP3
67	UP	chr7_32933645_32933792_-	1.1	4.6E-02	KBTBD2

---

Table S16. Significantly dysregulated genes with LTR-002 KD at 48h

Number	TYPE	TSS	Log2(FC)	FDR	Name
1	DOWN	chr10_77191031_77191479_+	-1.2	2.3E-12	C10ORF11
2	DOWN	chr2_145425267_145425610_+	-1.0	2.0E-03	AC074093
3	DOWN	chr6_120326805_120326902_+	-2.1	6.7E-03	LTR-002
4	DOWN	chr3_171758299_171758374_+	-0.7	4.6E-02	FNDC3B
5	UP	chr8_117658735_117658838_-	5.7	2.1E-06	EIF3H
6	UP	chr2_204055410_204055413_+	1.6	6.7E-03	AC011737.2
7	UP	chr1_22234544_22234549_+	1.0	6.7E-03	RP11-26H16.1

---

## III ÉTUDE 2 : CARACTÉRISATION DU CARCINOME HÉPATOCELLULAIRE À DOUBLE DIFFÉRENCIATION

### 1. Présentation de l'étude

Les CHC expriment au sein du tissu tumoral des marqueurs protéiques spécifiques des hépatocytes comme les cytokératines 8 et 18, alors que le cholangiocarcinome (deuxième tumeur primitive du foie la plus fréquente) dérivé de la transformation des cholangiocytes expriment des marqueurs biliaires, les cytokératines 7 et 19 (Lai *et al.*, 1989). Une première étude asiatique (Wu *et al.*, 1996) a classé environ 290 tumeurs primitives du foie selon des caractéristiques immunophénotypiques. La cytokératine 19 (KRT19, exprimée physiologiquement par les cholangiocytes et les cellules progénitrices bipotentes du foie) fut retrouvée exprimée dans 29 % des CHC étudiés et cette expression était corrélée à un pronostic sombre des patients comparé à celui de patients avec CHC n'exprimant pas KRT19. Ces résultats furent confirmés par une autre étude réalisée sur des échantillons de CHC provenant de patients caucasiens (Durnez *et al.*, 2006). Sur une cohorte de 109 patients, 16 % des CHC exprimaient la seule KRT19 et 28 % exprimaient les deux marqueurs biliaires KRT19 et CK7. Ces études suggèrent que la tumeur pourrait avoir pour origine la transformation maligne de cellules progénitrices hépatiques. Dans le même temps, une étude transcriptomique réalisée sur 61 échantillons de CHC humains, d'hépatoblastes et d'hépatocytes murins, a identifié un sous-groupe de CHC dont le profil d'expression était proche de celui des hépatoblastes, qui surexpriment des marqueurs de cellules ovals du foie ainsi que KRT19, et était associé à un mauvais pronostic des patients (Lee *et al.*, 2006).

D'autres études ont ensuite confirmé l'existence de ce sous-groupe moléculaire particulier (Komuta *et al.*, 2008; Cogliati *et al.*, 2010; Woo *et al.*, 2010). Toutes ces études ont associé l'expression du marqueur progéniteur KRT19 à différentes caractéristiques clinico-pathologiques telles que les métastases, la faible différenciation tumorale, la récurrence tumorale et la survie réduite des patients. L'ARNm KRT19 est inclus dans un score associant 4 autres gènes (*HNF1*, *RAN*, *RAMP3* et *TAF9*) qui aiderait à prédire le

pronostic des patients (Nault *et al.*, 2013a). KRT19 peut également être un outil intéressant dans la prédiction de la récurrence tumorale chez des patients ayant bénéficié d'une transplantation hépatique bien que leur CHC était au-delà des critères de Milan. En effet, dans une étude multicentrique comprenant 132 patients transplantés du foie, l'absence d'expression tumorale de KRT19 combinée à une signature moléculaire de type S2 (Hoshida *et al.*, 2009) était associée à une survie équivalente à celle observée chez des patients transplantés pour des tumeurs qui répondaient aux critères de Milan (67 % et 70 %, respectivement) (Mazzaferro *et al.*, 1996; Llovet *et al.*, 1999; Miltiadous *et al.*, 2015).

Ce sous-ensemble de CHC expriment non seulement d'autres marqueurs de cellules progénitrices (CK7, CD133, GSTpi), mais également des gènes impliqués dans l'invasion métastatique (Govaere *et al.*, 2014). Les expressions du facteur de transcription HNF4 $\alpha$  et celle de ses gènes cibles, impliqué dans la différenciation hépatocytaire, est réduite (Woo *et al.*, 2010; Coulouarn *et al.*, 2012; Govaere *et al.*, 2014). La voie NOTCH, impliquée dans la différenciation biliaire, est activée dans ces tumeurs ainsi que la voie de la transition épithélio-mésenchymateuse (EMT) (Kim *et al.*, 2011). L'inactivation de KRT19 dans des lignées cancéreuses dérivées de CHC réduisait la migration (Govaere *et al.*, 2014) et la prolifération cellulaires (Govaere *et al.*, 2014; Kawai *et al.*, 2015). Ces expériences ont également montré que, dans ces cellules cancéreuses, KRT19 activait la voie TGF $\beta$ /Smad (Kawai *et al.*, 2015) et qu'elles devenaient résistantes au traitement par le 5-fluorouracile. Ces observations ont été complétées par nos propres travaux montrant que la  $\gamma$ -SMA (Smooth Muscle Actine entérique) était exprimée dans des cellules hépatiques après activation de l'EMT par TGF $\beta$ , que l'expression de la  $\gamma$ -SMA était fortement corrélée à celles de KRT19 et d'EpCAM (et non de la vimentine) dans les CHC, et, enfin, que les CHC exprimant la  $\gamma$ -SMA présentaient les mêmes caractéristiques clinico-pathologiques que celles observées dans les CHC-KRT19 positif (Benzoubir *et al.*, 2015). Récemment, il a été montré que l'expression de KRT19 était régulée par des fibroblastes associés au cancer (CAFs) via les axes HGF-MET-ERK1/2-AP1 et SP1 (Rhee *et al.*, 2018).

Le rôle de KRT19 dans l'invasion tumorale est toutefois différent suivant le type de cancer considéré. Prenons l'exemple du cancer du sein dont certaines formes peuvent exprimer KRT19 et présenter une invasion tumorale. L'inhibition de l'expression de KRT19 par des siARN dans des lignées cellulaires mammaires a eu pour effet d'amplifier le phénomène d'invasion et de migration cellulaires. L'expression de marqueurs de cellules souches était également fortement augmentée. Le mécanisme de régulation sous-jacent impliquait le complexe  $\beta$ -caténine/RAC1 qui, en interagissant avec KRT19, entraînait une translocation nucléaire de  $\beta$ -caténine et une augmentation d'expression du gène cible NUMB qui est impliqué dans l'inhibition de la voie NOTCH (Saha *et al.*, 2016). Ce même type d'effet suite à l'inhibition de KRT19 dans le cancer du sein avait été préalablement observé dans une autre étude qui décrivait une voie de régulation différente, celle de la voie AKT (Ju *et al.*, 2013).

Notre étude a identifié à l'aide de méthodes d'analyse d'expression ARN et protéique combinées à des analyses *in silico* appliquées sur un total de 438 CHC une nouvelle voie de régulation des CHC KRT19-positif affectant le métabolisme énergétique des tumeurs. Nous mettons en évidence dans les CHC KRT19-positif une forte dérégulation des fonctions métaboliques hépatocytaires avec reprogrammation métabolique de type glycolytique liée à une activation de la réponse à l'hypoxie comparés aux CHC KRT19-négatif. Nous montrons également que les CHC KRT19-positif présentaient un excès de production de l'oncométabolite 2-hydroxyglutarate en absence de mutation des gènes IDH1 et IDH2 et que la présence tissulaire de 2-HG était associée à un profil aberrant hyperméthylé sur la lysine 9 de l'histone H3 (H3K9me3).

## 2. Matériel et méthodes

### 2.1 Sélection des échantillons

Plus de 400 CHC ont été immunophénotypés de manière systématique par 4 centres anatomo-pathologiques experts (Beaujon, Bordeaux, Paul-Brousse, Saint-Antoine) avec le marqueur hépatocyttaire HepPar1 et le marqueur biliaire KRT19. Tous les CHC analysés exprimaient HepPar1. Les CHC sélectionnés pour l'analyse protéomique devaient présenter une positivité de marquage de KRT19 dans au moins 50 % des hépatocytes tumoraux constituant le nodule. Six prélèvements congelés de CHC dont la proportion de cellules tumorales positives pour KRT19 était au moins supérieur à 50 % ont été sélectionnés. Des tumeurs négatives pour KRT19 ont également été retenues tenant compte de l'âge, du sexe des patients, de l'étiologie de la tumeur, du score METAVIR et du traitement pré-opératoire pour correspondre au mieux aux caractéristiques des patients avec CHC-KRT19 positif. Seuls cinq prélèvements de cette catégorie ont été retenus.

### 2.2 Microdissection Laser

Des coupes histologiques de 8 µm d'épaisseur ont été réalisées à -20 °C dans un cryotome et déposées sur des lames SuperFrost. Elles ont été par la suite conservées à -80 °C.

Les coupes furent ensuite colorées à l'hématoxyline/éosine à l'aide d'un protocole adapté à l'analyse des protéines (Dos Santos *et al.*, 2007). Brièvement, les coupes sont fixées pendant 1 minute dans de l'éthanol à 70 %, colorées à l'hématoxyline pendant 30 sec, immergées dans de l'eau de Scott (Sigma) puis colorées à l'éosine (0,25 % en solution alcoolique). Les coupes sont ensuite déshydratées dans des solutions d'éthanol de concentration croissante (30 sec à 70 %, 1 min à 100 %) puis dans deux bains de xylène de 5 min chacun. Chaque solution aqueuse contient un cocktail d'inhibiteurs de protéase (Complete Protease Inhibitor Cocktail – Roche Biochemicals, France).

Pour chaque échantillon, une surface de 20 mm<sup>2</sup> de tissu a été microdisséquée et catapultée dans le capuchon d'un tube de 0,5 mL contenant 10µL de SDS à 10 % à l'aide

du système PALM Microbeam (Carl Zeiss). Les échantillons ont été conservés à -80 °C jusqu'à solubilisation des protéines.

## 2.3 Spectrométrie de masse

### 2.3.1 Solubilisation des protéines

Quinze microlitres de tampon Laemlli (2.5X) ont été ajoutés aux cellules microdisséquées. Les cellules sont ensuite incubées 15 min à 37 °C puis 5 min à 95 °C. On ajoute 10 mg d'urée dans le tube et on ajuste à 40 µL avec de l'eau HPLC et on incube enfin à 37 °C pendant 15 minutes.

### 2.3.2 Séparation des protéines

Une électrophorèse 1D-PAGE est réalisée pour séparer les protéines selon leur taille moléculaire. Brièvement, les protéines sont dénaturées par chauffage de l'extrait à 95 °C pendant 5 min. Les 40µL de chaque extrait sont déposés sur un gel d'acrylamide à gradient 4-20 % (NuPage 4-12 % MES) et séparés par migration sur une longueur approximative de 3 cm.

Les gels sont ensuite immergés dans une solution de fixation contenant 30 % d'éthanol et 7,5 % d'acide acétique pendant 30 min puis colorés avec du bleu de Commasie (R-250, Bio-Rad) pendant 30 min. Les gels sont ensuite rincés avec la solution de fixation jusqu'à éclaircissement du gel.

### 2.3.3 Digestion tryptique et extraction des peptides

Seize bandes gel de taille régulière sont coupées pour chaque extrait et sont ensuite traitées de la manière suivante.

#### 2.3.3.1 Réduction et alkylation

Les bandes de gels sont lavées pendant 30 min sous agitation à température ambiante successivement avec (i) 200µL de bicarbonate d'ammonium 25mM, 50 % acétonitrile, (ii) 200µL d'eau et pour finir (iii) 200µL d'acétonitrile 100 %. Les bandes de gels sont ensuite passées aux ultrasons et séchées au SpeedVac.

S'ensuit une étape de réduction avec 100µL de Dithiothréitol (DTT) 10mM pendant 45 min à 56 °C. Les cystéines sont ensuite alkylées par iodoacétamide qui conduit à une carboxyamidométhylation. Chaque bande de gel est incubée avec 200µL d'une solution d'iodoacétamide 55mM pendant 35 min à température ambiante dans l'obscurité. Pour finir, la réaction d'alkylation est bloquée avec 200µL d'une solution de DTT 10mM pendant 10 min à température ambiante. Enfin, on rince les bandes de gels avec 200µl d'une solution de bicarbonate d'ammonium 25mM.

#### **2.3.3.2 Digestion tryptique et extraction des peptides**

Les protéines contenues dans le gel sont digérées dans une solution de trypsine 6 ng/µL pendant une nuit à 37 °C. Le surnageant de digestion est récupéré. Deux incubations successives de 15 min avec de l'acetonitrile à 50 et 100 % respectivement sont réalisées et les deux surnageants sont récupérés et ajoutés au premier. Les échantillons sont ensuite séchés au SpeedVac et conservés à -80 °C.

#### **2.3.3.3 Analyse en spectrométrie de masse**

Les échantillons peptidiques sont séparés sur une chaîne de chromatographie composée de deux colonnes de phase inverse C18 puis analysés sur un LTQ-Orbitrap XL (Thermo Fisher Scientific) selon une approche AMT (*Accurate Mass and Time*). Cette méthode consiste dans un premier temps à identifier les peptides présents dans un mélange complexe en utilisant une combinaison de techniques de fractionnement, nécessitant des analyses en mode MS/MS. Dans une deuxième phase, les peptides identifiés lors de la première phase sont quantifiés dans des analyses en mode MS. L'extraction des signaux MS est réalisée grâce à des logiciels adaptés (Decon2LS et VIPER).

#### **2.3.3.4 Base AMT**

Chaque échantillon a été analysé trois fois. Les masses des fragments peptidiques sont soumises au moteur de recherche Mascot (Matrix Science) et la base de données utilisée pour cette étude est la banque SwissProt TrEMBL (version xx) avec la taxonomie *Homo Sapiens*. Les peptides répondant aux critères suivants : (i) Premier rang ; (ii) Score d'identification supérieur au score d'homologie ; (iii)  $p < 0,05$  sont sélectionnés pour être



inclus dans la base. Une fois les peptides filtrés, les protéines couvertes par le même ensemble de peptides sont regroupés dans un *protein group*.

Chaque peptide filtré dans la base est caractérisé par (i) sa masse précise, (ii) son temps de rétention à la chromatographie (iii) et l'identifiant de la protéine dont il provient.

### 2.3.3.5 Quantification des peptides par MS (méthode sans marquage – Label free)

Une deuxième analyse est réalisée avec les données MS pour chaque peptide. L'intensité de chaque peptide reflétant son abondance est extraite des données brutes du spectromètre de masse ainsi que son identité à l'aide de la base AMT préalablement construite.

Afin de détecter les protéines différentiellement exprimées, on utilise une méthode statistique dérivée de celle décrite par Fu *et al.* (Fu *et al.*, 2008) et préalablement utilisée dans un précédent projet du groupe (Dos Santos *et al.*, 2010).

Brièvement, l'index d'abondance (AbI) est calculé selon la formule suivante :

$$AbI = \frac{\bar{A}}{\bar{A} + \bar{B}} \times \frac{N_A^D}{N_A^T} - \frac{\bar{B}}{\bar{A} + \bar{B}} \times \frac{N_B^D}{N_B^T}$$

$\bar{A}$  et  $\bar{B}$  sont les valeurs moyennes d'abondance des peptides dans les deux groupes étudiés.  $N_A^D$  et  $N_B^D$  représentent le nombre de fois où le peptide est détecté et  $N_A^T$  et  $N_B^T$  le nombre total de patients dans les deux conditions.

Après distribution et permutation des index d'abondance pour tous les peptides, nous pouvons définir les bornes au-delà desquelles les peptides peuvent être considérés comme étant différentiellement abondants entre les groupes.

Enfin on considère, qu'un *protein group* (soit la protéine) est considéré comme étant différentiellement exprimé lorsque plus de 50 % des peptides sont retrouvés différentiellement abondants entre les groupes.

## 2.4 Dosage tissulaire de la 2-Hydroxyglutarate

Cinq cents micromètres de tissus congelés sont broyés dans 1 mL de tampon de lyse (80 % méthanol/20 % eau) pendant 10 min à 25 Hz. Une extraction est réalisée à l'aide de

cartouches à phase inverse (STRATA XL-A SPE) et le groupe hydroxyle du carbone asymétrique avec du (+)-o,o'-diacetyl-L-tartaric anhydride à 80°C pendant 40min dans le but de séparer les énantiomères. L'extrait est repris dans une solution à 2mM d'ammonium formate et d'acetonitrile (4:1 v/v). Enfin, l'extrait est injecté dans un colonne de chromatographie C18 combiné à un spectromètre de masse (Xevo TQ MS Waters)

## 2.5 Détection des protéines par immunohistochimie

L'étude immunohistochimique a été réalisée sur des prélèvements tumoraux de CHC fixés au formol et inclus en paraffine. Des coupes histologiques de 4 µm sont déposées sur des lames SuperFrost Ultra Plus.

L'étude de l'expression et de la localisation des marqueurs listés ci-après est réalisée à l'aide d'une méthode sandwich et d'une révélation à la Diaminobenzidine (DAB).

Les coupes histologiques sont d'abord déparaffinées avec 3 immersions de Xylène de 15 min chacune puis réhydratées avec 3 bains d'éthanol de concentration décroissante (100 %, 90 % et 70 %, 5 min) et enfin dans de l'eau distillée.

Une restauration antigénique visant à hydrolyser les produits d'addition du formaldéhyde aux macromolécules tissulaires est pratiquée (cf. tableau ci-contre).

Anticorps	Référence	Restauration antigénique	Concentration
KRT19	RCK108 (Dako)	tampon citrate 10mM pH6	40µg/mL
MAPRE1	HPA003600 (Sigma)	EDTA 1mM pH8	0,5µg/mL
H6PD	HPA004824 (Sigma)	tampon citrate 10mM pH6	0,5µg/mL
HIF1α	ab1 (Abcam)	tampon citrate 10mM pH6	20µg/mL
Ki67	RBK027 (Diagomics)	tampon citrate 10mM pH6	Prêt à l'emploi
H3K9me3	07-442 (Millipore)	EDTA 1mM pH8	10µg/mL

Les peroxydases endogènes présentes dans le tissu sont inhibées par incubation dans une solution de peroxyde d'hydrogène à 3 % pendant 20 min. Les coupes sont ensuite incubées dans une solution d'anticorps (cf. tableau infra) pendant une nuit à 4°C en chambre humide.

Après deux rinçages au PBS-Tween 0,1 % et un au PBS seul, les coupes sont incubées avec un anticorps secondaire (portant une peroxydase) de chèvre spécifique des fragments constants des immunoglobulines de lapin et/ou de souris pendant 30 min.

Après deux rinçages au PBS-Tween 0,1 % et un au PBS seul, l'activité peroxydasique est révélée pendant 5 min maximum à l'abri de la lumière à l'aide d'une solution contenant un substrat de la peroxydase et la diaminobenzidine.

Les coupes sont enfin colorées à l'hématoxyline, déshydratées, immergées dans le xylène et recouvertes d'une lamelle montée à l'Eukitt.

La semi-quantification a été réalisé en combinant le pourcentage de cellules d'intérêt marqué et l'intensité de marquage (0 : négatif ; 1 : faible marquage ; 2 : intermédiaire ; 3 : fort)

## 2.6 Analyse des transcrits

Les données d'expression RNA Seq issues de données de CAGE (dbGAP :phs000885.v1.p1) du précédent travail (Hashimoto et al. 2015) ont été reprises afin d'identifier des groupes d'expression de la cytokératine 19. Une classification non supervisée est réalisé sur les 105 transcriptomes (50 CHC, 50 non tumeurs à distance et 5 foies normaux) et une analyse EdgeR est ensuite réalisé pour déterminer le profil différentiel.

## 2.7 Enrichissement fonctionnel

L'interprétation biologique d'une liste de gènes d'intérêt est aujourd'hui une nécessité pour rechercher les processus biologiques caractérisant les échantillons à partir des données d'expression. Dans notre étude, nous avons utilisé deux types d'algorithmes en fonction des données récoltées.

### 2.7.1 À partir des données protéiques

Les identifiants protéiques sont d'abord convertis selon la nomenclature HUGO puis soumis à l'analyse d'enrichissement fonctionnel EnrichR (Chen *et al.*, 2013a; Kuleshov *et al.*, 2016). Les bases de données KEGG (<https://www.genome.jp/kegg/>) et Reactome (<https://reactome.org/>) ont été utilisées pour la recherche de voies enrichies.

### 2.7.2 À partir des données transcriptionnelles

À la différence de la méthode précédente, nous utilisons la liste entière des gènes et de leur expression. Ce tableau de données est soumis à l'algorithme GSEA (Gene Set Enrichment Analysis, (Mootha *et al.*, 2003; Subramanian *et al.*, 2005)) dont l'objectif est d'identifier des groupes fonctionnels de gènes dont le différentiel d'expression est uniquement dû au caractère d'intérêt. Un score est ensuite attribué à chacun de ces groupes. Les groupes enrichis dont le taux de fausse découverte est inférieur à 5 % ont été sélectionnés.

## 2.8 Modèle in vitro

### 2.8.1 Culture de la lignée d'hépatocarcinome HuH7

La lignée de CHC utilisée est la lignée Huh7 issu d'un carcinome hépatocellulaire bien différencié. Les cellules sont mises en culture dans un milieu DMEM à 4,5 g/L de glucose (Gibco) complété avec 10 % de sérum de veau foetal et 1 % d'un mélange pénicilline/streptomycine.

### 2.8.2 Stabilisation du facteur de transcription HIF1- $\alpha$

Le facteur de transcription HIF1- $\alpha$  dans des conditions normoxiques (21 % d'oxygène) est exprimé de manière constitutive mais est rapidement dégradé (cf. Généralités sur le métabolisme des cellules tumorales p 51).

Le chlorure de cobalt ( $\text{CoCl}_2$ ) inhibe les prolyl-hydroxylases en remplaçant le fer par le cobalt rendant les enzymes inactives. HIF1- $\alpha$  n'est donc plus hydroxylé et donc n'est pas dégradé. Les cellules sont donc incubées avec 400 $\mu\text{M}$  de  $\text{CoCl}_2$  (Sigma-Aldrich) pendant 24h et 48h.

### 2.8.3 Immunodétection sur membranes

20µg de protéines ont été déposés sur gel de polyacrylamide réticulé à 10 % et transférés sur membrane de nitrocellulose avant immunodétection à l'aide des anticorps listés dans le tableau ci-contre.

Anticorps	Référence	Concentration
HIF1α	ab1 (Abcam)	5 µg/mL
KRT19	ab15463 (Abcam)	Dilution au 1/200
MAPRE1	sc-374474 (Santa Cruz)	1 µg/mL
PDCD4	sc-376430 (Santa Cruz)	1 µg/mL
HK2	2106S (Cell Signaling)	Dilution au 1/1000
Cyclin A	sc-596 (Santa Cruz)	1 µg/mL
α/β-tubulin	2148 (Cell Signaling)	Dilution au 1/1000

Les résultats ont été quantifiés en utilisant Image J et normalisés par rapport à la α/β-tubuline.

## 2.9 Analyses statistiques

Pour les comparaisons 2 à 2, un test de Mann-Whitney a été réalisé au lieu d'un test de Student si la normalité des distributions n'était pas respectée. Pour l'établissement du profil protéique différentiel, nous avons utilisé le calcul de l'abondance index. Pour l'analyse des transcrits séquencés par CAGE et l'analyse *in silico* des données du TCGA, l'algorithme edgeR a été utilisé pour comparer les deux populations. Les corrélations cliniques ont été réalisées à l'aide d'un test de Fisher. Tous les tests statistiques ont été réalisés via le logiciel R (v. 3.6.0).

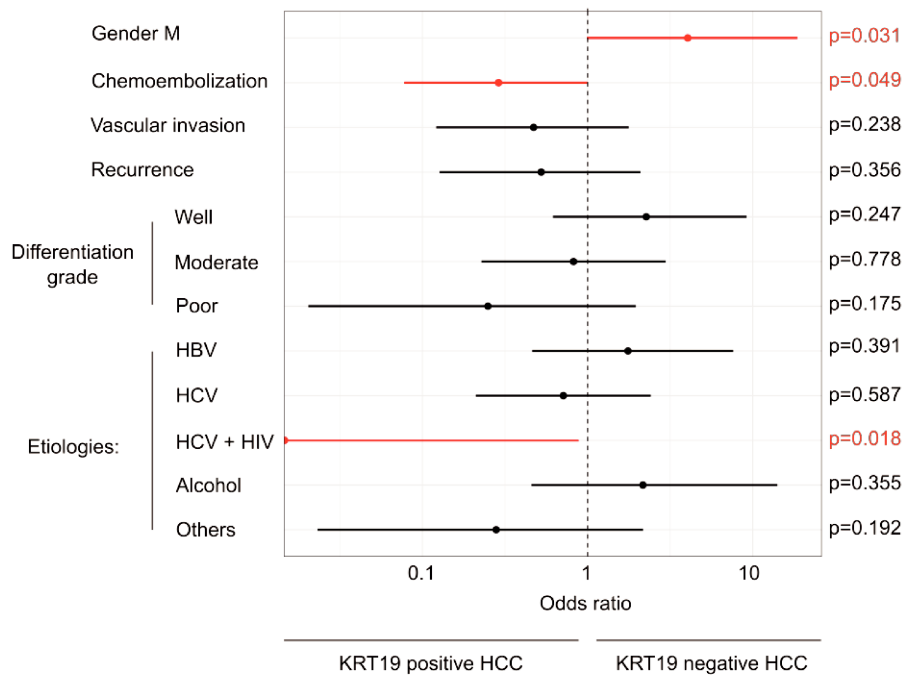


Figure 1: **Caractéristiques cliniques du CHC à double différenciation.** Rapport des côtes pour évaluer la corrélation de paramètres cliniques avec le groupe tumoral. Les points correspondent aux estimateurs de variables. La longueur des lignes représente l'intervalle de confiance (95%). En rouge les rapports de côtes correspondant aux variables statistiques. Autres étiologies : Hémochromatose, Syndrome métabolique, étiologie inconnue.

### 3. Résultats

#### 3.1 La protéomique quantitative lie les voies métaboliques aux CHC KRT19-positif

Plus de 400 échantillons de CHC ont été immunophénotypés de manière systématique avec le marqueur hépatocytaire HepPar1 et le marqueur biliaire KRT19. La proportion de CHC HepPar1 positif exprimant clairement le marqueur KRT19 était relativement faible de l'ordre de 8% des CHC analysés (supérieur au seuil de 5 % des cellules tumorales). L'intensité du signal KRT19 au sein des cellules cancéreuses était variable, allant de faible (seuil de 5% de positivité) à très forte (seuil > 50 % de positivité). La localisation des cellules cancéreuses exprimant KRT19 était variable, dispersées au sein de la masse tumorale avec quelques cellules positives ou formant des micro-foyers bien circonscrits. Nous avons recueilli les caractéristiques clinico-pathologiques disponibles pour un sous-groupe de patients (n=22) avec CHC KRT19-positif et comparé à celles de patients avec CHC KRT19-négatif (n=35). Une analyse univariée a montré que les CHC KRT19-positif étaient significativement associés à la co-infection par les virus de l'hépatite C et de l'immunodéficience humaine (VIH) et au traitement pré-opératoire par chimio-embolisation trans-artérielle (Figure 1). Nous n'avons pas retrouvé d'association significative avec le grade de différenciation tumorale et l'invasion vasculaire bien qu'une tendance se soit dégagée pour ces deux variables dans les CHC KRT19-positif. Les données de survie n'ont pas pu être exploitées car les patients des groupes KRT19-positif et -négatif ont bénéficié d'une résection chirurgicale (n=28) ou d'une transplantation hépatique (n=25). Cette dernière étant associée à une meilleure survie des patients, l'analyse statistique n'est pas réalisable. Le faible échantillonnage a empêché de grouper les patients en fonction du type d'acte chirurgical. Un total de 160 analyses par spectrométrie de masse en tandem ont été réalisées sur 6 CHCs microdisséqués KRT19-positif et 5 CHCs microdisséqués KRT19-négatif ce qui nous a permis de construire une base AMT contenant l'ensemble des tags peptidiques. Les échantillons KRT19-négatif ont été sélectionnés sur la base de l'âge, sexe, étiologie du CHC,

Gene symbol	Protein name	Fold change	Biological Activity
KRT19	Keratin, type I cytoskeletal 19	*	Intermediate filament. Organization of myofibers
GNAS	Guanine nucleotide-binding protein G(s) subunit alpha isoforms Xlas	*	Transducers in signaling pathways controlled by G protein-coupled receptors (GPCRs).
PTMA	Prothymosin alpha	*	Immune function. Play role in insulin resistance. Anti-apoptotic activity
C12orf10	UPF0160 protein MYG1, mitochondrial	*	Unknown function. May involved in stress-response
VIL1	Villin 1	18.2	Epithelial cell-specific Ca <sup>2+</sup> -regulated actin-modifying protein . Role in Actin nucleation and capping
MAPRE1	Microtubule-associated protein RP/EB family member 1	11.9	Plus-end tracking protein (+TIP). Regulates the dynamics of the microtubule cytoskeleton
CAST	Calpastatin	7.5	Inhibition of calpain, activated by hypoxia
MARC1	Mitochondrial amidoxime-reducing component 1	5.7	Reduction of N-oxydes. Role in oxidative stress
KRT5	Keratin, type II cytoskeletal 5	4.5	Cytoskeletal protein. Skin diseases
PKM2	Pyruvate kinase PKM	4.1	Last step of glycolysis. Main form in cancer
PCNA	Proliferating cell nuclear antigen	3.9	Role in DNA replication
CIRBP	Cold-inducible RNA-binding protein	3.9	Protective role in the genotoxic stress response
SNRPD1	Small nuclear ribonucleoprotein Sm D1	3.8	Role in the splicing of cellular pre-mRNAs.
RALY	RNA-binding protein Raly	3.7	Transcriptional cofactor for cholesterol biosynthetic genes in the liver
COPS7B	COP9 signalosome complex subunit 7b	3.2	Regulator of the ubiquitin conjugation pathway
CAPRN1	Caprin-1	3.0	Regulation the transport and translation of mRNAs of proteins. Binds MYC
NPM3	Nucleoplasmin-3	2.6	Nuclear chaperone
HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	2.6	Processing of pre-mRNAs
F11R	Junctional adhesion molecule A	2.5	Role in epithelial tight junction formation
RPL35	60S ribosomal protein L35	2.5	rRNA processing
CNPY2	Protein canopy homolog 2	2.3	secreted angiogenic factor regulated by hypoxia
PAFAH1B3	Platelet-activating factor acetylhydrolase IB subunit gamma	2.2	Ether lipid metabolism
MCM3	DNA replication licensing factor MCM3	2.1	helicase activity
RPL27A	60S ribosomal protein L27a	2.1	rRNA processing
MYADM	Myeloid-associated differentiation marker	2.0	Differentiation of myeloid cells
HNRNPA0	Heterogeneous nuclear ribonucleoprotein A	1.9	Post-transcriptional regulation of cytokines mRNAs.
UBE2V1	Ubiquitin-conjugating enzyme E2 variant 1	1.9	Transcriptional activation of the human FOS proto-oncogene
NHP2L1	NHP2-like protein 1	1.8	Spliceosome assembly
HIST1H2BC	Histone H2B type 1-C/E/F/G/I	1.8	Core component of nucleosome
SPAG9	C-Jun-amino-terminal kinase-interacting protein 4	1.7	Mediates JNK signaling
CMTM6	CKLF-like MARVEL transmembrane domain-containing protein 6	1.7	Neutrophil degranulation
BLMH	Bleomycin hydrolase	1.5	Physiological role unknown. Inactivation of the antitumor drug BLM
EIF5	Eukaryotic translation initiation factor 5	1.4	Hydrolysis of GTP bound to the 40S ribosomal initiation complex
MRPL28	39S ribosomal protein L28, mitochondrial	1.4	Mitochondrial protein synthesis
EEA1	Early endosome antigen 1	1.4	Endosomal trafficking
ATIC	Bifunctional purine biosynthesis protein PURH	1.3	Purine biosynthesis
TPD52L2	Tumor protein D54	1.1	Role in cell migration
TM9SF2	Transmembrane 9 superfamily member 2	1.1	Putative role in small molecule transport
EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	1.1	Early steps of protein synthesis

Tableau 1 : **Liste des protéines sur-exprimés dans le CHC à double différenciation.** Liste des protéines dont les peptides identifiés ont été retrouvés statistiquement plus abondant dans le CHC positifs pour la cytokératine 19. La première colonne est la dénomination en symbole de gène, a deuxième, le nom officiel des protéines (Uniprot), la troisième colonne représente le rapport d'expression entre les cancers cytokératine 19 positifs versus les CHC cytokératine 19 négatifs et la dernière représente les activités biologiques (selon Uniprot). \*: Pour ces 4 protéines, les peptides n'ont été identifiés que dans une seule condition. Le rapport ne peut donc pas être calculé.

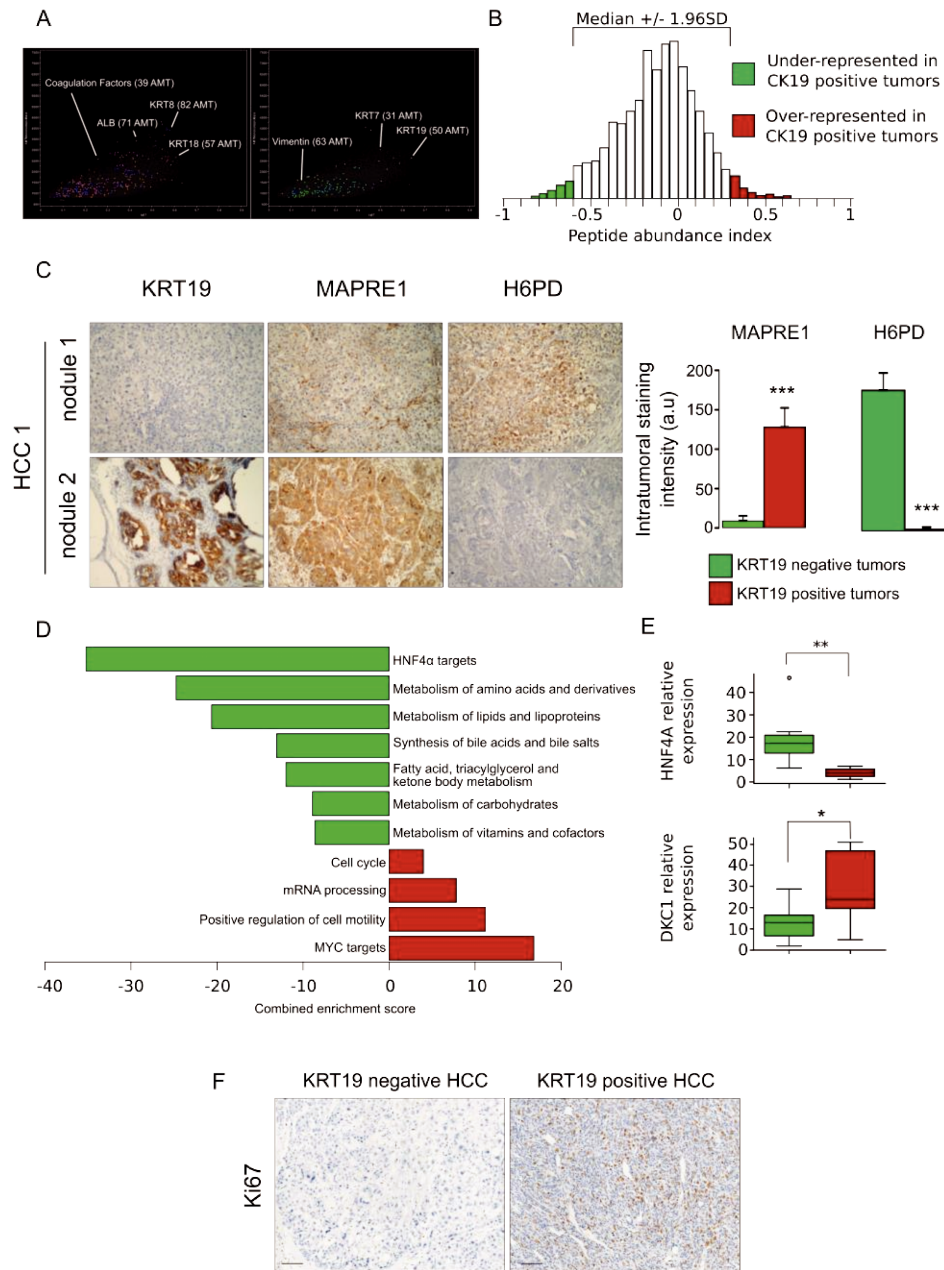


La base AMT incluait 46 903 tags peptidiques correspondant à 5 199 protéines. Parmi ces tags peptidiques, nous retrouvons à la fois des peptides provenant de marqueurs d'hépatocytes (KRT8, KRT18, albumine, facteurs de coagulation) et de marqueurs de cholangiocytes (KRT7, KRT19, Vimentine) (Figure 2A). L'analyse comparative des abondances peptidiques relatives entre les deux groupes microdisséqués de CHC KRT19-positif et -négatif a été réalisée à l'aide de l'index d'abondance (AbI) décrit dans le matériel et méthodes (2.3.3.5 p131 et Suppl Figure 1). Un total de 118 protéines sont retrouvées différentiellement dérégulées entre les deux catégories de CHC incluant 39 protéines sur-représentées et 79 protéines sous-représentées dans les CHC KRT19-positif (Tableau 1, Tableau supplémentaire 1 et Figure 2B). La plupart des protéines identifiées par l'analyse protéomique, bien que rapportées individuellement dans plusieurs travaux portant sur les CHC, n'ont pas été particulièrement associées au sous-type moléculaire KRT19 positif.

Les résultats protéomiques d'expression différentielle ont été validés par des analyses immunohistochimiques réalisées sur des tissus de CHC KRT19-positif (n=10) et KRT19-négatif (n=20). Les protéines cibles recherchées ont été celles pour lesquelles le rapport d'expression était le plus important entre les deux conditions KRT19-positif et KRT19-négatif. Il s'agit de Microtubule-associated protein RP/EB family member 1 (MAPRE1) et de GDH/6PGL endoplasmic bifunctional protein (H6PD).

Nous montrons que les nodules tumoraux KRT19-positif étaient tous positifs pour MAPRE1 avec de fréquentes et fortes intensités de marquage observées au sein des hépatocytes tumoraux. À l'inverse, les zones tumorales KRT19-positif n'exprimaient pas ou peu l'enzyme H6PD. Concernant les nodules tumoraux KRT19-négatif, le profil de marquage pour ces deux marqueurs était en miroir : on n'observait pas ou peu de marquage de MAPRE1, alors que H6PD présentait une forte intensité d'expression. L'analyse immunohistochimique quantitative a confirmé les données de l'analyse protéomique concernant ces deux marqueurs (Figure 2C).

Un enrichissement fonctionnel a été réalisé avec la liste des 118 protéines dérégulées dans les CHC KRT19-positif via l'algorithme EnrichR. Seuls les enrichissements de voies dont la valeur de p ajusté (obtenue par le test de Fisher) inférieur à 0,05 ont été sélectionnés



**Figure 2: Analyse protéomique des cellules microdisséquées du carcinome hépatocellulaire à double différenciation.** (A) Base AMT avec en couleurs les tags peptidiques correspondant à des marqueurs de CHC (à gauche) et de cholangiocarcinome (à droite). En abscisses est représenté le temps de rétention normalisé et en ordonnée la masse/charge ( $m/z$ ). (B) Distribution des index d'abondance (AbI) des 46 903 peptides identifiés par spectrométrie de masse en tandem. La permutation aléatoire permet l'établissement d'un intervalle de confiance pour déterminer le profil d'expression différentiel. Les peptides sur- et sous-abondants sont colorés en rouge et vert respectivement (C) Validation immunohistochimique de deux marqueurs identifiés lors de l'analyse. MAPRE1 (Microtubule-associated protein RP/EB family member 1), dont le rapport d'expression était l'un des plus élevés dans les carcinomes positifs à la cytokératine 19 et H6PD (GDH/6PGL endoplasmic bifunctional protein) dont le rapport était le plus faible. Images représentatives du marquage immunohistochimique (B) et quantification sur 10 échantillons positifs et 20 négatifs pour la cytokératine 19. analyse statistique réalisée avec le test de Mann-Whitney (\*\*\*) :  $p < 0,001$ ). (D) Les 39 et 79 protéines différemment exprimés dans le carcinome à double différenciation ont été soumis à l'algorithme d'enrichissement fonctionnel EnrichR. Les voies représentées sont celles dont le score d'enrichissement combiné (Écart par rapport à la moyenne combinée à la valeur  $p$  du test de Fisher) est le plus important dans les bases KEGG, REACTOME et ChEA. (E) Expression par PCR quantitative de HNF4 $\alpha$  et DCK1 dans des tumeurs humaines en fonction de l'expression de la cytokératine 19 (Tumeurs KRT19 positives  $n=5$ , tumeurs KRT19 négatives  $n=16$ ). L'expression est exprimée en unités arbitraires et correspond à la méthode delta-Ct par rapport à l'expression du gène de ménage TBP (Test de Mann-Whitney, \* :  $p < 0,05$  ; \*\* :  $p < 0,01$ ). (F) Images représentatives de l'expression de Ki67 par immunohistochimie (barre d'échelle : 100 $\mu$ m).

(figure 2D). Quatre voies moléculaires enrichies dans les CHC KRT19-positif (cibles de MYC, motilité cellulaire, machinerie des ARN messagers, cycle cellulaire) ont été identifiées. Le nombre réduit de protéines retrouvé sur-représentées dans la condition KRT19-positif rend compte du peu de voies moléculaires identifiées. Par ailleurs, nous notons une importante sous-représentation de voies métaboliques dans les CHC KRT19-positif correspondant au métabolisme hépatocytaire des glucides, lipides et acides aminés et à la synthèse des acides biliaires (Figure 2D).

Les enzymes qui constituent ces différentes voies métaboliques sont pour la plupart régulés par le facteur de transcription HNF4- $\alpha$  qui assure un rôle essentiel dans le développement, la différenciation et la polarité cellulaires dans le foie (Si-Tayeb *et al.*, 2010). Une analyse d'expression de HNF4- $\alpha$  et de DKC1 (cible de MYC) par PCR quantitative réalisée à partir d'échantillons congelés de CHC KRT19-positif (n=5) et KRT19-négatif (n=16) a montré une forte diminution de l'expression d'HNF4- $\alpha$  dans les CHC KRT19-positif comparé à celle des CHC KRT19-négatif (Figure 2E). Ce résultat est en accord avec des données de transcriptomique précédemment rapportées (Govaere *et al.*, 2014). L'expression de DKC1 était plus élevée dans les CHC KRT19-positif que dans les CHC KRT19-négatif suggérant une activation de la voie MYC dans les CHC à double différenciation (Figure 2E).

Nous montrons également par immunohistochimie à l'aide du marqueur de prolifération Ki67 que les CHCs KRT19-positif sont davantage prolifératifs que les CHC KRT19-négatif (Figure 2F) en accord avec la littérature (Govaere *et al.*, 2014)

### **3.2 La réponse à l'hypoxie et la glycolyse dont des marques distinctives des CHC KRT19-positif**

Une classification non supervisée a été réalisée à partir de données de RNA Seq que nous avons préalablement générées (dbGAP :phs000885.v1.p1) et obtenues par l'analyse de 50 couples tumeurs-non tumeurs de CHC et de 5 foies normaux. Trois sous-groupes moléculaires ont été identifiés comprenant un sous-groupe de 5 tumeurs de CHC présentant notamment une forte expression de KRT19 (« left cluster »), un sous-groupe de 20 tumeurs de CHC sans aucune expression de KRT19 (« right cluster ») et un sous-

groupe intermédiaire d'expression variable (Suppl Figure 2A). L'analyse différentielle par EdgeR des groupes « left cluster » et « right cluster » a identifié une liste de 420 gènes discriminant ces deux groupes de tumeurs, parmi lesquels des gènes de progéniteurs hépatiques (EPCAM, CD133), des gènes suppresseurs de tumeur (PDCD4) ainsi que de nombreux gènes (PKM2, GNAS, RALY, MAPRE1, CAPRIN1) codant des protéines que nous avons détectées dans notre analyse protéomique des CHC microdisséqués KRT19-positif et KRT19-négatif (Suppl Figure 2B). Cette liste de 420 gènes comprend également des gènes (S100A6, CD24, PPT1, PFKP) impliqués dans la récurrence tumorale (Woo *et al.*, 2010), des gènes surreprésentés dans la classe proliférative décrite par Chiang et col. (Chiang *et al.*, 2008) et corrélé à la survie (Lee *et al.*, 2004) (Suppl Figure 2C). L'analyse fonctionnelle des profils d'expression spécifiques des groupes KRT19-positif et KRT19-négatif formant la cohorte de 50 CHC et réalisée à partir des bases KEGG, Reactome et MsigDB a confirmé les données protéomiques de dérégulation majeure de plusieurs voies métaboliques hépatocytaires (Figure 2D, 3A). Comparativement aux CHC KRT19-négatif, les voies les plus sous-représentées dans les tumeurs KRT19-positif comprenaient le métabolisme des acides biliaires et celui des acides gras. Parmi les voies sur-représentées dans les CHC KRT19-positif, certaines d'entre elles ont déjà été rapportées dans le contexte KRT19 comme celles du TGF- $\beta$  et de NOTCH. La voie du TGF- $\beta$  est une voie majeure dans la transition épithélio-mésenchymateuse et la dédifférenciation cellulaire et est souvent retrouvée activée dans les CHC non différenciés (Coulouarn *et al.*, 2012; Govaere *et al.*, 2014; Rao *et al.*, 2017). L'activation de la voie NOTCH, connue pour son action dans la différenciation des cellules progénitrices hépatiques en cellules biliaires matures (Gil-García and Baladrón, 2016), est fortement corrélée à l'expression de KRT19 dans les CHC (Roskams *et al.*, 2010) est corrélé fortement à l'expression de la cytokératine 19 dans les CHC (Govaere *et al.*, 2014). Outre ces voies, notre analyse a identifié dans les CHC KRT19-positif une forte sur-représentation de cibles (lamine B1, récepteur transferrine, Ki67, Cycline B2, CDK4, MCM3, 4, 5) du facteur de transcription E2F qui est impliqué dans la régulation du cycle cellulaire et les mécanismes d'apoptose liés à TP53 (DeGregori *et al.*, 1997; Huntington *et al.*, 2016). Son expression, en particulier celle de E2F1, est augmentée dans les CHC (Midorikawa *et al.*, 2007) et est liée à une forte activation de l'oncogène MYC et de l'expression du TGF- $\alpha$  (Santoni-

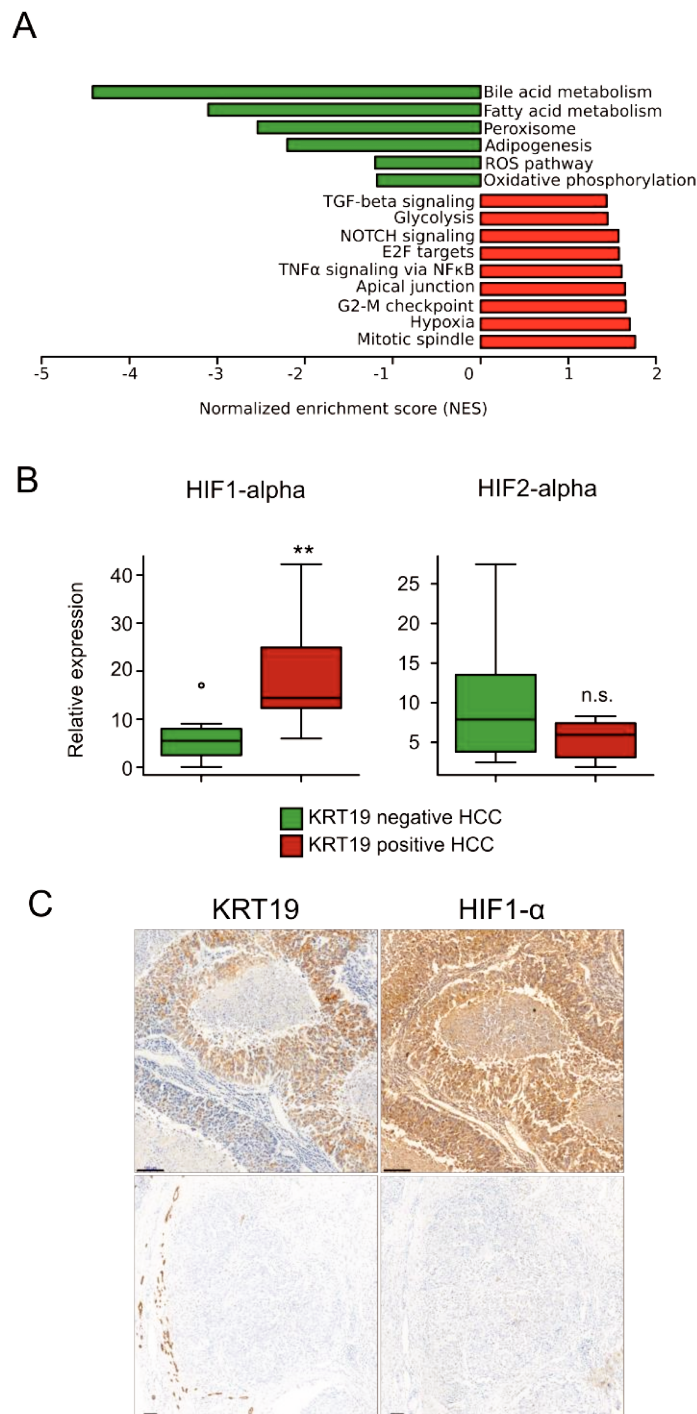


Figure 3 : **Analyse des transcrits ARN par séquençage CAGE.** A partir des données du premier travail (dbGAP :phs000885.v1.p1), une classification non supervisée a permis de définir deux groupes. Un groupe de forte expression (n=5) et un groupe dont l'expression est absente (n=20). (A) Enrichissement fonctionnel du transcriptome par l'algorithme GSEA. Les voies enrichies représentées sur le graphique ont un taux de fausse découverte inférieur à 5%. (B) Expression des facteurs HIF1α et HIF2α par PCR quantitative dans les deux groupes (Tumeurs KRT19 positives n=5, tumeurs KRT19 négatives n=16) (NS : non significatif, \*\* : p<0,01). (C) Expression de HIF1α par immunohistochimie dans des tumeurs humaines (barre d'échelle : 100μm).

Rugiu *et al.*, 1998). Une autre voie enrichie dans les CHC KRT19-positif est celle du TNF $\alpha$  bien connue pour ses effets dans la régénération hépatique et la pathogenèse du CHC notamment l'invasion et l'angiogenèse tumorale (Lewis and Pollard, 2006).

De manière intéressante, nous identifions plusieurs autres voies fortement dérégulées dans les CHC KRT19-positif, celles de la réponse à l'hypoxie, de la glycolyse et de la phosphorylation oxydative (Figure 3A). L'hypoxie est une caractéristique fréquente des tumeurs solides ainsi que des tumeurs traitées par voie systémique par des drogues de chimiothérapie (Méndez-Blanco *et al.*, 2018) ou localement par chimio-embolisation trans-artérielle (Petrillo *et al.*, 2018a). En réponse à un micro-environnement appauvri en oxygène, une réponse adaptative est déployée à l'échelon cellulaire avec activation de nombreuses voies de signalisation qui régulent des processus cellulaires variés (prolifération cellulaire, apoptose, angiogenèse, métabolisme, remodelage tissulaire), et ceci par l'intermédiaire de voies de réponse spécifiques comme celle des protéines de la famille HIF (Hypoxia Inducible Factor).

Nous montrons par PCR quantitative que l'expression de HIF1A, et non HIF2A, est augmentée dans les CHC KRT19-positif comparé à celle des CHC KRT19-négatif (Figure 3B, C). En accord avec ces résultats, nous observons une forte expression de plusieurs gènes cibles de la protéine HIF1- $\alpha$  parmi lesquels l'anhydrase carbonique CA9, le facteur d'angiogenèse VEGFA, l'hydrolase NDRG1 (N-myc downstream-regulated gene 1) et l'E3 ligase WSB1 (Suppl Figure 3A). La surexpression de HIF1A est également illustrée par immunomarquage tissulaire (Figure 3C).

Nous nous sommes ensuite demandés si la réponse à l'hypoxie était une marque associée spécifiquement à KRT19 en réalisant un clustering hiérarchique des 50 CHC à l'étude en fonction de l'expression d'un set (n=200) de gènes de réponse à l'hypoxie (HALLMARKS\_HYPOXIA; Broad Institute). Cette analyse a révélé l'existence de deux sous-groupes de CHC caractérisés par une expression franche différentielle des gènes cibles de réponse à l'hypoxie distinguant des tumeurs à forte (n=10) et faible (n=40) expression (Suppl Figure 3B). La comparaison du transcriptome global des deux sous-groupes par l'algorithme edgeR a placé KRT19 en première position des marqueurs discriminants entre sous-groupes tumoraux (Suppl Figure 3C, D) indiquant une réelle

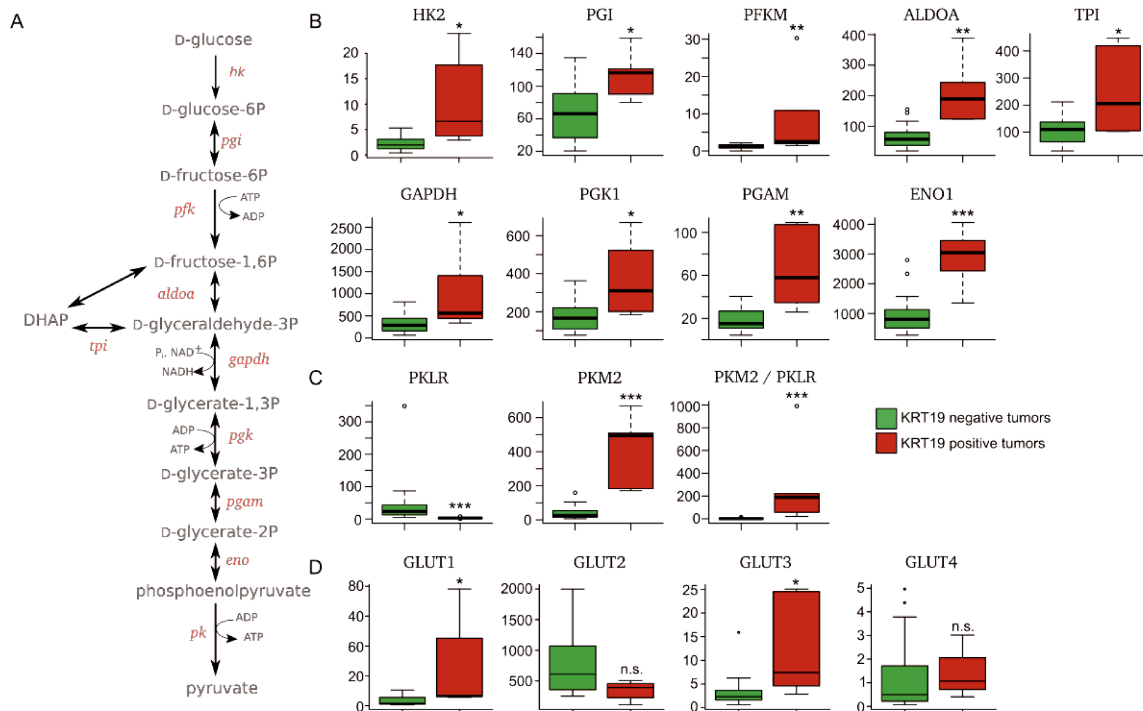


Figure 4 : **Les tumeurs à double différenciation sont hautement glycolytiques.** (A) Schéma décrivant la glycolyse avec les substrats, produits, enzymes et co-facteurs de la voie de la glycolyse. Expression des enzymes de la glycolyse (B) des pyruvates kinases PKLR et PKM2 (C) et des transporteurs de glucose GLUT1 à 4 (D) (Test de Mann Withney avec correction de Benjamini-Hochberg, \* :  $p < 0,05$  ; \*\* :  $p < 0,01$  ; \*\*\* :  $p < 0,001$ ).

association biologique entre réponse à l'hypoxie régulée par HIF1- $\alpha$  et expression hépatocytaire tumorale de KRT19.

Comme mentionné précédemment, les données transcriptomiques des 50 CHC analysés ont confirmé les données protéomiques en termes d'altérations de certaines voies métaboliques ce qui semble indiquer que la reprogrammation métabolique serait une deuxième marque moléculaire distinctive des CHC KRT19-positif (Figure 2D, 3A). Le taux d'expression de l'ensemble des enzymes impliqués dans la voie de la glycolyse est substantiellement augmenté dans les CHC KRT19-positif comparé aux CHC KRT19-négatif (Figure 4). L'analyse comparative des deux sous-types moléculaires de CHC KRT19-positif et négatif révèle en outre une interversion des formes foie-spécifique PKL (pyruvate kinase liver) et foetale PKM2 de la pyruvate kinase (enzyme convertissant le phosphoenolpyruvate en pyruvate au cours de la dernière étape de la glycolyse) au bénéfice de la forme foetale PKM2 dans les CHC KRT19-positif (Figure 4A, 4C). Il faut noter que l'analyse protéomique quantitative avait détecté PKM2 dans les nodules tumoraux KRT19-positif (Tableau 1). L'ensemble de ces résultats indique que les CHC KRT19-positif sont des tumeurs glycolytiques

### **3.3 Forte association entre réponse à l'hypoxie et expression de KRT19 dans la cohorte de CHC du TCGA**

Des analyses d'expression différentielle et d'enrichissement fonctionnel ont été appliquées sur les jeux de données transcriptomiques publiques du TCGA (The Cancer Genome Atlas) regroupant 371 CHC (Mai 2019). Les tumeurs ont été séparées suivant leur niveau d'expression de KRT19 ou du set (n=200) de gènes cibles de HIF1- $\alpha$  utilisé dans la Suppl. Figure 3B. Nous avons défini deux sous-groupes de tumeurs avec une expression extrême de KRT19 en extrayant le dernier quartile de CHC (expression maximale de KRT19, n=93) et les deux premiers quartiles (absence d'expression de KRT19 jusqu'à la médiane d'expression, n=185) (Figure 5A). Le groupe intermédiaire d'expression de KRT19 (n=93) n'a pas été retenu dans les analyses ultérieures. L'analyse du profil global d'expression différentiel via l'algorithme EdgeR a montré que les gènes principaux de la réponse à l'hypoxie (GLUT1, VEGF) et de la voie de la glycolyse (HK2, interversion



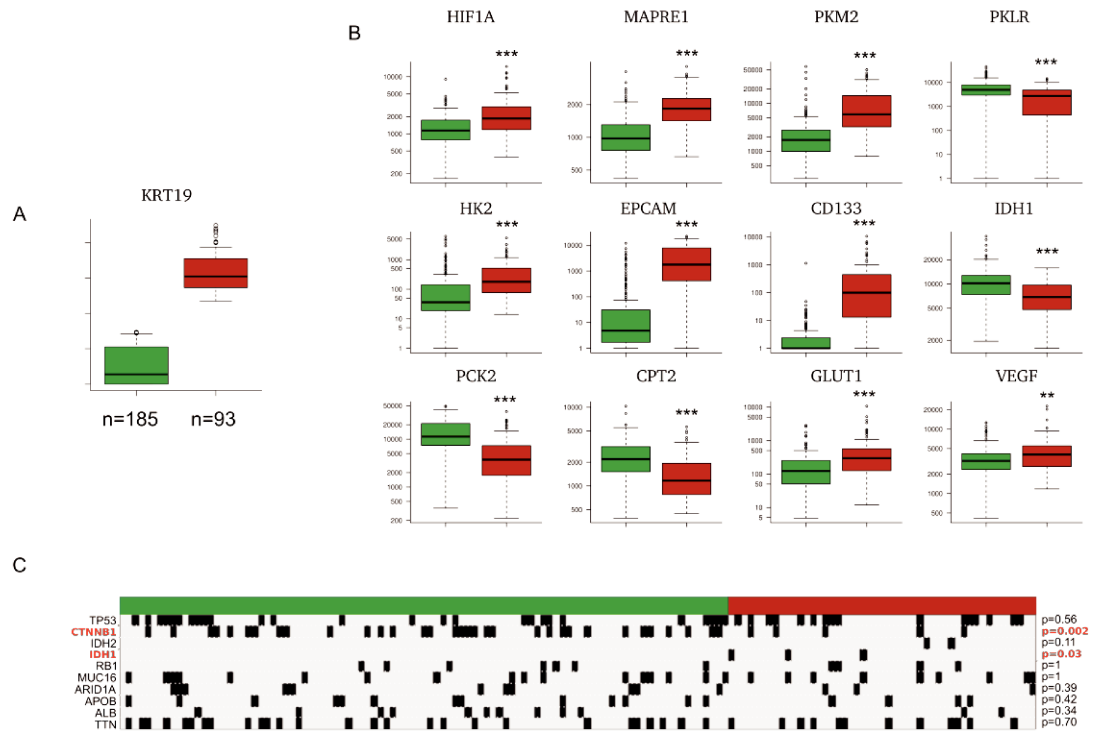


Figure 5 : **Analyse des transcrits de la base de données TCGA.** (A) Deux groupes ont été définis selon l'expression du transcrit KRT19 dans la base de données du TCGA (377 patients). Un groupe d'expression forte comprenant 25 % des tumeurs et un groupe faible comprenant la moitié des tumeurs de faible expression. (B) Expression de HIF1A, de gènes du métabolisme, de l'angiogenèse et de marqueurs progéniteurs (Test de Mann Withney avec correction de Benjamini-Hochberg, \* :  $p < 0,05$  ; \*\* :  $p < 0,01$  ; \*\*\* :  $p < 0,001$ ). (C) Distribution des mutations communes au CHC dans les deux groupes (Test de Fisher). En rouge les deux mutations (CNNTB1 et IDH1), dont la distribution est différente dans les groupes.

PKL/PKM2) ainsi que les marqueurs progéniteurs CD133 et EPCAM sont fortement surexprimés dans le sous-groupe de CHC à forte expression de KRT19. A l'inverse dans cette catégorie de tumeurs, des gènes impliqués dans le cycle de Krebs (PCK2, IDH1) et dans l'oxydation des acides gras (CPT2) sont fortement sous-exprimés (Figure 5B). De plus les gènes codant les protéines identifiées dans l'analyse protéomique quantitative des nodules tumoraux microdisséqués KRT19-positif ont été retrouvés surreprésentés dans le groupe de CHC du TCGA à forte expression de KRT19 (Figure 5B, Suppl Figure 4A). Inversement le sous-groupe de CHC à faible expression de KRT19 était enrichi de la liste des 79 gènes codant les protéines sous-représentées dans les CHC microdisséqués KRT19-positif (Suppl Figure 4B). Pris dans leur ensemble, ces résultats portant sur un total de 438 CHC intégrant des données de protéomique après microdissection laser et de transcriptomique indiquent que KRT19 est un marqueur moléculaire de la reprogrammation métabolique et de la réponse à l'hypoxie. L'analyse des mutations issues des données du TCGA a montré que la plupart des gènes mutés (TP53, ARID1A, RB1, APOB, ALB, TTN) dans les CHC étaient répartis de manière similaire entre les deux sous-groupes de CHC à expression forte ou faible de KRT19 à l'exception des gènes CNNTB1 et IDH1/2 (Figure 5C). Trente-trois CHC sur un total de 38 présentant une mutation du gène CNNTB1 appartenaient au sous-groupe de CHC à faible expression de KRT19 et donc probablement étaient des tumeurs non glycolytiques. Ceci est en accord avec un travail récent montrant que les CHC humains et murins dont la voie  $\beta$ -caténine est activée utilisent préférentiellement les acides gras, et non le glucose, pour leur croissance (Senni *et al.*, 2019). Nous montrons que les 3 CHC mutés pour le gène IDH1 et 2 autres mutés pour le gène IDH2 appartenaient tous au sous-groupe de CHC à expression forte de KRT19 suggérant que ces tumeurs mutées pour les gènes IDH étaient glycolytiques.

Un regroupement hiérarchique des 371 CHC du TCGA réalisé suivant le niveau d'expression d'un set (n=200) de gènes de réponse à l'hypoxie a séparé les tumeurs en deux sous-groupes de forte (n=38) et faible (n=333) réponse à l'hypoxie (Suppl. Figure 5). Le profil d'expression différentiel entre les 2 sous-groupes retrouvait une augmentation des marqueurs de la glycolyse (PKM2, PKLR, HK2), de l'angiogenèse (VEGF), du phénotype progéniteur (EPCAM, CD133, KRT19). Il est à noter que l'expression de certains marqueurs issus de l'analyse protéomique quantitative ont également été

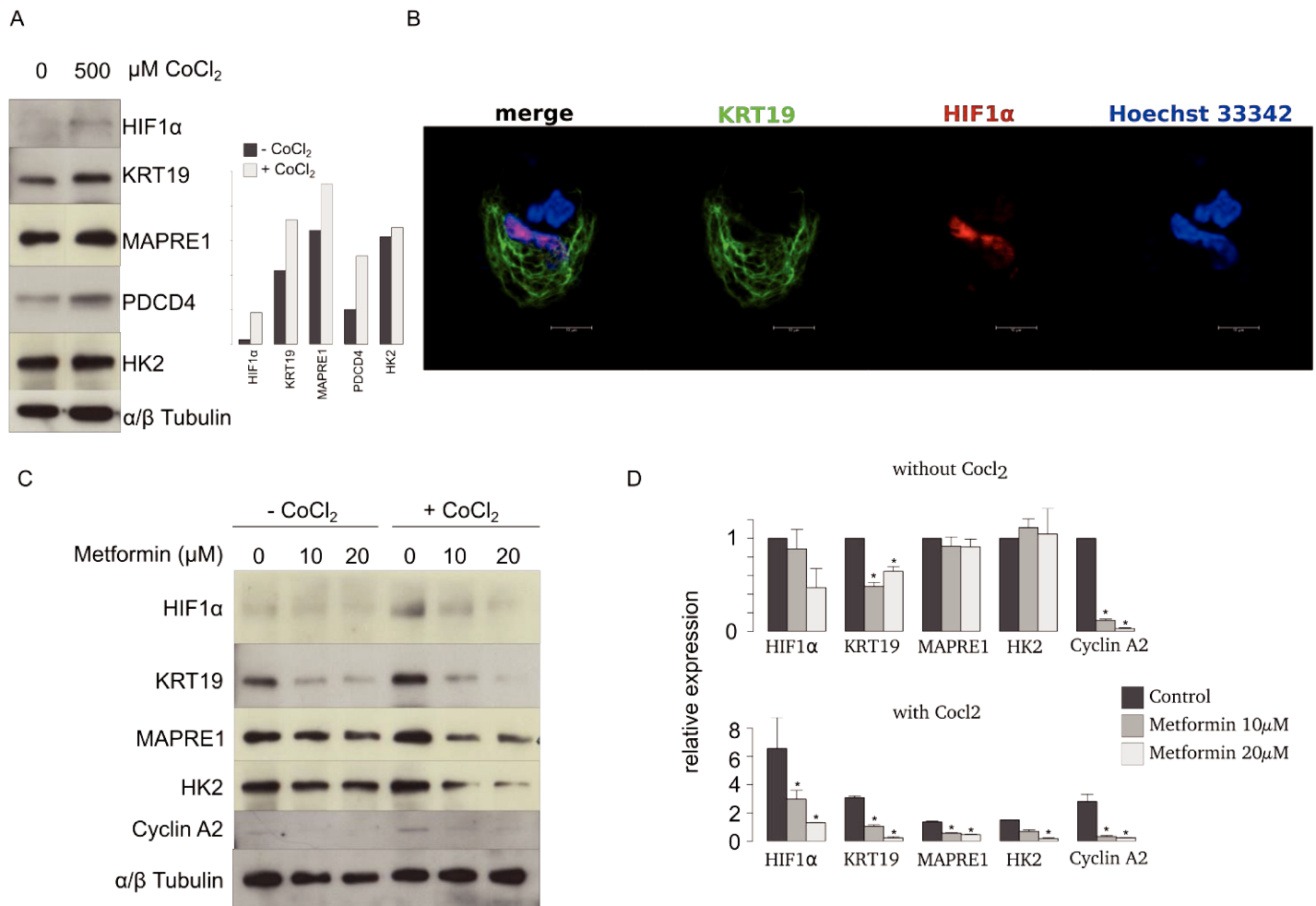


Figure 6 : **La réponse à l'hypoxie provoque l'activation de l'expression de KRT19.** (A) L'incubation pendant 48h de Chlorure de cobalt ( $\text{CoCl}_2$ ) stabilise HIF1 $\alpha$  et permet l'expression de la cytokératine 19, MAPRE1, PDCD4 et HK2 dans de cellules HuH7. (B) Co-marquage de HIF1 $\alpha$  (en rouge) et KRT19 (en vert) dans des HuH7 après incubation au  $\text{CoCl}_2$ . Les noyaux sont colorés en bleu. Barre d'échelle :  $10\mu\text{m}$ . (C) Immunodétection sur membrane de KRT19, MAPRE1, HK2 et Cyclin A après incubation à la metformin en présence ou non de HIF1 $\alpha$  stabilisé et quantification par densitométrie (D).

retrouvés augmentés dans le sous-groupe à forte réponse à l'hypoxie (KRT19, MAPRE1, PKM2). Au total, nous confirmons dans une cohorte indépendante de CHC l'association forte entre réponse à l'hypoxie et expression de KRT19 dans les CHC.

### 3.4 La réponse à l'hypoxie provoque l'activation de l'expression de KRT19

Nous avons étudié dans le modèle cellulaire d'hépatocarcinome humain HuH7 les effets de la stabilisation chimique du facteur HIF1 $\alpha$  par du chlorure de cobalt (CoCl<sub>2</sub>) sur l'expression de certaines protéines préalablement identifiées dans les analyses OMICS tissulaires. Les protéines étudiées ont été KRT19, MAPRE1 (Top 1 de la liste issue de la protéomique quantitative, PDCD4 (Top 1 de la liste issue du RNA Seq). L'immunodétection sur membrane réalisée 48h après traitement des cellules huH7 par CoCl<sub>2</sub> révèle bien la présence de la protéine HIF1 $\alpha$  dans les cellules traitées (Figure 6A). Nous montrons que les expressions de KRT19, MAPRE1, PDCD4 étaient augmentées dans les cellules traitées par le CoCl<sub>2</sub> comparé aux cellules contrôles non traitées. Nous montrons également que cet effet du CoCl<sub>2</sub> était observé dans des cellules HuH7 incubées dans un environnement hypoxique à 1% O<sub>2</sub> (données non présentées). L'imagerie par immunofluorescence a montré que les cellules HuH7 traitées par CoCl<sub>2</sub>, et non les cellules contrôles non traitées, présentaient à la fois la présence dans le noyau de HIF1 $\alpha$  et dans le cytoplasme du réseau cytoplasmique de filaments intermédiaires de KRT19 du cytosquelette (Figure 6B).

Nous avons ensuite testé si la metformine connue pour son action d'inhibition de HIF1 $\alpha$  et ses gènes cibles pouvait modifier le niveau d'expression des protéines étudiées dans des cellules HuH7 traitée ou non par CoCl<sub>2</sub>. Nous montrons que les deux concentrations de metformine utilisées (10 et 20 $\mu$ M) réduisent fortement l'expression des protéines KRT19, MAPRE1, HK2 et cycline A2 dans les cellules HuH7 traitées comparé aux cellules contrôles (Figure 6C). Cette diminution d'expression protéique était associée à une forte réduction de l'expression de HIF1 $\alpha$  en présence de metformine indiquant que la stabilisation de HIF1 $\alpha$  est le facteur déclencheur de l'activation de l'expression des marqueurs sélectionnés (notamment KRT19) et plausiblement des changements

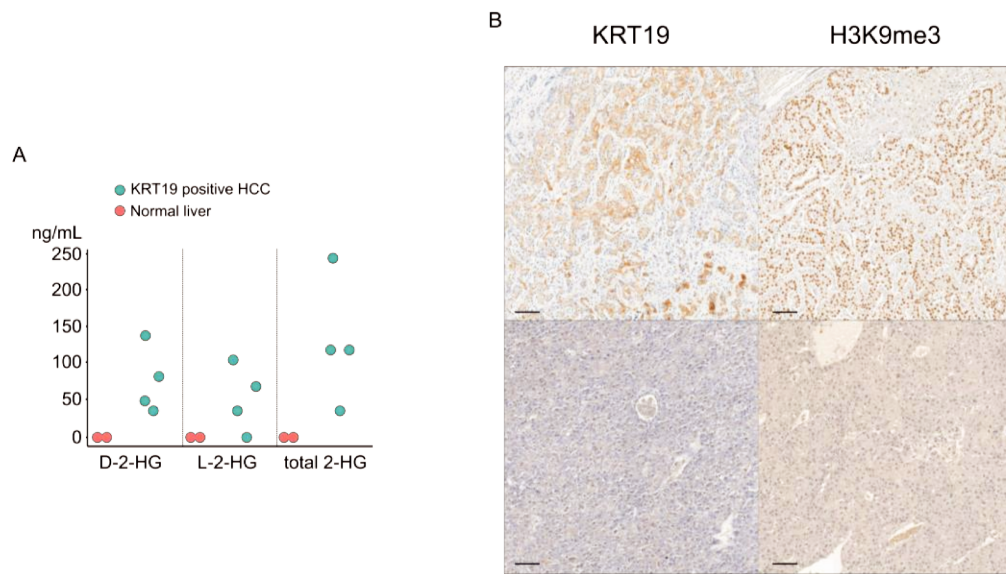


Figure 7 : **Production de 2-Hydroxyglutarate.** (A) Dosage par spectrométrie de masse de la 2-hydroxyglutarate dans des tumeurs à double différenciation et des foies normaux. (B) Images représentatives du marquage H3K9 tri-méthylé dans les CHC (barre d'échelle : 100 $\mu$ m).

moléculaires identifiées dans les CHC KRT19-positif. Cette affirmation est appuyée par l'existence d'un élément de réponse à l'hypoxie (HRE) à 4,6 Kb du site d'initiation de transcription du gène KRT19. Enfin, il est à noter que dans des conditions normales de culture sans CoCl<sub>2</sub>, la metformine réduisait l'expression protéique de KRT19 et de cycline A2, et non de MAPRE1 et HK2, dans les cellules contrôles suggérant l'existence de facteurs indépendants de HIF1 $\alpha$  dans la régulation négative de KRT19 (Figure 6C).

### 3.5 Les CHC hypoxiques produisent du 2-hydroxyglutarate

Sachant que l'hypoxie peut être responsable de la surproduction de 2-Hydroxyglutarate (2-HG) en absence de mutation dans les gènes IDH et que cette production contribue à la pathogenèse de certains cancers solides (altérations épigénétiques conduisant à la répression de la transcription, altérations de la différenciation tumorale, survie des cellules cancéreuses)(Intlekofer *et al.*, 2015; Oldham *et al.*, 2015). nous avons formulé l'hypothèse selon laquelle des CHC hypoxiques et non mutés IDH produisaient du 2-HG. L'oncométabolite 2-HG est un inhibiteur des dioxygénases dépendantes de l' $\alpha$ -cétoglutarate, parmi lesquelles des déméthylases des histones (KDM4C, JMJD2A, JMJD2C, JHDM1A/FBXL11) ou encore la prolyl hydroxylase PHD2 nécessaire à l'hydroxylation de HIF1 $\alpha$  conduisant à sa dégradation (Chowdhury *et al.*, 2011). Nous avons quantifié par spectrométrie de masse le taux tissulaire de 2-HG dans des CHC KRT19-positif qui n'étaient pas mutés pour les gènes IDH. Nos résultats préliminaires obtenus sur un très faible nombre de tumeurs (n=4) a montré la présence de 2-HG sous ses deux formes énantiomériques D (D2-HG) et L (L2-HG) à des taux variables (35-240 ng/mL). Ces deux formes D2-HG et L2-HG n'ont pas été détectées dans les 2 seuls foies non tumoraux analysés et collectés à distance d'une métastase colorectale (Figure 7A). Nous avons analysé par immunohistochimie le statut d'une marque particulière d'histone, la triméthylation de la lysine 9 de l'histone H3 (H3K9me3) qui résulte de l'inhibition de KDM4C par 2-HG dans une série de CHC hypoxiques KRT19-positif (n=3) ou KRT19-négatif (n=4) KRT19 et des CHC non hypoxiques (n=4). Nous montrons que toutes les tumeurs hypoxiques qu'elles soient KRT19-positif ou -négatif présentent une marque H3K9me3 nucléaire forte dans les nodules tumoraux comparés au foie non tumoral adjacent (Figure 7B). Par contre, nous ne détectons aucune accumulation de H3K9me3

dans les CHC n'exprimant ni KRT19 ni les gènes de réponse à l'hypoxie. Au total, ces données préliminaires indiquent que les CHC hypoxiques exprimant ou non KRT19 produisent du 2-HG indépendamment de l'existence ou non de mutations dans les gènes IDH ce qui conduit à une accumulation aberrante de H3K9me3 et vraisemblablement à des altérations du métabolisme énergétique et de la différenciation tumorale.

## 4. Discussion

Le grand nombre d'études moléculaires réalisées cette dernière décennie ont révélé plusieurs classes de CHC caractérisés par des signatures génomiques et protéomiques distinctes, qui pourraient résulter de l'activation de différentes voies de signalisation oncogéniques et/ou inflammatoires ou provenir de type cellulaire particulier. Un sous-groupe génomique, dont le pronostic est particulièrement mauvais, présente un profil d'expression génique comprenant des marqueurs d'hépatoblastes et de cellules ovales hépatiques, suggérant qu'il puisse dériver de cellules progénitrices bipotentes du foie ou d'altérations de la différenciation hépatocytaire au cours de la progression tumorale. Ce sous-groupe de CHC dit «progéniteur-like» a aussi une expression significativement plus élevée du marqueur de différenciation biliaire KRT19 que les autres sous-groupes moléculaires. Les découvertes génomiques récentes n'ont malheureusement eu jusqu'à présent qu'un impact clinique et thérapeutique limité.

L'objectif de l'étude a été de mieux caractériser au moyen de méthodes combinées de spectrométrie de masse quantitative et de transcriptomique le paysage moléculaire et phénotypique du sous-groupe agressif de CHC KRT19-positif avec l'espoir d'identifier des cibles thérapeutiques spécifiques à ce cancer ayant un réel impact sur le devenir des patients.

Ces méthodes associées à des analyses *in silico* appliquées à un grand nombre de CHC ont révélé une dérégulation de plusieurs voies moléculaires associées au statut KRT19 de la tumeur, et en particulier une reprogrammation métabolique des CHC KRT19-positif de type glycolytique liée à une activation de la réponse à l'hypoxie. Nous montrons également que l'expression tumorale de marqueurs dits progéniteurs tels que KRT19, EPCAM et CD133 était dépendante de la stabilisation de HIF1 $\alpha$ , et que cette activation d'expression médiée par HIF1 $\alpha$  était détectée *in vitro* dans des cellules d'hépatocarcinome.

Les CHC KRT19-positif sont caractérisés par (i) une forte inhibition du facteur de différenciation hépatocytaire HNF4 $\alpha$  entraînant une perte de la différenciation cellulaire et



(ii) une accumulation de HIF1 $\alpha$  et (iii) une activation des gènes cibles de MYC, provoquant des changements métaboliques substantiels affectant de nombreuses voies et rendant la tumeur glycolytique.

Certains oncogènes dont MYC participe à la reprogrammation métabolique de type glycolytique des cancers solides. L'analyse protéomique de cellules tumorales microdisséquées KRT19-positifs et -négatifs a révélé un grand nombre de cibles de MYC parmi lesquelles EIF2, EIF5, GNAS, NHP2L1, CAPRIN1, RPL27A, RPL35, PCNA, NPM3 et PTMA dans la condition KRT19 positif. Ce résultat est en accord avec les données de la littérature montrant que MYC est donc capable d'activer des gènes cibles de la voie de la glycolyse, du transport du glucose et la glutaminase (Yuneva *et al.*, 2007; DeBerardinis *et al.*, 2008). La réponse à l'hypoxie est fortement activée dans les CHC KRT19-positif suggérant un effet coopératif de MYC et de HIF1 $\alpha$  comme cela a été montré dans de nombreux cancers en conditions hypoxiques conduisant à l'expression de gènes de l'angiogenèse et de la glycolyse (Kim *et al.*, 2007; Dang *et al.*, 2008; Yeung *et al.*, 2008). MYC est aussi capable de stabiliser HIF1 $\alpha$  et d'augmenter son accumulation nucléaire en conditions normoxiques (Doe *et al.*, 2012). De plus il a été montré que la stabilisation de HIF1 $\alpha$  était nécessaire à l'activité proliférative de MYC dans des lignées de cancer du sein indiquant que MYC et HIF1 $\alpha$  sont donc deux protéines intimement liées dans la pathogenèse tumorale.

Les tumeurs hypoxiques sont connues pour être particulièrement résistantes aux traitements anticancéreux (chimiothérapie, radiothérapie, thérapies ciblées) (Ward, 1988; Karakashev and Reginato, 2015). Plusieurs mécanismes de résistance primaire ou secondaire entrent en jeu. 1- HIF1 $\alpha$  induit l'expression de glycoprotéines membranaires (MDR1, MRP2) responsables de l'efflux des xénobiotiques et donc de la diminution de leurs concentrations intracellulaires (Kruh and Belinsky, 2003; Krishnamurthy and Schuetz, 2005). 2- La réponse à l'hypoxie réduit l'apoptose liée aux dommages à l'ADN causés par les agents cytotoxiques en inhibant les protéines pro-apoptotiques BAX et BID (Erler *et al.*, 2004) ou en activant des protéines anti-apoptotiques comme la survivine (Chen *et al.*, 2009). Il a été montré dans des lignées d'hépatocarcinome et des modèles animaux de CHC que l'inhibition de HIF1 $\alpha$  permettait de resensibiliser les cellules

cancéreuses à la doxorubicine et au sorafenib (Zhu *et al.*, 2012; Cao *et al.*, 2013; Liang *et al.*, 2013) ou encore à la radiothérapie (Schwartz *et al.*, 2011; Yang *et al.*, 2011; Meijer *et al.*, 2012). 3- Le pyruvate et le lactate produits de la glycolyse ainsi que la génération de formes réduites de NAD(P)H et de glutathion (GSSH) réduisent le stress oxydatif des cellules tumorales traitées par irradiations, par un effet direct anti-oxydant participant ainsi aux mécanismes de radiorésistance (Cook *et al.*, 2004; Sattler and Mueller-Klieser, 2009). Par exemple, le lactate piège les radicaux superoxyde et hydroxyle (Groussard *et al.*, 2000) et l'inhibition du transporteur du lactate MCT4 par l'acide  $\alpha$ -cyano-4-hydroxycinnamique réduit la concentration intracellulaire de lactate et améliore la radiosensibilité de lignées de gliomes (Colen *et al.*, 2006). Tous ces éléments montrent, d'une part, que les thérapies anticancéreuses conventionnelles sont fréquemment inefficaces pour détruire des tumeurs glycolytiques, et que d'autre part, la combinaison à des traitements régulateurs du métabolisme tumoral pourrait améliorer la réponse anti-tumorale de ces tumeurs.

Nos données montrent également que les CHC KRT19-positifs présentent un profil d'expression génique similaire retrouvés dans le CHC mutés pour les gènes isocitrate déshydrogénase (IDH1 et IDH2) (Suppl Figure 6). L'absence de mutations IDH dans les tumeurs KRT19-positifs comme dans la plupart des CHC rapportés à la différence du cholangiocarcinome intra-hépatique ne peut donc expliquer ce profil particulier. Nous montrons que les CHC KRT19-positifs néanmoins produisent l'oncométabolite 2-HG, qui est le produit de la conversion de l' $\alpha$ -cetoglutarate par les enzymes IDH mutés et responsable des effets métaboliques à de la dédifférenciation hépatocytaire (Saha *et al.*, 2014). La présence tissulaire de 2HG a été corrélée à une aberration de la méthylation de la lysine 9 de l'histone H3 dans les nodules tumoraux KRT19-positifs et les nodules KRT19 négatifs à forte réponse à l'hypoxie. Ce même type d'observation avait été rapporté dans des cellules tumorales de gliome placées dans un environnement hypoxique (Intlekofer *et al.*, 2015) et présentant un profil aberrant hyperméthylé. L'excès de 2-hydroxyglutarate est capable de promouvoir la transformation maligne de cellules hématopoïétiques dans un modèle cellulaire (Losman *et al.*, 2013). Cette transformation cellulaire reposerait en partie sur l'inhibition fonctionnelle des dioxygénases dépendantes de l' $\alpha$ -cetoglutarate (comme les histones déméthylases) conduisant à une reprogrammation métabolique HIF1 $\alpha$

dépendante. Sur le plan clinique, le dosage non invasif du 2-HG pourrait être utile comme marqueur de progression et d'agressivité tumorale chez les patients atteints de CHC notamment ceux bénéficiant de TACE. Une élévation au cours du temps du taux sérique de 2HG chez les patients cancéreux serait le signe indirect d'une reprogrammation cellulaire vers une perte de différenciation hépatocytaire et l'acquisition d'un phénotype « progéniteur ».

Dans notre étude, les patients avec des tumeurs CHC glycolytiques KRT19-positif ont davantage bénéficié subi d'une chimioembolisation trans-artérielle (TACE) pré-opératoire (TACE) que les patients avec des CHC KRT19-négatif, suggérant que le micro-environnement nécrotique et hypoxique induit par TACE contribuerait à l'amplification de l'expression de KRT19, et peut-être à la sélection de cellules tumorales résistantes en cours de reprogrammation métabolique et phénotypique. Cette hypothèse est soutenue par des données cliniques de plusieurs études montrant que des CHC traités par TACE puis réséqués chirurgicalement présentaient des paramètres anatomopathologiques de franche agressivité et exprimaient des marqueurs progéniteurs dont KRT19 (Nishihara *et al.*, 2008, Nahm *et al.*, 2017a, Petrillo *et al.*, 2018b) L'expression par la tumeur de marqueurs progéniteurs de type KRT19, stigmatisme moléculaire de l'hypoxie comme montré dans notre étude, pourrait dans certains cas de CHC être un phénomène iatrogène qu'il serait important de contrôler pharmacologiquement.

Sur le plan thérapeutique, il existe des molécules enregistrées ou en développements capables de perturber l'état métabolique des cellules tumorales et de les rendre plus sensibles aux thérapies anticancéreuses classiques. La metformine, traitement classique des diabètes de type 2, est un bon exemple de candidat médicament pouvant potentialiser via ses effets métaboliques de réversion de la glycolyse de type Warburg les effets cytotoxiques des chimiothérapies anti-cancéreuses. Notre étude montre que la metformine est capable de réduire l'expression de gènes de la glycolyse et de marqueurs progéniteurs comme KRT19 dans des cellules d'hépatocarcinome dans un contexte d'activation de la réponse à l'hypoxie. Il serait intéressant de comparer les taux de survie et d'expression génique tissulaire (marqueurs progéniteurs, métabolisme) dans des cohortes de patients

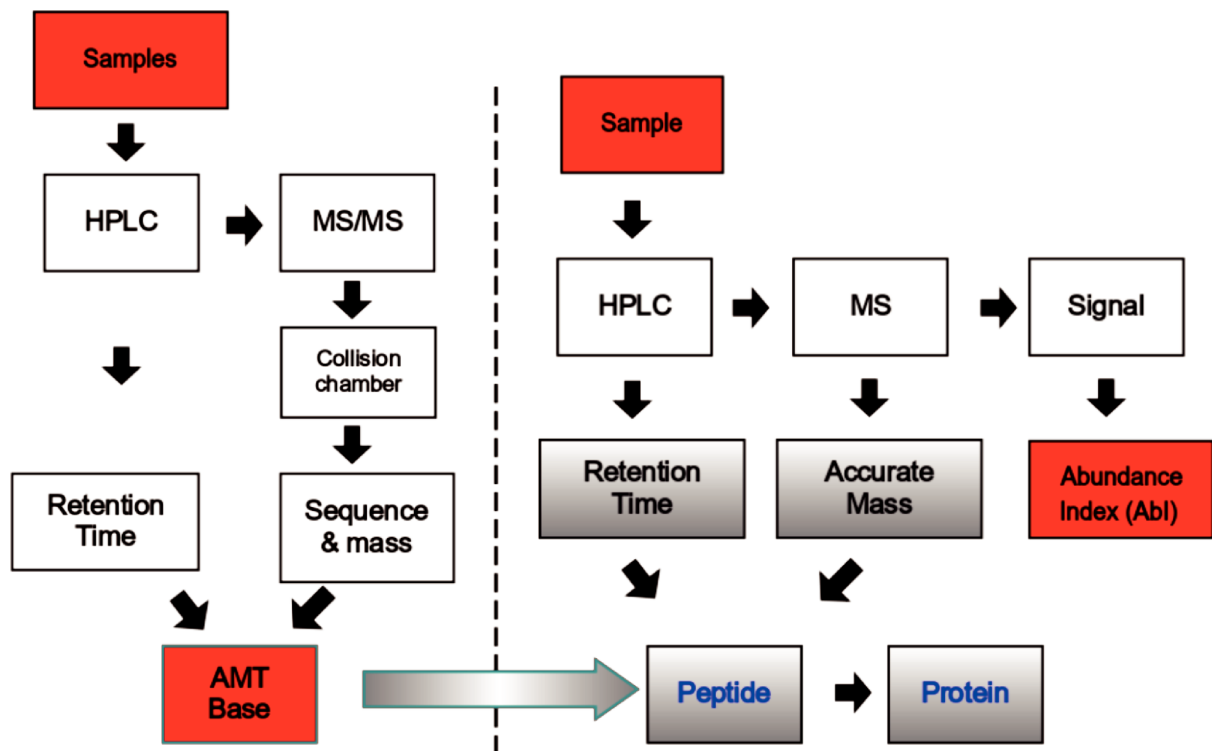
diabétiques et cancéreux traités par TACE et metformine comparé à des patients cancéreux non diabétiques.

Notre étude et d'autres études ont montré que les voies du TGF- $\beta$  et de l'EMT étaient activées dans les CHC KRT19-positif. L'EMT et l'angiogenèse sont deux des principaux facteurs de résistance au traitement par les thérapies ciblées comme le sorafenib (Huang *et al.*, 2013; Liang *et al.*, 2013). Il a été montré que la metformine avait également le pouvoir d'inhiber l'EMT (Cheng and Hao, 2016; Leonel *et al.*, 2017). Une combinaison associant le sorafenib et la metformine pourrait donc être une option thérapeutique valide pour lever la résistance tumorale au sorafenib. In vitro, l'association des deux médicaments réduisait davantage la prolifération tumorale que le sorafenib seul (Hsieh *et al.*, 2014; You *et al.*, 2016; Ling *et al.*, 2017; Chung *et al.*, 2018). Actuellement un essai clinique de phase 2 (NCT02672488) teste l'efficacité de cette combinaison chez des patients atteints de CHC de stade avancé (BCLC-C).

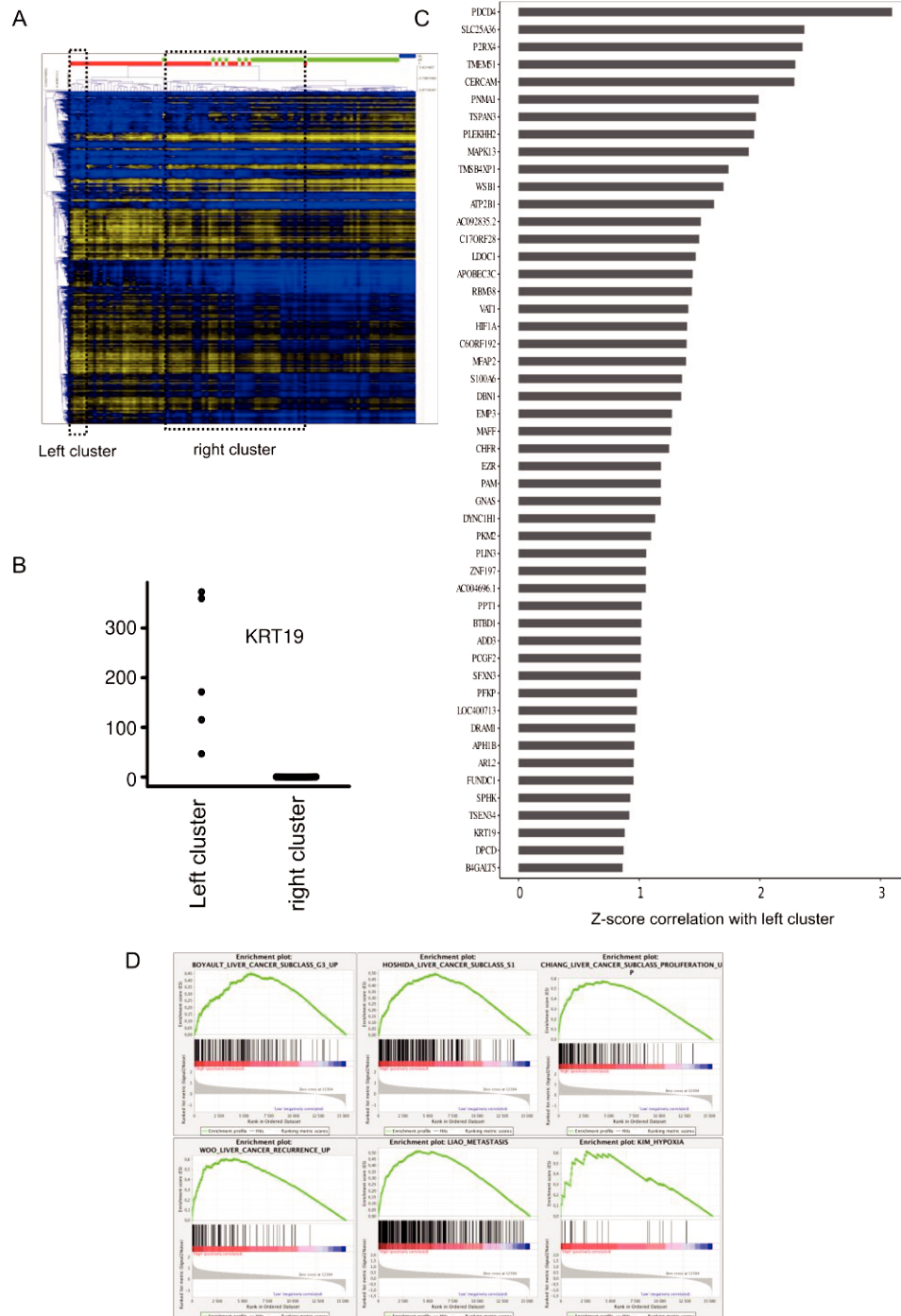
## 5. Conclusions et perspectives

Nous montrons dans cette étude que les CHC KRT19-positif sont des tumeurs fortement glycolytiques présentant une activation forte de la réponse à l'hypoxie et des aberrations épigénétiques liées à un excès de l'oncométabolite 2-HG. La caractérisation de la biologie de cette tumeur a dégagé plusieurs pistes d'investigation clinique : l'intérêt potentiel de la metformine en association avec les traitements anticancéreux conventionnels incluant TACE, le suivi non invasif du taux sérique du 2-HG chez les patients cancéreux voire peut-être cirrhotiques sans CHC. Nous pouvons envisager de colliger des données de patients diabétiques et traités par la metformine qui subissent une chimiothérapie pré-opératoire pour mettre en évidence un effet « préventif » du médicament sur la survie et la récurrence de ces patients. Le dosage du 2-hydroxyglutarate (2-HG) sérique chez ces patients, s'il s'avère être un marqueur efficace de la reprogrammation, pourrait permettre de suivre ces patients. Enfin à titre de prévention et de part sa nature « oncogénique », un dosage régulier du 2-HG sérique chez des patients cirrhotiques devrait être conduit dans une grande cohorte afin de tester le pouvoir diagnostique de cette molécule.

## 6. Figures et tableaux supplémentaires

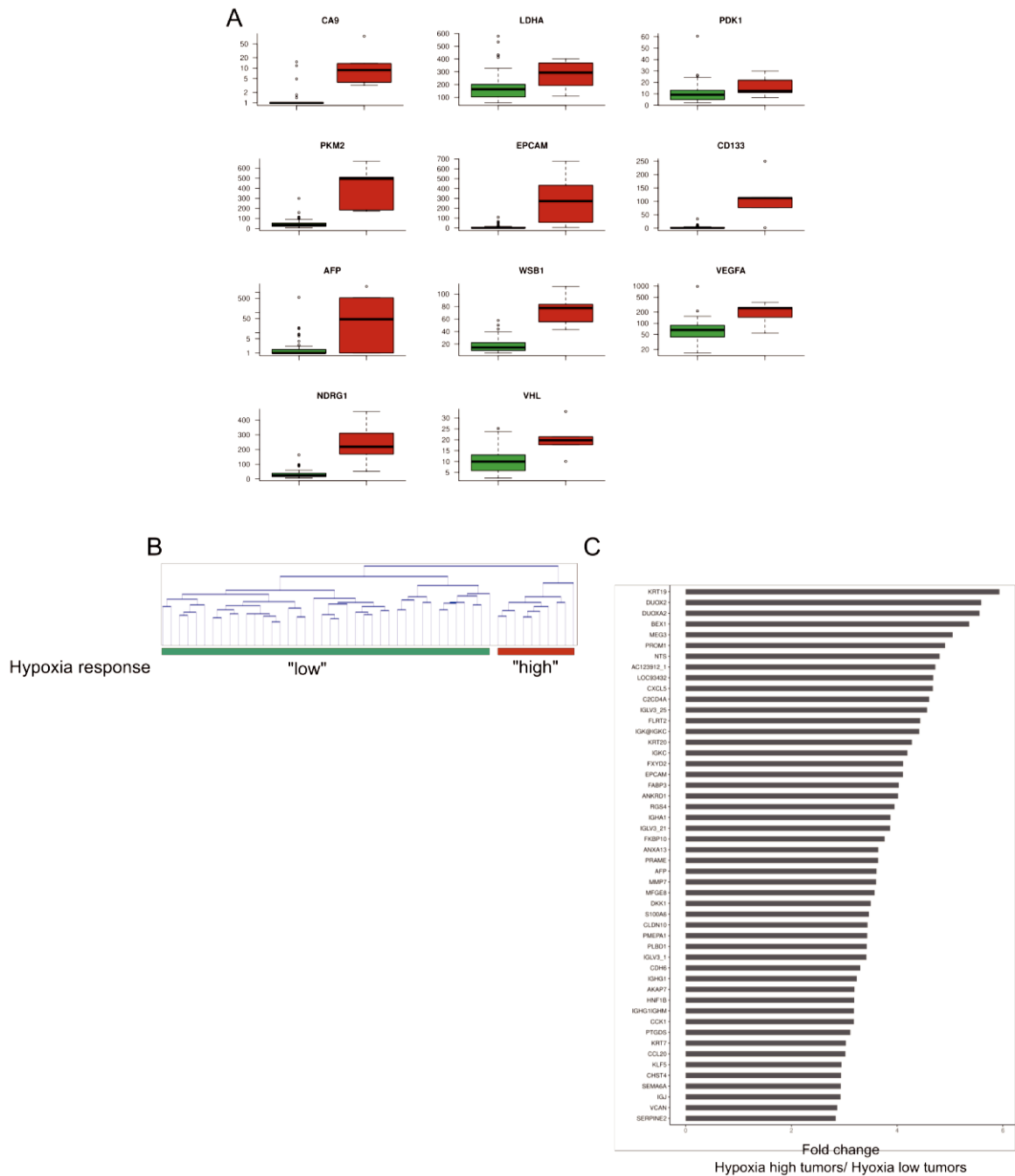


Suppl Figure 1 : **Constitution de la base AMT et quantification** : Dans un premier temps (à gauche), les échantillons peptidiques sont séparés dans une chromatographie liquide de haute performance (HPLC) et sont séquencés (par MS en tandem) pour obtenir une masse et une séquence peptidique. Le temps de rétention dans la chromatographie, la séquence et la masse précise du peptide sont utilisés pour créer la base AMT. Dans un deuxième temps (à droite), les mêmes échantillons sont à nouveau séparés dans l'HPLC et leur masse précise est mesurée par une MS simple. De plus l'intensité du signal peptidique est enregistrée. À l'aide du temps de rétention et de la masse, on se reporte à la base AMT, pour remonter à l'identité protéique et l'intensité du signal est utilisée pour calculer l'index d'abondance (AbI)



Suppl Figure 2 : **Analyse des transcrits ARN par séquençage CAGE.** (A) La classification hiérarchique non supervisée a permis l'identification de deux groupes extrêmes de CHC (« left cluster » et « right cluster ») caractérisé par une différence significative de l'expression de la KRT19 (B). Une analyse edgeR a été réalisée entre ces deux groupes et 420 gènes ont été retrouvés différemment exprimés. En C les cinquante gènes les plus discriminants sont représentés en fonction de la corrélation avec le groupe « left cluster ». (D) L'analyse par GSEA montre une corrélation entre le carcinome à double différenciation avec le groupe G3 de Boyault et al., S1 de Hoshida et al., le groupe « prolifératif » de Chiang et al., de CHC avec récurrence de Woo et al., de CHC métastatique de Liao et al. et de fibroblastes cultivés sous conditions hypoxiques de Kim et al.

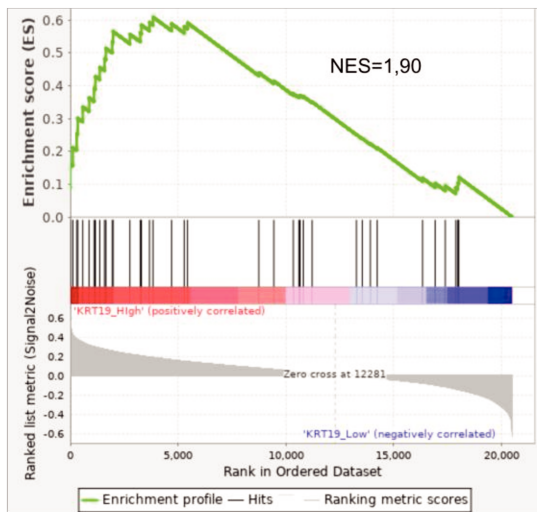




Suppl Figure 3 : **Expression de gènes de réponse à l'hypoxie.** (A) Expression de gènes cibles de HIF1 $\alpha$  et de marqueurs progéniteurs (Test de Mann Withney avec correction de Benjamini-Hochberg, \* :  $p < 0,05$  ; \*\* :  $p < 0,01$  ; \*\*\* :  $p < 0,001$ ). (B) A partir d'une signature de gènes impliqués dans la réponse à l'hypoxie (HALLMARKS\_HYPOXIA, Broad Institute), détermination de deux groupes fort et faible d'expression de gènes et rapport d'expression des 50 gènes les plus discriminants (C) entre ces deux groupes.

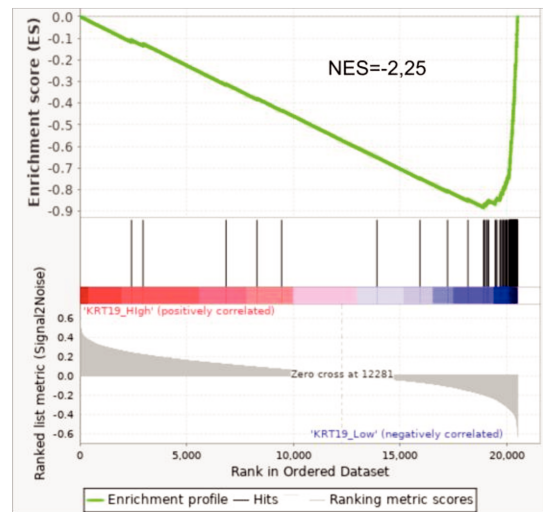
A

39 over-expressed on KRT19 positive HCC

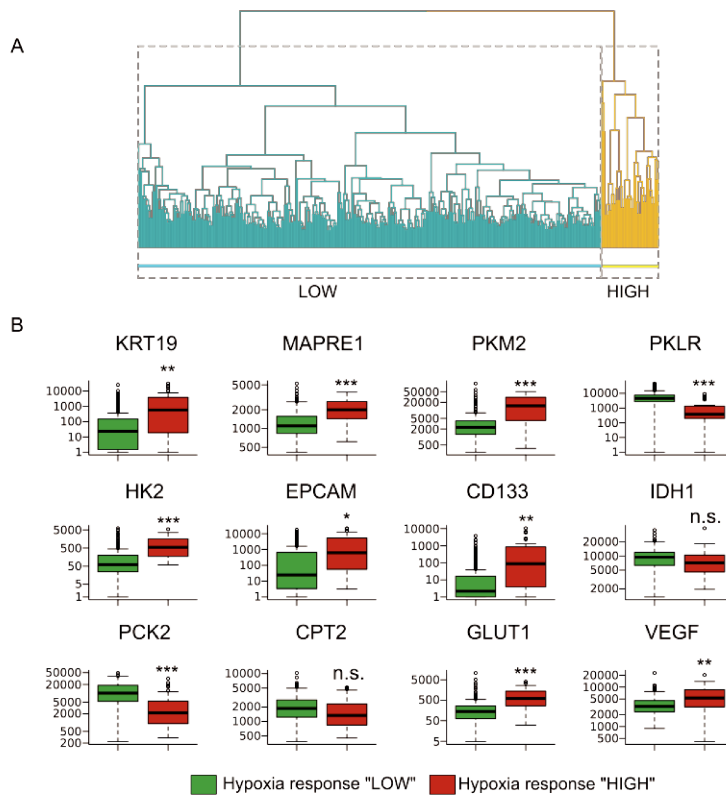


B

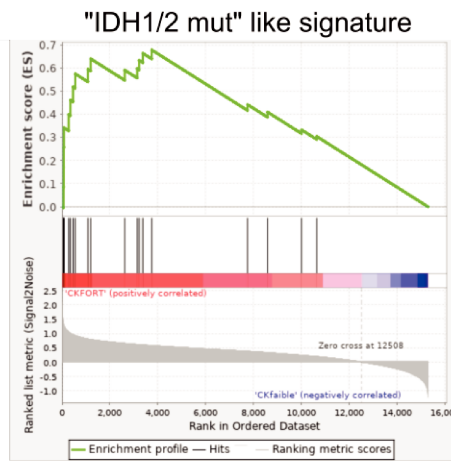
79 under-expressed on KRT19 positive HCC



Suppl Figure 4 : **Corrélation entre l'expression protéique et transcriptionnelle de la signature.** Les 39 (à gauche) et 79 (à droite) protéines sur- et sous-représentés respectivement dans le carcinome hépatocellulaire à double différenciation ont été corrélés aux deux groupes transcriptionnels via l'algorithme GSEA.



Suppl Figure 5 : **Expression de gènes de réponse à l'hypoxie de la cohorte TCGA** (A) Deux groupes ont été définis selon une signature de gènes dérégulés dans l'hypoxie (HALLMARKS\_HYPOXIA, Broad Institute) dans la base de données du TCGA (377 patients). (B) Expression de KRT19, de gènes du métabolisme, de l'angiogenèse et de marqueurs progéniteurs dans ces deux groupes (Test de Mann Withney avec correction de Benjamini-Hochberg, \* :  $p < 0,05$  ; \*\* :  $p < 0,01$  ; \*\*\* :  $p < 0,001$  ; n.s.: non significatif).



Suppl Figure 6 : **Signature IDH1/2 muté**. Le transcriptome des CHC positifs pour la cytokératine 19 (à gauche) montre un enrichissement de gènes de la signature « IDH1/1 muté » (Ally et al. 2017).

<b>Gene Symbol</b>	<b>Protein name</b>	<b>Fold Change</b>
HAO1	Hydroxyacid oxidase (Glycolate oxidase) 1	*
AOX1	Aldehyde oxidase	*
UGT1A4	UGT1A4 protein	*
FMO3	Dimethylaniline monooxygenase [N-oxide-forming] 3	*
PXMP2	Peroxisomal membrane protein 2	*
SLC25A15	solute carrier family 25 (mitochondrial carrier; ornithine transporter) member 15 (SLC25A15)	*
HSD17B6	17-beta-hydroxysteroid dehydrogenase type 6	*
IFITM1	Interferon-induced transmembrane protein 1	*
PON1	Serum paraoxonase/arylesterase 1	*
RDH16	Retinol dehydrogenase 16	*
SLC27A5	Bile acyl-CoA synthetase	*
TMEM85	Transmembrane protein 85	*
UGT2B4	UDP-glucuronosyltransferase 2B4	*
H6PD	GDH/6PGL endoplasmic bifunctional protein	465.34
GALK2	N-acetylgalactosamine kinase	125.26
OTC	Ornithine carbamoyltransferase	65.79
MUC5AC	Mucin 5AC	43.35
BAAT	Bile acid-CoA:amino acid N-acyltransferase	41.02
ALDOB	Fructose-bisphosphate aldolase B	30.65
ALDH1L1	10-formyltetrahydrofolate dehydrogenase	20.50
EHHADH	Enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase	19.23
AKR7A3	Aflatoxin B1 aldehyde reductase member 3	18.82
ADH1B	Alcohol dehydrogenase 1B	15.94
AKR1B10	Aldo-keto reductase family 1 member B10	14.81
EBP	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase	13.35

<b>Gene Symbol</b>	<b>Protein name</b>	<b>Fold Change</b>
PTGR1	Prostaglandin reductase 1	12.58
MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic	12.42
ABAT	4-aminobutyrate aminotransferase, mitochondrial	11.53
PCK2	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	10.45
ACOX1	Acyl-Coenzyme A oxidase 1, palmitoyl	8.99
IVD	Isovaleryl-CoA dehydrogenase, mitochondrial	8.52
CRYL1	Lambda-crystallin homolog	8.51
AADAC	Arylacetamide deacetylase	7.41
ALDH3A2	Fatty aldehyde dehydrogenase	7.28
SOD2	Superoxide dismutase	7.17
SHMT1	Serine hydroxymethyltransferase	6.66
BDH1	D-beta-hydroxybutyrate dehydrogenase, mitochondrial	6.29
SCP2	sterol carrier protein 2 (SCP2)	5.83
AKR1A1	Alcohol dehydrogenase [NADP+]	5.57
HIBADH	3-hydroxyisobutyrate dehydrogenase, mitochondrial	5.56
HGD	homogentisate 1,2-dioxygenase (homogentisate oxidase) (HGD)	5.35
ALDH1A1	Retinal dehydrogenase 1	5.17
ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	5.15
GOT1	Aspartate aminotransferase	4.88
ALDH4A1	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	4.83
PNPO	Pyridoxine 5'-phosphate oxidase	4.61
REEP6	Receptor expression-enhancing protein 6	4.60
CBR1	Carbonyl reductase [NADPH] 1	4.49
MAT1A	S-adenosylmethionine synthetase	4.28
UK114	Heat-responsive protein 12	4.20
ACSM2	acyl-CoA synthetase medium-chain family member 2	4.17
AGXT	Serine--pyruvate aminotransferase	4.07

<b>Gene Symbol</b>	<b>Protein name</b>	<b>Fold Change</b>
SULT2A1	Sulfotransferase family, cytosolic, 2A,	4.00
BLVRB	Flavin reductase	3.92
FAH	Fumarylacetoacetate hydrolase	3.91
UGP2	UTP--glucose-1-phosphate uridylyltransferase 2	3.80
ACAA1	3-ketoacyl-CoA thiolase, peroxisomal	3.80
ACOT13	Acyl-coenzyme A thioesterase 13	3.61
CPT2	Carnitine O-palmitoyltransferase 2, mitochondrial	3.57
DPYS	Dihydropyrimidinase	3.44
HMGCL	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase	3.40
HASPIN	Serine/threonine-protein kinase haspin	3.35
BPHL	Valacyclovir hydrolase	3.35
SORD	Sorbitol dehydrogenase	3.29
AKR1C2	Aldo-keto reductase family 1 member C2	3.12
GPD1	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	2.80
DDT	D-dopachrome decarboxylase	2.79
IQGAP2	IQ motif containing GTPase activating protein 2	2.57
SEC61B	Protein transport protein Sec61 subunit beta	2.52
POR	NADPH--cytochrome P450 reductase	2.31
TMEM97	Sigma intracellular receptor 2	2.21
TOLLIP	Toll interacting protein	2.05
B2M	Beta-2-microglobulin	1.60
RPL34	60S ribosomal protein L34	1.50
IFI30	Gamma-interferon-inducible lysosomal thiol reductase	1.50
UQCR10	Cytochrome b-c1 complex subunit 9	1.48
RHOA	Transforming protein RhoA	1.40
RAB10	Ras-related protein Rab-10	1.31
TFDP1	Transcription factor Dp-1	1.10

Suppl Tableau 1 : **Liste des protéines sous -exprimés dans le CHC à double différenciation.** Liste des protéines dont les peptides identifiés ont été retrouvés statistiquement moins abondant dans les CHC positifs pour la cytokératine 19. La première colonne est la dénomination en symbole de gène, a deuxième, le nom officiel des protéines (Uniprot) et la dernière colonne représente le rapport d'expression entre les cancers cytokératine 19 négatifs versus les CHC cytokératine 19 positifs. \* : Pour ces 13 protéines, les peptides n'ont été identifiés que dans une seule condition. Le rapport ne peut donc pas être calculé.



---

# IV ETUDE 3 : RÔLE DE LA LECTINE REG3A DANS L'HOMÉOSTASIE DU MICROBIOTE INTESTINAL : CONTRÔLE DE L'INFLAMMATION DANS UN MODÈLE DE COLITE

## 1. Présentation de l'étude

La dysbiose intestinale est un état modifié de la flore commensale soupçonnée d'être un facteur pathogène clé pour de nombreuses maladies inflammatoires chroniques. Une diminution de la diversité et un changement de composition du microbiote participeraient aux altérations de la barrière épithéliale, à la diffusion de produits bactériens pro-inflammatoires, et au déclenchement/amplification d'états immuno-inflammatoires dans toutes sortes de tissus. Il est connu que les cellules de Paneth situées dans la crypte intestinale libèrent dans la lumière intestinale des peptides et protéines antimicrobiens tels que les lectines de type C (REG3A humain; Reg3g et Reg3b murins). Il est plausible, mais non prouvé, que cet arsenal antimicrobien préserve l'homéostasie du microbiote et protège l'organisme contre des stress inflammatoire et métabolique.

Ce projet a porté sur une AMP qui est la lectine humaine REG3A (HIP/PAP) décrite comme étant une protéine antimicrobienne sécrétée qui lie certains polysaccharides. Elle présente *in vitro* une activité bactéricide en interagissant avec certaines cibles glucidiques comme le peptidoglycane de la paroi bactérienne des bactéries Gram positif (Hooper *et al.*, 2001; Cash *et al.*, 2006; Vaishnava *et al.*, 2011). L'expression des gènes REG3 est induite par les infections et l'inflammation. Notre équipe a précédemment montré que REG3A présentait une activité anti-oxydante en piégeant les espèces réactives de l'oxygène (ROS), notamment les radicaux hydroxyles ( $\text{OH}^\circ-$ ) et les anions superoxyde ( $\text{O}_2^{\circ-}$ ) (Moniaux *et al.*, 2011, 2015). Cette activité anti-oxydante explique pour partie les nombreuses observations décrivant une forte activité anti-inflammatoire des protéines REG3 dans une variété de types cellulaires et de tissus (foie, système nerveux central et

périphérique, pancréas, peau) où elles favorisent la régénération et la réparation tissulaires.

Notre équipe a effectué une étude approfondie des effets de la lectine humaine REG3A sur l'homéostasie du microbiote intestinal et la réponse à l'inflammation dans des modèles murins de colite. Nous avons montré que REG3A offrait une protection contre l'inflammation et le stress oxydatif à la fois pour l'épithélium intestinal et le microbiote intestinal. REG3A n'avait pas d'activité antibactérienne significative contre les bactéries Gram positif comme décrit dans la littérature mais au contraire augmentait la viabilité et la croissance de certains symbiotes bénéfiques du microbiote intestinal, en particulier des bactéries sensibles à l'oxygène appartenant à l'ordre des Clostridiales (*Ruminococcaceae*, *Lachnospiraceae*). Globalement, nous avons montré que REG3A modulait le microbiote intestinal, supprimant ainsi l'inflammation et préservant la fonction barrière de la muqueuse intestinale chez des souris atteintes de colite sévère. Nos résultats offrent un nouveau paradigme pour la régulation de l'interaction hôte-microbiote basée sur l'activité anti-oxydante d'une molécule de l'immunité innée.

Nous avons découvert que REG3A exerçait une pression de sélection via une activité anti-oxydante sur les communautés bactériennes formant le microbiote intestinal. Nos travaux ont révélé, d'une part, que cette activité biologique constitue un mécanisme nouveau de modulation du microbiote intestinal, et, que d'autre part, le stress oxydatif est un facteur d'environnement majeur dans la régulation des interactions hôte-microbiote au niveau des surfaces muqueuses. Notre travail expérimental sur les effets intestinaux de la lectine REG3A a positionné REG3A comme étant une entité biologique anti-inflammatoire protégeant du stress oxydatif l'épithélium intestinal et l'écologie microbienne. Comme il existe de grandes différences dans la tolérance à l'oxygène des bactéries intestinales, le spectre de l'activité anti-oxydante de REG3A modifie l'équilibre entre les différentes communautés bactériennes en faveur des bactéries anaérobies très sensibles à l'oxygène. Nos résultats suggèrent que certaines bactéries commensales anaérobies stricts sensibles à l'oxygène réagissent à la présence d'un antioxydant exogène tel que REG3A, ce qui leur confère un avantage sélectif de survie et de croissance sur les anaérobies aérotolestants, qui, eux, vont efficacement développer leur propre réponse

adaptative face à une toxicité oxygène ou ROS. Ce point de vue est soutenu par le fait qu'une protéine recombinante humaine REG3A améliore *in vitro* la croissance de symbionts sensibles à l'oxygène appartenant à l'ordre des Clostridiales. Un enrichissement en symbionts de cette nature, dont certains (*Faecalibacterium prausnitzii*) sont connus pour produire et libérer dans leur secrétome des petites molécules anti-inflammatoires, peut déclencher un cercle vertueux dans lequel une fonction barrière améliorée contribue à maintenir un équilibre bactérien symbiotique du fait de la réduction de la production de radicaux libres. Nos résultats démontrent que la pression exercée par REG3A sur le microbiote intestinal déclenche un tel changement vertueux dans la composition et la fonctionnalité du microbiote dans divers modèles murins conventionnels et axéniques.

## 2. Article

Darnaud M\*, **Dos Santos A\***, Gonzalez P\*, Augui S, Lacoste C, Desterke C, De Hertogh G, Valentino E, Braun E, Zheng J, Boisgard R, Neut C, Dubuquoy L, Chiappini F, Samuel D, Lepage P, Guerrieri F, Doré J, Bréchet C, Moniaux N, Faivre J.

*Enteric Delivery of Regenerating Family Member 3 alpha Alters the Intestinal Microbiota and Controls Inflammation in Mice With Colitis.*

Gastroenterology. 2018 Mar;154(4):1009-1023

# BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

## Enteric Delivery of Regenerating Family Member 3 alpha Alters the Intestinal Microbiota and Controls Inflammation in Mice With Colitis



Marion Darnaud,<sup>1,2,\*</sup> Alexandre Dos Santos,<sup>1,2,\*</sup> Patrick Gonzalez,<sup>1,2,\*</sup> Sandrine Augui,<sup>1,2</sup> Claire Lacoste,<sup>1,2</sup> Christophe Desterke,<sup>2</sup> Gert De Hertogh,<sup>3</sup> Emma Valentino,<sup>1,2</sup> Emilie Braun,<sup>1,2</sup> Jinzi Zheng,<sup>4,5</sup> Raphael Boisgard,<sup>4,5</sup> Christel Neut,<sup>6</sup> Laurent Dubuquoy,<sup>6</sup> Franck Chiappini,<sup>1,2</sup> Didier Samuel,<sup>1,2</sup> Patricia Lepage,<sup>7</sup> Francesca Guerrieri,<sup>8</sup> Joel Doré,<sup>7</sup> Christian Bréchet,<sup>1,2,9</sup> Nicolas Moniaux,<sup>1,2</sup> and Jamila Faivre<sup>1,2,10</sup>

<sup>1</sup>INSERM, U1193, Paul-Brousse University Hospital, Hepatobiliary Centre, Villejuif, France; <sup>2</sup>University Paris-Sud, Université Paris-Saclay, Faculté de Médecine Le Kremlin-Bicêtre, France; <sup>3</sup>Department of Imaging and Pathology, Unit of Translational Cell and Tissue Research, University of Leuven, Leuven, Belgium; <sup>4</sup>CEA, DSV, Institut d'Imagerie Biomédicale, Orsay, France; <sup>5</sup>INSERM, U1023, Université Paris-Sud, Orsay, France; <sup>6</sup>LIRIC-U995, University Lille, Inserm, CHU Lille, Lille, France; <sup>7</sup>Institut National de la Recherche Agronomique, UMR 1319 MICALIS, Jouy-en-Josas, France; <sup>8</sup>Center for Life NanoScience@Sapienza, Istituto Italiano di Tecnologia, Roma, Italy; <sup>9</sup>Pasteur Institute, Paris, France; and <sup>10</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Pôle de Biologie Médicale, Paul-Brousse University Hospital, Villejuif, France

**BACKGROUND & AIMS:** Paneth cell dysfunction causes deficiencies in intestinal C-type lectins and antimicrobial peptides, which leads to dysbiosis of the intestinal microbiota, alters the mucosal barrier, and promotes development of inflammatory bowel diseases. We investigated whether transgenic (TG) expression of the human regenerating family member 3 alpha gene (*REG3A*) alters the fecal microbiota and affects development of colitis in mice. **METHODS:** We performed studies with C57BL/6 mice that express human regenerating family member 3 alpha (hREG3A) in hepatocytes, via the albumin gene promoter. In these mice, hREG3A travels via the bile to the intestinal lumen. Some mice were given dextran sodium sulfate (DSS) to induce colitis. Feces were collected from mice and the composition of the microbiota was analyzed by 16S ribosomal RNA sequencing. The fecal microbiome was also analyzed from mice that express only 1 copy of human *REG3A* transgene but were fed feces from control mice (not expressing hREG3A) as newborns. Mice expressing hREG3A were monitored for DSS-induced colitis after co-housing or feeding feces from control mice. Colitis was induced in another set of control and hREG3A-TG mice by administration of trinitrobenzene sulfonic acid; some mice were given intrarectal injections of the hREG3A protein. Colon tissues were collected from mice and analyzed by histology and immunohistochemistry to detect mucin 2, as well as by 16S ribosomal RNA fluorescence in situ hybridization, transcriptional analyses, and quantitative polymerase chain reaction. We measured levels of reactive oxygen species (ROS) in bacterial cultures and fecal microbiota using 2',7'-dichlorofluorescein diacetate and flow cytometry. **RESULTS:** The fecal microbiota of mice that express hREG3A had a significant shift in composition, compared with control mice, with enrichment of Clostridiales (Ruminococcaceae, Lachnospiraceae) and depletion of Bacteroidetes (Prevotellaceae); the TG mice

developed less-severe colitis following administration of DSS than control mice, associated with preserved gut barrier integrity and reduced bacterial translocation, epithelial inflammation, and oxidative damage. A similar shift in the composition of the fecal microbiota occurred after a few months in TG mice heterozygous for *REG3A* that harbored a wild-type maternal microbiota at birth; these mice developed less-severe forms of colitis following DSS administration. Cohoused and germ-free mice fed feces from *REG3A*-TG mice and given DSS developed less-severe forms of colitis and had reduced lipopolysaccharide activation of the toll-like receptor 4 and increased survival times compared with mice not fed feces from *REG3A*-TG mice. *REG3A* TG mice developed only mild colonic inflammation after exposure to 2,4,6-trinitrobenzene sulfonic acid, compared with control mice. Control mice given intrarectal hREG3A and exposed to 2,4,6-trinitrobenzene sulfonic acid showed less colon damage and inflammation than mice not given intrarectal hREG3A. Fecal samples from *REG3A*-TG mice had lower levels of ROS than feces from control mice during DSS administration. Addition of hREG3A to bacterial cultures reduced levels of ROS and increased survival of oxygen-sensitive commensal bacteria (*Faecalibacterium prausnitzii* and *Roseburia intestinalis*). **CONCLUSIONS:** Mice with hepatocytes that express hREG3A, which travels to the intestinal lumen, are less sensitive to colitis than control mice. We found hREG3A to alter the colonic microbiota by decreasing levels of ROS. Fecal microbiota from *REG3A*-TG mice protect non-TG mice from induction of colitis. These findings indicate a role for reduction of oxidative stress in preserving the gut microbiota and its ability to prevent inflammation.

**Keywords:** Mouse Model; IBD; Bacteria; LPS.

## EDITOR'S NOTES

## BACKGROUND AND CONTEXT

Paneth's cells release a wide variety of C-type lectins and antimicrobial peptides into the intestinal lumen, but we do not know how they contribute to gut microbiota preservation and inflammation prevention.

## NEW FINDINGS

The researchers show that the lectin REG3A reduces oxidative stress for the gut epithelium and the commensal bacteria that form the gut microbiota and thus attenuates inflammation in mice with colitis.

## LIMITATIONS

This study did not determine which sites of REG3A are involved in free-radical scavenging.

## IMPACT

The enteric delivery of REG3A is a promising therapeutic method for the management of IBD.

An abnormal gut microbiota composition (dysbiosis) and defective antimicrobial mucosal barrier are important pathogenic factors for inflammatory bowel diseases (IBDs).<sup>1</sup> Dysbiosis was reported to occur before clinical symptoms in pediatric Crohn's disease, suggesting that dysbiosis contributed to disease onset in genetically susceptible individuals.<sup>2</sup> Dysbiosis associated with IBD generally consists of shifts in relative abundance of specific bacterial communities favoring potentially harmful aerobic and facultative anaerobic bacteria to the detriment of beneficial strict anaerobic ones.<sup>3</sup> There is accumulating evidence that these shifts result from the oxidative nature of the host inflammatory response.<sup>4,5</sup> The selection mechanism is that oxidized by-products of the host inflammatory response feed aerobic and facultative anaerobic bacteria, while the strict anaerobic bacteria, which lack reactive oxygen species (ROS) detoxifying enzymes, are killed. Oxidative stress in the gastrointestinal tract (GIT) is thus closely linked to dysbiosis during IBD and non-IBD gut disorders and might be an important therapeutic target.<sup>6,7</sup>

Another key contributor to dysbiosis and mucosal barrier impairment during IBD is a deficiency in the secretion of antimicrobial proteins (AMPs) such as defensins, cathelicidins, and C-type lectins resulting notably from Paneth cell dysfunction.<sup>8</sup> This enteric arsenal forms part of innate immunity, contributing to a balanced clearance of pathogens and tolerance of commensal microbes in the gut.<sup>8,9</sup> In mice, an overexpression, or an oral administration, of an enteric defensin led to significant changes in gut microbiota composition; however, the underlying mechanisms and the health benefits provided by these changes remain to be demonstrated.<sup>10,11</sup> In fact, AMPs display multiple biological activities, each of which might influence the host-microbiota interplay,<sup>12</sup> but those by which they actually regulate the gut microbiota composition and function are little known. In particular, the way in which the antimicrobial properties of AMPs interact with such a complex ecosystem as the gut microbial communities remains unclear, given that these

antimicrobial properties are counteracted by various mechanisms of bacterial resistance that maintain homeostasis.<sup>13,14</sup> Moreover, an individual antimicrobial molecule only targets a narrow bacterial spectrum, so that a broad range of them must act simultaneously to ensure selection pressure on the gut microbiota.<sup>15</sup> The structure and function of some antimicrobial molecules are sensitive to environmental factors, such as pH and redox potential,<sup>16-18</sup> entailing that a given set of molecules exerts different selection pressures in homeostatic or inflammatory states.

The human secreted C-type lectin hREG3A (human regenerating family member 3 alpha), or hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein, is a 16-kDa carbohydrate-binding protein exhibiting remarkable anti-inflammatory activities in many eukaryote cell types, where hREG3A promotes tissue regeneration and repair.<sup>19-25</sup>

It is expressed in Paneth's cells and the neuroendocrine cells of the small intestine physiologically, and in the large intestine in response to infection and inflammation. The mechanisms of reduction of inflammation by hREG3A are not fully understood. hREG3A is a potent ROS scavenger and its tissue-protective and regenerative effects are largely due to this antioxidant property.<sup>19,20,25</sup> hREG3A and its murine homologue Reg3g also display antibacterial activities against gram-positive bacteria by interacting with peptidoglycan carbohydrate.<sup>26</sup> Reg3b, another murine homologue of hREG3A, exerts bactericidal activity against gram-negative bacteria by binding to lipopolysaccharide.<sup>27-29</sup> Strain differences in susceptibility of gram-positive bacteria to hREG3A have been reported,<sup>30,31</sup> suggesting that the antibacterial spectrum of the REG3 (regenerating islet-derived protein 3) proteins might be relatively narrow.

In this study, we investigated the effects of an enteric delivery of the human REG3A protein on gut microbiota homeostasis and the response to inflammation in mice. We showed that hREG3A changed the gut microbiota composition (enrichment of Clostridiales, depletion of Bacteroidetes and Proteobacteria), decreased the oxidative stress and inflammatory response of the gut epithelium, and reduced host susceptibility to colitis. hREG3A lowered ROS levels in the gut microbiota of mice with colitis and increased the viability of some gram-positive oxygen-sensitive *Clostridia* in vitro, indicating that the antioxidant activity of hREG3A contributed to shaping the gut microbiota. This study offers

\*Authors share co-first authorship.

**Abbreviations used in this paper:** AMPs, antimicrobial peptides/proteins; DSS, dextran sodium sulfate; ExGF, formerly germ-free; [<sup>18</sup>F]FDG, 2-[<sup>18</sup>F] fluoro-2-deoxy-D-glucose; H<sub>2</sub>-DCFDA, 2',7'-dichlorofluorescein diacetate; hREG3A, human regenerating family member 3 alpha; IBD, inflammatory bowel disease; GIT, gastrointestinal tract; LPS, lipopolysaccharide endotoxin; PCA, principal component analysis; rcREG3A, recombinant human REG3A protein; REG3, regenerating islet-derived protein 3; ROS, reactive oxygen species; rRNA, ribosomal RNA; TG, transgenic; TNBS, 2,4,6-trinitrobenzene sulfonic acid; WT, wild-type.

 Most current article

© 2018 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0016-5085

<https://doi.org/10.1053/j.gastro.2017.11.003>

a new paradigm for regulation of the host-microbiota interplay based on the ROS scavenging activity of an innate immune molecule.

## Materials and Methods

### Animal Studies

Animal studies were performed in compliance with the institutional and European Union guidelines for laboratory animal care and approved by the Ethics Committee of CE2A-03-CNRS-Orléans (Accreditation N°01417.01). Homozygous transgenic (TG) C57BL/6 mice expressing *REG3A* in hepatocytes under the control of the mouse albumin gene promoter were derived from a strain previously produced in our laboratory.<sup>32</sup> The cohousing experiments were performed with 3-week-old weaned wild-type (WT) mice randomly cohoused with age-matched *REG3A*-TG mice at a 1:1 ratio for 8 weeks. Ten-week-old germ-free mice underwent 2 gastric gavages at 48-hour intervals using inocula consisting of pooled fresh feces from 8 *REG3A*-TG or WT mice.

### Dextran Sodium Sulfate Colitis

Colitis was induced in mice with a 3% 40 kDa dextran sodium sulfate (DSS) solution (TdB Consultancy AB, Uppsala, Sweden), which was added to their drinking water for 5 days followed by ad libitum access to tap water for 7 days.

### 2,4,6-Trinitrobenzene Sulfonic Acid Colitis

Overnight-fasted mice were anesthetized for 90 minutes and received intrarectally 150 mg/kg of 2,4,6-trinitrobenzene sulfonic acid (TNBS; Sigma-Aldrich, St Louis, MO) together with ethanol. WT mice received an intrarectal injection of 100  $\mu$ g of a recombinant hREG3A protein on the day before and on the day of TNBS administration.

### 16S Ribosomal RNA Analysis

DNA pyrosequencing of the V3-V4 hypervariable region of 16S ribosomal RNA (rRNA) gene was performed using the Illumina MiSeq instrument (San Diego, CA). The sequences thus generated were pre-processed using the mothur pipeline and then clustered into operational taxonomic units based on 97% of similarity. RDP release 11 was used as reference database. Principal component analysis (PCA) was performed based on bacterial family and genera compositions using the R packages ade4 and FactoMiner.

### Bacterial Strains, Culture Conditions, and ROS Quantification

The bacterial strains used were *Enterococcus faecalis* (ATCC-19433), *Faecalibacterium prausnitzii* A2-165 (DSM17677), and *Roseburia intestinalis* L1-82 (DSM14610T). Oxidative stress was induced by addition of 200 mM paraquat (1,1'-dimethyl-4,4'-bipyridinium dichloride; Sigma-Aldrich), which is commonly used experimentally to generate superoxide radical, hydrogen peroxide, and hydroxyl radical in cells.<sup>33</sup> Intracellular ROS levels and cell viability were determined by flow cytometry with 2',7'-dichlorofluorescein diacetate (H<sub>2</sub>-DCFDA) and propidium iodide in bacteria cultures. Bacterial growth was assessed by colony-forming unit counting on

agar plates. Bacterial ROS levels in fresh fecal microbiota were measured by flow cytometry with H<sub>2</sub>-DCFDA.

### Other Methods

For histology, immunohistochemistry, recombinant protein production, enzyme-linked immunosorbent assays, immunoblotting, dot-blot, fluorescence in situ hybridization, transcriptomics, and 2-[<sup>18</sup>F]fluoro-2-deoxy-D-glucose ([<sup>18</sup>F]FDG) positron emission tomography scan, please see details in the [Supplementary Materials and Methods](#).

### Statistics

Results are presented as means  $\pm$  SEM or box plots. Differences between sample groups were tested using nonparametric Wilcoxon rank sum (2-sided) for all the experiments reported, except for the following: 2-way analysis of variance tests were used for the in vivo quantifications of [<sup>18</sup>F]FDG and H<sub>2</sub>-DCFDA over time; the robustness of PCA clustering was assessed with a Monte-Carlo rank test; and the Kaplan-Meier survival analysis with a log-rank test was applied to compare the time course of survival between the groups. *P* values less than 0.05 were deemed statistically significant.

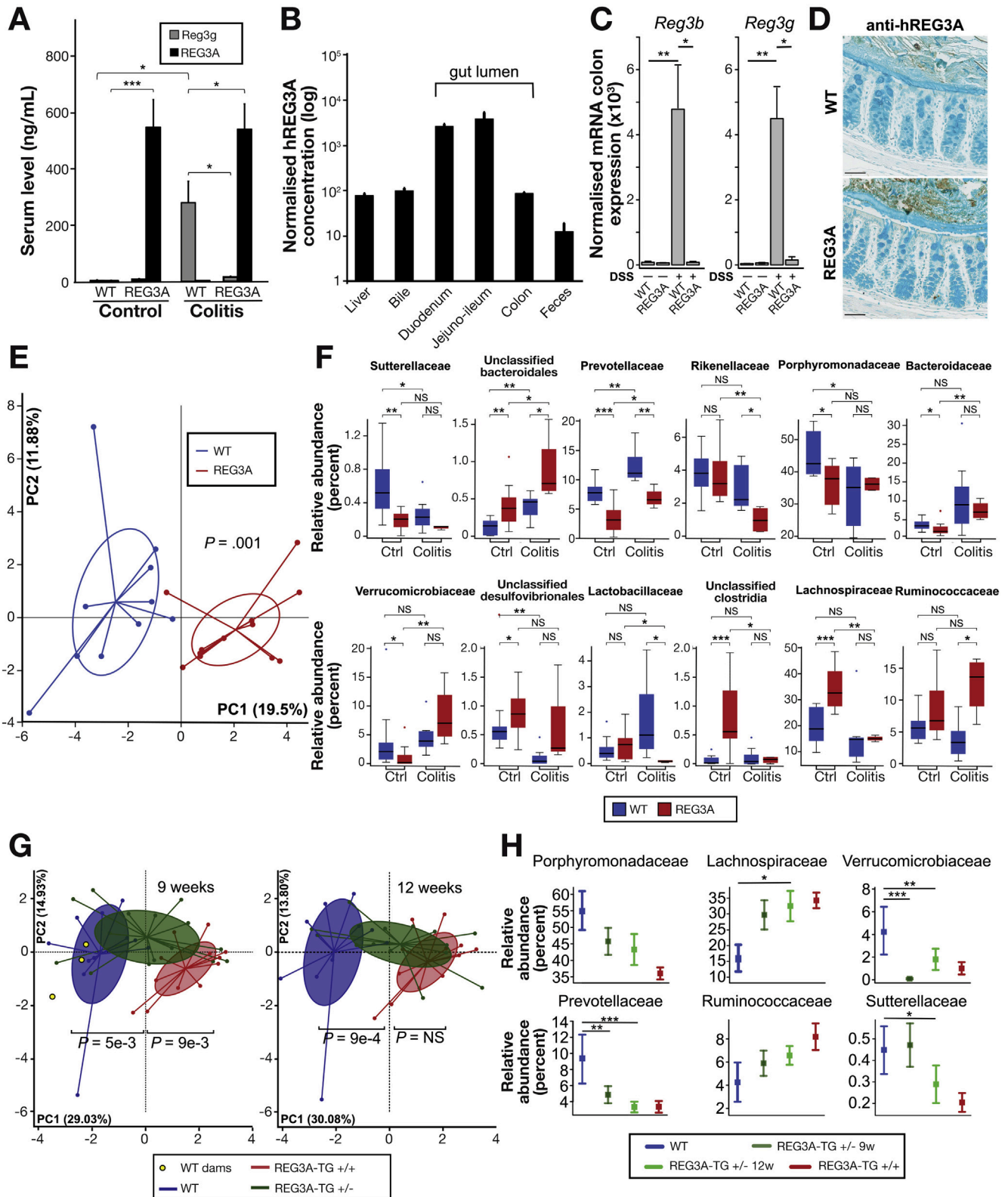
## Results

### Human REG3A Delivered into the Gut Lumen Shifts the Gut Microbiota Composition in Hepatocyte-targeted REG3A Transgenic Mice

We studied the effects of human REG3A on gut microbiota composition and gut barrier integrity in mice TG for *REG3A* under homeostatic and inflammatory conditions. Inflammation was induced by oral absorption of DSS. To prevent degradation of the hREG3A protein by acid and luminal proteases in the upper GIT, we used previously generated homozygous TG C57BL/6 mice expressing *REG3A* in hepatocytes under the control of the mouse albumin gene promoter<sup>32</sup> and showed that, in these mice, hREG3A flowed into bile ducts and the GIT lumen ([Figure 1A, B, and D](#)). Transgenic hepatocytes secreted hREG3A into blood vessels through the basolateral membranes, and into bile canaliculi through the apical membranes. The expression levels of endogenous Reg3b and Reg3g in the blood and colon tissues were negligible in the basal state of WT mice and high after DSS administration ([Figure 1A and C](#)). In *REG3A*-TG mice, a similar high serum level of hREG3A was found in the basal and inflammatory states ([Figure 1A](#)). This level was similar to that of the endogenous Reg3 in WT mice with colitis. The colon expression levels of the endogenous Reg3 mRNAs were negligible in both the basal and inflammatory states of *REG3A*-TG mice ([Figure 1C](#)). In other words, *REG3A*-TG mice did not display any up-regulation of Reg3b or Reg3g during colitis (perhaps because they were retro-controlled by the overexpressed *REG3A*), implying that Reg3b and Reg3g played a negligible part in *REG3A*-TG mice. We analyzed the fecal microbiota composition in healthy and DSS-given *REG3A*-TG mice using MiSeq 16S rDNA gene sequencing. Control WT and *REG3A*-TG mice were bred and processed under the same conditions. PCA revealed 2 clearly separate clusters for *REG3A*-TG and WT mice at the family level both

before ( $P = .001$ ) and after ( $P = .039$ ) exposure to DSS (Figure 1E and Supplementary Figure 1A). Twelve bacterial families displayed significant differences in their relative abundance between REG3A-TG and WT mice in either the

basal or inflammatory state, or both (Figure 1F). Sutterellaceae (phylum Proteobacteria), Prevotellaceae, Porphyromonadaceae, and Bacteroidaceae (phylum Bacteroidetes) were underrepresented, whereas Lachnospiraceae and



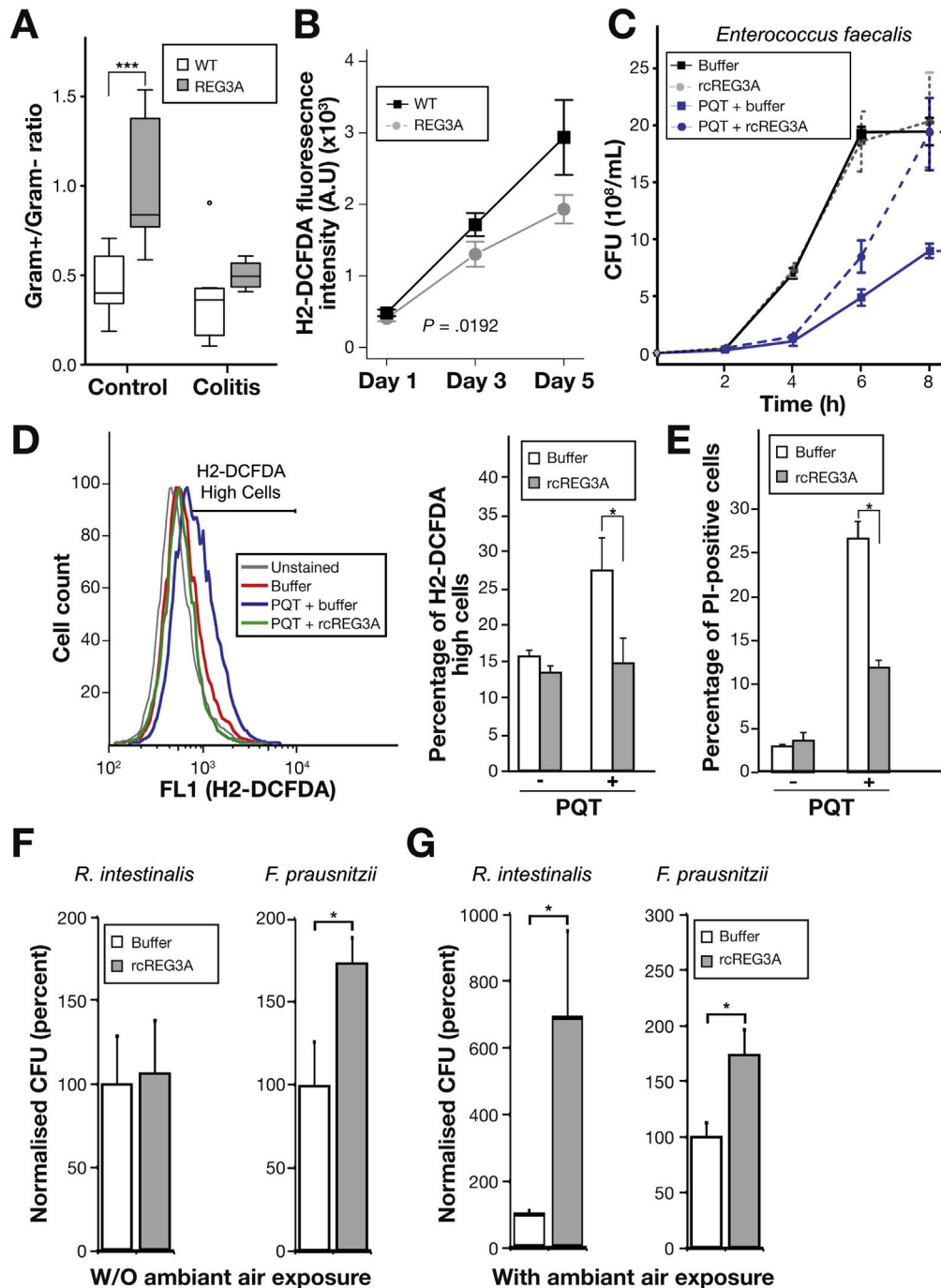


unclassified *Clostridia* (phylum Firmicutes) were over-represented in the basal state of *REG3A*-TG mice compared with that of WT mice (Figure 1F). The same shifts in microbial composition were found at the genus level (Supplementary Figure 1B). In the inflammatory, compared with basal, state, WT mice displayed an enrichment of Prevotellaceae and unclassified Bacteroidales, and a depletion of Porphyromonadaceae, Sutterellaceae, and unclassified Desulfovibrionales, whereas *REG3A*-TG mice displayed an increase in Prevotellaceae, unclassified Bacteroidales, Bacteroidaceae, and Verrucomicrobiaceae and a decrease in Lactobacillaceae, unclassified *Clostridia* and Lachnospiraceae. DSS-given *REG3A*-TG mice exhibited a lower level of Prevotellaceae and a much higher level of Ruminococcaceae (order Clostridiales) (Figure 1F and Supplementary Figure 1C) than DSS-given WT mice. We then investigated whether the gut microbiota composition of homozygous TG (*REG3A*-TG+/+) mice at baseline was due to the *REG3A* transgene or an extra-genetic effect.<sup>34</sup> We studied the time evolution of the microbiota in mice that expressed only 1 copy of human *REG3A* transgene (*REG3A*-TG+/-), were born to WT mothers and initially had the maternal microbiota. The fecal microbiomes of 15 *REG3A*-TG+/- pups from 3 different litters and cages were sequenced at the ages of 9 and 12 weeks and compared with those of their WT dams (n = 3) and unrelated WT mice of similar ages (n = 11). The fecal microbiomes of 12 *REG3A*-TG+/+ mice were analyzed for comparison. PCA plots at the family level showed that the microbiota from *REG3A*-TG+/- mice were grouped in-between those from (maternal and unrelated) WT and *REG3A*-TG+/+ mice at 9 weeks ( $P < .01$  for *REG3A*-TG+/- vs WT and *REG3A*-TG+/- vs *REG3A*-TG+/+) and moved close to those from *REG3A*-TG+/+ mice during the next 3 weeks ( $P < .001$  for *REG3A*-TG+/- vs WT;  $P > .05$  for *REG3A*-TG+/- vs *REG3A*-TG+/+) (Figure 1G). The microbiota composition shift of the *REG3A*-TG+/- mice compared with the microbiota of WT mice mostly consisted of a decrease in Prevotellaceae, Sutterellaceae, and Verrucomicrobiaceae, and an increase in Lachnospiraceae and Ruminococcaceae, to reach levels comparable to those of *REG3A*-TG+/+ mice after 12 weeks (Figure 1H). Thus, the WT microbiota of the *REG3A* heterozygous newborns evolved toward a TG-positive microbiota in a few months, which means that the *REG3A* transgene effect superseded the maternal legacy.

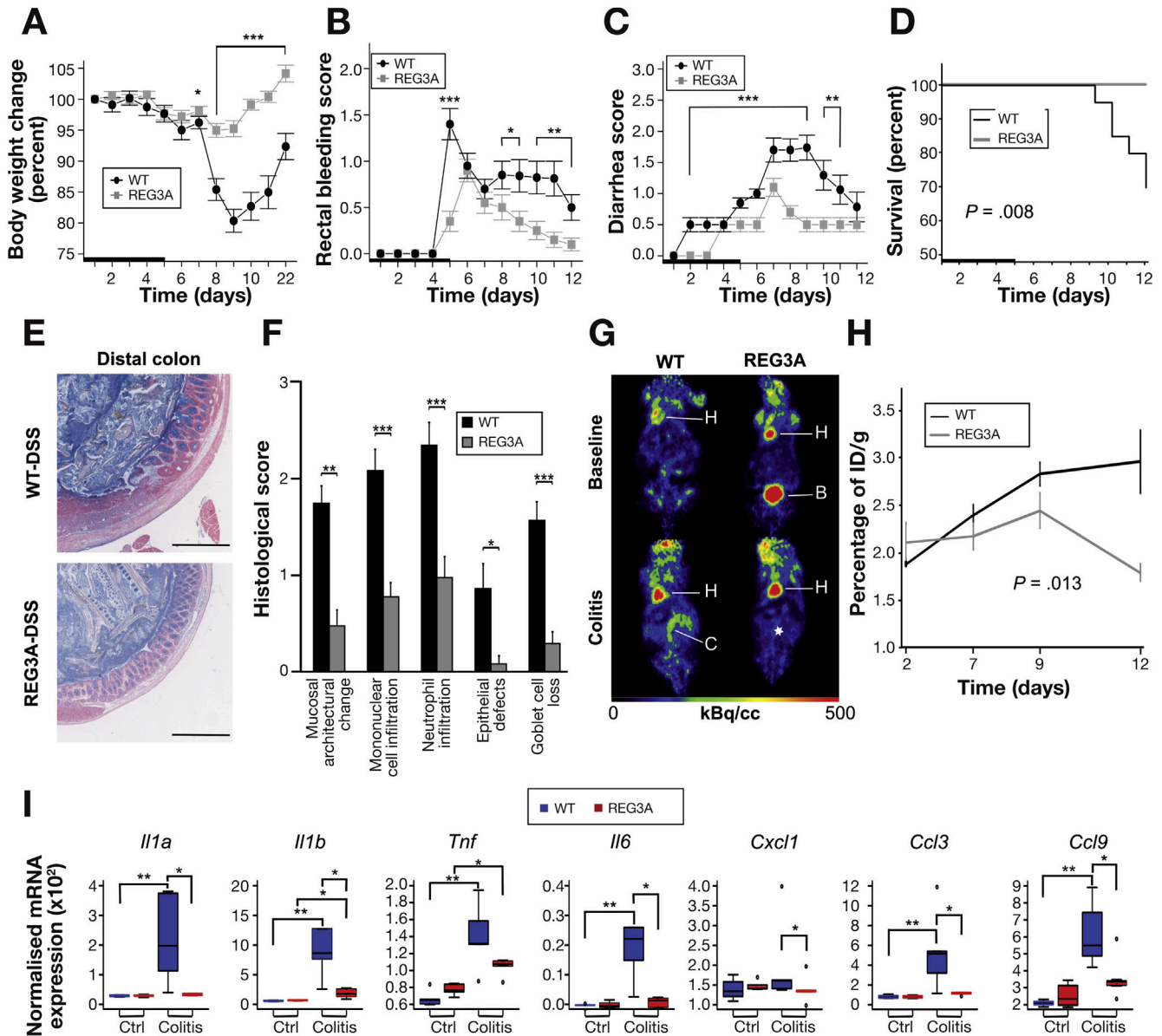
### *hREG3A* Increases the Viability of Some Highly Oxygen-Sensitive *Clostridia*

The shift in microbiota composition in *REG3A*-TG mice compared with WT mice corresponded to a large increase in the ratio between gram-positive and gram-negative bacteria (Figure 2A), difficult to reconcile with the reported selective anti-gram-positive bactericidal activity of hREG3A.<sup>26</sup> We conjectured that the antioxidant activity of hREG3A was a key factor in shifting the intestinal microbial ecology in *REG3A*-TG mice. The mechanism at play would be a selection pressure exerted by hREG3A in favor of strict anaerobic gram-positive bacteria, as are some *Clostridia*. To substantiate this view, we measured ROS levels in fresh fecal microbiota of WT and *REG3A*-TG mice using H<sub>2</sub>-DCFDA and flow cytometry over the 5-day period of DSS administration, during which the gut barrier was not yet damaged nor colitis established. The ROS levels at day 1 of DSS administration were similar to the baseline level and then increased over time in both mouse lines. However, the rate of increase was significantly lower for the microbiota of *REG3A*-TG mice ( $P = .02$  over the day 1–5 period), evidencing a ROS scavenging activity of the *REG3A* transgene in a microbiota subjected to an oxidative stress (Figure 2B). We also tested the antioxidant efficiency of a full-length recombinant human *REG3A* protein (rcREG3A) in prokaryote cells in vitro. The recombinant protein we used is chemically and biologically active in terms of anti-inflammatory properties in eukaryotic cells and carbohydrate-binding selectivity<sup>10,20,25</sup> (Supplementary Figure 2). We cultured gram-positive *Enterococcus faecalis* stressed with a ROS generator (paraquat) during the exponential phase of growth. Exposure to 200 mM paraquat had a strong bactericidal effect on *E faecalis*. The addition of 10 μM rcREG3A restored the exponential growth of *E faecalis*, suggesting that the ROS generated by paraquat were effectively reduced by rcREG3A (Figure 2C). This was demonstrated by flow cytometry using the ROS-specific fluorescent probe H<sub>2</sub>-DCFDA and propidium iodide DNA staining (Figure 2D and E and Supplementary Figure 3). We found some bacterial aggregation, but no bactericidal effect of rcREG3A on *E faecalis* (Figure 2C), at variance with previous reports.<sup>26</sup> This discrepancy may rely on many factors pertaining to the protein<sup>26</sup> or the bacterial strain,<sup>30,31</sup> but note that an absence of bactericidal activity of hREG3A would be consistent with the enrichment of gram-positive bacteria we

**Figure 1.** hREG3A shapes gut microbiota in *REG3A*-TG mice. (A) Serum concentrations of human *REG3A* and murine *Reg3g* in WT and *REG3A*-TG mice (*REG3A*) before (Control) and after (Colitis) administration of DSS (n = 6). (B) Digestive tract concentrations of hREG3A (n = 3). (C) Colonic *Reg3b* and *Reg3g* mRNA expression level (n = 5). (D) Alcian blue and anti-hREG3A stainings of colon tissue sections. Scale bar, 100 μm. (E) PCA of bacterial profiles at the family level in the fecal microbiota of *REG3A*-TG (n = 12) and WT (n = 10) mice in the basal state. Each dot represents 1 mouse.  $P = .001$  (Monte-Carlo rank test). (F) Relative abundances of bacterial families in the basal state (Ctrl) of the same mouse cohorts as in (E) and on day 12 after onset of exposure to DSS (Colitis) in *REG3A*-TG (n = 5) and WT (n = 7) mice. (G) PCA of bacterial profiles at the family level in heterozygous (*REG3A*-TG+/-; n = 15), homozygous (*REG3A*-TG+/+; n = 12), and unrelated WT (n = 11) mice of similar ages (Monte-Carlo rank test). Each dot represents 1 mouse. Yellow dots: WT dams of the *REG3A*-TG+/- mice (n = 3). (H) Relative abundances of bacterial families in *REG3A*-TG+/- mice aged 9 and 12 weeks compared with WT (n = 14) and to *REG3A*-TG+/+ mice. The data are averages ± SEM. Except for (E) and (G), the 2-sided Wilcoxon rank sum test was performed for analysis. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .



**Figure 2.** hREG3A rescues bacteria from oxidative stress. (A) Ratio of gram-positive to gram-negative bacteria in the fecal microbiota of REG3A-TG (REG3A) and WT mice in the basal state (control) and on day 12 of induced colitis. Same cohorts as in Figure 1E and F. (B) Bacterial ROS levels in the fecal microbiota of REG3A and WT mice ( $n = 6$  per group) measured by flow cytometry and H<sub>2</sub>-DCFDA during the period of DSS administration.  $P = .019$  (2-way analysis of variance). (C) Growth of gram-positive *Enterococcus faecalis* under oxidative stress. CFU, colony-forming units; PQT, paraquat. Assays were done in quadruplicate in 3 independent experiments. (D) Cellular ROS levels in *E. faecalis* cultures measured by flow cytometry with H<sub>2</sub>-DCFDA. Left: Representative flow cytometry curves. H<sub>2</sub>-DCFDA high cells: the 15% most strongly fluorescent cells in the PQT + buffer condition. Right: Percentage of bacteria in the high-intensity range in 3 independent experiments. (E) Cell viability by flow cytometry with propidium iodide (PI). Normalized CFU numbers of gram-positive *Roseburia intestinalis* and *Faecalibacterium prausnitzii* in strictly anaerobic cultures (F) and after exposure to ambient air (G). The numbers of CFUs were normalized to the average CFU number measured in control cultures (buffer). Assays were done in quadruplicate in 2 independent experiments. The data are averages  $\pm$  SEM. The 2-sided Wilcoxon rank sum test was performed for analysis except for (B). \* $P < .05$ , \*\*\* $P < .001$ .



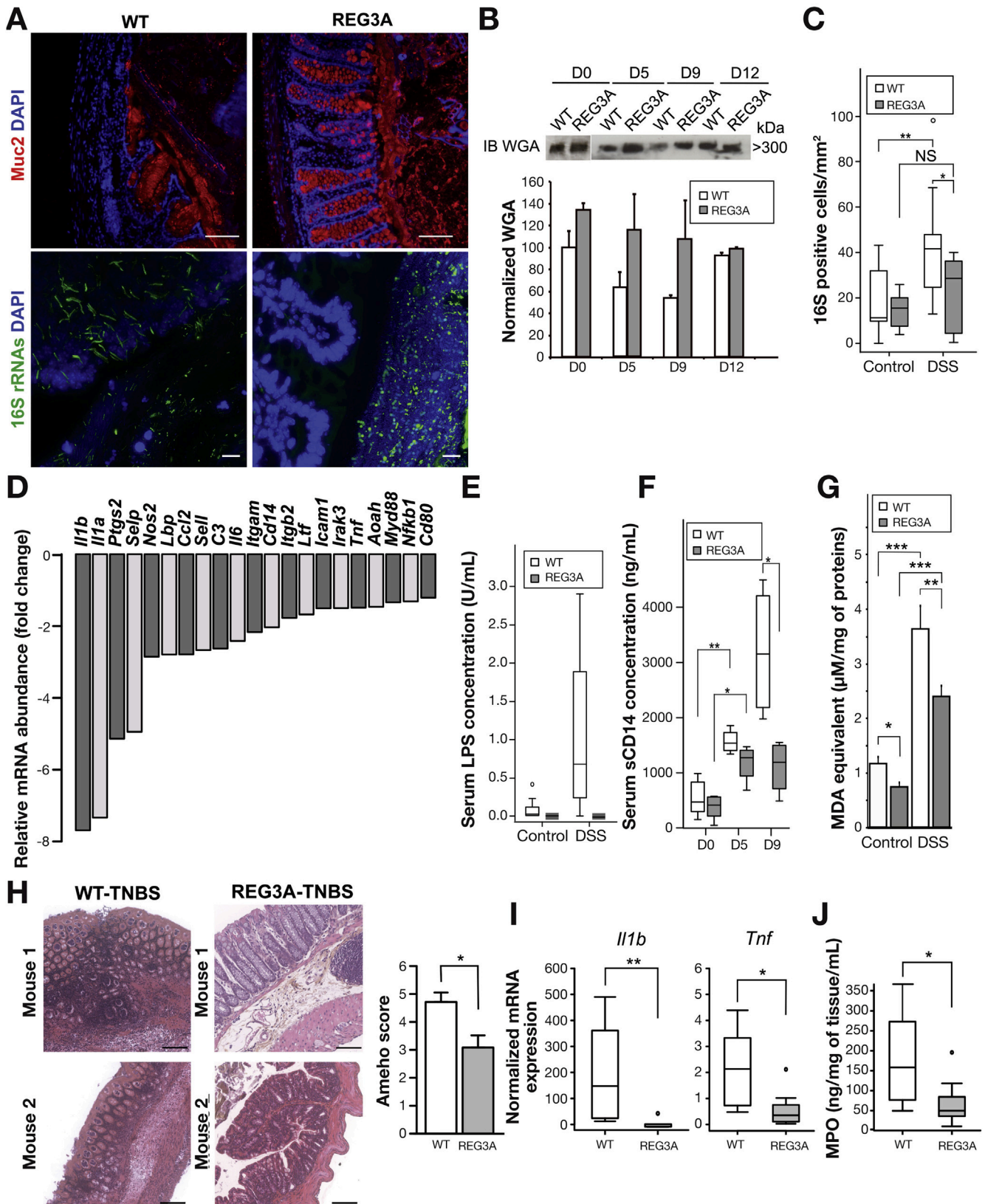
**Figure 3.** *REG3A*-TG mice have a lower susceptibility to DSS-induced colitis than WT mice. (A–D) Time evolution of DSS-induced colitis in *REG3A*-TG (*REG3A*) and WT mice (n = 20 for each group; 2 independent experiments). Heavy line: 5-day period of DSS administration. (A) Body weight change. (B) Rectal bleeding score. (C) Diarrhea score. (D) Kaplan-Meier survival plot. (E) Representative day-12 colon images stained with hematoxylin-eosin and Alcian blue. Scale bar, 50  $\mu$ m. (F) Histological assessment of gut epithelium (n = 14 for each group). (G) Representative whole-body [<sup>18</sup>F]FDG positron emission tomography images. The images are frontal slices of 3-dimensional volumes passing through heart, B, bladder; C, colon; H, heart. \*Absence of colon uptake. (H) Quantification of [<sup>18</sup>F]FDG over time (n = 3).  $P = .013$  (2-way analysis of variance). (I) mRNA expression level of the indicated inflammatory markers in colonic epithelial cells of WT and *REG3A*-TG mice (n = 5). The data are means  $\pm$  SEM. The 2-sided Wilcoxon rank sum test was performed for analysis except for (H). \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

observed in the gut microbiota of *REG3A*-TG mice compared with that of WT mice. We cultured of 2 well-documented extremely oxygen-sensitive gram-positive bacteria species, *Roseburia intestinalis* (Lachnospiraceae, Clostridium Cluster XIVa) and *Faecalibacterium prausnitzii* (Ruminococcaceae, Clostridium Cluster IV), which are major butyrate producers dramatically reduced during IBD.<sup>35–38</sup> *R. intestinalis* is known to survive for less than 2 minutes when exposed to air on the surface of agar plates.<sup>39</sup> *F. prausnitzii* is more sensitive to air exposure than *R. intestinalis* and grows

slowly even under anaerobic conditions.<sup>40</sup> We performed anaerobic cultures of *R. intestinalis* and *F. prausnitzii* followed, or not, by a 5-minute exposure to ambient air. We found that rcREG3A had a significant growth-promoting effect on *F. prausnitzii* under strict anaerobic conditions and a survival effect on both bacteria after exposure to oxygen (Figure 2F and G). The centrifugation of *F. prausnitzii* anaerobic cultures incubated with rcREG3A for 24 hours, followed by anti-hREG3A immunoblotting, showed that a 15-kDa rcREG3A co-sedimented with bacterial aggregates

(Supplementary Figure 4A). Slide-mounted imaging showed that *F. prausnitzii* incubated with rcREG3A survived, whereas control cultures were completely lysed, after 2

hours of exposure to ambient air (Supplementary Figure 4B). These findings establish that rcREG3A exerts a potent antioxidant activity on prokaryotic cells and is



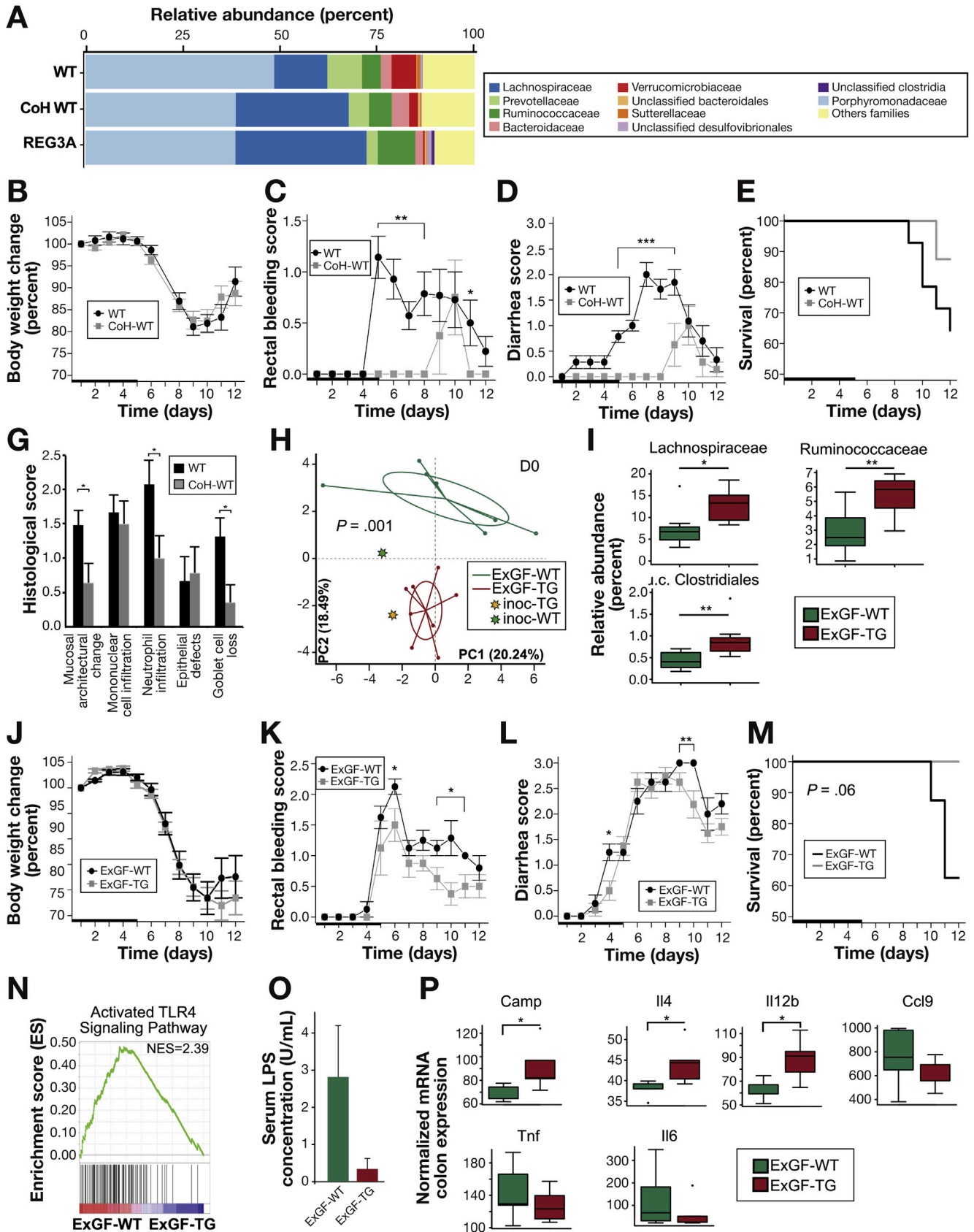
capable of increasing the viability and growth of some extremely oxygen-sensitive commensal Clostridia.

### REG3A-TG Mice Are Less Susceptible to DSS- and TNBS-induced Colitis Than WT Mice

The dysbiotic microbiota of patients with IBD is mostly characterized by an increase in Sutterellaceae, Prevotellaceae, and Enterobacteriaceae (phylum Proteobacteria), and a decrease in Ruminococcaceae and Lachnospiraceae (which include the principal butyrate-producing symbionts).<sup>3,35–38</sup> Similar shifts in the composition of the gut microbiota also are associated with colitis in genetically susceptible mouse models.<sup>41,42</sup> The fact that more or less reverse shifts occur in the gut microbiota of healthy REG3A-TG mice compared with WT mice suggests that a microbiota shaped by hREG3A might have a beneficial impact on health. We therefore studied the response to induced colitis in REG3A-TG mice, and also performed feces transfer experiments. REG3A-TG mice received oral administrations of 3% DSS for 5 days followed by normal water for 7 days. Control groups of WT mice were bred and processed under the same conditions. Body weight changes, stool consistency, and bleeding scores and survival rate were monitored over time. Histopathological features were scored on day 12. Strikingly, DSS-given REG3A-TG mice displayed only a very few signs of colitis. They had reduced weight loss, diarrhea, and rectal bleeding compared with WT mice and had a 100% survival rate on day 12 compared with 70% in WT mice (Figure 3A–D). Their colons displayed fewer barrier defects and less Goblet cell loss and inflammatory cell infiltration than those of WT mice (Figure 3E and F and Supplementary Figure 5). Twelve-week-old REG3A-TG+/- mice born to WT mothers were also exposed to DSS. They exhibited a relatively benign colitis (Supplementary Figure 6A–D), consistent with the fact that their gut microbiota was similar in composition (Figure 1G), and thus functionality, to those of REG3A homozygous TG mice. We ensured that the hREG3A lectin did not interact with dextran in vitro (Supplementary Figure 2),<sup>43</sup> and that, therefore, the observed phenotype could not be attributed to a direct blocking of DSS toxicity by hREG3A. Positron emission tomography imaging with [<sup>18</sup>F]FDG, a tracer for abnormally high glucose metabolism in inflammatory areas, was used to quantify intestinal inflammation. Colonic [<sup>18</sup>F]FDG uptake increased over time in WT mice with colitis, whereas it remained weak (ie, comparable to that

seen in healthy) in REG3A-TG mice ( $P = .013$  over the day 7–12 period) (Figure 3G and H). The severity of colitis also was assessed by measuring the colonic expression levels of inflammatory markers (*Il1a*, *Il1b*, *Tnf*, *Il6*, *Cxcl1*, *Ccl3*, *Ccl9*). Most of them showed a significant increase on DSS administration in WT mice, and not in REG3A-TG mice (Figure 3I). We studied the integrity of the gut mucosal barrier in healthy and DSS-given REG3A-TG mice using immunostaining for Muc2 mucin and 16S rRNA fluorescence in situ hybridization, transcriptome profiling, and gene set enrichment analysis with the KEGG and Reactome databases. In the basal state, mucosal barriers of REG3A-TG and WT mice were microscopically similar. However, several pathways related to the intestinal barrier, including modulators of cell-matrix interactions and mucin *O*-glycosylation, were up-regulated in REG3A-TG mice (Supplementary Figure 7A). During intestinal inflammation, the colon mucosal barrier remained intact over considerable areas in REG3A-TG mice, whereas it was largely disrupted in WT mice, leading to bacterial colonization across the epithelium (Figure 4A). The functional robustness of the mucosal barrier in REG3A-TG, compared with WT, mice was further highlighted as follows. An up-regulation of some tight-junction genes involved in the regulation of cell-cell interactions was observed (Supplementary Figure 7B). An increase in *O*-glycosylation of epithelial mucins was revealed by wheat germ agglutinin immunoblotting (Figure 4B). The level of bacterial translocation in mesenteric lymph nodes remained low, as shown by 16S rRNA fluorescence in situ hybridization analysis (Figure 4C). A number of genes related to the lipopolysaccharide endotoxin (LPS)-induced activation pathway were down-regulated in colon epithelial cells (Figure 4D). Unsupervised transcriptome analysis revealed a clustering of DSS-given REG3A-TG mice with healthy mice separately from DSS-given WT mice (Supplementary Figure 8). Consistent with this, the increase in inflammation-associated serum markers (LPS, soluble CD14) remained very small in these mice during intestinal inflammation (Figure 4E and F). Finally, the concentration of malondialdehyde, a biomarker of oxidative stress, in the colon tissue was substantially reduced in REG3A-TG, compared with WT, mice with colitis (Figure 4G). We then studied the susceptibility of REG3A-TG mice to an alternative type of colitis using a single intrarectal administration of trinitrobenzenesulfonic acid (TNBS) together with ethanol. In contrast with oral administration of DSS, which destroys colon epithelial cells, alters barrier function, and subsequently

**Figure 4.** Gut barrier integrity and weak inflammation in REG3A-TG mice exposed to DSS or TNBS. (A–G) Mice exposed to DSS. (A) Anti-MUC2 mucin immunofluorescence, bacterial 16S rRNA fluorescence in situ hybridization and 4',6-diamidino-2-phenylindole (DAPI) images of colon tissues from WT and REG3A-TG mice on day 12 of DSS-induced colitis. Scale bars, 20  $\mu$ m. (B) Representative anti-wheat germ agglutinin immunoblots (IB WGA) at the indicated time points of DSS-induced colitis. Bar chart: densitometric analysis ( $n = 5$  for each group; 2 independent experiments). (C) Bacterial translocation in mesenteric lymph nodes before ( $n = 4$ ) and during colitis ( $n = 5$ ). (D) Changes in expression level of the genes related to the LPS-induced activation pathway that were the most down-regulated in REG3A-TG mice. (E) Serum levels of LPS ( $n = 8$ ). (F) Serum levels of soluble CD14 ( $n = 8$ ). (G) Concentration of malondialdehyde (MDA) in gut epithelium extracts ( $n = 9$ –14). (H–J) Mice exposed to TNBS. REG3A-TG and WT mice were administered TNBS together with ethanol intrarectally ( $n = 15$  per group). (H) Ameho quantitative histological score and 2 representative histological examinations for each group 2 days after TNBS administration. Scale bar, 100  $\mu$ m. (I) mRNA expression level of the *Il1b* and *Tnf* inflammatory cytokines. (J) Myeloperoxidase (MPO) concentration in colon tissue extracts. The data are represented either as box plots or means  $\pm$  SEM. The 2-sided Wilcoxon rank sum test was performed for analysis. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .



causes inflammation, TNBS rapidly triggers a severe colonic inflammation through a T-cell immune response against haptenized proteins and luminal antigens.<sup>44</sup> We found that, 2 days after TNBS administration, *REG3A*-TG mice developed only mild colonic inflammation and mucosal damage (Figure 4H) associated with decreased levels of inflammatory cytokines (*Il1b*, *Tnf*) and myeloperoxidase (Figure 4I and J).

### The Anti-inflammatory Properties of Gut Microbiota Shaped by hREG3A Promote Survival to DSS-induced Colitis in Colonized WT Mice

To assess the protective action of hREG3A-shaped microbiota against colitis, we transferred fecal microbiota from *REG3A*-TG mice to WT mice by means of cohousing. Three-week-old weaned WT mice were cohoused with age-matched *REG3A*-TG mice for 8 weeks and then exposed to DSS. Control groups consisted of WT mice housed alone. At the end of the cohousing period, cohoused-WT mice displayed a significant shift in gut microbiota composition toward an *REG3A*-TG profile at the bacteria family level (Figure 5A) and a clear alleviation of DSS-induced colitis compared with single-housed WT mice (Figure 5B–G). Next, we colonized germ-free C57BL/6 mice with a fecal microbiota from *REG3A*-TG or WT mice for 3 weeks. Formerly (Ex) germ-free (GF), but now colonized, mice will be designated as ExGF-TG or ExGF-WT according to the origin of the colonizing microbiota. The fecal microbiota was then analyzed by sequencing (day 0) and colonized mice were exposed to DSS. The microbiota of ExGF-TG and ExGF-WT mice at day 0 had somewhat drifted from their respective inocula (Figure 5H). Nevertheless, the microbiota of ExGF-TG mice still harbored the same predominant bacterial communities as the inoculum, albeit with different relative abundance values, and remained far from that of ExGF-WT mice (Figure 5H and I and Supplementary Figure 9A–E). On DSS administration, ExGF-TG mice exhibited a complete colitis survival despite transient signs of colitis, whereas 37% of the ExGF-WT mice died (Figure 5J–M and Supplementary Figure 10). This was associated with a reduced inflammatory response in terms of TLR4 signaling activation, colonic inflammatory markers, and LPS-induced endotoxemia (Figure 5N–P). These results demonstrate a transmissible pro-survival action of the microbiota shaped by hREG3A probably due to a reduced inflammatory

response in the gut epithelium and less systemic dissemination of gram-negative LPS. This is consistent with the depletion of potentially aggressive gram-negative bacteria we observed in the gut microbiota shaped by hREG3A (Figure 2A).

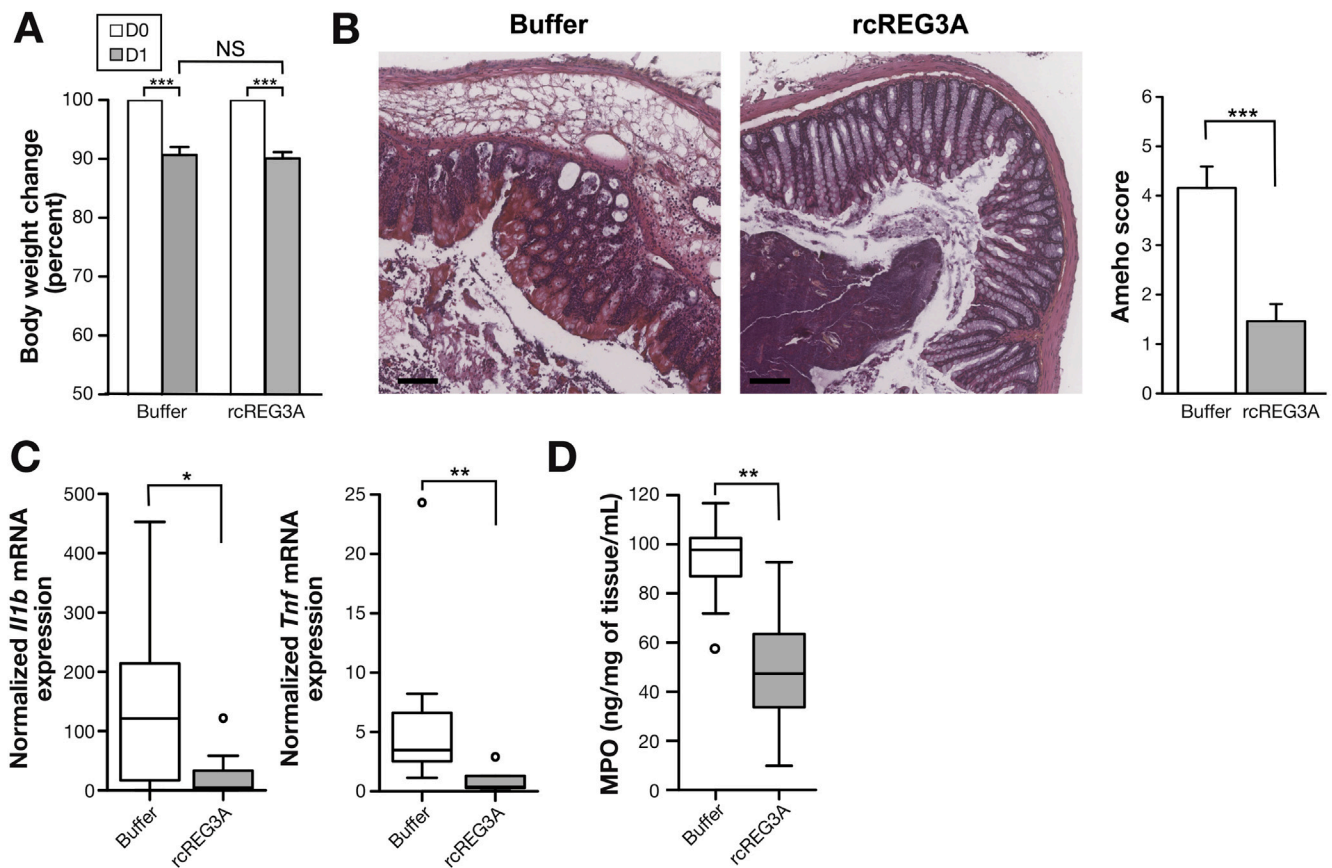
### Rectal Use of a Recombinant REG3A Protein Diminishes Colonic Inflammation in Mice With TNBS-induced Colitis

The fact that signs of colitis were partly reduced in cohoused-WT and ExGF-TG mice while they were fully suppressed in *REG3A*-TG mice suggests that, in the latter, a direct interaction between overexpressed hREG3A and the host contributed to the maintenance of gut barrier homeostasis, in addition to the anti-inflammatory effects of the gut microbiota shaped by hREG3A. To substantiate this view, we evaluated the effects of intrarectal administrations of rcREG3A in WT mice with colitis. An amount of 100  $\mu$ g of rcREG3A (or an equivalent volume of buffer) was delivered on the day before and on the day of TNBS administration. The body weight loss was the same in the 2 groups of mice (Figure 6A). Histological stainings revealed that mice given intrarectal rcREG3A showed milder colonic barrier defects and inflammation than mice not given intrarectal rcREG3A (Figure 6B and Supplementary Figure 11). A significant decrease of the inflammatory markers *Il1b*, *Tnf*, and myeloperoxidase was observed in the colon tissues of mice given intrarectal rcREG3A (Figure 6C). These positive effects of a local administration of a recombinant REG3A protein make it very likely that the lower susceptibility to colitis observed in *REG3A*-TG mice was mostly due to the hepatic transgene REG3A traveling to the gut lumen, and not to a liver-dependent effect. They indicate that an exogenous hREG3A can contribute to the preservation of gut barrier integrity during colitis and underlines the potential human-health relevance of this molecule.

## Discussion

Enteric innate immune molecules play an important role in gut barrier function and gut microbiota homeostasis through their pleiotropic activities. A reduced production of AMPs resulting from a deregulation of the function of Paneth's cell may lead to dysbiosis and ultimately IBD.<sup>3</sup>

**Figure 5.** Alleviated colitis severity in conventional and germ-free mice colonized with a hREG3A-shaped microbiota. (A) Relative abundances of bacterial families in the fecal microbiota of WT mice cohoused with *REG3A*-TG mice for 8 weeks (cohoused-WT; n = 8) compared with single-housed WT (n = 10) and *REG3A*-TG (n = 7) mice. (B–E) Time evolution of DSS-induced colitis in cohoused-WT and single-housed WT mice. Heavy line: 5-day period of DSS administration. (B) Body weight change. (C) Rectal bleeding score. (D) Diarrhea score. (E) Kaplan-Meier survival plot (P = nonsignificant). (G) Histological assessment of gut epithelium (n = 5). (H) PCA at the family level of germ-free mice colonized with microbiota from WT (ExGF-WT; n = 8) and *REG3A*-TG (ExGF-TG; n = 8) mice 3 weeks after mice were fed feces (day 0). inoc, inoculum used for oral gavage. P = .001 (Monte-Carlo rank test). (I) Relative abundances of the main bacterial families that are overrepresented in conventional *REG3A*-TG mice at day 0. (J–M) Time evolution of DSS-induced colitis in ExGF-WT and ExGF-TG mice. Heavy line: 5-day period of DSS administration. (J) Body weight change. (K) Rectal bleeding score. (L) Diarrhea score. (M) Kaplan-Meier survival plot. (N) Gene enrichment of TLR4 pathway on day 12 of colitis. NES, normalized enrichment score. (O) Serum LPS levels. (P) mRNA expression level of the indicated pro- and anti-inflammatory markers in colonic epithelial cells of ExGF-WT and ExGF-TG mice (n = 5). The data are averages  $\pm$  SEM. Except in (H), the 2-sided Wilcoxon rank sum test was performed for analysis. \*P < .05, \*\*P < .01, \*\*\*P < .001.



**Figure 6.** Intrarectal administration of hREG3A decreased colon damage and inflammation in WT mice with colitis. A total of 100  $\mu$ g of a recombinant REG3A protein (rcREG3A; n = 14) or an equivalent volume of buffer (n = 15) was administered intrarectally in WT mice with colitis on the day before (D0) and on the day (D1) of TNBS administration. (A) Body weight change. (B) Ameho quantitative histological score and representative colon images stained with hematoxylin and eosin. Scale bar, 100  $\mu$ m. (C) Colonic mRNA expression level of *Il1b* and *Tnf*. (D) Myeloperoxidase (MPO) in colon tissue extracts. The data are averages  $\pm$  SEM. The 2-sided Wilcoxon rank sum test was performed for analysis. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

A therapeutic approach consisting of increasing the intraluminal concentration of AMPs to preserve host-microbiota homeostasis and thus prevent intestinal inflammation was proposed.<sup>45</sup> However, preclinical models based on the enteric delivery of such biological agents were lacking because, among other things, of their proteolytic destruction in the upper digestive tract. The *REG3A*-TG mouse model we used enabled us to deliver a full-length human REG3A protein in the lower digestive tract lumen at sufficient doses to provide clear phenotypic effects in these mice, demonstrating the potential of an approach consisting of administering a large dose of exogenous REG3 to an organism already provided with an endogenous REG3-mediated response to intestinal inflammation.<sup>46</sup> The phenotypic features of *REG3A*-TG mice include significant changes in gut microbiota composition and a dramatic improvement of host resistance to intestinal inflammation, compared with control mice. *REG3A*-TG mice exposed to DSS exhibited very few signs of colitis, retained a tight mucosal barrier, and achieved complete survival. The inflammatory response and the oxidative stress in the colon epithelium was much reduced in *REG3A*-TG, compared with WT, mice with DSS-induced colitis. A low level of inflammation and mucosal damage was also found in *REG3A*-TG mice with

TNBS-induced colitis. The decrease in ROS levels in the fecal microbiota of *REG3A*-TG compared with control mice on DSS exposure and the increase of survival to ambient-air exposure of highly oxygen-sensitive bacteria cultured in the presence of rcREG3A demonstrate an antioxidant activity of hREG3A on prokaryotic cells that might modify host-microbiota interplay.

The microbial changes exhibited by *REG3A*-TG mice in homeostatic and inflammatory conditions mainly concerned an enrichment of Clostridiales (Ruminococcaceae, Lachnospiraceae) and a depletion of Bacteroidetes (Prevotellaceae). Mice heterozygous for *REG3A* that harbored a WT maternal microbiota at birth progressively acquired a gut microbiota composition close to that of *REG3A* homozygous TG mice, demonstrating the capability of hREG3A to shape the gut microbiota, and then exhibited a resistance to colitis similar to that of homozygous mice. The transfer of a hREG3A-shaped microbiota to WT mice by cohousing or feces gavage resulted in a less-severe disease, a reduced LPS-induced endotoxemia, and a higher survival to DSS exposure than in control groups, establishing the beneficial nature of the hREG3A-shifted microbiota composition. The phenotypic resistance was not fully transmitted by transfer experiments, suggesting that hREG3A acted also by other



ways, including direct mucosal effect, in *REG3A*-TG mice. Repeated intravenous administration of rcREG3A were ineffective, invalidating the hypothesis of a systemic anti-colitogenic effect, whereas rectal administrations of rcREG3A helped to preserve gut barrier integrity during induced colitis in control mice.

Thus, hREG3A provides protection against inflammation and oxidative stress for both the intestinal epithelium and the commensal communities that form the gut microbiota. Future studies will be necessary to determine the sites involved in the ROS scavenging activity of hREG3A. Whether murine *Reg3b* and *Reg3g* encode proteins that carry similar antioxidant properties also needs to be determined.

Regarding the role of hREG3A in the host-gut microbiota interaction, we suggest that the broad-spectrum antioxidant activity of hREG3A may change the balance between different bacterial communities with different oxygen tolerances. In our mouse model, some highly oxygen-sensitive commensal bacteria could be responsive to an exogenous antioxidant, such as hREG3A, giving them a selective advantage over aerotolerant anaerobic ones (ie, those developing an effective adaptive response to oxygen toxicity). In this hypothesis, hREG3A would shape the gut microbiota in the steady state through its antioxidant activity against the environmental stressors existing in the healthy gut and continue to exert this effect under inflammation and oxidative stress. This view is supported by the fact that hREG3A enhanced the growth of some highly oxygen-sensitive Clostridiales commensals. An enrichment of such symbionts may trigger a virtuous process by improving gut barrier function and reducing ROS production, which feeds back into benefiting symbionts. Conversely, an increase in gram-negative commensal bacteria may trigger macrophage activation, increase ROS production, and amplify mucosal injuries and dysbiosis in favor of aerotolerant commensals. Our results suggest that the pressure exerted by hREG3A on the gut microbiota triggers such a virtuous shift in composition and functionality.

From a clinical viewpoint, our findings suggest that an increase in intraluminal concentration of hREG3A, for instance via a colon-targeted delivery method, may be a valuable approach to attenuate intestinal inflammation. Regarding the potential benefits to patients, such an approach would be a more physiological and nontoxic strategy for the treatment of inflammatory outbreaks than the available therapies. Moreover, it might be most successful in the early stages of the inflammatory process and could therefore be useful for the maintenance of medically or surgically induced remission and even the prevention of IBD in high-risk individuals.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at [www.gastrojournal.org](http://www.gastrojournal.org), and at <https://doi.org/10.1053/j.gastro.2017.11.003>.

## References

1. Klag T, Stange EF, Wehkamp J. Defective antibacterial barrier in inflammatory bowel disease. *Dig Dis* 2013; 31:310–316.
2. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382–392.
3. Sartor RB, Wu GD. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. *Gastroenterology* 2017;152:327–339.e4.
4. Winter SE, Lopez CA, Bäumlner AJ. The dynamics of gut-associated microbial communities during inflammation. *EMBO Rep* 2013;14:319–327.
5. Albenberg L, Esipova TV, Judge CP, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* 2014; 147:1055–1063.e8.
6. Ballal SA, Veiga P, Fenn K, et al. Host lysozyme-mediated lysis of *Lactococcus lactis* facilitates delivery of colitis-attenuating superoxide dismutase to inflamed colons. *Proc Natl Acad Sci U S A* 2015; 112:7803–7808.
7. Dubourg G, Lagier J-C, Hüe S, et al. Gut microbiota associated with HIV infection is significantly enriched in bacteria tolerant to oxygen. *BMJ Open Gastroenterol* 2016;3:e000080.
8. Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature* 2002;415:389–395.
9. Vaishnav S, Yamamoto M, Severson KM, et al. The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine. *Science* 2011;334:255–258.
10. Salzman NH, Hung K, Haribhai D, et al. Enteric defensins are essential regulators of intestinal microbial ecology. *Nat Immunol* 2010;11:76–83.
11. Dupont A, Kacanis Y, Yang I, et al. Intestinal mucus affinity and biological activity of an orally administered antibacterial and anti-inflammatory peptide. *Gut* 2015; 64:222–232.
12. Wang G. Human antimicrobial peptides and proteins. *Pharmaceuticals (Basel)* 2014;7:545–594.
13. Cullen TW, Schofield WB, Barry NA, et al. Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* 2015;347:170–175.
14. Pang X, Xiao X, Liu Y, et al. Mosquito C-type lectins maintain gut microbiome homeostasis. *Nat Microbiol* 2016;1:16023.
15. Juneja P, Lazzaro BP. Population genetics of insect immune responses. In: Rolff J, Reynolds S, eds. *Insect infection and immunity*. Oxford, UK: Oxford University Press, 2009:206–224.
16. Goldman MJ, Anderson GM, Stolzenberg ED, et al. Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* 1997;88:553–560.
17. Schroeder BO, Wu Z, Nuding S, et al. Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. *Nature* 2011;469:419–423.

18. Abou Alaiwa MH, Reznikov LR, Gansemer ND, et al. pH modulates the activity and synergism of the airway surface liquid antimicrobials  $\beta$ -defensin-3 and LL-37. *Proc Natl Acad Sci U S A* 2014;111:18703–18708.
19. Moniaux N, Song H, Darnaud M, et al. Human hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein cures fas-induced acute liver failure in mice by attenuating free-radical damage in injured livers. *Hepatology* 2011;53:618–627.
20. Moniaux N, Darnaud M, Garbin K, et al. The Reg3 $\alpha$  (HIP/PAP) lectin suppresses extracellular oxidative stress in a murine model of acute liver failure. *PLoS ONE* 2015; 10:e0125584.
21. Lieu H-T, Batteux F, Simon M-T, et al. HIP/PAP accelerates liver regeneration and protects against acetaminophen injury in mice. *Hepatology* 2005;42:618–626.
22. Lieu H-T, Simon M-T, Nguyen-Khoa T, et al. Reg2 inactivation increases sensitivity to Fas hepatotoxicity and delays liver regeneration post-hepatectomy in mice. *Hepatology* 2006;44:1452–1464.
23. Lai Y, Li D, Li C, et al. The antimicrobial protein REG3A regulates keratinocyte proliferation and differentiation after skin injury. *Immunity* 2012;37:74–84.
24. Lv Y, Yang X, Huo Y, et al. Adenovirus-mediated hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein suppresses dextran sulfate sodium-induced acute ulcerative colitis in rats. *Inflamm Bowel Dis* 2012;18:1950–1960.
25. **Haldipur P, Dupuis N**, Degos V, et al. HIP/PAP prevents excitotoxic neuronal death and promotes plasticity. *Ann Clin Transl Neurol* 2014;1:739–754.
26. **Cash HL, Whitham CV**, Behrendt CL, et al. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 2006;313:1126–1130.
27. **Stelter C, Käppeli R**, König C, et al. Salmonella-induced mucosal lectin RegIII $\beta$  kills competing gut microbiota. *PLoS ONE* 2011;6:e20749.
28. **Ampting MTJ van, Loonen LMP**, Schonewille AJ, et al. Intestinally secreted C-type lectin Reg3b attenuates salmonellosis but not listeriosis in mice. *Infect Immun* 2012;80:1115–1120.
29. Miki T, Holst O, Hardt W-D. The bactericidal activity of the C-type lectin RegIII $\beta$  against Gram-negative bacteria involves binding to lipid A. *J Biol Chem* 2012;287: 34844–34855.
30. Medveczky P, Szmola R, Sahin-Tóth M. Proteolytic activation of human pancreatitis-associated protein is required for peptidoglycan binding and bacterial aggregation. *Biochem J* 2009;420:335–343.
31. Loonen LMP, Stolte EH, Jaklofsky MTJ, et al. REG3 $\gamma$ -deficient mice have altered mucus distribution and increased mucosal inflammatory responses to the microbiota and enteric pathogens in the ileum. *Mucosal Immunol* 2014;7:939–947.
32. Simon M-T, Pauloin A, Normand G, et al. HIP/PAP stimulates liver regeneration after partial hepatectomy and combines mitogenic and anti-apoptotic functions through the PKA signaling pathway. *FASEB J* 2003;17:1441–1450.
33. Bus JS, Gibson JE. Paraquat: model for oxidant-initiated toxicity. *Environ Health Perspect* 1984;55:37–46.
34. Ubeda C, Lipuma L, Gobourne A, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J Exp Med* 2012;209:1445–1456.
35. HSP de Souza, Fiocchi C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* 2016;13:13–27.
36. Willing BP, Dicksved J, Halfvarson J, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 2010;139: 1844–1854.e1.
37. **Quévrain E, Maubert MA, Michon C**, et al. Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* 2016;65:415–425.
38. Abbeele P Van den, Belzer C, Goossens M, et al. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J* 2013;7:949–961.
39. Duncan SH, Aminov RI, Scott KP, et al. Proposal of *Roseburia faecis* sp. nov., *Roseburia hominis* sp. nov. and *Roseburia inulinivorans* sp. nov., based on isolates from human faeces. *Int J Syst Evol Microbiol* 2006; 56:2437–2441.
40. Khan MT, Duncan SH, Stams AJM, et al. The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases. *ISME J* 2012;6:1578–1585.
41. **Elinav E, Strowig T**, Kau AL, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 2011;145:745–757.
42. Bloom SM, Bijanki VN, Nava GM, et al. Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* 2011;9:390–403.
43. Cash HL, Whitham CV, Hooper LV. Refolding, purification, and characterization of human and murine RegIII proteins expressed in *Escherichia coli*. *Protein Expr Purif* 2006;48:151–159.
44. Wirtz S, Popp V, Kindermann M, et al. Chemically induced mouse models of acute and chronic intestinal inflammation. *Nat Protoc* 2017;12:1295–1309.
45. Wehkamp J, Koslowski M, Wang G, et al. Barrier dysfunction due to distinct defensin deficiencies in small intestinal and colonic Crohn's disease. *Mucosal Immunol* 2008;1(Suppl 1):S67–S74.
46. Ogawa H, Fukushima K, Naito H, et al. Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm Bowel Dis* 2003; 9:162–170.

---

Author names in bold designate shared co-first authorship.

Received January 18, 2017. Accepted November 6, 2017.

#### Reprint requests

Address requests for reprints to: Jamila Faivre, MD, PhD, INSERM U1193, University Paris-Saclay, Paul-Brousse Hospital, Hepatobiliary Centre, 14 Av. Paul Vaillant-Couturier, Villejuif 94800, France. e-mail: jamila.faivre@inserm.fr; fax: +3314559026.

**Acknowledgments**

We warmly thank Dr H. Blottière, Professor F. Carbonnel, Professor J.F. Colombel, and Dr P. Langella for fruitful discussions. For experimental assistance, we thank R. Duchateau, Dr C. Martin, Dr K. Lipson (animal care), F. Levenez (gnotobiology, 16S DNA extraction), L. Calvo (16S rDNA libraries), C. Bridonneau (anaerobic bacterial strains), Dr P. Le Baccon (fluorescence microscopy), P. Duchambon (recombinant proteins), the Alfact Innovation Company (ALF5755 recombinant protein), the IGBMC Platform (microarray), the CLNS@Sapienza genomics Platform (sequencing), and the INRA GenoToul Bioinformatics Facility (computing and storage resources).

Jinzi Zheng's present address: Institute of Biomaterials and Biomedical Engineering, University of Toronto, Ontario M5G 1L7, Canada.

**Conflicts of interest**

The authors disclose no conflicts.

**Funding**

This work was supported by grants from the Fondation ARC pour la Recherche sur le Cancer (PJA 20111203952, PJA 20131200221, and PJA 20141202048 to J.F.; PJA 20111203952 to N.M.). Sandrine Augui was funded by an Association pour la Recherche sur le Cancer Fellowship (R12036LL). Jamila Faivre was supported by the European Union's Seventh Framework Programme (FP7) under grant agreement No. 259743 (MODHEP consortium), the Institut National du Cancer, the OSEO-BPI Programme d'Investissements d'Avenir (IMODI and HECAM consortiums; R14035LB and R15065LH, respectively).

## Supplementary Methods

### Animal Studies

Animal studies were performed in compliance with the institutional and European Union guidelines for laboratory animal care and approved by the Ethics Committee of CE2A-03 CNRS-Orléans (Accreditation N°01417.01). The number of *Mus musculus* mice used was in compliance with institutional ethical rules and consistent with common practice in the fields of microbiota analysis and experimental colitis. Conventional mice were produced and housed in the CNRS SEAT and Institut André Lwoff animal care facilities (Université Paris-Sud, Villejuif). Germ-free mice were housed at the Anaxem platform of the Micalis Institute (INRA, Jouy-En-Josas). All the conventionally bred mouse groups were homogeneous in terms of age (10–12 weeks), weight (25 g), sex (50% male and 50% female), and environmental factors. All the *REG3A*-TG mice had circulating hREG3A levels within the 200 to 600 ng/mL range. None was excluded from the study. WT and *REG3A*-TG mice had the same C57BL/6N genetic background. They were not littermates. WT female mice were crossed with *REG3A* homozygous males to generate heterozygous pups harboring the maternal microbiota at birth. For the cohousing experiments, 3-week-old weaned *REG3A*-TG mice were cohoused randomly with age-matched WT mice at a 1:1 ratio for 8 weeks. Some of these mice were exposed to DSS, and others were used for tissue analysis at the end of the cohousing period. For the feces transfer experiments, male germ-free C57BL/6 mice were housed in sterile isolators. Their germ-free status was checked weekly using microbiological assays on feces and drinking water. At the age of 10 weeks, the germ-free mice underwent 2 gastric gavages at 48-hour intervals, using 0.5 mL of inoculum prepared from conventional mice fecal microbiota diluted 1/100. Inocula consisted of pooled feces from 8 TG or 8 WT mice. To induce colitis, mice were fed with a 3% 40-kDa DSS solution (TdB Consultancy AB) added to their drinking water for 5 days (solution changed every day). The animals were then fed with tap water for 7 additional days. Weight loss, stool consistency, and bleeding were evaluated daily. Scores are defined as follows: stool consistency: 0 (normal), 1 (soft), 2 (very soft), 3 (diarrhea); stool bleeding: 0 (normal), 1 (red), 2 (dark red), 3 (gross bleeding). For colitis induction by TNBS (Sigma-Aldrich), overnight-fasted mice were anesthetized for 90 minutes and received a single intrarectal administration of TNBS (40  $\mu$ L, 150 mg/kg) dissolved in a 1:1 mixture of 0.9% NaCl with 100% ethanol. Control mice received a 1:1 mixture of 0.9% NaCl with 100% ethanol. The WT mice that were administered the recombinant hREG3A protein received an intrarectal injection of 100  $\mu$ g rcREG3A on the day before and on the day of TNBS administration. Macroscopic examination, histological analysis, and quantitative polymerase chain reaction (qPCR) were performed 2 days after TNBS administration. The different mouse groups (WT or TG, single-housed or cohoused, conventionally fed or germ-free) were used exhaustively.

### Composition of the Microbiota

DNA pyrosequencing of the V3-V4 hypervariable region of 16S rRNA gene was performed using the Illumina MiSeq instrument. The sequences thus generated were pre-processed using the mothur pipeline and then clustered into operational taxonomic units based on 97% of similarity. Using the RDP Classifier, we assigned operational taxonomic units to taxa (genus-level) with a confidence threshold of 80%. RDP release 11 was used as the reference database. PCA was performed based on bacterial family and genera compositions using the R packages *ade4* and *FactoMiner*. Families representing more than 0.5% of the total bacterial population were taken into account. The robustness of each clustering result was assessed using a Monte-Carlo rank test ( $B = 999$  repetitions,  $P < .05$ ). Differences between groups were analyzed using a nonparametric Wilcoxon rank-sum test. All the relevant data can be found in the text, except for the DNA sequence reads, which are available from the Sequence Read Archive (accession numbers pending) or by contacting the corresponding author.

### Bacterial Strains, Culture Conditions, and Flow Cytometry

Gram-positive *Enterococcus faecalis* ATCC-19433 were grown at 37°C under shaking in BHI (BD Bacto Brain Heart Infusion; Thermo Fisher Scientific, France) broth overnight and then in LB (Luria-Bertani; Becton Dickinson, Franklin Lakes, NJ). To induce peroxide stress on *E faecalis*,  $5.10^6$  bacteria per mL were incubated at 37°C with 200 mM paraquat (1,1'-dimethyl-4,4'-bipyridinium dichloride; Sigma-Aldrich) and 10  $\mu$ M of rcREG3A protein or equivalent volume of buffer. The reagents were added to the culture during the log phase. Colony-forming units (CFUs) were counted after plating serial dilutions on LB agar plates incubated at 37°C overnight. All cultures were performed in quadruplicate during 3 independent experiments. Intracellular ROS levels and cell viability were determined by flow cytometry with 2',7'-dichlorofluorescein diacetate (H<sub>2</sub>-DCFDA; Life Technologies) and propidium iodide (Sigma-Aldrich) in *E faecalis* cultures incubated with paraquat alone or paraquat supplemented with 10  $\mu$ M rcREG3A for 8 hours. H<sub>2</sub>-DCFDA was dissolved in dimethyl sulfoxide (Sigma-Aldrich) to 10 mM and added at a final concentration of 10  $\mu$ M. After incubating *E faecalis* cultures for 30 minutes in the dark at 37°C, the cells were pelleted, washed twice with phosphate-buffered saline (PBS), suspended in 1 mL PBS supplemented with 1  $\mu$ g/mL propidium iodide, and processed in a BD Accuri C6 flow cytometer (Becton Dickinson). For the measurements of bacterial ROS levels in fecal microbiota, fresh feces were suspended in cold PBS and centrifuged twice at 447 g for 5 minutes at 4°C. The supernatant was incubated with PBS containing 25  $\mu$ M H<sub>2</sub>-DCFDA for 15 minutes and then centrifuged at 2795 g for 5 minutes. The resulting pellet was suspended in cold PBS and processed in the flow cytometer. Analysis of the flow data was performed using Flow Jo software (Ashland, OR). *Faecalibacterium prausnitzii* A2-165 (DSM17677) and *Roseburia intestinalis* L1-82 (DSM14610T) were grown under strict anaerobic conditions at 37°C for 24 hours in LYHBHI medium (BHI broth supplemented with 0.5% yeast extract and

5 mg/L hemin) supplemented with 1 mg/mL maltose, 1 mg/mL cellobiose, and 0.5 mg/mL cysteine-HCl. Anaerobic growth of *F. prausnitzii* and *R. intestinalis* was performed in the presence of 5  $\mu$ M of rcREG3A for 7 or 24 hours, respectively, and followed by a 5-minute exposure to ambient air. Bacterial growth was assessed by CFU counting on agar plates. CFU numbers were normalized to the average CFU number measured in control cultures (buffer). Protein extracts from *F. prausnitzii* grown for 2 or 24 hours with rcREG3A or buffer were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Each experiment was performed in quadruplicate during 3 independent experiments. Wet mounts were observed using a Deltavision optical microscope (Applied Precision Inc, Issaquah, WA) with a  $\times 63$  objective lens.

### Recombinant REG3A Proteins

Two batches of recombinant proteins were used; one (ALF5755) supplied by the Alfact Innovation Company (Paris, France) and the other made in our laboratory. We subcloned the coding sequence corresponding to the secreted form of hREG3A within pET28-N-HIS-SUMO (Novagen, Molsheim, France). The resulting construct was verified by sequencing and then transformed into an engineered *Escherichia coli* BL21 (DE3) strain co-expressing the sulfhydryl oxidase, Erv1p, and the disulphide isomerase, DsbC. These modifications enabled the efficient production of natively folded hREG3A in the cytoplasm. The hREG3A proteins were produced in a 7L Infors-Labfors fermenter (Infors Ltd, Massy, France) at 20°C and pH 6.8  $\pm$  0.1 for 16 hours. These cells were then pelleted and lysed using a one-shot desintegrator (CellD, Roquemaure, France) at 1.5 kbar. The proteins were purified from the supernatant using Ni-NTA resin (Macherey Nagel, Duren, Germany), cleaved with a home-made SUMO protease. The His-tag and Sumo protease were removed from the protein suspension by a Ni-NTA resin. The resulting supernatant was dialyzed against 50 mM phosphate buffer pH 7.4.

### Enzyme-Linked Immunosorbent Assays

The serum concentrations of Reg3g, hREG3A, LPS, and soluble CD14 were measured using the E94676Mu (Uscn Life Science Inc, Houston, TX), PancrePAP (Dynabio, Lyon, France), LAL chromogenic endpoint (Hycult biotech, Plymouth Meeting, PA) and One Step ELISA (BioMetec, Greifswald, Germany) kits, respectively, according to the manufacturer's instructions. The myeloperoxidase concentration in colon tissues homogenates was determined using the mouse MPO ELISA kit (Hycult biotech).

### Histological Analysis

For each mouse studied, 2 samples from the distal and proximal colon were fixed in 4% paraformaldehyde acid and then embedded in paraffin; 4- $\mu$ m-thick sections were deparaffinized in xylene for 5 minutes and rehydrated for 5 minutes in 100%, 90%, and 70% ethanol successively. They were then stained with hematoxylin/eosin/Alcian blue, randomly sorted, and scored in a blind manner by an expert in pathology. The score used for DSS-induced colitis was as follows: Grade 0, normal mucosa; Grade 1, infiltration of

inflammatory cells; Grade 2, crypt abscesses and erosion; and Grade 3, destruction of epithelial cells (ulceration and loss of Goblet cells). The number of sections analyzed per colon sample ranged from 2 to 6. The Ameho score used for TNBS-induced colitis reads: Score 0, no lesion; Score 1, mild mucosal or submucosal inflammatory infiltrate with edema, punctuate mucosal erosions, muscularis mucosae intact; Score 2, score 1 changes involving 50% of the specimen; Score 3, prominent inflammatory infiltrate with deeper areas of ulceration extending through the muscularis mucosae into the submucosa; Score 4, score 3 changes involving 50% of the specimen; Score 5, extensive ulceration with coagulative necrosis extending deeply into the muscularis mucosae; and Score 6, score 5 changes involving 50% of the specimen.

### Immunohistochemistry

Formalin-fixed, paraffin-embedded colon sections (4  $\mu$ m) were de-waxed in xylene and rehydrated through graded alcohols. Tissue sections were pressure cooked in 10 mM citrate buffer pH 6 for 5 minutes, incubated with 3% H<sub>2</sub>O<sub>2</sub> for 20 minutes to block endogenous peroxidase, and then incubated with a laboratory-made purified rabbit polyclonal anti-hREG3A antibody.<sup>1</sup> Secondary anti-rabbit immunoglobulin G–horseradish peroxidase (HRP) was used according to the manufacturer's instructions (DAKO, Glostrup, Denmark). The sections were counterstained with Alcian blue (Sigma) before mounting the coverslips.

### Lipid Peroxidation

Malondialdehyde was dosed using the thiobarbituric acid method as previously described.<sup>1</sup> A standard curve was prepared using malondialdehyde tetrabutylammonium salt (Sigma-Aldrich, Lyon, France). The results were normalized in relation to the total protein content.

### In Situ Hybridization and Immunofluorescence

Unwashed segments of colon and mesenteric lymph nodes were fixed in methanol-Carnoy mixture, followed by embedding in paraffin. Paraffin sections (3 to 5  $\mu$ m) were de-waxed and washed in 100% ethanol. Fluorescence in situ hybridization (FISH) was performed as previously described.<sup>2</sup> The oligonucleotide probes used are EUB338 and univ1390 for universal 16S RNA (see the probeBase online resource). Bacterial translocation levels in mesenteric lymph nodes were determined on FISH images in at least 1 to 2 lymph nodes per mouse and 3 sections per node. Anti-Mucin2 antibody (H300, sc-15334; Santa Cruz Biotechnology, Heidelberg, Germany) was used at a dilution of 1:100 in 4% goat serum. The Muc2 signal was detected using a goat anti-rabbit immunoglobulin G–Alexa 598 (A11036; Molecular Probes, Eugene, OR). The tissues were counterstained with 4',6-diamidino-2-phenylindole, and fluorescent images were acquired using a Deltavision real-time microscope piloted by SoftWorx software (Applied Precision, GE Healthcare, Little Chalfont, UK), with a  $\times 40/1.4$ NA objective (Olympus, Tokyo, Japan) and a charge-coupled device CoolSNAP HQ2 camera (Photometrics, Tucson, AZ).

### Transcriptomics, qPCR

RNAs were purified from colon tissues using the mir-Vana miRNA isolation kit (Ambion, Austin, TX). RNA quality was evaluated with the Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). Samples with an RNA integrity number higher than 7 were processed for transcriptomic analysis. Biotinylated single strand cDNA targets were prepared from 200 ng total RNA using the Ambion WT Expression Kit and the Affymetrix (Santa Clara, CA) GeneChip WT Terminal Labeling Kit, according to the manufacturer's instructions. A total of 1.9  $\mu\text{g}$  of cDNAs were hybridized for 16 hours at 45°C on GeneChip Mouse Gene 1.0 ST arrays (Affymetrix) interrogating 28,853 genes represented by approximately 27 probes spread along the full length of the gene. The chips were washed and stained in the GeneChip Fluidics Station 450, and scanned with the GeneChip Scanner 3000 7G at a resolution of 0.7  $\mu\text{m}$ . CEL files were further processed with Affymetrix Expression Console software version 1.1 to calculate the probe set signal intensities using Robust Multi-array Average algorithms with default settings. Transcriptomic data have been deposited in Gene Expression Omnibus (National Center for Biotechnology Information, Bethesda, MD) under the accession number GSE64932. For qPCR, the sequences of the specific primers used were as follows. Tnf-F: GGGAG TAGACAAGGTACAAC; Tnf-R: TCTCATCAGTTCTATGGCCC. Il1-F: CAACCAACAAGTGATATTCTCCATG; Il1-R: GATCCA CACTCTCCAGCTGCA. Hprt-F: AGGACCTCTCGAAGTGT; Hprt-R: TCAATCCCTGAAGTACTCAT.

### Gene Set Enrichment Analysis

Gene set enrichment analysis was performed to explore the gene sets differentially expressed between WT and *REG3A*-TG mice. Gene set enrichment analysis version 2.0 software was downloaded from the Broad Institute (Cambridge, MA) Web site, working with version 4.0 of MsigDB. The most significant gene sets are displayed with their corresponding normalized enrichment scores and nominal *P* values. Medline/PubMed data mining for LPS-endotoxin functionality was performed using LPS and endotoxin as the keywords. A total of 356 genes with a raw *P* value less than .05 were retrieved from 4267 articles. Eighty-five genes were still related to LPS-induced cellular responses after correction of the multitesting error by the false discovery rate. Unsupervised hierarchical clustering (Euclidean distance and complete linkage) was performed using the expression profiles of the 85-gene set.

### Immunoblotting

The quantification of hREG3A protein along the gastrointestinal tract of *REG3A*-TG mice was performed using the anti-hREG3A immunoblotting of 50  $\mu\text{g}$  total protein. Immunodetection of the rcREG3A in bacterial cultures was performed by spinning down 200  $\mu\text{L}$  of a bacterial culture previously incubated with rcREG3A or buffer for 2 or 24 hours. The pellet was washed 3 times with PBS supplemented with a protease inhibitor cocktail (eComplete; Roche, Basel, Switzerland), resuspended in 30  $\mu\text{L}$  Laemmli

buffer, and stored at  $-20^{\circ}\text{C}$ . After denaturation for 5 minutes at  $95^{\circ}\text{C}$ , the sample extracts were separated on 15% polyacrylamide gel by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and then transferred onto a nitrocellulose membrane for immunoblotting using our anti-hREG3A antibody. Total protein lysates were extracted from mouse colon tissues with RIPA buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% NP40, 0.25% sodium deoxycholate, protease inhibitors). To detect O-glycoproteins containing Core 3, 500 ng protein was resolved on 4% to 15% precast polyacrylamide gel (Bio-Rad, Hercules, CA), electro-transferred onto a nitrocellulose membrane, and hybridized with wheat germ agglutinin conjugated with HRP (Sigma-Aldrich).

### Dot Blot

Monosaccharides (glucose, mannose) and polysaccharides (LPS, peptidoglycan, mannan, DSS) (Sigma-Aldrich) at 2 doses (30 ng, 150 ng) were spotted onto nitrocellulose membranes and cross-linked to the membrane using UV irradiation. The membranes were washed twice with PBS, and hybridized with 1  $\mu\text{g}/\text{mL}$  of rcREG3A at  $37^{\circ}\text{C}$  for 2 hours. Nonhybridized hREG3A proteins were washed away with TBST (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% Tween), The membranes were incubated with our anti-hREG3A antibody for 30 minutes at room temperature and then with HRP-conjugated goat anti-rabbit antibody to reveal the hREG3A proteins attached to saccharides.

### [ $^{18}\text{F}$ ]FDG Positron Emission Tomography Scan

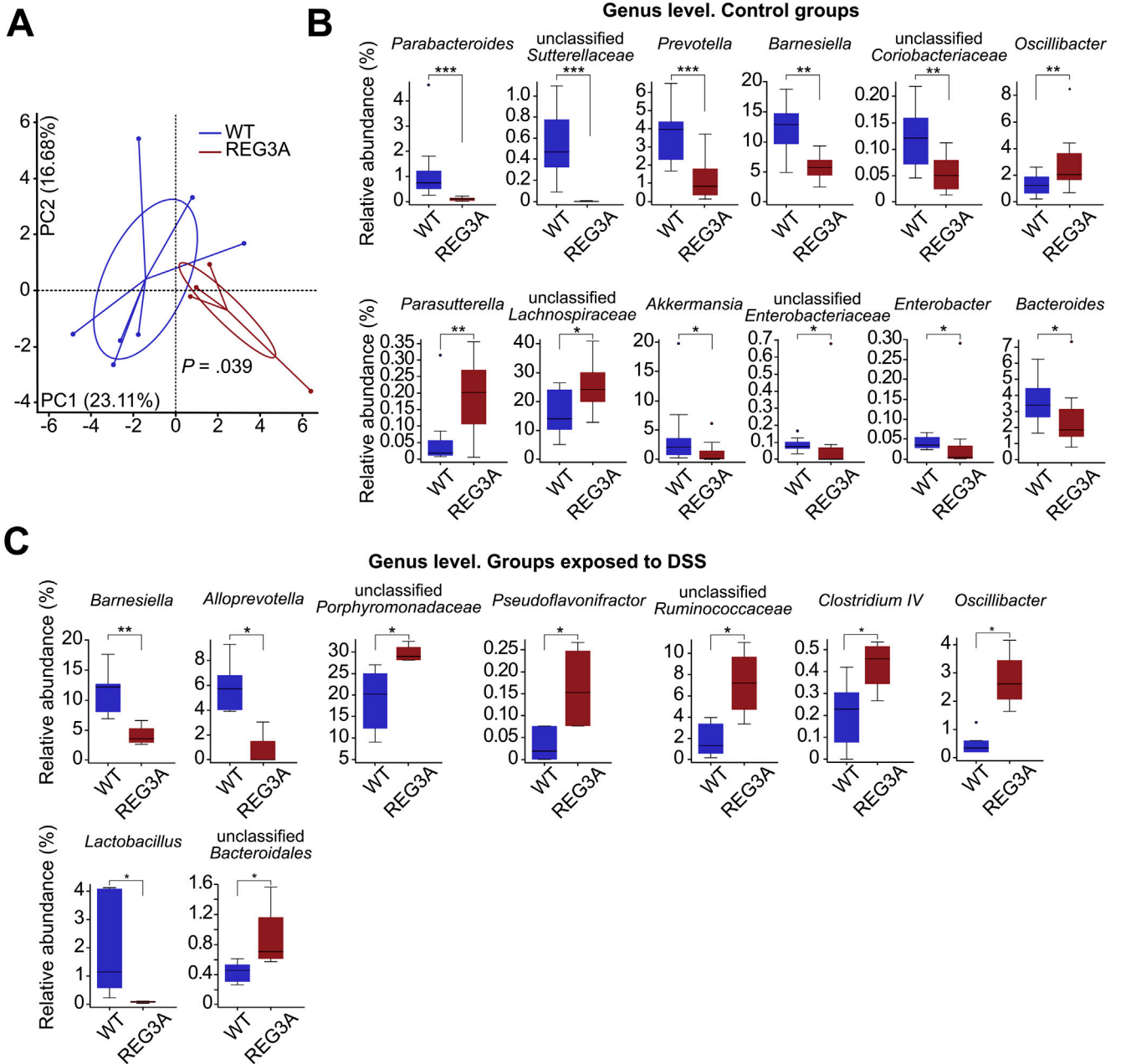
[ $^{18}\text{F}$ ]FDG was purchased from Cyclopharma S.A. (Clermont-Limagne, France). The mice to be imaged were anesthetized with oxygen-isoflurane (2% isoflurane, 0.4 L/min oxygen) and kept at a physiological body temperature on a heating pad. Three male WT and *REG3A*-TG mice were injected intravenously with 10 MBq [ $^{18}\text{F}$ ]FDG in 100  $\mu\text{L}$  PBS on different days of induced colitis. The images were acquired dynamically over 15 minutes as from 1 hour after the radioisotope injection, using a microPET Focus (Siemens, Munich, Germany). The uptake of radioactivity was quantified in regions of interest drawn in the abdominal areas of the reconstructed PET images. For each region of interest, the mean uptake was calculated as the percent-injected dose per gram tissue (%ID/g) assuming a tissue density of 1 g/ $\text{cm}^3$ .

### Statistics

Results are presented as means  $\pm$  SEM or box plots. Differences between sample groups were tested using nonparametric Wilcoxon rank sum (2-sided) for all the experiments reported, with the following exceptions: 2-way analysis of variance tests were used for the in vivo quantifications of [ $^{18}\text{F}$ ]FDG and  $\text{H}_2$ -DCFDA over time; the robustness of PCA clustering was assessed with a Monte-Carlo rank test; the Kaplan-Meier survival analysis with a log-rank test was applied to compare the time course of survival between the groups. *P* values less than .05 were considered to be statistically significant.

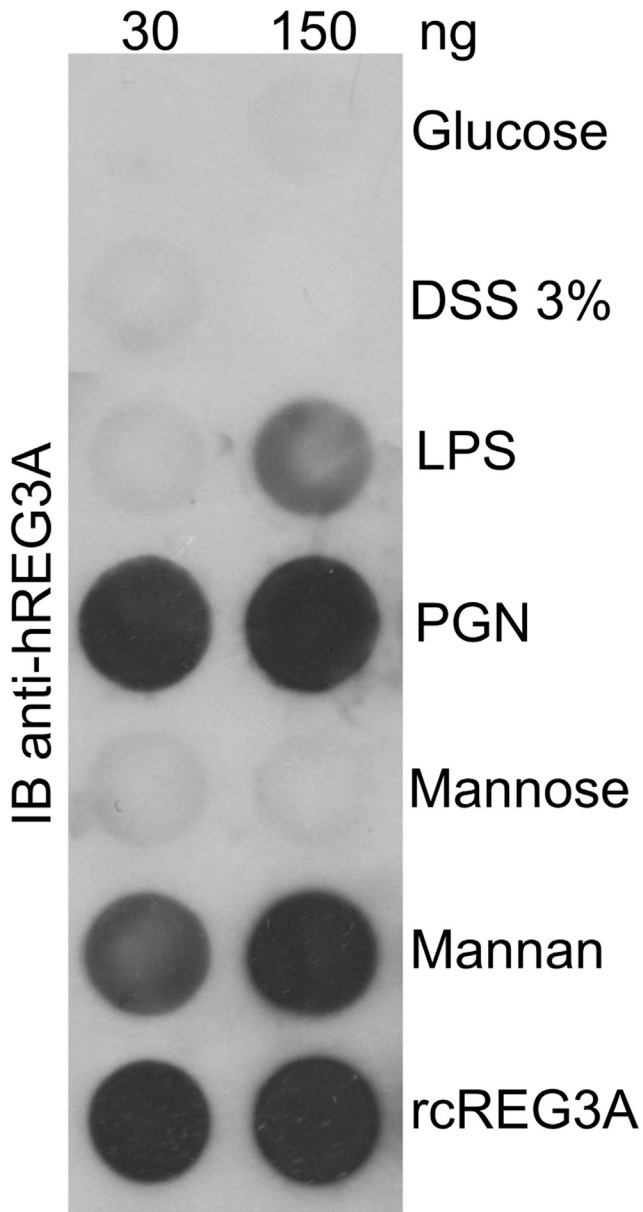
## Supplementary References

1. Moniaux N, Song H, Darnaud M, et al. Human hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein cures fas-induced acute liver failure in mice by attenuating free-radical damage in injured livers. *Hepatology* 2011;53:618–627.
2. Canny G, Swidsinski A, McCormick BA. Interactions of intestinal epithelial cells with bacteria and immune cells: methods to characterize microflora and functional consequences. *Methods Mol Biol* 2006;341:17–35.

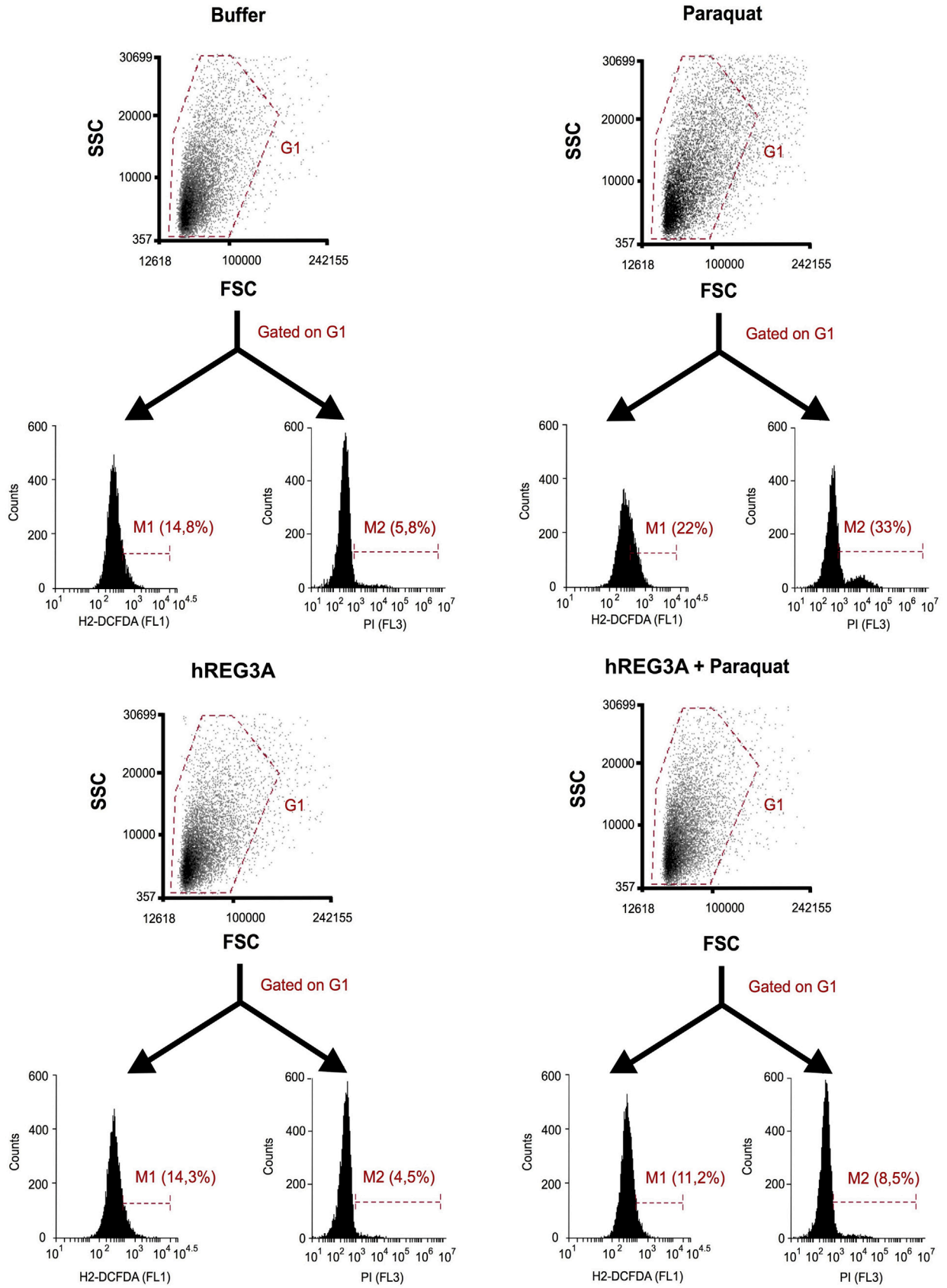


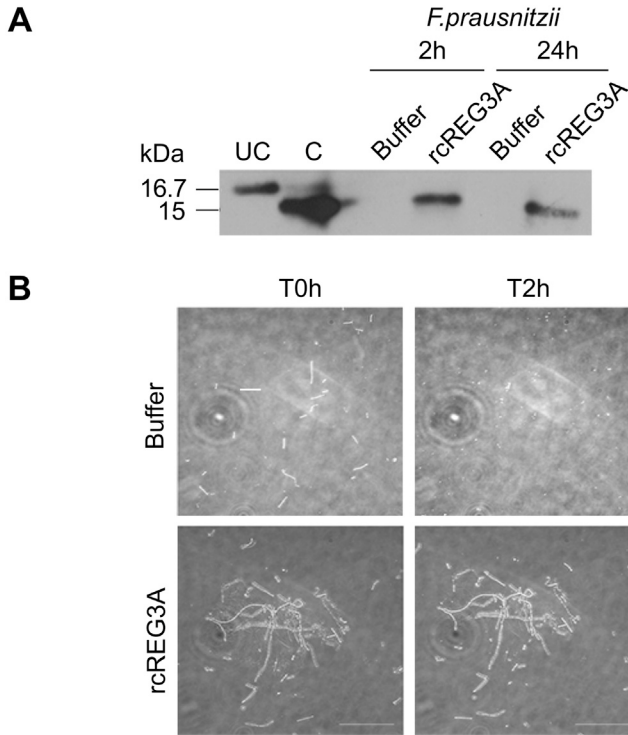
**Supplementary Figure 1.** Gut microbiota composition under homeostatic and inflammatory conditions in *REG3A*-transgenic mice. (A) PCA plots of bacterial profiles at the family level in *REG3A*-TG ( $n = 4$ ) and WT ( $n = 7$ ) mice on day 12 of induced colitis.  $P = .039$  (Monte-Carlo rank test). Each dot represents 1 mouse. (B) and (C) Relative abundances of bacterial genera in (B) the homeostatic (Control groups) and (C) inflammatory (Groups exposed to DSS) states. The data are averages  $\pm$  SEM. The Wilcoxon rank-sum test was performed for analysis.



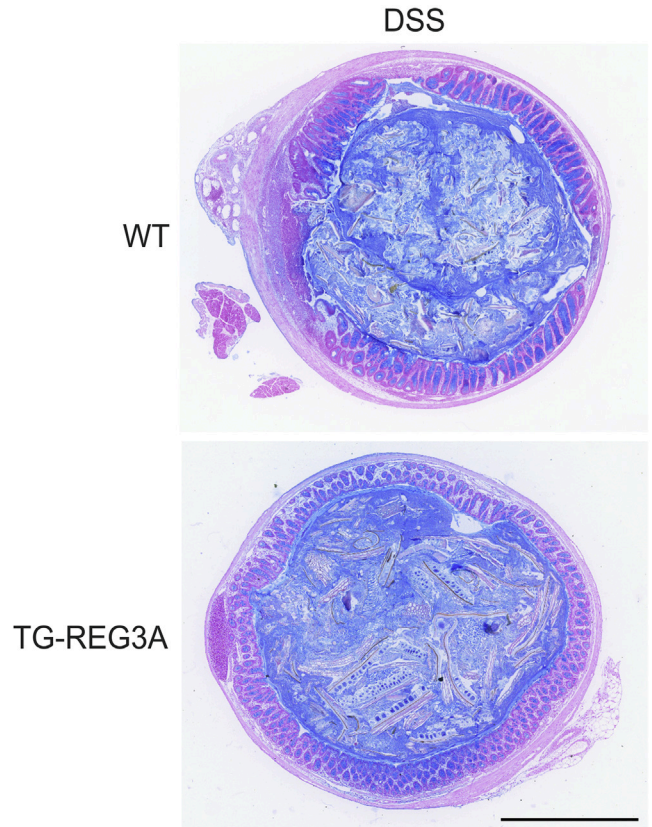


**Supplementary Figure 2.** The hREG3A lectin does not interact with DSS. Dot blot of hREG3A binding to the indicated immobilized mono- and polysaccharides followed by anti-hREG3A immunoblotting (IB). PGN, peptidoglycans. The lane labeled rcREG3A received no saccharides.



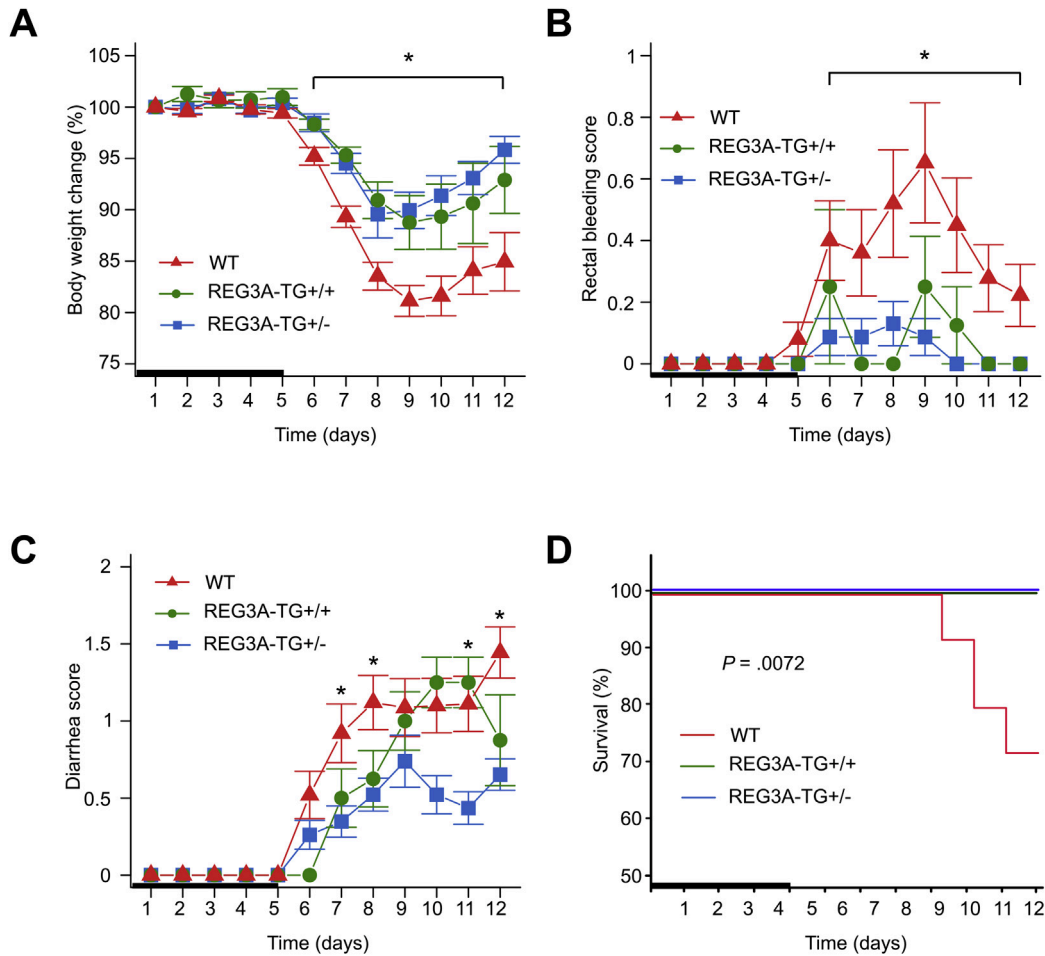


**Supplementary Figure 4.** hREG3A increases the viability of *F. prausnitzii* exposed to ambient air. (A) Anti-hREG3A immunoblot from *F. prausnitzii* anaerobic cultures incubated with a rcREG3A for 2 or 24 hours. Uncleaved (UC) and cleaved (C) rcREG3A: control signals. (B) Micrographs of slide-mounted *F. prausnitzii* colonies without or with rcREG3A before (T0h) and after (T2h) exposure to ambient air. Scale bar, 20  $\mu$ m.



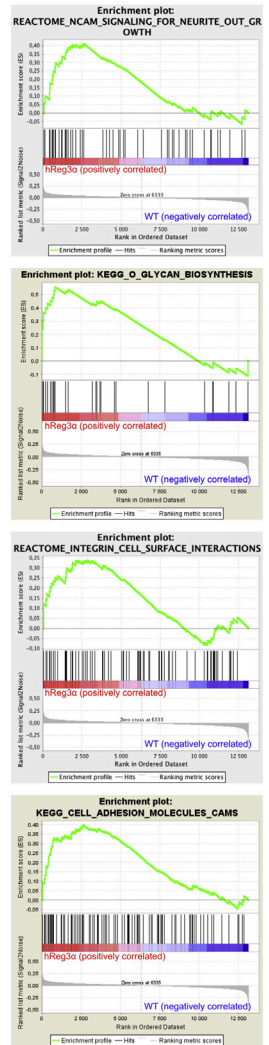
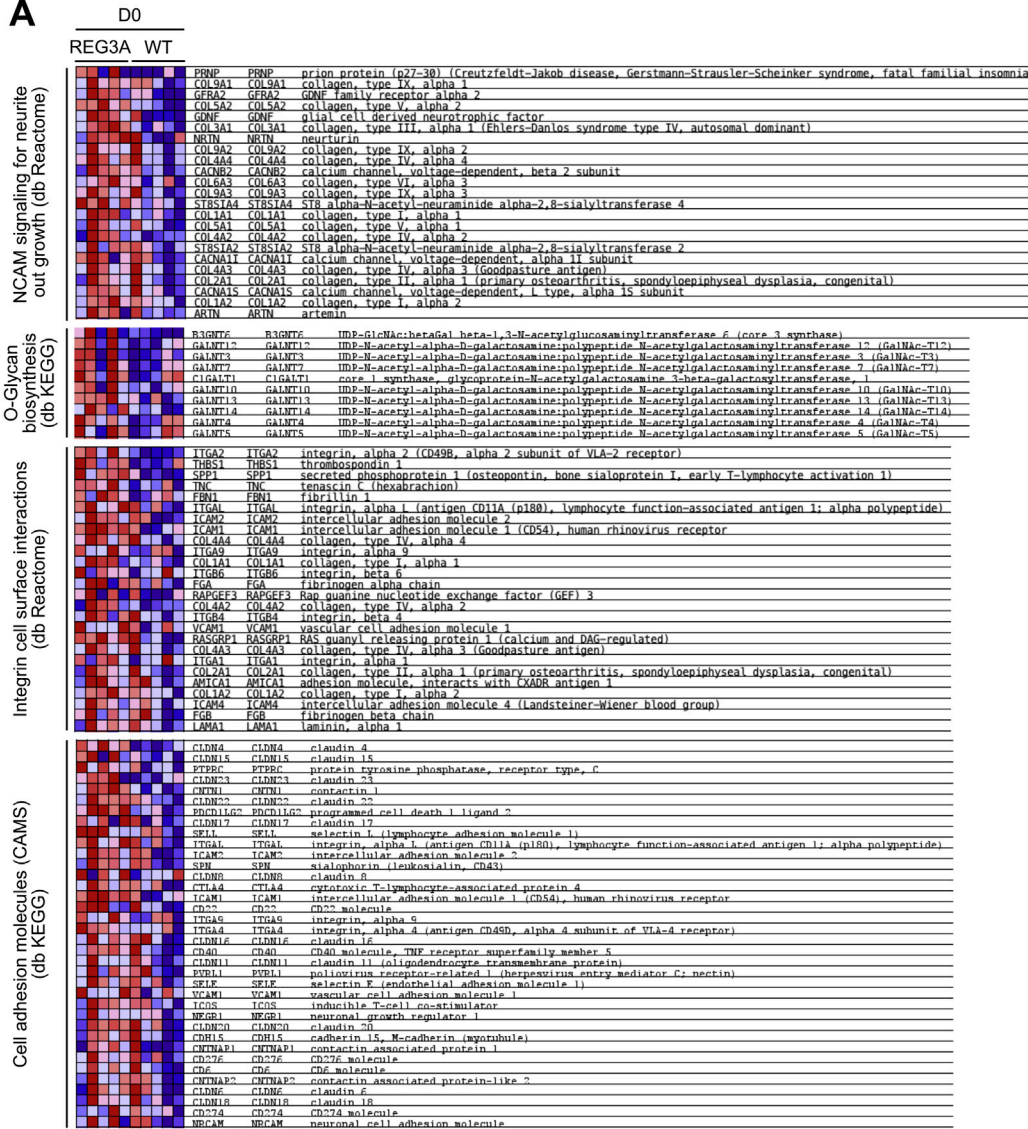
**Supplementary Figure 5.** Representative colon images stained with hematoxylin, eosin, and Alcian blue at day 12 of DSS-induced colitis in *REG3A*-transgenic (TG-REG3A) and WT mice. Same colon sections as in Figure 3E. Scale bar, 1 mm.

**Supplementary Figure 3.** Representative flow cytometry plots of cellular ROS levels ( $H_2$ -DCFDA dye; FL1 detector) and cell viability (propidium iodide [PI]; FL3 detector) in *E. faecalis* cultures exposed to paraquat and hREG3A. Gate G1 (alive cell population) representing 90% of the total cell population was defined in the forward (FSC) and side (SSC) scatter light plots in the cultures incubated with buffer. Gate M1 was set to include the 15% most strongly FL1 fluorescent cells and Gate M2 the 5% most strongly FL3 fluorescent cells in the buffer condition. The same G1, M1, and M2 gates were used in the analysis of cultures exposed to paraquat and hREG3A.

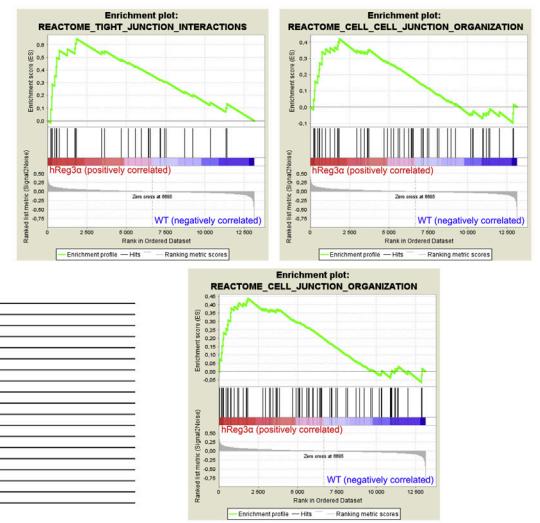
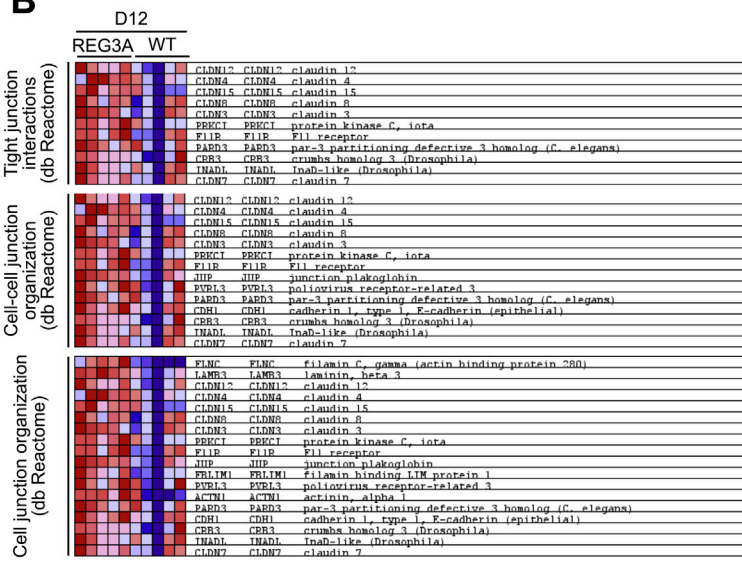


**Supplementary Figure 6.** Low susceptibility to induced colitis of *REG3A* heterozygous mice born to WT mothers compared with WT mice. (A–D) Time evolution of DSS-induced colitis in *REG3A* heterozygous (*REG3A-TG+/-*; n = 15), *REG3A* homozygous (*REG3A-TG+/+*; n = 12) and WT (n = 15) mice. (A) Body weight changes. (B) Rectal bleeding score. (C) Diarrhea score. (D) Kaplan-Meier survival plot. Heavy line: 5-day period of DSS administration. The data are means ± SEM. The Wilcoxon rank-sum test was performed for analysis. \*\**P* < .01, \*\*\**P* < .001.

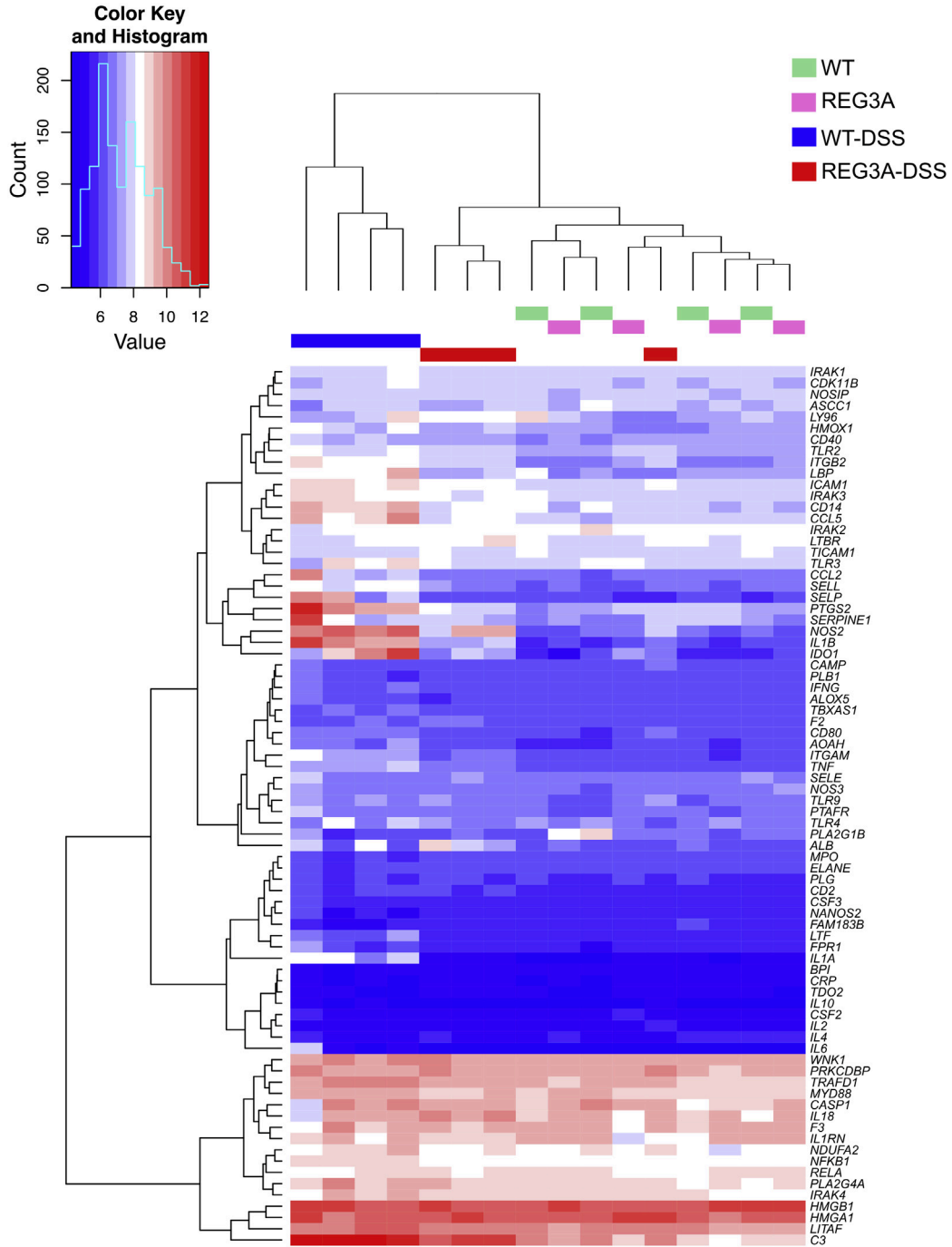
**A**



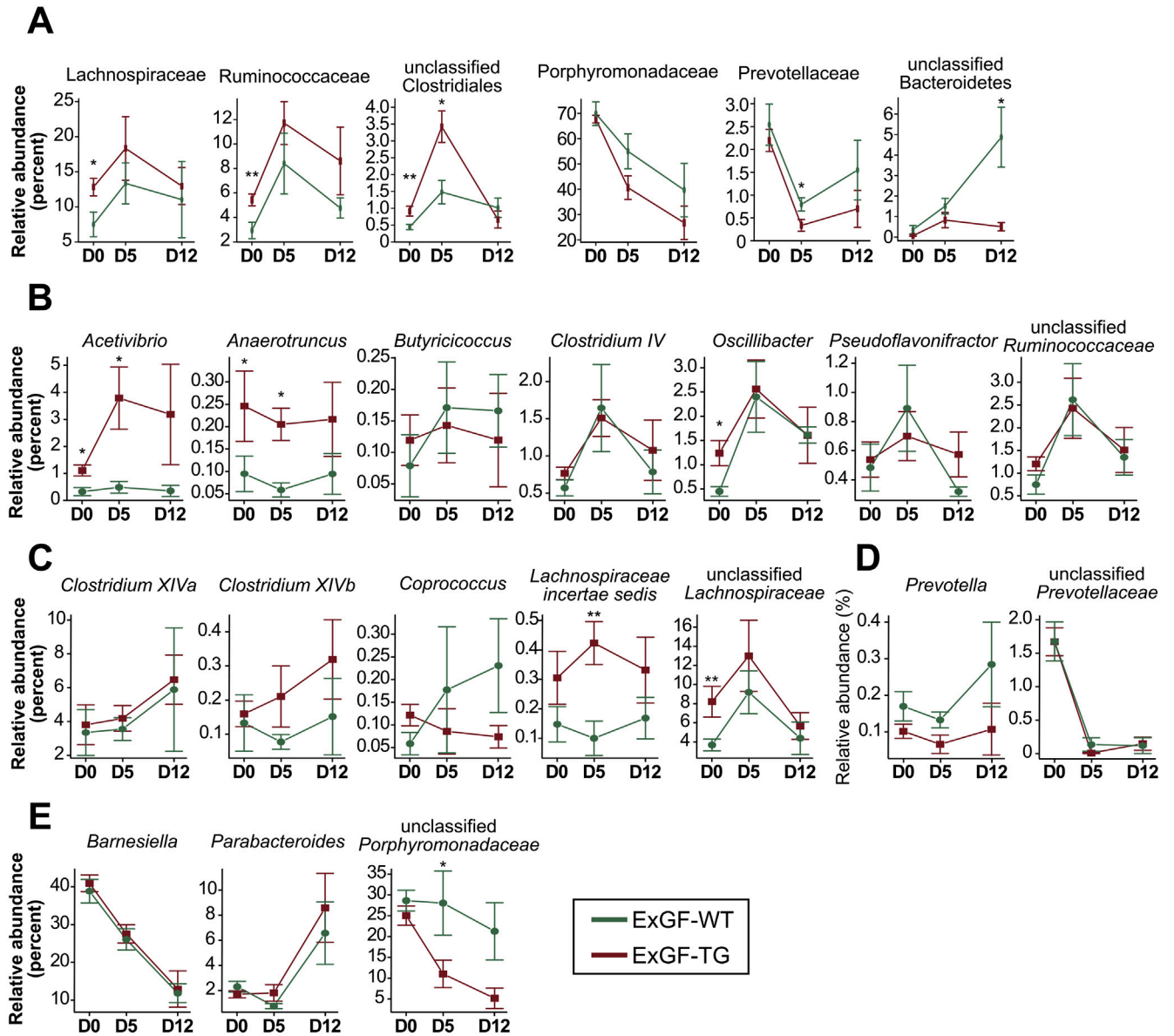
**B**



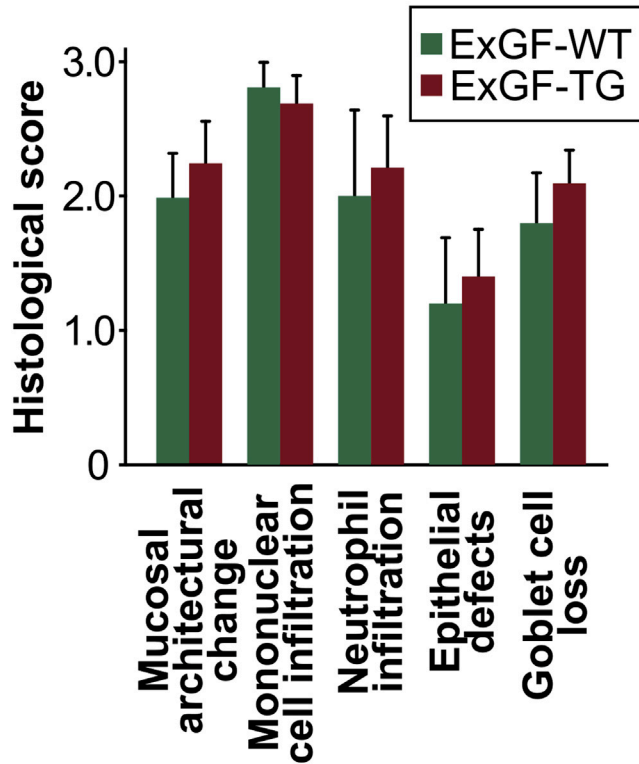
**Supplementary Figure 7.** Transcriptomic profiles of gene sets involved in the regulation of the gut barrier function. (A) and (B) Enrichment plots of genes related to gut barrier function and the corresponding heat maps for REG3A-TG and WT mice. (A) Homeostatic state (D0). (B) At day 12 of DSS-induced colitis (D12).  $P < .001$  for all the indicated pathways (empirical phenotype-based permutation test).



**Supplementary Figure 8.** Attenuation of the LPS activation pathway in the colon of *REG3A*-TG mice exposed to DSS. Hierarchical clustering for genes related to the LPS-induced activation pathway in healthy and DSS-given *REG3A*-transgenic and WT mice (n = 4 per group).

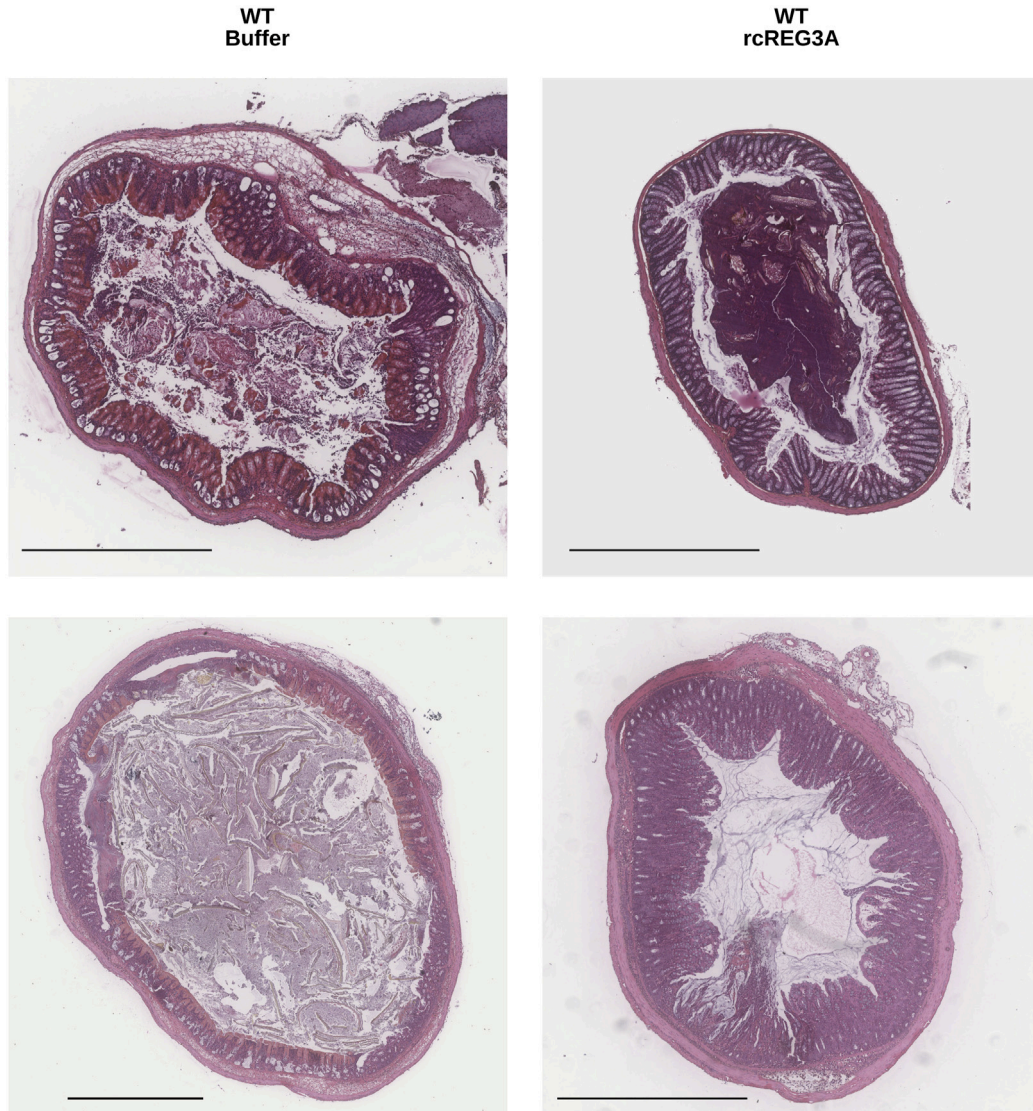


**Supplementary Figure 9.** Time evolution of bacterial communities in colonized germ-free mice during DSS-induced colitis. (A–E) Relative abundances of bacterial communities in germ-free mice colonized with the fecal microbiota from *REG3A*-TG (ExGF-TG; n = 8) and WT mice (ExGF-WT; n = 8) on days 0, 5, and 12 of DSS-induced colitis. (A) Bacterial families. (B–E) Bacterial genera of the following families: (B) Ruminococcaceae. (C) Lachnospiraceae. (D) Prevotellaceae. (E) Porphyromonadaceae. The data are means ± SEM. The 2-sided Wilcoxon rank-sum test was performed for analysis. \**P* < .05, \*\**P* < .01.



**Supplementary Figure 10.** Gut barrier impairment in germ-free mice colonized with fecal microbiota from *REG3A*-transgenic (ExGF-TG;  $n = 8$ ) and WT (ExGF-WT;  $n = 8$ ) mice during DSS-induced colitis. Histological assessment of gut epithelium ( $n = 8$  for each group). The data are means  $\pm$  SEM.  $P =$  nonsignificant (Wilcoxon rank-sum test).





**Supplementary Figure 11.** Representative colon images from WT mice exposed to TNBS and administered 100  $\mu$ g of rcREG3A or an equivalent volume of buffer intrarectally. Staining: hematoxylin and eosin. Top row: full views of the same colon sections as in [Figure 6B](#). Scale bar, 1 mm.

---

## V DISCUSSION GÉNÉRALE

L'intérêt croissant de ces dernières années pour l'exploration du génome non codant, la multiplication d'études multi-OMICS ainsi que le développement de nombreux outils de bio-analyse ont ouvert un immense champ d'investigation dans le domaine de l'oncologie. L'étude des micro-ARN en tant que régulateurs fondamentaux des cellules normales et pathologiques a pris une très grande ampleur et leurs rôles dans la biologie tumorale affecte tous les « Hallmarks » (capacités distinctives) du cancer décrits par Hanahan et Weinberg. Une autre part du génome non codant que sont les éléments transposables et en particulier les rétrovirus endogènes (HERV) humains peuvent également contribuer à la pathogenèse de maladies complexes incluant la carcinogenèse. Ces éléments transposables intégrés dans l'ADN chromosomique à la suite de leurs insertions ancestrales dans la lignée germinale ont pour la grande majorité d'entre eux perdu leur pouvoir de mobilisation par rétrotransposition suite à l'accumulation de mutations, insertions ou délétions qui aboutissent à la formation de pseudoéléments. Néanmoins, certaines séquences rétrovirales endogènes contenues dans les domaines 5' (promoteur et séquences activatrices) ou 3' (signaux de polyadénylation) sont restées fonctionnelles et codent des protéines dont l'expression a été associée à de nombreuses maladies neurodégénératives, auto-immunes ou tumorales. Durant ma thèse, nous avons étudié extensivement certaines altérations des génomes codant et non codant dans plusieurs cohortes d'échantillons de CHC humains. Nous avons établi une carte complète des promoteurs des ARN non codant (ARNnc) transcrits activement dans un ensemble de 50 paires tumeurs/non tumeurs de CHC humains en utilisant la méthode CAGE (Cap Analysis of Gene Expression) qui permet l'identification des sites d'initiation de la transcription (TSS) des ARNnc et des ARN messagers en capturant les extrémités 5' coiffées des ARN. Nous avons révélé une activation forte et dispersée dans tout le génome tumoral de séquences promotrices LTR. Le groupe de CHC qui présentait une expression forte d'un répertoire de 43 LTR particuliers (« LTR-High ») était plutôt des tumeurs viro-induites, peu différenciées et présentaient un fort taux de récurrence tumorale. Sur les 16 tumeurs constituant le groupe « LTR-High », seules 2 tumeurs exprimaient le marqueur

progéniteur KRT19 et une troisième tumeur appartenait au sous-type hypoxique que nous avons défini dans l'étude portant sur les CHC-KRT19 positif.

Nous avons dans ce travail également identifié un grand nombre de transcrits non-codant dont la fonctionnalité reste à étudier. Alors que pour les protéines, l'analyse de la séquence en acides aminés révèle la fonction moléculaire par la nature des domaines présents permettant ainsi de découvrir de nouvelles fonctions, ce type d'analyse n'est pour l'instant pas possible pour les ARN non codant. Il est donc nécessaire de caractériser de manière fonctionnelle ces ARN par des expériences de surexpression ou d'inhibition dans des modèles cellulaires et *in vivo*. Il est aussi essentiel de répertorier ces ARN non codant dans un bien plus grand nombre d'échantillons humains, de lignées cellulaires et, tout comme les projets TCGA ou ENCODE, mettre à la disposition de la communauté scientifique ce répertoire.

Nous avons au cours de cette thèse réalisé un second travail ciblant les CHC KRT19-positif qui sont des tumeurs particulièrement agressives avec l'objectif de mieux cerner la biologie de ces tumeurs et d'identifier des signatures spécifiques de ce sous-groupe phénotypique de CHC, et potentiellement découvrir des cibles thérapeutiques nouvelles. Nous montrons que les CHC KRT19-positif sont des tumeurs fortement glycolytiques en relation probablement avec des événements hypoxiques comme ceux pouvant être induits par des traitements néo-adjuvants comme la chimioembolisation trans-artérielle. Nous pensons que cette observation mériterait d'être complétée par l'étude du micro-environnement tumoral (TME) des CHC KRT19-positif étant donné les liens fonctionnels décrits entre immunité et métabolisme tumoral. Nous pouvons bénéficier de l'expérience acquise précédemment dans les cholangiocarcinomes intrahépatiques (ICC) où nous avons été capables d'établir la première analyse moléculaire complète du TME des ICC basée sur le profil d'expression génique et des marqueurs de type cellulaire immunitaire et stromal dans un grand ensemble de tumeurs ICC (Job\*, Rapoud\* et al. manuscrit en révision). Cette procédure analytique des TME, notamment immunitaire, grâce au logiciel MCP-counter pourrait être appliquée au TME des CHC KRT19-positif. Nous pourrions ainsi corrélérer le phénotype métabolique tumoral et le TME immunitaire, et peut-être ainsi proposer des traitements combinant des immunothérapies anti-checkpoints et des

traitements métaboliques (comme la metformine). Sur le plan clinique, nos données préliminaires suggèrent un impact de la chimioembolisation trans-artérielle sur l'émergence dans un environnement hypoxique/nécrotique de clones tumoraux KRT-19 positif pouvant contribuer à la résistance aux traitements anti-cancéreux. Ces données, si elles se confirmaient, nécessiteraient un traitement métabolique des patients en amont de l'intervention de chimio-embolisation. Pour corroborer notre hypothèse d'événement hypoxique dans la survenue de clones résistants KRT19-positif, il est envisageable de travailler dans des modèles murins de carcinogenèse induite (diéthylnitrosamine, DEN par exemple) en administrant des drogues stabilisant HIF1 $\alpha$  (comme la diméthylloxaloylglycine ; DMOG, en intrapéritonéal) ou des cycles d'ischémie/reperfusion.

L'analyse de la littérature montre qu'il n'y a quasiment pas de travaux démontrant un lien entre l'activation des LTR et la réponse à l'hypoxie dans le cancer. En effet seules deux équipes ont analysé l'expression des LTR dans le carcinome rénal et montré que: (i) HIF1 $\alpha$  se fixait sur des éléments LTR actifs en amont du gène POU5F1 (Oct4), facteur de transcription clé de la pluripotence (Siebenthal *et al.*, 2019); et (ii) l'activation d'un rétrovirus endogène de la famille des HERV-E était capable d'inhiber l'expression du suppresseur de tumeur von Hippel–Lindau (VHL) et donc de stabiliser HIF1 $\alpha$  (Cherkasova *et al.*, 2011). Ces résultats suggèrent que le génome non codant tumoral peut être profondément remanié par un micro-environnement tumoral hypoxique comme celui des CHC KRT19-positif.

Enfin, les classifications moléculaires exposées dans ce manuscrit sont en quelque sorte des condensés analytiques d'études globales ou spécifiques de sous-classes de CHC. Décrivant principalement des altérations moléculaires comme des mutations de gènes, des voies de signalisation enrichies et des corrélations clinico-pathologiques, l'intégration des altérations du génome tumoral non-codant au génome codant identifiées dans nos études et par d'autres équipes permettrait d'accroître notre connaissance de la grande hétérogénéité des CHC et d'affiner les stratifications des patients. L'intérêt de ces classifications est également d'identifier des cibles diagnostiques, pronostiques et thérapeutiques bien que malheureusement dans le cas du CHC les études OMICS n'ont eu jusqu'à présent que peu d'impact clinique. L'une des prochaines étapes pourrait donc être

d'évaluer l'utilité clinique pronostique de certains ARN non codants dans le suivi des patients.

Une troisième étude présentée dans ce mémoire concerne les effets biologiques de la protéine anti-microbienne REG3A sur le microbiote intestinal. Les peptides et protéines anti-microbiens (AMPs) sont des molécules de la première ligne de défense de l'hôte appartenant au système immunitaire inné. Ils comprennent entre autres les défensines, les cathélicidines et les lectines de type 3. Les AMPs déploient, outre une activité antibiotique endogène, de nombreuses autres fonctions immunomodulatrices, chémo-attractantes, mitogènes et anti-apoptotiques, toutes ces fonctions pouvant contribuer à l'homéostasie du microbiote intestinal et au dialogue hôte – microbiote. Les AMPs humains dont l'expression est constitutive ou induite par l'inflammation et les infections sont produits par un nombre limité de types cellulaires, et leur régulation tissu-spécifique est finement contrôlée. Les AMPs sont largement étudiés dans le contexte des maladies inflammatoires chroniques de l'intestin (MICI) bien que leur rôle dans la pathogenèse des MICI soit inconnue. Les MICI sont caractérisées par une fonction barrière altérée, une dysbiose intestinale avec diminution de la richesse et de la diversité microbiennes conduisant à une altération de la réponse immunitaire intestinale vis-à-vis du microbiote intestinal chez des hôtes génétiquement prédisposés. Ces maladies sont en constante progression dans le monde en particulier dans le monde industrialisé avec une dégradation importante au cours du temps de la qualité de vie des patients. Nous montrons que la lectine humaine REG3A connue pour son activité bactéricide vis-à-vis de bactéries Gram positif, était capable de protéger certaines familles de bactéries anaérobies strictes, qu'elles soient Gram positif ou négatif, en conditions de culture basale ou oxydante via son activité anti-oxydante et donc de modifier l'écologie bactérienne intestinale en faveur de bactéries symbionts ayant des activités anti-inflammatoires.

Le microbiote particulier hébergé par des souris transgéniques pour REG3A humain conférait à l'hôte une réelle résistance phénotypique vis-à-vis de processus inflammatoires intestinaux. Ce mécanisme de résistance attribué au microbiote intestinal et mettant en jeu une régulation du potentiel redox intestinal est encore peu connu. Il sera donc nécessaire d'identifier (i) les espèces et souches bactériennes responsables de la résistance

phénotypique observée et (ii) les métabolites secrétés par ces bactéries et d'en étudier le mécanisme moléculaire dans des modèles expérimentaux d'inflammation induite. Il est possible par ailleurs d'étendre le champ d'investigation à d'autres pathologies comme certains cancers qui se développent sur une composante inflammatoire comme le cancer colorectal associé à une colite. En effet, il serait intéressant d'utiliser REG3A comme outil de modulation de l'écologie microbienne intestinale afin d'identifier les bactéries et produits bactériens d'un « bon » microbiote, résilient à la carcinogenèse épithéliale. Comme nous l'avons vu en introduction, le dialogue foie – microbiote est important dans l'homéostasie cellulaire et de l'organisme tout entier. Les patients atteints de CHC souffrent généralement d'une dysbiose intestinale avec une réduction importante de la diversité microbienne. Nous pourrions évaluer l'impact de sous-types bactériens dans la promotion et la progression tumorales au moyen de modèles animaux d'induction avec le DEN par exemple, de comparer ces populations bactériennes avec celles retrouvées chez des patients ayant un CHC et potentiellement s'en servir comme outils diagnostiques précoces chez des patients cirrhotiques à risque de développer un CHC. En parallèle de ces expérimentations, le laboratoire souhaiterait développer une formulation médicamenteuse orale à délivrance entérique composée d'une protéine recombinante REG3A capable de moduler une dysbiose intestinale dans une direction proche de la composition et fonctionnalité du microbiote décrit dans notre modèle transgénique et d'offrir une option thérapeutique alternative de lutte contre les maladies chroniques inflammatoires.

Les avancées récentes du séquençage à haut débit et de spectrométrie de masse permettent d'étudier des systèmes cellulaires complexes. L'intégration de données massives et complexes fournit une excellente plateforme pour établir des réseaux entre les altérations génomiques et épigénomiques à la réponse cellulaire transcriptionnelle, protéique et métabolique. Ces techniques sont de formidables opportunités pour mieux comprendre la biologie cellulaire et tumorale. Malheureusement, ces travaux reflètent la biologie cellulaire à un instant donné de la progression tumorale. La prochaine étape qui doit être réalisée en parallèle du développement de ces méthodologies, est l'élaboration de modèles cellulaires multi-dimensionnels (génome, épigénome, transcrits, protéines, métabolites, microbes) capables de reproduire la dynamique moléculaire décrite par les analyses multi-

OMICS. Néanmoins, le principal défaut de ces bases est l'absence quasi-totale de données clinico-pathologiques. Il est aujourd'hui impératif, que ces données figurent dans ces bases de données. L'objectif sera de concevoir des modèles prédictifs visant à valider les classifications et à proposer une médecine personnalisée aux patients.

---

## VI RÉFÉRENCES BIBLIOGRAPHIQUES

- Abdel-Hady M, Kelly D. Chronic Hepatitis B in Children and Adolescents: Epidemiology and Management. *Pediatr Drugs* 2013; 15: 311–317.
- Adachi Y, Tsuchihashi J, Shiraishi N, Yasuda K, Etoh T, Kitano S. AFP-producing gastric carcinoma: multivariate analysis of prognostic factors in 270 patients. *Oncology* 2003; 65: 95–101.
- Aft RL, Zhang FW, Gius D. Evaluation of 2-deoxy-D-glucose as a chemotherapeutic agent: mechanism of cell death. *Br J Cancer* 2002; 87: 805–812.
- Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level. *JAMA Oncol* 2017; 3: 1683–1691.
- Allain C, Angenard G, Clément B, Coulouarn C. Integrative Genomic Analysis Identifies the Core Transcriptional Hallmarks of Human Hepatocellular Carcinoma. *Cancer Res* 2016; 76: 6374–6381.
- Altekruse SF, McGlynn KA, Reichman ME. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. *J Clin Oncol Off J Am Soc Clin Oncol* 2009; 27: 1485–1491.
- Aravanis AM, Lee M, Klausner RD. Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection. *Cell* 2017; 168: 571–574.
- Armbruster V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, et al. A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res* 2002; 8: 1800–1807.
- Arnaud C, Sebbagh M, Nola S, Audebert S, Bidaut G, Hermant A, et al. MCC, a new interacting protein for Scrib, is required for cell migration in epithelial cells. *FEBS Lett* 2009; 583: 2326–2332.
- Arthur MJ, Stanley A, Iredale JP, Rafferty JA, Hembry RM, Friedman SL. Secretion of 72 kDa type IV collagenase/gelatinase by cultured human lipocytes. Analysis of gene expression, protein synthesis and proteinase activity. *Biochem J* 1992; 287 ( Pt 3): 701–707.
- Ascione T, Di Flumeri G, Boccia G, De Caro F. Infections in patients affected by liver cirrhosis: an update. *Infez Med Riv Period Eziologia Epidemiol Diagn Clin E Ter Delle Patol Infett* 2017; 25: 91–97.
- Baffy G, Brunt EM, Caldwell SH. Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace. *J Hepatol* 2012; 56: 1384–1391.
- Bai X, Liu Z, Shao X, Wang D, Dong E, Wang Y, et al. The heterogeneity of plasma miRNA profiles in hepatocellular carcinoma patients and the exploration of diagnostic circulating miRNAs for hepatocellular carcinoma. *PloS One* 2019; 14: e0211581.
- Banales JM, Iñarrairaegui M, Arbeláiz A, Milkiewicz P, Muntané J, Muñoz-Bellvis L, et al. Serum Metabolites as Diagnostic Biomarkers for Cholangiocarcinoma, Hepatocellular Carcinoma, and Primary Sclerosing Cholangitis. *Hepatol Baltim Md* 2019; 70: 547–562.
- Barthel A, Okino ST, Liao J, Nakatani K, Li J, Whitlock JP, et al. Regulation of GLUT1 gene transcription by the serine/threonine kinase Akt1. *J Biol Chem* 1999; 274: 20281–20286.



Bast RC, Boyer CM, Jacobs I, Xu FJ, Wu S, Wiener J, et al. Cell growth regulation in epithelial ovarian cancer. *Cancer* 1993; 71: 1597–1601.

Battaglia S, Benzoubir N, Nobilet S, Charneau P, Samuel D, Zignego AL, et al. Liver cancer-derived hepatitis C virus core proteins shift TGF-beta responses from tumor suppression to epithelial-mesenchymal transition. *Plos One* 2009; 4: e4355.

Behari J, Zeng G, Otruba W, Thompson M, Muller P, Micsenyi A, et al. R-ETODOLAC DECREASES BETA-CATENIN LEVELS ALONG WITH SURVIVAL AND PROLIFERATION OF HEPATOMA CELLS. *J Hepatol* 2007; 46: 849–857.

Benzoubir N, Mussini C, Lejamtel C, Dos Santos A, Guillaume C, Desterke C, et al. Gamma-smooth muscle actin expression is associated with epithelial-mesenchymal transition and stem-like properties in hepatocellular carcinoma. *Plos One* 2015; 10: e0130559.

Bergström A, Pisani P, Tenet V, Wolk A, Adami HO. Overweight as an avoidable cause of cancer in Europe. *Int J Cancer* 2001; 91: 421–430.

Blanc J-F, Barbare, JC, Baumann, AS, Boige, V, Boudjema K, M. Bouattour, et al. « Carcinome hépatocellulaire » [Internet]. In: *Thésaurus National de Cancérologie Digestive*. 2019[cited 2019 Jun 24] Available from: <http://www.tncd.org>

Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, et al. Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein. *Oncogene* 2000; 19: 4328–4336.

Borkham-Kamphorst E, van Roeyen CRC, Ostendorf T, Floege J, Gressner AM, Weiskirchen R. Pro-fibrogenic potential of PDGF-D in liver fibrosis. *J Hepatol* 2007; 46: 1064–1074.

Bouchard MJ, Wang LH, Schneider RJ. Calcium signaling by HBx protein in hepatitis B virus DNA replication. *Science* 2001; 294: 2376–2378.

Bourette RP, Mouchiroud G, Ouazana R, Morlé F, Godet J, Blanchet JP. Expression of human colony-stimulating factor-1 (CSF-1) receptor in murine pluripotent hematopoietic NFS-60 cells induces long-term proliferation in response to CSF-1 without loss of erythroid differentiation potential. *Blood* 1993; 81: 2511–2520.

Bournat JC, Brown CW. Mitochondrial dysfunction in obesity. *Curr Opin Endocrinol Diabetes Obes* 2010; 17: 446–452.

Bouwmeester T, Bauch A, Ruffner H, Angrand P-O, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* 2004; 6: 97–105.

Boyault S, Rickman DS, de Reyniès A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007; 45: 42–52.

Boyer N, Marcellin P. Traitement de l'hépatite C. *Med Sci* 2002; 18: 343–352.

Bret PM, Labadie M, Bretagnolle M, Paliard P, Fond A, Valette PJ. Hepatocellular carcinoma: diagnosis by percutaneous fine needle biopsy. *Gastrointest Radiol* 1988; 13: 253–255.

Brooks PJ, Theruvathu JA. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. *Alcohol Fayettev N* 2005; 35: 187–193.

Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004; 40: 1387–1395.

Bruix J, Reig M, Sherman M. Evidence-Based Diagnosis, Staging, and Treatment of Patients With Hepatocellular Carcinoma. *Gastroenterology* 2016; 150: 835–853.

Budhu A, Roessler S, Zhao X, Yu Z, Forgues M, Ji J, et al. Integrated Metabolite and Gene Expression Profiles Identify Lipid Biomarkers Associated With Progression of Hepatocellular Carcinoma and Patient Outcomes. *Gastroenterology* 2013; 144: 1066-1075.e1.

Burns KH. Transposable elements in cancer. *Nat Rev Cancer* 2017; 17: 415–424.

Buscarini L, Fornari F, Bolondi L, Colombo P, Livraghi T, Magnolfi F, et al. Ultrasound-guided fine-needle biopsy of focal liver lesions: techniques, diagnostic accuracy and complications. A retrospective study on 2091 biopsies. *J Hepatol* 1990; 11: 344–348.

Caballero F, Fernández A, De Lacy AM, Fernández-Checa JC, Caballería J, García-Ruiz C. Enhanced free cholesterol, SREBP-2 and StAR expression in human NASH. *J Hepatol* 2009; 50: 789–796.

Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* 2003; 348: 1625–1638.

Cao Q, Imbeaud S, Datta S, Zucman-Rossi J. Authors' response: virus-host interactions in HBV-related hepatocellular carcinoma: more to be revealed? *Gut* 2015; 64: 853–854.

Cao Y, Eble JM, Moon E, Yuan H, Weitzel DH, Landon CD, et al. Tumor Cells Upregulate Normoxic HIF-1 in Response to Doxorubicin. *Cancer Res* 2013; 73: 6230–6242.

Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 2006; 313: 1126–1130.

Chan HL-Y, Hui AY, Wong ML, Tse AM-L, Hung LC-T, Wong VW-S, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut* 2004; 53: 1494–1498.

Chandel NS, Maltepe E, Goldwasser E, Mathieu CE, Simon MC, Schumacker PT. Mitochondrial reactive oxygen species trigger hypoxia-induced transcription. *Proc Natl Acad Sci U S A* 1998; 95: 11715–11720.

Chapoutot C, Perney P, Fabre D, Taourel P, Bruel JM, Larrey D, et al. [Needle-tract seeding after ultrasound-guided puncture of hepatocellular carcinoma. A study of 150 patients]. *Gastroenterol Clin Biol* 1999; 23: 552–556.

Chaturvedi VK, Singh A, Dubey SK, Hetta HF, John J, Singh MP. Molecular mechanistic insight of hepatitis B virus mediated hepatocellular carcinoma. *Microb Pathog* 2019; 128: 184–194.

Chen C-J, Yang H-I, Su J, Jen C-L, You S-L, Lu S-N, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *J Am Med Assoc* 2006; 295: 65–73.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; 14: 128.

Chen F, Xue J, Zhou L, Wu S, Chen Z. Identification of serum biomarkers of hepatocarcinoma through liquid chromatography/mass spectrometry-based metabonomic method. *Anal Bioanal Chem* 2011; 401: 1899–1904.

Chen T, Meng Z, Gan Y, Wang X, Xu F, Gu Y, et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating  $\beta$ -catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia* 2013; 27: 1469–1478.

Chen Y-Q, Zhao C-L, Li W. Effect of hypoxia-inducible factor-1 $\alpha$  on transcription of survivin in non-small cell lung cancer. *J Exp Clin Cancer Res* 2009; 28: 29.

Cheng A-L, Kang Y-K, Chen Z, Tsao C-J, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol* 2009; 10: 25–34.

Cheng A-L, Kang Y-K, Lin D-Y, Park J-W, Kudo M, Qin S, et al. Sunitinib versus sorafenib in advanced hepatocellular cancer: results of a randomized phase III trial. *J Clin Oncol Off J Am Soc Clin Oncol* 2013; 31: 4067–4075.

Cheng K, Hao M. Metformin Inhibits TGF- $\beta$ 1-Induced Epithelial-to-Mesenchymal Transition via PKM2 Relative-mTOR/p70s6k Signaling Pathway in Cervical Carcinoma Cells [Internet]. *Int J Mol Sci* 2016; 17[cited 2019 Aug 31] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5187800/>

Cheng S-T, Ren J-H, Cai X-F, Jiang H, Chen J. HBx-elevated SIRT2 promotes HBV replication and hepatocarcinogenesis. *Biochem Biophys Res Commun* 2018; 496: 904–910.

Cherkasova E, Malinzak E, Rao S, Takahashi Y, Senchenko VN, Kudryavtseva AV, et al. Inactivation of the von Hippel-Lindau tumor suppressor leads to selective expression of a human endogenous retrovirus in kidney cancer. *Oncogene* 2011; 30: 4697–4706.

Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res* 2008; 68: 6779–6788.

Chiappini F, Coilly A, Kadar H, Gual P, Tran A, Desterke C, et al. Metabolism dysregulation induces a specific lipid signature of nonalcoholic steatohepatitis in patients. *Sci Rep* 2017; 7: 46658.

Chiaradonna F, Sacco E, Manzoni R, Giorgio M, Vanoni M, Alberghina L. Ras-dependent carbon metabolism and transformation in mouse fibroblasts. *Oncogene* 2006; 25: 5391–5404.

Chiba T, Saito T, Yuki K, Zen Y, Koide S, Kanogawa N, et al. Histone lysine methyltransferase SUV39H1 is a potent target for epigenetic therapy of hepatocellular carcinoma. *Int J Cancer* 2015; 136: 289–298.

Chiba T, Suzuki E, Negishi M, Saraya A, Miyagi S, Konuma T, et al. 3-Deazaneplanocin A is a promising therapeutic agent for the eradication of tumor-initiating hepatocellular carcinoma cells. *Int J Cancer* 2012; 130: 2557–2567.

Chinese Human Proteome Project (CNHPP) Consortium, Jiang Y, Sun A, Zhao Y, Ying W, Sun H, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma [Internet]. *Nature* 2019[cited 2019 Mar 4] Available from: <http://www.nature.com/articles/s41586-019-0987-8>

Cho I-R, Oh M, Koh SS, Malilas W, Srisuttee R, Jhun BH, et al. Hepatitis B virus X protein inhibits extracellular IFN- $\alpha$ -mediated signal transduction by downregulation of type I IFN receptor. *Int J Mol Med* 2012; 29: 581–586.

Chowdhury R, Yeoh KK, Tian Y-M, Hillringhaus L, Bagg EA, Rose NR, et al. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* 2011; 12: 463–469.

Chung Y-G, Tak E, Hwang S, Lee J-Y, Kim J-Y, Kim Y-Y, et al. Synergistic effect of metformin on sorafenib in in vitro study using hepatocellular carcinoma cell lines. *Ann Hepato-Biliary-Pancreat Surg* 2018; 22: 179–184.

Cillo U, Vitale A, Bassanello M, Boccagni P, Brolese A, Zanus G, et al. Liver transplantation for the treatment of moderately or well-differentiated hepatocellular carcinoma. *Ann Surg* 2004; 239: 150–159.

Cirera I, Bauer TM, Navasa M, Vila J, Grande L, Taurá P, et al. Bacterial translocation of enteric organisms in patients with cirrhosis. *J Hepatol* 2001; 34: 32–37.

Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* 2013; 58: 1693–1702.

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.

Cogliati B, Aloia TPA, Bosch RV, Alves VAF, Hernandez-Blazquez FJ, Dagli MLZ. Identification of hepatic stem/progenitor cells in canine hepatocellular and cholangiocellular carcinoma. *Vet Comp Oncol* 2010; 8: 112–121.

Colen CB, Seraji-Bozorgzad N, Marples B, Galloway MP, Sloan AE, Mathupala SP. Metabolic remodeling of malignant gliomas for enhanced sensitization during radiotherapy: an in vitro study. *Neurosurgery* 2006; 59: 1313–1323; discussion 1323-1324.

Colletti M, Cicchini C, Conigliaro A, Santangelo L, Alonzi T, Pasquini E, et al. Convergence of Wnt signaling on the HNF4 $\alpha$ -driven transcription in controlling liver zonation. *Gastroenterology* 2009; 137: 660–672.

Collins P, Jones C, Choudhury S, Damelin L, Hodgson H. Increased expression of uncoupling protein 2 in HepG2 cells attenuates oxidative damage and apoptosis. *Liver Int Off J Int Assoc Study Liver* 2005; 25: 880–887.

Colnot S, Decaens T, Niwa-Kawakita M, Godard C, Hamard G, Kahn A, et al. Liver-targeted disruption of Apc in mice activates  $\beta$ -catenin signaling and leads to hepatocellular carcinomas. *Proc Natl Acad Sci* 2004; 101: 17216–17221.

Cook JA, Gius D, Wink DA, Krishna MC, Russo A, Mitchell JB. Oxidative stress, redox, and the tumor microenvironment. *Semin Radiat Oncol* 2004; 14: 259–266.

Coulouarn C, Cavard C, Rubbia-Brandt L, Audebourg A, Dumont F, Jacques S, et al. Combined hepatocellular-cholangiocarcinomas exhibit progenitor features and activation of Wnt and TGF $\beta$  signaling pathways. *Carcinogenesis* 2012; 33: 1791–1796.

von der Crone S, Deppe C, Barthel A, Sasson S, Joost HG, Schürmann A. Glucose deprivation induces Akt-dependent synthesis and incorporation of GLUT1, but not of GLUT4, into the plasma membrane of 3T3-L1 adipocytes. *Eur J Cell Biol* 2000; 79: 943–949.

Cui M, Xiao Z, Wang Y, Zheng M, Song T, Cai X, et al. Long noncoding RNA HULC modulates abnormal lipid metabolism in hepatoma cells through an miR-9-mediated RXRA signaling pathway. *Cancer Res* 2015; 75: 846–857.

Cui W, Gu F, Hu K-Q. Effects and mechanisms of silibinin on human hepatocellular carcinoma xenografts in nude mice. *World J Gastroenterol* 2009; 15: 1943–1950.

Cusi K. Role of obesity and lipotoxicity in the development of nonalcoholic steatohepatitis: pathophysiology and clinical implications. *Gastroenterology* 2012; 142: 711-725.e6.

Dang CV, Kim J, Gao P, Yustein J. The interplay between MYC and HIF in cancer. *Nat Rev Cancer* 2008; 8: 51–56.

Dapito DH, Mencin A, Gwak G-Y, Pradere J-P, Jang M-K, Mederacke I, et al. Promotion of Hepatocellular Carcinoma by the Intestinal Microbiota and TLR4. *Cancer Cell* 2012; 21: 504–516.

Debebe A, Medina V, Chen C-Y, Mahajan IM, Jia C, Fu D, et al. Wnt/ $\beta$ -catenin activation and macrophage induction during liver cancer development following steatosis. *Oncogene* 2017; 36: 6020–6029.

DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab* 2008; 7: 11–20.

DeGregori J, Leone G, Miron A, Jakoi L, Nevins JR. Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci U S A* 1997; 94: 7245–7250.

DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* 2011; 475: 106–109.

Denko NC. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat Rev Cancer* 2008; 8: 705–713.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012; 22: 1775–1789.

Désert R, Rohart F, Canal F, Sicard M, Desille M, Renaud S, et al. Human hepatocellular carcinomas with a periportal phenotype have the lowest potential for early recurrence after curative resection. *Hepatology* 2017; 66: 1502–1518.

Desjardins P, Du T, Jiang W, Peng L, Butterworth RF. Pathogenesis of hepatic encephalopathy and brain edema in acute liver failure: Role of glutamine redefined. *Neurochem Int* 2012; 60: 690–696.

Ding D, Lou X, Hua D, Yu W, Li L, Wang J, et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet* 2012; 8: e1003065.

Doe MR, Ascano JM, Kaur M, Cole MD. Myc Posttranscriptionally Induces HIF1 Protein and Target Gene Expression in Normal and Cancer Cells. *Cancer Res* 2012; 72: 949–957.

Donadon V, Balbi M, Gherseti M, Grazioli S, Perciaccante A, Della Valentina G, et al. Antidiabetic therapy and increased risk of hepatocellular carcinoma in chronic liver disease. *World J Gastroenterol* 2009; 15: 2506–2511.

Donadon V, Balbi M, Mas MD, Casarin P, Zanette G. Metformin and reduced risk of hepatocellular carcinoma in diabetic patients with chronic liver disease. *Liver Int Off J Int Assoc Study Liver* 2010; 30: 750–758.

Dos Santos A, Court M, Thiers V, Sar S, Guettier C, Samuel D, et al. Identification of cellular targets in human intrahepatic cholangiocarcinoma using laser microdissection and accurate mass and time tag proteomics. *Mol Cell Proteomics* 2010; 9: 1991–2004.

Dos Santos A, Thiers V, Sar S, Derian N, Bensalem N, Yilmaz F, et al. Contribution of laser microdissection-based technology to proteomic analysis in hepatocellular carcinoma developing on cirrhosis. *Proteomics Clin Appl* 2007; 1: 545–554.

Duffy AG, Ulahannan SV, Makorova-Rusher O, Rahma O, Wedemeyer H, Pratt D, et al. Tremelimumab in combination with ablation in patients with advanced hepatocellular carcinoma. *J Hepatol* 2017; 66: 545–551.

Dumortier J, Lombard-Bohas C, Valette PJ, Boillot O, Scoazec JY, Berger F, et al. Needle tract recurrence of hepatocellular carcinoma after liver transplantation. *Gut* 2000; 47: 301.

Durnez A, Verslype C, Nevens F, Fevery J, Aerts R, Pirenne J, et al. The clinicopathological and prognostic relevance of cytokeratin 7 and 19 expression in hepatocellular carcinoma. A possible progenitor cell origin. *Histopathology* 2006; 49: 138–151.

Duvic M, Talpur R, Ni X, Zhang C, Hazarika P, Kelly C, et al. Phase 2 trial of oral vorinostat (suberoylanilide hydroxamic acid, SAHA) for refractory cutaneous T-cell lymphoma (CTCL). *Blood* 2007; 109: 31–39.

Duvoux C, Roudot-Thoraval F, Decaens T, Pessione F, Badran H, Piardi T, et al. Liver transplantation for hepatocellular carcinoma: a model including  $\alpha$ -fetoprotein improves the performance of Milan criteria. *Gastroenterology* 2012; 143: 986-994.e3; quiz e14-15.

Edmondson HA, Steiner PE. Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies. *Cancer* 1954; 7: 462–503.

El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet Lond Engl* 2017; 389: 2492–2502.

El-Mir MY, Badia MD, Luengo N, Monte MJ, Marin JJ. Increased levels of typically fetal bile acid species in patients with hepatocellular carcinoma. *Clin Sci Lond Engl* 1979 2001; 100: 499–508.

Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt stimulates aerobic glycolysis in cancer cells. *Cancer Res* 2004; 64: 3892–3899.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57–74.

Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 1996; 334: 77–81.

Erler JT, Cawthorne CJ, Williams KJ, Koritzinsky M, Wouters BG, Wilson C, et al. Hypoxia-mediated down-regulation of Bid and Bax in tumors occurs via hypoxia-inducible factor 1-dependent and -independent mechanisms and contributes to drug resistance. *Mol Cell Biol* 2004; 24: 2875–2889.

Espina N, Lima V, Lieber CS, Garro AJ. In vitro and in vivo inhibitory effect of ethanol and acetaldehyde on O6-methylguanine transferase. *Carcinogenesis* 1988; 9: 761–766.

Fan H, Lu X, Wang X, Liu Y, Guo B, Zhang Y, et al. Low-dose decitabine-based chemoimmunotherapy for patients with refractory advanced solid tumors: a phase I/II report. *J Immunol Res* 2014; 2014: 371087.

Farges O, Dokmak S. Malignant transformation of liver adenoma: an analysis of the literature. *Dig Surg* 2010; 27: 32–38.

Farges O, Ferreira N, Dokmak S, Belghiti J, Bedossa P, Paradis V. Changing trends in malignant transformation of hepatocellular adenoma. *Gut* 2011; 60: 85–89.

Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019; 144: 1941–1953.

Finn RS, Kang Y-K, Mulcahy M, Polite BN, Lim HY, Walters I, et al. Phase II, open-label study of brivanib as second-line therapy in patients with advanced hepatocellular carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res* 2012; 18: 2090–2098.

Fitian AI, Nelson DR, Liu C, Xu Y, Ararat M, Cabrera R. Integrated metabolomic profiling of hepatocellular carcinoma in hepatitis C cirrhosis through GC/MS and UPLC/MS-MS. *Liver Int Off J Int Assoc Study Liver* 2014; 34: 1428–1444.

Flamm SL. Chronic hepatitis C virus infection. *J Am Med Assoc* 2003; 289: 2413–2417.

Forner A, Da Fonseca LG, Díaz-González Á, Sanduzzi-Zamparelli M, Reig M, Bruix J. Controversies in the management of hepatocellular carcinoma. *JHEP Rep* 2019; 1: 17–29.

Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet Lond Engl* 2012; 379: 1245–1255.

Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* 2014; 46: 558–566.

Fred C, Törnqvist M, Granath F. Evaluation of cancer tests of 1,3-butadiene using internal dose, genotoxic potency, and a multiplicative risk model. *Cancer Res* 2008; 68: 8014–8021.

- Fu X, Gharib SA, Green PS, Aitken ML, Frazer DA, Park DR, et al. Spectral index for assessment of differential protein expression in shotgun proteomics. *J Proteome Res* 2008; 7: 845–854.
- Fujimoto A, Furuta M, Shiraishi Y, Gotoh K, Kawakami Y, Arihiro K, et al. Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. *Nat Commun* 2015; 6: 6120.
- Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012; 44: 760–764.
- Fukuyama R, Nicolaita R, Ng KP, Obusez E, Sanchez J, Kalady M, et al. Mutated in colorectal cancer, a putative tumor suppressor for serrated colorectal cancer, selectively represses beta-catenin-dependent transcription. *Oncogene* 2008; 27: 6044–6055.
- Fung KYC, Cosgrove L, Lockett T, Head R, Topping DL. A review of the potential mechanisms for the lowering of colorectal oncogenesis by butyrate. *Br J Nutr* 2012; 108: 820–831.
- Gaddikeri S, McNeeley MF, Wang CL, Bhargava P, Dighe MK, Yeh MMC, et al. Hepatocellular Carcinoma in the Noncirrhotic Liver. *Am J Roentgenol* 2014; 203: W34–W47.
- Gailhouse L, Liew LC, Yasukawa K, Hatada I, Tanaka Y, Nakagama H, et al. Differentiation Therapy by Epigenetic Reconditioning Exerts Antitumor Effects on Liver Cancer Cells. *Mol Ther J Am Soc Gene Ther* 2018; 26: 1840–1854.
- Gane EJ, Stedman CA, Hyland RH, Ding X, Svarovskaia E, Symonds WT, et al. Nucleotide polymerase inhibitor sofosbuvir plus ribavirin for hepatitis C. *N Engl J Med* 2013; 368: 34–44.
- Gao X, Qu J, Chang X, Lu Y, Bai W, Wang H, et al. Hypomethylation of long interspersed nuclear element-1 promoter is associated with poor outcomes for curative resected hepatocellular carcinoma. *Liver Int* 2014; 34: 136–146.
- Garcia M, de Thé H, Tiollais P, Samarut J, Dejean A. A hepatitis B virus pre-S-retinoic acid receptor beta chimera transforms erythrocytic progenitor cells in vitro. *Proc Natl Acad Sci U S A* 1993; 90: 89–93.
- Garcia-Tsao G, Bosch J. Varices and Variceal Hemorrhage in Cirrhosis: A New View of an Old Problem. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc* 2015; 13: 2109–2117.
- Gil-García B, Baladrón V. The complex role of NOTCH receptors and their ligands in the development of hepatoblastoma, cholangiocarcinoma and hepatocellular carcinoma. *Biol Cell* 2016; 108: 29–40.
- Ginsberg HN. Is the slippery slope from steatosis to steatohepatitis paved with triglyceride or cholesterol? *Cell Metab* 2006; 4: 179–181.
- Goering W, Ribarska T, Schulz WA. Selective changes of retroelement expression in human prostate cancer. *Carcinogenesis* 2011; 32: 1484–1492.
- Goldberg IJ, Ginsberg HN. Ins and outs modulating hepatic triglyceride and development of nonalcoholic fatty liver disease. *Gastroenterology* 2006; 130: 1343–1346.



- Gonzalez-Angulo AM, Morales-Vasquez F, Hortobagyi GN. Overview of resistance to systemic therapy in patients with breast cancer. *Adv Exp Med Biol* 2007; 608: 1–22.
- Gotoh M, Sakamoto M, Kanetaka K, Chuuma M, Hirohashi S. Overexpression of osteopontin in hepatocellular carcinoma. *Pathol Int* 2002; 52: 19–24.
- Gottlob K, Majewski N, Kennedy S, Kandel E, Robey RB, Hay N. Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes Dev* 2001; 15: 1406–1418.
- Gougelet A, Sartor C, Senni N, Calderaro J, Fartoux L, Lequoy M, et al. Hepatocellular Carcinomas With Mutational Activation of Beta-Catenin Require Choline and Can Be Detected by Positron Emission Tomography. *Gastroenterology* 2019: S0016508519409888.
- Govaere O, Komuta M, Berkers J, Spee B, Janssen C, Luca F de, et al. Keratin 19: a key role player in the invasion of human hepatocellular carcinomas. *Gut* 2014; 63: 674–685.
- Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007; 446: 153–158.
- Groussard C, Morel I, Chevanne M, Monnier M, Cillard J, Delamarche A. Free radical scavenging and antioxidant effects of lactate ion: an in vitro study. *J Appl Physiol Bethesda Md* 1985 2000; 89: 169–175.
- Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, 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.
- Guzy RD, Hoyos B, Robin E, Chen H, Liu L, Mansfield KD, et al. Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. *Cell Metab* 2005; 1: 401–408.
- Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* 2009; 10: 32–42.
- Hadad SM, Coates P, Jordan LB, Dowling RJO, Chang MC, Done SJ, et al. Evidence for biological effects of metformin in operable breast cancer: biomarker analysis in a pre-operative window of opportunity randomized trial. *Breast Cancer Res Treat* 2015; 150: 149–155.
- Hama N, Totoki Y, Miura F, Tatsuno K, Saito-Adachi M, Nakamura H, et al. Epigenetic landscape influences the liver cancer genome architecture. *Nat Commun* 2018; 9: 1643.
- Han Y-P, Yan C, Zhou L, Qin L, Tsukamoto H. A matrix metalloproteinase-9 activation cascade by hepatic stellate cells in trans-differentiation in the three-dimensional extracellular matrix. *J Biol Chem* 2007; 282: 12928–12939.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646–674.
- Harada K, Baba Y, Ishimoto T, Chikamoto A, Kosumi K, Hayashi H, et al. LINE-1 methylation level and patient prognosis in a database of 208 hepatocellular carcinomas. *Ann Surg Oncol* 2015; 22: 1280–1287.

- Hassan MM, Curley SA, Li D, Kaseb A, Davila M, Abdalla EK, et al. Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma. *Cancer* 2010; 116: 1938–1946.
- Haybaeck J, Zeller N, Wolf MJ, Weber A, Wagner U, Kurrer MO, et al. A lymphotoxin-driven pathway to hepatocellular carcinoma. *Cancer Cell* 2009; 16: 295–308.
- Hayes CN, Chayama K. MicroRNAs as Biomarkers for Liver Disease and Hepatocellular Carcinoma. *Int J Mol Sci* 2016; 17: 280.
- He S, Lin B, Chu V, Hu Z, Hu X, Xiao J, et al. Repurposing of the antihistamine chlorcyclizine and related compounds for treatment of hepatitis C virus infection. *Sci Transl Med* 2015; 7: 282ra49.
- Hendriks HF, Brouwer A, Knook DL. The role of hepatic fat-storing (stellate) cells in retinoid metabolism. *Hepatology* 1987; 7: 1368–1371.
- Hiraga T, Nakamura H. Imatinib mesylate suppresses bone metastases of breast cancer by inhibiting osteoclasts through the blockade of c-Fms signals. *Int J Cancer* 2009; 124: 215–222.
- Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular analysis of commensal host-microbial relationships in the intestine. *Science* 2001; 291: 881–884.
- Horie Y, Suzuki A, Kataoka E, Sasaki T, Hamada K, Sasaki J, et al. Hepatocyte-specific Pten deficiency results in steatohepatitis and hepatocellular carcinomas. *J Clin Invest* 2004; 113: 1774–1783.
- Horiguchi J, Sherman ML, Sampson-Johannes A, Weber BL, Kufe DW. CSF-1 and C-FMS gene expression in human carcinoma cell lines. *Biochem Biophys Res Commun* 1988; 157: 395–401.
- Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* 2009; 69: 7385–7392.
- Hsiang C-Y, Wu S-L, Chen J-C, Lo H-Y, Li C-C, Chiang S-Y, et al. Acetaldehyde induces matrix metalloproteinase-9 gene expression via nuclear factor-kappaB and activator protein 1 signaling pathways in human hepatocellular carcinoma cells: Association with the invasive potential. *Toxicol Lett* 2007; 171: 78–86.
- Hsieh S-C, Tsai J-P, Yang S-F, Tang M-J, Hsieh Y-H. Metformin inhibits the invasion of human hepatocellular carcinoma cells and enhances the chemosensitivity to sorafenib through a downregulation of the ERK/JNK-mediated NF-κB-dependent pathway that reduces uPA and MMP-9 expression. *Amino Acids* 2014; 46: 2809–2822.
- Hsu F-T, Liu Y-C, Chiang I-T, Liu R-S, Wang H-E, Lin W-J, et al. Sorafenib increases efficacy of vorinostat against human hepatocellular carcinoma through transduction inhibition of vorinostat-induced ERK/NF-κB signaling. *Int J Oncol* 2014; 45: 177–188.
- Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 1991; 350: 427–428.
- Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. *Cell* 2008; 134: 703–707.
- Hu J, Xu Y, Hao J, Wang S, Li C, Meng S. MiR-122 in hepatic function and liver diseases. *Protein Cell* 2012; 3: 364–371.

- Hu T-H, Huang C-C, Lin P-R, Chang H-W, Ger L-P, Lin Y-W, et al. Expression and prognostic role of tumor suppressor gene PTEN/MMAC1/TEP1 in hepatocellular carcinoma. *Cancer* 2003; 97: 1929–1940.
- Hu W, Feng Z, Eveleigh J, Iyer G, Pan J, Amin S, et al. The major lipid peroxidation product, trans-4-hydroxy-2-nonenal, preferentially forms DNA adducts at codon 249 of human p53 gene, a unique mutational hotspot in hepatocellular carcinoma. *Carcinogenesis* 2002; 23: 1781–1789.
- Hu X, Zhao Y, He X, Li J, Wang T, Zhou W, et al. Ciliary neurotrophic factor receptor alpha subunit-modulated multiple downstream signaling pathways in hepatic cancer cell lines and their biological implications. *Hepatology* 2008; 47: 1298–1308.
- Huang X-Y, Ke A-W, Shi G-M, Zhang X, Zhang C, Shi Y-H, et al.  $\alpha$ B-crystallin complexes with 14-3-3 $\zeta$  to induce epithelial-mesenchymal transition and resistance to sorafenib in hepatocellular carcinoma. *Hepatology* 2013; 57: 2235–2247.
- Huntington JT, Tang X, Kent LN, Schmidt CR, Leone G. The Spectrum of E2F in Liver Disease—Mediated Regulation in Biology and Cancer. *J Cell Physiol* 2016; 231: 1438–1449.
- Hutchinson SJ, Bird SM, Goldberg DJ. Influence of alcohol on the progression of hepatitis C virus infection: a meta-analysis. *Clin Gastroenterol Hepatol* 2005; 3: 1150–1159.
- Ide H, Seligson DB, Memarzadeh S, Xin L, Horvath S, Dubey P, et al. Expression of colony-stimulating factor 1 receptor during prostate development and prostate cancer progression. *Proc Natl Acad Sci U S A* 2002; 99: 14404–14409.
- Iizuka N, Sakaida I, Moribe T, Fujita N, Miura T, Stark M, et al. Elevated levels of circulating cell-free DNA in the blood of patients with hepatitis C virus-associated hepatocellular carcinoma. *Anticancer Res* 2006; 26: 4713–4719.
- Ikeda K, Saitoh S, Suzuki Y, Kobayashi M, Tsubota A, Koida I, et al. Disease progression and hepatocellular carcinogenesis in patients with chronic viral hepatitis: a prospective observation of 2215 patients. *J Hepatol* 1998; 28: 930–938.
- Inagaki Y, Okazaki I. Emerging insights into Transforming growth factor beta Smad signal in hepatic fibrogenesis. *Gut* 2007; 56: 284–292.
- Intlekofer AM, Dematteo RG, Venneti S, Finley LWS, Lu C, Judkins AR, et al. Hypoxia Induces Production of L-2-Hydroxyglutarate. *Cell Metab* 2015; 22: 304–311.
- Iwakiri Y. Pathophysiology of Portal Hypertension. *Clin Liver Dis* 2014; 18: 281–291.
- Jéhannin-Ligier K, Dantony E, Bossard N, Molinié F, Defossez G, Daubisse-Marliac L, et al. Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017. *Rapp Tech St-Maurice Santé Publique Fr* 2017: 1–80.
- Jenuwein T. Translating the Histone Code. *Science* 2001; 293: 1074–1080.
- Jiang P, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong VW-S, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 2015; 112: E1317–1325.

Jiang Y, Chen J, Yue C, Zhang H, Tong J, Li J, et al. The Role of miR-182-5p in Hepatocarcinogenesis of Trichloroethylene in Mice. *Toxicol Sci Off J Soc Toxicol* 2017; 156: 208–216.

Johnson PJ, Pirrie SJ, Cox TF, Berhane S, Teng M, Palmer D, et al. The Detection of Hepatocellular Carcinoma Using a Prospectively Developed and Validated Model Based on Serological Biomarkers. *Cancer Epidemiol Prev Biomark* 2014; 23: 144–153.

Jones RG, Thompson CB. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev* 2009; 23: 537–548.

Ju J-H, Yang W, Lee K-M, Oh S, Nam K, Shim S, et al. Regulation of cell proliferation and migration by keratin19-induced nuclear import of early growth response-1 in breast cancer cells. *Clin Cancer Res* 2013; 19: 4335–4346.

Jueliger S, Lyons J, Cannito S, Pata I, Pata P, Shkolnaya M, et al. Efficacy and epigenetic interactions of novel DNA hypomethylating agent guadecitabine (SGI-110) in preclinical models of hepatocellular carcinoma. *Epigenetics* 2016; 11: 709–720.

Jung K, Fleischhacker M, Rabien A. Cell-free DNA in the blood as a solid tumor biomarker--a critical appraisal of the literature. *Clin Chim Acta Int J Clin Chem* 2010; 411: 1611–1624.

Kakiyama G, Hylemon PB, Zhou H, Pandak WM, Heuman DM, Kang DJ, et al. Colonic inflammation and secondary bile acids in alcoholic cirrhosis. *Am J Physiol Gastrointest Liver Physiol* 2014; 306: G929-937.

Karakashev SV, Reginato MJ. Progress toward overcoming hypoxia-induced resistance to solid tumor therapy [Internet]. *Cancer Manag Res* 2015[cited 2019 Aug 28] Available from: <https://www.dovepress.com/progress-toward-overcoming-hypoxia-induced-resistance-to-solid-tumor-t-peer-reviewed-article-CMAR>

Kawada N, Tran-Thi TA, Klein H, Decker K. The contraction of hepatic stellate (Ito) cells stimulated with vasoactive substances. Possible involvement of endothelin 1 and nitric oxide in the regulation of the sinusoidal tonus. *Eur J Biochem FEBS* 1993; 213: 815–823.

Kawai T, Yasuchika K, Ishii T, Katayama H, Yoshitoshi EY, Ogiso S, et al. Keratin 19, a Cancer Stem Cell Marker in Human Hepatocellular Carcinoma. *Clin Cancer Res* 2015; 21: 3081–3091.

Kawauchi K, Araki K, Tobiume K, Tanaka N. p53 regulates glucose metabolism through an IKK-NF-kappaB pathway and inhibits cell transformation. *Nat Cell Biol* 2008; 10: 611–618.

Kim H, Choi GH, Na DC, Ahn EY, Kim GI, Lee JE, et al. Human hepatocellular carcinomas with ‘Stemness’-related marker expression: keratin 19 expression and a poor prognosis. *Hepatology* 2011; 54: 1707–1717.

Kim H, Yu SJ, Yeo I, Cho YY, Lee DH, Cho Y, et al. Prediction of Response to Sorafenib in Hepatocellular Carcinoma: A Putative Marker Panel by Multiple Reaction Monitoring-Mass Spectrometry (MRM-MS). *Mol Cell Proteomics MCP* 2017; 16: 1312–1323.

Kim J, Gao P, Liu Y-C, Semenza GL, Dang CV. Hypoxia-Inducible Factor 1 and Dysregulated c-Myc Cooperatively Induce Vascular Endothelial Growth Factor and Metabolic Switches Hexokinase 2 and Pyruvate Dehydrogenase Kinase 1. *Mol Cell Biol* 2007; 27: 7381–7393.

Kneteman NM, Oberholzer J, Al Saghier M, Meeberg GA, Blitz M, Ma MM, et al. Sirolimus-based immunosuppression for liver transplantation in the presence of extended criteria for hepatocellular carcinoma. *Liver Transplant Off Publ Am Assoc Study Liver Dis Int Liver Transplant Soc* 2004; 10: 1301–1311.

Ko YH, Verhoeven HA, Lee MJ, Corbin DJ, Vogl TJ, Pedersen PL. A translational study ‘case report’ on the small molecule ‘energy blocker’ 3-bromopyruvate (3BP) as a potent anticancer agent: from bench side to bedside. *J Bioenerg Biomembr* 2012; 44: 163–170.

Kobori M, Shinmoto H, Tsushida T, Shinohara K. Phloretin-induced apoptosis in B16 melanoma 4A5 cells by inhibition of glucose transmembrane transport. *Cancer Lett* 1997; 119: 207–212.

Komuta M, Spee B, Vander Borgh S, De Vos R, Verslype C, Aerts R, et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. *Hepatology* 2008; 47: 1544–1556.

Krishnamurthy P, Schuetz JD. The ABC transporter Abcg2/Bcrp: role in hypoxia mediated survival. *Biometals Int J Role Met Ions Biol Biochem Med* 2005; 18: 349–358.

Kruh GD, Belinsky MG. The MRP family of drug efflux pumps. *Oncogene* 2003; 22: 7537–7552.

Kuang Y, El-Khoueiry A, Taverna P, Ljungman M, Neamati N. Guadecitabine (SGI-110) priming sensitizes hepatocellular carcinoma cells to oxaliplatin. *Mol Oncol* 2015; 9: 1799–1814.

Kudo Y, Tanaka Y, Tateishi K, Yamamoto K, Yamamoto S, Mohri D, et al. Altered composition of fatty acids exacerbates hepatotumorigenesis during activation of the phosphatidylinositol 3-kinase pathway. *J Hepatol* 2011; 55: 1400–1408.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016; 44: W90-97.

de La Coste A, Romagnolo B, Billuart P, Renard CA, Buendia MA, Soubrane O, et al. Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc Natl Acad Sci U S A* 1998; 95: 8847–8851.

Lai YS, Thung SN, Gerber MA, Chen ML, Schaffner F. Expression of cytokeratins in normal and diseased livers and in primary liver carcinomas. *Arch Pathol Lab Med* 1989; 113: 134–138.

Lalwani ND, Reddy MK, Qureshi SA, Reddy JK. Development of hepatocellular carcinomas and increased peroxisomal fatty acid beta-oxidation in rats fed [4-chloro-6-(2,3-xylydino)-2-pyrimidinylthio] acetic acid (Wy-14,643) in the semipurified diet. *Carcinogenesis* 1981; 2: 645–650.

Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med* 2010; 16: 571–9, 1p following 579.

Laurent-Puig P, Legoix P, Bluteau O, Belghiti J, Franco D, Binot F, et al. Genetic alterations associated with hepatocellular carcinomas define distinct pathways of hepatocarcinogenesis. *Gastroenterology* 2001; 120: 1763–1773.

Leavens KF, Easton RM, Shulman GI, Previs SF, Birnbaum MJ. Akt2 is required for hepatic lipid accumulation in models of insulin resistance. *Cell Metab* 2009; 10: 405–418.

- Lee J-S, Chu I-S, Heo J, Calvisi DF, Sun Z, Roskams T, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 2004; 40: 667–676.
- Lee J-S, Heo J, Libbrecht L, Chu I-S, Kaposi-Novak P, Calvisi DF, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med* 2006; 12: 410–416.
- Lee Y, Hirose H, Ohneda M, Johnson JH, McGarry JD, Unger RH. Beta-cell lipotoxicity in the pathogenesis of non-insulin-dependent diabetes mellitus of obese rats: impairment in adipocyte-beta-cell relationships. *Proc Natl Acad Sci U S A* 1994; 91: 10878–10882.
- Lee Y-H, Seo D, Choi K-J, Andersen JB, Won M-A, Kitade M, et al. Antitumor effects in hepatocarcinoma of isoform-selective inhibition of HDAC2. *Cancer Res* 2014; 74: 4752–4761.
- Leonel C, Ferreira LC, Borin TF, Moschetta MG, Freitas GS, Haddad MR, et al. Inhibition of Epithelial-mesenchymal Transition in Response to Treatment with Metformin and Y27632 in Breast Cancer Cell Lines [Internet]. *Anticancer Agents Med Chem* 2017[cited 2019 Aug 31] Available from: <http://www.eurekaselect.com/148895/article>
- Lewis CE, Pollard JW. Distinct Role of Macrophages in Different Tumor Microenvironments. *Cancer Res* 2006; 66: 605–612.
- Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 2005; 102: 11070–11075.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006; 444: 1022–1023.
- Li J, Sung CYJ, Lee N, Ni Y, Pihlajamäki J, Panagiotou G, et al. Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. *Proc Natl Acad Sci* 2016; 113: E1306–E1315.
- Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, et al. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* 2011; 43: 828–829.
- Li S-P, Xu H-X, Yu Y, He J-D, Wang Z, Xu Y-J, et al. LncRNA HULC enhances epithelial-mesenchymal transition to promote tumorigenesis and metastasis of hepatocellular carcinoma via the miR-200a-3p/ZEB1 signaling pathway. *Oncotarget* 2016; 7: 42431–42446.
- Li Z-Y, Xi Y, Zhu W-N, Zeng C, Zhang Z-Q, Guo Z-C, et al. Positive regulation of hepatic miR-122 expression by HNF4 $\alpha$ . *J Hepatol* 2011; 55: 602–611.
- Liang H-W, Wang N, Wang Y, Wang F, Fu Z, Yan X, et al. Hepatitis B virus-human chimeric transcript HBx-LINE1 promotes hepatic injury via sequestering cellular microRNA-122. *J Hepatol* 2016; 64: 278–291.
- Liang Y, Zheng T, Song R, Wang J, Yin D, Wang L, et al. Hypoxia-mediated sorafenib resistance can be overcome by EF24 through Von Hippel-Lindau tumor suppressor-dependent HIF-1 $\alpha$  inhibition in hepatocellular carcinoma [Internet]. *Hepatology* 2013[cited 2019 Aug 28]
- Liao B, Zhang Y, Sun Q, Jiang P. Vorinostat enhances the anticancer effect of oxaliplatin on hepatocellular carcinoma cells. *Cancer Med* 2017; 7: 196–207.

- Liao W, Mao Y, Ge P, Yang H, Xu H, Lu X, et al. Value of quantitative and qualitative analyses of circulating cell-free DNA as diagnostic tools for hepatocellular carcinoma: a meta-analysis. *Medicine (Baltimore)* 2015; 94: e722.
- Ling S, Song L, Fan N, Feng T, Liu L, Yang X, et al. Combination of metformin and sorafenib suppresses proliferation and induces autophagy of hepatocellular carcinoma via targeting the mTOR pathway. *Int J Oncol* 2017; 50: 297–309.
- Liu M, Chen L, Ma N-F, Chow RKK, Li Y, Song Y, et al. CHD1L promotes lineage reversion of hepatocellular carcinoma through opening chromatin for key developmental transcription factors. *Hepatology* 2016; 63: 1544–1559.
- Livraghi T, Meloni F, Di Stasi M, Rolle E, Solbiati L, Tinelli C, et al. Sustained complete response and complications rates after radiofrequency ablation of very early hepatocellular carcinoma in cirrhosis: Is resection still the treatment of choice? *Hepatology* 2008; 47: 82–89.
- Llovet JM, Fuster J, Bruix J. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. *Hepatology* 1999; 30: 1434–1440.
- Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J-F, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008; 359: 378–390.
- Long YC, Zierath JR. AMP-activated protein kinase signaling in metabolic regulation. *J Clin Invest* 2006; 116: 1776–1783.
- Losman J-A, Looper R, Koivunen P, Lee S, Schneider RK, McMahon C, et al. (R)-2-Hydroxyglutarate is Sufficient to Promote Leukemogenesis and its Effects are Reversible [Internet]. *Science* 2013; 339[cited 2019 Sep 2] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836459/>
- Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014; 12: 661–672.
- Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* 2014; 21: 423–425.
- Lu Z, Xiao Z, Liu F, Cui M, Li W, Yang Z, et al. Long non-coding RNA HULC promotes tumor angiogenesis in liver cancer by up-regulating sphingosine kinase 1 (SPHK1). *Oncotarget* 2016; 7: 241–254.
- Mamane Y, Petroulakis E, LeBacquer O, Sonenberg N. mTOR, translation initiation and cancer. *Oncogene* 2006; 25: 6416–6422.
- Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 2009; 137: 110–118.
- Maschek G, Savaraj N, Priebe W, Braunschweiger P, Hamilton K, Tidmarsh GF, et al. 2-deoxy-D-glucose increases the efficacy of adriamycin and paclitaxel in human osteosarcoma and non-small cell lung cancers in vivo. *Cancer Res* 2004; 64: 31–34.

- Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* 1996; 334: 693–699.
- Meijer TWH, Kaanders JHAM, Span PN, Bussink J. Targeting Hypoxia, HIF-1, and Tumor Glucose Metabolism to Improve Radiotherapy Efficacy. *Clin Cancer Res* 2012; 18: 5585–5594.
- Melegari M, Wolf SK, Schneider RJ. Hepatitis B virus DNA replication is coordinated by core protein serine phosphorylation and HBx expression. *J Virol* 2005; 79: 9810–9820.
- Méndez-Blanco C, Fondevila F, García-Palomo A, González-Gallego J, Mauriz JL. Sorafenib resistance in hepatocarcinoma: role of hypoxia-inducible factors. *Exp Mol Med* 2018; 50: 134.
- Midorikawa Y, Makuuchi M, Tang W, Aburatani H. Microarray-based analysis for hepatocellular carcinoma: from gene expression profiling to new challenges. *World J Gastroenterol* 2007; 13: 1487–1492.
- Milosevic I, Vujovic A, Barac A, Djelic M, Korac M, Radovanovic Spurnic A, et al. Gut-Liver Axis, Gut Microbiota, and Its Modulation in the Management of Liver Diseases: A Review of the Literature. *Int J Mol Sci* 2019; 20: 395.
- Miltiados O, Sia D, Hoshida Y, Fiel MI, Harrington AN, Thung SN, et al. Progenitor cell markers predict outcome of patients with hepatocellular carcinoma beyond Milan criteria undergoing liver transplantation. *J Hepatol* 2015; 63: 1368–1377.
- Min H-K, Kapoor A, Fuchs M, Mirshahi F, Zhou H, Maher J, et al. Increased hepatic synthesis and dysregulation of cholesterol metabolism is associated with the severity of nonalcoholic fatty liver disease. *Cell Metab* 2012; 15: 665–674.
- Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. *J Clin Gastroenterol* 2013; 47 Suppl: S2-6.
- Miyamoto S, Murphy AN, Brown JH. Akt mediates mitochondrial protection in cardiomyocytes through phosphorylation of mitochondrial hexokinase-II. *Cell Death Differ* 2008; 15: 521–529.
- Møller H, Møllemaard A, Lindvig K, Olsen JH. Obesity and cancer risk: a Danish record-linkage study. *Eur J Cancer Oxf Engl* 1990 1994; 30A: 344–350.
- Monga SP.  $\beta$ -Catenin Signaling and Roles in Liver Homeostasis, Injury, and Tumorigenesis. *Gastroenterology* 2015; 148: 1294–1310.
- Moniaux N, Darnaud M, Garbin K, Dos Santos A, Guettier C, Samuel D, et al. The Reg3 $\alpha$  (HIP/PAP) Lectin Suppresses Extracellular Oxidative Stress in a Murine Model of Acute Liver Failure. *Plos One* 2015; 10: e0125584.
- Moniaux N, Song H, Darnaud M, Garbin K, Gigou M, Mitchell C, et al. Human hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein cures fas-induced acute liver failure in mice by attenuating free-radical damage in injured livers. *Hepatology* 2011; 53: 618–627.
- Montasser MF, Shaker MK, Albreedy AM, Montasser IF, El Dorry A. Risk factors for early intrahepatic distant recurrence after radiofrequency ablation for hepatocellular carcinoma in Egyptian patients. *J Dig Dis* 2014; 15: 676–683.



Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. *Curr Med Chem* 2013; 20: 257–271.

Montgomery MK, Turner N. Mitochondrial dysfunction and insulin resistance: an update. *Endocr Connect* 2015; 4: R1–R15.

Moore KP, Aithal GP. Guidelines on the management of ascites in cirrhosis. *Gut* 2006; 55 Suppl 6: vi1-12.

Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003; 34: 267–273.

Morofuji N, Ojima H, Hiraoka N, Okusaka T, Esaki M, Nara S, et al. Antibody-based proteomics to identify an apoptosis signature for early recurrence of hepatocellular carcinoma. *Clin Proteomics* 2016; 13: 28.

Nabi K, Le A. The Intratumoral Heterogeneity of Cancer Metabolism. *Adv Exp Med Biol* 2018; 1063: 131–145.

Nahm JH, Rhee H, Kim H, Yoo JE, Lee JS, Jeon Y, et al. Increased expression of stemness markers and altered tumor stroma in hepatocellular carcinoma under TACE-induced hypoxia: A biopsy and resection matched study. *Oncotarget* 2017; 8: 99359–99371.

Nault J-C, De Reyniès A, Villanueva A, Calderaro J, Rebouissou S, Couchy G, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology* 2013; 145: 176–187.

Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun* 2013; 4: 2218.

Nault J-C, Martin Y, Caruso S, Hirsch TZ, Bayard Q, Calderaro J, et al. Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatol Baltim Md* 2019

NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *The Lancet* 2016; 387: 1377–1396.

Neuveut C, Wei Y, Buendia MA. Mechanisms of HBV-related hepatocarcinogenesis. *J Hepatol* 2010; 52: 594–604.

Ng CKY, Di Costanzo GG, Tosti N, Paradiso V, Coto-Llerena M, Roscigno G, et al. Genetic profiling using plasma-derived cell-free DNA in therapy-naïve hepatocellular carcinoma patients: a pilot study. *Ann Oncol Off J Eur Soc Med Oncol* 2018; 29: 1286–1291.

Niraula S, Dowling RJO, Ennis M, Chang MC, Done SJ, Hood N, et al. Metformin in early breast cancer: a prospective window of opportunity neoadjuvant study. *Breast Cancer Res Treat* 2012; 135: 821–830.

Nishihara Y, Aishima S, Kuroda Y, Iguchi T, Taguchi K, Asayama Y, et al. Biliary phenotype of hepatocellular carcinoma after preoperative transcatheter arterial chemoembolization. *J Gastroenterol Hepatol* 2008; 23: 1860–1868.

- Nolsøe C, Nielsen L, Torp-Pedersen S, Holm HH. Major complications and deaths due to interventional ultrasonography: a review of 8000 cases. *J Clin Ultrasound JCU* 1990; 18: 179–184.
- Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci U S A* 2014; 111: 12426–12431.
- Okochi O, Hibi K, Sakai M, Inoue S, Takeda S, Kaneko T, et al. Methylation-mediated silencing of SOCS-1 gene in hepatocellular carcinoma derived from cirrhosis. *Clin Cancer Res Off J Am Assoc Cancer Res* 2003; 9: 5295–5298.
- Oldham WM, Clish CB, Yang Y, Loscalzo J. Hypoxia-Mediated Increases in L-2-hydroxyglutarate Coordinate the Metabolic Response to Reductive Stress. *Cell Metab* 2015; 22: 291–303.
- Orimo T, Ojima H, Hiraoka N, Saito S, Kosuge T, Kakisaka T, et al. Proteomic profiling reveals the prognostic value of adenomatous polyposis coli-end-binding protein 1 in hepatocellular carcinoma. *Hepatology* 2008; 48: 1851–1863.
- Paradis V, Dargere D, Bonvoust F, Vidaud M, Segarini P, Bedossa P. Effects and regulation of connective tissue growth factor on hepatic stellate cells. *Lab Invest* 2002; 82: 767–774.
- Paradis V, Youssef N, Dargère D, Bâ N, Bonvoust F, Deschatrette J, et al. Replicative senescence in normal liver, chronic hepatitis C, and hepatocellular carcinomas. *Hum Pathol* 2001; 32: 327–332.
- Paradis V, Zalinski S, Chelbi E, Guedj N, Degos F, Vilgrain V, et al. Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: a pathological analysis. *Hepatology* 2009; 49: 851–859.
- Parikh RM, Mohan V. Changing definitions of metabolic syndrome. *Indian J Endocrinol Metab* 2012; 16: 7–12.
- Park EJ, Lee JH, Yu G-Y, He G, Ali SR, Holzer RG, et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell* 2010; 140: 197–208.
- Park J-W, Finn RS, Kim JS, Karwal M, Li RK, Ismail F, et al. Phase II, open-label study of brivanib as first-line therapy in patients with advanced hepatocellular carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res* 2011; 17: 1973–1983.
- Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006; 118: 3030–3044.
- Paterlini-Bréchet P, Saigo K, Murakami Y, Chami M, Gozuacik D, Mugnier C, et al. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* 2003; 22: 3911–3916.
- Patsialou A, Wyckoff J, Wang Y, Goswami S, Stanley ER, Condeelis JS. Invasion of human breast cancer cells in vivo requires both paracrine and autocrine loops involving the colony-stimulating factor-1 receptor. *Cancer Res* 2009; 69: 9498–9506.
- Patterson AD, Maurhofer O, Beyoglu D, Lanz C, Krausz KW, Pabst T, et al. Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer Res* 2011; 71: 6590–6600.

Pavio N, Battaglia S, Boucreux D, Arnulf B, Sobesky R, Hermine O, et al. Hepatitis C virus core variants isolated from liver tumor but not from adjacent non-tumor tissue interact with Smad3 and inhibit the TGF-beta pathway. *Oncogene* 2005; 24: 6119–6132.

Peng C, Ye Y, Wang Z, Guan L, Bao S, Li B, et al. Circulating microRNAs for the diagnosis of hepatocellular carcinoma. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver* 2019; 51: 621–631.

Petrillo M, Patella F, Pesapane F, Suter MB, Ierardi AM, Angileri SA, et al. Hypoxia and tumor angiogenesis in the era of hepatocellular carcinoma transarterial loco-regional treatments. *Future Oncol Lond Engl* 2018; 14: 2957–2967.

Petrillo M, Patella F, Pesapane F, Suter MB, Ierardi AM, Angileri SA, et al. Hypoxia and tumor angiogenesis in the era of hepatocellular carcinoma transarterial loco-regional treatments. *Future Oncol* 2018; 14: 2957–2967.

Petroulakis E, Mamane Y, Le Bacquer O, Shahbazian D, Sonenberg N. mTOR signaling: implications for cancer and anticancer therapy. *Br J Cancer* 2006; 94: 195–199.

Piguet A-C, Semela D, Keogh A, Wilkens L, Stroka D, Stoupis C, et al. Inhibition of mTOR in combination with doxorubicin in an experimental model of hepatocellular carcinoma. *J Hepatol* 2008; 49: 78–87.

Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, et al. NF-kappaB functions as a tumour promoter in inflammation-associated cancer. *Nature* 2004; 431: 461–466.

Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol Cancer* 2017; 16: 41.

Puri P, Baillie RA, Wiest MM, Mirshahi F, Choudhury J, Cheung O, et al. A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology* 2007; 46: 1081–1090.

Purohit V, Rapaka R, Kwon OS, Song BJ. Roles of alcohol and tobacco exposure in the development of hepatocellular carcinoma. *Life Sci* 2013; 92: 3–9.

Raedsch R, Lauterburg BH, Hofmann AF. Altered bile acid metabolism in primary biliary cirrhosis. *Dig Dis Sci* 1981; 26: 394–401.

Ramachandran A, Snehalatha C. Rising burden of obesity in Asia. [Internet]. *J Obes* 2010; 2010[cited 2016 Sep 5] Available from: <http://dx.doi.org/10.1155/2010/868573>

Rao S, Zaidi S, Banerjee J, Jogunoori W, Sebastian R, Mishra B, et al. Transforming growth factor- $\beta$  in liver cancer stem cells and regeneration. *Hepatology* 2017; 1: 477–493.

Ren N, Qin L-X, Tu H, Liu Y-K, Zhang B-H, Tang Z-Y. The prognostic value of circulating plasma DNA level and its allelic imbalance on chromosome 8p in patients with hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2006; 132: 399–407.

Rhee H, Kim H-Y, Choi J-H, Woo HG, Yoo JE, Nahm JH, et al. Keratin 19 Expression in Hepatocellular Carcinoma Is Regulated by Fibroblast-Derived HGF via a MET-ERK1/2-AP1 and SP1 Axis. *Cancer Res* 2018; 78: 1619–1631.

- Ridlon JM, Kang D-J, Hylemon PB, Bajaj JS. Gut microbiota, cirrhosis, and alcohol regulate bile acid metabolism in the gut. *Dig Dis Basel Switz* 2015; 33: 338–345.
- Rocha M, Avenaoud P, Ménard A, Le Bail B, Balabaud C, Bioulac-Sage P, et al. Association of *Helicobacter* species with hepatitis C cirrhosis with or without hepatocellular carcinoma. *Gut* 2005; 54: 396–401.
- Ronot M, Vilgrain V. Hepatocellular carcinoma: diagnostic criteria by imaging techniques. *Best Pract Res Clin Gastroenterol* 2014; 28: 795–812.
- Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol* 2007; 1: 19–25.
- Roskams T, Katoonizadeh A, Komuta M. Hepatic progenitor cells: an update. *Clin Liver Dis* 2010; 14: 705–718.
- Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Annu Rev Biochem* 2001; 70: 81–120.
- Rubbia-Brandt L, Leandro G, Spahr L, Giostra E, Quadri R, Malé PJ, et al. Liver steatosis in chronic hepatitis C: a morphological sign suggesting infection with HCV genotype 3. *Histopathology* 2001; 39: 119–124.
- Rubbia-Brandt L, Quadri R, Abid K, Giostra E, Malé PJ, Mentha G, et al. Hepatocyte steatosis is a cytopathic effect of hepatitis C virus genotype 3. *J Hepatol* 2000; 33: 106–115.
- Runyon BA, AASLD. Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with ascites due to cirrhosis 2012. *Hepatol Baltim Md* 2013; 57: 1651–1653.
- Safaei A, Arefi Oskouie A, Mohebbi SR, Rezaei-Tavirani M, Mahboubi M, Peyvandi M, et al. Metabolomic analysis of human cirrhosis, hepatocellular carcinoma, non-alcoholic fatty liver disease and non-alcoholic steatohepatitis diseases. *Gastroenterol Hepatol Bed Bench* 2016; 9: 158–173.
- Saha SK, Choi HY, Kim BW, Dayem AA, Yang G-M, Kim KS, et al. KRT19 directly interacts with  $\beta$ -catenin/RAC1 complex to regulate NUMB-dependent NOTCH signaling pathway and breast cancer properties. *Oncogene* 2016; 36: 332–349.
- Saha SK, Parachoniak CA, Ghanta KS, Fitamant J, Ross KN, Najem MS, et al. Mutant IDH inhibits HNF-4 $\alpha$  to block hepatocyte differentiation and promote biliary cancer. *Nature* 2014; 513: 110–114.
- Sahin F, Kannangai R, Adegbola O, Wang J, Su G, Torbenson M. mTOR and P70 S6 kinase expression in primary liver neoplasms. *Clin Cancer Res Off J Am Assoc Cancer Res* 2004; 10: 8421–8425.
- Salloum R, Hummel TR, Kumar SS, Dorris K, Li S, Lin T, et al. A molecular biology and phase II study of imetelstat (GRN163L) in children with recurrent or refractory central nervous system malignancies: a pediatric brain tumor consortium study. *J Neurooncol* 2016; 129: 443–451.
- Santoni-Rugiu E, Jensen MR, Thorgeirsson SS. Disruption of the pRb/E2F pathway and inhibition of apoptosis are major oncogenic events in liver constitutively expressing c-myc and transforming growth factor alpha. *Cancer Res* 1998; 58: 123–134.

- Sato T, Issa J-PJ, Kropf P. DNA hypomethylating drugs in cancer therapy. *Cold Spring Harb Perspect Med* 2017; 7: a026948.
- Sato Y, Sekine T, Ohwada S. Alpha-fetoprotein-producing rectal cancer: calculated tumor marker doubling time. *J Surg Oncol* 1994; 55: 265–268.
- Sattler UGA, Mueller-Klieser W. The anti-oxidant capacity of tumour glycolysis. *Int J Radiat Biol* 2009; 85: 963–971.
- Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015; 47: 505–511.
- Schuppan D, Afdhal NH. Liver cirrhosis. *The Lancet* 2008; 371: 838–851.
- Schwartz DL, Bankson J, Bidaut L, He Y, Williams R, Lemos R, et al. HIF-1-Dependent Stromal Adaptation to Ischemia Mediates In Vivo Tumor Radiation Resistance. *Mol Cancer Res* 2011; 9: 259–270.
- Scoccianti C, Cecchini M, Anderson AS, Berrino F, Boutron-Ruault M-C, Espina C, et al. European Code against Cancer 4th Edition: Alcohol drinking and cancer. *Cancer Epidemiol* 2016; 45: 181–188.
- Semela D, Piguet A-C, Kolev M, Schmitter K, Hlushchuk R, Djonov V, et al. Vascular remodeling and antitumoral effects of mTOR inhibition in a rat model of hepatocellular carcinoma. *J Hepatol* 2007; 46: 840–848.
- Semenza GL, Roth PH, Fang HM, Wang GL. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J Biol Chem* 1994; 269: 23757–23763.
- Sengupta S, Parikh ND. Biomarker development for hepatocellular carcinoma early detection: current and future perspectives. *Hepatic Oncol* 2017; 4: 111–122.
- Senni N, Savall M, Cabrerizo Granados D, Alves-Guerra M-C, Sartor C, Lagoutte I, et al.  $\beta$ -catenin-activated hepatocellular carcinomas are addicted to fatty acids. *Gut* 2019; 68: 322–334.
- Shanbhogue AK, Prasad SR, Takahashi N, Vikram R, Sahani DV. Recent advances in cytogenetics and molecular biology of adult hepatocellular tumors: implications for imaging and management. *Radiology* 2011; 258: 673–693.
- Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, et al. Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *EBioMedicine* 2018
- Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 2013; 153: 101–111.
- Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* 2017; 153: 812–826.
- Siebenthal KT, Miller CP, Vierstra JD, Mathieu J, Tretiakova M, Reynolds A, et al. Integrated epigenomic profiling reveals endogenous retrovirus reactivation in renal cell carcinoma. *EBioMedicine* 2019; 41: 427–442.

- Sigglekow ND, Pangon L, Brummer T, Molloy M, Hawkins NJ, Ward RL, et al. Mutated in colorectal cancer protein modulates the NF $\kappa$ B pathway. *Anticancer Res* 2012; 32: 73–79.
- Singh S, Kaye S, Francis N, Peston D, Gore M, McClure M, et al. Human endogenous retrovirus K (HERV-K) rec mRNA is expressed in primary melanoma but not in benign naevi or normal skin. *Pigment Cell Melanoma Res* 2013; 26: 426–428.
- Si-Tayeb K, Lemaigre FP, Duncan SA. Organogenesis and Development of the Liver. *Dev Cell* 2010; 18: 175–189.
- Smit AF. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res* 1993; 21: 1863–1872.
- Soga T, Sugimoto M, Honma M, Mori M, Igarashi K, Kashikura K, et al. Serum metabolomics reveals  $\gamma$ -glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J Hepatol* 2011; 55: 896–905.
- Sonveaux P, Végran F, Schroeder T, Wergin MC, Verrax J, Rabbani ZN, et al. Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J Clin Invest* 2008; 118: 3930–3942.
- Subramaniam D, Thombre R, Dhar A, Anant S. DNA methyltransferases: a novel target for prevention and therapy. *Front Oncol* 2014; 4: 80.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102: 15545–15550.
- Sun T, Tang Y, Sun D, Bu Q, Li P. Osteopontin versus alpha-fetoprotein as a diagnostic marker for hepatocellular carcinoma: a meta-analysis. *OncoTargets Ther* 2018; 11: 8925–8935.
- Sung W-K, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012; 44: 765–769.
- Szabo G, Bala S. MicroRNAs in liver disease. *Nat Rev Gastroenterol Hepatol* 2013; 10: 542–552.
- Takamori R, Wong LL, Dang C, Wong L. Needle-tract implantation from hepatocellular cancer: is needle biopsy of the liver always necessary? *Liver Transplant Off Publ Am Assoc Study Liver Dis Int Liver Transplant Soc* 2000; 6: 67–72.
- Tang LSY, Covert E, Wilson E, Kottlilil S. Chronic Hepatitis B Infection: A Review. *JAMA* 2018; 319: 1802–1813.
- Toffanin S, Hoshida Y, Lachenmayer A, Villanueva A, Cabellos L, Minguez B, et al. MicroRNA-based classification of hepatocellular carcinoma and oncogenic role of miR-517a. *Gastroenterology* 2011; 140: 1618-1628.e16.
- Tokuhisa Y, Iizuka N, Sakaida I, Moribe T, Fujita N, Miura T, et al. Circulating cell-free DNA as a predictive marker for distant metastasis of hepatitis C virus-related hepatocellular carcinoma. *Br J Cancer* 2007; 97: 1399–1403.
- Torre C, Perret C, Colnot S. Transcription dynamics in a physiological process:  $\beta$ -catenin signaling directs liver metabolic zonation. *Int J Biochem Cell Biol* 2011; 43: 271–278.

- Toso C, Meeberg G, Hernandez-Alejandro R, Dufour J-F, Marotta P, Majno P, et al. Total tumor volume and alpha-fetoprotein for selection of transplant candidates with hepatocellular carcinoma: A prospective validation. *Hepatology* 2015; 62: 158–165.
- Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet* 2014; 46: 1267–1273.
- Trevisani F, Frigerio M, Santi V, Grignaschi A, Bernardi M. Hepatocellular carcinoma in non-cirrhotic liver: a reappraisal. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver* 2010; 42: 341–347.
- Tsai M-C, Manor O, Wan Y, Mosammamparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010; 329: 689–693.
- Tseng C-H. Metformin and risk of hepatocellular carcinoma in patients with type 2 diabetes. *Liver Int Off J Int Assoc Study Liver* 2018; 38: 2018–2027.
- Tsukada S, Parsons CJ, Rippe RA. Mechanisms of liver fibrosis. *Clin Chim Acta* 2006; 364: 33–60.
- Tumanov AV, Koroleva EP, Christiansen PA, Khan MA, Ruddy MJ, Burnette B, et al. T cell-derived lymphotoxin regulates liver regeneration. *Gastroenterology* 2009; 136: 694-704.e4.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006; 444: 1027–1031.
- Uccello M, Malaguarnera G, Corriere T, Biondi A, Basile F, Malaguarnera M. Risk of Hepatocellular Carcinoma in Workers Exposed to Chemicals [Internet]. *Hepat Mon* 2012; 12[cited 2019 Apr 11] Available from: <http://hepatmon.com/en/articles/70401.html>
- Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, et al. The antibacterial lectin RegIII $\gamma$  promotes the spatial segregation of microbiota and host in the intestine. *Science* 2011; 334: 255–258.
- Venook AP, Papandreou C, Furuse J, de Guevara LL. The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. *The Oncologist* 2010; 15 Suppl 4: 5–13.
- Villanueva A, Chiang DY, Newell P, Peix J, Thung S, Alsinet C, et al. Pivotal role of mTOR signaling in hepatocellular carcinoma. *Gastroenterology* 2008; 135: 1972–1983, 1983.e1–11.
- Vona G, Sabile A, Louha M, Sitruk V, Romana S, Schütze K, et al. Isolation by size of epithelial tumor cells: a new method for the immunomorphological and molecular characterization of circulating tumor cells. *Am J Pathol* 2000; 156: 57–63.
- Wan L, Xu K, Wei Y, Zhang J, Han T, Fry C, et al. Phosphorylation of EZH2 by AMPK Suppresses PRC2 Methyltransferase Activity and Oncogenic Function. *Mol Cell* 2018; 69: 279-291.e5.
- Wang F, Wang R, Li Q, Qu X, Hao Y, Yang J, et al. A transcriptome profile in hepatocellular carcinomas based on integrated analysis of microarray studies. *Diagn Pathol* 2017; 12: 4.
- Wang J, Zindy F, Chenivresse X, Lamas E, Henglein B, Bréchet C. Modification of cyclin A expression by hepatitis B virus DNA integration in a hepatocellular carcinoma. *Oncogene* 1992; 7: 1653–1656.

- Wang Q, Wang G, Liu C, He X. Prognostic value of CpG island methylator phenotype among hepatocellular carcinoma patients: A systematic review and meta-analysis. *Int J Surg Lond Engl* 2018; 54: 92–99.
- Wang W, Xu L, Liu P, Jairam K, Yin Y, Chen K, et al. Blocking Wnt Secretion Reduces Growth of Hepatocellular Carcinoma Cell Lines Mostly Independent of  $\beta$ -Catenin Signaling. *Neoplasia N Y N* 2016; 18: 711–723.
- Warburg O. THE METABOLISM OF TUMORS IN THE BODY. *J Gen Physiol* 1927; 8: 519–530.
- Ward JF. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Prog Nucleic Acid Res Mol Biol* 1988; 35: 95–125.
- Wei Y, Van Nhieu JT, Prigent S, Srivatanakul P, Tiollais P, Buendia M-A. Altered expression of E-cadherin in hepatocellular carcinoma: correlations with genetic alterations, beta-catenin expression, and clinical features. *Hepatol Baltim Md* 2002; 36: 692–701.
- Wice BM, Reitzer LJ, Kennell D. The continuous growth of vertebrate cells in the absence of sugar. *J Biol Chem* 1981; 256: 7812–7819.
- Wiest R, Das S, Cadelina G, Garcia-Tsao G, Milstien S, Groszmann RJ. Bacterial translocation in cirrhotic rats stimulates eNOS-derived NO production and impairs mesenteric vascular contractility [Internet]. 1999[cited 2019 Feb 5] Available from: <https://www.jci.org/articles/view/7458/pdf>
- Wiest R, Lawson M, Geuking M. Pathological bacterial translocation in liver cirrhosis. *J Hepatol* 2014; 60: 197–209.
- Wilson S, Fan L, Sahgal N, Qi J, Filipp FV. The histone demethylase KDM3A regulates the transcriptional program of the androgen receptor in prostate cancer cells. *Oncotarget* 2017; 8: 30328–30343.
- Witzigmann D, Quagliata L, Schenk SH, Quintavalle C, Terracciano LM, Huwyler J. Variable asialoglycoprotein receptor 1 expression in liver disease: Implications for therapeutic intervention. *Hepatol Res Off J Jpn Soc Hepatol* 2016; 46: 686–696.
- Wolk A, Gridley G, Svensson M, Nyrén O, McLaughlin JK, Fraumeni JF, et al. A prospective study of obesity and cancer risk (Sweden). *Cancer Causes Control CCC* 2001; 12: 13–21.
- Wong TCL, Lo CM. Resection strategies for hepatocellular carcinoma. *Semin Liver Dis* 2013; 33: 273–281.
- Wong VW-S, Wong GL-H, Choi PC-L, Chan AW-H, Li MK-P, Chan H-Y, et al. Disease progression of non-alcoholic fatty liver disease: a prospective study with paired liver biopsies at 3 years. *Gut* 2010; 59: 969–974.
- Woo HG, Lee J-H, Yoon J-H, Kim CY, Lee H-S, Jang JJ, et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. *Cancer Res* 2010; 70: 3034–3041.
- Wouters BG, Koritzinsky M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer* 2008; 8: 851–864.



Wrobel CN, Debnath J, Lin E, Beausoleil S, Roussel MF, Brugge JS. Autocrine CSF-1R activation promotes Src-dependent disruption of mammary epithelial architecture. *J Cell Biol* 2004; 165: 263–273.

Wu C-H, Ho Y-S, Tsai C-Y, Wang Y-J, Tseng H, Wei P-L, et al. In vitro and in vivo study of phloretin-induced apoptosis in human liver cancer cells involving inhibition of type II glucose transporter. *Int J Cancer* 2009; 124: 2210–2219.

Wu L-M, Yang Z, Zhou L, Zhang F, Xie H-Y, Feng X-W, et al. Identification of histone deacetylase 3 as a biomarker for tumor recurrence following liver transplantation in HBV-associated hepatocellular carcinoma. *PLoS One* 2010; 5: e14460.

Wu PC, Fang JW, Lau VK, Lai CL, Lo CK, Lau JY. Classification of hepatocellular carcinoma according to hepatocellular and biliary differentiation markers. Clinical and biological implications. *Am J Pathol* 1996; 149: 1167–1175.

Wynn A, Vacheron A, Zuber J, Solomon SS. Metformin Associated With Increased Survival in Type 2 Diabetes Patients With Pancreatic Cancer and Lymphoma. *Am J Med Sci* 2019

Xie Z, Wang J, Liu M, Chen D, Qiu C, Sun K. CC-223 blocks mTORC1/C2 activation and inhibits human hepatocellular carcinoma cells in vitro and in vivo [Internet]. *PLoS ONE* 2017; 12[cited 2019 Feb 4] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5363890/>

Xu H, Barnes GT, Yang Q, Tan G, Yang D, Chou CJ, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest* 2003; 112: 1821–1830.

Xu W, Cao L, Chen L, Li J, Zhang X-F, Qian H-H, et al. Isolation of circulating tumor cells in patients with hepatocellular carcinoma using a novel cell separation strategy. *Clin Cancer Res Off J Am Assoc Cancer Res* 2011; 17: 3783–3793.

Xu Y, Huang J, Ma L, Shan J, Shen J, Yang Z, et al. MicroRNA-122 confers sorafenib resistance to hepatocellular carcinoma cells by targeting IGF-1R to regulate RAS/RAF/ERK signaling pathways. *Cancer Lett* 2016; 371: 171–181.

Yamashita T, Budhu A, Forgues M, Wang XW. Activation of hepatic stem cell marker EpCAM by Wnt-beta-catenin signaling in hepatocellular carcinoma. *Cancer Res* 2007; 67: 10831–10839.

Yang J, Jin X, Yan Y, Shao Y, Pan Y, Roberts LR, et al. Inhibiting histone deacetylases suppresses glucose metabolism and hepatocellular carcinoma growth by restoring FBP1 expression. *Sci Rep* 2017; 7: 43864.

Yang JD, Addissie BD, Mara KC, Harmsen WS, Dai J, Zhang N, et al. GALAD Score for Hepatocellular Carcinoma Detection in Comparison with Liver Ultrasound and Proposal of GALADUS Score. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2019; 28: 531–538.

Yang W, Sun T, Cao J, Fan S. Hypoxia-Inducible Factor-1 $\alpha$  Downregulation by Small Interfering RNA Inhibits Proliferation, Induces Apoptosis, and Enhances Radiosensitivity in Chemical Hypoxic Human Hepatoma SMMC-7721 Cells. *Cancer Biother Radiopharm* 2011; 26: 565–571.

Yang Y, Li C, Nie X, Feng X, Chen W, Yue Y, et al. Metabonomic studies of human hepatocellular carcinoma using high-resolution magic-angle spinning  $^1\text{H}$  NMR spectroscopy in conjunction with multivariate data analysis. *J Proteome Res* 2007; 6: 2605–2614.

Yeung SJ, Pan J, Lee M-H. Roles of p53, MYC and HIF-1 in regulating glycolysis - the seventh hallmark of cancer. *Cell Mol Life Sci CMLS* 2008; 65: 3981–3999.

Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol* 2014; 2: 901–910.

Yoshida T, Maulik N, Engelman RM, Ho YS, Das DK. Targeted disruption of the mouse Sod I gene makes the hearts vulnerable to ischemic reperfusion injury. *Circ Res* 2000; 86: 264–269.

Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* 2013; 499: 97–101.

You A, Cao M, Guo Z, Zuo B, Gao J, Zhou H, et al. Metformin sensitizes sorafenib to inhibit postoperative recurrence and metastasis of hepatocellular carcinoma in orthotopic mouse models. *J Hematol Oncol* *J Hematol Oncol* 2016; 9: 20.

Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatol Baltim Md* 2016; 64: 73–84.

Yu G, Chen X, Chen S, Ye W, Hou K, Liang M. MiR-19a, miR-122 and miR-223 are differentially regulated by hepatitis B virus X protein and involve in cell proliferation in hepatoma cells. *J Transl Med* 2016; 14: 122.

Yu L-X, Yan H-X, Liu Q, Yang W, Wu H-P, Dong W, et al. Endotoxin accumulation prevents carcinogen-induced apoptosis and promotes liver tumorigenesis in rodents. *Hepatology* 2010; 52: 1322–1333.

Yun J, Rago C, Cheong I, Pagliarini R, Angenendt P, Rajagopalan H, et al. Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells. *Science* 2009; 325: 1555–1559.

Yuneva M. Finding an “Achilles’ heel” of cancer: The role of glucose and glutamine metabolism in the survival of transformed cells. *Cell Cycle* 2008; 7: 2083–2089.

Yuneva M, Zamboni N, Oefner P, Sachidanandam R, Lazebnik Y. Deficiency in glutamine but not glucose induces MYC-dependent apoptosis in human cells. *J Cell Biol* 2007; 178: 93–105.

Zetterberg A, Engström W. Glutamine and the regulation of DNA replication and cell multiplication in fibroblasts. *J Cell Physiol* 1981; 108: 365–373.

Zhan T, Digel M, Küch E-M, Stremmel W, Füllekrug J. Silybin and dehydrosilybin decrease glucose uptake by inhibiting GLUT proteins. *J Cell Biochem* 2011; 112: 849–859.

Zhang C, Guo X, Jiang G, Zhang L, Yang Y, Shen F, et al. CpG island methylator phenotype association with upregulated telomerase activity in hepatocellular carcinoma. *Int J Cancer* 2008; 123: 998–1004.

Zhang C, Li Z, Cheng Y, Jia F, Li R, Wu M, et al. CpG island methylator phenotype association with elevated serum alpha-fetoprotein level in hepatocellular carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res* 2007; 13: 944–952.

Zhang H, Gao P, Fukuda R, Kumar G, Krishnamachary B, Zeller KI, et al. HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell* 2007; 11: 407–420.

Zhang H-L, Yu L-X, Yang W, Tang L, Lin Y, Wu H, et al. Profound impact of gut homeostasis on chemically-induced pro-tumorigenic inflammation and hepatocarcinogenesis in rats. *J Hepatol* 2012; 57: 803–812.

Zhang Y, Huang F, Wang J, Luo H, Wang Z. 2-DG-Regulated RIP and c-FLIP Effect on Liver Cancer Cell Apoptosis Induced by TRAIL. *Med Sci Monit Int Med J Exp Clin Res* 2015; 21: 3442–3448.

Zhu AX, Abrams TA, Miksad R, Blaszkowsky LS, Meyerhardt JA, Zheng H, et al. Phase 1/2 study of everolimus in advanced hepatocellular carcinoma. *Cancer* 2011; 117: 5094–5102.

Zhu AX, Finn RS, Edeline J, Cattani S, Ogasawara S, Palmer D, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol* 2018; 19: 940–952.

Zhu AX, Kudo M, Assenat E, Cattani S, Kang Y-K, Lim HY, et al. Effect of everolimus on survival in advanced hepatocellular carcinoma after failure of sorafenib: the EVOLVE-1 randomized clinical trial. *JAMA* 2014; 312: 57–67.

Zhu C, Utsunomiya T, Ikemoto T, Yamada S, Morine Y, Imura S, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. *Ann Surg Oncol* 2014; 21 Suppl 4: S729-35.

Zhu H, Luo S, Wang J, Li X, Wang H, Pu W, et al. Effect of environmental factors on chemoresistance of HepG2 cells by regulating hypoxia-inducible factor-1 $\alpha$ . *Chin Med J (Engl)* 2012; 125: 1095–1103.

Zucman-Rossi J, Benhamouche S, Godard C, Boyault S, Grimber G, Balabaud C, et al. Differential effects of inactivated Axin1 and activated  $\beta$ -catenin mutations in human hepatocellular carcinomas. *Oncogene* 2007; 26: 774–780.

Zytoon AA, Ishii H, Murakami K, El-Kholy MR, Furuse J, El-Dorry A, et al. Recurrence-free survival after radiofrequency ablation of hepatocellular carcinoma. A registry report of the impact of risk factors on outcome. *Jpn J Clin Oncol* 2007; 37: 658–672.

**Titre :** Identification de régulateurs clés de la carcinogenèse hépatique humaine : Etudes clinico-pathologiques, moléculaires et fonctionnelles

**Mots clés :** Cancer, Microdissection laser, Omiques, Métabolisme, Génome non codant

**Résumé :** Le carcinome hépatocellulaire (CHC) est la forme la plus fréquente de cancer du foie et l'une des principales causes de mortalité par cancer dans le monde. Il s'agit d'une maladie de mauvais pronostic, aux ressources thérapeutiques limitées, hétérogène sur le plan immunophénotypique et génomique, qui se développe très souvent sur un foie remanié cirrhotique. Les études moléculaires ont révélé plusieurs sous-classes de CHC caractérisés par des signatures génomiques et protéomiques distinctes. Au cours de mon travail de thèse, nous avons contribué à améliorer notre compréhension de la biologie des CHC et des classifications moléculaires en cartographiant le génome non-codant de tumeurs de CHC induites par des virus hépatotropes (VHB, VHC) et en étudiant la sous-classe moléculaire de CHC la plus agressive KRT19-positif. Nous avons établi la première carte de transcriptome à ARN non codants du CHC et révélé une forte

activation intra-tumorale des rétrotransposons à LTR, qui sont principalement inhibés dans les cellules hépatiques normales, dans des CHC induits par les VHB et VHC. Certains des transcrits dérivés de LTR se sont révélés être des régulateurs clés de l'expression génique et donc activer la croissance cellulaire. Dans la deuxième étude, nous identifions une nouvelle voie de régulation des CHC KRT19-positif affectant le métabolisme énergétique de ces tumeurs. Les CHC KRT19-positif sont des tumeurs fortement glycolytiques liée à une activation de la réponse à l'hypoxie. L'excès de production par les CHC KRT19-positif de l'oncométabolite 2-hydroxyglutarate en absence de mutation des gènes IDH1 et IDH2 était associé à un profil aberrant hyperméthylé sur la lysine 9 de l'histone H3 (H3K9me3) suggérant une répression de la transcription notamment des gènes impliqués dans la différenciation cellulaire

**Title :** Key regulators identification of human hepatocarcinogenesis : Clinical, pathological, molecular and functional studies

**Keywords :** Cancer, Laser microdissection, OMICS, metabolism, non-coding genome

**Abstract :** Hepatocellular carcinoma (HCC) is the main primary liver cancer and one of the most leading cause of cancer-related death worldwide. This heterogeneous disease with a worse prognosis has been subjected of numerous studies aimed to establish global phenotypic profiles. During my thesis, I dedicated my work to improve these classifications by identifying signatures on the non-coding genome and working on a very aggressive form of HCC expressing progenitor markers. With help of a Japanese team, we demonstrated that LTR-derived ncRNAs were active in HCC and that correlation correlates with expression of common cancer markers (GPC3) and TP53 mutations. This signature can also be used to discriminate HCCs at high risk of recurrence. Finally, we have showed that

these LTRs are detectable on preneoplastic stages in the Mdr2 KO mouse model. In parallel, I worked on HCC that expresses progenitor markers such as cytokeratin 19. Using proteomic and transcriptional approaches and in silico analyses, we propose that the occurrence of this type of cancer is due to an hypoxic event likely related to trans-arterial chemoembolization. These tumors have a highly glycolytic phenotype with production of an oncometabolite (2-hydroxyglutarate) that has been generally found in IDH1/2 mutated cholangiocarcinomas. Finally we suggest the use of metformin, type 2 diabetes drug, to reverse metabolic reprogramming and restore sensitivity to chemotherapy..

