UNIVERSITAT DE
BARCELONA

Molecular Characterization and Novel Therapeutic Approaches in Hepatocellular Carcinoma

Laura Torrens Fontanals

UNIVERSITAT DE BARCELONA

Faculty of Medicine

Molecular Characterization and Novel Therapeutic Approaches in Hepatocellular Carcinoma

Doctoral thesis report submitted by

Laura Torrens Fontanals

To obtain a doctoral degree by the University of Barcelona

Supervised by:

Prof. Josep M Llovet Bayer

Full Professor of Medicine – University of Barcelona

Professor of Research – Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS),

Hospital Clínic de Barcelona

Doctoral Programme in Medicine and Translational Research

Faculty of Medicine and Health Sciences. University of Barcelona

December 2021

IDIBAPS
Institut
D'Investigacions
Biomèdiques
August Pi i Sunyer

CLÍNIC
BARCELONA
Hospital Universitari

Icahn School
of Medicine at
Mount
Sinai

Professor of Research
IDIBAPS
Liver Unit, Hospital Clínic Barcelona
Villaroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

Barcelona, 1 de Desembre de 2021

Jo, Dr. Josep M. Llovet Bayer, cap del "Laboratori de Recerca Translacional en Oncologia Hepàtica" de l'Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) – Universitat de Barcelona – Hospital Clínic, Professor ICREA i director de Máster Oficial de Medicina Translacional de la Universitat de Barcelona,
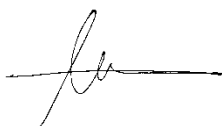
CERTIFICO:

Que la tesi doctoral titulada "Molecular Characterization and Novel Therapeutic Approaches in Hepatocellular Carcinoma" presentada per Laura Torrens Fontanals per optar al títol de Doctor en Medicina i Recerca Translacional per la Universitat de Barcelona s'ha realitzat sota la meva direcció i compleix tots el requisits necessaris per ser defensada davant el Tribunal d'Avaluació corresponent.

Dr. Josep M Llovet i Bayer

Conformitat de l'estudiant de doctorat:

Laura Torrens Fontanals

Universitat de Barcelona

Vigila, esperit, vigila;
no perdis mai el teu nord;
no et deixis dur a la tranquil·la
aigua mansa de cap port.

Gira, gira els ulls enlaire,
no miris les platges roïns,
dóna el front en el gran aire;
sempre, sempre mar endins.

Sempre amb les veles suspeses
del cel al mar transparent;
sempre entorn aigües esteses
que es moguin eternament.

Fuig-ne de la terra immoble;
fuig dels horitzons mesquins;
sempre al mar, al gran mar noble:
sempre, sempre mar endins.

Fora terres, fora platja;
oblida't de tot regrés;
no s'acaba el teu viatge;
no s'acabarà mai més.

Joan Maragall

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Josep M Llovet, my thesis director, for his support during my PhD, for the invaluable opportunity to participate in the research projects included in this thesis, and for always pushing me to be better. I also want to thank him for providing the ideal environment both in Barcelona and New York for me to grow as a researcher.

I have been so fortunate to have the most supportive and amazing labmates, who have made the most enjoyable experience out of this PhD, even when experiments fail, grants get rejected and reviewer 3 doesn't like your data. For this, I want to give my most heartfelt thank you to everyone at the Liver Cancer Translational Research laboratory at IDIBAPS. I would like to begin by thanking Agrin for guiding me in my first steps as a disoriented "little grasshopper". To Roser, Victoria, and Catherine, for their continuous support and predisposition to answering all my questions. To Iris, Sara, and Robert for being my "senior PhD students" and reference during this journey. To Carmen, for being my partner in crime both inside and outside the lab and for making me smile even when there is an ocean in between us. To Florian, for all the priceless laughs, advice, and scientific discussions in the office. To all my "junior PhD students", Jordi, Ugne, Agavni, Roger, Marta, and Júlia for making going to work a true pleasure, and particularly to the lenva/pembro and hepatoblastoma teams, for being a fundamental part of this thesis. To Judit and Laia, thanks for teaching me all the techniques and for being so patient with my absent-minded self – sorry for forgetting to turn off the centrifuge for the millionth time. Also to Ari, Marta, and Sara for always being so helpful and making the administrative side of research much easier.

My stay in New York has been one of the most enriching experiences of this thesis, both professionally and personally. I would like to thank everyone at the Mount Sinai Liver Cancer Program and the ISMMS Department of Liver Diseases for their support and for making me feel at home since day one. To Daniela, for her mentoring and for taking the time to make sure that my projects were moving forward. To the relentless Mongolian team, Marc and Miguel, for cheering with me for all the victories and defeats – or just Fridays at Earl's. Especially to Marc, *per ser Nova York*. To Philipp for all his advice and conversations on our way to the animal facility, and also for showing me how to sneak into Nowadays. To Carla and Miho for making all the hours spent at the lab more fun.

Equally important for my cherished time in New York have been all those friends that reminded me there is a whole city outside of Mount Sinai. For this, I want to thank Vinit and Sara for

including me in countless adventures and introducing me to the loveliest and most eclectic group of New Yorkers. Also, to Yago and Sami for completing the Earl's gang and for helping me master my Jenga skills week after week.

Com explicava el meu professor d'universitat Dr. David Bueno, la llengua materna està estretament lligada a l'emoció. Certament, no podria escriure els següents agraïments de cap altra manera que en català. A tots els amics i companys que m'heu acompanyat durant aquesta etapa, us estaré per sempre més agraïda. Gràcies, Paula, Irina, Irene, Jordi, Elena i Ania per ser el meu ATG. En especial a tu, Paula, per ser un pilar i la millor companya en els plans més descabellats. A la Blanca per ser-hi, des de fa gairebé 15 anys, cada estiu davant del Ceferino a les sis i cada cop que he necessitat una amiga. I també a la Júlia per les seves paraules d'ànims acompanyades sempre dels millors gin-tònics amb cardamom.

Per últim, vull expressar el més profund agraïment a la meva família. En especial, als meus pares, Jordi i Núria, pel seu amor incondicional i per haver-me donat totes les eines i els valors que m'han permès arribar fins aquí. A la Mariona per ser la meva *other half* i per fer-me sentir sempre compresa. Aquest camí hauria sigut impossible si no l'hagués emprès amb tu. A l'Anna, pels seus bons consells de germana gran, i a la petita Berta per portar-nos joia i ajudar-me a retrobar la bellesa en les coses més essencials. Per últim, als avis Jordi i Carme, que ja no hi son, per fer-me sentir sempre estimada i encoratjada, i als avis Ernest i Carme per ser el meu referent i per construir, amb afecte i tenacitat, la nostra família que és el millor tresor que podria tenir.

TABLE OF CONTENTS

INDEX

FIGURES

TABLES

ANNEX

ABBREVIATIONS

AAV: Adeno-associated virus

AAV2: Adeno-associated virus type 2

AFP: Alpha fetoprotein

ALT: Alanine aminotransferase

BCLC: Barcelona Clinic Liver Cancer

BCP: Basal core promoter

CNV: Copy number variation

CTLA4: Cytotoxic T-lymphocyte-associated protein 4

CTNNB1: Catenin beta 1

DC: Dendritic cell

DC1: Type 1 dendritic cell

DDR: DNA damage response

DMS: Dimethyl sulfate

FC: Fold change

FDA: Food and Drug Administration

FGFR: Fibroblast growth factor receptor

FOXP3: Forkhead box protein 3

GEMM: Genetically engineered mouse model

GSEA: Gene set enrichment analysis

HBsAg: HBV surface antigen

HBV: Hepatitis B virus

HCC: Hepatocellular carcinoma

HCV: Hepatitis C virus

HDV: Hepatitis delta virus

HFD: High-fat diet

HGDN: High-grade dysplastic nodules

ICGC: International Cancer Genome Consortium

ICI: Immune checkpoint inhibitor

IFN- α: Interferon α

IHC: Immunohistochemistry

LGDN: Low-grade dysplastic nodule

MDSC: Myeloid-derived suppressor cell

NAFLD: Non-alcoholic fatty liver disease

NASH: Non-alcoholic steatohepatitis

NGS: Next generation sequencing

ABBREVIATIONS

NK: natural killer

NMFc: Non-negative matrix factorization

NTP: Nearest template prediction

ORR: Objective response rate

PDX: Patient-derived xenograft

PD1: Programmed cell death protein 1

rAAV: Recombinant adeno-associated virus

RET: RET proto-oncogene

RNA-seq: RNA sequencing

SBS: Single base substitution

SBSM: SBS Mongolia

SNP: Single-nucleotide polymorphism

SNV: Single nucleotide variant

ssGSEA: Single sample gene set enrichment analysis

TCGA: The Cancer Genome Atlas

TGF-β: Transforming growth factor β

TAM: Tumor-associated macrophage

TAN: Tumor-associated neutrophils

TIM3: T cell immunoglobulin and mucin domain containing-3

TKI: Tyrosine-kinase inhibitor

TMB: Tumor mutational burden

T_{reg}: Regulatory T cell

WES: Whole-exome sequencing

WGS: Whole-genome sequencing

LIST OF ARTICLES COMPRISED IN THE THESIS

Thesis in the form of a collection of published articles. The thesis includes 3 articles and 3 aims.

Study 1 – Molecular Characterization of Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features

Torrens L, Puigvehí M, Torres-Martín M, Wang H, Maeda M, Haber PK, Leonel T, García-López M, Leow WQ, Montironi C, Torrecilla S, Varadarajan AR, Taik P, Campreciós G, Enkhbold C, Taivanbaatar E, Yerbolat A, Villanueva A, Pérez-del-Pulgar S, Thung S, Chinburen J, Letouzé E, Zucman-Rossi J, Uzilov A, Neely J, Forns X, Roayaie S, Sia D, Llovet JM. Molecular Characterization of Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features. Submitted to Proc Natl Acad Sci U S A. 2021.

Impact factor: 11.205, 1st quartile. Subject Areas: Multidisciplinary.

Aim:

1. To provide a molecular characterization of Mongolian HCC and identify its unique genomic features compared to Western HCC.

Study 2 – Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice

Dalwadi D, **Torrens L**, Abril-Fornaguera J, Pinyol R, Willoughby C, Posey J, Llovet JM, Lanciault C, Russell DW, Grompe M, Naugler WE. Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice. Mol Ther. 2021;29:680–90.

Impact factor: 11.454, 1st quartile. Subject Areas: Genetics, Molecular Biology, Molecular Medicine, Drug Discovery, Pharmacology, Medicine (miscellaneous).

Aim:

2. To assess whether NAFLD-associated liver damage increase the risk of AAV integration inducing HCC.

Study 3 – Immunomodulatory Effects of Lenvatinib Plus Anti-Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma

Torrens L, Montironi C, Puigvehí M, Mesropian A, Leslie J, Haber PK, Maeda M, Balaseviciute U, Willoughby CE, Abril-Fornaguera J, Piqué-Gili M, Torres-Martín M, Peix J, Geh D, Ramon-Gil E, Saberi B, Friedman SL, Mann DA, Sia D, Llovet JM. Immunomodulatory Effects of Lenvatinib Plus Anti–Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma. Hepatology. 2021;74:2652–69.

Impact factor: 17.425, 1st quartile. Subject Areas: Hepatology, Medicine (miscellaneous).

Aim:

3. To identify the immunomodulatory effects of lenvatinib in combination with anti-PD1 and provide a mechanistic rationale for this treatment in advanced HCC.

THESIS SUMMARY (CATALAN)

Títol

Caracterització Molecular i Noves Estratègies Terapèutiques pel Carcinoma Hepatocel·lular

Introducció

El carcinoma hepatocel·lular (CHC) representa un important problema de salut pública degut a la seva elevada incidència i mortalitat[1,2]. La incidència mundial de CHC és heterogènia, sent Mongòlia el país amb més casos per habitant, gairebé 10 vegades per sobre de la mitjana global[2]. Els principals factors de risc del CHC són la infecció pel virus de l'hepatitis B o pel virus de l'hepatitis C i la malaltia del fetge gras no alcohòlica (NAFLD, de les seves sigles en anglès). També s'han proposat altres agents que podrien estar associats a hepatocarcinogènesi, inclosa la infecció per virus adenoassociat (VAA)[3,4], però és necessària més informació per determinar en quines condicions aquest virus pot promoure el CHC.

Al voltant del 40% dels pacients amb CHC són diagnosticats en etapes avançades de la malaltia, en les quals són elegibles per a teràpies sistèmiques incloent inhibidors multiquinases (p. ex., lenvatinib) i inhibidors de punts de control immunitaris (ICI; p. ex., anticossos anti-PD1). Tot i que els ICI estan revolucionant el tractament del CHC, només aconsegueixen respostes en el ~15% dels pacients en monoteràpia[5,6]. Per contra, noves teràpies combinades estan incrementant la taxa de resposta fins al ~30%. Considerant això, és necessari identificar noves teràpies que puguin fer sinergia amb ICI per tal d'augmentar la supervivència dels pacients.

Durant els últims anys, la caracterització molecular i immunològica de tumors mitjançant estudis translacionals ha proporcionat una millor comprensió de la patogènesi molecular del CHC[5,6]. Aquest coneixement té una gran rellevància clínica, ja que, entre altres coses, permet 1) identificar determinants genètics i moleculars que afavoreixen l'hepatocarcinogènesi en poblacions específiques de pacients; i 2) dissenyar nous tractaments incloent teràpies combinades.

Hipòtesis

Les hipòtesis d'aquesta tesi són les següents:

- L'anàlisi exhaustiva de les característiques moleculars i immunològiques del CHC proporcionarà nova informació sobre els determinants genètics i moleculars associats al CHC en poblacions d'alt risc, afavorint l'estudi de noves estratègies terapèutiques. Aquestes dades podrien tenir implicacions fonamentals per a la presa de decisions clíniques en CHC.

- L'avaluació de les alteracions genòmiques i transcripcionals del CHC a Mongòlia pot proporcionar nova informació que ajudi a identificar factors genètics i ambientals propis associats a l'elevada incidència en aquesta població.
- La malaltia hepàtica crònica i NAFLD promouen la integració oncogènica del VAA i el desenvolupament de CHC, la qual cosa podria ser un condicionant per a l'ús de la teràpia gènica mitjançant VAA en pacients.
- L'inhibidor multiquinasa lenvatinib té potencial immunomodulador i la seva combinació amb ICI anti-PD1 podria incrementar l'efecte anti-tumoral en models experimentals i en pacients amb CHC.

Objectius

Els objectius específics d'aquesta tesi doctoral són:

- Realitzar una caracterització molecular del CHC a Mongòlia i identificar trets moleculars específics en comparació amb tumors de pacients occidentals.
- Analitzar si el NAFLD incrementa el risc d'integració del VAA en hepatòcits, induint així el CHC.
- Investigar l'efecte immunomodulador de lenvatinib en combinació amb l'anticòs anti-PD1 per proporcionar una base racional pel seu ús com a tractament pel CHC avançat.

Mètodes

- **Estudi #1** (Torrens *et al.*, en revisió)**:** Un total de 192 mostres de tumors hepàtics CHC de pacients provinents de Mongòlia i 187 mostres de CHC de pacients occidentals (Europa i EEUU) van ser analitzades mitjançant seqüenciació completa d'exoma i d'ARN. Per identificar les característiques moleculars úniques del CHC de Mongòlia, es van comparar les característiques clinico-patològiques dels tumors de pacients mongols i occidentals, així com els seus perfils mutacionals i transcripcionals.
- **Estudi #2** (Dalwadi *et al.*, Mol Ther 2021)**:** Per tal de determinar si la infecció pel virus VAA dels hepatòcis promou l'hepatocarcinogènesi en condicions de malaltia hepàtica, es va fer servir un model animal murí on ratolins adults i nounats de la soca C57BL/6 van ser infectats amb un vector VAA dirigit contra el locus genètic *Rian* (VAA-Rian). A continuació, els animals van ser a) alimentats amb una dieta rica en greixos per generar NAFLD o b) sotmesos a hepatectomia parcial per generar dany hepàtic i regeneració. Es va monitoritzar el

desenvolupament tumoral i els tumors generats es van caracteritzar mitjançant seqüenciació d'ARN.

- **Estudi #3** (Torrens *et al.*, Hepatology 2021)**:** Per avaluar l'activitat anti-tumoral de la combinació de l'inhibidor multiquinasa lenvatinib amb anti-PD1 es van generar tres models murins singènics de CHC. L'impacte dels tractaments sobre el perfil molecular i immunitari dels tumors van ser analitzats per citometria de flux i perfilats a nivell transcripcional i immunohistoquímic. Finalment, es va explorar el perfil d'expressió gènica de 228 tumors humans per identificar pacients amb CHC que es podrien beneficiar de la teràpia combinada de lenvatinib i anti-PD1.

Resultats

- **Estudi #1** (Torrens *et al.*, en revisió)**:** Els resultats del nostre estudi indiquen que els tumors CHC de pacients de Mongòlia presenten les següents particularitats:
  1. Una elevada taxa de mutacions, que gairebé dobla la taxa de mutacions de la cohort occidental (121 davant de 70 mutacions per tumor).
  2. Un perfil mutacional diferent, amb una major freqüència de mutacions de *TP53*, *APOB*, *TSC2* i la família de gens *KMT2*.
  3. La presència d'una nova firma mutacional (SBS Mongòlia) al 25% dels casos, que està associada a la firma del genotòxic dimetil sulfat derivat de la combustió del carbó.
  4. A nivell d'expressió gènica, els tumors de la cohort de Mongòlia es classifiquen en tres classes moleculars (MGL1, MGL2 i MGL3), dues de les quals (MGL2 al 26% dels CHC i MGL3 al 30%) i presenten característiques moleculars i clínico-patològiques no observades en mostres de CHC occidental.
- **Estudi #2** (Dalwadi *et al.*, Mol Ther 2021)**:**
  1. Només el 5% dels ratolins adults infectats amb VAA-Rian van desenvolupar CHC. Per contra, l'elevada proliferació cel·lular en nounats va promoure un increment en la integració viral i generació de tumors en el 100% dels animals.
  2. Tant la lesió hepàtica per la dieta rica en greix com la derivada de l'hepatectomia parcial van conduir a un increment en l'aparició de CHC en ratolins adults infectats amb VAA-Rian en comparació amb animals infectats però no sotmesos a estrès hepàtic.
  3. L'anàlisi transcriptòmica de tumors de CHC murins de ratolins originats degut a la infecció amb VAA-Rian va revelar similituds moleculars amb tumors de CHC humans

amb sobreexpressió del locus *Rian*, associats amb proliferació, fenotip agressiu i mal pronòstic.

4. El fetge de ratolins alimentats amb una dieta rica en greixos presentaven un perfil immunològic protumorigènic, que podria ser el motiu de l'increment del risc de CHC induït per la infecció de VAA.

- **Estudi #3** (Torrens *et al.*, Hepatology 2021)**:**

1. La combinació de lenvatinib i anti-PD1 aconsegueix una major taxa de resposta, un temps de resposta més curt i una reducció de la viabilitat del tumor en comparació amb els tractaments administrats en monoteràpia.

2. Lenvatinib exerceix un potent efecte immunomodulador sobre l'infiltrat tumoral caracteritzat per una reducció en la proporció de cèl·lules T reguladores i inhibició de la senyalització immunosupressora mitjançant TGFß.

3. El tractament anti-PD1 modifica l'infiltrat immunològic del tumor augmentant les cèl·lules T i la subpoblació de cèl·lules dendrítiques de tipus 1 al tumor.

4. La combinació de lenvatinib amb anti-PD1 és l'únic dels tractaments avaluats capaç de generar una resposta immune anti-tumoral activada.

5. El 22% dels pacients amb CHC presenten un perfil transcripcional que podria estar associat a resistència primària a anti-PD1 i resposta al règim combinat. Aquests pacients podrien beneficiar-se de l'efecte de reforç del tractament de combinació.

Conclusions

- El CHC a Mongòlia presenta característiques moleculars úniques comparat amb els tumors de pacients occidentals. El seu perfil mutacional podria ser degut a exposició a factors ambientals responsables de l'elevada incidència en aquest país.

- La infecció amb el virus VAA promou el desenvolupament de CHC en models murins amb fetge gras degut a un augment en la proliferació d'hepatòcits i a la presència d'un infiltrat protumorogènic. Per tant, la teràpia gènica amb VAA podria resultar oncogènica en pacients amb inflamació hepàtica crònica.

- La combinació de lenvatinib i anti-PD1 té un efecte immunomodulador al CHC caracteritzat per la generació d'un perfil immunològic activat. Aquest tractament podria maximitzar el benefici clínic en un subgrup de pacients de CHC.

- Els estudis translacionals basats en models preclínics i anàlisis òmiques tenen la capacitat de revelar les característiques moleculars del CHC, potencials factors de risc i nous enfocaments terapèutics.

INTRODUCTION

1. Principles of Cancer

1.1. Worldwide Impact of Cancer

The term **cancer** refers to a group of diseases involving uncontrollable cell growth with the potential to spread to other parts of the body. Cancer is one of the leading causes of death worldwide and supposes an important barrier to increasing life expectancy[7]. In 2020, there was an estimated 19.3 million new cases of cancer and 10 million cancer-related deaths worldwide. The burden of cancer **incidence and mortality** is rapidly increasing, partly due to aging of the population and changes in the prevalence and distribution of cancer risk factors associated with socioeconomic development (e.g., smoking, unhealthy diet, excess body weight, and physical inactivity)[2]. Worldwide, an estimated 28.4 million new cancer cases are projected to occur in 2040, corresponding to a 47% increase from the current number of cases[2]. The most commonly diagnosed cancer types are breast (11.7% of total cases), lung (11.4%), colorectal (10.0%), prostate (7.3%), stomach (5.6%) and liver (4.7%) cancers. In terms of mortality, the most common causes of cancer-related death are lung (18.0%), colorectal (9.4%), liver (8,3%), stomach (7.7%), breast (6.9%) and esophagus (5.5%) cancer[2]. Notably, the incidence and mortality of cancer are not homogeneous around the world and present **regional and gender variations** reflecting societal, economic, and lifestyle disparities[2].

Many efforts have been invested to identify risk factors for cancer development and effective therapeutic strategies to reduce its incidence and mortality. In this context, understanding the biological principles of the disease and its clinical implications is crucial.

1.2. Biological Basis of Cancer

1.2.1. Hallmarks of Cancer

Tumorigenesis is a **multistep process** that drives the progressive transformation of normal human cells into malignant cells[8]. Observations of human tumors and animal models suggest that all cancers are a manifestation of eight essential **hallmark capabilities** consisting in alterations in cell physiology that collectively permit malignant growth[8,9] (**Figure 1**):

- *Sustaining Proliferative Signaling*: Cancer cells have the capacity to increase the production and release of growth-promoting signals that instruct progression through

the cell division cycle, thereby deregulating the homeostasis of cell number and achieving an aberrant tissue architecture and function. These signals are conveyed in large part by growth factors that bind to cell-surface receptors, typically containing intracellular tyrosine kinase domains.

- *Evading Growth Suppressors*: Cancer cells can also circumvent inhibitory signaling that negatively regulates cell proliferation; many of which depend on the actions of tumor suppressor genes.

- *Resisting cell death*: Programmed cell death by apoptosis occurs physiologically to maintain cell population and acts as a natural barrier to cancer development. When cells become old or damaged, they are genetically programmed to die in order to prevent the propagation of DNA errors. Through the introduction of abnormalities in sensors that play a key role in triggering apoptosis, tumor cells can avoid cell death.

- *Enabling replicative Immortality*: Most normal cell lineages can undergo only a limited number of successive cell growth-and-division cycles. Cancer cells, on the other hand, require unlimited replicative potential to generate macroscopic tumors. Multiple lines of evidence indicate that telomeres protecting the ends of chromosomes are centrally involved in this capability.

- *Inducing angiogenesis*: The tumor-associated neovasculature, generated by the process of angiogenesis, provides essential nutrients and oxygen, and evacuates metabolic wastes and carbon dioxide. By synthesizing angiogenic factors, tumors promote the production of aberrant vessels enabling cell proliferation and metastasis.

- *Activating invasion and metastasis*: The capability for invasion enables cancer cells to escape the primary tumor mass and colonize other tissues. In this process, the primary tumor spawn cancer cells that move out, invade adjacent tissues, and travel to distant sites where they may form new colonies. These distant settlements of tumor cells — metastases — are the cause of 90% of human cancer deaths.

- *Deregulation of cellular energetics*: The uncontrolled cell proliferation in cancer requires adjustments of energy metabolism to fuel cell growth and division. Cancer cells can reprogram their glucose metabolism, and thus their energy production, by limiting their energy metabolism largely to glycolysis.

- *Avoiding immune destruction*: Tumors have developed several strategies to avoid detection by the immune system, which constantly monitors cells and tissues through a process known as immune surveillance. These processes include escaping immune recognition, suppressing immune reactivity, or preventing T-cell infiltration.

1.2.2. Enabling Characteristics

Enabling characteristics are necessary for the acquisition of the eight hallmark capabilities (**Figure 1**):

- *Genome instability and mutation*: The acquisition of oncogenic traits largely depends on a succession of alterations in the genomes of cancer cells. Certain mutant genotypes confer a selective advantage on subsets of cancer cells, enabling their outgrowth and eventual dominance in a local tissue environment.

- *Tumor-promoting inflammation*: Inflammatory cells infiltrated in the tumor can contribute to multiple hallmark capabilities by supplying bioactive molecules to the microenvironment, including growth factors that sustain proliferative signaling, survival factors that limit cell death, proangiogenic factors, extracellular matrix-modifying enzymes that facilitate invasion, etc.
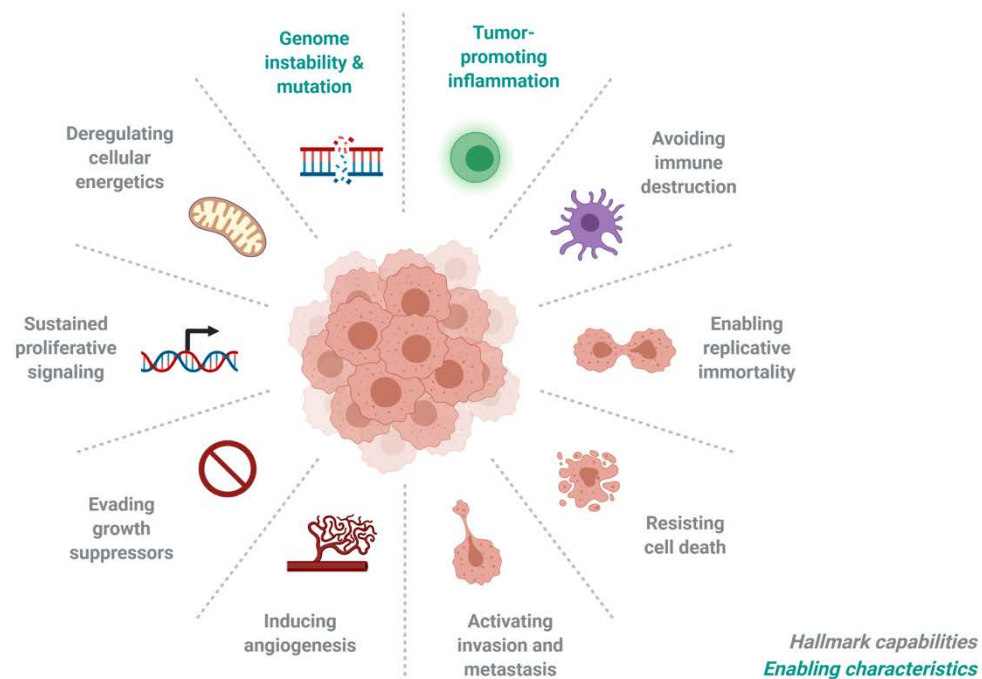
Figure 1. Hallmarks of cancer and enabling characteristics. Representation of the eight hallmark capabilities and two enabling characteristics originally proposed by

Hanahan and Weinberg in 2000 and further expanded in 2011. Modified from Hanahan D & Weinberg D, *Cell* 2011[9] using biorender.com.

The following section expands on the implications of these enabling characteristics – **genomic mutations and inflammation** – in shaping cancer and how the mutation and inflammatory profile of a tumor can provide valuable information about the molecular basis of the disease and optimal therapeutic approaches.

1.3. Genomic Basis of Cancer

As stated by the eminent oncologist Bert Vogelstein, "cancer is, in essence, a genetic disease"[10], since it originates due to changes in the DNA that are acquired and maintained with each somatic cell replication. These acquired changes, known as **somatic alterations**, are the result of cell-intrinsic processes (e.g., DNA damage or spontaneous decay of nucleotides throughout time) or extrinsic exposures (e.g., ultraviolet radiation or cigarette smoke). Although less frequent, **germline alterations** associated with hereditary cancer-predisposition are also a mechanism of cancer development[11].

1.3.1. Structural alterations

Structural alterations correspond to all the molecular aberrations that involve modification of the underlying DNA sequence.

- *Mutations*: These alterations involve changes in the DNA sequence. Depending on the position and specific change, mutations may alter the protein sequence, known as **non-synonymous mutations**. Most adult solid tumors (e.g., colon, breast, brain, and liver cancer) present an average of 33 to 66 non-synonymous mutations (**Figure 2**)[12]. Conversely, certain tumor types display more mutations than average, reflecting the involvement of a potent mutagen in their pathogenesis. This is the case of tobacco smoking in lung cancer and ultraviolet light in melanoma.

    About 95% of the non-synonymous mutations in a tumor correspond to single-base substitutions (e.g., C>G), also known as **single nucleotide variants** (SNV). Out of these non-synonymous SNVs, 90.7% result in missense changes, 7.6% result in nonsense changes, and 1.7% result in alterations of splice sites or untranslated regions immediately adjacent to the start and stop codons[12]. The remaining 5% of non-

synonymous mutations comprise **deletions and insertions** of one or a few bases (e.g., CTT>CT)[12].

Only a small fraction of non-synonymous mutations (3-8 mutations) participate in the transformation of a normally functioning cell into a cancer cell[13]. These mutations are known as "**drivers**". The remaining alterations – termed "**passengers**" – occur coincidentally as a byproduct of the operative mutational processes during tumorigenesis, without supposing a functional advantage to the cell. To date, only ~500 of the 20,000 genes in the human genome have been shown to act as driver genes[13,14].
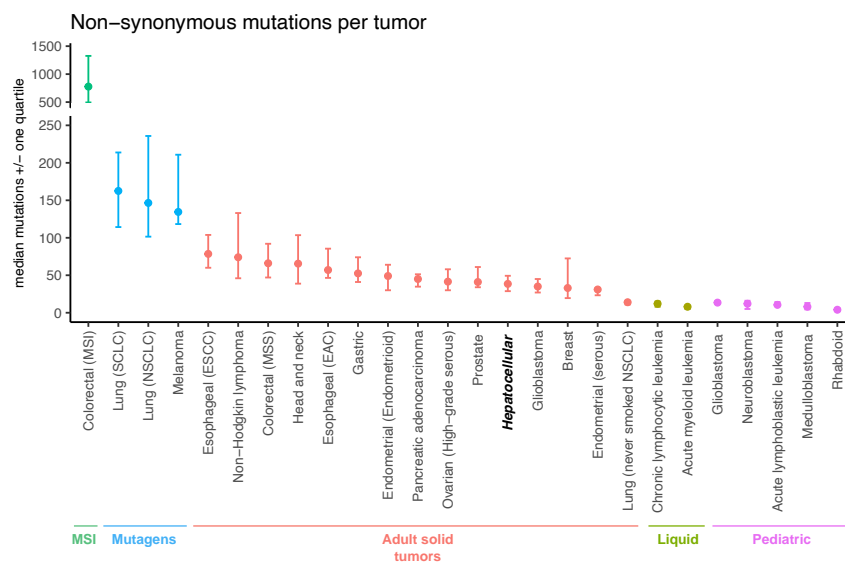
Figure 2. Number of somatic mutations in representative human cancers. The median number of non-synonymous mutations per tumor in a variety of tumor types. Horizontal bars indicate the 25 and 75% quartiles. MSI, microsatellite instability; SCLC, small cell lung cancers; NSCLC, non–small cell lung cancers; ESCC, esophageal squamous cell carcinomas; MSS, microsatellite stable; EAC, esophageal adenocarcinomas. Data obtained from Vogelstein B et al., Science 2013[12].

- *Copy-Number Variations (CNV):* These changes refer to an increase (copy-number gain) or decrease (copy-number loss) in the number of gene copies present in the normal diploid genome[15]. According to its size, a CNV can be classified as **broad CNV**, spanning the length of a chromosome arm or more; or **focal CNV**, which affects shorter regions. Adult solid tumors present a median of 11 focal amplifications and 12 focal deletions[16]. Copy number gains may range from one copy to several extra copies (high-level amplifications)[15]. In terms of copy-number losses, the deletion might generate the loss of one of the gene alleles (**loss of heterozygosity**) or both alleles (**homozygous deletions**).

- Insertional mutagenesis: This is the phenomenon by which an exogenous DNA sequence integrates within the human genome, resulting in the deregulation of neighboring genes that can potentially lead to oncogenic development[17]. Such insertional mutations can be mediated by **viruses**. Although not all viral integrations are oncogenic, they can lead to cell transformation if the insertion occurs in an essential gene or a gene that is involved in cellular replication or programmed cell death. In this regard, viruses such as human papilloma virus, Epstein Barr virus, hepatitis B virus (HBV), human T lymphotropic virus 1, and human herpes virus 8, are known to contribute to the genesis of one or more types of cancer[15].

- Chromosomal *rearrangements:* Rearrangements originate due to a DNA break that is then rejoined to another DNA segment that was not originally contiguous[15], which results in the formation of **gene fusions**. Fusions can exert tumorigenic potential due to overexpression of one of the involved genes or the generation of a hybrid gene. The first gene fusion described was the so-called "Philadelphia chromosome'' in chronic lymphocytic leukemia, where a rearrangement between *BCR* (chromosome 22) and *ABL* (chromosome 9) genes renders constitutive activation of ABL leading to a deregulated cell cycle function[15].

1.3.3. Epigenetic alterations

Epigenetic alterations are heritable changes in gene expression that are not due to modifications in the DNA sequence. The most widely studied epigenetic modifications in cancer are the following:

- *DNA methylation*: The addition of methyl groups in the DNA molecule is a mechanism of gene expression regulation that occurs normally but can be deregulated in cancer. This modification appears almost exclusively in the context of cytosine-guanine (CpG) dinucleotides, which tend to cluster in regions called **CpG islands**. In most cases. methylation in CpG islands is associated with gene silencing. Cancer cells are characterized by a massive global loss of DNA methylation (20–60% less overall 5-methyl-cytosine). However, hypermethylation at the CpG islands of certain promoters is also frequent[18].

- *Nucleosome positioning*: A nucleosome is a segment of DNA that is wrapped around a core of proteins, which blocks the access of activators and transcription factors. Occlusion of the transcription start site due to aberrant nucleosome positioning leads

to transcription repression. Nucleosome remodeling complexes SWI/SNF are commonly altered in cancer, although in most cases, the molecular mechanisms underlying their function remain unclear[18].

- *Histone modifications*: Histone tails are subject to post-transcriptional modification occur such as acetylation, methylation, phosphorylation, and ubiquitination among others. As a consequence, nucleosome architecture is altered, thus exposing or hiding specific regions of the DNA for transcriptional regulation, which ultimately affect gene expression. A common hallmark of human cancer is the reduction in the acetylation and methylation of **histone H4**, which is associated with the hypomethylation of DNA repetitive sequences[18].

1.3.2. Mutational Signatures

Despite most somatic mutations in cancer do not have a functional role, they can be extremely informative for the mutational processes that have been operative over the lifetime of a patient, including exogenous or endogenous mutagens, errors of the DNA replication machinery, and defective DNA repair. Each process generates unique **nucleotide substitution patterns** on the cancer genome, termed "mutational signatures" (**Figure 3**). Consequently, the presence of mutational signatures provides relevant information that helps unveil the genetic factors and environmental exposures in cancer[19].
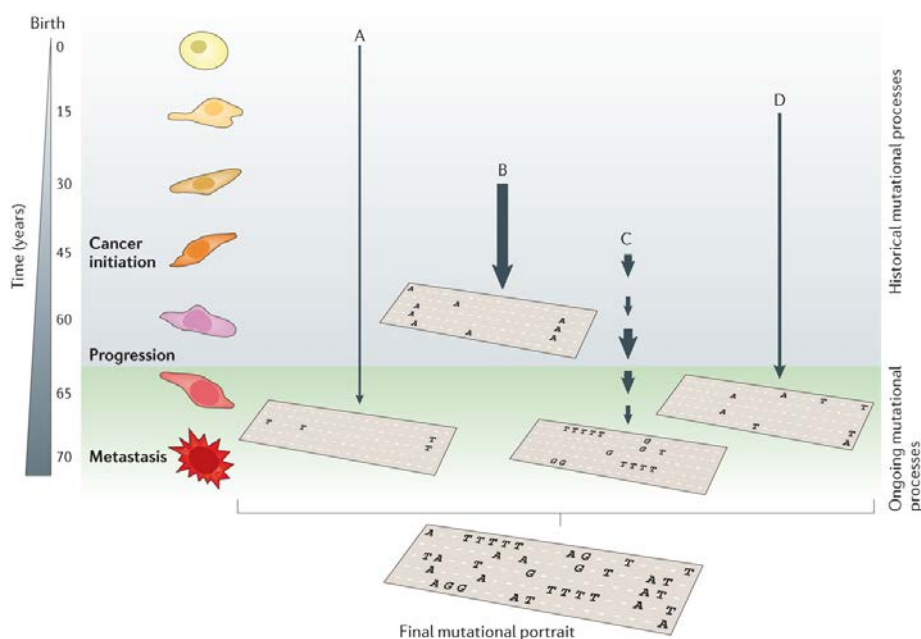
Figure 3. Active mutational processes over the course of cancer development.
Each mutational process leaves a characteristic imprint — a mutational signature

— in the cancer genome. In this hypothetical cancer genome, arrows indicate the duration and intensity of exposure to a mutational process. The final mutational portrait is the sum of all the different mutational processes (A–D) that have been active in the entire lifetime. Extracted from Helleday T et al., Nat Rev Genet 2014[19].

The link between **environmental exposures** and cancer has been known since the 18th century, when English physicians observed an increased incidence of nasal polyps among users of snuff and associated scrotal cancer in chimney sweeps with chronic exposure to soot[20,21]. Subsequent associations between environmental agents and tumorigenesis include tobacco smoking and lung cancer, aniline dyes and bladder cancer, asbestos and mesothelioma, aflatoxin and liver cancer, and benzene and leukemia among many others. Many of these agents cause DNA damage that results in **carcinogen-specific mutational signatures** (**Table 1**). For instance, C·G → A·T transversions are related to smoking in lung cancer samples, and C·G → T·A transitions are significantly over-represented in skin cancers related to UV light exposure[19] (**Figure 4**). Furthermore, the trinucleotide context of a mutation (i.e., the bases located 5' and 3' of the mutated nucleotide) is known to affect mutation rates in the genome and must also be taken into consideration when defining a mutational signature. Combining the six possible substitutions (C·G → A·T, C·G → G·C, C·G → T·A, T·A → A·T, T·A → C·G, and T·A → G·C) together with their trinucleotide contexts, all SNVs can be classified into 96 combinations. This classification has been used to extract multiple different mutational signatures in cancer[21–23]. Collective efforts analyzing the mutational fingerprints in thousands of tumor samples have allowed the identification of more than 40 mutational signatures, comprised in the COSMIC catalog[23] (**Figure 4**). Also, experimental analyses in cancer models have unveiled more than 50 additional signatures associated with exposure to known environmental carcinogens[21].

The presence of these mutational signatures in cancer serves as an archaeological record of the multiple mutational processes that have been operative, including the exposure to environmental agents[24,25]. In addition, some mutational signatures are associated with a distinct clinical outcome and are emerging as potential biomarkers for novel targeted therapies[24].
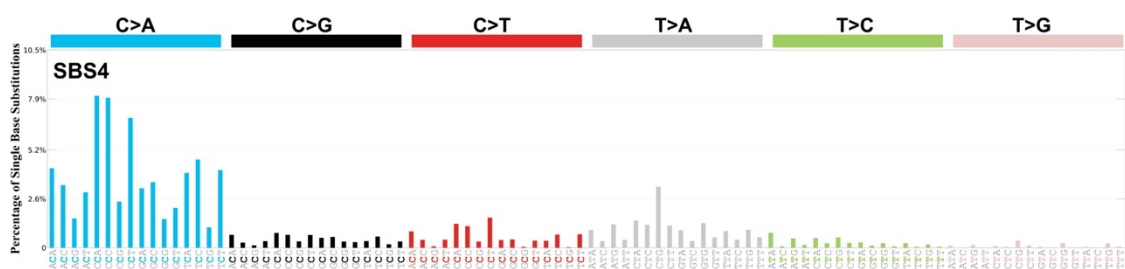
Figure 4. Example of mutational signature in cancer. COSMIC Signature 4, associated with tobacco smoking, is displayed according to the 96 substitution

classification defined by the substitution type and trinucleotide context. The probability bars for the six types of substitutions are displayed in different colors. The mutation types are on the horizontal axes, and vertical axes depict the percentage of mutations attributed to a specific mutation type. Modified from Alexandrov LB et al., Nature 2013[22].

Table 1. Mutational signatures included in the COSMIC catalog associated with known or suspected carcinogens in humans.

Signature	Carcinogen	IARC category	Main tumor type
SBS4	Tobacco smoking	1	Lung, liver, head
SBS7a-d	Ultraviolet light	1	Skin, head
SBS11	Temozolomide (alkylating agent)	1	Central nervous system, pancreas
SBS16	Alcohol	1	Liver, head
SBS18	Reactive oxygen species	3	Stomach, myeloid, colorectal
SBS22	Aristolochic acid	1	Liver, kidney, biliary
SBS24	Aflatoxin	1	Liver, biliary
SBS29	Tobacco chewing	1	Liver, kidney, bone
SBS31	Platinum treatment	2A	Liver, myeloid
SBS32	Azathioprine	1	Biliary, myeloid

SBS, single base substitution; IARC, International Agency for Research on Cancer. Data obtained from Alexandrov LB et al., Nature 2020[23] and the IARC Monograph on the Identification of Carcinogenic Hazards to Humans[26].

1.4. Role of the Immune System in Cancer

The immune system is a complex network composed of immune cells and lymphoid organs that defend the body against microorganisms – such as bacteria, viruses, and parasites – or cancer cells. In the cancer setting, mutations that accumulate during tumorigenesis give rise to tumor-specific antigens. The function of the immune system is to detect and eliminate cells producing tumor antigens[27].

Immunity can be classified into innate and adaptive, which cooperate to detect and eliminate non-self bodies[28]. Immune cells from both the adaptive and innate immune systems are present in the tumor microenvironment and contribute to the modulation of tumor progression. The **innate response** is the first line of defense and involves cells ready to face foreign bodies. It exerts a tumor suppressive function, either by directly killing tumor cells or by triggering adaptive immune responses. Innate immune cells include natural killer (NK) cells, eosinophils, basophils, and phagocytic cells such as mast cells, neutrophils, monocytes, macrophages, and

dendritic cells[28,29]. Other components of the innate immune system are plasmatic proteins (complement system) and epithelial barriers, with fewer implications in cancer.

The **adaptive response**, on the other hand, is highly specific and capable of discriminating between healthy and cancerous tissues. It is initiated when the innate immunity is not able to clear the non-self bodies (i.e., infection or cancer). This leads to the triggering of warning signals that activate the immune adaptive machinery through antigen presentation by the dendritic cells. The main players in adaptive immunity are **lymphocytes**, including B cells and T cells. B cells participate in humoral immune responses through the generation of antibodies, whereas T cells are mostly involved in cell-mediated immunity[28]. Among those, CD8$^+$ cytotoxic T lymphocytes and T helper lymphocytes are the most specialized antitumor cells[30].

2. Translational Research in Cancer

2.1. Translational Research

Translational medicine is an interdisciplinary branch within the biomedical field that aims to translate discoveries and technologies from basic research to the clinical setting in order to improve the prevention, diagnosis, and treatment of human diseases[31,32] (**Figure 5**). To pursue this goal, the translational medicine field incorporates the use of cutting-edge technologies, implements interdisciplinary approaches, and encourages collaboration between institutions. Translational medicine is based on the feedback loop "**from bench to bedside and back again**" which allows complementing the search for a cure with the understanding of the biological basis of a disease (**Figure 5**). Specifically, the key elements of this discipline are[33]:

- **Basic science** studies defining biological aspects of a disease and effects of therapies.
- Human research to define the molecular basis of a disease and provide the rationale for the development of therapeutics for human diseases.
- **Pre-clinical experimental studies** that develop principles for application of therapies to human disease or aimed at advancing the research of bringing therapies to the clinic.
- Appropriately designed **clinical trials** originated from the above-mentioned studies with efficacy or toxicity as an endpoint.
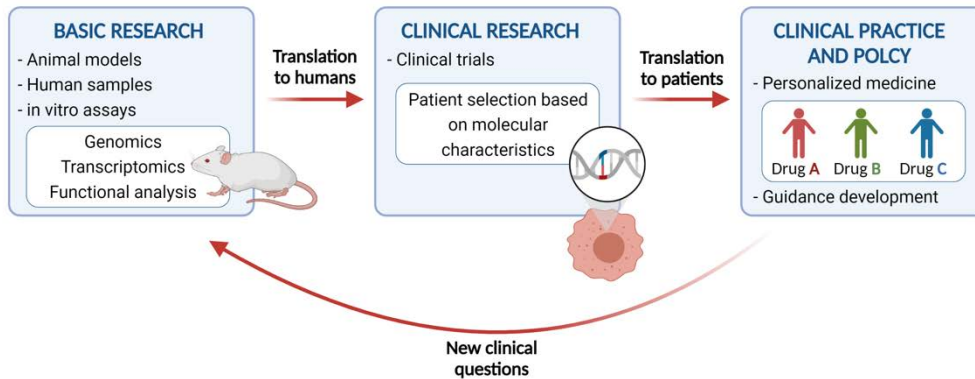
Figure 5. The translational medicine pipeline. Basic research discoveries are used in the design of clinical trials, which in turn will impact patient management policies. Adapted from Torrens L et al., Overview of Translational Medicine. Handbook of Translational Medicine, 2016[33].

Translational studies in the oncology field have led to the identification of driver genes and targetable molecular alterations which eventually have brought to the development of targeted therapies with proven anti-tumoral efficacy and subsequently improved survival of cancer patients. These discoveries also have implications for the identification of patients who are at high-risk for certain solid tumors or who may benefit from specific treatment approaches[34]. In this regard, the identification of targetable alterations is paving the way for the use of **molecular therapies** and, consequently, standard-of-care treatments are moving from a clinical-pathological basis to a targeted approach based on the molecular characteristics of the tumors. Targeted therapies are designed to disrupt specific molecular targets in tumor cells, thus increasing specificity and decreasing toxicity[35]. For instance, this is the case of FDA-approved gefitinib, a small molecule that specifically blocks the EGF receptor thus providing better progression-free survival than chemotherapy to lung cancer patients with EGFR mutations[36]. In addition to this, translational research studies have elucidated the mechanisms that regulate the anti-tumor immune response, enabling the development of **immunotherapies** in oncology. Such therapies, designed to boost the action of the immune system against cancer cells, are showing unprecedented clinical benefits in terms of tumor regression and long-term responses in advanced cases[37].

The implementation of precision medicine requires the use of **biomarkers** as tools for the selection of patients who might benefit from therapies (**predictive biomarkers**) or as indicators of the long-term outcome of the patient (**prognostic biomarkers**). Biomarkers are molecular or cellular markers that can be measured in tumor tissue or body fluids for a wide variety of materials (e.g., DNA, methylation, RNA, protein, and circulating tumor cells)[38]. Their use as

surrogate markers of a pathogenic process or drug efficacy is changing the paradigm of patient management, allowing the implementation of targeted therapeutic approaches. Clear examples of this are the detection of *ALK* fusion protein in non-small cell lung cancer (targetable by crizotinib) or *BRAF* mutations in melanoma (targetable by vemurafenib), which enable treating only patients that will benefit from a given therapy[33].

Overall, by impacting on the diagnosis, prognosis, and patient treatment, translational medicine has enabled the development of more **individualized treatments** for cancer patients[33]. Because of this, precision medicine is already a reality in some tumor types such as melanoma, small-cell lung cancer, or colorectal cancer[39].

2.2. Commonly Used Tools for Cancer Research

2.2.1. Omic techniques

Translational research has revolutionized clinical management in oncology by providing an in-depth understanding of the molecular basis of cancer. This has been largely achieved thanks to the use of "**omic techniques**". Omics refers to the collection and analysis of large data sets of biological variables and is regularly applied to genes (genomics), RNA molecules (transcriptomics), proteins (proteomics), and their subunits, such as nucleotides and amino acids (metabolomics)[40]. These new technologies present three clear advantages: a) they work at a **high-throughput level**, thus presenting an increased speed for tumor profiling; a) they allow researchers to obtain information of **multiple data points** of a biological sample in a single analysis; and c) they reduce the **cost** per analyzed unit[40,41]. A clear example of the implementation of omics is the establishment of two major international collaborative projects – The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC)[42] – that have achieved the molecular profiling of thousands of human samples from the 38 most common cancer types[14].

Genomic analysis

The study of the cancer genome comprises the characterization of mutations, copy number variations, and chromosomal rearrangements. Before the arrival of **Next Generation Sequencing** (NGS) technologies, the techniques used to study all these structural genomic aberrations were limited to the characterization of a single genome region per analysis, mainly

by Sanger sequencing for mutations and FISH karyotyping for CNVs. In contrast, NGS technologies allow for massively parallel sequencing reactions at the same time[41]. This includes a step of DNA fragmentation, library generation for DNA amplification, and the use of potent sequencer machines able to generate an output consisting of a collection of DNA sequences for each spatially separated cluster of DNA fragments[43]. Compared to previous methods, NGS allows obtaining multiple readings for the same nucleotide. The number of times that a single position of the genome sequence is read is referred to as **depth** of sequencing[41]. According to the region of interest that the sequencing technology covers, we can distinguish different assays, such as **whole-genome sequencing** (WGS), **whole-exome sequencing** (WES), and **targeted sequencing**[19]. These sequencing approaches yield many thousands of genomic alterations that can be used to study the mutation profile of each patient and to extract mutational signatures. Alternatively to NGS, single-nucleotide polymorphism arrays (**SNP arrays**) are also a common high-throughput tool for genome-wide CNV characterization[44]. Contrary to NGS, SNP arrays only provide nucleotide information at specific positions instead of all nucleotides of an aligned region. Hence, arrays offer lower resolution and imprecise delineation of the breakpoints of CNVs, than NGS[44].

The challenging step of genomic analysis is now to translate this knowledge into clinically-meaningful information, such as the identification of causal biological changes that drive cancer phenotypes or to guide the selection of therapy for patients[45]. NGS has become the foundational technology for modern diagnostic testing in oncology, with several laboratory-developed tests recently achieving recognition by the US Food and Drug Administration (FDA)[45].

Transcriptomic analysis

Tumor phenotypes are determined by aberrant transcriptional patterns, which are the consequence of somatic mutations and epigenetic alterations[46]. In contrast to DNA, which is mostly identical across all cells of an organism, the transcribed RNA is highly dynamic, reflecting the diversity of cell types, cellular states, and regulatory mechanisms. Therefore, the transcriptome profile of a tumor sample can be regarded as a snapshot of its underlying biological status[46].

Transcriptomic studies also draw upon NGS strategies. Using the same technology as genomic approaches, **RNA sequencing** (RNA-seq) is based on the sequencing of cDNA derived from messenger RNA, total RNA, and microRNAs among others[47]. RNA-seq can be used to assess the expression level of genes, study their association with a given phenotype and compare the

expression levels between phenotypes. It also provides information about splicing variants, allelic expression, and RNA editing[47]. Alternatively, the transcriptomic profile of tumors can be assessed with **microarray-based techniques**. Microarrays consist of a collection of microbeads containing DNA probes corresponding to known sequences. RNA is isolated from the control and the target samples, undergo reverse transcription, and labeling and then cDNA is hybridized to the array. The abundance of hybridization is quantified by fluorescently labeled probes, which is a readout of the RNA expression levels. Despite microarrays do not allow the detection of gene fusions or non-previously known transcripts, they are a widely used tool due to their lower cost.

Gene expression profiling can be used to identify cancer biomarkers and aberrant expression of cancer drivers, as well as to unveil molecular targets for anticancer therapies[48]. In addition, during the past 20 years, gene expression profiling has focused on the identification of clinically useful molecular signatures. **Gene expression signatures** are alterations in the expression of sets of genes with an association with disease prognosis, therapeutic benefit, or molecular profile[48]. Signatures showing activation of specific signaling pathways or cell functions are tools for molecular classification of tumors, such as the well-established five classes of breast cancer, with biological and clinical relevance[40]. The combination of tumor transcriptomic profiling and novel analysis algorithms allow the assessment of the relative cell composition of a tumor. In this regard, transcriptome profiling can be used to infer the immune-cell composition and to classify tumors according to their immune profile, which could have implications for the prediction of immunotherapy responses[49,50]. Furthermore, RNA-based prognostic panels are now available and are clinically used for major cancer types, including breast (MammaPrint, Oncotype DX, and Prosigna), lung (GeneFx), prostate (Prolaris), and colon (ColoPrint)[46].

2.2.3. Preclinical Models of Cancer

Preclinical models aim at understanding the pathogenesis and mechanisms involved in disease initiation and progression and are essential to building the groundwork for the development of clinical therapies[51]. A significant amount of our current understanding of the pathophysiology, prevention, and treatment of cancer is based on preclinical studies that exhibit some of the clinical or molecular features observed in humans. Unlike the processing of human data, these models allow the generation of mechanistic studies. Among other merits, preclinical models of cancer have achieved the identification of oncogenic mutations and pathways involved in tumorigenesis and have provided important clues regarding targetable alterations for cancer treatment.

In Vitro Experimental Models

In vitro models allow researchers to recapitulate aspects of tumor biology using specific cell types, extracellular matrices, and/or soluble factors outside of a living organism. **Cancer cell lines** are, by far, the most used model in cancer research and a particularly useful tool for screening of anticancer compounds[51]. The main advantage of cell lines resides in the fact that they allow a thorough control of experimental conditions and provide good reproducibility of results. However, they fail to recapitulate the characteristics of a whole tumor, including genomic heterogeneity of tumor cells and the presence of a tumor microenvironment. A popular alternative to cell lines are **organoids**, which refer to three-dimensional tissue culture that recapitulate the tumor architecture and maintain multilineage differentiation. Currently, great efforts are being invested in developing tumor organoids that maintain a functional immune microenvironment.

In Vivo Experimental Models

Laboratory models typically involve animals in which processes that are present in human diseases are either induced or occur naturally. The most widely used animal in cancer research is the **mouse** (*Mus musculus*). Remarkably, 90% of the mouse genome can be linked up with a region of the human genome and 99% of the mouse genes have analogs in humans. Other advantages of the murine model include rapid tumor development, low housing cost, and good-sized litters[51].

Murine models of cancer encompass a wide spectrum of options. The selection of the optimal model depends on the biological question that needs to be tackled and often represents a compromise between time, complexity, and clinical relevance[51].

- *Xenograft models*: The xenograft model is based on the **implantation of tumor cell lines** onto mice to grow tumors of human origin. this implantation can be produced either **orthotopically** (i.e., in the same organ as the tumor would develop) or **subcutaneously**. This model is relatively inexpensive, highly reproducible, and applicable to many cancer types[51]. Furthermore, subcutaneous models are useful in studying the response to therapy, as tumors are easily accessible, and growth can be monitored with serial tumor measurement. The main limitation of this model is that it requires immunocompromised mice to avoid rejection of the foreign tissue[52].

- *Syngeneic models*: This model utilizes mouse tumor cells implanted orthotopically or subcutaneously in recipient mice with a fully **functional immune system**[52]. Therefore, it can mimic a comprehensive antitumor immune response and recapitulate its implications in tumor development and response to treatment. This is especially relevant when assessing the anti-tumoral effect of drugs targeting the anti-tumor immune response (i.e., immunotherapies). However, it might be less representative of the human disease due to the use of mouse tumors.

- *Patient-derived xenografts (PDX)*: The PDX model consists in the implantation of **primary human tumor** samples into immunocompromised mice. Therefore, they retain the architecture and microenvironment of the primary human tumor and provide an accurate representation of the original molecular landscape and heterogeneity. Similar to the conventional xenograft, its main limitation is the lack of a fully-functional immune response[51].

- *Genetically engineered mouse models (GEMMs)*: Advances in genetic engineering have led to the genetic program of organ-specific activation of oncogenes or inactivation of tumor suppressor genes to induce tumor formation, recapitulating the genetic lesions found in human cancers. GEMMs develop spontaneous and autochthonous tumors including tumor microenvironment and immune infiltrate[51]. GEMMs enable the study of tumorigenesis and biological tumor development, but they often involve long latency and a lower tumor incidence[52]. **Hydrodynamic tail-vein injection** of cells or DNA plasmids carrying the mutations of interest has opened a new avenue for delivering genetic materials to a mouse model[52].

- *Environmentally-induced cancer models:* Tumor development in mouse models can be achieved following various environmental exposures, including carcinogen agents, radiation, pathogenic viruses, or specific diets[53]. The type and route of administration of the carcinogen determine the location in which the tumor is formed (e.g., topical, intramuscular, or oral administration). Some models mimic the setting of chronic inflammation and organ damage which accompany or lead to tumorigenesis. This is the case of **high-fat diet** (HFD) models, which induce liver disease and steatosis, that can be easily combined with different methods to induce liver cancer (e.g., GEMM, cell implantation, or additional carcinogens). Overall, the main caveat of environmentally-induced models is that only a restricted subset of tumor types can be developed, and tumor penetrance can be often limited.

2.3. Current Challenges in Translational Oncology

Translational medicine is playing a pivotal role in broadening our knowledge about the molecular pathogenesis of cancer. The emergence of omic techniques and collective initiatives to sequence tumor samples has allowed the identification of key molecular alterations and has provided insight into the biological mechanisms involved in oncogenesis. At the same time, functional studies in preclinical models have enabled a more detailed characterization of tumor tissues. Overall, this has led to the identification of drug targets, enabling the development of novel therapies and personalized treatment approaches. In this regard, the use of biomarkers facilitates the selection of patients who would benefit from each treatment, but only a few biomarkers for solid tumors are routinely tested in the clinical setting at the moment[34]. In the future, the development of **more affordable** high-throughput sequencing approaches will allow us to increase the assessment of molecular alterations to all patients and will lead to more tailored therapies. New and **less invasive** methods for obtaining tumor material such as liquid biopsies will enable more frequent monitoring of tumor response to therapy, leading to an improved therapeutic benefit to toxicity ratio[34]. Finally, the **incorporation of novel biomarkers in clinical trials** may also improve the efficacy of experimental drug testing.

3. Hepatocellular Carcinoma

3.1. Epidemiology and Risk Factors

Primary liver cancer is the sixth most commonly diagnosed cancer and the third leading cause of cancer-related death worldwide, with approximately 906,000 new cases and 830,000 deaths in 2020. It is estimated that, by 2025, more than 1 million individuals will be affected by liver cancer annually[2,3]. In most geographical regions, rates of both incidence and mortality are 2 to 3 times higher among men than women (**Figure 6A**). The highest incidence and mortality rates are observed mainly in transitioning countries from East Asia and Africa, with **Mongolia** exhibiting the highest incidence of HCC worldwide[2] (**Figure 6B**). Nonetheless, liver cancer incidence and mortality are also increasing in different parts of Europe and the USA[54].

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer and accounts for ~90% of cases[2,3]. HCC typically arises in the setting of **chronic liver damage** generated by known risk factors[2,3]. The worldwide prevalence of such factors is highly heterogeneous, which

determines the global distribution of HCC cases (**Figure 6B**)[2,3,55]. **Well-established risk factors** of HCC are:

- *Hepatitis B virus infection*: HBV is a DNA virus that can integrate into the host genome, leading to oncogene activation. HBV infection is the main etiologic agent, in Asia, Africa, Melanesia, and Polynesia[2,3,55].

- *Hepatitis C virus infection:* Unlike HBV, hepatitis C virus (HCV) is an RNA virus that does not integrate into the host genome. Therefore, the risk of HCC in HCV-infected patients is due to the development of cirrhosis or chronic liver damage[3]. HCV infection is the predominant etiology in the North of Africa, Europe, North America, Japan, and Central Asia[2,3,56].

- *Alcohol abuse:* Excessive alcohol intake causes alcoholic liver disease, cirrhosis, and HCC. In addition, there is growing evidence supporting alcohol-specific protumorigenic mechanisms besides cirrhosis[3]. Alcohol consumption is the second most common risk factor in most European countries and also in North America[2,3].

- *Non-alcoholic fatty liver disease:* Non-alcoholic fatty liver disease (NAFLD) is a spectrum of chronic liver diseases normally occurring in patients with diabetes mellitus and obesity. NAFLD ranges from excessive hepatocyte triglyceride accumulation and steatosis (nonalcoholic fatty liver) to hepatic triglyceride accumulation plus inflammation and hepatocyte injury (nonalcoholic steatohepatitis [NASH]), which finally leads to hepatic cirrhosis and HCC[57]. NAFLD is the fastest growing etiology of HCC, particularly in Western countries [2,3,55].

Other **emerging factors** have been reported to contribute to disease risk, such as:

- *Hepatitis D virus infection (HDV):* HDV is an RNA virus that requires the presence of HBV surface antigens for its replication and infectivity. Despite being associated with a more severe course of liver disease and increased risk of HCC compared with HBV infection alone[58], HDV is not currently included in the roster of carcinogenic agents since data supporting this relationship remains scarce[26,59].

- *Adeno-associated virus type 2 (AAV2):* Adeno-associated virus (AAV) is a defective DNA virus that causes frequent non-pathogenic infections in the general population[60]. However, recurrent clonal integrations of certain AAV strains – mostly AAV2 – have been

identified in a subgroup of HCC (~5%)[3,4], and preclinical data points towards HCC development after AAV2 infection in murine models[3,4], suggesting that this virus could be hepatocarcinogenic in certain circumstances.

Finally, several carcinogenic **cofactors** can exacerbate the risk of HCC in patients exposed to these factors[61]. For instance, the fungal compound **aflatoxin B1**, a common food contaminant in Southeast Asia and Sub-Saharan Africa, has been reported to act synergistically with HBV to induce HCC. Other known cofactors in HCC are **aristolochic acid**, which is used in Asian traditional medicine, and **tobacco** smoke[3,61].
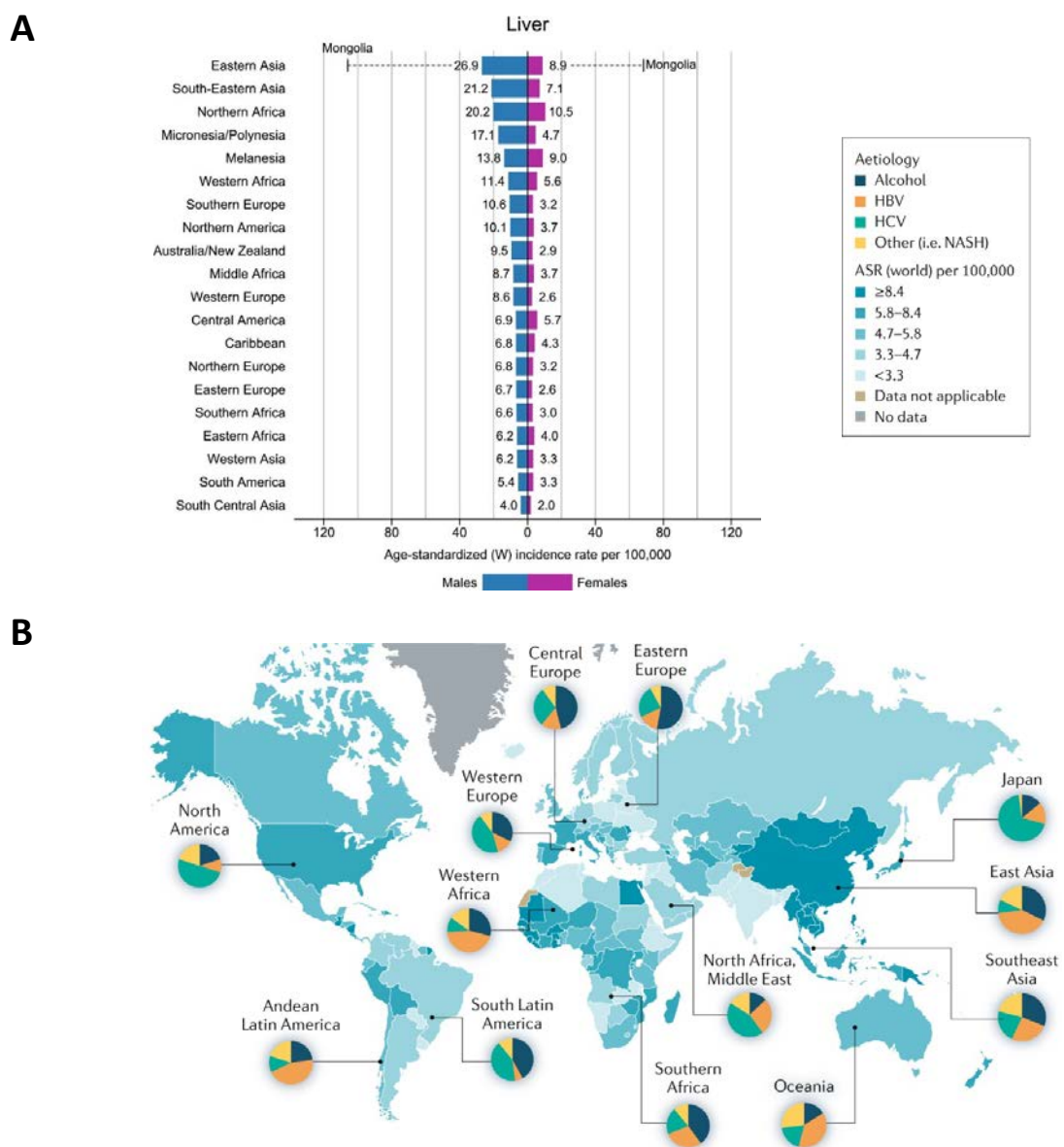
Figure 6. Age-standardized incidence rates in liver cancer. A-B. Age-standardized incidence by sex (**A**) and geographical area (**B**). The major etiological factors involved in hepatocarcinogenesis are depicted in figure B. NASH, nonalcoholic

steatohepatitis; ASR, age standardized rate. Modified from Llovet JM et al., Nat Rev Dis Prim 2021[3] and Sung CA et al., Cancer J Clin 2021 (Global Cancer Statistics 2020)[2].

3.1.1. Particularities of HCC in Mongolia

Mongolia, a landlocked East Asian country between Russia and China, shows the **world's highest incidence** of HCC, with a burden of 85.6 cases per 100,000 inhabitants (106.0 and 68.4 cases per 100,000 inhabitants in males and females, respectively) [62]. This incidence far exceeds that of the surrounding countries such as China, with 18.2 cases per 100,000 inhabitants (5-fold higher in Mongolia) and Russia, with 4.2 (>20-fold higher in Mongolia), or any other country worldwide[62,63]. The extremely high incidence observed in Mongolia has been largely attributed to the unique combination of HCC risk factors, with a historical high prevalence of both **HBV** (10.6%) and **HCV** (6.4%) viruses, and **alcohol** consumption[56,64,65]. Indeed, 90% of Mongolian HCC cases are positive for HBV, HCV, or both[66], and **HDV** is present in 50-80% of Mongolian HBV-infected individuals[67,68]. In this regard, despite the implementation of universal infant HBV vaccination in 1991, which achieved a marked decline in HBV prevalence[64], and the higher control in HCV blood-based transmission, the burden of HCC in Mongolia is increasing year after year, and the incidence is now ~10 times higher compared to the 1960s[69,70]. Other reasons such as **low screening** and treatment rates due to **financial barriers** have also been proposed, suggested by the fact that 75% of cases are diagnosed at advanced stages and have poor 1-year survival of less than 25%[69,71]. However, it is likely that earlier diagnosis would even increase the actual burden, as probably many cases remain undiagnosed, especially in remote areas.

Besides the abovementioned risk factors, Mongolia has many particularities that might eventually play a role in HCC burden. It is a huge but sparsely populated country (1.9 people per km^2), but half of its 3-million population lives in Ulaanbaatar, an overpopulated capital with **dismal environmental conditions**, whereas the rest is still predominantly nomad. Mongolian individuals are descendants of the Genghis Khan Empire, which spread from East Asia to Europe during the 13th century[72]. This explains why modern Mongolian populations have a certain degree of genotypic similarity with Siberian, Finn, Chinese Han, and Japanese populations, and are more similar to European than to the majority of Asian populations[73]. Interestingly, its neighbor countries also show a high prevalence of HCC risk factors (in Russia HBV is 2%, HCV is 2.7%, and alcohol consumption is a major issue; and in China, HBV prevalence is 6.1%), but, as stated, its HCC burden is much lower[56,64]. Around 60% of the 10 million Mongols live in the Northern region of China called Inner Mongolia, representing 17% of the population in that

region. Interestingly, this area shows a prevalence of HBV infection below 4% and a low incidence of HCC[74,75].

Regarding HCC onset, the **male/female ratio** in Mongolia is 1.5/1, as opposed to that observed in the surrounding countries (2.6/1 in Russia, 3.4/1 in China, and 3/1 in East Asia)[76], which could potentially be linked to distinct patterns of exposure to HCC risk factors in Mongolia compared to other populations. Overall, it is unclear whether HCC incidence in Mongolia is completely explained by the unique combination of risk factors, or eventually, **other genetic or environmental factors** might play a role.

3.2. Molecular Pathogenesis of HCC

HCC development is a complex **multistep process** involving the interplay between **genetic** factors and **environmental** exposures[77,78]. The combination of these factors triggers the transformation of the normal liver to an inflammatory and fibrogenic procarcinogenic field, which constitutes the background for HCC development[77]. indeed, 70–80% of HCC cases occur in the context of established liver cirrhosis, the last stage of this underlying **chronic liver disease**[79]. The natural history of HCC in cirrhosis follows a sequence of events starting with pre-cancerous cirrhotic nodules, called low-grade dysplastic nodules (LGDN) and high-grade dysplastic nodules (HGDN), which can finally transform into HCC[80]. However, HCC might also develop in the context of chronic liver disease without cirrhosis. For instance, HBV is able to insert into the genome in cancer genes, triggering the formation of HCC without the need for a cirrhotic background[78]. In normal liver, HCC can also arise from the malignant transformation of hepatocellular adenoma, a rare benign lesion[81].

3.2.1. Molecular Landscape

Over the last decade, translational genomic studies have provided an overview of the molecular landscape of HCC[82–86]. HCC tumors present 40-60 non-silent somatic mutations accumulated in coding regions[78,83] (**Figure 2**). The **most commonly mutated driver genes** include *TERT*, *CTNNB1*, and *TP53* (~55%, 29%, and 27% of patients, respectively) (**Table 2**)[78,83]. **CNAs** are also recurrent in HCC, with 21-30 focal alterations per tumor[87]. Characteristic focal CNAs in HCC include gains in 8q24.21 involving *MYC* (12%) and 11q13.3 affecting *CCND1* and *FGF19* (6-7%), as well as losses in 9p21.3 affecting *CDKN2A* (**Table 2**)[83,87]. Finally, few recurrent fusion proteins have been described in HCC, including the *ABCB11- LRP2* fusion in only 2% of the patients[88].

INTRODUCTION

Each HCC tumor is a unique combination of genetic and epigenetic alterations, underlining the complexity and diversity in HCC. As a result, the major **signaling pathways** disrupted in HCC progression include the following:

Table 2. Recurrent somatic driver alterations in resected HCCs.

Altered pathway	Altered gene	Type of alteration	Percentage (range)
Mutations			
Telomere maintenance	*TERT$^\$$*	Promoter Activating mutation	55 (44–59)
Cell cycle regulation	*TP53$^\$$*	Loss of function mutation	27 (18–31)
	ATM	Loss of function mutation	4 (2–5)
	RB1	Loss of function mutation	4 (3–5)
	CDKN2A	Loss of function mutation	2 (1–3)
Wnt / β-catenin signaling	*CTNNB1$^\$$*	Activating mutation	29 (23–36)
	AXIN1	Loss of function mutation	7 (4–10)
	APC	Loss of function mutation	2 (0–3)
Chromatin remodeling	*ARID1A*	Loss of function mutation	8 (4–12)
	ARID2	Loss of function mutation	7 (3–10)
	KMT2A	Loss of function mutation	3 (0–4)
	KMT2C	Loss of function mutation	3 (2–5)
	KMT2B	Loss of function mutation	2 (0–4)
	BAP1	Loss of function mutation	2 (0–5)
	ARID1B	Loss of function mutation	1 (0–3)
Ras/PI3K/mTOR	*RPS6KA3*	Unclassified	4 (3–6)
	PIK3CA$^\#$	Activating mutation	2 (1–4)
	KRAS$^\#$	Activating mutation	1 (0–1)
	NRAS	Activating mutation	0 (0–1)
	PDGFRA$^\#$	Mutation	1 (0–4)
	EGFR$^\#$	Activating mutation	1 (0–2)
	PTEN	Loss of function mutation	1 (0–2)
Oxidative stress	*NFE2L2$^\&$*	Activating mutation	4 (2–6)
	KEAP1$^\&$	Activating mutation	3 (2–5)
Hepatocyte differentiation	*ALB*	Mutation	9 (5–13)
	APOB	Mutation	8 (1–10)
JAK–STAT	*IL6ST*	Mutation	2 (0–3)
	JAK1$^\$$	Mutation	1 (0–3)
TGFβ signaling$^\$$	*ACVR2A*	Loss of function mutation	4 (1–10)
IGF signaling$^\$$	*IGF2R*	Mutation	1 (0–2)
Copy number alterations			
Telomere maintenance	*TERT$^\$$*	High-level focal amplification	6 (1–9)
Cell cycle regulation	*MYC*	High-level focal amplification	12 (4–18)
	CCND1$^\$$	High-level focal amplification	7 (5–7)
	CDKN2A	Homozygous deletion	5 (4–6)
	RB1	Homozygous deletion	5 (4–6)
	TP53$^\$$	Homozygous deletion	2 (0–2)
RTK-RAS-PI3K signaling	*FGF19$^\#$*	High-level focal amplification	6 (5–6)
	VEGFA$^\#$	High-level focal amplification	5 (1–8)
Viral insertions			
Telomere maintenance	*TERT$^\$$*	HBV insertion	3 (1 – 5)
Cell cycle regulation	*CCNA2*	HBV insertion	5 (1-6)
	CCNE	HBV insertion	3 (1-6)
Chromatin remodeling	*KMT2B*	HBV insertion	1

HCC, hepatocellular carcinoma; IGF, insulin growth factor; mTOR, mammalian Target of Rapamycin; STAT, signal transducer and activator of transcription; TGFβ, transforming growth factor β. #: targetable by an FDA-approved drug. $: targetable by a drug in testing phases. &: targetable using mTOR inhibitors in testing phases. Adapted from Llovet JM et al., Nat Cancer 2021[89] and Bayard Q et al., Nat Commun 2018[90].

- *Telomere maintenance*: Approximately 90% of human HCCs harbor increased telomerase expression, the enzyme responsible for the maintenance of telomere length[3]. Telomerase prevents the erosion of the chromosomes that physiologically occur at each cell division during aging. In HCC, overexpression occurs mainly due to somatic **TERT promoter mutations** (55%)[91,92], **HBV insertion** in the promoter (3%)[93], and **copy-number amplification** (6%)[84]. While these alterations are mutually exclusive, *TERT* promoter mutations are frequently associated with *CTNNB1* mutations, suggesting cooperation between telomerase maintenance and the β-catenin pathway in liver tumorigenesis[83,91]. Remarkably, 19% of HGDN exhibit these mutations, suggesting TERT as a "**gatekeeper**" during hepatocarcinogenesis[91].

- *WNT/β-catenin signaling*: This pathway is implicated in physiologic embryogenesis, zonation, and metabolic control in the liver. **CTNNB1**, a gene coding for β-catenin, presents activating mutations in 29% of HCC. Inactivating mutations or deletions have also been identified in **AXIN1** (7%) and **APC** (2%)[82–86].

- *Cell cycle control*: **TP53** – a key tumor suppressor participating in cell cycle regulation – present inactivating mutations in 27% of HCC patients[82–86]. The only recurrent hotspot identified so far in *TP53* is R249S, related to aflatoxin exposure[94,95]. Additionally, the retinoblastoma pathway that controls progression from G1 to S phase is often inactivated in HCC mainly by homozygous deletions in **CDKN2A** (5%) or **RB1** mutations (4%), both associated with poor prognosis[82,83]. Finally, recurrent **HBV insertions** in *CCNE1* (5%)[93] and amplification of the **CCND1/FGF19 locus** (6-7%)[96,97], two key proteins involved in cell cycle progression, have been reported in HCC.

- *Epigenetic modifiers:* Mutations in epigenetic modifiers from the **SWI/SNF chromatin remodeling complex** are recurrent in HCC, including inactivating mutations in **ARID1A** (8%) and **ARID2** (7%)[82,83,98]. Less frequent mutations have been described in the histone methylation writer family, including **KMT2A** and **KMT2C** (3%)[82,83]. Furthermore, **KMT2B** can be affected by mutations (2%) or recurrent HBV insertions (10%)[82,83,93]. Altogether, the functional consequences of *ARID1A*, *ARID2*, and *KMT2* family mutations in hepatocarcinogenesis remain to be further explored.

- *Oxidative stress pathway:* The oxidative stress pathway is deregulated by activating mutations in **NFE2L2** – coding for NRF2 – and inactivating mutations in **KEAP1** in (3-4%)[82,83,98]. Interestingly, NRF2 pathway activation was previously shown to protect

against liver tumor initiation, but its constitutive activation can drive tumor progression in late-stage HCC[99].

- *Tyrosine kinase receptor-RAS-PI3K signaling*: The RAS/RAF/mitogen-activated protein kinase pathway is activated by loss-of-function mutations in ***RP6SKA3*** (4%), coding for the RAS inhibitor RSK2[98]. In contrast to other tumors, activating mutations of genes belonging to the RAS family are rarely observed in HCC (<2%)[77]. On the other hand, PI3K/AKT/mTOR signaling is stimulated by activating mutations in ***PIK3CA*** (2%) and homozygous-deletions in ***PTEN*** (1%)[82,83], as well as ***FGF19*** focal amplification (6%)[97,100]. Notably, some HCCs with activation of the PI3K/AKT/MTOR cascade have no genetic alterations in this pathway. In this sense, indirect upstream signaling through the insulin growth factor (IGF) pathway has been proposed as an alternative mechanism of activation of this cascade[77,101].

Other mutated genes of HCC belong to the including **JAK/STAT** signaling (*IL6ST* and *JAK1*, 1-2%) and **TGFβ** (*ACVR2A*, 4%) signaling pathways. Genes related to liver function such as hepatic differentiation ***ALB*** and ***APOB*** also present recurrent mutations in 8-9% of cases[79] (**Table 2**).

3.2.2. Genomic Alterations Associated with Viral Infections

Two DNA viruses – HBV and AAV2 have been reported to induce **insertional mutagenesis** leading to HCC[4,102]. On the other hand, HCV is a single-strand DNA that is unable to insert into the host DNA and does not present a clear direct mechanism of carcinogenesis[103].

- *Hepatitis B virus:* Oncogenic HBV-mediated insertional mutagenesis can occur within the *TERT* promoter, leading to an overexpression of **telomerase**[3,102]. Other recurrent HBV insertions have been mapped in several oncogenes including the cyclins *CCNA2* and *CCNE1*, which activate in **cell cycle** progression[90] (**Table 2**).

- *Adeno-associated virus* type 2*:* AAV2 infection is frequent in the human population, and no specific diseases have yet been associated with natural infection[4]. This, together with its high cell infectivity, has supported extensive development of AAV2-derived vectors for **gene therapy** for the last 30 years[4]. Indeed, AAV2 vectors have been largely considered safe, given that they generally persist episomally with infrequent integration[104]. The growing popularity of AAV2 viral vectors has resulted in the FDA and EMA approval of two gene therapies for the treatment of spinal muscular atrophy and retinal dystrophy, and ~130 active clinical trials are currently testing AAV-based gene

therapy for a great variety of human diseases[104,105]. However, mounting recent evidence supports that AAV2 vectors may have **genotoxic potential** and AAV2 insertions have been identified in a small set of HCC patients (2-5%)[4,60]. Preclinical data points towards the development of HCC after AAV transfer in murine models as a result of integration in the chromosome 12 locus that includes the noncoding RNA gene *Rian* (analogous to the human *DLK1-DIO3* locus in 14q32.2)[106,107]. Furthermore, Molecular profiling of human HCC revealed a subclass of 6-19% HCC with overexpression of this microRNA cluster, associated with an aggressive phenotype and poor prognosis[108,109]. Similar to HBV, other recurrent AAV2 insertion points also include *TERT*, *CCNA2,* and *CCNE1* genes[4]. Finally, considering that AAV2 integration capacity is enhanced in cells undergoing cell cycle progression[110], it remains to be studied whether hepatocyte replication in response to liver injury and chronic inflammation could potentially favor oncogenic AAV integration[111]. Adverse events in these patients could potentially be a concern for the use of AAV gene therapy.

3.2.3. Mutational Signatures in HCC

During the development of chronic liver disease and cirrhosis, hepatocytes progressively accumulate genetic mutations, which constitute mutational signatures that can be associated with specific risk factors. In this regard, genome and exome sequencing analyses of HCC have allowed the identification of mutational signatures from the COSMIC catalog that suggest the presence of intrinsic and extrinsic mutational processes[23,83,112] (**Figure 7**). For instance, SBS5 and SBS1, which are among the most common signatures in this tumor type, recapitulate **clock-like** endogenous mutations that occur in cell division and accumulate with age. Other signatures can be linked to specific HCC risk factors. For instance, SBS16 has been associated with **alcohol** intake, and signatures SBS4 and SBS29, with **tobacco** smoking. Interestingly, some of these signatures show clear geographical differences due to distinct exposure to known carcinogens[23]. This is the case of SBS22 and SBS24, mostly detected in HCC patients from Asia and Africa exposed to **aristolochic acid** and **aflatoxin B1**, respectively[23,83,113]. Finally, some signatures with unknown etiology can be routinely found in the HCC genome, such as SBS12 and SBS40, and future studies will be required to elucidate the mutational mechanisms causing their mutational patterns[23].

Overall, these observations align with the role of the liver in detoxifying numerous metabolites, which can damage the hepatocyte genome[3]. In this context, assessing the mutational landscape

of HCC can provide relevant information that may help unveil the genetic factors and environmental exposures underlying hepatocarcinogenesis.

Signature	Proposed etiology
SBS5	Clock-like
SBS1	Clock-like
SBS12	
SBS29	Tobacco chewing
SBS4	Tobacco smoking
SBS40	
SBS16	Alcohol
SBS22	Aristolochic acid exposure
SBS24	Aflatoxin B1 exposure
SBS26	Defective DNA mismatch repair
SBS6	Defective DNA mismatch repair
SBS31	Platinum treatment
SBS35	Platinum treatment
SBS18	Reactive oxygen species
SBS19	
SBS19	Mismatch repair deficiency
SBS30	Defective base excision repair
SBS28	
SBS9	Polymerase activity
SBS17b	
SBS17a	

Proportion of tumors with the signature
0 0.5 1.0

Median mutations per Mb due to signature (among tumors with the signature)
2.5
1
0.5
0.25
0.1
0.05
0

Figure 7. Mutational signatures identified in HCC. Single base mutational signatures from the COSMIC catalog version 3.0 reported in hepatocellular carcinoma (HCC) samples (n = 323). Signatures are sorted by frequency and median mutations per megabase (Mb) explained by the signature in positive samples. The proposed etiology associated with each signature (if any) is indicated. Modified from Alexandrov LB et al., Nature 2020[23].

3.2.4. Molecular and Immune Classes

Integrative efforts involving genomic, transcriptomic, and epigenetic data have established a molecular classification of HCC[3,77,114]. These molecular classes reflect specific genomic alterations, histopathological fingerprints, and clinical outcomes, with potential implications in patient prognostication and therapy selection. The molecular profile of HCC can be roughly divided into **two major molecular types**, each accounting for ~50% of patients with this disease – the proliferation class and the non-proliferation class[3,80] (**Figure 8**):

- *Proliferation class:* HCC tumors belonging to this class are associated with mutations in *TP53*, chromosomal instability, and enrichment in HBV-associated HCC [3,115–117]. They also present enrichment of **poor prognosis signatures** and clinical characteristics of **aggressive tumors** (e.g., vascular invasion)[3,118]. The proliferation class can be further subdivided into two subclasses[3]: 1) The **proliferation-progenitor cell group**, characterized by the activation of classic cell proliferation pathways (e.g., PI3K–AKT–mTOR, RAS–MAPK or MET) and expression of progenitor cell markers (e.g., EPCAM and

α-fetoprotein); and 2) the **proliferation–WNT–TGFβ** group, characterized by non-canonical activation of Wnt.
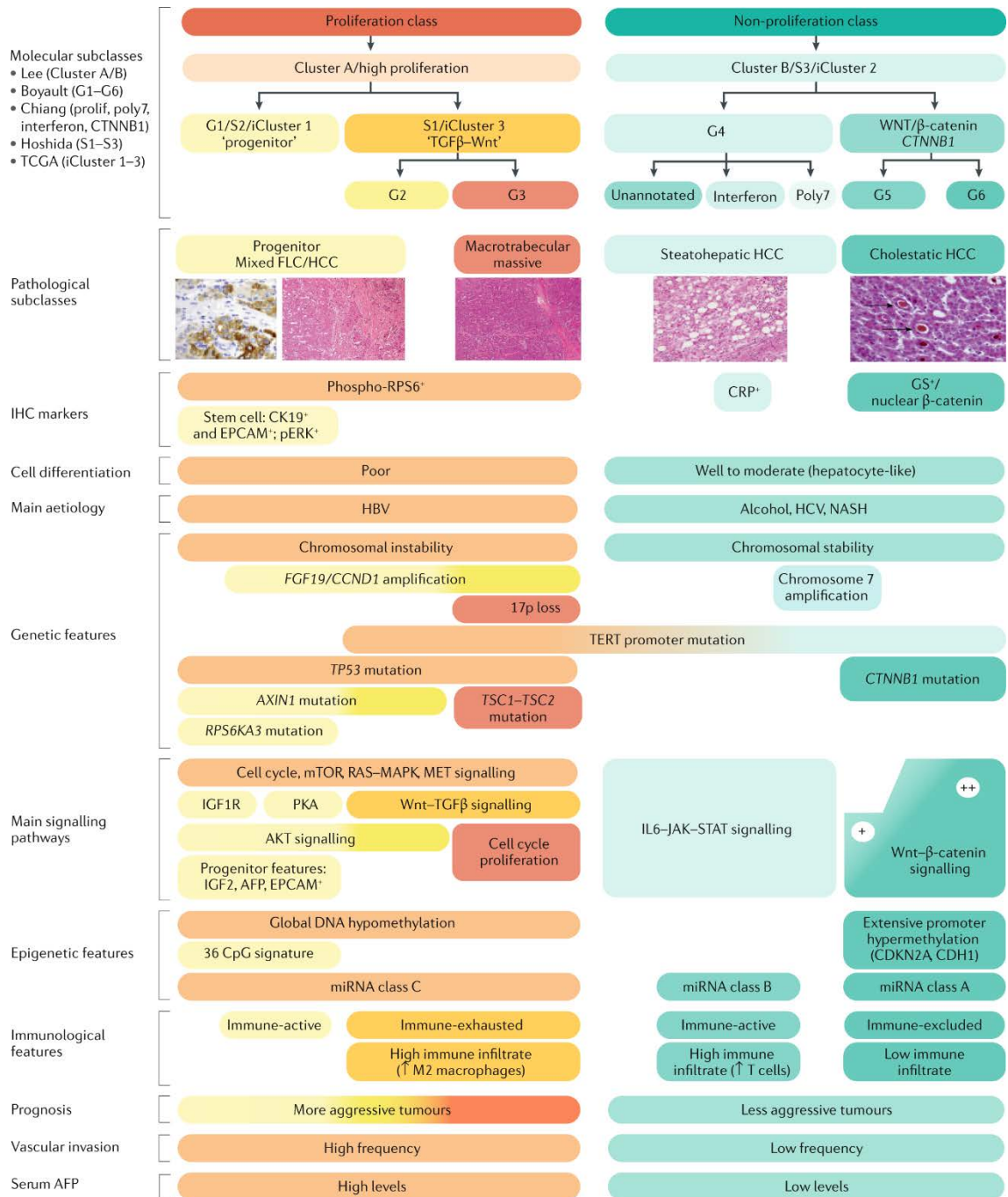
Figure 8. Molecular classification of HCC. Hepatocellular carcinoma (HCC) can be classified into two major molecular groups based on transcriptomic features, which present association with genomic, histopathological, and clinical characteristics. FLC, fibrolamellar carcinoma; IHC, immunohistochemistry; FLC, fibrolamellar carcinoma; TCGA, The Cancer Genome Atlas; IHC, immunohistochemistry; AFP; α-fetoprotein; HBV, hepatitis B virus; HCV, hepatitis C virus; NASH, non-alcoholic steatohepatitis; miRNA, microRNA. Obtained from Llovet JM et al., Nat Rev Dis Prim 2021[3].

- *Non-proliferation class*: Tumors from this class lack strong proliferative signaling and retain molecular features resembling **normal hepatic physiology** (e.g., metabolic functions and protein synthesis). This class is enriched in alcohol-associated and HCV-related HCC and is associated with better outcomes[3]. Although the non-proliferation class is heterogeneous, it can be divided into 2 main subclasses – one characterized by **canonical Wnt signaling** activation and *CTNNB1* mutations, and another characterized by the **activation of IFNα signaling**[3,100,119].

HCCs can also be classified based upon their immune microenvironment profile using transcriptomic data[120,121] (**Figure 9**). Around ~25% of the tumors belong to the **immune class**, characterized by high levels of immune infiltrate, high cytolytic activity, and expression of PD1/PDL1 immune checkpoints. This class can be further divided into **immune-active** tumors, presenting an enrichment in cytotoxic T cell infiltrate and signatures of response to immunotherapy; and **immune-exhausted** tumors, with TGFβ signaling activation driving immune exhaustion. Recently, the concept of "hot tumors" in HCC – i.e., tumors with high immune infiltration – has been further expanded with the **inflamed class** (~35% of cases), which encompasses both the immune class and the novel immune-like class[121]. Tumors from the immune-like class are dominated by high interferon-γ signaling coexisting with *CTNNB1* mutations. Whether inflamed HCCs or other immune-related biomarkers are associated with response to ICI is currently being investigated.

On the other side of the spectrum, the non-inflamed class (~65%) encompasses two further subclasses based on their mechanisms of immune escape[121]. First, the **immune excluded class** (~25%) is characterized by low immune cell infiltrate and enrichment in *CTNNB1* mutations, which has been associated with resistance to immunotherapy[122]. Finally, the **intermediate class** (~45%) presents enrichment in *TP53* mutations and frequent deletions in genomic regions harboring genes related to interferon signaling or antigen presentation.

3.3. Tumor Microenvironment

The tumor microenvironment of HCC is a complex and spatially structured mixture of tumor cells, immune cells and tumor-associated fibroblasts, and hepatic non-parenchymal resident cells. All these populations dynamically interact and influence the inflammatory profile of the tumor[123].

Approximately 90% of HCC cases are associated with **chronic inflammatory processes** due to viral hepatitis, alcohol intake, or NAFLD. Consequently, HCC is a prototypical inflammation-associated cancer, and the immune microenvironment plays a pivotal role in hepatocarcinogenesis[3,124]. Remarkably, immune-related gene expression patterns in the non-tumoral liver parenchyma have been associated with enhanced risk of HCC development in patients with cirrhosis[125]. In fully-developed HCC, the presence of an intra-tumoral immune infiltrate is associated with good prognosis, likely due to the activation of a more effective antitumor immunity[3,126].
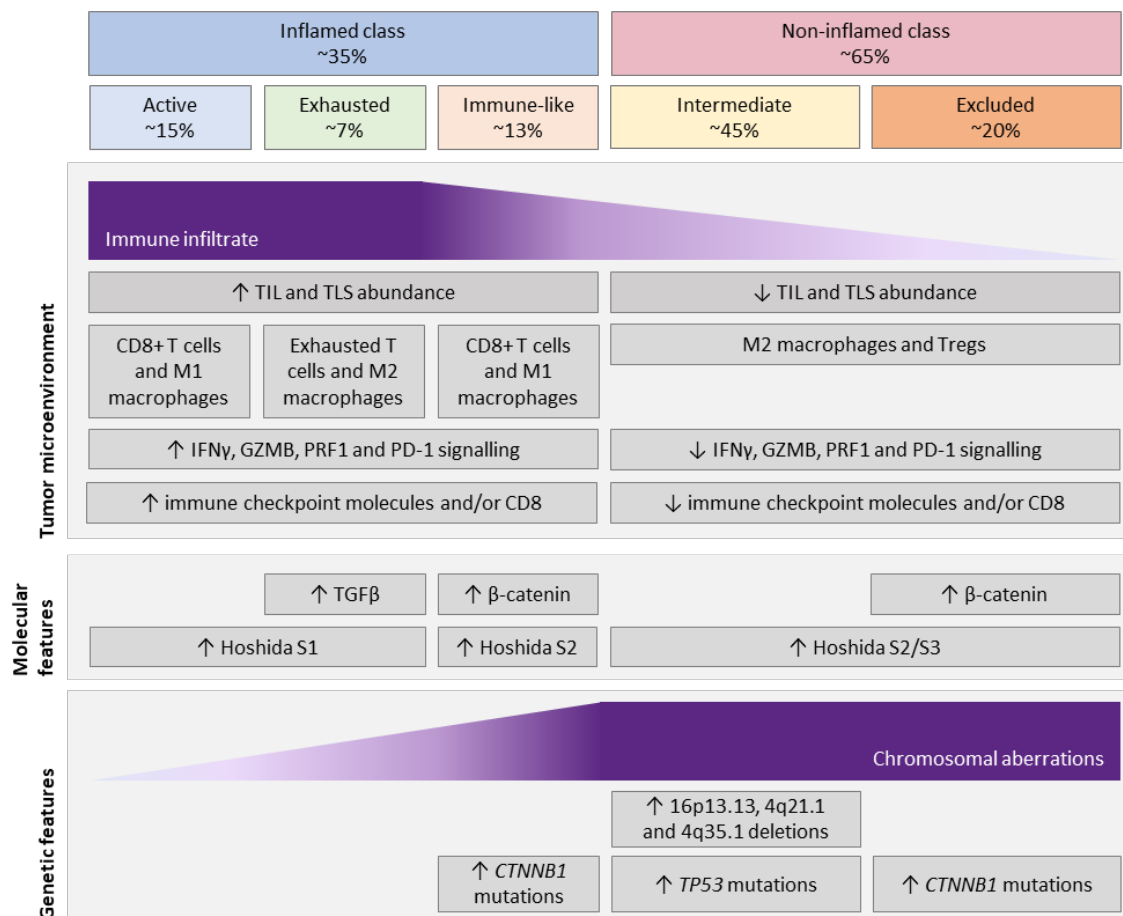
Figure 9. HCC immune-based classification. Classification of hepatocellular carcinoma (HCC) based on immune-related parameters. TCR, T cell receptor; TIL, tumor-infiltrating lymphocyte; TLS, tertiary lymphoid structures; T_{reg}, regulatory T. Adapted from Llovet JM et al., Nat Rev Clin Oncol 2021[121].

3.3.1. Immune Cell Infiltrate

Immune cells from the innate and adaptive immune systems interact in the tumor microenvironment to enable or suppress anticancer immune surveillance. The main players in the HCC immune infiltrate are the following[121,127,128]:

- *CD8+ cytotoxic T cells:* CD8+ lymphocytes exert **effector anti-tumoral functions** by eliciting cytotoxic activity through the release of granzyme B and perforin, and by producing **proinflammatory cytokines** such as IFN-γ, which inhibit tumor cell growth. Association of CD8+ T cell infiltrate with good overall survival has been widely demonstrated in several tumors including HCC[128]. However, there is a need to further explore additional markers that define the functional state of the CD8 infiltrate to improve its prognostic value (e.g., expression of immune checkpoints).

- *CD4+ helper T cells:* CD4+ cells encompass different subtypes with opposing effects. **T$_{H1}$ cells** and their derived cytokines (e.g., IFNγ) are strongly associated with good clinical outcomes in most cancer types, whereas cytokines generated by **T$_{H2}$ cells** (e.g., IL-4 and IL-10) are upregulated in advanced HCC with vascular invasion and metastasis.

- *Regulatory T cells (T$_{reg}$):* T$_{reg}$ cells are a subset of CD4+ T cells characterized **that inhibit immune responses** through several mechanisms including suppression of CD8+ T cells via TGF-β and IL-10 signaling. High tumor T$_{reg}$ infiltrate has been proposed as an independent prognostic factor for poor overall survival in HCC.

- *B cells:* B cells are at the center of the humoral adaptive immunity and are responsible for mediating the production of antibodies directed against tumor antigens. Despite they being abundant in the tumor microenvironment, no clear prognostic value has been assigned to this population.

- *Macrophages:* Tumor-associated macrophages (TAMs) arise from two distinct lineages. **Tissue-resident macrophages**, which self-renew locally, and short-lived **monocyte-derived macrophages** that infiltrate into the tumor[129]. **Kupffer cells**, which constitute up to 90% of tissue-resident liver macrophages, and other TAMs can contribute to hepatocarcinogenesis and immune evasion. An abundance of these TAMs has been associated with a poor prognosis in HCC[121]. Furthermore, TAMs have been traditionally classified into M1 and M2 macrophages on basis of their functional role. **M1 macrophages** exert a proinflammatory role by secreting cytokines with anti-tumoral effects, such as IL-12. Conversely, **M2 macrophages** produce anti-inflammatory cytokines that promote HCC tumor growth, invasion, and metastasis (e.g., IL-5, IL-6, TGF-β). Several studies have shown that a greater proportion of M2 macrophages in the HCC microenvironment results in worse clinical outcomes[121,127,128].

- *Dendritic cells (DCs):* DCs are antigen-presenting cells that exert pro-immunogenic functions by promoting T cell activation and differentiation. Subsets of DCs with distinct functions and morphology have been identified, including **anti-tumoral type 1 DCs**, and **regulatory type 2 DCs**.

- *Myeloid-derived suppressor cells (MDSCs):* MDSCs comprise a heterogeneous population of **immature and immunosuppressive myeloid cells** with protumoral capacities. They have been reported to suppress adaptive antitumor immunity by impairing CD4+ and CD8+ T cell responses and promoting T_{reg} cell expansion. Increased numbers of MDSCs have been found in tumor tissue and peripheral blood from patients with HCC, and elevated cell counts have been associated with tumor progression.

- *Neutrophils:* Tumor-associated neutrophils (TANs) release a plethora of factors exerting mostly protumoral functions, including promotion of tumor growth and invasion[128]. An increase TAN count has been associated with poor clinical outcomes in most cancer types. Notably, TANs can be further subdivided into **N1** and **N2** neutrophils, which represent the extreme of a wide spectrum displaying intermediate phenotypes. N2 neutrophils are likely responsible for the protumoral role of TANs, as opposed to the antitumoral and immunostimulatory N1 phenotype.

3.3.2. Mechanisms of Immune Evasion

Dysfunctional tumor-immune system interactions lead to immune evasion through different mechanisms of immune escape[29], including:

- *Secretion or expression of immunosuppressive factors:* Tumor cells can secrete immunosuppressive cytokines such as **TGF-β** and **IL-10**. This promotes a permissive tumor microenvironment and negatively regulates cytotoxic T lymphocytes effector activity, while also favoring tumor cell proliferation and survival. In addition, tumor cells express **co-inhibitory receptors** that negatively regulate T cell function, such as cytotoxic T lymphocyte-associated antigen 4 (CTLA4), Programmed cell death protein 1 (PD1), and T cell immunoglobulin and mucin domain containing-3 (TIM3) among others[123].

- *Recruitment of immunosuppressive cells*: Infiltration of cells with negative regulatory immune activity to the tumor microenvironment can counteract the effector function

of T cells[27,29]. This is the case of **regulatory T_reg**, **myeloid-derived suppressor cells**, and M2-polarized **TAMs**.

- *Antigen loss*: Tumor cells can downregulate the expression of tumor antigens, making it harder for the immune cells to identify them as non-self[29].

- *Reduced lymphocytic extravasation:* **Aberrant tumor vasculature** restricts the entry of immune cells to the tumor microenvironment[130]. Furthermore, tumor blood vessels lack adhesion molecules, thus reducing lymphocytic extravasation and infiltrations[29].

Any attempt to overcome these barriers to achieve an effective anti-tumor immune activation can potentially represent a therapeutic approach for HCC treatment. For instance, recent studies indicate that **VEGF expression** by malignant hepatocytes exerts pro-angiogenic effects and generates an immune-tolerant, pro-tumorigenic microenvironment by decreasing cytotoxic T cell and dendritic cell function, and promoting the recruitment of immunosuppressive cells such as T_reg, myeloid-derived suppressor cells, and tumor-associated macrophages[130–132] (**Figure 10**). This suggests suggesting that inhibition of the VEGF/VEGFR pathway could be an effective approach to boost the anti-tumoral immune response[3,100,130].
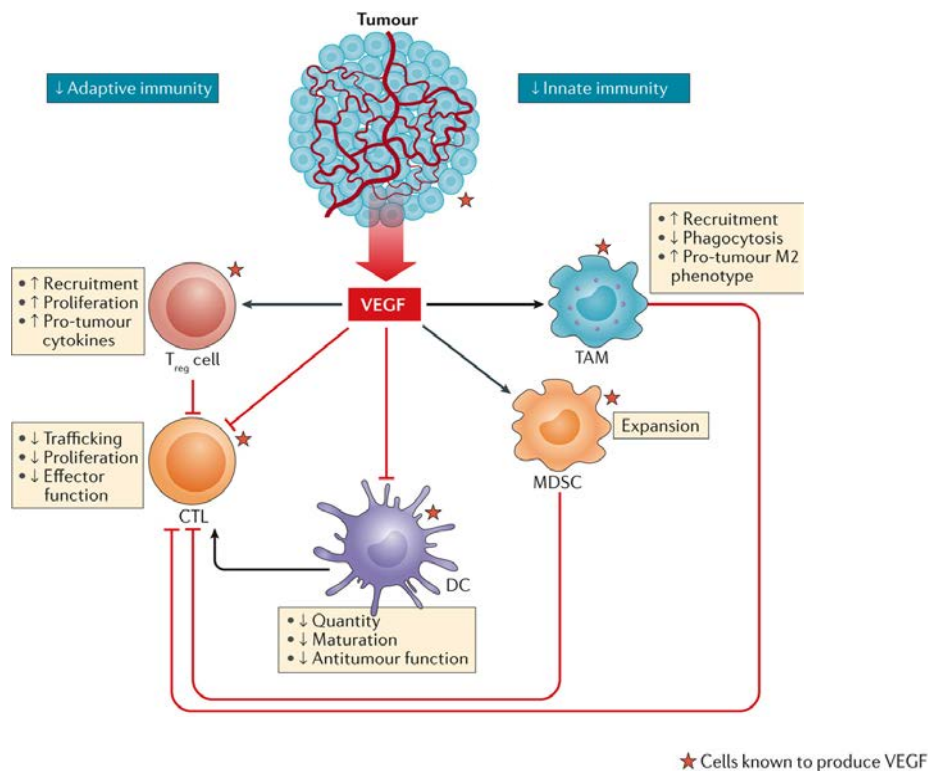
Figure 10. Direct effects of VEGF pathway activation on tumor-infiltrated immune cells. VEGF produced by tumor and immune cells modulates the functions of innate

and adaptive immune cells towards immunosuppression. Modified from Fukumura
D et al., Nat Rev Clin Oncol 2018[130].

3.4. Clinical Management of HCC Patients

3.4.1. Diagnosis, Staging, and Management

HCC development is a multistep process with a prolonged subclinical course, and it occurs in the context of a diversity of etiologies and liver disease. In light of this, surveillance protocols have been developed for HCC detection in high-risk patients[3,55] (**Table 3**). Early diagnosis of HCC in patients who have been enrolled in **surveillance programs** relies on the identification of a liver nodule by abdominal ultrasound and its confirmation using non-invasive radiological approaches or liver biopsy. Nevertheless, diagnosis at symptomatic advanced stages occurs in ~50% of cases globally, particularly in developing countries[3].

Table 3. Summary of surveillance strategies.

Patient population	Expected incidence per population
Cirrhosis from any etiology, Child-Pugh A or B	
Hepatitis B cirrhosis	3–8% per year
Hepatitis C cirrhosis	3–5% per year
Alcohol-related cirrhosis	1.3–3% per year
NASH cirrhosis	Unknown, but probably 1–2% per year
Hemochromatosis and cirrhosis	Unknown, but probably >1.5% per year
α1 antitrypsin deficiency and cirrhosis	Unknown, but probably >1.5% per year
Stage 4 primary biliary cirrhosis	3–5% per year
Other cirrhosis	Unknown
Non-cirrhotic hepatitis B	
Asian male hepatitis B carriers >40 years	0.4–0.6% per year
Asian female hepatitis B carriers >50 years	0.3–0.6% per year
Hepatitis B carrier with family history of HCC	Incidence higher than without family history
African Black people with hepatitis B	HCC occurs at a younger age (<40 years)
Patients with sufficient risk by risk score such as PAGE-B	>3% cumulative 5-year incidence if score >10
Other causes	
Patients with NASH in the absence of cirrhosis	<1.5% per year
Hepatitis C infection without cirrhosis (including F3)	<1.5% per year

HCC, hepatocellular carcinoma; NASH, non-alcoholic steatohepatitis. Adapted from
Llovet et al., Nat Rev Dis Prim 2021[3].

INTRODUCTION

Since most HCC patients suffer from underlying chronic liver damage, the clinical management of the disease is complex and needs to take into account the patient's overall health status as well as anti-tumor benefits. The **Barcelona Clinic Liver Cancer** (BCLC) – endorsed by the European Guidelines[133] – is a worldwide recognized HCC clinical algorithm for the stratification of patients.   It considers **clinical variables** such as performance status (i.e., ECOG), liver dysfunction (i.e., Child-Pugh), and **tumor-related features** such as size, number of nodules, or portal invasion[80,134]. The BCLC staging system defines five prognostic subclasses and allocates **specific treatments** for each one, based on the levels of evidence defined by the National Cancer Institute (**Figure 11**)[134,135]:

- *BCLC 0 (or very early HCC):* Patients at this stage are asymptomatic with well-preserved liver function and present low tumor burden (i.e., single tumor of <2 cm without vascular invasion). These patients are candidates for **local curative treatments**, including resection and ablation, and present very low recurrence rates.

- *BCLC A (or early HCC):* This stage includes asymptomatic patients with preserved liver function that present a single tumor >2 cm or up to 3 nodules measuring <3 cm. These patients are also considered for curative treatments. Specifically, they can be treated with **resection**, **transplantation**, or **ablation** depending on liver-related variables, can extend median survival beyond 60 months. However, recurrence rates are high (70% in 5 years), and no adjuvant therapies have demonstrated efficacy to date.

- *BCLC B (or intermediate stage):* These patients present multinodular disease with large nodules (>3 cm) or more than 3 nodules of any size. At this stage, patients are asymptomatic and maintain adequate liver function. The established standard of care for this stage is **transarterial chemoembolization**, which achieves a median survival of 26–30 months.

- *BCLC C (or advanced stage):* At this stage, patients present an advanced disease with macrovascular invasion or extrahepatic spread. These patients are eligible for **systemic therapies**, including both tyrosine-kinase inhibitors (TKI) and immune checkpoint inhibitors (ICI), which provide a median overall survival of around 19 months in first line.

- *BCLC D (or terminal stage):* This stage includes patients with impaired liver function or relevant tumor-related symptoms, for which **best supportive care** is recommended.
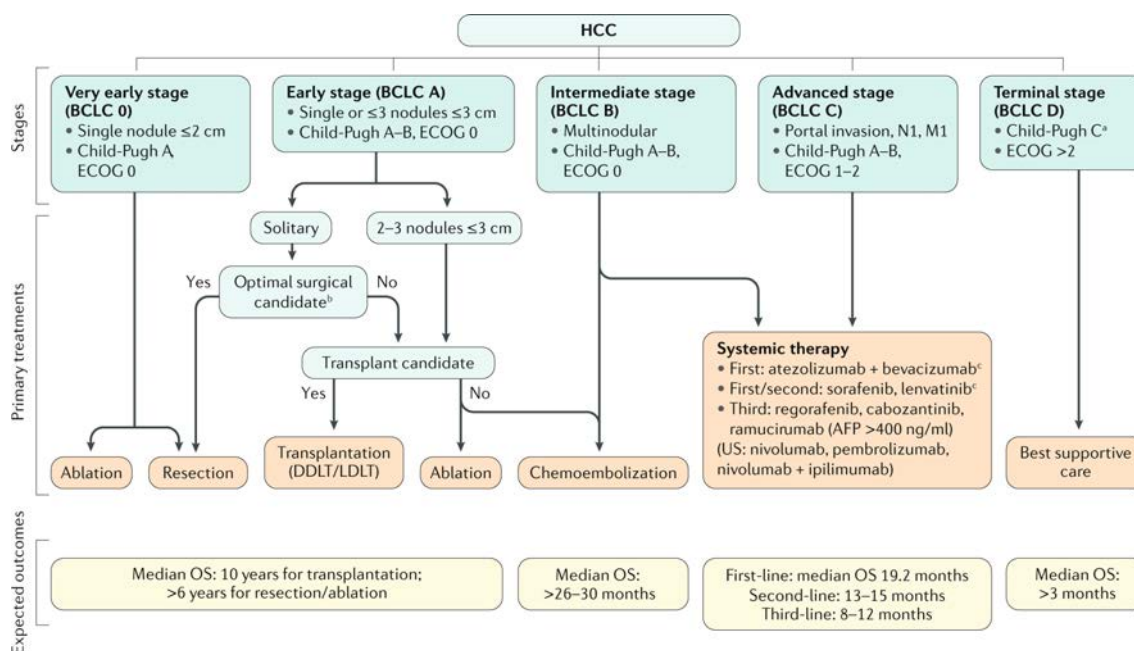
Figure 11. The Barcelona Clinic Liver Cancer (BCLC) staging system. The algorithm based on BCLC classifies patients in five stages depending on disease extension, liver function and performance status. AFP, α-fetoprotein; DDLT, deceased-donor liver transplantation; ECOG, Eastern Cooperative Oncology Group; HCC, hepatocellular carcinoma; LDLT, living-donor liver transplantation; M1, distant metastasis; N1, lymph node metastasis; OS, overall survival; RCT, randomized controlled trial; TACE, transarterial chemoembolization. Adapted from Llovet JM et al., Nat Rev Dis Prim 2021[3].

3.4.2. Systemic Therapies in Advanced HCC

Translational research has deeply improved our understanding of the pathophysiology and molecular alterations that drive HCC[3,80]. It is estimated that ~25% of HCC tumors present actionable mutations, but the low prevalence of such mutations (<10% in most cases) hampers the design of proof-of-concept studies[79,83]. Furthermore, the most common mutations in HCC are not targetable with existing drugs (e.g., *TERT*, *TP53,* and *CTNNB1*). This knowledge is yet to be translated into clinical practice. However, several molecular therapies and immunotherapy-based approaches have been implemented in the treatment of advanced HCC during the last decades.

Tyrosine-kinase inhibitors

TKI are compounds that target tyrosine kinase proteins, thus blocking many signaling pathways deregulated in cancer such as proliferation and angiogenesis[136]. In HCC, the approval of the TKI **sorafenib** in 2007 paved the way for implementing molecular therapies[137]. It wasn't until 10 years later that another TKI, **lenvatinib**, showed non-inferiority compared to sorafenib and

received FDA approval for the treatment of advanced HCC in the first-line setting[138]. The REFLECT trial, a global open-label randomized phase III study established an improved median overall survival for lenvatinib (13.6 months) compared with sorafenib (12.3 months), as well as improved ORR (24.1% versus 9.2%)[3,138] (**Table 4**). The main targets of lenvatinib include VEGF receptors 1-3, FGF receptors 1-4, PDGF receptor α, RET, and KIT, and its difference with sorafenib corresponds to a higher potency blocking VEGF receptors and the FGFR family[139].

In the second-line setting, three TKI – **regorafenib**, **cabozantinib**, and **ramucirumab** – have been approved for the treatment of HCC patients progressing to sorafenib (**Table 4**). Importantly, ramucirumab is the only biomarker-guided therapy for HCC, which is indicated for patients with baseline α-fetoprotein levels of ≥400 ng/dl[140]. Overall, TKI inhibition in second-line confers a median overall survival of 13-15 months[3].

Table 4. Summary of main outcomes and adverse events among systemic therapies approved for advanced HCC.

Treatment	Study name	Median overall survival (months)	Median PFS (months)	ORR	
				mRECIST	RECIST
First-line therapies					
Atezolizumab + bevacizumab	IMbrave150	19.2	6.8	35.4%;	29.8%
Sorafenib	SHARP (IMbrave150, REFLECT)	10.7–13.4	3.7–4.3	NA	2%
Lenvatinib	REFLECT	13.6	7.4	24.1%	18.8%
Second-line therapies					
Regorafenib	RESORCE	10.6	3.1	11%	7%
Cabozantinib	CELESTIAL	10.2	5.2	NA	4%
Ramucirumab	REACH-2	8.5	2.8	NA	5%
Second-line therapies based on FDA accelerated approval					
Pembrolizumab	Keynote 240	13.9	3.0	NA	18.3%
Ipilimumab + nivolumab	Checkmate 040*	22.8	n.a.	34%	32%

AST, aspartate aminotransferase; BR, bilirubin; HCC, hepatocellular carcinoma; ORR, overall response rate; PFS, progression-free survival; mRECIST, modified Response Evaluation Criteria In Solid Tumors; RECIST, Response Evaluation Criteria In Solid Tumors. *Data from the Checkmate 404 trial corresponds to phase Ib/II. Adapted from Llovet et al., Nat Rev Dis Prim 2021[3] and Bruix et al., J Hep 2021[141].

Immune Checkpoint Inhibitors

Immunotherapies with ICIs have emerged as promising treatment options for multiple solid tumors including HCC. ICIs are monoclonal antibodies directed against negative regulators of T cell immune function such as PD1, PDL1, and CTLA4, expressed by tumor or immune cells

present in the tumor milieu. By blocking these immune checkpoint proteins, ICIs induce an expansion of CD8[+] T cells infiltrate, resulting in the activation of the anti-tumor immune response[142].

In HCC, the combination of **atezolizumab** (anti-PDL1 antibody) and **bevacizumab** (anti-VEGF antibody) was the first regimen to improve overall survival compared with sorafenib[143], achieving a median survival of 19.2 months and 30% ORR (**Table 4**)[144]. As a consequence of these findings, this combination has become the standard of care in first-line therapies for advanced HCC[135,141]. Based upon promising phase Ib/II studies, two additional therapies – **pembrolizumab** and **nivolumab plus ipilimumab** – have received accelerated approval by the FDA in the second-line setting[145,146]. Despite the promising results of the anti-PD1 monoclonal antibody pembrolizumab, with an ORR of 18% and a median overall survival of 13.9 months, phase III studies failed to demonstrate that it prolongs overall survival compared to sorafenib but showed a non-significant trend with improved long-term survival rates[141,146,147]. This raised concerns that single-agent ICI may not have sufficient activity to show significant improvements in median OS in an unselected population. The combination of nivolumab (anti-PD1) plus ipilimumab (anti-CTLA4) achieved an ORR of 31% with a median overall survival of 23 months[145], thus providing great hopes for treatment combinations in HCC.

Emerging Combination Regimens

Although ICIs are changing the landscape of HCC treatment, monotherapy approaches elicit responses in only ~15% of patients, while the majority are primarily resistant[5,6]. Thus, much effort has been invested into identifying existing kinase inhibitors that can effectively synergize with ICI. Indeed, the combination of ICIs with **VEGF inhibitors** has shown promising activity in many solid tumors[121,130], including atezolizumab plus bevacizumab in HCC. The rationale for these combinations is that the VEGFA/VEGFR pathway has a direct immunosuppressive effect on the tumor infiltrate[130–132] (**Figure 10**). It also favors tumor growth, progression, and aberrant vasculature formation[130]. Therefore, therapies targeting the VEGF pathway likely mitigate the local immunosuppressive effects of VEGF signaling and promote T cell infiltration. In this context, the combination of **lenvatinib plus pembrolizumab** is currently being tested as first-line therapy in unresectable HCC in a phase III trial. Phase Ib data showed an encouraging objective response rate (ORR) of 46% by mRECIST, with a median overall survival of 22 months and a median progression-free survival of 9.5 month[148]. Lenvatinib could boost the effects of ICIs on the antitumor immune response by 'releasing the brake' on inflammation. However, the

immunomodulatory capacity of lenvatinib alone or in combination with anti-PD1 still remains poorly characterized.

In addition, several trials testing the combinations of a variety of other multi-kinase inhibitors plus ICIs are underway, including **cabozantinib plus atezolizumab** (NCT03755791). Notably, the inhibition of other tyrosine kinase receptors besides VEGFR, such as FGFR1-4, RET, and PDGF could potentially have immunological and molecular implications.

Finally, **combinations of different ICIs** present another promising strategy, as evidenced by the recent FDA accelerated approval of nivolumab + ipilimumab in second line[145]. This regimen is now being tested in a phase III trial versus sorafenib or lenvatinib as first-line treatment in patients with advanced-stage HCC (NCT04039607). Furthermore, the combination of the anti-PD-L1 durvalumab with the anti-CTLA4 tremelimumab recently demonstrated superior efficacy compared to sorafenib first-line therapy (NCT03298451)[121].

3.4.3. Biomarkers for Patient Selection

To date, the translation of predictive biomarkers that guide systemic therapies in HCC is under investigation. **Few candidate biomarkers** of response to TKI have been reported, with the sole exception of **α-fetoprotein** serum levels in ramuricumab (>400 ng/ml)[3,140]. For instance, FGFR4 inhibition elicited promising responses in a subset of patients with tumors overexpressing FGF19 in a phase I trial, but this discovery has not been translated in phase III investigations[149]. Similarly, MET inhibition with tivantinib did not improve survival in patients with tumor-MET overexpression[150].

A variety of candidate biomarkers for ICI are under investigation across different solid tumors, including HCC. **PD-L1 expression** by immunohistochemistry has been approved as companion diagnostics or complementary test of anti-PD-1 treatments in other malignancies, but its predictive role in HCC is unclear[3,151]. Furthermore, pembrolizumab has been FDA-approved for the treatment of advanced-stage cancers with **microsatellite instability** or **mismatch repair-deficiency** irrespectively of tumor type or histology, but this alteration is found in only ~3% of HCC[79,152]. Tumor lymphocytic **infiltration**, **gene signatures** of immune activity, and *CTNNB1* mutation status also warrant examination for predictive value in patients treated with ICIs in HCC[37].

Overall, most patients receiving ICIs do not derive benefit, and there is an **urgent need** to identify and develop **predictive biomarkers** of response to these therapies, both to enable a precision medicine approach in HCC and to better understand and overcome mechanisms of resistance.

HYPOTHESES AND AIMS

1. Hypotheses

Primary liver cancer is the sixth most commonly diagnosed cancer and the second leading cause of cancer-related death worldwide[2,3]. Around 90% of cases correspond to **hepatocellular carcinoma** (HCC)[2,3], which arises almost unfailingly in the setting of chronic liver diseases. The worldwide incidence of HCC cases is heterogeneous, reflecting the distribution of risk factors, with **Mongolia** exhibiting the highest HCC burden worldwide[2]. Well-established HCC **risk factors** are hepatitis B virus, hepatitis C virus, and non-alcoholic fatty liver disease (NAFLD). Other agents have also been reported to promote hepatocarcinogenesis, including **adeno-associated virus** (AAV) integration[3,4], but more studies are needed to determine in which conditions this virus can induce HCC.

Approximately, 50–60% of HCC patients are exposed to **systemic therapies** in their lifespan, particularly in advanced stages of the disease[3]. These mostly include multikinase inhibitors such as **lenvatinib** and immune checkpoint inhibitors (ICI) such as **anti-PD1 antibodies**. Although ICIs are revolutionizing HCC treatment, monotherapy approaches elicit responses in only 15% of patients, while the majority are primarily resistant[5,6]. Thus, identifying existing multikinase inhibitors that can effectively synergize with ICI is urgently needed.

During the last decade, molecular profiling of tumors using translational approaches has significantly contributed to the understanding of the molecular pathogenesis of HCC[3,83]. This knowledge has provided important opportunities in clinical oncology by 1) identifying novel genetic and environmental features associated with hepatocarcinogenesis in specific patient populations; and 2) unraveling new treatment strategies.

Considering all this, the **hypotheses** of this thesis are that:

1. Performing a comprehensive analysis of the molecular and immunological features of HCC will provide relevant information about the genetic and molecular determinants associated with HCC in specific populations and will unveil novel potential therapeutic approaches. This could result in fundamental implications for the clinical decision-making in HCC.

2. Assessing the distinct genomic and transcriptomic alterations of Mongolian HCC could provide relevant information that may help unveil the genetic factors and environmental exposures underlying the high incidence in this population.

1. Hypotheses

3. Hepatocyte replication in response to liver injury and NAFLD could promote oncogenic AAV integration and HCC development, which could be a concern for the use of AAV gene therapy in patients with chronic liver.

4. The multikinase inhibitor lenvatinib has immune-modulating potential and its combination with anti-PD1 checkpoint inhibitors could improve its anti-tumoral effect in HCC.

2. Aims

Considering the background and hypotheses exposed above, the specific aims of this doctoral thesis were the following:

1. To provide a molecular characterization of Mongolian HCC and identify its unique genomic features compared to Western HCC.

2. To assess whether NAFLD-associated liver damage increase the risk of AAV integration inducing HCC.

3. To identify the immunomodulatory effects of lenvatinib in combination with anti-PD1 and provide a mechanistic rationale for this treatment in advanced HCC.

2. Aims

RESULTS

Study 1 – Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features

Laura Torrens*, Marc Puigvehí*, Miguel Torres-Martín, Huan Wang, Miho Maeda, Philipp K. Haber, Thais Leonel, Mireia García-López, Wei Qiang Leow, Carla Montironi, Sara Torrecilla, Ajay Ramakrishnan Varadarajan, Patricia Taik, Genís Campreciós, Chinbold Enkhbold, Erdenebileg Taivanbaatar, Amankyeldi Yerbolat, Augusto Villanueva, Sofía Pérez-del-Pulgar, Swan Thung, Jigjidsuren Chinburen, Eric Letouzé, Jessica Zucman-Rossi, Andrew Uzilov, Jaclyn Neely, Xavier Forns, Sasan Roayaie, Daniela Sia, Josep M. Llovet

* Contributed equally

Summary

Mongolia has the **world's highest incidence of HCC** (~100 cases per 100,000 inhabitants)[2], far exceeding that of the surrounding countries[2]. The Mongolian population presents many particularities that might play a role in HCC burden, including specific risk factors, socioeconomic particularities and genetic background[72,153]. Despite a strikingly high prevalence of HBV (10.6%), HCV (6.4%), and HDV (50-80% of HBV-positive individuals) infections and alcohol consumption[56,64,65], **it is unclear whether HCC incidence is completely explained by the combination of risk factors**, or eventually other unknown factors might play a role.

In this context, we hypothesize that assessing the distinct genomic and transcriptomic alterations of Mongolian HCC could provide relevant information that may help unveil the **genetic factors and environmental exposures** underlying the high incidence in this population. Indeed, large-scale next-generation sequencing studies conducted during the last decade have been key in deciphering the molecular alterations and transcriptomic-based subtypes occurring in HCC[77,83]. Particularly, the presence of mutational signatures consisting in nucleotide substitution patterns has allowed to track the exposure of endogenous and exogenous agents in HCC samples[22].

With the aim to provide a molecular characterization of Mongolian HCC and identify its unique genomic features compared to Western HCC, we collected 192 **Mongolian HCC** and 187 **Western HCC** from Europe and US. Whole exome and RNA sequencing were conducted, and the mutational landscape, mutational signatures and transcriptomic profiles were evaluated and compared between cohorts. Furthermore, viral characteristics were assessed, including HBV and HDV viral genotypes and presence of HBV pro-oncogenic mutations.

We were able to identify distinct clinical and molecular features of Mongolian HCC compared to Western cases, including:

1. High prevalence in **females** (up to 46% of the cohort), consistent with the reported male/female 1.5/1 ratio in Mongolia[76], as opposed to that observed in the Western cohort (20% females) and globally[3]. In addition, Mongolian patients were **younger**, with less advanced hepatic fibrosis, and higher rate of **HBV-HDV co-infection** (84% of HBV-infected patients).

2. **HBV characteristics** associated with low oncogenic potential, including **genotype D1** and low prevalence of **HBV precore and basal core promoter mutations** compared to Western cases[154,155].

3. High rate of **protein-coding mutations**, almost doubling that in the Western in-house cohort (121 vs 70 mutations per tumor) and publicly available datasets[82,83,85,153]. This could potentially be explained by the presence of intrinsic and/or extrinsic factors promoting mutagenesis in Mongolian HCC.

4. Higher mutation rates in **HCC drivers** such as *TP53*, *APOB,* and the *KMT2* gene family. Furthermore, *TSC2* mutations were identified as a potential drivers in 9% of Mongolian tumors by *in silico* tools revealing positive selection of damaging alterations in this gene[156,157].

5. Presence of a novel mutational signature (**SBS Mongolia**) in 25% of cases associated with the carcinogenic **dimethyl sulfate** (DMS) genotoxic signature. This potentially suggests that long-term exposure to DMS generated from coal combustion could be a risk factor for HCC development in Mongolia.

6. A **distinct transcriptomic profile** consisting in three molecular clusters (MGL1-3), of which MGL2 (26%) and MGL3 classes (30%) were **specific for Mongolian tumors** and presented molecular and clinico-pathological features not observed in HCC samples from Western countries, including enrichment in HBV-HDV infection, female gender, and inflamed profile ($p < 0.05$).

In conclusion, Mongolian HCC is characterized by unique molecular classes, high mutational burden, and a distinct mutational signature associated with environmental factors. These

findings could pave the way for the identification of environmental or genetic factors associated with the increased incidence in this country.

Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features

Laura Torrens[1,2#], Marc Puigvehí[1,3#], Miguel Torres-Martín[1,2], Huan Wang[4], Miho Maeda[1], Philipp K. Haber[1], Thais Leonel[5], Mireia García-López[5], Wei Qiang Leow[1,6], Carla Montironi[1,2], Sara Torrecilla[2], Ajay Ramakrishnan Varadarajan[4], Patricia Taik[4], Genís Campreciós[1,5], Chinbold Enkhbold[7], Erdenebileg Taivanbaatar[8], Amankyeldi Yerbolat[8], Augusto Villanueva[1], Sofía Pérez-del-Pulgar[5], Swan Thung[1], Jigjidsuren Chinburen[8], Eric Letouzé[9], Jessica Zucman-Rossi[9], Andrew Uzilov[4,10], Jaclyn Neely[11], Xavier Forns[5], Sasan Roayaie[12], Daniela Sia[1], Josep M. Llovet[1,2,13*]

[1]Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [2]Translational research in Hepatic Oncology, Liver Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, University of Barcelona, Barcelona, Spain; [3]Hepatology Section, Gastroenterology Department, Parc de Salut Mar, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Catalonia, Spain[4]Sema4, Stamford, Connecticut, USA; [5]Liver Unit, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), CIBEREHD, University of Barcelona, Barcelona, Spain; [6]Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore; [7]Hepato-Pancreatico-Biliary Surgery Department, National Cancer Center, Ulaanbaatar, Mongolia; [8]National Cancer Center, Ulaanbaatar, Mongolia; [9]Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, Université Paris 13, Functional Genomics of Solid Tumors laboratory, F-75006, Paris, France; [10]Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; [11] Bristol Myers Squibb, Princeton, New Jersey, USA; [12]Department of Surgery, White Plains Hospital, White Plains, New York, USA; [13]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

[#] These authors contributed equally; * Corresponding author

ABSTRACT

Background and aims: Mongolia has the world's highest incidence of hepatocellular carcinoma (HCC), with ~100 cases/10^5 inhabitants/year. Here, we aimed to provide a molecular characterization of Mongolian HCC and unveil unique genomic and environmental features compared to Western HCC.

Methods: We collected 192 well-annotated paired fresh-frozen HCC/non-tumoral samples from Mongolian patients and compared its molecular profile using whole exome and RNA sequencing with a newly collected and unreported Western cohort (n=187). Mutational calling, mutational signature analysis, tumor mutational burden, and transcriptome analysis were conducted. Viral genotypes were assessed in HBV+ and HDV+ samples by direct sequencing.

Results: Mongolian patients, compared to Western, were significantly younger, with higher female predominance, and presented higher rates of HBV-HDV co-infected non-cirrhotic livers (all p<0.001). Mongolian HCCs present three unique molecular features: a) higher rates of protein-coding mutations (121 vs 70 mutations per tumor in Western) with higher mutations rates in known (e.g., *APOB*) and putative HCC drivers (*TSC2*); b) a novel mutational signature (SBS Mongolia) identified in 25% of cases (vs 4% in Western samples) that was enriched in a signature of genotoxic exposure to dimethyl sulfate, a byproduct of coal combustion; and c) a distinct transcriptomic profile consisting in three molecular clusters (MGL1-3), of which MGL2 (26%) and MGL3 classes (30%) were specific for Mongolian tumors and were enriched in HBV-HDV infection, female gender, and inflamed profile (p<0.05).

Conclusion: Mongolian HCC is characterized by unique molecular classes, high mutational burden, and a unique mutational signature associated with environmental factors.

INTRODUCTION

Liver cancer is the third leading cause of cancer-related mortality, and its global burden has increased in recent years [1,2]. Hepatocellular carcinoma (HCC) accounts for 90% of liver cancer cases and arises almost unfailingly in the setting of chronic liver diseases. The worldwide incidence of HCC cases is heterogeneous, reflecting the distribution of known liver disease risk factors such as hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol consumption, and non-alcoholic steatohepatitis [2]. Mongolia, a landlocked East Asian country between Russia and China, shows the world's highest incidence of HCC, with a burden of 86 cases per 100,000

inhabitants [1]. This incidence far exceeds that of the surrounding countries (4x and >20x compared to China and Russia, respectively) or any other country worldwide [1], and has been attributed to a historical high prevalence of HBV (10.6%) and HCV (6.4%) viruses, as well as alcohol consumption [3–5]. Indeed, 90% of Mongolian HCC cases are positive for HBV, HCV, or both [6], and co-infection with hepatitis delta virus (HDV), a defective virus that needs HBV for its replication cycle and has been associated with liver fibrosis and HCC development, occurs in 50-80% of HBV-infected individuals [7]. Despite the implementation of universal infant HBV vaccination in 1991 [3] and the improved control in HCV blood-based transmission, the burden of HCC in Mongolia is increasing year after year, and the incidence is now ~10x compared to the 1960s [1]. Overall, the reasons for this extreme incidence have never been thoroughly understood.

During the last 10 years, large-scale next-generation sequencing studies have been key in deciphering the transcriptomic-based HCC subtypes and the molecular alterations occurring in HCC [8,9]. The presence of mutational signatures consisting in unique nucleotide substitution patterns has allowed to track the exposure of endogenous and exogenous factors in cancer [10]. For instance, HCC risk factors such as aristolochic acid, aflatoxin, alcohol, or tobacco smoking can be related to specific signatures [8,11]. In this context, assessing the genetic alterations, mutational frequencies and transcriptomic profile of Mongolian HCC could provide relevant information that may help unveil the genetic factors and environmental exposures underlying the high incidence in this population. Previous studies have provided valuable data about the molecular landscape in Mongolian HCC [12]. However, further analyses comparing Mongolian tumors with a Western cohort are needed to understand the relevance of the molecular traits in Mongolian HCC. Additionally, viral characteristics in Mongolian HCC remain underexplored.

To identify the unique molecular features of Mongolian HCC, we performed whole exome (WES) and RNA sequencing (RNA-seq) in 379 HCC tumors of Mongolian and Western origin. Herein, we provide a comprehensive characterization of the molecular profile in Mongolian HCC and reveal unique features consisting in an increased number of mutations, as well as specific mutational and transcriptomic patterns. These findings could pave the way for the identification of environmental or genetic factors associated with the increased incidence in this country.

RESULTS

MATERIALS AND METHODS

Study design

A total of 219 paired HCC/non-tumoral liver samples from distinct patients undergoing HCC resection were collected at the National Cancer Center, Ulaanbaatar, Mongolia (**Fig. 1a**). Samples were collected from October 2015 to October 2017, in accordance with Mongolian regulations, the National Cancer Center, and the Ministry of Health of Mongolia. Written informed consent was obtained from all participants.

A Western cohort was used as internal control, including tumor and matched non-tumoral liver samples (n = 187) from patients undergoing resection (**Fig. 1a**). Samples were collected from two institutions of the HCC Genomic Consortium: IRCCS Istituto Nazionale dei Tumori (Milan, Italy; n = 110) and Icahn School of Medicine at Mount Sinai (New York, USA; n = 77). Samples were collected with written informed consent upon Institutional Review Board approval.

Whole exome sequencing

WES analysis was run in NovaSeq 6000 (Illumina, San Diego, CA) in the New York Genome Center facilities. WES data was used for mutation calling, mutational signature analysis, and tumor mutational burden (TMB) evaluation. TMB was calculated based on protein-coding mutations assuming an average exome size of 30 Mb, in accordance with previously published studies [10]. Mutations were called comparing the tumor with its paired non-tumoral counterpart. Molecular variant calling was performed by Sema4 (Stamford, CT, USA), using the Tigris pipeline (v2.0.1), which carries out modified GATK4 (4.0.11.0) best practices (https://software.broadinstitute.org/gatk/).

In addition, WES data from a European (n = 241) [8], Korean (n = 231) [13], TCGA (n = 363) [14] and Mongolian NCI (n = 71) [12] HCC cohorts were used for mutation calling and TMB evaluation.

Mutational signature analysis

Somatic SNVs in exome region (defined by coding exons in Ensembl GRCh37 built) at allelic frequency cutoff of 0.05 and with gnomAD population frequency or ethnic-specific frequency ≤ 0.5% were selected. Tumor samples with total SNV count ≥ 50 (after aforementioned filtering) were used in downstream mutational signature analyses, resulting in a total of 254 samples (148 from the Mongolian cohort plus 106 from the Western cohort). The remaining samples (n = 9) were considered negative for the signatures. All signature fitting and *de novo* signature

extraction analyses were performed using exome region SNVs and trinucleotide frequencies normalized via `exome2genome` approach in the deconstructSigs R package [15].

RNA sequencing

RNA data was processed by the RAPiD pipeline at the Mount Sinai Genomic Core Facility. Briefly, Fastq files were aligned using STAR (v 2.7.0f) [16] to hg19 with gencode annotation v19. QoRTs (v1.3.6) [17] was used for QC and obtaining raw counts. Batch correction was performed using RUVSeq [18]. Empirical method (RUVg) with 10.000 low expressed genes and normalization for subsequent analysis was performed using VST method from DESeq2 [19]. Unsupervised clustering analysis of the whole cohort (n = 224) was performed using the Non-negative matrix factorization (NMFc) module from GenePattern and Euclidean distance and Ward's agglomerative procedure. Clustering of the Mongolian (n = 118) and NCI Mongolian cohort [12] (n = 70) was also performed using NMFc. Gene expression characterization was performed using Nearest Template Prediction (NTP), Gene Set Enrichment Analysis (GSEA), and single sample GSEA (ssGSEA) modules from GenePattern. To this end, Molecular Signature Database (MSigDB, www.broadinstitute.org/msigdb) and previously reported gene sets were used (**Supplementary Table 1**). Class comparison between molecular clusters was performed using subclass mapping analysis. Finally, the stromal infiltration and relative tumor purity were assessed using the ESTIMATE R package [20].

Statistical analysis

Statistical analyses were performed using either SPSS software package (version 24.0; SPSS Inc, Chicago, IL, USA) or scipy (v 1.2.1) and matplotlib (v3.0.3) modules from Python (v3.7.3). Differences between qualitative variables were assessed with the Fisher exact test and corrected for multiple comparisons using false discovery rate. Differences between quantitative variables were analyzed with a two-sided non-parametric test (Mann-Whitney or Kruskal-Wallis, were appropriate), and adjustments for multiple comparison analysis were performed using Dunn's test.

Data Availability

Original whole exome sequencing and RNA sequencing data are available at the European Genome-Phenome Archive (EGAS00001005364). The remaining data are available in the Article, Supplementary Information, or upon request.

Additional detailed information is provided in the Supplementary Materials and Methods.

RESULTS

Clinico-pathological characteristics of the cohorts

HCC patients from the Mongolian cohort were younger (61 vs 66 years old, p < 0.001), with a higher rate of HBV/HDV co-infection (84% of HBV infected vs 7% in the Western cohort, p < 0.001), and lower rate of non-infected cases (15% vs 40%, p < 0.001) (**Table 1, Supplementary Table 1**). In line with previous data [21]. 54% of the HCC patients in Mongolia were male compared to 80% in the Western cohort (p < 0.001). The rate of advanced hepatic fibrosis (F3-4) and cirrhosis (F4) was significantly lower in Mongolian compared to Western cases (38% vs 79% and 16% vs 60%, respectively), independently of etiology (**Supplementary Fig. 1a**). Tumor characteristics were similar in both cohorts, with most tumors within Barcelona Clinic Liver Cancer (BCLC) stages 0-A, and with alpha-fetoprotein (AFP) < 400 (IU/mL). However, tumors in the Mongolian cohort showed a lower differentiation grade (**Table 1**). No significant survival differences were observed between cohorts.

Viral characterization of Mongolian and Western individuals

To understand whether the particular clinico-pathological features of the Mongolian HCCs were due to unique viral characteristics, we analyzed the phenotypes of HBV and HDV in infected patients. HBV genomes can be classified into 9 genotypes (A to I), according to differences in nucleotide sequences. HBV genotypes have a characteristic geographical and ethnic distribution, and HBV genotype D is known to be almost universal in Mongolia [22]. In the Western cohort, patients were infected mostly by genotypes C and D (27% and 43%, respectively, p < 0.001) whereas in the Mongolian cohort, all HBV-infected individuals were genotype D, including 2 patients with recombinant forms of genotype C and D (**Supplementary Table 2**). Interestingly, genotype D has been previously associated with reduced HCC development as compared to genotype C [23]. The most frequent HBV sub-genotypes in Mongolian and Western cases were D1 (77%) and D3 (39%), respectively (**Supplementary Fig. 1b**). Among Western cohort samples, the most prevalent genotype in European patients was genotype D3 (71%), whereas in patients from the USA it was genotype C (60%) (**Supplementary Fig. 1c**). Regarding HDV infection, all Mongolian patients were genotype 1 (**Supplementary Table 2**). We then evaluated the presence of basal core promoter (BCP) and precore HBV mutations, which have been associated with liver disease progression and HCC development [24]. It has been previously suggested that these mutations are more frequent in HBV-infected patients with genotype D, with only 11% of patients with such genotype being wild-type [25]. We detected BCP A1762T, BCP G1764A, and precore G1896A mutations in 14%, 21% and 29% of HBV-infected Mongolian patients,

respectively (**Supplementary Fig. 1d, Supplementary Table 2**). Interestingly, the prevalence of these mutations was significantly higher in the Western cohort (60%, 66%, and 46%, respectively), ($p < 0.001$, $p < 0.001$, and $p = 0.089$, respectively, **Supplementary Fig. 1d**). The percentages observed in our Western patients were similar to those previously reported [26]. In line with previously reported data [27], precore G1896A mutations were less frequent in HBV/HDV co-infected versus HBV mono-infected in Mongolia (20% vs 64%, $p = 0.002$); however, we found no differences for both BCP A1762T and G1764A mutations (14% vs 14%, and 19% vs 29%, respectively, $p = ns$). Overall, these data suggest that the rate of BCP and precore HBV pro-oncogenic mutations in Mongolia is particularly low, despite the predominance of genotype D in this population.

The interaction between HBV and HDV viruses is complex and not fully understood. It has been proposed that HDV can suppress HBV [28,29], but HBV and HDV levels can fluctuate over time [29]. To further investigate this, we determined HBV-DNA and HDV-RNA levels in tumor-adjacent liver tissue in Mongolian and Western samples. Of note, high HBV-DNA levels in blood have been associated with more aggressive liver disease, including HCC development [30]. First, we analyzed the differences between HDV positive and negative samples. Intrahepatic HBV-DNA load was significantly higher in samples with HBV/HDV co-infection than those with HBV mono-infection (5.0 vs 3.8 log copies/μg total DNA, $p = 0.001$, **Supplementary Fig. 2a**). Furthermore, high HBV-DNA levels were associated with advanced liver fibrosis, advanced tumor stage, and worse survival in Mongolian individuals (**Supplementary Table 3, Supplementary Fig. 2b**). Conversely, patients with high HDV-RNA levels showed significantly higher alanine aminotransferase levels, suggesting greater inflammation (**Supplementary Table 4**). Finally, no significant differences in HBV-DNA load were found between Mongolian and Western samples (4.85 vs. 4.97 log copies/μg total DNA, respectively $p = 0.23$, **Supplementary Fig. 2c, Supplementary Table 5**). Taken together, our analysis suggests that HBV and HDV viral characteristics in Mongolia are highly homogeneous and with low oncogenic potential.

Analysis of the genomic landscape in Mongolian HCC

To gain further insights into the molecular landscape of Mongolian HCC, we performed mutation and copy number variation (CNV) analysis. The pattern of broad gains and losses in our Western cohort was consistent with previous reports in HCC [8], with 1q and 8q gains and 8p losses being the most common alterations (**Supplementary Fig. 3a-b, Supplementary Table 6**). When we compared the broad chromosomal variation profiles between Mongolian and Western HCC patients, no difference was observed in terms of overall CNV burden (**Supplementary Fig. 3c-e**).

RESULTS

Nonetheless, Mongolian HCCs showed a significantly higher occurrence of 1p gains, 9q gains, as well as fewer 9q losses, 1q losses, and 8p losses (**Supplementary Table 6**).

The average number of mutations per tumor was significantly higher in Mongolian patients compared to Western, with a median of 121 and 70 mutations/tumor, respectively ($p < 0.001$) (**Fig. 1b**). Accordingly, the median tumor mutational burden (TMB) in the Mongolian and Western cohorts was 4.0 and 2.3 mutations/Mb, respectively ($p < 0.001$). No significant differences were observed depending on the origin of Western samples ($p = 0.645$, **Fig. 1c**). To rule out the possibility that the observed difference was due to random selection bias, we then compared the median protein-coding mutations with previously published cohorts in Western and Asian countries, applying the same filtering criteria [8,12–14]. The median number of mutations per tumor was 111 in the Mongolian NCI cohort [12], 76 in TCGA [14], 61 in the European [8] cohort and 63 in the Korean cohort [13], corresponding to a TMB of 3.7, 2.5, 2.0, and 2.1, respectively (**Fig. 1b, Supplementary Table 7**) confirming that the number of mutations in both Mongolian HCC cohorts was significantly higher than in other countries (all 0.001). There was a positive association between the number of mutations and tumor grade in Mongolian cases, with a median of 100, 120, and 171 mutations/tumor in samples with good, moderate or poor differentiation grade ($p = 0.004$). We did not observe any significant difference in the median of mutations according to etiology in either cohort (**Supplementary Fig. 4a-c**).

Previous studies in HCC have suggested that highly mutated tumors (TMB ≥4 mutations/Mb) are enriched with mutations in DNA damage response (DDR) genes [31]. In our Mongolian cohort, 84 (56%) of samples showed ≥ 4 mutations/Mb, compared to 7 (6.3%) in the Western cohort ($p < 0.001$). However, no association was observed with presence of mutations in DDR genes (**Supplementary Table 8**).

Mutational profile of Mongolian HCC

We then explored whether the higher frequency of mutations in the Mongolian cohort was due to enrichment in specific genes. Among the 250 most frequently mutated genes considering either cohort, 225 (90%) were more mutated in Mongolian HCCs, suggesting that the higher number of mutations was broadly occurring across the whole genome and not concentrated in specific loci (**Fig. 1c**). Overall, we detected a significant mutational increase in multiple known HCC driver genes in Mongolia compared to the Western cohort, including *TP53* (46% vs 32%), *APOB* (15% vs 5%), *TSC2* (9% vs 1%), and *NFE2L2* (6% vs 1%), ($p < 0.05$, **Fig. 2, Supplementary Fig. 5, Supplementary Tables 9**). The mutation rate of HCC drivers was also assessed in the NCI Mongolian cohort for comparison [12] (**Fig. 2a, Supplementary Table 10**). Furthermore, genes

84

belonging to the *KMT2* histone lysine methyltransferase family were significantly more mutated in Mongolian HCC (34% vs 18%, p = 0.005, **Fig. 3a**). Mongolian HCC with mutations in the *KMT2* gene family displayed higher TMB than patients without these mutations in both Mongolian cohorts (median TMB 4.6 and 4.7 vs 3.9, p = 0.005 and p < 0.0001, **Fig. 3b**).

Next, potential drivers in Mongolian HCC were further assessed by OncodriveCLUSTL and dN/dScv algorithms [32,33]. Among the genes significantly more mutated in the Mongolian cohort, 6 were enriched for damaging alterations, suggesting that they could exert a drivel role in Mongolian HCC (q < 0.05; **Fig. 3c-d**), including *TSC2* (9%). Similar *TSC2* mutation rates were confirmed in the Mongolian NCI cohort (7%) (**Fig. 3c**).

Finally, differences in the mutation profile depending on etiology were detected in the Mongolian cohort (**Supplementary Fig. 4d-e**). *CDKN2A* mutations were enriched in patients with HBV (7.4% vs 0%) and HDV (8.0% vs 0%) infection, while *ARID2* mutations were more common in HCV-infected patients (11.9% vs 1.2%) (p < 0.05, **Supplementary Fig. 4d**). No significant differences between Western samples from Europe and USA were observed (**Fig. 2b, Supplementary Table 9**).

Overall, our results indicate that Mongolian HCC shows a significantly higher tumor mutational burden and, although the mutational and chromosomal spectrum highly resembled that of Western HCC, significant differences were observed for key driver genes including *APOB, KMT2* family, and *TSC2*.

Mutational signature analysis

We then analyzed the pattern of single base substitutions in the Mongolian and Western cohorts. Notably, Mongolian HCC was characterized by a higher proportion of T>G substitutions compared to Western tumors (**Supplementary Fig. 6a-c**). This was also observed in the Mongolian NCI cohort (**Supplementary Fig. 6d**).

De novo mutational signature extraction from the Mongolian and Western cohorts revealed four signatures (**Fig. 4a**), three of which were mapped to COSMICv3 signatures previously found in liver cancer: SBS22, a combination of SBS6-SBS40 and of SBS16-SBS26 (cosine similarity >0.90 in all cases) [34]. The fourth signature did not present strong similarities with any of the COSMIC signatures and was therefore considered novel (**Supplementary Table 11-12**). We then performed signature fitting using the *de novo* signature 4 and HCC-specific COSMICv3 signatures by applying a bootstrap approach (exposure cutoff ≥ 0.1). Interestingly, the Mongolian cohort was enriched in the *de novo* signature 4, henceforth renamed SBS Mongolia (SBSM) (25.2%

RESULTS

[38/151 vs 4.5% [5/112] in Mongolian vs Western cohorts; p < 0.0001). Notably, Mongolian HCC presenting the SBSM signature showed a distinct substitution profile consisting in a high proportion of T>G substitutions (14% vs 8% in SBSM positive and negative samples, respectively [p < 0.001], and 6% in Western HCC, **Fig. 4c-d**). Other COSMICv3 signatures previously reported in HCC [34] presented similar prevalence between cohorts (**Fig. 4c-d, Supplementary Fig. 6e-f, Supplementary Table 13**).

To further characterize SBSM positive samples, we assessed the presence of mutational signatures linked to the effects of known or suspected environmental mutagens from the Compendium of Mutational Signatures of Environmental Agents [35]. Samples presenting SBSM were significantly enriched for the mutational signature associated with exposure to dimethyl sulfate (DMS) (71.1% [27/38] vs 26.5% [30/113], p < 0.0001, **Fig. 4b**). In line with this, DMS was the only environmental-related signature significantly enriched in Mongolian HCC compared to Western (37.7% [57/151] vs 18.8% [21/112], **Supplementary Table 14**). Patients presenting the DMS signature were older (64.3 vs 59.5 years) and predominantly HCV-positive; **Supplementary Fig. 7**). No association between SBSM and TMB, etiology, fibrosis, or other clinical and molecular variables were found (**Fig. 4b**).

Finally, we investigated the mutational profile in adjacent matched liver tissue of Mongolia and Western HCCs. Due to the small number of SNVs present in the adjacent tissues, mutational signature fitting was performed on pooled variants from the Mongolian and Western cohorts using HCC-specific COSMICv3 signatures plus SBSM (**Supplementary Fig. 8**). SBSM was the only dominant signature in adjacent tissue from the Mongolian cohort (**Supplementary Fig. 8c-d**), suggesting that non-tumoral liver tissue in Mongolia presents the signature before HCC arises. SBS5, associated with age-related clock-like mutations [34], was the main signature in Western non-tumoral tissue.

Overall, Mongolian HCC presents a unique substitution profile characterized by a novel mutational signature, SBSM, which is associated with the DMS-related signature. Considering this, DMS exposure warrants further investigation as a potential environmental factor for HCC in Mongolia.

Identification of unique gene expression patterns in Mongolian HCC

We then investigated the transcriptome profiling of HCC samples using RNA-seq data to define the molecular patterns in Mongolian HCC tumors. Unsupervised clustering analysis of Mongolian and Western samples using non-negative matrix factorization (NMFc) identified two robust

clusters (**Supplementary Fig. 9a-c**). Notably, 80% of Western tumors were included in one cluster, while the second cluster showed a strong enrichment of Mongolian HCC samples, suggesting that Mongolian HCC may present a distinct molecular profile. Furthermore, NMFc analysis of the Mongolian HCC samples alone revealed three main clusters -MGL1, MGL2, and MGL3- (**Supplementary Fig. 9b-c**), which overlapped with the classification of the whole cohort. Unsupervised clustering showed similar results, thus confirming the robustness of our findings (**Supplementary Fig. 9d**).

To elucidate the transcriptomic differences between Mongolian and Western HCC, we performed single-sample Gene Set Enrichment Analysis (ssGSEA) and Nearest Template Prediction (NTP) (**Supplementary Fig. 10a**). The Mongolian cohort presented an enrichment in Hoshida S1 and Proliferation classes (39% vs 20% and 36% vs 14%, $p < 0.01$). In addition, Mongolian tumors showed enhanced inflammatory signaling (i.e., IFN and HCC Immune class), response to viral infection, and growth factor-related pathways (all $p < 0.05$). Comparatively, the Western cohort was enriched in the Hoshida S3 class (26% vs 47%, $p < 0.01$) and liver-related metabolic activation.

We then characterized each one of the MGL clusters. Patients belonging to MGL1 class (44% of the cohort) were more frequently HCV-infected, older, and mostly males (female:male ratio of 1:2), (**Supplementary Table 15**) and with a molecular profile closer to Western HCCs than the rest of Mongolian HCC (**Fig. 5**). On the other hand, patients of the MGL2 (26%) and MGL3 clusters (30%) were significantly younger than MGL1 and enriched in HBV/HDV infection and triple infections (HBV/HDV/HCV). Interestingly, while patients of the MGL3 class showed a female:male ratio similar to MGL1 (1:2), the MGL2 class showed a female:male ratio of 2:1 and higher AFP levels. None of the MGL clusters was associated with specific outcomes (**Fig. 5, Supplementary Fig. 11, Supplementary Table 15**).

In terms of molecular features (**Fig. 5**), the MGL1 cluster was characterized by enrichment in *CTNNB1* class and activation of metabolic and liver-specific pathways (all $p < 0.05$). MGL2 HCC cases displayed a proliferative HCC phenotype (i.e enrichment of Proliferation, G3, and Cluster A gene signatures), higher rates of *RB1* mutations, and lower rates of *CTNNB1* mutations (**Fig. 5**). Finally, MGL3 were characterized by lower rates of *TP53* mutations and fewer broad chromosomal alterations and were particularly enriched in immune-related features (i.e immune class, interferon and inflammatory pathways, and PD1 signaling).

The clinico-pathological and molecular features of each MGL cluster were further validated using RNA-seq data from the Mongolian NCI cohort [12] (**Supplementary Fig. 12a, Supplementary**

RESULTS

Table 16). In addition, subclass mapping and NTP analysis in the in-house Mongolian cohort indicated a good overlap of the MGL clusters with the previously published MO classification of Mongolian HCC (**Supplementary Fig. 13**) [12]. Specifically, MGL1 aligned with the MO1 class, MLG2 with MO4; and MGL3 with MO2 and MO3 (all FDR < 0.05).

Overall, our results suggest that three distinct gene expression patterns characterize Mongolian HCC of whom two presents unique clinico-pathological features not observed in HCC samples from Western countries.

Immune characterization of the molecular classes of Mongolian HCC

We then further explored the inflammatory profile of Mongolian HCC using ESTIMATE [20] and ssGSEA in our in-house Mongolian cohort and the NCI cohort [12]. In both Mongolian cohorts, MGL3 showed a significantly higher immune enrichment score, presence of both an innate and adaptive immune response, and signatures predicting response to immunotherapy (**Fig. 6, Supplementary Fig. 12b**). A certain degree of inflammation was also observed in MGL2, even if significantly lower than MGL3. Our previously reported HCC immune class was detected in most patients belonging to MGL3 (88% in the Mongolian cohort, 100% in the NCI Mongolian cohort [12], **Fig. 6, Supplementary Fig. 12b**). Compared to Western HCC, the immune class in Mongolian tumors was larger (42% versus 29%, p = 0.05) with an inverted ratio of Exhausted/Active subtypes (65/35 vs 30/70, **Supplementary Fig. 10b**), indicating a more prominent immunosuppressive phenotype. Overall, we observed a higher presence of immune signaling in the Mongolia clusters MGL2 and MGL3. Considering that most patients belonging to these clusters were HBV/HDV infected and younger, this suggests that HDV infection could accelerate disease progression through inflammatory mechanisms.

DISCUSSION

Our study entails a comprehensive characterization of the molecular profile of HCC in Mongolia, the country with the highest global incidence. Mongolia has many particularities that might play a role in HCC burden, including specific risk factors, socioeconomic particularities, and genetic profiles [12,36]. Despite a strikingly high prevalence of HBV (10.6%), HCV (6.4%), and HDV (70% of HBV-positive individuals) infections and alcohol consumption [3–5], it is unclear whether HCC incidence is completely explained by the unique combination of risk factors, or eventually, other unknown factors might be responsible. In addition, a direct comparison with a comprehensive Western cohort is necessary to understand the relevance of newly-identified molecular traits.

By analyzing WES and RNA-seq data from 379 new Mongolian and Western HCC samples, we identified unique genomic and transcriptomic footprints in Mongolian tumors that suggest a role of specific genetic and environmental factors in the country. The study provides novel information in three major areas: a) Clinical characteristics of Mongolian cases, b) High tumor mutational burden and mutational profile associated with environmental agents, and c) Unique transcriptomic-based molecular classes.

Regarding the clinico-pathological particularities of Mongolian HCC patients, we confirmed the high prevalence in females (up to 46% of the cohort), consistent with the reported male/female 1.5/1 ratio [21], as opposed to that observed globally [2] and in the surrounding countries (2.6/1 in Russia, 3.4/1 in China, and 3/1 in East Asia) [21]. Mongolian HCC also occurred in younger patients with milder underlying liver fibrosis (F1-2 stages in >60% of cases) and with a dominant viral-related etiology (85% of either HBV, HBV-HDV, or HCV-positive).

In our study, HBV characteristics such as genotype D1 and precore mutations in < 30% of cases revealed traits associated with a low oncogenic potential of the virus in Mongolia [23,25], whereas HBV load was similar to Western samples. Thus, other factors not associated with HBV infection might be responsible for the high HCC incidence rates in Mongolia [23,25]. In this sense, 84% of the HBV-infected Mongolian patients showed co-infection with HDV, which contrasts with the Western data (less than 7% of coinfection) [37]. This unique co-infection profile was associated with two molecular subclasses only identified in Mongolian patients (MGL2 and MGL3), thus pointing towards a potential role of HDV in the oncogenic process, despite it is not currently considered a carcinogenic agent [38].

From the genomic standpoint, Mongolian HCCs had a high rate of protein-coding mutations, which almost doubled that in the Western in-house cohort (121 vs 70 mutations per tumor) and publicly available datasets [8,12–14]. This suggests the presence of intrinsic and/or extrinsic factors promoting mutagenesis in Mongolian HCC. Notably, while several HCC driver genes presented similar mutation rates in the Mongolian cohort compared to the Western tumors and previous studies [8,13,14] (e.g., *CTNNB1* and *ARID1A*), others such as *APOB*, and the *KMT2* gene family were significantly more mutated in Mongolian HCC (e.g., *APOB* in 15.2% vs 4.5% in Western). For instance, *KMT2* family mutations (34% vs 17% in Western) were associated with higher TMB in Mongolian HCC, suggesting that they could be partially responsible for the increased mutational burden. In line with this, loss of function in *KMT2* methyltransferases has been proposed to induce DNA damage due to aberrant chromatin remodeling [39]. Finally, *TSC2* mutations (9%) were identified as potential drivers in Mongolian tumors. *TSC2*, a known cancer-

related gene [8,40] participating in the mTOR oncogenic pathway, has been proposed as an actionable alteration with level 2B evidence, as it could be a predictor of response to the FDA-approved drug everolimus [40].

To understand whether Mongolian HCCs present differential genomic and genotoxic footprints, we assessed the presence of distinct mutational signatures [10,35]. We identified a new mutational signature (SBS Mongolia) with no similarities with previously reported COSMIC signatures. SBS Mongolia was significantly enriched in Mongolian HCC (25% vs 4.5%), indicating distinct substitution patterns characterized by increased T>G substitutions. Overall, this mutational landscape points towards specific exposure to environmental factors in Mongolian patients. In this regard, Mongolian HCC samples presenting the novel SBS Mongolia were significantly enriched for the mutational signature associated with exposure to DMS (71.1% vs 26.5%). DMS has been classified as a probable carcinogenic hazard to humans by the International Agency for Research on Cancer [41] (category 2A carcinogen) and its production has been associated with coal combustion. Most of the Mongolian population is currently exposed to coal combustion, as coal is used to fight against the intense cold weather both in urban and rural areas. Half of the 3-million population of Mongolia lives in Ulaanbaatar, an overpopulated capital with dismal environmental conditions [42], whereas the rest is still predominantly nomad and lives in traditional tents or *gers*, where coal is used both for cooking and heating. This fact has been recognized by international organizations as a major health threat in this country [42]. Considering this, we hypothesize that long-term exposure to DMS from coal combustion could be a risk factor for HCC development in Mongolia. In this regard, the DMS signature was associated with older patients, potentially due to a longer exposure time.

The transcriptomic profile of Mongolian tumors was consistent with known HCC features [2]. Nonetheless, we observed two striking differences a) Mongolian HCC was characterized by an enhanced proliferative and immunological signaling, with a proportion of tumors belonging to proliferative/progenitor HCC classes (39%) doubling the one in Western HCC (20%) and previously published studies [43]; and b) Two out of the three identified molecular classes presented distinct molecular features compared to Western HCC, and thus are deemed unique for Mongolian tumors. Effectively, MGL2 (26%) and MGL3 classes (30%) were specific for Mongolian tumors but not for Western HCC and were enriched in HBV/HDV infection. Interestingly, the MGL2 class was associated with both clinical and molecular features of aggressiveness and showed a female:male ratio of 2:1. The presence of this molecular class could be due to the increased HCC incidence in females in this population [21]. Finally, MGL3 presented an inflamed profile, potentially linked to the response to HBV/HDV infection [7].

In conclusion, we provided an exhaustive comparison of the genomic and transcriptomic characteristics of Mongolian HCC with an in-house Western cohort. Mongolian HCC is characterized by high mutational rates, a distinct mutational signature profile, and the presence of two unique molecular subclasses. Finally, environmental factors such as DMS need to be further explored as a potential risk factor in this population.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.

2. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. Nat Rev Dis Prim. 2021;7:7.

3. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. Lancet Gastroenterol Hepatol. 2018;3:383–403.

4. Blach S, Zeuzem S, Manns M, et al. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. Lancet Gastroenterol Hepatol. 2017;2:161–76.

5. Demaio AR, Dugee O, de Courten M, et al. Exploring knowledge, attitudes, and practices related to alcohol in Mongolia: a national population-based survey. BMC Public Health. 2013;13:178.

6. Dondog B, Lise M, Dondov O, et al. Hepatitis B and C virus infections in hepatocellular carcinoma and cirrhosis in Mongolia. Eur J Cancer Prev. 2011;20:33–9.

7. Puigvehí M, Moctezuma-Velázquez C, Villanueva A, et al. The oncogenic role of hepatitis delta virus in hepatocellular carcinoma. JHEP Reports. 2019;1:120–30.

8. Schulze K, Imbeaud S, Letouzé E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet. 2015;47:505–11.

9. Zucman-Rossi J, Villanueva A, Nault JC, et al. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology. 2015;149:1226–39.

10. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

11. Letouzé E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8:1–13.

12. Candia J, Bayarsaikhan E, Tandon M, et al. The genomic landscape of Mongolian hepatocellular carcinoma. Nat Commun. 2020;11.

13. Ahn S-M, Jang SJ, Shim JH, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology. 2014;60:1972–82.

14. Ally A, Balasundaram M, Carlsen R, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327-1341.e23.

15. Rosenthal R, McGranahan N, Herrero J, et al. deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17.

16. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

17. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics. 2015;16:224.

18. Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32:896–902.

19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

20. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4.

21. Petrick JL, Braunlin M, Laversanne M, et al. International trends in liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J cancer. 2016;139:1534–45.

22. Baatarkhuu O, Gerelchimeg T, Munkh-Orshikh D, et al. Epidemiology, Genotype Distribution, Prognosis, Control, and Management of Viral Hepatitis B, C, D, and Hepatocellular Carcinoma in Mongolia. Euroasian J Hepato-Gastroenterology. 2018;8:57–62.

23. Wong GL-H, Chan HL-Y, Yiu KK-L, et al. Meta-analysis: The association of hepatitis B virus genotypes and hepatocellular carcinoma. Aliment Pharmacol Ther. 2013;37:517–26.

24. Wei F, Zheng Q, Li M, et al. The association between hepatitis B mutants and hepatocellular carcinoma: A meta-analysis. Medicine (Baltimore). 2017;96:e6835.

25. Sonneveld MJ, Rijckborst V, Zeuzem S, et al. Presence of precore and core promoter mutants limits the probability of response to peginterferon in hepatitis B e antigen-positive chronic hepatitis B. Hepatology. 2012;56:67–75.

26. Chen B-F, Liu C-J, Jow G-M, et al. High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. Gastroenterology. 2006;130:1153–68.

27. Jardi R, Rodriguez F, Buti M, et al. Role of hepatitis B, C, and D viruses in dual and triple infection: Influence of viral genotypes and hepatitis B precore and basal core promoter mutations on viral replicative interference. Hepatology. 2001;34:404–10.

28. Giersch K, Homs M, Volz T, et al. Both interferon alpha and lambda can reduce all intrahepatic HDV infection markers in HBV/HDV infected humanized mice. Sci Rep. 2017;7.

29. Schaper M, Rodriguez-Frias F, Jardi R, et al. Quantitative longitudinal evaluations of hepatitis delta virus RNA and hepatitis B virus DNA shows a dynamic, complex replicative profile in chronic hepatitis B and D. J Hepatol. 2010;52:658–64.

30. Chen CJ, Yang HI, Su J, 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.

31. Totoki Y, Tatsuno K, Covington KR, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet. 2014;46:1267–73.

32. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

33. Arnedo-Pac C, Mularoni L, Muiños F, et al. OncodriveCLUSTL: A sequence-based clustering method to identify cancer drivers. Bioinformatics. 2019;35:4788–90.

34. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

35. Kucab JE, Zou X, Morganella S, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019;177:821-836.e16.

36. Bai H, Guo X, Zhang D, et al. The Genome of a Mongolian Individual Reveals the Genetic Imprints of Mongolians on Modern Human Populations. Genome Biol Evol. 2014;6:3122–36.

37. Tsatsralt-Od B, Takahashi M, Nishizawa T, et al. High prevalence of dual or triple infection of hepatitis B, C, and delta viruses among patients with chronic liver disease in Mongolia. J Med Virol. 2005;77:491–9.

38. Rao RC, Dou Y. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. Vol. 15, Nature Reviews Cancer. Nature Publishing Group; 2015. p. 334–46.

39. Chakravarty D, Gao J, Phillips S, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. 2017;2017:1–16.

40. International Agency for Research on Cancer. Agents Classified by the IARC Monographs, Volumes 1–127 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans [Internet]. 2018 [cited 2020 Nov 5]. Available from: https://monographs.iarc.fr/agents-classified-by-the-iarc

41. WHO. Air pollution in Mongolia. Bull World Health Organ. 2019;97:79:80.

42. Chiang DY, Villanueva A, Hoshida Y, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res. 2008;68:6779–88.

FIGURES

a

b

Protein-coding mutations

c

d
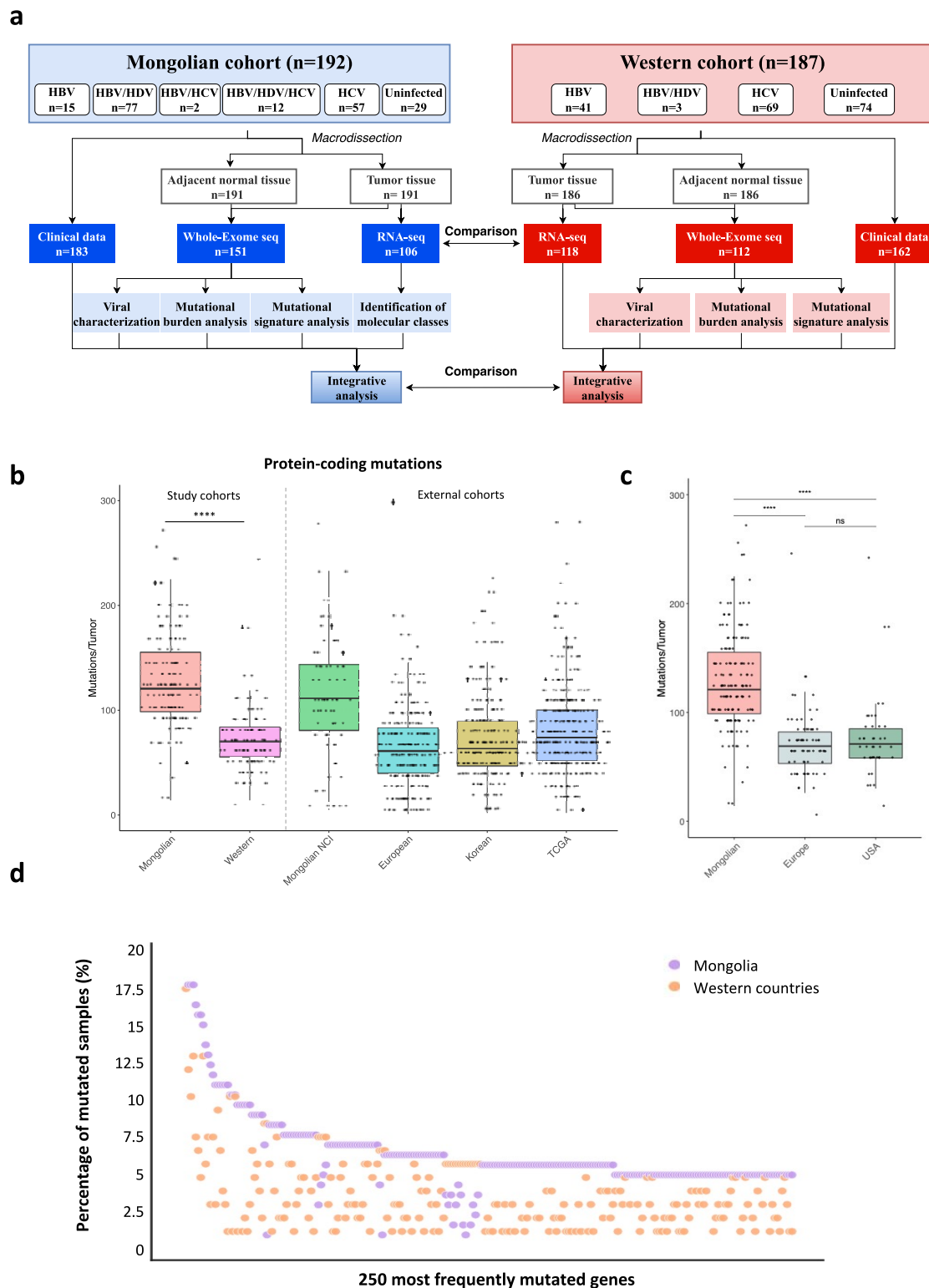
Figure 1. Flow chart of the study and mutational profile in Mongolian HCC. a A total of 192 HCC samples from Mongolia were used in this study. A Western cohort including 187 HCCs was used as internal control. **b** Mutations per tumor in the Mongolian (n = 151) and Western (n = 112) cohorts. Mongolian NCI (n = 71), European (n = 241), Korean (n = 231) and TCGA (n = 363) external cohorts are shown

as reference. **c** Mutations per tumor in Mongolian HCC (n = 151) compared to Western HCC from Europe (n = 69) and USA (n = 43). Y axis was cut at 300 mutations/tumor to facilitate data interpretation. P-value corresponds to Kruskal-Wallis test. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers). **d** Percentage of mutated samples for the top 250 most frequently altered genes in the Mongolian and Western cohort. Y axis was cut at 20% to improve readability.
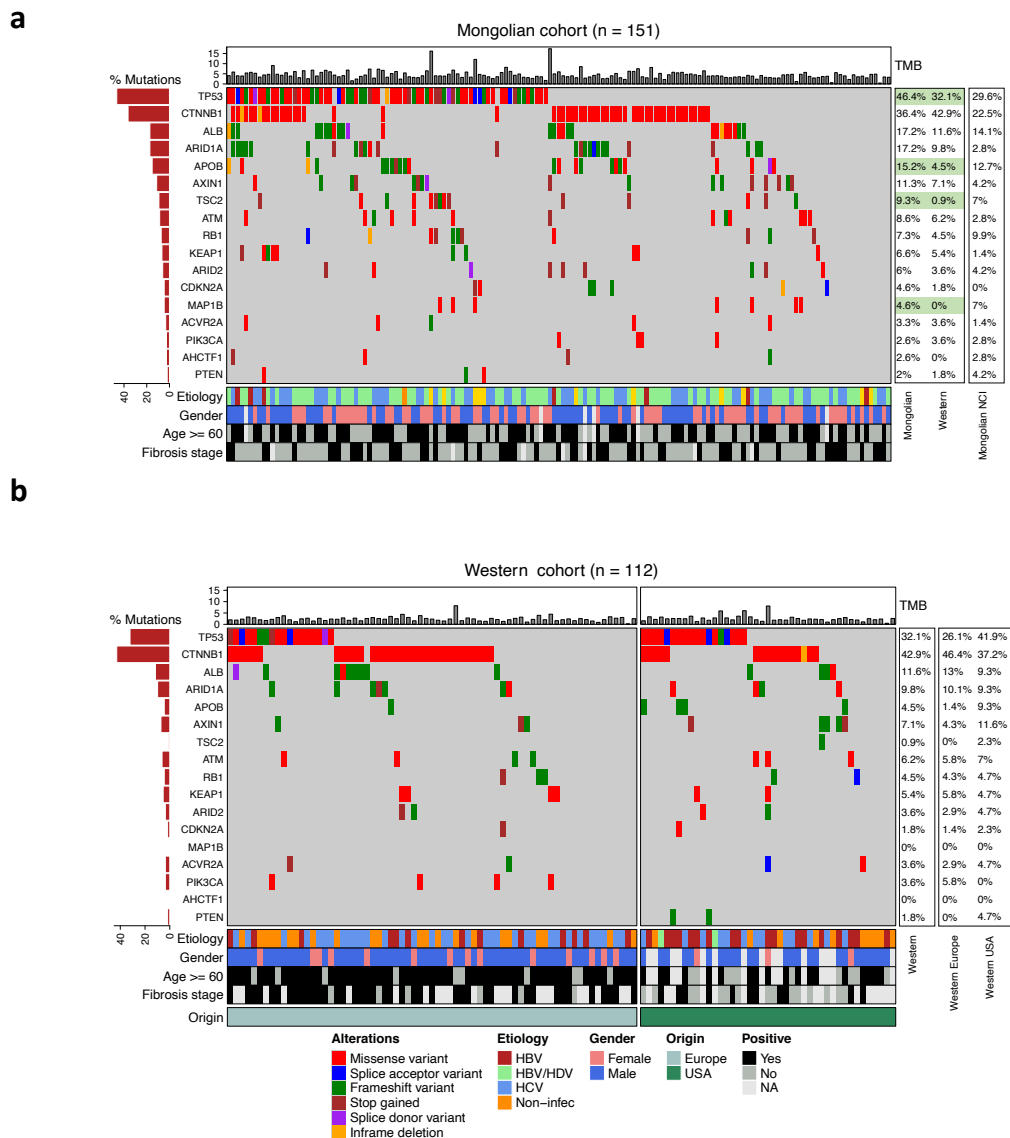
Figure 2. Mutational landscape of Mongolian and Western HCC. a Mutations present in HCC samples from the Mongolian cohort (n = 151). The frequency of mutations in the Western and Mongolian NCI cohort are indicated for comparison (left). Genes with significant differences between the Mongolian and Western cohorts are highlighted in green (Fisher p < 0.05). **b** Mutations present in HCC samples from the Western cohort (n = 112), sorted by sample origin. Overall frequency of mutations and frequencies in Western samples from Europe (n = 69) and USA (n = 43) are shown (left). No significant differences between samples from Europe and USA were found (Fisher test). Top panel shows tumor mutational burden (TMB, mutations/Mb) per sample. Middle panel indicates the presence of mutations per sample (right) and overall percentage (left) in the most frequently mutated genes. Bottom panel details clinico-pathological parameters.
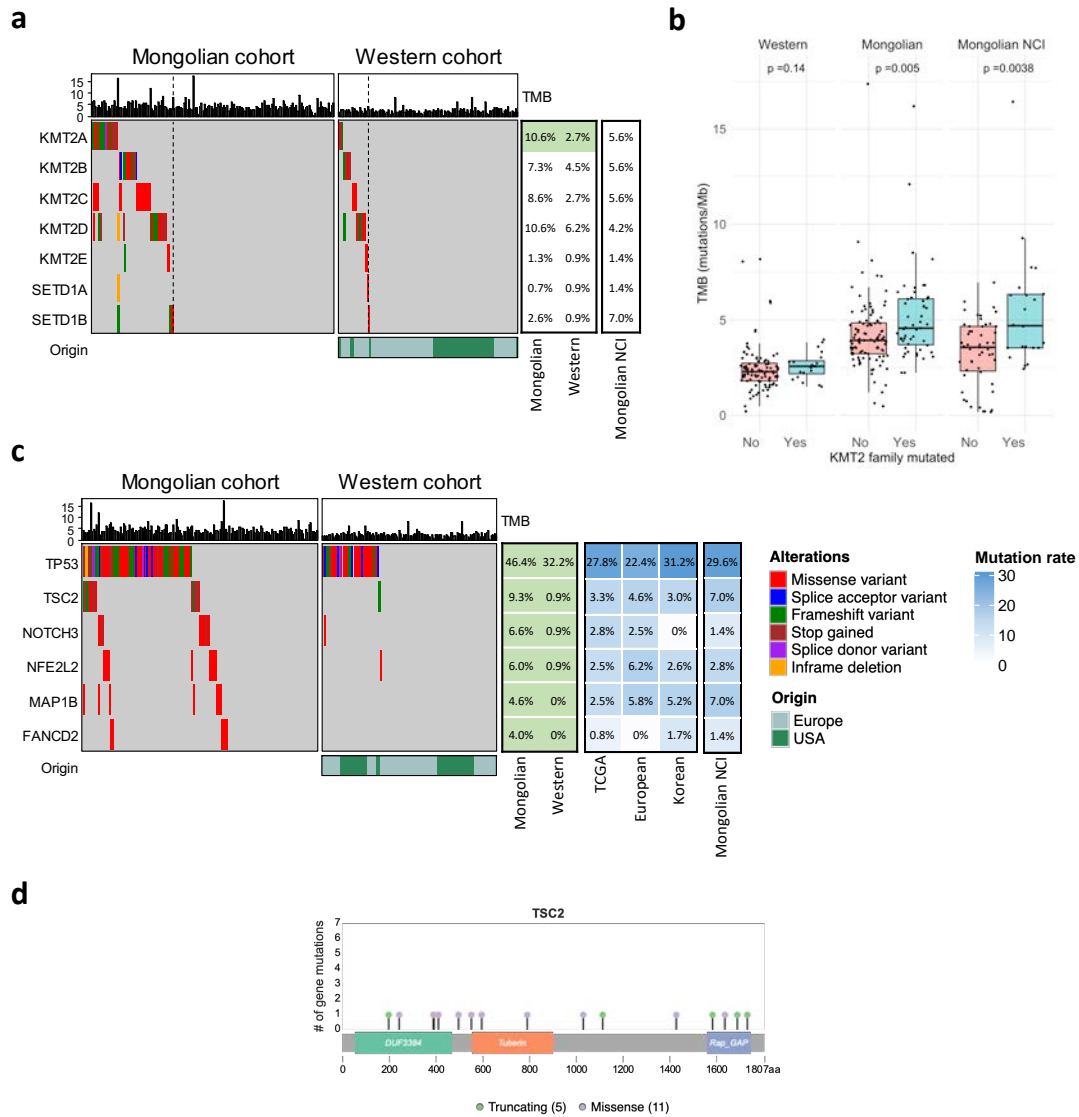
Figure 3. Potential driver and *KMT2* family mutations in Mongolian and Western cohorts. a Mutations in the *KMT2* gene family in the Mongolian (n = 151) and Western cohorts (n = 112). **b** TMB in samples with *KMT2* family mutations and wild type (wt). Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers). P-value corresponds to Kruskal-Wallis test. **c** Mutations in potential driver genes in the Mongolian and Western cohorts according to enrichment in damaging alterations. Percentage of mutations in external cohorts is indicated in the right panel. Top panel shows tumor mutational burden (TMB, mutations/Mb) per sample. Middle panel indicates the presence of mutations per sample (left) and overall percentage (right) in the Mongolian and Western cohorts. Bottom panel details clinico-pathological characteristics. Significant differences between the Mongolian and Western cohorts are indicated in green (Fisher p < 0.05). **d** Mutation distribution in the *TSC2* gene.

Figure 4. Single-base substitution signatures in Mongolian and Western HCC. a *De novo* single-base substitution (SBS) signatures found in Mongolian (n = 151) and Western HCC (n = 112) samples. For each signature, the 96-substitution classification is displayed, including the substitution type and sequence context. **b** Distribution of samples harboring SBSM in the Mongolian cohort. Presence of dimethyl sulfate (DMS) mutational signature, main clinical variables and T>G substitution frequency are shown for each sample. FDR-adjusted p-values comparing SBSM positive and negative samples are indicated. Percentages in the lower panel refer to median T>G frequency. **c-d** Signature fitting results in Mongolian (**c**) and Western (**d**) HCC using HCC-specific COSMIC signatures and SBS Mongolia (SBSM).

Figure 5. Molecular classification of Mongolian HCC. Consensus-clustered classification of Mongolian HCC samples using Non-negative matrix factorization. In the heatmap, clinico-pathological characteristics, nearest template prediction and gene set enrichment in each sample are shown.
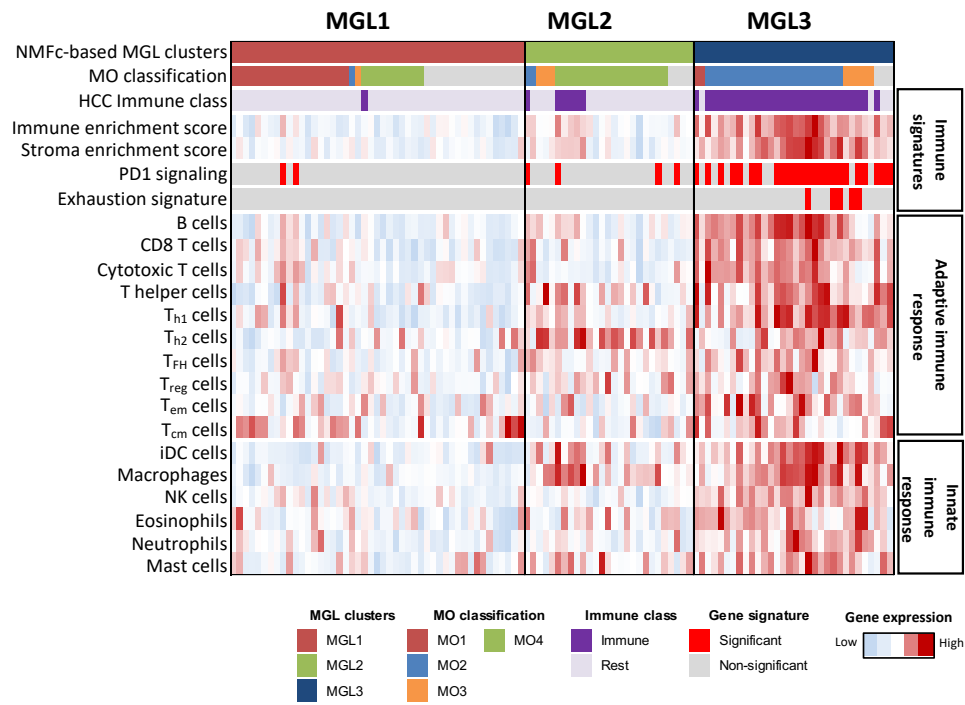
Figure 6. Inflammatory profile of Mongolian HCC. Characterization of inflammatory profile in the MGL clusters assessed by ESTIMATE analysis and single sample gene set enrichment analyses capturing distinct immune populations. T_{h1}, Type 1 helper; T_{h2}, type 2 helper; T_{FH}, T follicular helper; T_{reg}, regulatory T; T_{em}, effector memory T, T_{cm} central memory T.

RESULTS

TABLES

Table 1. Baseline characteristics of the Mongolian and Western cohorts

	Mongolian Cohort (n = 192)	Western Cohort (n = 187)	p value
Age (years)	61 (18-80)	66 (29-91)	<0.001
< 60 years (n, %)	82 (44.6)	32 (18.6)	<0.001
< 50 years (n, %)	20 (10.9)	7 (4.1)	0.017
Gender (male, %)	98 (53.6)	137 (79.7)	<0.001
Etiology			<0.001
HBV+ (n, %)	15 (7.8)	41 (21.9)	
HBV/HDV+ (n, %)	77 (40.1)	3 (1.6)	
HBV/HCV/HDV+ (n, %)	12 (6.3)	0 (0)	
HCV+ (n, %)	57 (29.7)	69 (36.9)	
HBV/HCV (n, %)	2 (1)	0 (0)	
Non-infected (n, %)	29 (15)	74 (39.6)	
Bilirubin (mg/dL)	0.6 (0.1-3.7)	0.9 (0.3-3.8)	<0.001
Albumin (g/L)	41 (29-49)	40 (22-54)	ns
Platelets (10^9/L)	181 (76-574)	160 (27-493)	<0.001
< 150x10^9/L (n, %)	50 (28.6)	81 (47.4)	<0.001
AFP (IU/mL)	22 (1-121000)	12 (1-311190)	ns
> 400 IU/mL (n, %)	32 (20.9)	26 (16.4)	ns
Tumor size	5 (0.8-20)	4.2 (1-20)	0.001
> 5 cm (n, %)	93 (52.8)	67 (41.4)	0.039
BCLC stage (0-A, %)	132 (78.1)	129 (79.6)	ns
Multinodular disease (n, %)	26 (15.4)	42 (25.8)	0.021
Advanced liver fibrosis (F3-4, %)	64 (38.1)	106 (78.5)	<0.001
Cirrhosis (F4, %)	27 (16.1)	81 (60)	<0.001
Microvascular invasion (yes, %)[#]	72 (47.7)	80 (46.5)	ns
Tumor grade (G3-4, %)	11 (10.6)	41 (28.7)	<0.001

HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; AFP, alfa-fetoprotein; BCLC, Barcelona Clinic Liver Cancer The following variables have missing values for the Mongolian and Western cohorts, respectively: Age: 8 and 15 patients. Gender: 9 and 15 patients. Bilirubin, albumin, and platelets: 19 and 17 patients. AFP: 39 and 28 patients. Tumor size: 16 and 25 patients. BCLC stage: 23 and 25 patients. Tumor number: 23 and 24 patients. Liver fibrosis in 24 and 52 patients. Microvascular invasion in 41 and 15 patients. Tumor grade in 88 and 44 patients.

Financial support

Conflict of interest

Author contributions

L.T. and M.P. contributed equally to this work. M.P., D.S., and J.M.L. developed the study concept and design. L.T., M.P., M.T.M., H.W., M.M.., T.L., M.G.L., W.Q.L., C.M., S.T., A.R., P.T., C.E., E.T., A.Y., G.C., S.P.P., S.T., and D.S. acquired the experimental data. L.T., M.P., M.T.M., H.W., M.M., P.K.H., T.L., M.G.L., W.Q.L., C.M., S.T., A.R.V., P.T., G.C., S.P.P., A.V., E.L., J.Z.R., A.U., J.N., X.F., S.R., D.S., and J.M.L. conducted the analysis and interpretation of the data. L.T., M.P., D.S., M.T.M., H.W., and A.U. performed statistical analyses. L.T., M.P., M.T.M., P.K.H., D.S., and J.M.L. drafted the manuscript. L.T., M.P., M.T.M., P.K.H., S.P.P., A.V., S.T., J.C., E.L., J.Z.R., A.U., J.N., X.F., S.R., D.S., and J.M.L. provided critical revision of the manucript. D.S. and J.M.L. supervised the study.

Acknowledgmenents

We thank Clara Rossi, Fellow at Mount Sinai, for her help.

Study 2 – Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice

Dhwanil A Dalwadi, **Laura Torrens**, Jordi Abril-Fornaguera, Roser Pinyol, Catherine Willoughby, Jeffrey Posey, Josep M Llovet, Christian Lanciault, David W Russell, Markus Grompe, Willscott E Naugler

Summary

Recombinant **AAV** is a widely used platform for gene replacement, silencing, and editing which holds great promise for the implementation of **gene therapies** in the clinical setting. In the past few years, two AAV-based gene therapies have been approved by regulatory agencies. Furthermore, ~130 active clinical trials are currently ongoing and could result in groundbreaking therapeutic advances for a great variety of medical conditions[4,105]. However, compelling studies provide evidence that AAV infection could induce **hepatocarcinogenesis due to integration** into oncogenic genomic sites, which poses serious safety concerns[105,106]. In this regard, AAV2 insertions have been identified in a small set of HCC patients (2-5%)[4,60] and murine models, often occurring in the ***Rian* locus**, a cluster of oncogenic microRNA located at the murine chromosome 12[106,107,158], which is analogous to the human *DLK1-DIO3* locus in chromosome 14. Furthermore, overexpression of the *DLK1-DIO3* locus has been identified in a subclass of 6-19% HCC tumors, associated with an aggressive phenotype and poor prognosis[108,109].

Common causes of chronic liver disease such as NAFLD could potentially favor AAV integration in the genome due to increased liver damage, inflammation, and regenerative proliferation of hepatocytes[110]. Considering the high prevalence of NAFLD, affecting up to 30% of the U.S. population[159], adverse events in these patients could be a concern for the use of AAV gene therapy. Herein **we aimed to assess whether NAFLD-associated liver damage increase the risk of AAV integration inducing HCC.**

To determine whether hepatocyte proliferation impacts AAV-induced oncogenesis, wildtype C57BL/6 mice were infected with an **AAV editing vector targeting the *Rian* locus** (AAV-Rian). Specifically, the AAV-Rian vector consisted of a strong promoter flanked by regions homologous to the *Rian* gene, which allow recombination and promoter insertion in that genomic region[158]. AAV infection was performed both at neonatal and adult mice. Furthermore, animals were treated either with HFD to induce **NAFLD-like liver injury**, or with partial **hepatectomy** to

promote hepatocyte proliferation. The molecular and inflammatory profiles in HCC and background liver samples from each group were characterized by **RNA sequencing**.

Our results revealed that:

1. **AAV targeting the *Rian* locus** (AAV-Rian) led to HCC in all male mice infected as **neonates**, likely due to growth-related hepatocyte proliferation in young mice. Conversely, only 5% of untreated male mice infected with AAV as adults developed HCC.

2. Liver injury through **HFD** and **partial hepatectomy** led to increased HCC development in adult AAV-Rian infected mice compared to untreated animals (100% vs 5%).

3. Transcriptomic analysis of murine tumors from AAV-Rian infected mice revealed remarkable molecular **similarities with human HCC tumors** overexpressing the *Rian* locus, associated with proliferation, aggressive phenotype, and poor prognosis.

4. NAFLD mice infected with a **non-targeted control AAV** developed HCC, though only half as frequently as those exposed to the targeted AAV-Rian (50% vs 100% in adults). These tumors presented increased expression of genes located in the *Rian* locus as well as a similar gene expression profile to AAV-Rian tumors. This suggests **random integrations** into the *Rian* locus, in line with results from previous studies[158].

5. Transcriptomic analysis of background liver samples from our murine model showed that HFD elicited a protumorigenic immune cancer field as well as activated lipid metabolism pathways. This could contribute to the AAV-induced in this model.

6. Female mice were less susceptible to develop AAV-Rian-induced HCC compared to males (29% vs 100% female and male neonates, respectively). Furthermore, livers from male mice with NAFLD treated with estrogen exhibited a reduction in aberrant immune exhaustion signaling compared to untreated males.

In conclusion, this study shows that AAV gene therapy increase HCC development in murine models of NAFLD, likely due to enhanced hepatocyte proliferation and the generation of a protumorigenic immune cancer field effect in the liver. This study raises concerns about the risks of AAV gene therapy causing HCC, particularly in patients with chronic liver injury such as NAFLD.

Molecular Therapy

Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice

Dhwanil A. Dalwadi,[1,2] Laura Torrens,[3] Jordi Abril-Fornaguera,[3] Roser Pinyol,[3] Catherine Willoughby,[3] Jeffrey Posey,[2] Josep M. Llovet,[3,4,5] Christian Lanciault,[6] David W. Russell,[7,8] Markus Grompe,[2] and Willscott E. Naugler[1]

[1]Department of Medicine, Division of Gastroenterology and Hepatology, Oregon Health and Science University, Portland, OR 97239, USA; [2]Papé Family, Pediatric Research Institute, Department of Pediatrics, Oregon Health and Science University, Portland, OR 97239, USA; [3]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS-Hospital Clínic, University of Barcelona, Catalonia, Spain; [4]Mount Sinai Liver Cancer Program, Divisions of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, NY, NY, USA; [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; [6]Department of Pathology, Oregon Health and Science University, Portland, OR 97239, USA; [7]Department of Medicine, University of Washington, Seattle, WA 98195, USA; [8]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Adeno-associated virus (AAV) integrates into host genomes at low frequency, but when integration occurs in oncogenic hotspots it can cause hepatocellular carcinoma (HCC). Given the possibility of recombinant AAV (rAAV) integration leading to HCC, common causes of liver inflammation like non-alcoholic fatty liver disease (NAFLD) may increase the risk of rAAV-induced HCC. A rAAV targeting the oncogenic mouse *Rian* locus was used, and as expected led to HCC in all mice infected as neonates, likely due to growth-related hepatocyte proliferation in young mice. Mice infected with rAAV as adults did not develop HCC unless they were fed a diet leading to NAFLD, with increased inflammation and hepatocyte proliferation. Female mice were less susceptible to rAAV-induced HCC, and male mice with NAFLD treated with estrogen exhibited less inflammation and immune exhaustion associated with oncogenesis compared to those without estrogen. Adult NAFLD mice infected with a non-targeted control rAAV also developed HCC, though only half as frequently as those exposed to the *Rian* targeted rAAV. This study shows that adult mice exposed to rAAV gene therapy in the context of chronic liver disease developed HCC at high frequency, and thus warrants further study in humans given the high prevalence of NAFLD in the population.

INTRODUCTION

Recombinant adeno-associated virus (rAAV) is a promising gene therapy tool and is currently widely used for this purpose in the laboratory and increasingly in the clinic.[1] Key advantages of rAAV as a gene therapy vector include its apparent lack of pathogenicity, ability to infect both dividing and quiescent cells, and a very mild immune response, though this is to some extent dependent on the dose.[2] There are more than a hundred ongoing rAAV clinical trials along with at least two FDA approved therapies in the US. Although rAAV is a promising gene therapy vector, concern remains regarding potential side-effects, notably insertional mutagenesis in oncogenes or tumor suppressor genes with resultant carcinogenesis. While random rAAV insertions might result in cancer in any tissue, both human and mouse data suggest that hepatocellular carcinoma (HCC) is the most likely malignancy.[3,4]

The majority of experimental evidence to date suggests that rAAV gene therapy is safe, but there are some compelling rodent studies that provide evidence of potential genotoxicity of rAAV vectors.[5] For example, when neonatal mice were infected with a rAAV gene therapy vector, the mice that received the vector were four to seven times more prone to developing HCC than control mice, and the oncogenic vector integration site was mapped to the *Rian* locus on chromosome 12.[6,7] When a gene-editing rAAV vector was used to insert a strong promoter specifically into the *Rian* locus, 100% of neonatal mice infected with the vector developed HCC.[4] However, to date no evidence shows that adult mice develop HCC when treated with rAAV vectors.[8,9] It is hypothesized that proliferating hepatocytes in the neonatal liver are responsible for the high frequency of homologous and non-homologous rAAV integrations. It follows, then, that the absence of proliferating hepatocytes in the healthy adult liver would diminish rAAV integration and resultant oncogenesis.

While quiescent under healthy homeostatic conditions, however, hepatocytes readily proliferate in the adult liver in response to injury.[10–13] The clearest example of such proliferation occurs after partial hepatectomy, wherein hepatocytes in the remnant liver have a striking response to the injury—marked hepatocyte DNA

RESULTS
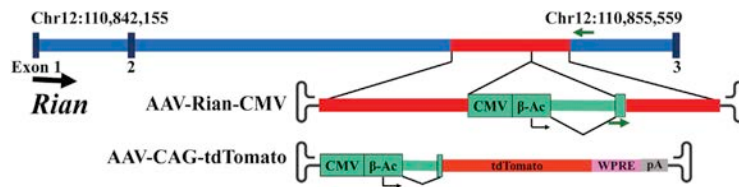
Figure 1. Schematic of *Rian* Targeting Vector and a Control tdTomato Vector

The AAV-Rian vector consists of a CAG promoter flanked by arms of homology to *Rian*. The green arrows indicate the positions of primers, one inside the CAG promoter and one outside of the arm of homology, used to detect CAG integration by homologous recombination. AAV-tdTomato is a non-editing vector that expresses the tdTomato transgene driven by the CAG promoter.

synthesis and proliferation.[10,14] More commonly seen (especially in humans) are conditions of chronic liver inflammation and injury, which are themselves associated with compensatory proliferation as the body seeks to repair the damaged liver.[11,12] Multiple conditions of chronic liver inflammation/injury/proliferation make up the bulk of human liver disease and include entities such as alcoholic hepatitis, hepatitis B and C infections, and various auto-immune conditions. The most common cause of chronic liver inflammation in the US and Western countries is non-alcoholic fatty liver disease (NAFLD), affecting up to 30% of the US population.[15] The high prevalence of obesity and NAFLD makes it likely to be present in a significant number of patients who could benefit from gene therapy.

In this study we assessed whether hepatocyte replication and liver injury increase HCC formation in adult mice who receive rAAV gene therapy. Given the high prevalence of inflammatory liver conditions such as NAFLD, adverse events in these patients could be a concern for the use of rAAV gene therapy. To test this hypothesis, we infected wild-type C57BL/6 mice with a rAAV editing vector targeting the *Rian* locus by homologous recombination.[4] Partial hepatectomy was used to induce hepatocyte proliferation and a high-fat diet (HFD) to induce liver injury.[10,16] The results of this study demonstrate that both partial hepatectomy and HFD-induced liver injury are sufficient to cause HCC if a strong promoter integration occurs in an oncogenic locus. This study raises concerns about the risks of rAAV gene therapy causing HCC in patients with chronic liver injury, particularly patients suffering from non-alcoholic steatohepatitis (NASH) or other inflammatory liver conditions.

RESULTS

Neonatal mice infected with AAV-Rian-CMV (AAV-Rian), a gene editing vector that inserts a strong CAG promoter in the *Rian* locus, all developed HCC as a result of a *Rian*-specific rAAV integration.[4] The development of HCC in neonates was attributed to the hepatocyte proliferation expected in the neonatal liver. However, there are currently no reports of adult mice developing HCC when infected with rAAV. In this study we report for the first time that rAAV can cause HCC in adult mice under conditions causing hepatocyte proliferation. Two models were utilized, a neonatal model where the mice were infected with rAAV followed by initiation of HFD or regular diet (RD) and an adult model where mice were first started on HFD or RD, followed by rAAV infection.

HFD-Induced Chronic Liver Injury Does Not Increase Hepatocarcinogenesis in rAAV-Infected Neonates

1-day-old neonatal mice were injected with 3×10^{10} vg of AAV-Rian or AAV-CAG-tdTomato (AAV-tdTomato) (Figure 1) via the temporal vein. At 3 weeks old, they were either started on a HFD or RD, and at 6 months of age the mice were sacrificed and assessed for tumors (Figure 2A). Gross inspection of livers revealed multiple nodules in vector-infected mice (Figure 2D), whereas mice that received vehicle injections (No rAAV, Dulbecco's PBS [DPBS], n = 6, 3 males, 3 females) did not develop tumors regardless of diet (data not shown). Mice that received the tdTomato virus and were on the RD did not develop tumors (n = 6, 3 males, 3 females); however, 20% of males and 25% of females that were on the HFD developed tumors (n = 9, 4 males, 5 females), though this difference was not statistically significant compared to RD by chi-square test (p = 0.54 for females and p = 0.41 for males; Figure 2B). All males that received the *Rian* virus developed tumors, regardless of diet (n = 21, 14 HFD, 7 RD), whereas only 29% of females developed tumors when on the RD, and 20% developed tumors when on the HFD (Figure 2B, n = 13, 5 HFD, 8 RD). The average number of tumors per mouse in mice that received the tdTomato virus on the HFD was 0.25 for male (n = 4) and 0.2 for females (n = 5; Figure 2B). In mice that received the *Rian* virus, the tumor burden was higher in males than females with an average of 9 tumors per mouse when on the RD (n = 7), and 10 when on the HFD (n = 14; Figure 2C). Tumor burden in female mice was lower, 0.25 tumors per mouse for mice on RD (n = 8) and 5.4 for mice on HFD (n = 5), however, the variance was such that the difference was not statistically significant (p > 0.05; Figure 2B). Histological analysis showed that the tumors had characteristic features of HCC such as cytologic atypia with bizarre, enlarged hepatocytes and wide trabeculae, and pseudoacinar formation (Figure 2E). At the chosen time points inflammatory liver injury did not change AAV-induced HCC initiation or progression in the neonatal model.

Hepatocyte Proliferation Caused by Chronic Liver Injury or Partial Hepatectomy Leads to rAAV-Induced HCC in Adult Mice

To determine whether liver injury and regeneration affects rAAV-induced HCC development in adult mice, we started 3-week-old male mice on HFD or RD and infected them at full adult maturity (10 weeks old) with either AAV-Rian or an AAV-tdTomato control virus. To stimulate hepatocyte proliferation, a subset of rAAV-infected mice on the RD underwent a 2/3 partial hepatectomy 1 week after viral infection. 6 months after viral infection, the mice were euthanized and assessed for tumors (Figures 3A and 3D). None of
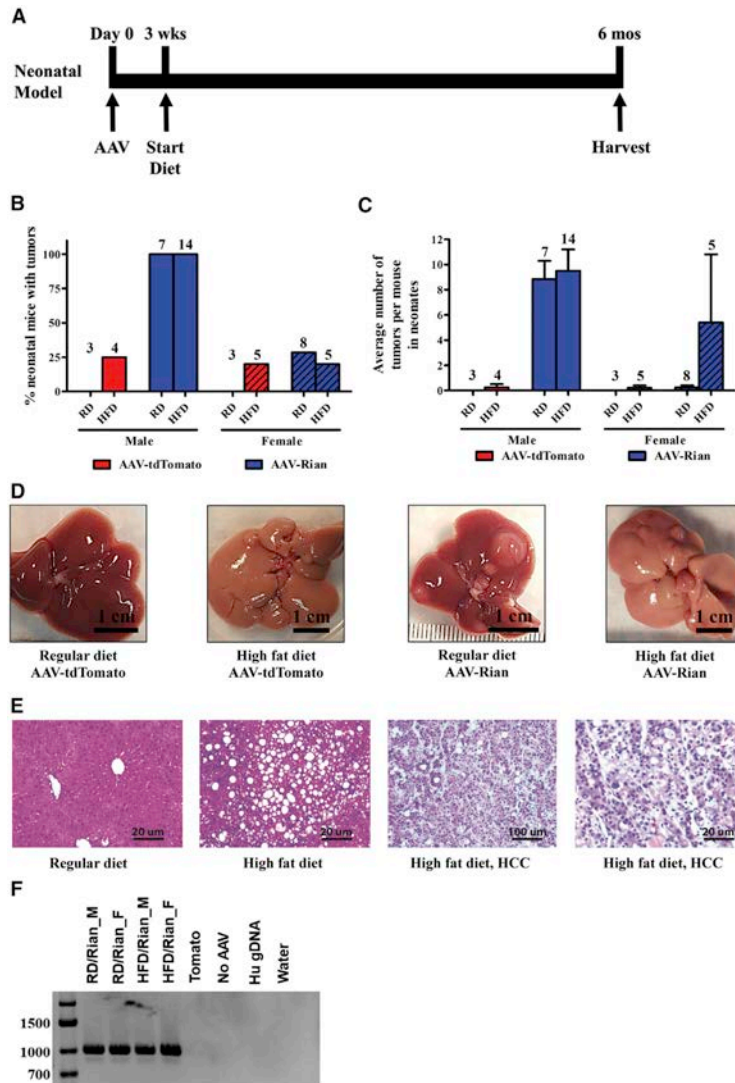
Figure 2. HFD-Induced Liver Injury Does Not Exacerbate AAV-Induced HCC in the Neonatal Mouse Model

(A) Outline of experimental design (n = 3–14 per group). (B) Tumor incidence in neonatal mice infected with AAV-Rian or AAV-tdTomato in presence or absence of diet induced liver injury. Tumor incidence was not statistically significant between diet groups by chi-square analysis (p > 0.05). Numbers above the bars represent the sample size. (C) Tumor burden at 6 months. No statistical significance was achieved between diet groups by one-way ANOVA followed by Bonferroni post hoc (p > 0.05). Numbers above the bars represent the sample size. (D) Images of gross liver specimens from 6-month-old mice. (E) Representative H&E staining of livers from mice on RD and HFD. (F) Representative gel of CAG insertion in Rian tumors by integration PCR, where one primer targets the CAG region and the other primer targets a region outside of the homology arm. Four tumors, two from female and two from male tumors, were selected. No integration was detected in mice on RD infected with AAV-tdTomato.

average tumor burden of 16 (n = 5, Figures 3B and 3C). Both HFD and partial hepatectomy had a significant effect on tumor incidence and burden compared to mice on RD that received the *Rian* virus (p < 0.01). Histologic findings of HCC were similar to the neonatal livers (Figure 3E).

CAG Integration in the *Rian* Locus Activates Oncogenic Pathways

Several tumors and background liver were dissected and assessed for integration of the CAG promoter into the *Rian* locus by homologous recombination. Primers were designed such that one primer was in the 3′ end of the CAG promoter and another in the flanking gDNA outside the homology arm. All of the Rian tumors from the neonatal experiment and the adult mice experiment (AAV-Rian tumors from mice on HFD and those that underwent PH) had a targeted CAG integrations (Figures 2F and 3F), which were confirmed by

the mice on the RD that received the tdTomato virus developed tumors (n = 5), but 50% of the mice on the HFD (n = 10) developed tumors, with an average tumor burden of approximately 3 tumor nodules per mouse (Figures 3B and 3C). However, the tumor incidence (Figure 3B, chi-square test p = 0.10) and burden (Figure 3C, one-way ANOVA p > 0.05) were not statistically different. Only 5% of the mice on the RD that received the *Rian* vector developed tumors, with an average tumor burden of 0.05 (n = 20), whereas 100% of the mice on the HFD developed tumors with a tumor burden of 6 (n = 10, p < 0.001; Figures 3B and 3C). Similarly, all of the mice on the RD that underwent a 2/3 partial hepatectomy developed tumors, with an

sequencing. Bands seen in the RD/tdTomato and RD/Rian group were sequenced and did not contain Rian or CAG sequences, suggesting nonspecific amplification.

In order to determine whether the tumors from this study presented a Rian gene-expression profile similar to prior studies, a gene signature (AAV-Rian signature) recapitulating the top differentially expressed genes was generated.[4] Single-sample gene set enrichment analysis (ssGSEA) and nearest template prediction (NTP) analysis were performed to assess whether the tumors had the AAV-Rian signature using non-tumor samples as controls (Figure 4A). ssGSEA analysis showed

RESULTS

Figure 3. HFD-Induced Liver Injury and Partial Hepatectomy Exacerbates AAV-Induced HCC in the Adult Mouse Model

(A) Outline of experimental design (n = 5–20 per group). (B) Tumor incidence in adult mice infected with AAV. HFD increased Rian induced tumor incidence (chi-square test, p < 0.05). Numbers above the bars represent the sample size. (C) Tumor burden in adult mice infected with AAV. Both HFD and partial hepatectomy increased tumor burden in AAV-Rian-CMV infected mice (one-way ANOVA, Bonferroni post hoc, p < 0.01). Numbers above the bars represent the sample size. (D) Images of gross liver specimens

(legend continued on next page)

110

an enrichment of the AAV-Rian signature in the tumor samples. The "AAV-Rian-UP" set of genes was significantly enriched (p < 0.0001) in tumor samples and the "AAV-Rian-DOWN" set of genes were significantly enriched in the non-tumoral samples (p < 0.0001). The NTP approach also demonstrated a similar AAV-Rian signature in the tumors. Interestingly, the tumor from the RD-Rian-Female mouse did not have the Rian signature, in line with the different clustering of female tumors (n = 2) revealed by non-negative matrix factorization (NMF) analysis (Figure S1) but did have a discernibly reduced expression of the main HCC-related oncogenic pathways. Nevertheless, the RD-Rian-Female tumor did have an increase in gene expression in/ near the *Rian* locus where CAG integrated compared to background liver (Figure 4A). In contrast to the female tumors, the HFD-Tomato tumor (n = 1) clustered with the AAV-Rian tumors (Figure S1) and had a similar AAV-Rian signature (Figure 4A; Figure S2). The gene-expression pattern of the tomato tumor strongly suggests random integration into the *Rian* locus, given that no other known genomic targets would produce the same pattern. Overall, these results indicate that the gene-expression profile of these tumors highly resemble the gene expression profile of the *Rian* tumors reported previously.[4]

Gene-expression profiles of the tumors was assessed to determine whether they recapitulated any of the molecular classes that have been described for human HCC. Interestingly, the *Rian* tumors expressed a molecular profile similar to a subclass of human HCC associated with proliferative and progenitor like features and a poor prognosis (Figure 4B).[17–19] Additional HCC signaling pathways involved enrichment of a gene signature associated with poor outcome (i.e., poor-survival signature), including positivity for the Met signature (associated with MET activation), EpCAM-positivity, AKT signaling, IGFR1 positivity, WNT/transforming growth factor β (TGF-β) activation, and NOTCH1 signaling.[20–25] The above described molecular profile of the *Rian* tumors parallels a subtype of human HCCs with overexpression of a cluster of microRNAs (miRNAs) located in the *MEG8* locus.[26] Moreover this group of human HCC is associated with molecular signatures of poor prognosis similar to that seen in *Rian* tumors.[27] Enrichment of EpCAM and NOTCH1 signaling signatures is probably due to certain biological differences between murine tumors and human HCC. Murine tumors tend to be very aggressive, which is linked to more deregulated pathways than human tumors.

Estrogen Partially Ameliorates HFD-Induced Inflammation

Results of this study confirmed a delay in initiation of HCC in females compared to males as seen in other studies.[4,28] Since estrogen has anti-inflammatory properties, it was hypothesized that estrogen may be involved in delaying the initiation of HCC. To test this, 3-week-old male mice were started on HFD or RD, received tri-weekly

intraperitoneal (i.p.) injections of 80 μg/kg estrogen for 1 month, and were then sacrificed for analysis. H&E staining showed a small but noticeable reduction in liver damage and fat in male mice on HFD that received estrogen (Figure 5B). Further, estrogen treatment resulted in a small (∼4%) but significant reduction in hepatocyte proliferation (Figure 6). This reduction in proliferation was not captured in the RNA sequencing (RNA-seq) analysis, likely due to the overall small proportion of proliferating hepatocytes in the liver.

RNA-seq of background livers from mice on RD and HFD, with or without estrogen did show differences in immune-related pathways, oncogenic pathways, lipid metabolism, and liver function. ssGSEA indicated that background livers from mice fed the HFD presented an increased immune cancer field, recently defined by our group in adjacent liver tissue from HCC patients, which is associated with inflammatory signaling and higher risk of HCC development.[29] This is in line with the increased inflammation of the liver observed by the enrichment in inflammatory-related signaling pathways such as hallmarks of inflammatory response, interferon-α (IFN-α) or IFN-γ response, signatures of T cell exhaustion or presence of T regs (Figure 5C). The gene expression pathways in HFD mice treated with E2 are similar in most respect in HFD mice not treated with E2. This likely indicates that estrogen has little role in ameliorating the NASH condition at the dose tested. However, with the estrogen treatment we observed an amelioration of the immune exhaustion features, including a decrease in TNF-α, TGF-β, and Wnt/β-catenin signaling, an increase of antigen processing and presentation, Toll-like receptor signaling, and activation of the adaptive immune system including general activation of B and T cells. Livers from mice fed the HFD also presented upregulation of some oncogenic pathways (Myc, MTORC, or EGFR) and increased lipid metabolism, and downregulation of other metabolic pathways (i.e., bile acid or drug metabolism).

DISCUSSION

Here we show for the first time that adult mice that received a rAAV gene targeting vector developed HCC where hepatocyte proliferation is increased due to NAFLD and partial hepatectomy. Adult mice typically do not develop rAAV-induced liver cancer in the absence of injury.[8,9] We hypothesized that the age-dependence of gene therapy induced cancer is due to the rate of hepatocyte proliferation, which is high in neonates due to natural liver growth. To determine whether the proliferation rate impacts rAAV induced oncogenesis, we evaluated two injury regimens. Hepatocyte cell division is increased after both partial hepatectomy and fat-induced liver injury. Both conditions led to the development of HCC in adult mice, which raises concerns for the use of rAAV therapy in the general human population, where significant numbers have chronic inflammatory liver diseases. In the US, for example, 0.34% of the population have HBV infection,

from 9-month-old mice. (E) Representative H&E staining of livers from mice on RD and HFD. (F) Representative gel of CAG insertion in Rian tumors by integration PCR. Target PCR product was not observed in mice on RD infected with either Rian or tdTomato virus but was observed in mouse on HFD infected with Rian and in mice that had received a partial hepatectomy. The bands seen in the RD/tdTomato and RD/Rian lanes were sequenced and did not show CAG or Rian sequences, suggesting that these bands are artifacts.
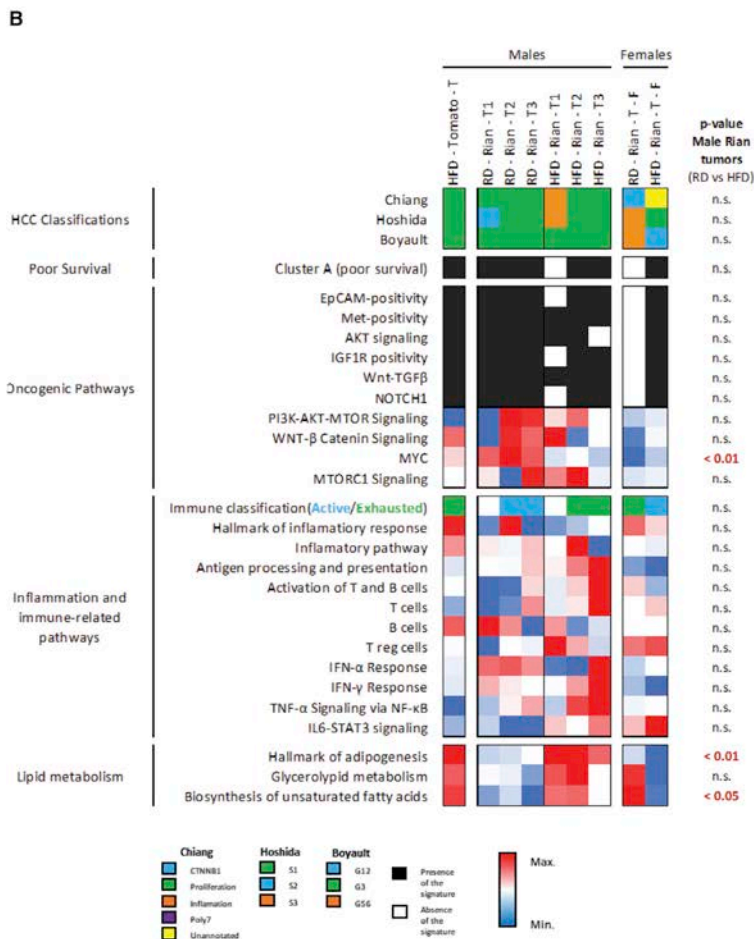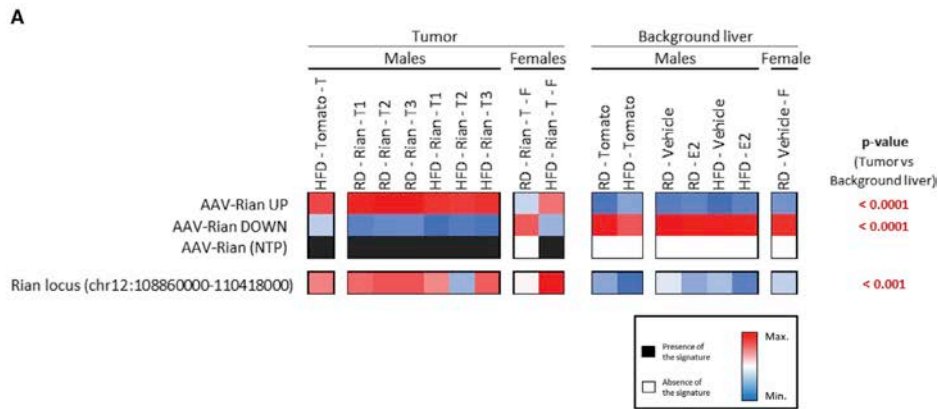
(legend on next page)
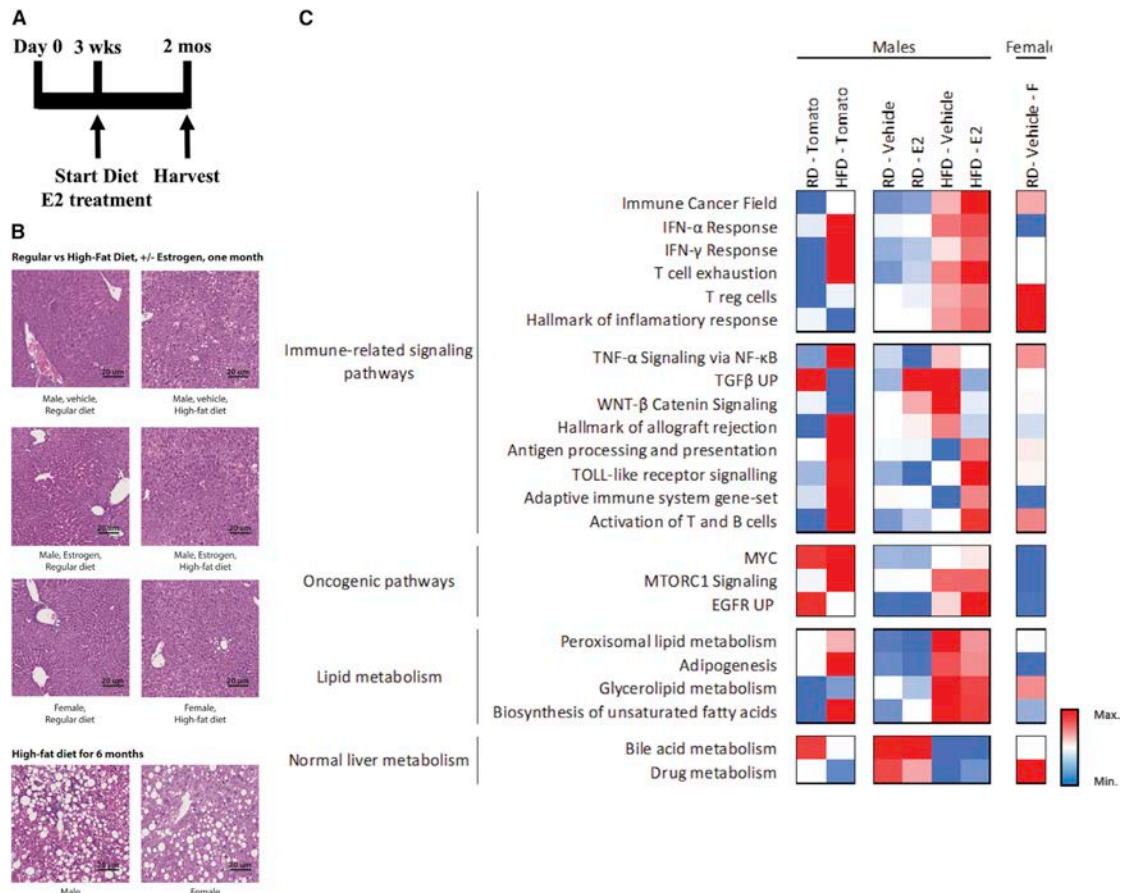
Figure 5. Estrogen Partially Ameliorates HFD-Induced Liver Injury

(A) Outline of experimental design. (B) Representative H&E liver images of mice on RD or HFD for 1 month or 6 months. Females have slightly lower HFD-induced injury at both time points, and male mice on HFD+E2 also had less injury than males on HFD that didn't receive E2. (C) Heatmap showing the expression of immune-related pathways, oncogenic pathways, lipid metabolism, and liver function in the background liver samples.

1.7% have HCV infection, 4.3% have alcoholic liver disease, and up to 30% have fatty liver disease.[15,30–32] These inflammatory conditions alone are risk factors for HCC development, thus adding a rAAV-associated oncogenic risk may significantly diminish the possible therapeutic benefits of gene therapy.

Inflammation has long been associated with carcinogenesis, and hepatocarcinogenesis is no exception.[33] Most human HCCs arise in the background of chronic liver diseases characterized by injury and inflammation. Inflammation itself may promote hepatocarcinogenesis through several mechanisms, such as DNA damage from reactive oxygen species, changes in the immune system milieu, and an increase in hepatocyte proliferation.[34] This study demonstrates an increase in the risk of rAAV-induced HCC in adult mice fed with HFD to approximate NAFLD, a common chronic liver disease in humans. While the exact mechanism for this increase is unclear, the most likely reason is an increase in hepatocyte proliferation. AAV is more likely to integrate into proliferating hepatocytes as

Figure 4. Tumors Exhibit a Prototypical Rian Signature and Have a Similar Transcriptomic Profile to Human HCC Subclassifications

(A) Heatmap showing the AAV-Rian signature and the "AAV-Rian UP" and "AAV-Rian Down" gene sets generated from the Wang et al.[4] study, and analysis of the expression of the genes in the *Rian* Locus (GRCm38/mm10, chr12:108860000–110418000). (B) HCC classification, survival signatures, oncogenic pathways, and lipid metabolism activation in the tumor samples of the study. p Values were calculated comparing diets within male Rian tumor samples.
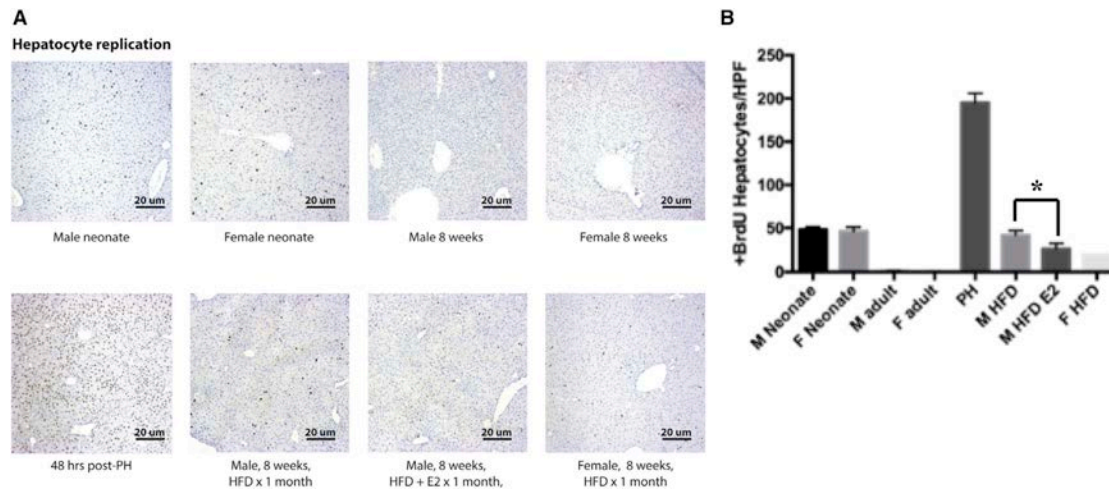
Figure 6. Estrogen Treatment Partially Suppressed Hepatocyte Proliferation
(A) BrdU staining (brown dots) identifies dividing hepatocytes. (B) Morphometric analysis of BrdU hepatocytes. Estrogen partially reversed HFD-induced injury (Student's t test, p < 0.05).

seen in the adult HFD model.[13,35,36] These findings raise concern regarding rAAV gene therapy in humans, where resultant HCC has not been detected. In patients with chronic liver disease like NAFLD, and with other viral infections like hepatitis B and C, the risk of random rAAV integration is greater due to the proliferating hepatocytes associated with chronic liver injury. The correlation between AAV gene therapy and other viral infection is not extensively studied and is an area that would benefit from further investigation. One caveat to generalizing these mouse data to humans, however, is the possibility that mice may have a lower threshold for HCC development than humans.

Males have long been known to have a significantly higher incidence of HCC compared to females, and the mechanism is likely related to a relative increase in hepatic inflammation for a given injury.[28,37] Here we confirm that female mice are less susceptible to rAAV-induced HCC compared to males. We show that estrogen treatment of male mice fed a HFD partially ameliorates inflammation and proliferation. Interestingly, estrogen also altered the immune milieu in males with NAFLD, decreasing immune exhaustion associated with activation of oncogenic pathways. CD4+ lymphocyte loss in patients with NAFLD was found to promote hepatocarcinogenesis, and the link here found between estrogen and liver inflammation suggests a further connection explaining gender disparity seen in hepatocarcinogenesis.[38]

We observed a high incidence of HCC in adult mice on a HFD that received the control tdTomato AAV. Unlike the *Rian* AAV, the tdTomato virus does not have the capability for targeted integration into the genome. Nonetheless, we found 5 of 10 mice on the HFD that received the tdTomato AAV developed HCC, some of which had the same tumor signature as the AAV-Rian-induced HCC. There

are numerous reports that show that AAV randomly integrates in host genomes and that these sites are scattered throughout the genome.[3,39–42] It is likely that the virus integrated in other sites as well, some of which could have been in oncogenic loci; however, we are not aware of any other oncogenic loci that have a similar oncogenic signature as *Rian*. Although there is not sufficient data in this report talk about the frequency of integration in oncogenic loci, the *Rian* locus appears to be a hotspot for random integration since the tomato tumors analyzed had a similar, but slightly shifted gene activation profile as the tumors derived from a targeted insertion (Figure S2). The fact that tumors developed only in those on the HFD points toward the same factors noted above, namely that inflammation and increased hepatocyte proliferation increase HCC incidence. Whether the human liver is similarly susceptible to random rAAV integrations in the *MEG8* locus (analogous to mouse *Rian*) leading to HCC is unclear but should be studied.

We did not observe a difference in initiation or progression of HCC in mice infected with AAV as neonates, though this may be a result of assessing the mice at a time point where all of the control mice already had cancer. Indeed, a small number of mice on a HFD who had neonatal AAV-infection died before the pre-determined 6-month period (data not shown, but necropsied mice all had liver tumors). We suspect that acceleration of hepatocarcinogenesis occurred in the AAV-infected neonates on HFD, but that our pre-determined experimental end (6 months) was too late to document changes such as the earlier appearance of HCC. We chose the 6-month time point in the neonates based on prior studies, which led to 100% of mice developing HCC.[4] To understand if HCC developed earlier in AAV-infected neonates on a HFD, earlier time points would need to be assessed, prior to 100% of mice developing HCC. Indeed, there

was a trend toward higher tumor incidence and burden in both genders and ages of AAV-infected mice fed the HFD but did not reach statistical significance. Given the many conditions that were tested, the sample size per group became relatively small, which could explain why some of the trends did not reach significance. Viral load is another parameter that could have influenced hepatocarcinogenesis in adult mice. It is possible that a higher AAV dose could result in higher HCC incidence, however, we did not assess whether there might be a dose response effect of the virus on hepatocarcinogenesis.

In conclusion, this study demonstrates that adult mice infected with both targeted and non-targeted rAAV develop HCC when there is also a stimulus for hepatocyte proliferation such as fatty liver induced by a HFD. Given the high prevalence of inflammatory liver conditions in the general population, this should raise an alarm about the use of rAAV, an otherwise promising vector used for diverse gene therapy application. Female mice are less susceptible to rAAV-induced HCC compared to males, part of which is due to a more favorable immune milieu related to estrogen and likely applies to humans given the documented gender disparity. More studies are needed to assess the risks of rAAV-induced hepatocarcinogenesis given the promise and widespread use of rAAV for gene therapies.

MATERIALS AND METHODS

Vector Production

The AAV-Rian-CMV plasmid was obtained from David W. Russell (University of Washington).[4] The vector was packaged in the AAV-DJ serotype by cotransfection of HEK293 cells with the vector plasmid pAAV-Rian-CMV, pDJ (capsid plasmid), and pHelper.[43] Briefly, 7×10^6 HEK293T cells were plated in 10-cm dishes, cultured in DMEM complete (10% FBS, 100 U/mL penicillin, 100 μg/mL streptomycin, 2 mM GlutaMax, 1 mM sodium pyruvate). After 16 to 20 h, media was replaced with reduced serum DMEM (2% FBS) followed by polyethylenimine (PEI)-based triple transfection. Equimolar quantities of each plasmid was combined with a total mass of 24 μg in OptiMEM (GIBCO, Grand Island, NY) media and 3 times the plasmid weight of PEI (72 μg from 1 mg/mL stock solution, pH 7.1, Polysciences, Warrington, PA) were combined, with a final volume of 1.5 mL per 10-cm dish. The transfection mixture was incubated at room temperature for 30 min followed by addition to cells. 5 days post transfection the media was collected and centrifuged twice at $1,900 \times g$ for 30 min. The supernatant was treated with benzonase (Sigma, St. Louis, MO) at 37°C for 1 h and then centrifuged at $1,900 \times g$ for 30 min. The supernatant was filtered through 0.45 μm PES vacuum filter, precipitated with PEG −8,000 (Spectrum Chemical, Gardena, CA; 8% w/v PEG-8000, 0.5 M NaCl) overnight, and then centrifuged at $6,000 \times g$ for 1 h. The viral pellet was resuspended in DPBS. The viral titer was determined by a dot-blot assay as described previously.[44]

Animal Care

C57BL/6J neonatal mice were bred at OHSU and 3-week-old C57BL/6J male mice were purchased from Jackson Laboratory. All animal experiments were approved by the Oregon Health & Science University Institutional Animal Care and Use Committee (Portland, OR) and performed in accordance with the approved protocols.

Effect of Diet-Induced Liver Injury on AAV Induced HCC in Neonates

1-day-old mice received either 3×10^{10} viral genomes of AAV-Rian-CMV (AAV-Rian) or AAV-CAG-tdTomato (AAV-tdTomato; tdTomato reporter gene driven by the CAG promoter) vectors in 16 μL DPBS or DPBS alone via superficial temporal vein injections. This dose has been shown to cause to HCC in neonatally infected mice.[4] At 3 weeks of age, mice were either placed on a RD (5LOD) or a HFD (Envigo, Madison, WI). Table S1 summarizes the different groups and sample size. Mice were sacrificed at 6 months of age and their livers were harvested. Tumor burden was assessed by counting visible nodules and calculating the average number per mouse. Sections of the liver were fixed in 4% PFA (v/v) or frozen at −80°C for histological and molecular analysis.

Effect of Age- and Diet-Induced Liver Injury on AAV-Induced HCC

3-week-old mice (purchased from Jackson Laboratory) were placed on either RD or HFD. When the mice were 10 weeks old, they received 8×10^{11} AAV-Rian or AAV-tdTomato vectors in 100 μL DPBS via retro-orbital injection. This dose reflects a weight-based calculation of virus analogous to the dose given to the neonatal mice. A week later (11 weeks), a subset of mice on the regular diet received a 2/3 partial hepatectomy (PH). The mice were sacrificed and their livers were harvested at 9.5 months of age. Table S2 summarizes the different groups and sample size. Tumor burden assessment and liver sample processing was performed as described in the previous section.

Effect of Estrogen on HFD-Induced Liver Injury

3-week-old mice (purchased from Jackson Laboratory) were either on a HFD or a RD and received either estrogen (80 μg/kg dissolved in corn oil, Sigma, St. Louis, MO) or vehicle (corn oil) every other day for 1 month (5 mice per group). 2 h before tissue harvest, mice were injected with 100 mg/kg bromodeoxyuridine (BrdU; Alfa Aesar, Ward Hill, MA) to label actively dividing hepatocytes. At the end of treatment period, livers were harvested and sections were fixed in 4% PFA or frozen for histology and molecular analysis.

Histology

Tissues were fixed in 4% PFA (v/v) and then transferred to 70% (v/v) EtOH. H&E staining was performed by the Digestive Diseases Center at Texas Children's Hospital.

DNA Isolation and Vector Integration Analysis

1 mg of tissue was used for DNA isolation using the MasterPure Complete DNA and RNA Purification kit (Lucigen, Middleton, WI) per manufacturer's protocol, followed by phenol/chloroform extraction and ethanol precipitation. The DNA pellet was resuspended in TE buffer and the concentration was measured using the

RESULTS

Qubit Fluorometer. Integration of AAV-Rian was assessed by PCR and sequencing. Primers were designed such that the forward primer (5′-GCTCCTGGGCAACGTGCTGGT-3′) was complementary to the CAG promoter and the reverse primer (5′-TGGAAGAGCCGG-GAAGCCTTTGA-3′) was outside of the AAV-Rian homology arm (Figure 1A). PCR was performed using the 2× PrimeStar polymerase mix (Takara, Mountain View, CA) per manufacturer's protocol. Briefly, 100 ng of gDNA and 0.2 μM of each primer was used in a 50 μL reaction. Thermocycler condition was as follows: denatured 98°C for 10 s, followed by extension at 68°C for 45 s, for 30 cycles; the expected product size is 951 base pairs. PCR products were analyzed on a 0.8% agarose gel and verified by sequencing (GeneWiz, South Plainfield, NJ).

RNA-Seq and Analysis

1 mg of tissue was homogenized in RNAzol RT (MRC, Cincinnati, OH) and RNA was extracted per manufacturer's protocol. RNA was resuspended in nuclease free water and sent to GeneWiz for library prep and sequencing. Three independent Rian tumors from the neonatal model on HFD and RD were used for library preparation. RNA from three individual livers were pooled to prepare one library for RD/tdTomato and HFD/tdTomato background liver. Only a single nodule was used to prepare libraries for HFD/tdTomato male tumor and HFD/tdTomato female tumor. RNA from liver sections of three individual mice were pooled for each of the following groups from the estrogen experiment: RD/vehicle male, RD/vehicle female, HFD/vehicle male, RD/estrogen male, and HFD/estrogen male (a single library was prepared for each of the five groups). The data was standardized to fragments per kilobase of transcript per million mapped reads (FPKM). See Supplemental Materials and Methods sections for a detailed description of data analysis.

Statistical Analysis

All data presented as mean ± SEM and statistical significance was assessed by one-way ANOVA followed by Bonferroni post hoc test or Student's t test. Chi-square test was used to assess significance of tumor incidence. For RNA-seq data, correlations for categorical variables were analyzed by Fisher exact test. In all cases, $p < 0.05$ (two-tailed) or FDR < 0.05 was considered significant. Methods used in the statistical analysis of RNA-seq data are noted in the Supplemental Materials and Methods section.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ymthe.2020.10.018.

AUTHOR CONTRIBUTIONS

D.A.D.: Designed and performed experiments, analyzed data, and wrote the manuscript. D.W.R., M.G., and W.E.N.: Conceived the project, helped design experiments, and helped write the manuscript. J.P.: Quantified AAV titers. C.L.: Analyzed HCC histology. L.T., J.A.-F., R.P., C.W., and J.M.L.: Performed bioinformatics analysis and helped review the manuscript.

REFERENCES

1. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. 18, 358–378.

2. Hinderer, C., Katz, N., Buza, E.L., Dyer, C., Goode, T., Bell, P., Richman, L.K., and Wilson, J.M. (2018). Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. 29, 285–298.

3. Nault, J.C., Datta, S., Imbeaud, S., Franconi, A., Mallet, M., Couchy, G., Letouzé, E., Pilati, C., Verret, B., Blanc, J.F., et al. (2015). Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. Nat. Genet. 47, 1187–1193.

4. Wang, P.R., Xu, M., Toffanin, S., Li, Y., Llovet, J.M., and Russell, D.W. (2012). Induction of hepatocellular carcinoma by in vivo gene targeting. Proc. Natl. Acad. Sci. USA 109, 11264–11269.

5. Chandler, R.J., LaFave, M.C., Varshney, G.K., Trivedi, N.S., Carrillo-Carrasco, N., Senac, J.S., Wu, W., Hoffmann, V., Elkahloun, A.G., Burgess, S.M., and Venditti, C.P. (2015). Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. J. Clin. Invest. 125, 870–880.

6. Donsante, A., Miller, D.G., Li, Y., Vogler, C., Brunt, E.M., Russell, D.W., and Sands, M.S. (2007). AAV vector integration sites in mouse hepatocellular carcinoma. Science 317, 477.

7. Donsante, A., Vogler, C., Muzyczka, N., Crawford, J.M., Barker, J., Flotte, T., Campbell-Thompson, M., Daly, T., and Sands, M.S. (2001). Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther. 8, 1343–1346.

8. Bell, P., Wang, L., Lebherz, C., Flieder, D.B., Bove, M.S., Wu, D., Gao, G.P., Wilson, J.M., and Wivel, N.A. (2005). No evidence for tumorigenesis of AAV vectors in a large-scale study in mice. Mol. Ther. 12, 299–306.

9. Li, H., Malani, N., Hamilton, S.R., Schlachterman, A., Bussadori, G., Edmonson, S.E., Shah, R., Arruda, V.R., Mingozzi, F., Wright, J.F., et al. (2011). Assessing the potential for AAV vector genotoxicity in a murine model. Blood 117, 3311–3319.

10. Taub, R. (2004). Liver regeneration: from myth to mechanism. Nat. Rev. Mol. Cell Biol. 5, 836–847.

11. Michalopoulos, G.K. (2017). Hepatostat: Liver regeneration and normal liver tissue maintenance. Hepatology 65, 1384–1392.

12. Forbes, S.J., and Newsome, P.N. (2016). Liver regeneration - mechanisms and models to clinical application. Nat. Rev. Gastroenterol. Hepatol. 13, 473–485.

13. Cast, A., Kumbaji, M., D'Souza, A., Rodriguez, K., Gupta, A., Karns, R., Timchenko, L., and Timchenko, N. (2019). Liver Proliferation Is an Essential Driver of Fibrosis in Mouse Models of Nonalcoholic Fatty Liver Disease. Hepatol. Commun. 3, 1036–1049.

14. Michalopoulos, G.K. (2010). Liver regeneration after partial hepatectomy: critical analysis of mechanistic dilemmas. Am. J. Pathol. 176, 2–13.

15. Cotter, T.G., and Rinella, M. (2020). Nonalcoholic Fatty Liver Disease 2020: The State of the Disease. Gastroenterology. 158, 1851–1864.

16. Tsuchida, T., Lee, Y.A., Fujiwara, N., Ybanez, M., Allen, B., Martins, S., Fiel, M.I., Goossens, N., Chou, H.I., Hoshida, Y., and Friedman, S.L. (2018). A simple diet- and chemical-induced murine NASH model with rapid progression of steatohepatitis, fibrosis and liver cancer. J. Hepatol. 69, 385–395.

17. Boyault, S., Rickman, D.S., de Reyniès, 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.

18. Hoshida, Y., Nijman, S.M., Kobayashi, M., Chan, J.A., Brunet, J.P., Chiang, D.Y., 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.

19. Llovet, J.M., Montal, R., Sia, D., and Finn, R.S. (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. Nat. Rev. Clin. Oncol. 15, 599–616.

20. Kaposi-Novak, P., Lee, J.S., Gòmez-Quiroz, L., Coulouarn, C., Factor, V.M., and Thorgeirsson, S.S. (2006). Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. J. Clin. Invest. 116, 1582–1595.

21. Yamashita, T., Forgues, M., Wang, W., Kim, J.W., Ye, Q., Jia, H., Budhu, A., Zanetti, K.A., Chen, Y., Qin, L.X., et al. (2008). EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res. 68, 1451–1461.

22. Majumder, P.K., Febbo, P.G., Bikoff, R., Berger, R., Xue, Q., McMahon, L.M., Manola, J., Brugarolas, J., McDonnell, T.J., Golub, T.R., et al. (2004). mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. Nat. Med. 10, 594–601.

23. Tovar, V., Cornella, H., Moeini, A., Vidal, S., Hoshida, Y., Sia, D., Peix, J., Cabellos, L., Alsinet, C., Torrecilla, S., et al. (2017). Tumour initiating cells and IGF/FGF signalling contribute to sorafenib resistance in hepatocellular carcinoma. Gut 66, 530–540.

24. Lachenmayer, A., Alsinet, C., Savic, R., Cabellos, L., Toffanin, S., Hoshida, Y., Villanueva, A., Minguez, B., Newell, P., Tsai, H.W., et al. (2012). Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. Clin. Cancer Res. 18, 4997–5007.

25. Villanueva, A., Alsinet, C., Yanger, K., Hoshida, Y., Zong, Y., Toffanin, S., Rodriguez-Carunchio, L., Sole, M., Thung, S., Stanger, B.Z., et al. (2012). Notch signaling is activated in human hepatocellular carcinoma and induces tumor formation in mice. Gastroenterology 143, 1660–1669.

26. Toffanin, S., Hoshida, Y., Lachenmayer, A., Villanueva, A., Cabellos, L., Minguez, B., Savic, R., Ward, S.C., Thung, S., Chiang, D.Y., et al. (2011). MicroRNA-based classification of hepatocellular carcinoma and oncogenic role of miR-517a. Gastroenterology 140, 1618–1628.

27. Lee, J.S., Chu, I.S., Heo, J., Calvisi, D.F., Sun, Z., Roskams, T., Durnez, A., Demetris, A.J., and Thorgeirsson, S.S. (2004). Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 40, 667–676.

28. Naugler, W.E., Sakurai, T., Kim, S., Maeda, S., Kim, K., Elsharkawy, A.M., and Karin, M. (2007). Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science 317, 121–124.

29. Moeini, A., Torrecilla, S., Tovar, V., Montironi, C., Andreu-Oller, C., Peix, J., Higuera, M., Pfister, D., Ramadori, P., Pinyol, R., et al. (2019). An Immune Gene Expression Signature Associated With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. Gastroenterology 157, 1383–1397.

30. Kim, H.S., Rotundo, L., Yang, J.D., Kim, D., Kothari, N., Feurdean, M., Ruhl, C., and Unalp-Arida, A. (2017). Racial/ethnic disparities in the prevalence and awareness of Hepatitis B virus infection and immunity in the United States. J. Viral Hepat. 24, 1052–1066.

31. Rosenberg, E.S., Hall, E.W., Sullivan, P.S., Sanchez, T.H., Workowski, K.A., Ward, J.W., and Holtzman, D. (2017). Estimation of State-Level Prevalence of Hepatitis C Virus Infection, US States and District of Columbia, 2010. Clin. Infect. Dis. 64, 1573–1581.

32. Wong, T., Dang, K., Ladhani, S., Singal, A.K., and Wong, R.J. (2019). Prevalence of Alcoholic Fatty Liver Disease Among Adults in the United States, 2001-2016. JAMA 321, 1723–1725.

33. Kasai, Y., Takeda, S., and Takagi, H. (1996). Pathogenesis of hepatocellular carcinoma: a review from the viewpoint of molecular analysis. Semin. Surg. Oncol. 12, 155–159.

34. Yang, Y.M., Kim, S.Y., and Seki, E. (2019). Inflammation and Liver Cancer: Molecular Mechanisms and Therapeutic Targets. Semin. Liver Dis. 39, 26–42.

35. Trobridge, G., Hirata, R.K., and Russell, D.W. (2005). Gene targeting by adeno-associated virus vectors is cell-cycle dependent. Hum. Gene Ther. 16, 522–526.

36. Vansaun, M.N., Mendonsa, A.M., and Lee Gorden, D. (2013). Hepatocellular proliferation correlates with inflammatory cell and cytokine changes in a murine model of nonalchoholic fatty liver disease. PLoS ONE 8, e73054.

37. Fattovich, G., Stroffolini, T., Zagni, I., and Donato, F. (2004). Hepatocellular carcinoma in cirrhosis: incidence and risk factors. Gastroenterology 127 (5, Suppl 1), S35–S50.

38. Ma, C., Kesarwala, A.H., Eggert, T., Medina-Echeverz, J., Kleiner, D.E., Jin, P., Stroncek, D.F., Terabe, M., Kapoor, V., ElGindi, M., et al. (2016). NAFLD causes selective CD4(+) T lymphocyte loss and promotes hepatocarcinogenesis. Nature 531, 253–257.

39. Miller, D.G., Trobridge, G.D., Petek, L.M., Jacobs, M.A., Kaul, R., and Russell, D.W. (2005). Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J. Virol. 79, 11434–11442.

40. Nakai, H., Wu, X., Fuess, S., Storm, T.A., Munroe, D., Montini, E., Burgess, S.M., Grompe, M., and Kay, M.A. (2005). Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. J. Virol. 79, 3606–3614.

41. Chandler, R.J., Sands, M.S., and Venditti, C.P. (2017). Recombinant Adeno-Associated Viral Integration and Genotoxicity: Insights from Animal Models. Hum. Gene Ther. 28, 314–322.

42. La Bella, T., Imbeaud, S., Peneau, C., Mami, I., Datta, S., Bayard, Q., Caruso, S., Hirsch, T.Z., Calderaro, J., Morcrette, G., et al. (2020). Adeno-associated virus in the liver: natural history and consequences in tumour development. Gut 69, 737–747.

43. Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol. 82, 5887–5911.

44. Powers, J.M., Chang, X.L., Song, Z., and Nakai, H. (2018). A Quantitative Dot Blot Assay for AAV Titration and Its Use for Functional Assessment of the Adeno-associated Virus Assembly-activating Proteins. J. Vis. Exp. 12, 56766.

Study 3 – Immunomodulatory Effects of Lenvatinib Plus Anti-Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma

Laura Torrens, Carla Montironi, Marc Puigvehí, Agavni Mesropian, Jack Leslie, Philipp K Haber, Miho Maeda, Ugne Balaseviciute, Catherine E Willoughby, Jordi Abril-Fornaguera, Marta Piqué-Gili, Miguel Torres-Martín, Judit Peix, Daniel Geh, Erik Ramon-Gil, Behnam Saberi, Scott L Friedman, Derek A Mann, Daniela Sia, Josep M Llovet

Summary

Approximately, 50–60% of patients with HCC are estimated to be exposed to systemic therapies in their lifespan, particularly in advanced stages of the disease[78]. Recently, the tyrosine kinase inhibitor **lenvatinib** showed efficacy in an open-label randomized phase III trial and became the first new **FDA-approved drug for advanced-stage HCC** in the first-line setting in over 10 years[138]. On the other hand, immunotherapies with **immune checkpoint inhibitors** (ICIs) are emerging as promising treatment options, but responses in HCC are only observed in 15-20% of patients, while the majority are **primarily resistant**.

Much effort has been invested into identifying existing kinase inhibitors that can effectively synergize with ICI. The combination of **lenvatinib plus** the anti-PD1 ICI **pembrolizumab** has shown unprecedented phase Ib results (objective response rate of 46% and median survival of 22 months)[148] and is currently being assessed in a phase III trial. This strategy is based on the hypothesis that lenvatinib could boost the antitumor immune response and improve the clinical benefit of anti-PD1. However, a deeper understanding of the immunomodulatory capacity of these treatments is still needed. Considering this, **we aimed to identify the immunomodulatory effects of lenvatinib in combination with anti-PD1 and provide a mechanistic rationale for this treatment in advanced HCC.**

To fill this gap, we generated three murine immunocompetent **syngeneic models** of HCC – two subcutaneous and an orthotopic – and assessed the anti-tumoral activity of lenvatinib alone or in combination with anti-PD1 and its effects on the **systemic and tumor-infiltrating immune cells**. By performing flow cytometry, transcriptomic, and immunohistochemistry analyses, we explored the immunomodulatory effect of each treatment and unveiled unique effects of the combination. Next, we explored the gene expression analysis of murine and human tumors to

identify **HCC patients** who are likely to exhibit **primary resistance** to single agents but could potentially be **rescued** with the combination treatment.

Our results showed that:

1. Anti-PD1 and lenvatinib in monotherapy were able to improve survival and reduce tumor growth in our murine models. However, the **combination** treatment achieved a higher **response rate**, shorter **time to response**, and a reduction of **tumor viability**.

2. **Lenvatinib** exerted a potent immunomodulatory effect on the tumor infiltrate by reducing the T_{reg} proportion and altering their histological distribution by eliciting T_{reg} exclusion from the intratumoral region. In addition, lenvatinib blocked immunosuppressive signaling including TGFß pathway inhibition.

3. **Anti-PD1** treatment was also able to modify the immune infiltrate by increasing the **T cells** and **type 1 dendritic cells** in the tumor. However, transcriptomic profiling of the tumors revealed that the treatment elicited an immune exhausted phenotype in the infiltrate.

4. The **combined effect** of lenvatinib plus anti-PD1 treatments induced an increase in the **proinflammatory component** in the tumor. Consequently, only the combination treatment generated a specific activated antitumoral immune response.

5. No significant alterations were detected in **circulating immune cells**, suggesting that the effect of the treatments may be tumor-specific.

6. Transcriptomic analysis in a human HCC cohort revealed that **22% of tumors** presented down-regulation of genes associated with the molecular effect of the combination, along with reduced pro-inflammatory signaling, high T_{reg} levels, and VEGF pathway activation. Thus, these patients could harbor **primary resistance to anti-PD1** and potentially benefit from the **booster effect** of the combination treatment.

In conclusion, these findings provide a comprehensive understanding of the immune-remodeling capacities of the combination of lenvatinib plus anti-PD1, with important implications for patient selection to ultimately maximize the clinical benefit from this treatment regimen.

HEPATOLOGY

Immunomodulatory Effects of Lenvatinib Plus Anti–Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma

Laura Torrens [ID],[1,2] Carla Montironi [ID],[1,2] Marc Puigvehí [ID],[1,3] Agavni Mesropian [ID],[2] Jack Leslie [ID],[4] Philipp K. Haber,[1] Miho Maeda,[1] Ugne Balaseviciute [ID],[2] Catherine E. Willoughby [ID],[2] Jordi Abril-Fornaguera [ID],[2] Marta Piqué-Gili [ID],[2] Miguel Torres-Martín [ID],[1,2] Judit Peix [ID],[2] Daniel Geh [ID],[4] Erik Ramon-Gil [ID],[4] Behnam Saberi,[1] Scott L. Friedman [ID],[1] Derek A. Mann [ID],[4] Daniela Sia [ID],[1] and Josep M. Llovet [ID] [1,2,5]

BACKGROUND AND AIMS: Lenvatinib is an effective drug in advanced HCC. Its combination with the anti-PD1 (programmed cell death protein 1) immune checkpoint inhibitor, pembrolizumab, has generated encouraging results in phase Ib and is currently being tested in phase III trials. Here, we aimed to explore the molecular and immunomodulatory effects of lenvatinib alone or in combination with anti-PD1.

APPROACH AND RESULTS: We generated three syngeneic models of HCC in C57BL/6J mice (subcutaneous and orthotopic) and randomized animals to receive placebo, lenvatinib, anti-PD1, or combination treatment. Flow cytometry, transcriptomic, and immunohistochemistry analyses were performed in tumor and blood samples. A gene signature, capturing molecular features associated with the combination therapy, was used to identify a subset of candidates in a cohort of 228 HCC patients who might respond beyond what is expected for monotherapies. In mice, the combination treatment resulted in tumor regression and shorter time to response compared to monotherapies ($P < 0.001$). Single-agent anti-PD1 induced dendritic and T-cell infiltrates, and lenvatinib reduced the regulatory T cell (Treg) proportion. However, only the combination treatment significantly inhibited immune suppressive signaling, which was associated with the TGFß pathway and induced an immune-active microenvironment ($P < 0.05$ vs. other therapies). Based on immune-related genomic profiles in human HCC, 22% of patients were identified as potential responders beyond single-agent therapies, with tumors characterized by Treg cell infiltrates, low inflammatory signaling, and VEGFR pathway activation.

CONCLUSIONS: Lenvatinib plus anti-PD1 exerted unique immunomodulatory effects through activation of immune pathways, reduction of Treg cell infiltrate, and inhibition of TGFß signaling. A gene signature enabled the identification of ~20% of human HCCs that, although nonresponding to single agents, could benefit from the proposed combination. (Hepatology 2021;74:2652-2669).

Liver cancer is the second-leading cause of cancer-related death and a major health problem globally.[1] HCC is the most common form of liver cancer, accounting for >90% of

cases.[2] Around 40% of HCC patients are diagnosed at advanced stages of the disease, in which the tyrosine kinase inhibitor (TKI) sorafenib, has been the only approved treatment for >10 years.[3] Only recently, lenvatinib has shown noninferiority compared to sorafenib and received U.S. Food and Drug Administration (FDA) approval for the treatment of advanced HCC.[4] On the other hand, immunotherapies with immune checkpoint inhibitors (ICIs) are emerging as promising treatment options. For example, the new combination of the ICI, atezolizumab, plus bevacizumab (VEGFA inhibitor) has been FDA approved as first-line therapy[5] and pembrolizumab and nivolumab plus ipilimumab as second-line treatments.[6-8]

Although ICIs are changing the landscape of cancer medicine, responses in HCC are only observed in 15%-20% of patients, whereas the majority are primarily resistant. Thus, much effort has been invested into identifying existing kinase inhibitors that can effectively synergize with ICIs. In this context, lenvatinib plus the anti-PD1 ICI, pembrolizumab, is currently being tested in unresectable HCC in phase Ib and III trials, with an encouraging objective response rate (ORR) of 46%, 22-month median survival, and 9.5-month median progression-free survival.[9] This strategy is based on the hypothesis that lenvatinib could inhibit the immunosuppressive and proangiogenic effect of the VEGFA-VEGFR pathway on the tumor microenvironment,[10,11] thus boosting antitumor immune response and improving the clinical benefit of anti-PD1. In this regard, experimental studies conducted in HCC models have recently suggested a link between lenvatinib and inflammation as well as a greater antitumoral effect of the combination treatment.[12,13] However, a deeper understanding of the immunomodulatory capacity of these treatments is still needed.

To fill this gap, we generated three murine immunocompetent models of HCC—two subcutaneous and an orthotopic—and assessed the antitumoral activity of lenvatinib alone or in combination with anti-PD1 and its effects on systemic and tumor-infiltrating immune cells. By performing flow cytometry, transcriptomic, and immunohistochemistry (IHC) analyses, we explored the immunomodulatory

ARTICLE INFORMATION:

From the [1]Mount Sinai Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY; [2]Translational Research in Hepatic Oncology, Liver Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)-Hospital Clínic, Universitat De Barcelona, Barcelona, Spain; [3]Hepatology Section, Gastroenterology Department, Parc de Salut Mar, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain; [4]Newcastle Fibrosis Research Group, Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom; [5]Institució Catalana De Recerca i Estudis Avançats, Barcelona, Spain.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Josep M. Llovet, M.D.
Liver Cancer Translational Research Laboratory, Liver Unit
IDIBAPS-Hospital Clinic, Faculty of Medicine
University of Barcelona

Rosselló 153
08036, Barcelona, Catalonia, Spain
E-mail: jmllovet@clinic.cat
Tel.: 0034-932-279-155

2653

effect of each treatment and unveiled noteworthy effects of the combination. Gene expression analysis of murine and human tumors allowed the identification of a subset of HCC patients (~22%) who are likely to exhibit primary resistance to single agents, but could potentially be rescued with the combination treatment.

Overall, these findings provide a comprehensive understanding of the immune-remodeling capacities of the combination of lenvatinib plus anti-PD1, with important implications for patient selection to ultimately maximize the clinical benefit from this treatment regimen.

Materials and Methods

SUBCUTANEOUS SYNGENEIC MOUSE MODELS

Two subcutaneous syngeneic HCC models were generated by injecting 5×10^6 Hepa1-6 cells (ATCC, Manassas, VA) in 100 μL of PBS in 5- to 6-week-old female C57BL/6J mice (n = 59; Charles River Laboratories, Wilmington, MA) and 5×10^6 Hep53.4 cells (CLS, Eppelheim, Germany) in 5- to 6-week-old male C57BL/6J mice (n = 40).[14] Animals were weighed, and tumor volume was assessed three times per week. Once tumors reached 200 mm³, animals were randomly assigned to receive lenvatinib (Eisai, Ibaraki, Japan), anti-PD1 (anti-murine PD-1 monoclonal antibody clone J43 BioXCell, San Diego, CA, BE0033-2), combination therapy (lenvatinib plus anti-PD1), or placebo (drug vehicle plus polyclonal IgG, BioXCell BE0091; Fig. 1A).

Mice from the Hepa1-6 model were euthanized at day 13 postrandomization (early time point, n = 20), once a tumor volume of 1,000 mm³ was reached or at study termination (late time point, n = 39). The Hep53.4 model was used as validation, and all animals were euthanized at day 13 postrandomization. Tumor and blood samples were collected and processed for subsequent analyses (Figs. 1A and 2A). Assessment of the tumorigenic and histological features in an additional set of male and female mice (n = 30) revealed no sex differences in our model (Supporting Fig. S1). Studies were performed in compliance with guidelines for the use of animals established by the institution ethical committee and the Guide for the Care and Use of Laboratory Animals.

ORTHOTOPIC SYNGENEIC MOUSE MODEL

To generate the orthotopic model, 5×10^6 luciferase-transfected Hep53.4 cells were implanted in liver of 5- to 6-week-old male C57BL/6J mice. Tumor growth was serially assessed using the In Vivo Imaging System once per week. Only mice with bioluminescence values $>10^7$ counts before randomization were included in the study (n = 33). Animals were randomized 2 weeks after implantation to receive lenvatinib, anti-PD1, combination therapy, or placebo. Mice were euthanized at day 13 postrandomization (Fig. 2A), and tumor burden was assessed *ex vivo* in liver samples.

MULTICOLOR FLOW CYTOMETRY ANALYSIS

Flow cytometry analysis was performed on tumor and blood samples from the Hepa1-6 model collected at the early time point. After sample processing, ~1×10^6 freshly prepared cells were stained with fluorochrome-coupled antibodies targeting cell markers. Three antibody panels were designed to detect lymphocyte and myeloid cell populations of interest (Supporting Table S1; Supporting Fig. S2). Cells were stained according to standard flow cytometer protocols (Supporting Tables S2 and S3). Fluorescence data from 50,000 events per sample were collected on a LSRII cytometer (BD Biosciences, Franklin Lakes, NJ), available at Icahn School of Medicine at Mount Sinai (New York, NY) facilities and analyzed using FlowJo Flow Cytometry analysis software.

HISTOLOGICAL AND IHC ANALYSIS OF TUMOR SAMPLES

Tumor samples were fixed in buffered 4% paraformaldehyde for 24 hours and underwent tissue processing and embedding in paraffin to create formalin-fixed, paraffin-embedded blocks. In the Hepa1-6 model, only tumors with sufficient tumoral material were processed for histological analysis (n = 14 and 17 in the early and late time points, respectively). Twenty-four tumors from the Hep53.4 subcutaneous model were used for validation.

Tumor viability, defined as the proportion of tumor-presenting viable cells in a sample (i.e., excluding necrotic regions or granulation tissue), was assessed on

TORRENS ET AL.

A

Hepa1-6 subcutaneous model (n = 59)

B Tumor growth

C Survival (time to 1000 mm³)

D Time to objective response

E Response to treatment (early timepoint)

Tumor response rate			
	PD	SD	OR
Placebo	11	4	0
Lenvatinib	3	9	3
Anti-PD1	1	4	10
Combination	0	0	14
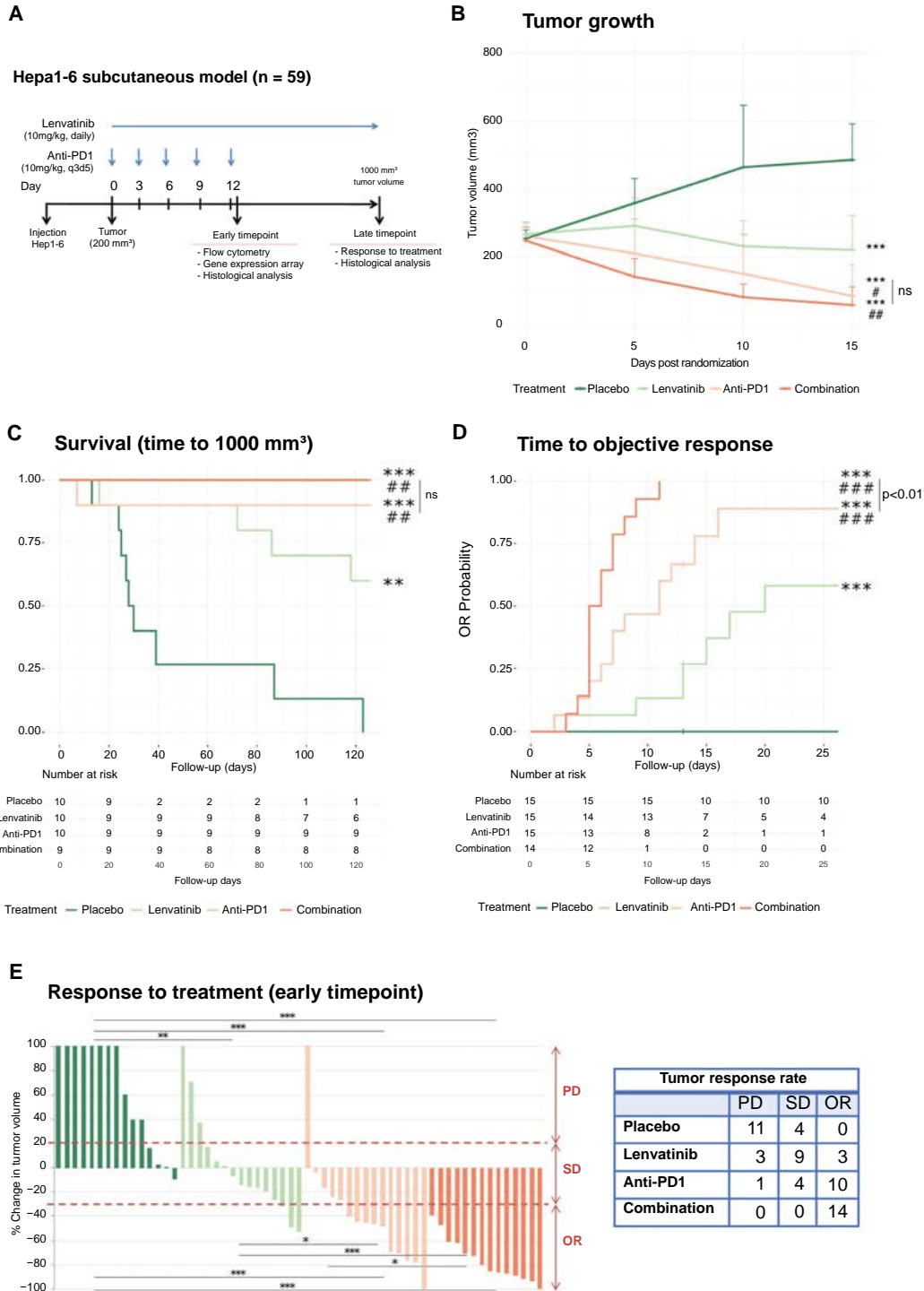
FIG. 1. Antitumoral effect of lenvatinib plus anti-PD1 in the Hepa1-6 model. (A) Timeline of the study. (B) Tumor growth, (C) survival, and (D) time to objective response of treated mice. (E) Response to treatment at early time point (n = 59). Upper part indicates differences in progressive disease rate, and lower part the differences in objective response rate. Right table shows the number of mice per group and response. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ versus placebo (unless indicated); #$P < 0.05$; ##$P < 0.01$; ###$P < 0.001$ versus lenvatinib. Abbreviations: Cum, cumulative; OR, objective response; PD, progressive disease; SD, stable disease.

hematoxylin and eosin slides. Further histopathological examination was performed by IHC (Supporting Table S3). All analyses were performed by an expert pathologist blinded to the treatment arms.

TRANSCRIPTOMIC ANALYSIS OF TUMOR SAMPLES

Tumor samples from the Hepa1-6 model collected at the early time point or from mice reaching the survival endpoint at day 13 postrandomization (n = 21) were processed for transcriptome analysis. Gene expression microarray studies were conducted using the Clariom S Mouse Array (GSE153203; Affymetrix, Santa Clara, CA).

The combination rescue signature was generated by selecting the top differentially expressed genes between tumors from the combination and placebo arms (Bonferroni, $P < 0.05$; fold change [FC], >3 or <0.33). Genes significantly enriched in the monotherapy groups compared to placebo according to the same criteria were eliminated from the signature to capture only the combination-specific transcriptomic effect.

IDENTIFICATION OF POTENTIAL RESPONDERS TO COMBINATION TREATMENT

We analyzed gene expression data from a cohort of 228 surgically resected fresh-frozen HCC samples (Heptromic data set, GSE63898) previously collected in the setting of the HCC Genomic Consortium.[15,16] A gene signature was generated to identify candidates that might respond to the combination beyond what is expected for single treatment effect alone.

STATISTICAL ANALYSES

Statistical analysis was conducted using R (version 3.6.2; R Foundation for Statistical Computing, Vienna, Austria) or GraphPad Prism software (version 5.01; GraphPad Software Inc., San Diego, CA). Comparison of continuous variables was performed using Kruskal-Wallis and Dunn's tests for nonparametric distributions or ANOVA and Tukey tests for parametric distributions. Correlations for categorical variables were analyzed by Fisher's exact test. Survival and time to response were assessed with Kaplan-Meier estimates and the log-rank test.

Additional detailed information is provided in the Supporting Materials and Methods.

Results

ANTITUMOR ACTIVITY OF LENVATINIB AND ANTI-PD1 COMBINATION IN A SYNGENEIC MURINE MODEL

To evaluate the antitumor activity of lenvatinib, anti-PD1 immunotherapy, and its combination, we generated three syngeneic murine HCC models[14] (Figs. 1A and 2A). Median tumor volume at randomization was equal in all treatment arms, and no significant differences in body weight or other toxicity signs were observed, indicating that all treatments were well tolerated (Supporting Fig. S3A,B). In the Hepa1-6 model, mice treated with lenvatinib, anti-PD1, or its combination exhibited a significant reduction in tumor growth and improved survival compared to the placebo arm (median survival of 29 days in placebo and not reached [NR] in the remaining groups; $P < 0.0001$; Fig. 1B,C). Both anti-PD1 and combination treatments showed higher antitumor efficacy than lenvatinib, but no significant differences were observed between the anti-PD1 and combination arms. However, mice receiving combination treatment required a shorter time to objective responses compared to anti-PD1, lenvatinib, and placebo (median 5, 11, and 20 days and NR, respectively; $P < 0.0001$; Fig. 1D).

Notably, the combination treatment achieved a significantly higher ORR compared to both lenvatinib and anti-PD1 at the early time point ($P < 0.05$; Fig. 1E). At the late time point, the best ORR was also achieved by the combination treatment ($P < 0.001$ vs.

A

B Tumor growth

C Time to objective response

D Viable tumor

E Tumor growth

F Tumor burden
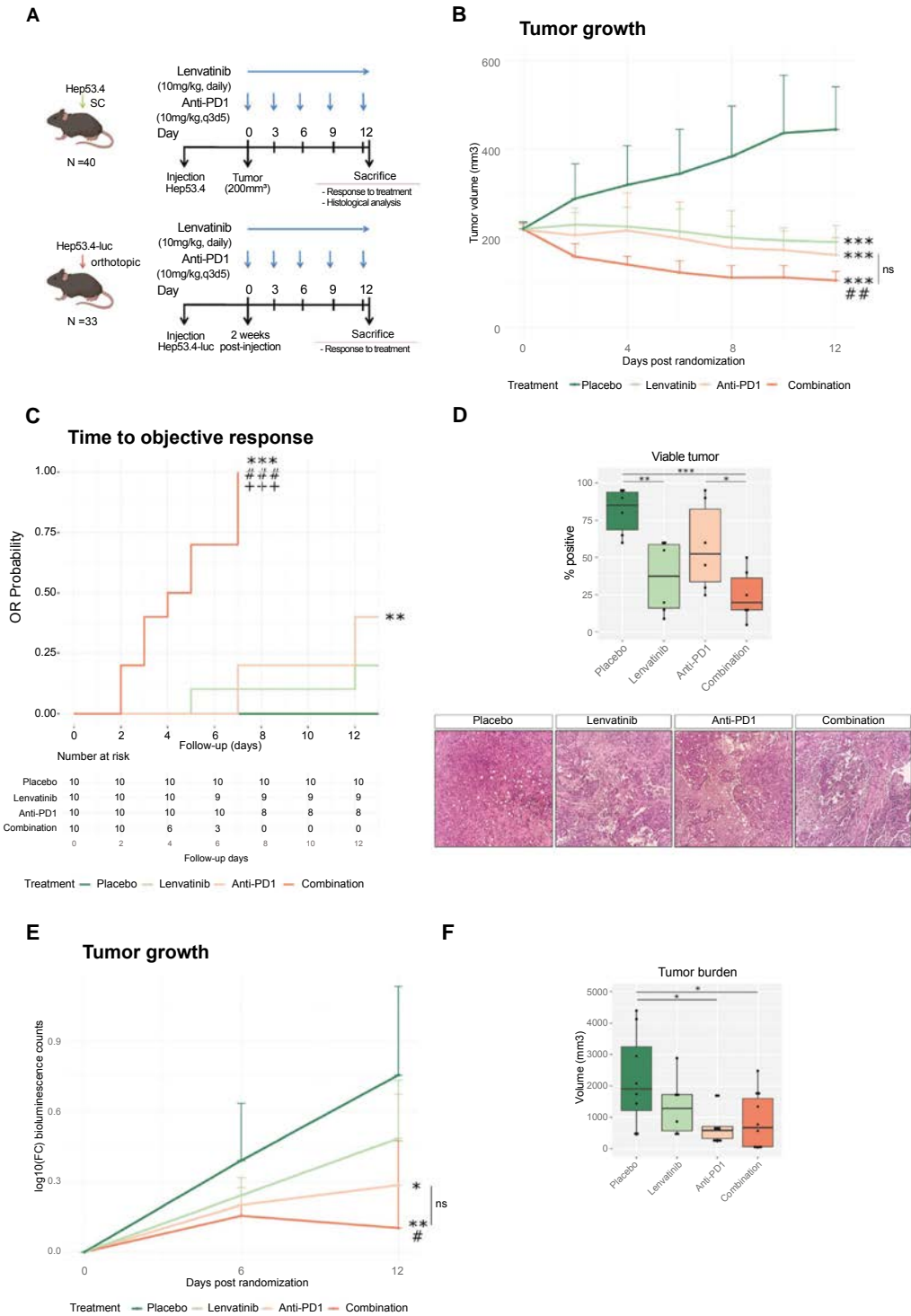
FIG. 2. Antitumoral effect of lenvatinib plus anti-PD1 in the subcutaneous and orthotopic Hep53.4 models. (A) Timeline of the subcutaneous and orthotopic studies. (B) Tumor growth and (C) time to objective response of treated mice from the subcutaneous model. (D) Tumor viability assessed in H&E slides from the subcutaneous model. Representative images captured at 20×. (E) Tumor growth in the orthotopic model, measured as changes in bioluminescence compared to the start of the treatment. (F) Tumor volume measured *ex vivo* in liver samples. Box plots indicate median and quartiles. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ versus placebo (unless indicated); #$P < 0.05$; ##$P < 0.01$; ###$P < 0.001$ versus lenvatinib; +++$P < 0.001$ versus anti-PD1. Abbreviations: FC, fold change; OR, objective response; SC, subcutaneous.

placebo; Supporting Fig. S3C). All treatments induced a reduction in the progressive disease rate compared to placebo at both time points. The subcutaneous Hep53.4 model confirmed the significant enhanced antitumoral effect, fastest response, and higher ORR in the combination treatment arm (Fig. 2B,C and Supporting Fig. S4A). Similarly, in the orthotopic Hep53.4 model, the combination treatment achieved the greatest reduction in tumor growth, although differences compared to anti-PD1 were nonsignificant (mean FC, 2.7 and 1.6 in anti-PD1 and combination, respectively; $P = 0.156$; Fig, 2E,F and Supporting Fig. S4B).

Finally, tumor viability was assessed to further understand the antitumoral effect of the treatments. Only the combination treatment significantly reduced tumor viability at both time points of the Hepa1-6 model ($P < 0.01$ vs. placebo; Supporting Fig. S3D-F) and in the subcutaneous Hep53.4 model ($P < 0.001$; Fig. 2D). Tumor viability was also significantly lower than in the anti-PD1 arm at the Hep53.4 model and Hepa1-6 late endpoint (all $P < 0.05$). The combination treatment reduced cell proliferation compared to placebo and anti-PD1 ($P < 0.05$). Therefore, although no differences in tumor growth were detected between the two treatments, tumors from the combination arm were less viable and contained higher proportions of necrotic and granulation tissue.

Overall, all treatments presented antitumoral activity, but the combination of lenvatinib plus anti-PD1 reduced the time to treatment response and tumor viability, and improved ORR compared to the monotherapies.

MECHANISM OF ACTION OF LENVATINIB AND ANTI-PD1 COMBINATION

The Combination Treatment Elicits an Immune-Activating Lymphocytic Infiltrate

The presence of the main lymphocytic populations in tumor samples from each treatment arm in the Hepa1-6 model was assessed by flow cytometry. Tumors from animals treated with anti-PD1 and combination therapies contained a significantly increased intratumoral T-cell infiltrate and a reduction in the PD1+CD8+ T-cell proportion, indicating that the anti-PD1 treatment reached the tumor. Both lenvatinib and combination treatments reduced the intratumoral regulatory T cell (Treg) infiltrate ($P < 0.05$ vs. placebo; Fig. 3A; Supporting Table S4). Only the combination treatment was able to enhance the CD8+ T-cell to Treg ratio and proportion of proliferating CD8 T cells, ultimately indicating an increase in the tumoral proinflammatory immune cell component. No differences in the CD4+ T-cell and CD8+ T-cell or B-cell populations were detected between treatments. Finally, the analysis of the immune cell populations in blood samples did not reveal differences in systemic immune response among treatment arms (Supporting Fig. S5A; Supporting Table S4).

To study the mid- and long-term effects of the treatments, CD8, CD4 T cells, and Treg cells were analyzed by immunostaining in tumor samples collected at the early and late time points. Interestingly, CD4 staining was exclusively located in the intratumoral area in the combination arm compared to a mainly peritumoral location in the placebo and lenvatinib arms ($P < 0.05$; Supporting Fig. S6), which could potentially induce a more effective antitumoral immune response.[17] CD8 lymphocytes were mostly intratumoral, with no significant differences between treatment arms. Next, forkhead box protein 3 (FOXP3) staining indicated a significant Treg decrease in the combination group at the late time point and Treg exclusion from the intratumoral region in the combination arm compared to placebo ($P < 0.05$; Supporting Fig. S7A-C). Together with the flow cytometry data, this indicates that lenvatinib and combination treatments might alter the proportion of Treg cells in the lymphocytic infiltrate and its location. No clear association between Treg cell localization and tumor vasculature was observed (Supporting Fig. S7D,E).
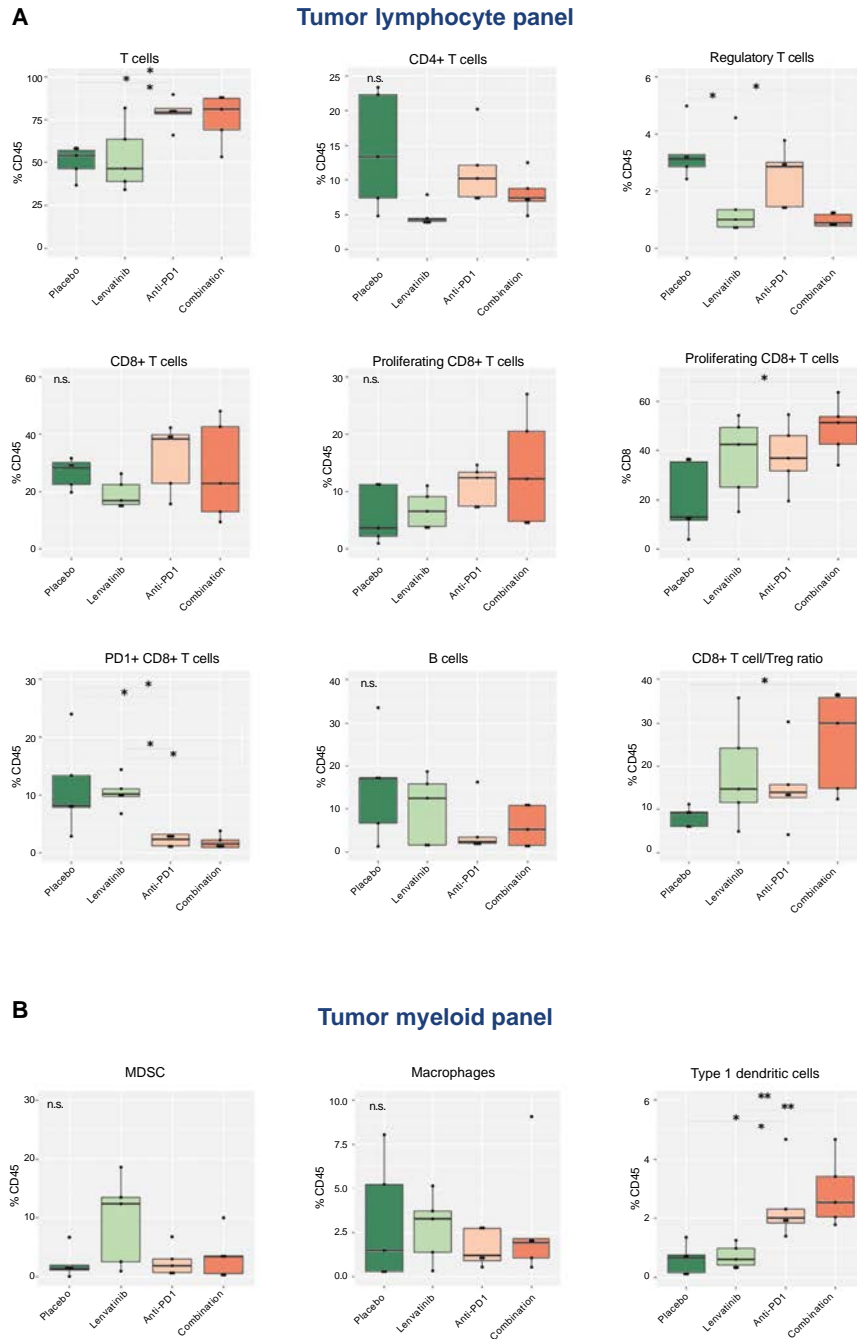
FIG. 3. Immune cell populations in tumor samples detected by flow cytometry analysis. (A) Lymphoid and (B) myeloid immune cell populations from tumor samples collected at the early time point. Results for each treatment arm are shown (n = 5 samples per arm). Box plots indicate median and quartiles. *P < 0.05; **P < 0.01. Abbreviations: MDSC, myeloid-derived suppressor cells; Treg, regulatory T cell.
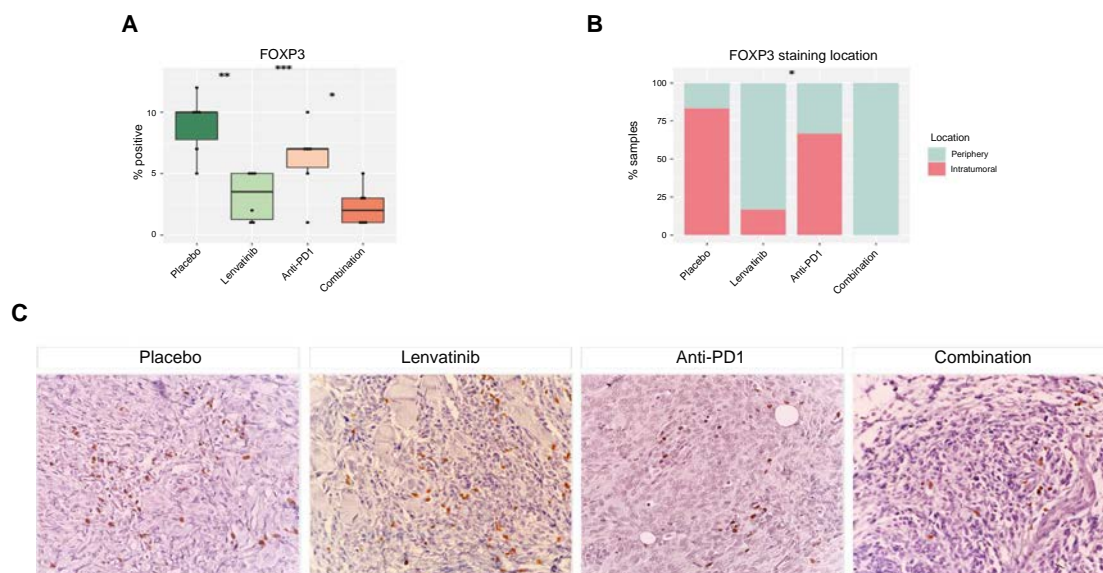
FIG. 4. Histological analysis of Treg tumor infiltrate in the Hep53.4 subcutaneous model. (A) Percentage of positive cells for FOXP3 staining in tumor samples from treated animals. (B) Percentage of samples with intratumoral or peripheral FOXP3 staining. (C) Representative images of FOXP3 staining captured with 40× magnification. Box plots indicate median and quartiles. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

This histological profile was validated in the Hep53.4 subcutaneous model, which confirmed a reduction in FOXP3 staining in both lenvatinib and combination arms compared to placebo ($P < 0.01$) and shift toward a peritumoral localization in the combination group ($P < 0.05$ vs. placebo; Fig. 4). Despite that no significant differences in CD3, CD8, and CD4 staining were detected, there was an increase in CD8 intratumoral location in tumors from the anti-PD1 and combination arms (Supporting Fig. S6D-F). This immunological profile aligns with a more active intratumoral infiltrate in the combination group.

Taken together, the treatments were able to modify the lymphoid infiltrate in the tumor. Interestingly, anti-PD1 increased the T-cell infiltrate, and lenvatinib reduced the proportion and intratumoral location of Treg cells, associated with immune suppression.[18] Consequently, the combination treatment achieved a greater tumoral proinflammatory immune cell component.

Impact of Treatments on Intratumoral and Systemic Myeloid Populations

The analysis of the intratumoral myeloid populations revealed that anti-PD1 and combination treatments increased the type 1 dendritic cell (DC1) infiltrate compared to placebo ($P < 0.01$; Fig. 3B; Supporting Table S4). No significant differences were observed in the proportion of infiltrating macrophages and myeloid-derived suppressor cells (MDSCs). However, further IHC analysis revealed that anti-PD1 increased the percentage of M2 macrophages at the early time point ($P < 0.05$ vs. placebo), whereas the combination treatment did not alter this immunosuppressive population compared to placebo (Supporting Fig. S8A). This effect was not observed at the late time point. The circulating myeloid populations were not significantly modified by the treatments (Supporting Fig. S5B).

Overall, we observed that anti-PD1 and combination treatments increased tumor infiltrating DC1, the main stimulator of T-cell function.[19] The immune-suppressive

2660

M2 macrophage component was increased by anti-PD1 treatment only at the early time point, indicating that this effect could be lost once treatment is suspended.

Combination Treatment Significantly Alters the Tumor Inflammatory and Proliferative Gene Expression Profile

We next sought to investigate the molecular profile of Hepa1-6 tumors from each treatment arm. Principal component analysis showed that murine tumors clustered together with HCC human tumors from a cohort of 228 surgically resected fresh-frozen HCCs (Heptromic data set),[16] indicating that the syngeneic model was able to recapitulate the transcriptomic characteristics of human HCC (Supporting Fig. S9A). In mice treated with the combination, there were 1,265, 1,621, and 30 significantly differentially expressed genes compared to placebo, lenvatinib, and anti-PD1, respectively (Supporting Fig. S9B; Supporting Table S5). Notably, 68.8% (870 of 1,265) of the differentially expressed genes in the combination arm compared to placebo were not altered by the monotherapies. This noteworthy molecular effect of the combination treatment could be attributable to a synergistic effect at the transcriptional level. Pathway analysis revealed that the genes solely deregulated by the combination treatment were associated with activation of proinflammatory pathways (i.e., T-cell, B-cell, and chemokine signaling; Supporting Table S6).

By comparing the expression profile of each treatment arm, we observed that tumors from the combination arm had fewer differentially expressed genes associated with proliferation and cell-cycle progression ($P < 0.05$ vs. placebo; Fig. 5A). In addition, in both anti-PD1 and combination treatments, there was enhanced inflammatory signaling (e.g., T-cell receptor and chemokine signaling, inflammatory response, and dendritic cell [DC] chemotaxis). Lenvatinib also showed a more modest but significant enrichment of proinflammatory gene sets ($P < 0.05$ vs. placebo). Of note, lenvatinib was able to inhibit signaling pathways downstream of its main targets, including VEGFR, RET proto-oncogene (RET), and fibroblast growth factor receptor (FGFR) 2 (Supporting Fig. S9C).

Combination Therapy Elicits Immune Activation and Down-Regulation of TGFß Signaling

We then assessed the immunological profile of our samples using the recently described HCC immune class.[15,20] Interestingly, only the tumors from the combination arm were enriched in the immune-active class (4 of 5 vs. 0 of 5 in anti-PD1; $P < 0.05$), associated with an adaptive T-cell response activation, whereas all anti-PD1 tumors showed an immune-exhausted profile (5 of 5 vs. 1 of 5 in combination; $P < 0.05$), which is characterized by activation of immunosuppressive signaling hampering the antitumoral immune response (Fig. 5A). Samples from the placebo and lenvatinib groups were nonimmune, consistent with the lower T-cell infiltration observed by flow cytometry. Subclass mapping analysis confirmed that tumors from the combination group showed genetic similarity to human HCCs belonging only to the immune-active class, whereas anti-PD1-treated tumors showed a similarity with exhausted tumors as well (Fig. 5B).

Gene expression data were then used to characterize the impact of the treatments on the tumor composition. Anti-PD1 and combination treatments reduced the malignant component of the tumors and increased the immune infiltrate (Fig. 6A,B), whereas tumors from the placebo and lenvatinib arms had higher estimated tumor purity. In line with flow cytometry and IHC results, tumors from the anti-PD1 and combination arms contained an enrichment of gene sets associated with T-cell and DC infiltrates (Fig. 6A). The macrophage expression profile revealed an increased M1 proportion in the combination arm whereas anti-PD1 tumors were enriched in M2. The analysis of immunosuppressive signaling in tumors revealed that the combination treatment inhibited transforming growth factor β (TGFß) signaling and reduced the expression of Wnt ligands, likely impairing Wnt/β-catenin signaling in the microenvironment (Fig. 5A). Lenvatinib alone also displayed an immunomodulatory effect by reducing TGFß signaling ($P < 0.05$ vs. placebo). The reduced TGFß signaling in these treatment arms could be linked to the decreased proportion of Tregs.[18] In addition, gene set enrichment analysis (GSEA) confirmed an increase in proinflammatory signaling by
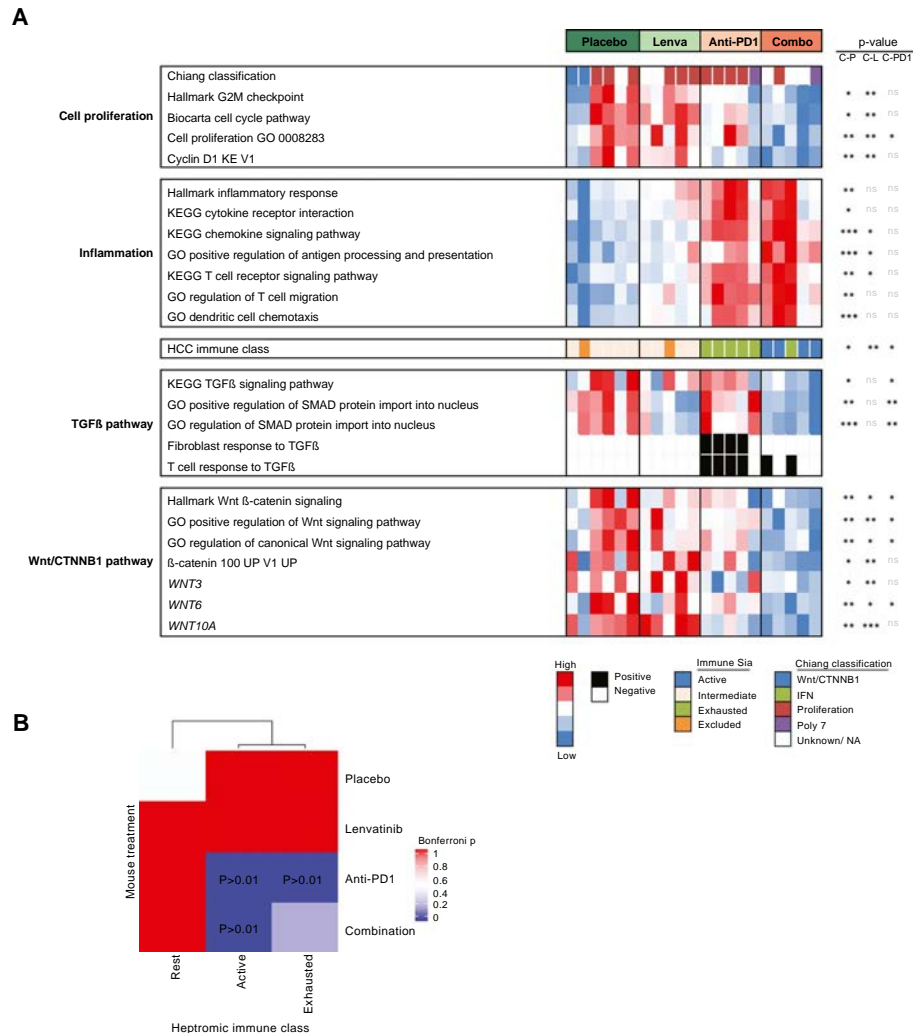
FIG. 5. Gene expression profile of treated tumors. (A) Transcriptomic and immunological profile of tumors from each treatment arm. (B) Subclass mapping analysis showing the transcriptomic similarity between tumors from each treatment arm and human HCC classified according to the HCC immune class. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Abbreviations: C, combination; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; L, lenvatinib; NA, not available; P, placebo; PD1, anti-PD1.

all treatments (false discovery rate [FDR], <0.05 vs. placebo). Compared to anti-PD1, the combination induced down-regulation of gene sets associated with resistance to ICI and TGFß signaling and increased proinflammatory signaling (FDR < 0.05; Supporting Fig. S9D-G; Supporting Table S7).

To investigate the impact of the treatments on the tumoral expression of checkpoint inhibitors, the expression of PD1, programmed death ligand 1 (PDL1), and cytotoxic T-lymphocyte-associated protein 4 (CTLA4) was assessed by IHC (Supporting Fig. S8B-D). CTLA4 positivity was increased in tumors from the anti-PD1 and combination groups at the early time point ($P < 0.05$ in combination vs. placebo; n.s. in anti-PD1 vs. placebo). At the late time point, increased CTLA4 expression was maintained only in the anti-PD1 arm ($P < 0.001$ vs. placebo). No significant differences were observed in PD1 and PDL1 expression between groups.

2662
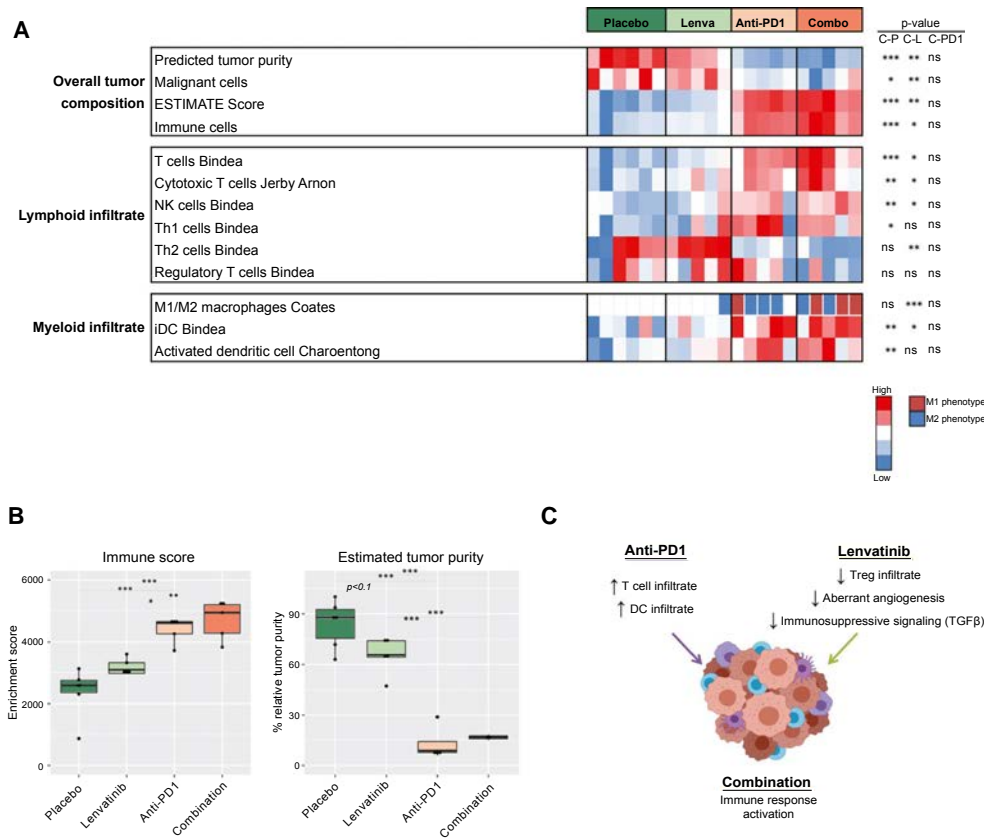
FIG. 6. Characterization of tumor composition based on gene expression data. (A) Tumor composition assessed by ESTIMATE analysis or ssGSEA capturing distinct cell populations. (B) Immune score and estimated tumor purity in tumors from each treatment arm measured by ESTIMATE analysis. Box plots indicate median and quartiles. (C) Summary of the molecular and immunological effects of lenvatinib, anti-PD1, and combination treatment on the tumor. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Abbreviations: C, combination; ESTIMATE, estimation of stromal and immune cells in malignant tumor tissues using expression data; iDC, immature dendritic cell; L, lenvatinib; NK, natural killer; P, placebo; PD1; anti-PD1; ssGSEA, single-sample gene set enrichment analysis; Th, T helper; Treg, regulatory T cell.

Overall, though all treatments were able to impact the molecular and immunological profile of tumors, only the combination treatment induced an immuno-modulatory active profile associated with proinflammatory signaling and TGFß inhibition (Fig. 6C).

Lenvatinib Suppresses Aberrant Angiogenesis in Tumors

Given that lenvatinib is a known antiangiogenic agent, the capacity of the treatments to reduce aberrant vasculature in tumors was investigated. In the GSEA analysis, lenvatinib and combination groups showed a reduction in the expression of endothelial-related genes, which could be indicative of decreased angiogenesis (Supporting Fig. S9D,F). IHC analysis showed a reduction in the amount of vessels encapsulating tumor clusters (VETCs) in the lenvatinib and combination arms, reaching significance at the late time point ($P < 0.05$ vs. placebo; Supporting Fig. S8E,F). VETCs have been proposed as a predictor of aggressive HCC[21] and could be linked to the antitumoral effect of lenvatinib. CD31 staining showed similar results (Supporting Fig. 8G). At the late time point, only the combination group maintained a significant reduction in CD31 expression ($P < 0.05$ vs. placebo). Globally, our results indicate that lenvatinib and combination treatments exerted an antiangiogenic effect (Fig. 6C).
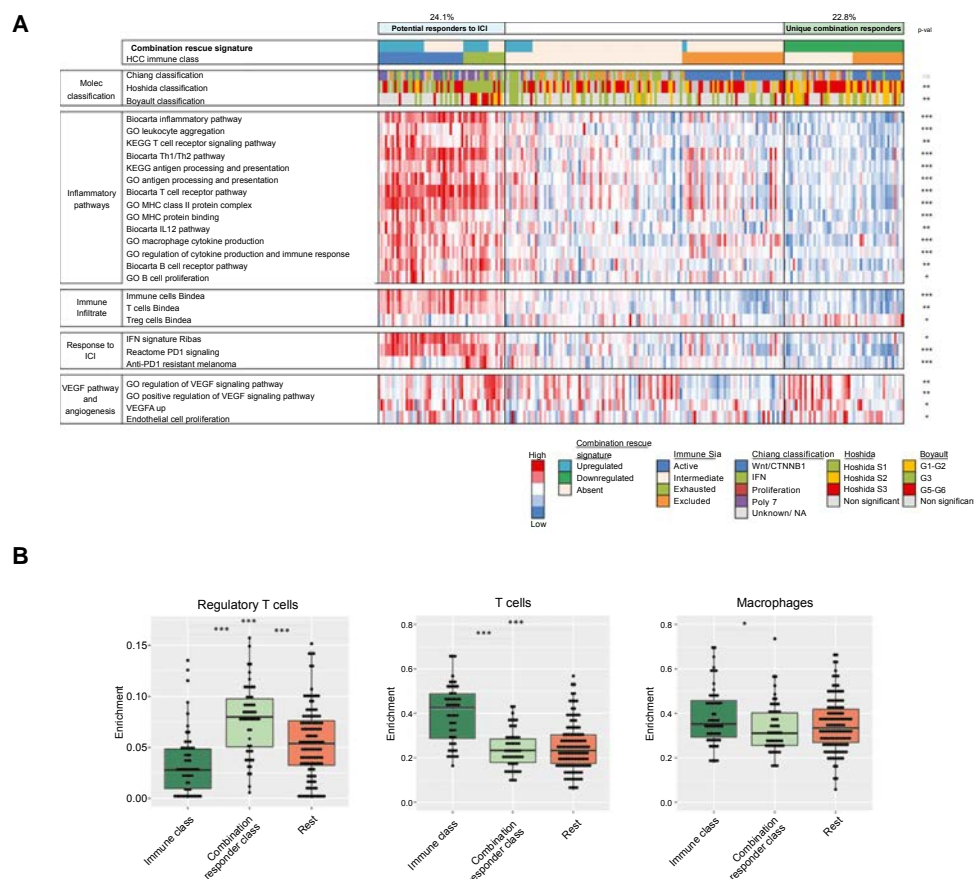
FIG. 7. Identification of HCC human tumors expressing the combination rescue signature. (A) Transcriptomic profile of HCC samples classified as HCC immune class (potential responders to ICIs)[15] or combination-only responder class. *P* values reflect the comparison of the combination-only responder class (green) and nonimmune tumors (white). (B) Estimated proportion of immune cells (CIBERSORT) in human tumors classified according to its potential response to therapies. Box plots indicate median and quartiles. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Abbreviations: GO, Gene Ontolog; ICI, immune checkpoint inhibitors; KEGG, Kyoto Encyclopedia of Genes and Genomes; NA, not available.

ENRICHMENT OF POTENTIAL RESPONDERS TO COMBINATION TREATMENT IN A HUMAN COHORT OF HCC

A Gene Signature Capturing the Molecular Effects of the Combination Treatment Identifies Potential Responders

Previous studies from our group suggested that the HCC immune class could be able to predict response to ICIs.[15] Based on our results defining the molecular impact of the combination therapy, we sought a potential biomarker capable of identifying patients that are not likely to respond to anti-PD1 alone, but could benefit from the combination treatment. To this end, the combination rescue signature was generated by selecting the top differentially expressed genes in the combination group compared to placebo, but not altered by stand-alone monotherapies (Supporting Table S8). We hypothesize that human tumors with a similar gene expression profile to the combination-treated murine tumors according to our

2664

signature (i.e., those with high immune infiltration and activation) do not need the booster effect of the combination treatment because they could potentially respond to anti-PD1 or lenvatinib monotherapies.[15] On the other hand, samples with down-regulation of genes associated with the molecular effect of the combination therapy could benefit from the immune-activating effect of the combination treatment.

Human tumors resembling the combination-treated murine tumors according to our signature were enriched in the HCC immune class (33 of 55 immune vs. 15 of 173 nonimmune; *P* < 0.0001; Fig. 7A). The HCC immune class accounted for 24.1% of the cohort (55 of 228) as reported.[15] Conversely, 22.8% of the tumors (52 of 228) were classified as combination-only responders according to the combination rescue signature, indicating that they are likely to be resistant to ICIs, but could be rescued with the combination treatment. There was no overlap of tumors expressing the HCC immune class and the combination-only responder class. Even if just patients in the immune class (24.1%) and combination-only responder class (22.8%) were considered, 46.9% (107 of 228) of the patients could potentially benefit from the combination treatment.

Characterization of Tumors From the Combination-Only Responder Class

Compared to the other nonimmune tumors, the combination-only responder class presented an enrichment of the S2 and G1 molecular classes,[22,23] associated with proliferation and progenitor-like features. There was no significant overlap with the immune excluded or catenin beta 1 (*CTNNB1*) classes,[20,24] indicating that the combination rescue signature is not recapitulating Wnt/ß-catenin activation (Figs. 7A and 8). On the other hand, the combination-only responder class displayed a significant activation of endothelial cell proliferation and VEGFR pathway. Thus, VEGF inhibition with lenvatinib could potentially boost the inflammatory status in these tumors.[10] These samples also presented activation of FGFR1-4 and RET downstream pathways, which are also targeted by lenvatinib, although no significant differences compared to other nonimmune samples were found (Supporting Fig. S10).

The combination-only responder class was also characterized by a significant decrease in proinflammatory signaling (e.g., T-cell receptor [TCR] activation,

cytokine production, and antigen presentation) and in gene sets associated with response to ICIs and T-cell infiltrate. Interestingly, we observed an enrichment of gene sets associated with Treg cells, which further highlights the relevance of this cell population in the immune-remodeling effect of the combination treatment. Cell-type abundance analysis by digital cytometry (CIBERSORT) confirmed the differences in Treg infiltrate between the combination-only responder class and the other nonimmune samples (Fig. 7B). A reduction of T-cell and macrophage infiltrates compared to the immune class was also observed. According to our data, the increased VEGF and Treg signaling in this group could be inhibited by the combination treatment, whereas other nonimmune tumors may present different mechanisms of immune suppression, which cannot be corrected with this therapeutic approach. Of note, the combination-only responder class was not associated with survival or other clinicopathological variables (Supporting Table S9).

Overall, our signature identified ~22% of human HCC patients with gene expression deregulations that could potentially be restored by the combination therapy. These samples were characterized by reduced proinflammatory signaling, high Treg levels, and VEGF signaling.

Discussion

Over the past several years, checkpoint-inhibitor–based immunotherapies have achieved unprecedented success for cancer treatment. However, responses are only observed in ~15%-20% of HCC patients, highlighting the need to identify either biomarkers of response or combination treatments that could increase their clinical benefit. Here, we generated three syngeneic models to assess the antitumoral, immunological, and molecular effects of combining anti-PD1 ICIs with the TKI, lenvatinib, currently approved for first-line treatment of advanced HCC.[9] We demonstrated that lenvatinib exerts an immunomodulatory effect, which, together with anti-PD1, could induce antitumoral immune response activation, thus providing a mechanistic rationale for this combination. We also identified a group of human HCC tumors that display features of primary resistance to ICIs, but could potentially benefit from the combination treatment.

In the quest to identify combination strategies for cancer treatment, antiangiogenic agents and

2665

multikinase inhibitors are among the most notable candidates because of their immunomodulatory capacities.[10,25] The lenvatinib plus pembrolizumab combination is a promising approach for HCC, with phase Ib data showing an unprecedented ORR of 46% and median survival of 22 months.[9] This combination is currently being assessed in a phase III trial compared to lenvatinib alone (NCT03713593). In our models, anti-PD1 and lenvatinib alone improved survival and decreased tumor growth; however, the combination treatment achieved a higher response rate, shorter time to response, and a reduction of tumor viability in accordance with a previous publication.[13]

The VEGFR family is one of the main targets of lenvatinib. VEGFA-VEGFR pathway activation favors tumor growth, progression, and aberrant vasculature formation.[10] More important, the VEGF pathway has a direct immunosuppressive effect on the tumor infiltrate by decreasing cytotoxic T-cell and DC function and promoting the recruitment of immunosuppressive cells, such as Tregs, M2 macrophages, and MDSCs.[10,26,27] In addition, the inhibition of other lenvatinib targets, such as FGFR1-4, RET, and platelet-derived growth factor, could potentially have other immunological and molecular implications. Considering this, lenvatinib could boost the effects of ICIs on antitumor immune response by "releasing the brake" on inflammation. However, the immunomodulatory capacity of lenvatinib alone or in combination with anti-PD1 still remains poorly characterized.

Here, we demonstrate that lenvatinib exerts a potent effect on the immune infiltrate by reducing the Treg proportion and altering their intratumoral location. On the other hand, anti-PD1 induced an increase of T-cell and DC1 infiltrate. This is in accordance with human studies reporting T-cell recruitment following PD1 blockade and association between DC1 and T-cell function.[19,28] The combined effect of the two therapeutic approaches induced an increase in the proinflammatory component in the tumor, associated with enhanced antitumor immunity.[29] Data from treated cancer patients indicate that intratumoral Tregs might limit anti-PD1 efficacy.[30,31] Therefore, the effect of lenvatinib in reducing this immunosuppressive population could be responsible for the greater antitumoral capacity of the combination therapy.

Previous experimental studies have suggested a link between lenvatinib and antitumor immune responses.[12,13] The combination treatment was able to increase the CD8 T-cell infiltrate and decrease the monocyte/macrophage component in an HCC murine model,[13] in line with previous studies assessing the effect of lenvatinib alone.[12] The study also reported a DC decrease, suggesting that although anti-PD1 promotes DC1 recruitment according to our data, this may not happen in all DC subpopulations. Our data indicate that the combination regimen elicited a reduction of the protumorigenic M2 phenotype compared to the anti-PD1 arm. This could be linked to the decreased Treg proportion in the combination arm, which reportedly promotes the M2 phenotype in tumor-infiltrated macrophages.[32] The reprogramming of the macrophage phenotype corresponds to *in vivo* data linking lenvatinib treatment with a reduction in M2 macrophages in colon cancer.[12] Besides its effects in tumor, it has been proposed that VEGF can cause systemic immunosuppression.[10] However, no significant alterations were detected in circulating immune cells, suggesting that the effect of the treatments may be tumor specific.

Transcriptomic analysis from tumor samples allowed us to assess the molecular and inflammatory changes induced by the treatments. Tumors from animals receiving the combination treatment showed a reduction of pathways associated with proliferation, in line with its greater antitumoral capacity and histological analysis. Interestingly, lenvatinib blocked immunosuppressive signaling through TGFß pathway inhibition, likely by decreasing the Treg proportion given that TGFß signaling is one of its key immunosuppressive mechanisms.[18] On the other hand, anti-PD1 increased the T-cell infiltrate and elicited an immune exhausted phenotype in tumor characterized by expression of immune-suppressive pathways and Treg infiltrate,[18] in accordance with human studies reporting an increase in the exhausted T-cell component following PD1 blockade.[28] Therefore, the combined effect of lenvatinib plus anti-PD1 generated both an enhanced T-cell infiltrate and a reduction of immunosuppressive signaling (i.e., Treg infiltrate and TGFß pathway). Consequently, only the combination treatment generated a specific activated antitumoral immune response.

Anti-PD1 treatment also induced a deregulation in expression of checkpoint inhibitors. CTLA4 expression was increased in tumors treated with anti-PD1 and combination treatment, probably attributable to feedback mechanisms induced by TCR stimulation.[33,34] Increased CTLA4 expression has also been associated
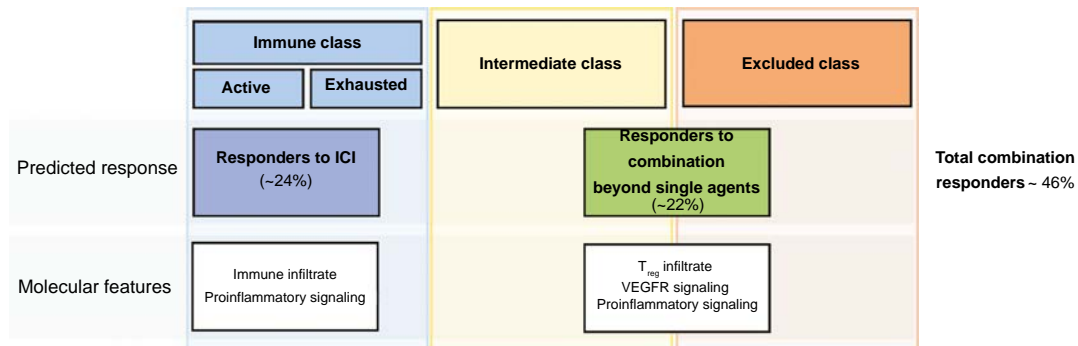
FIG. 8. HCC classification according to its immunological features and potential response to the combination therapy. Diagram summarizing the HCC immune classification and potential response to ICIs or combination treatment beyond single agents according to the combination rescue signature. Percentage of HCCs belonging to each class is shown in brackets. Abbreviations: ICI, Immune checkpoint inhibitors; Treg, regulatory T cell.

with T-cell exhaustion.[18] Overall intratumoral PD1 positivity was not altered by any treatment. Of note, anti-PD1 and the combination treatment reduced the $CD8^+PD1^+$ T-cell infiltrate. This population has been recently proposed to limit the efficacy of ICIs in NASH-related HCC.[35] Therefore, the decrease of this population could be an important mechanism promoting response to these treatments. On the other hand, this result could also be influenced by the blockade of the PD1 epitope by anti-PD1 treatment, rather than a reduction of this cell population in the infiltrate.

The identification of the molecular and immunological effects induced by lenvatinib plus anti-PD1 is a step toward in identifying patients with primary resistance to ICI monotherapies who could benefit from the combination treatment. In a previous study, we defined a gene-expression–based immune classifier able to identify 24% of HCC patients (immune class) with markers of T-cell infiltrate and molecular features similar to melanoma tumors most responsive to ICIs.[15] Using gene expression data from our murine model, we here generated a signature capturing the transcriptomic modifications induced by the combination therapy, but not by monotherapies as stand-alone therapies. The assessment of our signature in a human cohort of 228 samples showed that 22% of patients presented down-regulation of genes associated with the molecular effect of the combination, along with reduced proinflammatory signaling, high Treg levels, and VEGF signaling. Thus, tumors sharing these hallmarks could harbor primary resistance to

anti-PD1, but potentially benefit from the booster effect of the combination treatment. Altogether, around half of patients could respond to combination therapies (Fig. 8). Considering that the gene signature has been generated from on-treatment murine tumor samples, further studies extrapolating these findings to pretreatment HCC profiles will be required. Therefore, the predictive capacity of response of the signature will need to be validated in a prospective human cohort of HCC patients receiving lenvatinib plus an anti-PD1 ICI.

In summary, our study provides a characterization of the antitumor, immunological, and molecular effects of lenvatinib, anti-PD1, and its combination in murine HCC models. We demonstrated that lenvatinib exerts an immunomodulatory effect on the tumor infiltrate associated with a reduction in Tregs and inhibition of immune-suppressive pathways. Its combination with anti-PD1 favored the generation of an activated immune profile and faster response to treatment. Finally, the identification of the mechanisms underlying a beneficial effect of the combination therapy led to the generation of a gene signature present in ~20% of human HCC that correlates with high Treg infiltrate, VEGFR pathway activation, and low inflammatory signaling. This signature might recognize patients likely to benefit most from this combination. Thus, further investigations are warranted to confirm whether the signature can be a tool to identify HCC patients who may respond to the combination therapy beyond their responses to single-agent therapies.

2667

Author Contributions: L.T., D.S., and J.M.L. were involved in the study conceptualization. L.T., C.M., M.P., A.M., J.L., P.K.H., M.M., U.B., C.E.W., J.A.F., M.P.G., and J.P. contributed to the investigation and methodology. C.M., A.M., and C.E.W. contributed to pathological characterization of tumor samples. L.T. and M.T.M. contributed to data curation and software analysis. L.T., C.M., M.P., P.K.H., A.M., C.E.W., J.A.F., M.T.M., and D.S. contributed to formal analysis. C.E.W., M.T.M., B.S., S.L.F., D.A.M., D.S., and J.M.L. provided supervision. J.M.L. contributed to funding acquisition. L.T., D.S., and J.M.L. wrote the manuscript. All authors were involved in the critical review and editing of the manuscript.

REFERENCES

1) Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-249.

2) Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2021;7:6.

3) Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378-390.

4) Cheng AL, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, et al. Phase III trial of lenvatinib (LEN) vs sorafenib (SOR) in first-line treatment of patients (pts) with unresectable hepatocellular carcinoma (uHCC). J Clin Oncol 2017;35(Suppl.):Abstract 4001.

5) Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N Engl J Med 2020;382:1894-1905.

6) Llovet JM, Villanueva A, Marrero JA, Schwartz M, Meyer T, Galle PR, et al. Trial design and endpoints in hepatocellular carcinoma: AASLD Consensus Conference. HEPATOLOGY 2021;73:158-191.

7) Finn RS, Ryoo BY, Merle P, Kudo M, Bouattour M, Lim HY, et al. Results of KEYNOTE-240: phase 3 study of pembrolizumab (Pembro) vs best supportive care (BSC) for second line therapy in advanced hepatocellular carcinoma (HCC). J Clin Oncol 2019;37(Suppl. 15):Abstact 4004.

8) Yau T, Kang YK, Kim TY, El-Khoueiry AB, Santoro A, Sangro B, et al. Efficacy and safety of nivolumab plus ipilimumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib: the CheckMate 040 Randomized Clinical Trial. JAMA Oncol 2020;6:e204564.

9) Finn RS, Ikeda M, Zhu AX, Sung MW, Baron AD, Kudo M, et al. Phase ib study of lenvatinib plus pembrolizumab in patients with unresectable hepatocellular carcinoma. J Clin Oncol 2020;38:2960-2970.

10) Fukumura D, Kloepper J, Amoozgar Z, Duda DG, Jain RK. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. Nat Rev Clin Oncol 2018;15:325-340.

11) **Longo V**, **Gnoni A**, Gardini AC, Pisconti S, Licchetta A, Scartozzi M, et al. Immunotherapeutic approaches for hepatocellular carcinoma. Oncotarget 2017;8:33897-33910.

12) Kato YU, Tabata K, Kimura T, Yachie-Kinoshita A, Ozawa Y, Yamada K, et al. Lenvatinib plus anti-PD-1 antibody combination treatment activates CD8+ T cells through reduction of tumor-associated macrophage and activation of the interferon pathway. PLoS One 2019;14:e0212513.

13) Kimura T, Kato YU, Ozawa Y, Kodama K, Ito J, Ichikawa K, et al. Immunomodulatory activity of lenvatinib contributes to anti-tumor activity in the Hepa1-6 hepatocellular carcinoma model. Cancer Sci 2018;109:3993-4002.

14) Brown ZJ, Heinrich B, Greten TF. Mouse models of hepatocellular carcinoma: an overview and highlights for immunotherapy research. Nat Rev Gastroenterol Hepatol 2018;15:536-554.

15) 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.

16) **Villanueva A**, **Portela A**, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. HEPATOLOGY 2015;61:1945-1956.

17) **Galon J**, **Costes A**, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960-1964.

18) Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol 2015;15:486-499.

19) **Böttcher JP**, **Reis e Sousa C**. The role of type 1 conventional dendritic cells in cancer immunity. Trends Cancer 2018;4:784-792.

20) Pinyol R, Sia D, Llovet JM. Immune exclusion-WNT/CTNNB1 class predicts resistance to immunotherapies in HCC. Clin Cancer Res 2019;25:2021-2023.

21) **Renne SL**, **Woo HY**, **Allegra S**, Rudini N, Yano H, Donadon M, et al. Vessels encapsulating tumor clusters (VETC) is a powerful predictor of aggressive hepatocellular carcinoma. HEPATOLOGY 2020;71:183-195.

22) Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet JP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385-7392.

23) 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.

24) 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.

25) Khan KA, Kerbel RS. Improving immunotherapy outcomes with anti-angiogenic treatments and vice versa. Nat Rev Clin Oncol 2018;15:310-324.

26) Gavalas NG, Tsiatas M, Tsitsilonis O, Politi E, Ioannou K, Ziogas AC, et al. VEGF directly suppresses activation of T cells from ascites secondary to ovarian cancer via VEGF receptor type 2. Br J Cancer 2012;107:1869-1875.

27) Gabrilovich DI, Chen HL, Girgis KR, Cunningham HT, Meny GM, Nadaf S, et al. Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells. Nat Med 1996;2:1096-1103.

28) **Yost KE**, **Satpathy AT**, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med 2019;25:1251-1259.

29) Arce Vargas F, Furness AJS, Litchfield K, Joshi K, Rosenthal R, Ghorani E, et al. Fc effector function contributes to the activity of human anti-CTLA-4 antibodies. Cancer Cell 2018;33:649-663.e4.

30) Vignali DAA, Collison LW, Workman CJ. How regulatory T cells work. Nat Rev Immunol 2008;8:523-532.

31) Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T cells and immune tolerance. Cell 2008;133:775-787.

32) Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat Rev Cancer 2019;19:133-150.

33) **Kamphorst AO**, **Pillai RN**, Yang S, Nasti TH, Akondy RS, Wieland A, et al. Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. Proc Natl Acad Sci U S A 2017;114:4993-4998.

34) Lindsten T, Lee KP, Harris ES, Petryniak B, Craighead N, Reynolds PJ, et al. Characterization of CTLA-4 structure and expression on human T cells. J Immunol 1993;151: 3489-3499.

35) Pfister D, Núñez NG, Pinyol R, Govaere O, Pinter M, Szydlowska M, et al. NASH limits anti-tumour surveillance in immunotherapy-treated HCC. Nature 2021;592: 450-456.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.32023/suppinfo.

2669

DISCUSSION

Molecular profiling of tumors using translational approaches has significantly contributed to the understanding of the molecular pathogenesis of HCC[3,83]. The three articles presented in this doctoral thesis provide a comprehensive analysis of the molecular and immunological features of HCC based on multi-omic techniques and animal models of cancer. Specifically, they provide a characterization of genetic and molecular determinants associated with HCC in specific populations or groups of patients, and mechanistic rationale for novel combination therapies for this devastating disease.

1. Molecular Features of HCC in Mongolia

Liver cancer presents large geographical variations in incidence in relation to the prevalence of risk factors for liver diseases such as HBV and HCV infection, alcohol consumption and NASH[2,3] (**Figure 6**). Considering this, the biological characterization of cancer patients from regions with different incidences using next-generation sequencing technologies provides an opportunity to study the **heterogeneity among populations** and unveil distinct molecular particularities or and presence of putative risk factors. An example of this was provided by the TIGER-LC initiative, which unraveled previously unknown biological features of liver cancer in Thailand. Following this approach, in the **1st study** of this doctoral thesis, we provide a comprehensive characterization of the molecular profile of **HCC in Mongolia**, the country the highest incidence worldwide (85.6 cases per 100,000 inhabitants)[62] compared to Western HCC.

In Mongolia, HCC is the most **frequently diagnosed cancer** in both sexes and in all regions of the country[63] (**Figure 12**). Despite a strikingly high prevalence of HBV (10.6%), HCV (6.4%), and HDV (70% of HBV-positive individuals) infections and alcohol consumption[56,64,65], it is unclear whether HCC incidence is completely explained by the unique combination of **risk factors**, or eventually, other unknown factors might be responsible. By analyzing WES and RNA-seq data from two new cohorts consisting of 192 Mongolian HCC and 187 Western HCC (European and American), we identified distinct **genomic and transcriptomic footprints** in Mongolian tumors that suggest the presence of specific genetic factors in the country that could contribute to the higher HCC incidence in this country. Notably, a previous study provided valuable data about the molecular landscape in Mongolian HCC[153], but a comparison with an in-house Western cohort and in-depth analysis of potential environmental agents based on mutational fingerprints and virological characterization were still required. Thus, further analyses comparing Mongolian HCC with Western tumors were still needed to better understand potential associations between

molecular traits and high HCC prevalence in this country. Herein, we were able to unveil novel clinical and virological characteristics, mutation profile, and transcriptomic-based molecular classes in Mongolian HCC patients.
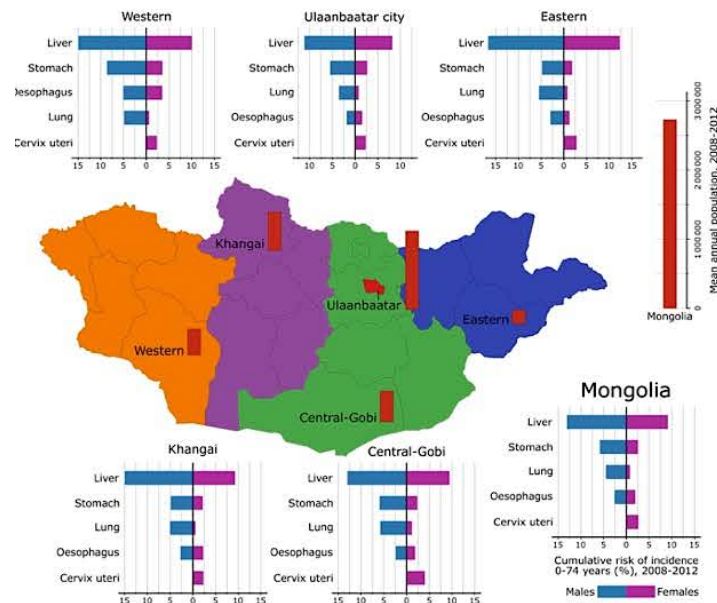
Figure 12. Cancer incidence in Mongolia. Mean annual population and cumulative risk (0–74 years, %) of most common cancer sites, by sex and region in Mongolia (2008–2012). Extracted from Chimed T et al., Int J Cancer 2017[63].

HCC in Mongolia has a strikingly **high prevalence among females**, with a male to female ratio of 1.5/1[76]. This contrasts with the strong male predominance observed in HCC patients globally (male to female ratio of 2–3:1), likely related to differential exposition to risk factors as well as differences in sex hormones[3]. Consistently, our Mongolian cohort presented 46% of females, as opposed to 20% observed in the Western cohort[3]. Other remarkable clinical characteristics in the Mongolian cohort were younger age, earlier liver fibrosis stages. and higher viral infection rates, all in accordance with previous studies[66,153]. In this regard, the prevalence of **HBV and HCV** in Mongolia is strikingly high, with about 20% of the population being infected by at least one or more types of viruses, and many of them not being aware of their status[63]. Mongolia also has the highest prevalence of **HBV-HDV coinfection** in the world[59]. In our study, we observed a higher rate of HBV-HDV co-infection (84% of HBV-infected patients) in Mongolian patients compared to less than 7% of coinfection in the Western cohort. HDV infection has been associated with a more severe course of liver disease and increased risk of HCC compared to HBV infection alone and could contribute to hepatocarcinogenesis through oncogenic mechanisms independent from HBV[122]. Finally, regarding the potential oncogenic role of AAV in this population, a previous study reported a very low prevalence of oncogenic AAV integration

in Mongolian and Thai HCC patients (<1%), likely due to epidemiological differences between Asian and European patients[60].

We further evaluated viral characteristics of the Mongolian and Western datasets to assess whether HBV infection in Mongolia presented features associated with elevated oncogenic potential. Notably, HBV-infected individuals in our Mongolian cohort were genotype D, which is known to be almost universal in Mongolia[160], while Western patients presented both genotypes C and D. Genotype D has been previously associated with reduced HCC development as compared to genotype C, suggesting lower oncogenic potential in HBV from Mongolia[154]. Another relevant characteristic of HBV infection is the presence of basal core promoter (BCP) and precore **HBV mutations**, which have been associated with liver disease progression and a higher risk of HCC development[161]. The prevalence of these mutations was significantly lower in the Mongolian cohort compared to Western, with mutational frequencies similar to those previously reported[162]. These data show that the rate of BCP and precore HBV pro-oncogenic mutations in Mongolia is particularly low despite the predominance of genotype D in this population, which has been associated with a higher rate of HBV mutations[155]. Overall, HBV viral characteristics in Mongolia are highly homogeneous and with **low oncogenic potential**, suggesting that other features besides HBV infections – such as HDV or other agents – might be contributing to the high HCC incidence rates in this country[154,155].

From the genomic standpoint, we unveiled a higher rate of **protein-coding mutations** in Mongolian HCC, almost doubling that in the Western in-house cohort (121 vs 70 mutations per tumor) and publicly available datasets[82,83,85,153], which could be due to intrinsic and/or extrinsic factors promoting mutagenesis in Mongolian HCC. The most commonly mutated genes in both cohorts aligned with previous studies in HCC[83] (e.g., mutations in *CTNNB1*, *TP53*, *ARID1A, ALB*…) (**Table 2**) indicating that the mutational spectrum of Mongolian HCC resembled that of Western HCC. Nevertheless, several **HCC drivers** were significantly more mutated in Mongolian HCC, including *APOB* (15% vs 5% in Western HCC), *TSC2* (9% vs 1%), and *NFE2L2* (6% vs 1%), and the *KMT2* gene family (34% vs 17%) (**Figure 13**). Notably, we detected an enrichment in damaging mutations affecting the ***TSC2* gene** in Mongolian tumors. This pattern of mutations suggests a positive selection of this alteration and a potential driver role[156]. *TSC2* is a known cancer-related gene participating in the mTOR oncogenic pathway[83,163] and it has been proposed as an actionable alteration with level 2B evidence, as it could be a predictor of response to the FDA-approved drug everolimus[163]. Thus, these patients could likely benefit from everolimus treatment, despite this drug was not effective in all-comer HCC patients according to phase III

data[164]. Considering that systemic therapies for HCC treatment are not routinely used in Mongolia to this day[160], the potential implementation of a targeted approach using everolimus in this group of patients faces many challenges. Overall, the mutation profile in HCC driver genes aligned with data from a previously described cohort of 71 Mongolian HCC patients[153]. Compared to this study, herein we were able to 1) identify a higher mutational burden in Mongolian HCC, and 2) confirm statistically significant differences in the mutation rate of HCC drivers in Mongolia compared to an in-house Western cohort.
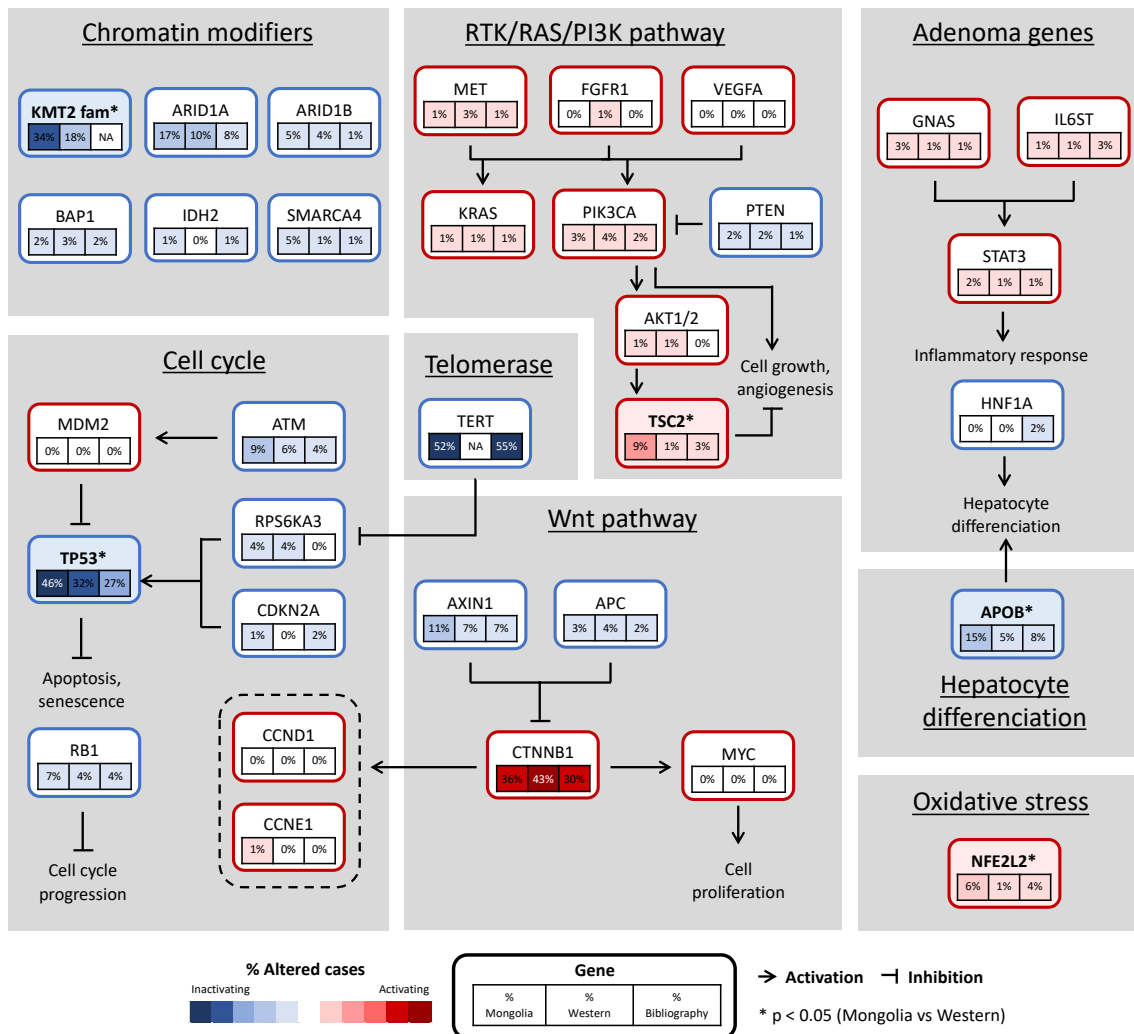
Figure 13. Somatic mutation in HCC-driving signaling pathways. Mutated genes in HCC grouped by the main deregulated signaling pathways proposed by Ally A et al., Cell 2017[82]. Molecular interactions between genes are represented. The percentage of mutations in the Mongolian cohort, Western cohort, and previous bibliography[85,89] are indicated for each gene. Genes with significant differences between Mongolian and Western HCC are highlighted. Original figure.

To investigate whether Mongolian HCC presents distinct genomic footprints, we assessed the presence of **mutational signatures**[21,22]. This approach has been used in other cancers to propose

potential exogenous exposure capable of explaining differences in incidence. For instance, a recent study performed mutational signature analyses in squamous cell carcinoma samples from eight countries with varying incidence, which unveiled very similar exposure profiles between countries, including tobacco, alcohol, and opium[165]. In our study, *de novo* signature analysis revealed the presence of a new mutational signature (**SBS Mongolia**) significantly enriched in Mongolian HCC (25% vs 4.5% in Western HCC), indicating unique substitution patterns characterized by increased frequency of T>G substitutions. Other mutational signatures previously associated with HCC such as SBS5, SBS40, and SBS22 were also detected, with no differences between cohorts. The mutational signature profile reported in a previous study of Mongolian HCC also aligned with these results, including signatures associated with tobacco, alcohol consumption, or aristolochic acid[153]. However, the presence of *de novo* signatures was not assessed, and the study did not include a comparison of the mutational landscape with a Western HCC cohort.

Interestingly, Mongolian HCC samples from our cohort presenting SBS Mongolia were significantly enriched in a mutational signature associated with exposure to **dimethyl sulfate** (DMS, 71.1% in positive tumors for SBS Mongolia vs 26.5% in negative tumors). The International Agency for Research on Cancer classifies DMS as a probable carcinogenic hazard to humans (category 2A carcinogen)[26], which is a byproduct of **coal combustion**. The DMS signature was also more common in Mongolian HCC compared to Western, potentially indicating higher exposure to DMS in this country. In this regard, most of the Mongolian population is currently exposed to coal combustion, which is used to fight against the intense cold weather both in urban and rural areas. Half of the 3-million population of Mongolia lives in Ulaanbaatar, an overpopulated capital with one of the **highest levels of air pollution in the world**[166]. The rest is still predominantly nomad and lives in traditional tents or *gers*, where coal is used both for cooking and heating. This fact has been recognized by international organizations as a **major health threat** in this country[166,167]. In fact, air pollution from coal combustion has been reported to account for 40% of lung cancer deaths in Ulaanbaatar, corresponding to almost 10% of total deaths in the city[168]. Considering all this, our results suggest that long-term exposure to DMS from coal combustion could also be a risk factor for HCC development in Mongolia. In this regard, the DMS signature was associated with older patients, potentially due to a longer exposure time.

Further studies will be required to gain a **mechanistic understanding** of how these signatures arise. Specifically, two major streams of investigation are required to confirm the association

between a mutational signature and a mutagen[22]. First, the presence of the signatures needs to be assessed in model systems exposed to the genotoxic. For instance, it should be confirmed whether DMS and/or byproducts of coal combustion in animal or *in vitro* models originate the DMS signature, SBS Mongolia or both. In second place, epidemiological studies assessing the onset of tumors presenting such mutational footprints in a population exposed to the mutagen would be required. In this case, a causal relationship between DMS and HCC in Mongolia would have to be supported by a positive association between 1) history of exposure to coal combustion and 2) HCC tumors with a strong contribution of the DMS and SBS Mongolia signatures. This has been the case for previous signatures of exposure to HCC risk factors such as aristolochic acid signature (COSMIC signature SBS22), for which the association between the signature identified in human cancers and the genotoxic was validated *in vitro*[169], and exposure of patients presenting the signature was confirmed epidemiologically[113,170].

Finally, our transcriptomic analysis revealed that Mongolian tumors presented a **distinct transcriptomic profile** that did not fit into the classical proliferation and non-proliferation subgroups reported in HCC[3,100]. Specifically, Mongolian HCC was characterized by an enhanced proliferative and immunological signaling, with a proportion of tumors belonging to **proliferative HCC classes** (39%) doubling the one in Western HCC and previously published studies (~20%)[100]. Mongolian HCC clustered into **three molecular clusters** (MGL1-3). Two of these classes (MGL2 in 26% of the patients and MGL3 in 30%) presented **distinct molecular and clinical features** compared to Western HCC and thus were deemed unique for Mongolian tumors. Both **MGL2** and **MGL3** classes were enriched in HBV/HDV infection and younger patients. Interestingly, the MGL2 class was associated with clinical and molecular features of aggressiveness and showed a **male to female ratio** of 1:2, while MGL3 presented an inflamed profile, potentially due to an immunological response to HBV/HDV infection[59]. The increased HCC incidence among females in Mongolia[76] may be due to higher exposure to risk factors compared to males (e.g., environmental agents or viral hepatitis), thus compensating the conventional gender imbalance found in this tumor type. However, the reason why females are enriched in the aggressive HCC cluster needs to be further understood. The molecular profile of the identified MGL1-3 classes aligned with previously-proposed transcriptomic-based clusters in Mongolian HCC[153]. Overall, in this study, we were able to identify transcriptomic differences compared to Western HCC and novel clinic-pathological characteristics associated with our classification.
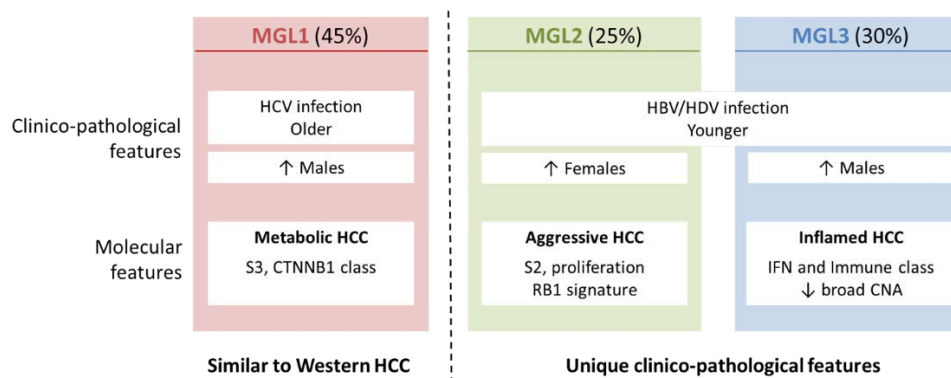
Figure 14. Molecular classification of Mongolian HCC. Mongolian HCC can be classified into three molecular clusters with distinct clinico-pathological and molecular features. HCV, hepatitis C virus; HBV, hepatitis B virus; HDV, hepatitis D virus; CNA, copy number alterations. Original figure.

In conclusion, we provided an exhaustive **comparison** of the genomic and transcriptomic characteristics of **Mongolian HCC** with an in-house **Western cohort**. We were able to identify novel features of Mongolian tumors, including 1) **virological traits** associated with low oncogenic potential; 2) high **mutational rates**; 3) a distinct **mutational signature** associated with environmental agents; and 4) a **transcriptomic profile** characterized by two molecular classes not present in Western HCC. Based on our results, environmental factors such as DMS need to be further explored as a **potential risk factor** in this population.

2. Risk of AAV Integration in NAFLD Patients Undergoing Gene Therapy

Chronic viral is one of the main risk factors leading to HCC[3], which occurs mainly through the induction of chronic liver disease and cirrhosis due to persistent inflammation and oxidative stress[103] (**Figure 15**). This mechanism explains most HCV-related HCC, which do not present a clear genetic mechanism of carcinogenesis. In contrast, direct oncogenic effects associated with HCC development have been linked to HBV infections and, more recently, to **AAV2**[3,4]. Both viruses have the capacity to integrate into the human genome, thus giving rise to **insertional mutagenesis** and subsequent deregulation of neighboring genes leading to HCC[17] (**Figure 15**).

Despite mounting evidence linking AAV insertional mutagenesis with HCC, this virus is currently considered non-pathogenic in humans[103]. **Recombinant AAV** (rAAV) – and especially the AAV2 serotype – is a widely used vector for gene replacement, silencing, and editing which holds great promise for the implementation of **gene therapies** in the clinical setting. In the past few years,

two AAV-based gene therapies have been approved by regulatory agencies for the treatment of spinal muscular atrophy and retinal dystrophy[105]. Furthermore, ~140 active clinical trials are currently ongoing and could result in groundbreaking therapeutic advances for a great variety of medical conditions including genetic disorders, neurological diseases, and cancer (**Figure 16**). However, compelling studies linking AAV2 infection with HCC development are posing serious safety concerns[4,105].
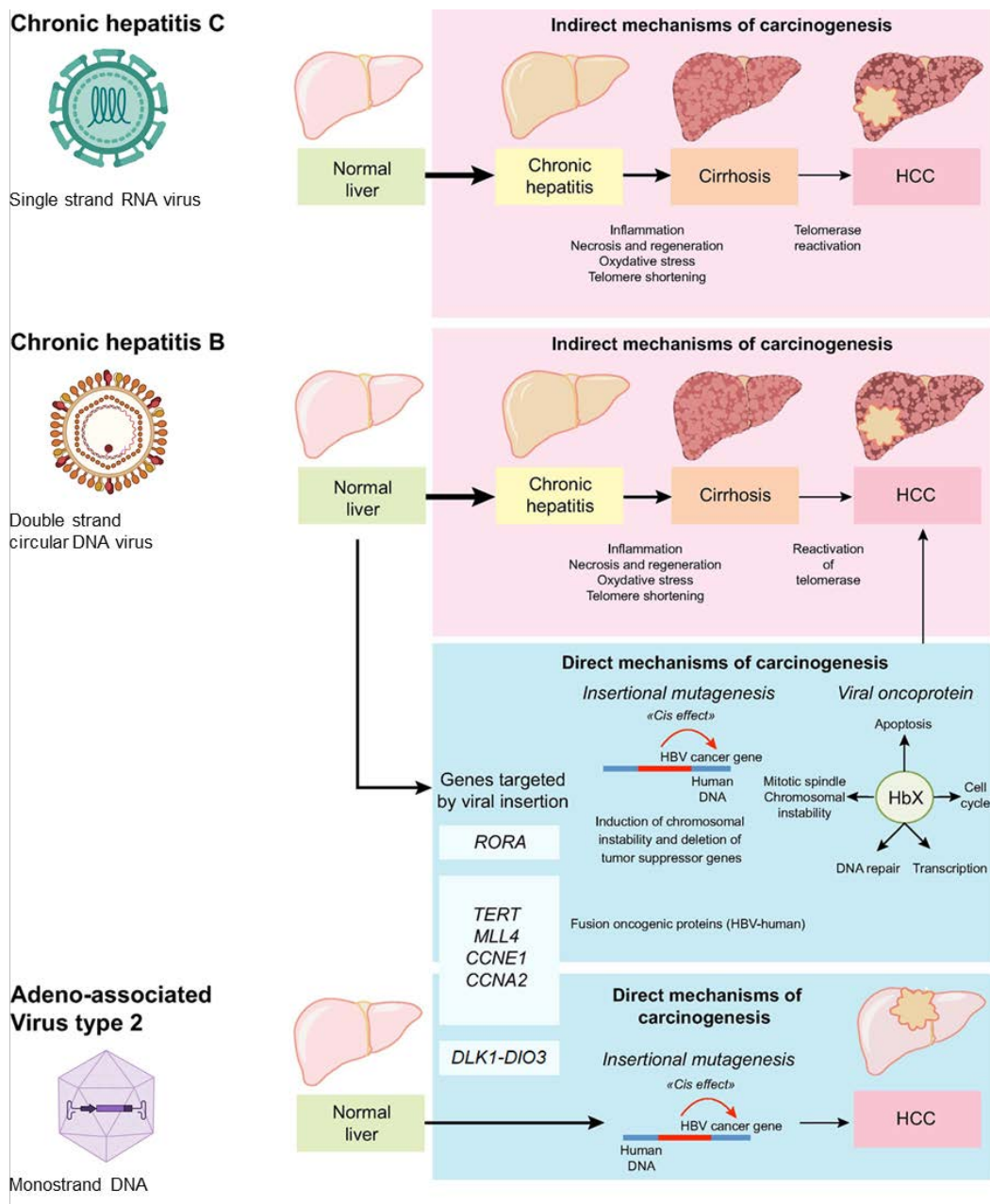
Figure 15. Viral mechanisms of liver carcinogenesis. Direct and indirect mechanisms of viral related liver carcinogenesis are represented for hepatitis C virus (HCV), hepatitis B virus (HBV), and adeno-associated virus type 2 (AAV2). Indirect mechanisms are related to the development of cirrhosis triggered by chronic inflammation and oxidative stress induced by chronic viral hepatitis. Direct

oncogenic mechanisms are mainly due to action of viral oncoproteins (Hbx in HBV), chromosomal instability induced by HBV integration and insertional mutagenesis (HBV and AAV2) with aberrant regulation of gene expression. The genes targeted by clonal viral integrations are represented in the blue box. Adapted from Schulze K et al. J Hep 2016[103] using biorender.com.
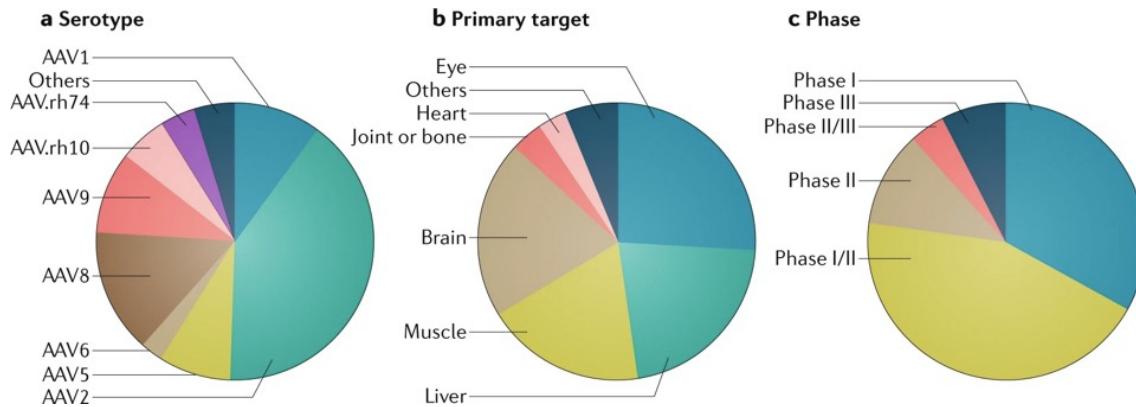
Figure 16. Overview of recombinant AAV interventional gene therapy clinical trials. Clinical trials registered in ClinicalTrials.gov, accessed on 13 November 2018 (n = 145). Trials are categorized based on adeno-associated virus (AAV) capsid serotype (a), primary tissue target for gene delivery (b), and clinical trial phase (c). Extracted from Wang D et al., Nat Rev Drug Discov 2019[105].

AAV2 integration capacity has been reported to be enhanced in cells undergoing cell cycle progression[110]. While adult hepatocytes are quiescent under homeostatic conditions, they undergo proliferation in response to liver injury[111], thus potentially favoring such integration. The clearest example of this phenomenon occurs after partial hepatectomy. Importantly, conditions of **chronic liver injury and inflammation** are also associated with compensatory proliferation as the body seeks to repair the damaged liver. Considering this, in the **2nd study** of this doctoral thesis, we assessed whether common causes of chronic liver disease such as **NAFLD** could potentially favor AAV2 integration in the genome due to increased liver damage, inflammation, and regenerative proliferation of hepatocytes[111]. To do so, neonatal and adult mice were infected with an AAV editing vector targeting the oncogenic *Rian* locus (AAV-Rian). Animals were treated with HFD to induce NAFLD-like liver injury, or partial hepatectomy was performed to promote hepatocyte proliferation. Herein, we showed that hepatocyte proliferation and NAFLD-associated liver damage increased HCC formation in a murine model treated with rAAV gene targeting.

Previous studies by our group[158] and others[106] showed that AAV2 was able to induce HCC in neonatal mice through random vector integration within the oncogenic ***Rian* locus** in the murine chromosome 12. The integration site corresponded to a cluster of oncogenic microRNA genes,

which become activated by the promoter contained within the recombinant virus[158,171]. HCC tumors generated by AAV integration into the *Rian* locus were molecularly and histologically similar to the C3 subclass of human HCCs present in 6-19% of HCC patients[108,109]. This miRNA-based subclass is characterized by overexpression of the human ortholog of the *Rian* cluster (*DLK-DIO3* locus), as well as an aggressive phenotype[158]. To mimic AAV integration in this locus, a rAAV targeting the murine *Rian* locus was used in this study. Interestingly, RNA-sequencing analysis of murine tumors from our model revealed remarkable molecular similarities with this HCC subclass, associated with proliferation, aggressive phenotype, and poor prognosis. Furthermore, deep sequencing analyses of HCC tumors have previously revealed AAV integration in known cancer driver genes in 2-5% of HCC cases, thus providing further evidence of the **oncogenic potential** of oncogenic AAV integration in the human genome[4,60]. Further analyses will be required to determine whether sequence motifs, chromatin states, and vector characteristics influence integration preferences in the *Rian* locus and other oncogenic sites[106].

According to previous data, adult mice typically do not develop AAV-induced liver cancer in the absence of injury[172,173]. To determine whether hepatocyte proliferation impacts rAAV-related oncogenesis, **two injury regimens** were evaluated: **NAFLD** induced by the administration of a HFD and **partial hepatectomy**. Both conditions led to HCC development in 100% of the rAAV-infected adult mice, compared to only 5% in untreated infected animals. Conversely, HCC rates in neonates were significantly higher than in adults irrespectively of whether they received high-fat or normal diets (i.e., 100% incidence in all male groups). The age-dependence of rAAV-induced cancer is likely linked to the rate of hepatocyte proliferation, which is high in neonates due to natural liver growth. This enhanced hepatocyte proliferation would explain the higher HCC incidence in models with liver injury. Overall, these results raise **concerns for the use of rAAV** therapy in the general human population, where chronic inflammatory liver diseases are very prevalent and could promote rAAV integration. For instance, up to 30% of the population in the US have fatty liver disease, 0.34% have HBV infection, 1.7% have HCV infection, and 4.3% have alcoholic liver disease[159]. These inflammatory conditions alone are risk factors for HCC development, thus adding a rAAV-associated oncogenic risk may significantly diminish the possible therapeutic benefits of gene therapy.

Most human HCCs arise in the background of chronic liver diseases characterized by injury and inflammation[3,9]. Inflammation itself may promote hepatocarcinogenesis through several mechanisms such as DNA damage from reactive oxygen species, changes in the immune system milieu, and an increase in hepatocyte proliferation. This is especially true in NAFLD, which has

been linked to alterations in the liver immune infiltrate, leading to hepatocyte damage and HCC[3,174]. In line with this, transcriptomic analysis of background liver samples from our murine model showed that HFD promoted a protumorigenic immune cancer field as well as activated lipid metabolism pathways in the background liver[125]. This aberrant immune field has been associated with activation of pro-oncogenic **immunosuppressive signaling** (e.g., TGFβ signaling and T cell exhaustion) and a higher risk of HCC develoment[125]. Therefore, this immunosuppressive milieu induced by HFD and NAFLD could have a key role in the increased risk of rAAV-induced HCC in our model. Notably, previous studies from our group revealed that NASH-HCC showed a significantly higher prevalence of the immunosuppressive cancer field compared to other etiologies[175].

As previously discussed in this doctoral thesis, HCC has a strong male predominance, which can be partially attributed to anti-inflammatory properties from **estrogens** in females[3]. In our study, we confirmed that female mice are less susceptible to AAV-induced HCC compared to males. We showed that estrogen treatment in male mice fed with a HFD reduced liver damage and fat deposition. This aligns with human studies showing lower NASH prevalence and less severe NAFLD stages in female individuals compared to males and postmenopausal women due to a protective role of estrogens[176]. In our model, estrogen also altered the immune milieu in males with NAFLD, decreasing pro-oncogenic immune exhaustion signaling (e.g., TGF-β and Wnt/β-catenin pathways) and promoting an adaptive immune response. No information regarding sex differences in oncogenic AAV integration has been currently reported in humans, likely due to the low availability of AAV-related HCC samples.

We also investigated whether AAV infection could result in **spontaneous viral integration** in the *Rian* locus as indicated by previous studies[171]. We observed a high incidence of HCC in adult mice receiving HFD who were infected with the control AAV (tdTomato AAV) which, unlike the *Rian* AAV, does not have the capability for targeted integration into the genome. Specifically, 5 out of 10 mice on the HFD who received the control AAV developed HCC. These tumors presented increased expression of genes located in the *Rian* locus in line with results from AAV-Rian-induced HCC, as well as a similar gene expression profile and pathway activation, indicating potential integration in the *Rian* locus. This suggests that random integrations into the *Rian* locus are common and highly carcinogenic in mice. The fact that only control AAV-infected mice receiving HFD developed tumors (as opposed to control AAV-infected mice receiving normal diet) further suggests that inflammation and hepatocyte proliferation increase HCC incidence. Whether the **human liver** is similarly susceptible to random rAAV leading to HCC needs to be

further investigated. In this regard, the reasons for the difference between the high rate of AAV2 infection in humans (40-80% seropositivity in the general population) and the low rate of AAV2 related HCC is still unclear[103].

Overall, this study demonstrates that adult **mice infected with both targeted and non-targeted rAAV present increased development HCC in the presence of liver injury** such as fatty liver induced by HFD. This was likely due to increased hepatocyte proliferation and the generation of a protumorigenic **immune cancer field** effect in the liver. Given the high prevalence of inflammatory liver conditions in the general population such as NAFLD and NASH, this should raise an alarm about the use of rAAV, an otherwise promising vector used for diverse **gene therapy** applications. Female mice were less susceptible to rAAV-induced HCC compared to males, part of which is due to a more favorable immune milieu related to **estrogen**, and likely applies to humans given the documented gender differences. Given the promise and widespread use of rAAV for gene therapies, more studies are needed to assess the risks of rAAV-induced hepatocarcinogenesis, including the monitoring of patients treated with rAAV vectors in search of viral integration and cancer development[104].

3. Immunomodulatory Effects of Lenvatinib Plus Anti-PD1 Combination

During the last decade, major breakthroughs have dramatically improved the landscape of **advanced HCC treatment** thanks to the approval of novel systemic therapies including TKI and ICI. Despite this, survival benefits observed in clinical trials for single agents are modest, and responses for ICI monotherapy are only observed in ~15-20% of patients[3]. Therefore, there is an **urgent need** to identify existing therapies that can effectively synergize with ICI. In this regard, **antiangiogenic drugs** constitute the basis of the standard of care treatment in advanced HCC (**Figure 17**) and have been proposed as ideal candidates for **combination with ICI** due to their immunomodulatory capacities[130,177]. Effectively, the new combination of the ICI atezolizumab plus bevacizumab (VEGFA inhibitor) demonstrated objective responses in 36% of patients and has been FDA-approved as first-line therapy[143]. Similarly, the combination of **lenvatinib plus pembrolizumab** is a promising approach for HCC, with phase Ib data showing an unprecedented ORR of 46% and median survival of 22 months[148]. This combination is currently being assessed in a phase III trial compared to lenvatinib alone (LEAP-002 trial, NCT03713593).
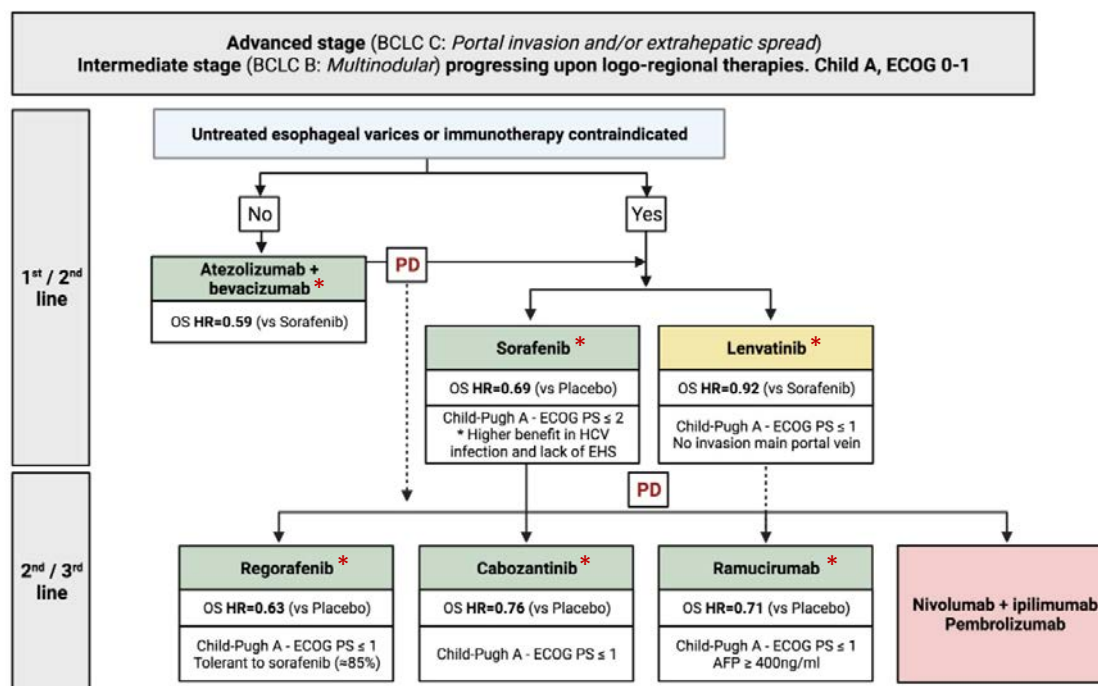
Figure 17. Treatment strategy for advanced HCC. Green boxes designate drugs with positive results from phase III trials with a superiority design. Yellow boxes designate drugs with positive results from phase III trials with a non-inferiority design. Drugs in red boxes have received accelerated approval from the FDA following promising efficacy results in phase II trials. Antiangiogenic agents are indicated with a *. AFP, alpha-fetoprotein; BCLC, Barcelona Clinic Liver Cancer; ECOG PS, Eastern Cooperative Oncology Group performance status; EHS, extrahepatic spread; HCC, hepatocellular carcinoma; HR, Hazard Ratio; PD, Progressive Disease. Adapted from Llovet JM et al., Nat Cancer 2021 (in press)[121].

In the **3rd study** of this doctoral thesis, we explored the **anti-tumoral and immunomodulatory effects of lenvatinib in combination with anti-PD1 ICI**. To achieve this, we generated three syngeneic models of HCC and assessed the anti-tumoral activity of lenvatinib alone or in combination with anti-PD1 and its effects on the systemic and tumor-infiltrating immune cells. In our models, anti-PD1 and lenvatinib in monotherapy improved survival and decreased tumor growth; however, the combination treatment achieved a **higher response rate**, shorter time to response, and a reduction of tumor viability. This is in accordance with a previous publication showing an enhanced anti-tumoral potential of the combination compared to monotherapies *in vivo*[178]. Furthermore, we demonstrated that lenvatinib exerts an **immunomodulatory effect**, which together with anti-PD1, elicits an anti-tumoral immune response activation, thus providing a **mechanistic rationale** for this combination.

The immunomodulatory capacity of lenvatinib alone or in combination with anti-PD1 remains poorly characterized. Notably, the main targets of this TKI include VEGFR and FGFRs. The **VEGFA-**

DISCUSSION

VEGFR pathway activation has been reported to favor tumor growth, progression, and aberrant vasculature formation[130]. More importantly, the VEGF pathway has a direct immunosuppressive effect on the tumor infiltrate by decreasing cytotoxic T cell and DC function and promoting the recruitment of immunosuppressive cells such as T_{reg}, M2 macrophages, and MDSC (**Figure 10**).[130–132]. Additionally, recent studies propose that FGFR4 inhibition by lenvatinib could also promote immune activation in HCC models[179,180]. Our study revealed that **lenvatinib** exerts a potent effect on the immune infiltrate by reducing the **T_{reg}** proportion and altering their intra-tumoral location, which is in accordance with the potential effects of VEGF inhibition (**Figure 10**). Considering this, lenvatinib could boost the effects of ICIs on the antitumor immune response by targeting the VEGFR and FGFR pathways. In addition, our results indicate that lenvatinib treatment reduced the aberrant vasculature in the tumors, in line with the anti-angiogenic nature of this drug. The abnormal tumor vasculature contributes to immunosuppression through several direct and indirect mechanisms[130], and thus vascular normalization by lenvatinib could also contribute to an immunomodulatory capacity of lenvatinib (**Figure 18**).

The role of **anti-PD1** inhibition activating the anti-tumor immune response is well known. In our models, anti-PD1 treatment induced an increase of **T cell** and **type 1 dendritic cell** (DC1) **infiltrate**. This is in accordance with human studies reporting T cell recruitment following PD1 blockade and association between DC1 and T cell function[181,182]. Importantly, the **combined effect** of the two therapeutic approaches induced an increase in the **pro-inflammatory component** in the tumor as shown by the high CD8 to T_{reg} ratio, associated with enhanced antitumor immunity[183]. Data from treated cancer patients indicate that intra-tumoral T_{reg} cells might limit anti-PD1 efficacy[184,185]. Therefore, the effect of lenvatinib in reducing this immunosuppressive population could be responsible for the greater anti-tumoral capacity of the combination therapy.

Other recent experimental studies have suggested a link between the lenvatinib plus anti-PD1 combination and antitumor immune responses[178,179,186]. For instance, the combination treatment increased the CD8 T cell infiltrate and decreased the **monocyte/macrophage** and **DC component** in an HCC murine model[178], in line with previous studies assessing the effect of lenvatinib as monotherapy[186]. Therefore, while anti-PD1 promotes DC1 recruitment according to our study, this may not be the case for all DC subpopulations. Furthermore, our data indicate that the combination elicited a reduction of the protumorigenic **M2 phenotype** compared to the anti-PD1 arm. Considering that T_{reg} cells promote the M2 phenotype in tumor-infiltrated

macrophages[37], this could be linked to the decreased T_{reg} proportion in the combination arm. The reprogramming of the macrophage phenotype agrees with *in vivo* data linking lenvatinib treatment with a reduction in M2 macrophages in colon cancer[186]. Considering the complexity of the myeloid component and different functional states of each sub-population[128], the effect of the treatments on additional myeloid subtypes remains to be investigated. Finally, despite VEGF signaling has been proposed to elicit systemic immunosuppression[130], no significant alterations were detected in the **circulating immune cells**, suggesting that the effect of the treatments may be tumor-specific.
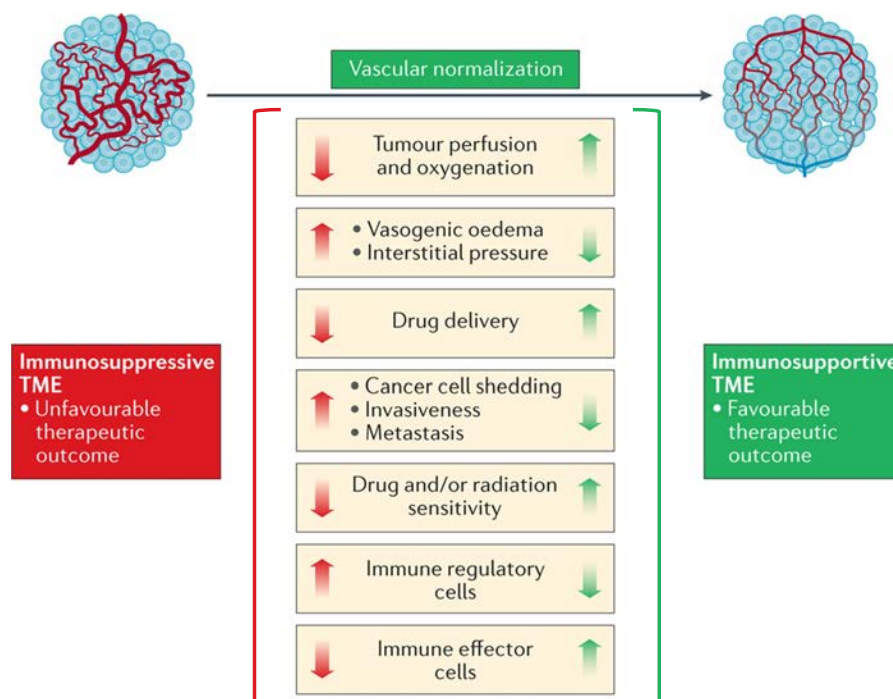
Figure 18. Vascular-normalizing therapies can reprogram the immunosuppressive tumor microenvironment. The structural and functional abnormalities of tumor blood vessels lead to impaired blood flow thus resulting in a hypoxic tumor microenvironment (TME). Hypoxic conditions in tumors exert numerous immune-suppressive effects and limit the delivery and effectiveness of therapies. TME, tumor microenvironment. Modified from Fukumura D et al., Nat Rev Clin Oncol 2018[130].

In addition to the described changes in the tumoral infiltrate, the treatments also elicited changes in the **immune-related molecular signaling**. Notably, lenvatinib blocked immunosuppressive signaling through **TGFß pathway** inhibition. Considering that TGFß signaling is one of its key immunosuppressive mechanisms of T_{reg} cells[187], this is likely a consequence of the decrease in T_{reg} proportion in this treatment arm. Conversely, despite anti-PD1 enhanced the intratumoral T cell infiltrate, this presented an **immune exhausted** phenotype characterized

by expression of immune-suppressive signaling[187], in accordance with human studies reporting an increase in the exhausted T cell component following PD1 blockade[181]. The combined effect of lenvatinib plus anti-PD1 generated both an increase in the T cell infiltrate and a suppression of immunosuppressive signaling (i.e., T_{reg} infiltrate and TGFß pathway). Consequently, only the **combination treatment** generated a specific **activated anti-tumoral immune response**.

The identification of the molecular and immunomodulatory effects induced by lenvatinib plus anti-PD1 can provide relevant information to identify which patients are likely to benefit from this treatment based on their immunological profile. Using gene expression data from our murine model, we generated a **molecular signature** capable of identifying 22% of human HCC patients with downregulation of genes associated with the molecular effect of the combination but not by monotherapies as stand-alone therapies (**Figure 19**). Tumors from these patients presented reduced pro-inflammatory signaling, high T_{reg} levels, and VEGF signaling. We hypothesize that patients with these characteristics are not likely to respond to anti-PD1 alone but could be **responders to the combination treatment**. On the other hand, a previous study from our group identified 24% of HCC patients belonging to the HCC immune class, which presented markers of T cell infiltrate and molecular features indicative of good response to ICI monotherapy[120]. Considering both groups altogether, we hypothesize that about half of HCC patients could respond to combination therapies. These numbers align with the ORR of 46% observed in patients receiving lenvatinib plus anti-PD1 in clinical trials[148]. Notably, our gene signature has been generated from on-treatment murine tumor samples, and therefore further studies extrapolating these findings to pretreatment HCC profiles will be required. The predictive capacity of response of the signature will need to be validated in a prospective human cohort of HCC patients receiving lenvatinib plus an anti-PD1 ICI.

In conclusion, the current study revealed that lenvatinib exerts an immunomodulatory effect on the tumor infiltrate associated with a reduction in the intratumoral T_{reg} infiltrate and inhibition of immune-suppressive pathways. Its combination with anti-PD1 favored the generation of an activated immune profile and a faster response to treatment. Furthermore, we generated a molecular signature present in ~20% HCC patients that correlates with high T_{reg} infiltrate, VEGFR pathway activation, and low inflammatory signaling, and might recognize patients likely to benefit most from this combination. Further investigations are warranted to confirm if the signature can be a tool to identify HCC patients who may respond to the combination therapy beyond their responses to single-agent therapies.
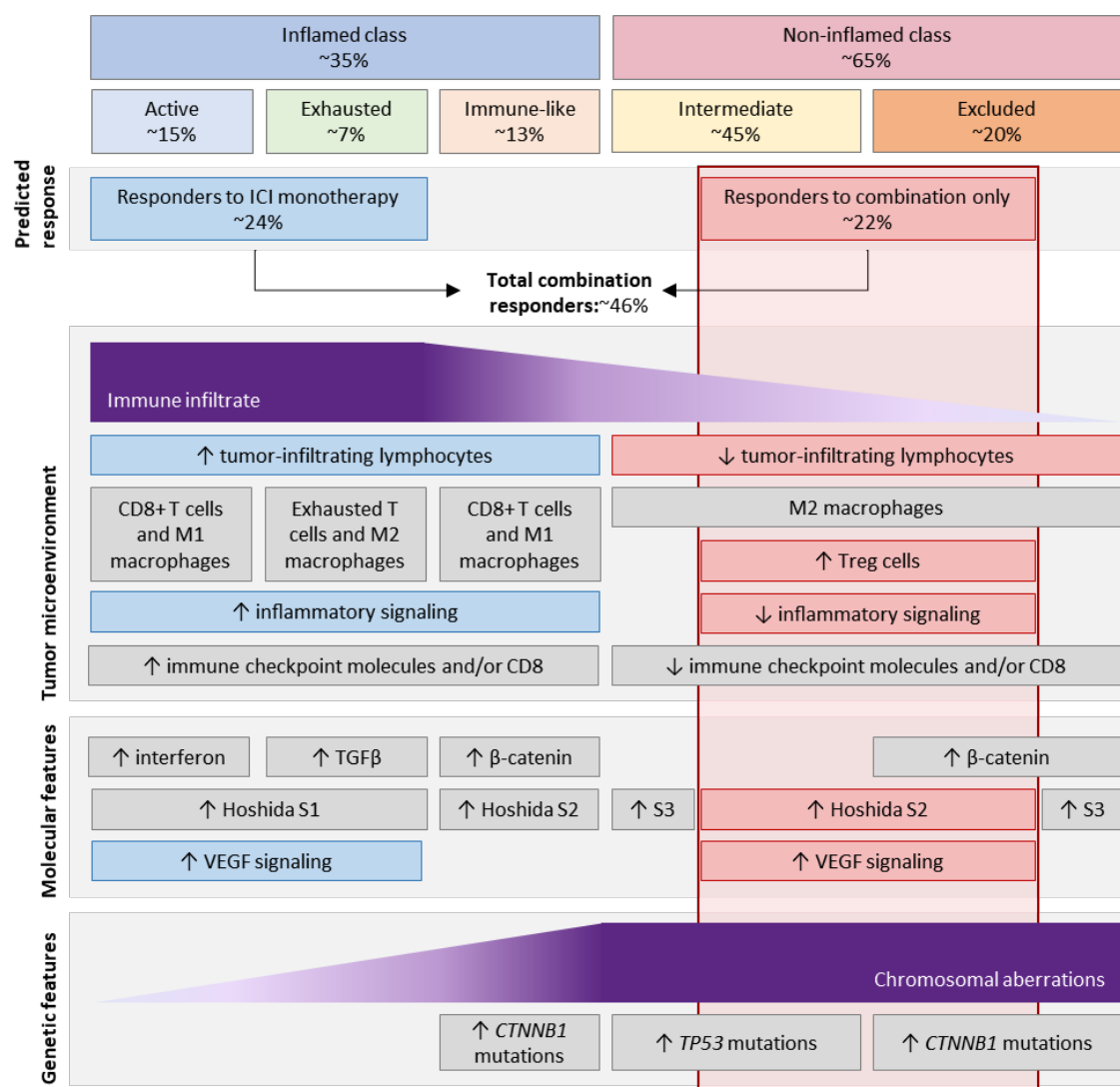
Figure 19. HCC classification according to its immunological features and potential response to the combination therapy. Diagram summarizing the HCC immune classification and potential response to ICIs or combination treatment beyond single agents. Features observed in Study #3 are highlighted in blue and red. Additional features of each subgroup are depicted in grey ICI, Immune checkpoint inhibitors; T_{reg}, regulatory T cell. Original figure. Data extracted from Torrens L *et al.,* Hepatology 2021[188] and Llovet JM et al., Nat Cancer 2021[121].

4. Improving the Understanding of HCC Pathogenesis with Translational Approaches

Using translational approaches based on multi-omic analysis and preclinical models of HCC, the three studies comprised in this thesis provide relevant information in several areas and tackle **unmet needs of HCC** (**Figure 20**). First of all, this thesis investigates the **molecular heterogeneity**

between HCC patients from two geographical regions with different HCC burdens and risk factors. Specifically, distinct virological, genomic, and transcriptomic features of Mongolian HCC patients compared to Western were revealed. Furthermore, our studies elucidate potential **risk factors of HCC** in specific populations. We uncovered mutational fingerprints in the genome of Mongolian HCC constituting a novel mutational signature (SBS Mongolia) associated with the signature of exposure to carcinogenic DMS. This suggests exposure to DMS from coal combustion in this population, which could contribute to the increased HCC incidence. In addition, we provide evidence suggesting a high risk of AAV2 integration in patients with fatty liver disease, thus promoting HCC development. Finally, we propose **novel therapeutic approaches** for HCC patients. In this regard, our data shows enhanced anti-tumoral and immune-modulatory effects of the combination of lenvatinib plus anti-PD1 compared to monotherapies. In addition, the identification of targetable drivers in Mongolia such as *TSC2* suggests that targeted therapeutic approaches could benefit a subgroup of patients.

The abovementioned findings could have implications in **advancing the scientific knowledge and clinical management** in HCC (**Figure 20**). Specifically: 1) The improved understanding of clinical, virological, and molecular pathogenesis of HCC in Mongolia provides novel data regarding the heterogeneous distribution of HCC burden in the world. 2) The identification of potential risk factors of HCC could have an impact on the management of the disease. For instance, our data suggest a need to prevent exposure to risk factors in Mongolia, including viral infections and potentially coal combustion. In addition, our second study raises concerns for the use of gene therapy in patients with NAFLD and chronic liver inflammation. 3) We provide a mechanistic rationale for the use of lenvatinib plus pembrolizumab treatment in advanced HCC, which is currently being assessed in phase III trials[148], and propose a subgroup of HCC patients which could benefit from this approach. Furthermore, the understanding of the molecular profile of Mongolian HCC could also have implications for the treatment of these patients in the future.
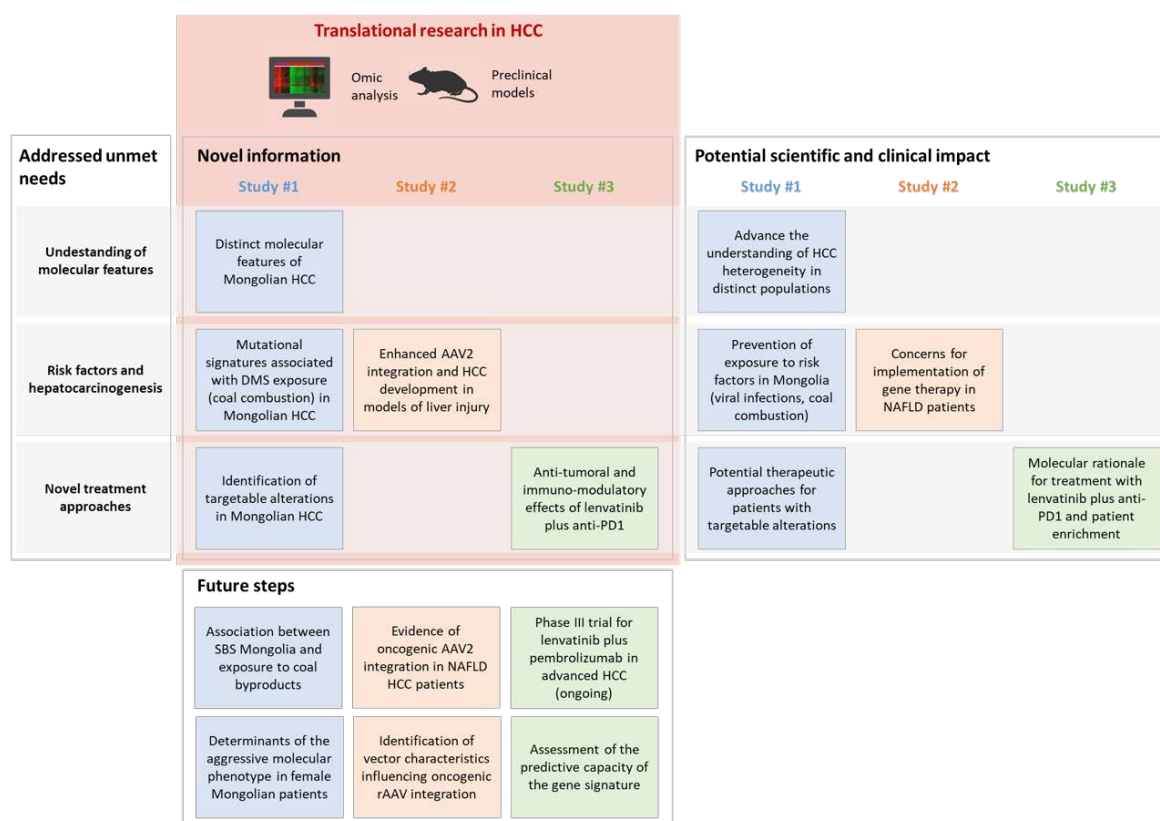
Figure 20. Summary of the studies comprised in this doctoral thesis. This doctoral thesis aims at providing relevant information to advance current unmet needs in hepatocellular carcinoma (HCC). With the use of translational approaches, the three studies herein discussed have potential and clinical impact in the field. AAV2, adeno-associated virus type 2; rAAV, recombinant adeno-associated virus; NAFLD, nonalcoholic fatty liver disease; DMS, dimethyl sulfate. Original figure.

Overall, translational studies such as the ones included in the current thesis are key to advancing our knowledge of this devastating disease. As discussed in prior sections of this discussion and summarized in **Figure 20**, future research will be required to validate the results herein presented and translate this knowledge into actual **clinical applications** that result in survival benefits for the patients. A joint multidisciplinary effort involving basic, translational, and clinical research will be necessary to keep advancing towards this common goal.

CONCLUSIONS

The main conclusions arising from the work presented in this thesis are the following:

- HCC in Mongolia presents **unique genomic and transcriptomic footprints** consisting in an increased number of mutations, as well as specific mutational and transcriptomic patterns. This includes the presence of a newly identified mutational signature (**SBS Mongolia**) in 25% of Mongolian HCC cases, which is associated with a signature of genotoxic DMS exposure. These molecular features suggest a role of **environmental factors** that might explain the high HCC burden in this country.

- **NAFLD**-related chronic inflammation and liver injury promoted the development of **rAAV-induced HCC** in mice due to integration in an oncogenic locus. This raises concerns about the risks of rAAV **gene therapy** causing HCC in patients with chronic liver diseases.

- The combination of **lenvatinib plus anti-PD1** in HCC elicits **immune-modulatory** capacities characterized by an **activated immune profile**. Our gene signature captured ~20% of human HCC which might benefit from the combination treatment. This could have implications for **patient selection** to ultimately maximize the clinical benefit from this treatment.

- Novel **molecular features** of HCC, potential **risk factors**, and **therapeutic approaches** can be unraveled through translational studies based on preclinical models and omics analysis of human tumors. This approach has the potential to impact the clinical management in HCC.

REFERENCES

1.   GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388:1459–544.

2.   Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.

3.   Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Prim. 2021;7:7.

4.   Nault J, Datta S, Imbeaud S, Franconi A, Mallet M, Couchy G, et al. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. Nat Genet. 2015;47:1187–93.

5.   Finn RS, Ryoo B-Y, Merle P, Kudo M, Bouattour M, Lim H-Y, et al. Results of KEYNOTE-240: phase 3 study of pembrolizumab (Pembro) vs best supportive care (BSC) for second line therapy in advanced hepatocellular carcinoma (HCC). J Clin Oncol. 2019;37:Abstact 4004 (Suppl 15).

6.   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. 2017;389:2492–502.

7.   Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. Cancer. 2021;127:3029–30.

8.   Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

9.   Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144:646–74.

10.   Vogelstein B, Kinzler K. Cancer genes and the pathways they control. Nat Med. 2004;10:789–99.

11.   Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. Cancer Cell. 2004;6:433–8.

12.   Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., et al. Cancer Genome Landscapes. Science. 2013;339:1546–58.

13.   Vogelstein B, Kinzler KW. The Path to Cancer — Three Strikes and You're Out. N Engl J Med. 2015;373:1895–8.

14.   Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20:555–72.

REFERENCES

15.   Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458:719–24.

16.   Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45:1134–40.

17.   Talbot S, Crawford D. Viruses and tumours--an update. Eur J Cancer. 2004;40:1998–2005.

18.   Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28:1057–68.

19.   Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15:585–98.

20.   Brown J, Thornton J. Percivall Pott (1714-1788) and chimney sweepers' cancer of the scrotum. Br J Ind Med. 1957;14:68–70.

21.   Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019;177:821-836.e16.

22.   Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJRJR, Behjati S, Biankin A V., et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

23.   Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

24.   Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering Signatures of Mutational Processes Operative in Human Cancer. Cell Rep. 2013;3:246.

25.   Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. . Nat Rev Cancer. 2021;21:619–37.

26.   International Agency for Research on Cancer. Agents Classified by the IARC Monographs, Volumes 1–127 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans [Internet]. 2018 [cited 2020 Nov 5]. Available from: https://monographs.iarc.fr/agents-classified-by-the-iarc

27.   Tovar V, Llovet JM. Molecular Basis of Cancer. In: Llovet JM, editor. Handbook of Translational Medicine. 2016. p. 79–101.

28.   Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. Cell Mol Immunol. 2020;17:807–21.

29.   Greten FR, Grivennikov SI. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. Immunity. 2019;51:27–41.

30.   Grivennikov SI, Greten FR, Karin M. Immunity, Inflammation, and Cancer. Cell. 2010;140:883–99.

31.   Sánchez-Serrano I. Success in translational research: lessons from the development of bortezomib. Nat Rev Drug Discov. 2006;5:107–14.

32. Woolf SH. The Meaning of Translational Research and Why It Matters. JAMA. 2008;299:211–3.

33. Torrens L, Pinyol R, Jimenez W, Llovet J. Overview of Translational Medicine. In: Llovet J, editor. Handbook of Translational Medicine. 2016. p. 21–30.

34. McPake M. ESMO: Handbook of Principles of Translational Research. Vol. 19, European journal of cancer care. 2010. e6 p.

35. Gharwan H, Groninger H. Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. Nat Rev Clin Oncol. 2016;13:209–27.

36. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Gemma A, et al. Gefitinib or Chemotherapy for Non-Small-Cell Lung Cancer with Mutated EGFR. N Engl J Med. 2010;362: 2380–8.

37. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat Rev Cancer. 2019;19:133–50.

38. Martin KJ, Fournier M V., Reddy GPV, Pardee AB. A Need for Basic Research on Fluid-Based Early Detection Biomarkers. Cancer Res. 2010;70:5203–6.

39. Kurnit KC, Dumbrava EEI, Litzenburger B, Khotskaya YB, Johnson AM, Yap TA, et al. Precision Oncology Decision Support: Current Approaches and Strategies for the Future. Clin Cancer Res. 2018;24:2719–31.

40. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA Cancer J Clin. 2016;66:75–88.

41. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010;11:685–96.

42. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.

43. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135–45.

44. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet. 2007;39:S16-21.

45. Donoghue MTA, Schram AM, Hyman DM, Taylor BS. Discovery through clinical sequencing in oncology. Nat Cancer. 2020;1:774–83.

46. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2017;19:93–109.

47. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.

48. Supplitt S, Karpinski P, Sasiadek M, Laczmanska I. Current achievements and applications of transcriptomics in personalized cancer medicine. Int J Mol Sci. 2021; 22:1422.

REFERENCES

49. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160:48–61.

50. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78.

51. Soucek L, Torrens L, Pujades C, Clària J. Experimental Models. In: Llovet J, editor. Handbook of Translational Medicine. 2016. p. 118–26.

52. Brown ZJ, Heinrich B, Greten TF. Mouse models of hepatocellular carcinoma: an overview and highlights for immunotherapy research. Nat Rev Gastroenterol Hepatol. 2018;15:536–54.

53. Frese K, Tuveson D. Maximizing mouse cancer models. Nat Rev Cancer. 2007;7:645–58.

54. McGlynn K, Petrick J, London W. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. Clin Liver Dis. 2015;19:223–38.

55. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol. 2019;16:589–604.

56. Blach S, Zeuzem S, Manns M, Altraif I, Duberg AS, Muljono DH, et al. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. Lancet Gastroenterol Hepatol. 2017;2:161–76.

57. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. Nat Rev Gastroenterol Hepatol. 2019;16:411–28.

58. Wedemeyer H, Manns MP. Epidemiology, pathogenesis and management of hepatitis D: Update and challenges ahead. Nat Rev Gastroenterol Hepatol. 2010;7:31–40.

59. Puigvehí M, Moctezuma-Velázquez C, Villanueva A, Llovet JM. The oncogenic role of hepatitis delta virus in hepatocellular carcinoma. JHEP Reports. 2019;1:120–30.

60. La Bella T, Imbeaud S, Peneau C, Mami I, Datta S, Bayard Q, et al. Adeno-associated virus in the liver: natural history and consequences in tumour development. Gut. 2020;69:737–47.

61. European Association For The Study Of The Liver, European Organisation For Research And Treatment Of Cancer. EASL–EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol. 2012;56:908–43.

62. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Cancer J Clin. 2018;68:394–424.

63. Chimed T, Sandagdorj T, Znaor A, Laversanne M, Tseveen B, Genden P, et al. Cancer incidence and cancer control in Mongolia: Results from the National Cancer Registry 2008-12. Int J cancer. 2017;140:302–9.

64. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. Lancet Gastroenterol Hepatol. 2018;3:383–403.

65. Demaio AR, Dugee O, de Courten M, Bygbjerg IC, Enkhtuya P, Meyrowitsch DW. Exploring knowledge, attitudes, and practices related to alcohol in Mongolia: a national population-based survey. BMC Public Health. 2013;13:178.

66. Dondog B, Lise M, Dondov O, Baldandorj B, Franceschi S. Hepatitis B and C virus infections in hepatocellular carcinoma and cirrhosis in Mongolia. Eur J Cancer Prev. 2011;20:33–9.

67. Oyunsuren T, Kurbanov F, Tanaka Y, Elkady A, Sanduijav R, Khajidsuren O, et al. High frequency of hepatocellular carcinoma in Mongolia; association with mono-, or co-infection with hepatitis C, B, and delta viruses. J Med Virol. 2006;78:1688–95.

68. Tsatsralt-Od B, Takahashi M, Nishizawa T, Endo K, Inoue J, Okamoto H. High prevalence of dual or triple infection of hepatitis B, C, and delta viruses among patients with chronic liver disease in Mongolia. J Med Virol. 2005;77:491–9.

69. Jazag A, Puntsagdulam N, Chinburen J. Status Quo of Chronic Liver Diseases, Including Hepatocellular Carcinoma, in Mongolia. Korean J Intern Med. 2012;27:121.

70. Davaalkham D, Ojima T, Uehara R, Watanabe M, Oki I, Wiersma S, et al. Impact of the universal hepatitis B immunization program in Mongolia: achievements and challenges. J Epidemiol. 2007;17:69–75.

71. Kim YA, Estevez J, Le A, Israelski D, Baatarkhuu O, Sarantuya T, et al. Screening and management of viral hepatitis and hepatocellular carcinoma in Mongolia: results from a survey of Mongolian physicians from all major provinces of Mongolia. BMJ Open Gastroenterol. 2016;3:e000119.

72. Bai H, Guo X, Zhang D, Narisu N, Bu J, Jirimutu J, et al. The Genome of a Mongolian Individual Reveals the Genetic Imprints of Mongolians on Modern Human Populations. Genome Biol Evol. 2014;6:3122–36.

73. Bai H, Guo X, Narisu N, Lan T, Wu Q, Xing Y, et al. Whole-genome sequencing of 175 Mongolians uncovers population-specific genetic architecture and gene flow throughout North and East Asia. Nat Genet. 2018;50:1696–704.

74. Yin J, Zhang H, He Y, Xie J, Liu S, Chang W, et al. Distribution and hepatocellular carcinoma-related viral properties of hepatitis B virus genotypes in Mainland China: a community-based study. Cancer Epidemiol Biomarkers Prev. 2010;19:777–86.

REFERENCES

75. Chen JG, Zhang SW. Liver cancer epidemic in China: past, present and future. Semin Cancer Biol. 2011;21:59–69.

76. Petrick JL, Braunlin M, Laversanne M, Valery PC, Bray F, McGlynn KA. International trends in liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J cancer. 2016;139:1534–45.

77. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology. 2015;149:1226–39.

78. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Prim. 2016;2:16018.

79. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Vol. 15, Nature Reviews Clinical Oncology. 2018. p. 599–616.

80. Villanueva A. Hepatocellular Carcinoma. Longo DL, editor. N Engl J Med. 2019;380:1450–62.

81. Nault J, Couchy G, Balabaud C, Morcrette G, Caruso S, Blanc J, et al. Molecular Classification of Hepatocellular Adenoma Associates With Risk Factors, Bleeding, and Malignant Transformation. Gastroenterology. 2017;152:880-894.e6.

82. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327-1341.e23.

83. Schulze K, Imbeaud S, Letouzé E, Alexandrov LBL, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet. 2015;47:505–11.

84. 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;46:1267–73.

85. Ahn S-M, Jang SJ, Shim JH, Kim D, Hong S-M, Sung CO, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology. 2014;60:1972–82.

86. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res. 2013;23:1422–33.

87. Bassaganyas L, Pinyol R, Esteban-Fabro R, Torrens L, Torrecilla S, Willoughby CE, et al. Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. Clin Cancer Res. 2020;26:6350–61.

88. Fernandez-Banet J, Lee NP, Chan KT, Gao H, Liu X, Sung WK, et al. Decoding complex patterns of genomic rearrangement in hepatocellular carcinoma. Genomics.

2014;103:189–203.

89.   Llovet J, Pinyol R, Kelley R, El-Khoueiry A, Reeves H, Wang X, et al. Molecular pathogenesis and systemic therapies for hepatocellular carcinoma. Nat Cancer. 2021;In press.

90.   Bayard Q, Meunier L, Peneau C, Renault V, Shinde J, Nault J, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. Nat Commun. 2018;9:5235.

91.   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.

92.   Nault JC, Calderaro J, Tommaso L Di, Balabaud C, Zafrani ES, Bioulac-Sage P, et al. TERT promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. Hepatology. 2014;60:6589–604.

93.   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;44:765–9.

94.   Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC. Mutational hot spot in the p53 gene in human hepatocellular carcinomas. Nature. 1991;350:427–8.

95.   Bressac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature. 1991;350:429–31.

96.   Wang K, Lim HY, Shi S, Lee J, Deng S, Xie T, et al. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. Hepatology. 2013;58:706–17.

97.   Sawey ET, Chanrion M, Cai C, Wu G, Zhang J, Zender L, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell. 2011;19:347–58.

98.   Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad I Ben, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet. 2012;44:694–8.

99.   Sporn MB, Liby KT. NRF2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer. 2012;12:564–71.

100.   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;68:6779–88.

101.   Martinez-Quetglas I, Pinyol R, Dauch D, Torrecilla S, Tovar V, Moeini A, et al. IGF2 Is Up-regulated by Epigenetic Mechanisms in Hepatocellular Carcinomas and Is an Actionable

Oncogene Product in Experimental Models. Gastroenterology. 2016;151:1192–205.

102. Paterlini-Bréchot 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;22:3911–6.

103. Schulze K, Nault JC, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol. 2016;65:1031–42.

104. Davé U, Cornetta K. AAV Joins the Rank of Genotoxic Vectors. Mol Ther. 2021;29:418–9.

105. Wang D, Tai P, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nat Rev Drug Discov. 2019;18:358–78.

106. Chandler R, Sands M, Venditti C. Recombinant Adeno-Associated Viral Integration and Genotoxicity: Insights from Animal Models. Hum Gene Ther. 2017;28:314–22.

107. Donsante A, Vogler C, Muzyczka N, Crawford J, Barker J, Flotte T, et al. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther. 2001;8:1343–6.

108. 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;140:1618-28.e16.

109. Luk J, Burchard J, Zhang C, Liu A, Wong K, Shek F, et al. DLK1-DIO3 genomic imprinted microRNA cluster at 14q32.2 defines a stemlike subtype of hepatocellular carcinoma associated with poor survival. J Biol Chem. 2011;286:30706–13.

110. Trobridge G, Hirata RRK, Russell DWD. Gene Targeting by Adeno-Associated Virus Vectors Is Cell-Cycle Dependent. Hum Gene Ther. 2005;16:522–6.

111. Cast A, Kumbaji M, D'Souza A, Rodriguez K, Gupta A, Karns R, et al. Liver Proliferation Is an Essential Driver of Fibrosis in Mouse Models of Nonalcoholic Fatty Liver Disease. Hepatol Commun. 2019;3:1036–49.

112. Letouzé E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8:1315.

113. Ng A, Poon S, Huang S, Lim J, Boot A, Yu W, et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. Sci Transl Med. 2017;9:eaan6446.

114. Rebouissou S, Nault J. Advances in molecular classification and precision oncology in hepatocellular carcinoma. J Hepatol. 2020;72:215–29.

115. Villanueva A, Chiang DY, Newell P, Peix J, Thung S, Alsinet C, et al. Pivotal role of mTOR signaling in hepatocellular carcinoma. Gastroenterology. 2008;135:1972–83.

174

116. Kaposi-Novak P, Lee J-S, Gòmez-Quiroz L, Coulouarn C, Factor VM, Thorgeirsson SS. Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. J Clin Invest. 2006;116:1582–95.

117. Newell P, Toffanin S, Villanueva A, Chiang DY, Minguez B, Cabellos L, et al. Ras pathway activation in hepatocellular carcinoma and anti-tumoral effect of combined sorafenib and rapamycin in vivo. J Hepatol. 2009;51:725–33.

118. Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. Gastroenterology. 2011;140:1501-1512.e2.

119. Lachenmayer A, Alsinet C, Savic R, Cabellos L, Toffanin S, Hoshida Y, et al. Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. Clin Cancer Res. 2012;18:4997–5007.

120. 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;153:812–26.

121. Llovet JMM, Castet F, Heikenwalder M, Maini MKK, Mazzaferro V, Pinato DJJ, et al. Immunotherapies for hepatocellular carcinoma. Nat Rev Clin Oncol. 2021.

122. Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, Lindblad KE, Maier B, Sia D, et al. β-Catenin Activation Promotes Immune Escape and Resistance to Anti–PD-1 Therapy in Hepatocellular Carcinoma. Cancer Discov. 2019;9:1124–41.

123. Sangro B, Sarobe P, Hervás-Stubbs S, Melero I. Advances in immunotherapy for hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol 2021 188. 2021;18:525–43.

124. Ringelhan M, Pfister O, O'Connor T, Pikarsky E, Heikenwalder M. The immunology of hepatocellular carcinoma. Nat Immunol. 2018;19:222–32.

125. Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, et al. An Immune Gene Expression Signature Associated With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. Gastroenterology. 2019; 157:1383-1397.e11.

126. Wada Y, Nakashima O, Kutami R, Yamamoto O, Kojiro M. Clinicopathological study on hepatocellular carcinoma with lymphocytic infiltration. Hepatology. 1998;27:407–14.

127. Prieto J, Melero I, Sangro B. Immunological landscape and immunotherapy of hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol. 2015;12:681–700.

128. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nat Rev Cancer. 2020;20:662–80.

129. Lavin Y, Mortha A, Rahman A, Merad M. Regulation of macrophage development and

function in peripheral tissues. Nat Rev Immunol. 2015;15:731–44.

130. Fukumura D, Kloepper J, Amoozgar Z, Duda DG, Jain RK. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. Nat Rev Clin Oncol. 2018;15:325–40.

131. Gavalas NG, Tsiatas M, Tsitsilonis O, Politi E, Ioannou K, Ziogas AC, et al. VEGF directly suppresses activation of T cells from ascites secondary to ovarian cancer via VEGF receptor type 2. Br J Cancer. 2012;107:1869–75.

132. Gabrilovich DI, Chen HL, Girgis KR, Cunningham HT, Meny GM, Nadaf S, et al. Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells. Nat Med. 1996;2:1096–103.

133. Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul JL, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol. 2018;69:182–236.

134. Llovet J, Brú C, Bruix J. Prognosis of Hepatocellular Carcinoma: The BCLC Staging Classification. Semin Liver Dis. 1999;19:329–38.

135. Llovet JM, Villanueva A, Marrero JA, Schwartz M, Meyer T, Galle PR, et al. Trial Design and Endpoints in hepatocellular carcinoma: AASLD Consensus Conference. Hepatology. 2021;73:158–91.

136. Ferguson F, Gray N. Kinase inhibitors: the road ahead. Nat Rev Drug Discov. 2018;17:353–76.

137. Llovet J, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J-FJJ, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008;359:378–90.

138. Cheng A-L, Finn RS, Qin S, Han K-H, Ikeda K, Piscaglia F, et al. Phase III trial of lenvatinib (LEN) vs sorafenib (SOR) in first-line treatment of patients (pts) with unresectable hepatocellular carcinoma (uHCC). J Clin Oncol. 2017;35:Abstract 4001 (Suppl).

139. Matsui J, Yamamoto Y, Funahashi Y, Tsuruoka A, Watanabe T, Wakabayashi T, et al. E7080, a novel inhibitor that targets multiple kinases, has potent antitumor activities against stem cell factor producing human small cell lung cancer H146, based on angiogenesis inhibition. Int J Cancer. 2008;122:664–71.

140. Zhu AX, Kang Y-K, Yen C-J, Finn RS, Galle PR, Llovet JM, et al. Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α-fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2019;20:282–96.

141. Bruix J, Chan SL, Galle PR, Rimassa L, Sangro B. Systemic treatment of hepatocellular carcinoma: An EASL position paper. J Hepatol. 2021;75:960–74.

176

142. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348:124–8.

143. Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim T-Y, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. 2020;382:1894–905.

144. Cheng A, Qin S, Ikeda M, Galle P, Ducreux M, Zhu A, et al. IMbrave150: Efficacy and safety results from a ph III study evaluating atezolizumab (atezo) + bevacizumab (bev) vs sorafenib (Sor) as first treatment (tx) for patients (pts) with unresectable hepatocellular carcinoma (HCC). Ann Oncol. 2019;30 (Sup 9).

145. Yau T, Kang YK, Kim TY, El-Khoueiry AB, Santoro A, Sangro B, et al. Efficacy and Safety of Nivolumab plus Ipilimumab in Patients with Advanced Hepatocellular Carcinoma Previously Treated with Sorafenib: The CheckMate 040 Randomized Clinical Trial. JAMA Oncol. 2020;6: e204564.

146. Zhu AX, Finn RS, Edeline J, Cattan 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;19:940–52.

147. Finn RS, Ryoo BY, Merle P, Kudo M, Bouattour M, Lim HY, et al. Pembrolizumab As Second-Line Therapy in Patients With Advanced Hepatocellular Carcinoma in KEYNOTE-240: A Randomized, Double-Blind, Phase III Trial. J Clin Oncol. 2020;38:193–202.

148. Finn RS, Ikeda M, Zhu AX, Sung MW, Baron AD, Kudo M, et al. Phase Ib Study of Lenvatinib Plus Pembrolizumab in Patients With Unresectable Hepatocellular Carcinoma. J Clin Oncol. 2020;38:JCO.20.00808.

149. Kim RD, Sarker D, Meyer T, Yau T, Macarulla T, Park J-W, et al. First-in-Human Phase I Study of Fisogatinib (BLU-554) Validates Aberrant Fibroblast Growth Factor 19 Signaling as a Driver Event in Hepatocellular Carcinoma. Cancer Discov. 2019;CD-19-0555.

150. Rimassa L, Assenat E, Peck-Radosavljevic M, Pracht M, Zagonel V, Mathurin P, et al. Tivantinib for second-line treatment of MET-high, advanced hepatocellular carcinoma (METIV-HCC): a final analysis of a phase 3, randomised, placebo-controlled study. Lancet Oncol. 2018;19:682–93.

151. Masucci G V., Cesano A, Hawtin R, Janetzki S, Zhang J, Kirsch I, et al. Validation of biomarkers to predict response to immunotherapy in cancer: Volume I — pre-analytical and analytical validation. J Immunother Cancer. 2016;4:76.

152. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch-repair

deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357:409.

153. Candia J, Bayarsaikhan E, Tandon M, Budhu A, Forgues M, Tovuu LO, et al. The genomic landscape of Mongolian hepatocellular carcinoma. Nat Commun. 2020;11:4383.

154. Wong GL-H, Chan HL-Y, Yiu KK-L, Lai JW-Y, Chan VK-K, Cheung KK-C, et al. Meta-analysis: The association of hepatitis B virus genotypes and hepatocellular carcinoma. Aliment Pharmacol Ther. 2013;37:517–26.

155. Sonneveld MJ, Rijckborst V, Zeuzem S, Heathcote EJ, Simon K, Senturk H, et al. Presence of precore and core promoter mutants limits the probability of response to peginterferon in hepatitis B e antigen-positive chronic hepatitis B. Hepatology. 2012;56:67–75.

156. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

157. Arnedo-Pac C, Mularoni L, Muiños F, Gonzalez-Perez A, Lopez-Bigas N, Schwartz R. OncodriveCLUSTL: A sequence-based clustering method to identify cancer drivers. Bioinformatics. 2019;35:4788–90.

158. Wang PR, Xu M, Toffanin S, Li Y, Llovet JM, Russell DW. Induction of hepatocellular carcinoma by in vivo gene targeting. Proc Natl Acad Sci U S A. 2012;109:11264–9.

159. Cotter T, Rinella M. Nonalcoholic Fatty Liver Disease 2020: The State of the Disease. Gastroenterology. 2020;158:1851–64.

160. Baatarkhuu O, Gerelchimeg T, Munkh-Orshikh D, Batsukh B, Sarangua G, Amarsanaa J. Epidemiology, Genotype Distribution, Prognosis, Control, and Management of Viral Hepatitis B, C, D, and Hepatocellular Carcinoma in Mongolia. Euroasian J Hepato-Gastroenterology. 2018;8:57–62.

161. Wei F, Zheng Q, Li M, Wu M. The association between hepatitis B mutants and hepatocellular carcinoma: A meta-analysis. Medicine (Baltimore). 2017;96:e6835.

162. Chen B-F, Liu C-J, Jow G-M, Chen P-J, Kao J-H, Chen D-S. High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. Gastroenterology. 2006;130:1153–68.

163. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. 2017;PO.17.00011.

164. Zhu AX, Kudo M, Assenat E, Cattan S, Kang Y-KK, Lim HY, et al. Effect of Everolimus on Survival in Advanced Hepatocellular Carcinoma After Failure of Sorafenib. JAMA. 2014;312:57–67.

165. Moody S, Senkin S, Islam SMA, Wang J, Nasrollahzadeh D, Cortez Cardoso Penha R, et al. Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. Nat Genet. 2021;53:1553–63.

166. WHO. Air pollution in Mongolia. Bull World Health Organ. 2019;97:79:80.

167. UNICEF. Environment & air pollution | UNICEF Mongolia [Internet]. [cited 2020 Nov 23]. Available from: https://www.unicef.org/mongolia/environment-air-pollution

168. World Bank. Air quality analysis of Ulaanbaatar : improving air quality to reduce health impacts. 2011. 1–128 p.

169. Nik-Zainal S, Kucab JE, Morganella S, Glodzik D, Alexandrov LB, Arlt VM, et al. The genome as a record of environmental exposure. Mutagenesis. 2015;30:763.

170. Zhang W, He H, Zang M, Wu Q, Zhao H, Lu L ling, et al. Genetic Features of Aflatoxin-Associated Hepatocellular Carcinoma. Gastroenterology. 2017;153:249-262.e2.

171. Donsante A, Miller DGG, Li Y, Vogler C, Brunt EMM, Russell DWW, et al. AAV Vector Integration Sites in Mouse Hepatocellular Carcinoma. Science. 2007;317:477.

172. Bell P, Wang L, Lebherz C, Flieder DB, Bove MS, Wu D, et al. No Evidence for Tumorigenesis of AAV Vectors in a Large-Scale Study in Mice. Mol Ther. 2005;12:299–306.

173. Li H, Malani N, Hamilton SRR, Schlachterman A, Bussadori G, Edmonson SEE, et al. Assessing the potential for AAV vector genotoxicity in a murine model. Blood. 2011;117:3311–9.

174. Ma C, Kesarwala AH, Eggert T, Medina-Echeverz J, Kleiner DE, Jin P, et al. NAFLD causes selective CD4+T lymphocyte loss and promotes hepatocarcinogenesis. Nature. 2016;531:253–7.

175. Pinyol R, Torrecilla S, Wang H, Montironi C, Piqué-Gili M, Torres-Martin M, et al. Molecular characterisation of hepatocellular carcinoma in patients with non-alcoholic steatohepatitis. J Hepatol. 2021;75:865–78.

176. Lefebvre P, Staels B. Hepatic sexual dimorphism — implications for non-alcoholic fatty liver disease. Nat Rev Endocrinol. 2021;17:662–70.

177. Khan KA, Kerbel RS. Improving immunotherapy outcomes with anti-angiogenic treatments and vice versa. Nat Rev Clin Oncol. 2018;15:310–24.

178. Kimura T, Kato Y, Ozawa Y, Kodama K, Ito J, Ichikawa K, et al. Immunomodulatory activity of lenvatinib contributes to antitumor activity in the Hepa1-6 hepatocellular carcinoma model. Cancer Sci. 2018;109:3993–4002.

179. Yi C, Chen L, Lin Z, Liu L, Shao W, Zhang R, et al. Lenvatinib Targets FGF Receptor 4 to Enhance Antitumor Immune Response of Anti–Programmed Cell Death-1 in HCC. Hepatology. 2021;74:2544–60.

180. Deng H, Kan A, Lyu N, Mu L, Han Y, Liu L, et al. Dual Vascular Endothelial Growth Factor Receptor and Fibroblast Growth Factor Receptor Inhibition Elicits Antitumor Immunity

and Enhances Programmed Cell Death-1 Checkpoint Blockade in Hepatocellular Carcinoma. Liver cancer. 2020;9:338–57.

181.   Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med. 2019;25:1251–9.

182.   Böttcher JP, Reis e Sousa C. The Role of Type 1 Conventional Dendritic Cells in Cancer Immunity. Trends in Cancer. 2018;4:784–92.

183.   Arce Vargas F, Furness AJS, Litchfield K, Joshi K, Rosenthal R, Ghorani E, et al. Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. Cancer Cell. 2018;33:649-663.

184.   Vignali DAA, Collison LW, Workman CJ. How regulatory T cells work. Nat Rev Immunol. 2008;8:523–32.

185.   Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T Cells and Immune Tolerance. Cell. 2008;133:775–87.

186.   Kato Y, Tabata K, Kimura T, Yachie-Kinoshita A, Ozawa Y, Yamada K, et al. Lenvatinib plus anti-PD-1 antibody combination treatment activates CD8 + T cells through reduction of tumor-associated macrophage and activation of the interferon pathway. PLoS One. 2019;14:e0212513.

187.   Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015;15:486–99.

188.   Torrens L, Montironi C, Puigvehí M, Mesropian A, Leslie J, Haber PK, et al. Immunomodulatory Effects of Lenvatinib Plus Anti–Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma. Hepatology. 2021;74:2652–69.

ANNEX A

1. Publications

List of scientific articles and book chapters written during the period of the doctoral thesis (2016-2021)

Published articles

- **Torrens L**, Montironi C, Puigvehí M, Mesropian A, Leslie J, Haber PK, et al. Immunomodulatory Effects of Lenvatinib Plus Anti–Programmed Cell Death Protein 1 in Mice and Rationale for Patient Enrichment in Hepatocellular Carcinoma. Hepatology. 2021 Jun;74:2652–2669.

- Dalwadi DA, **Torrens L**, Abril-Fornaguera J, Pinyol R, Willoughby C, Posey J, et al. Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice. Mol Ther. 2021 Feb;29(2):680-690.

- Carrillo-Reixach J, **Torrens L**, Simon-Coma M, Royo L, Domingo-Sàbat M, Abril-Fornaguera J, et al. Epigenetic footprint enables molecular risk stratification of hepatoblastoma with clinical implications. J Hepatol. 2020 Aug;73(2):328-341.

- Bassaganyas L, Pinyol R, Esteban-Fabró R, **Torrens L**, Torrecilla S, Willoughby C, et al. Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. Clin Cancer Res. 2020 Dec 26(23)6350-6361.

- Moeini A, Sia D, Zhang Z, Camprecios G, Stueck A, Dong H, Montal R, **Torrens L**, et al. Mixed hepatocellular cholangiocarcinoma tumors: Cholangiolocellular carcinoma is a distinct molecular entity. J Hepatol. 2017 May;66(5):952-961.

Submitted articles

- **Torrens L***, Puigvehí M*, Torres-Martín M, Wang H, Maeda M, Haber PK, et al. Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features Associated with Environmental Agents. Submitted to Proc Natl Acad Sci U S A. 2021. **Contributed equally.*

- **Torrens L***, Abril-Fornaguera J*, Carrillo-Reixach J, Balaseviciute U, Rialdi A, Del Río-Álvarez A, et al. Identification of IGF2 as Genomic Driver and Actionable Therapeutic Target in Hepatoblastoma. Submitted to Cancer Res. 2021. **Contributed equally.*

- Montironi C*, Castet F*, Haber PK*, Pinyol R, Torres-Martin M, **Torrens L**, et al. Inflamed and non-inflamed classes of HCC: a revised immunogenomic classification. Submitted to Gut. 2021. *Contributed equally.

- Esteban-Fabró R*, Willoughby CE*, Piqué-Gili, M, Montironi C, Abril-Fornaguera J, Judit Peix, **Torrens L**, et al. Cabozantinib enhances anti-PD1 activity and elicits a neutrophil-based immune response in hepatocellular carcinoma. Submitted to Clin Cancer Res. 2021. *Contributed equally.

Book chapters

- **Torrens L**, Pinyol R, Jimenez W, Llovet JM (2016). Overview of Translational Medicine. In Llovet JM (Ed.), Handbook of Translational Medicine (pp 21-28). Barcelona, Spain: Edicions de la Universitat de Barcelona. ISBN: 978-84-475-4030-3.

- Soucek L**, Torrens L**, Pujades C, Clària J (2016). Experimental Models. In Llovet JM (Ed.), Handbook of Translational Medicine (pp 118-126). Barcelona, Spain: Edicions de la Universitat de Barcelona. ISBN: 978-84-475-4030-3.

2. Communications to Scientific Meetings

Participation in national and international meetings with poster or oral communications. Presenter name is underlined.

- **Torrens L**, Puigvehí M, Torres-Martín M, Wang H, Maeda M, Haber P, et al. Molecular Characterization of HCC in Mongolia Delineates Unique Genomic Features. EASL International Liver Congress 2021. Poster presentation.

- **Torrens L**, Abril-Fornaguera J, Carrillo J, Balaseviciute U, Rialdi A, Haber P, et al. Identification of IGF2 as Genomic Driver and Actionable Therapeutic Target in Hepatoblastoma. ILCA 16th Annual Conference. 2021 Virtual Conference. Poster presentation.

- **Torrens L**, Puigvehí M, Torres-Martín M, Wang H, Maeda M, Haber P, et al. Molecular Characterization of HCC in Mongolia Delineates Unique Genomic Features. ILCA 16th Annual Conference. 2021 Virtual Conference. Poster presentation.

- **Torrens L**, Abril-Fornaguera J, Carrillo J, Balaseviciute U, Rialdi A, Haber P, et al. Identification of IGF2 as Genomic Driver and Actionable Therapeutic Target in Hepatoblastoma. EASL International Liver Congress 2021. Oral presentation.

- Esteban-Fabró R, Willoughby CE, Piqué-Gili M, Montironi C, Abril-Fornaguera J, Peix J, **Torrens L**, et al. Cabozantinib enhances anti-pd1 efficacy and elicits a neutrophil-based immune response in murine models: implications for human HCC. ILCA 16th Annual Conference. 2021 Virtual Conference. Oral presentation.

- **Torrens L**, Abril-Fornaguera J, Carrillo-Reixach J, Balaseviciute U, Rialdi A, Haber P et al. Identificación de IGF2 como diana terapéutica en Hepatoblastoma. 46 Congreso AEEH. 2021. Madrid, Spain. Poster presentation.

- **Torrens L**, Montironi C, Mesropian A, Haber PK, Maeda M, Puigvehí M, et al. Immune-Remodeling Effects of Lenvatinib Plus Anti-PD1 in a Murine Model of Hepatocellular Carcinoma. ILCA 15th Annual Conference. 2020 Virtual Conference. Poster presentation (Best Basic / Translational poster session).

- Esteban-Fabró R, Willoughby CE, Piqué-Gili M, Peix J, Montironi C, Abril-Fornaguera J, **Torrens L**, et al. Cabozantinib enhances the efficacy and immune activity of anti-PD1

therapy in a murine model of hepatocellular carcinoma. ILCA 15th Annual Conference. 2020 Virtual Conference.

- Sia D, Puigvehi M, **Torrens L**, Wang H, Torres-Martin M, Maeda M, et al. Molecular Characterization of Hepatocellular Carcinoma in Mongolia Delineates Unique Genomic Features. AASLD: The Liver Meeting. 2020 Virtual Conference. Poster presentation.

- **Torrens L**, Montironi C, Haber PK, Maeda M, Puigvehi M, Kamphorst A, et al. Efecto de la combinación de lenvatinib y anti-PD1 sobre el sistema inmune en un modelo experimental de carcinoma hepatocelular. 45 Congreso AEEH. 2020. Madrid, Spain. Oral poster presentation.

- Esteban-Fabró R, Willoughby CE, Piqué-Gili M, Peix J, Montironi C, Abril-Fornaguera J, **Torrens L**, et al. Cabozantinib enhances the efficacy and immune modulatory activity of anti-PD1 therapy in a syngeneic mouse model of hepatocellular carcinoma. EASL International Liver Congress 2020. Oral presentation.

- Esteban-Fabró R, Willoughby CE, Piqué-Gili M, Peix J, Montironi C, Abril-Fornaguera J, **Torrens L**, et al. Cabozantinib aumenta la eficacia y actividad inmunomoduladora de la terapia con anti-PD1 en un modelo murino singénico de carcinoma hepatocelular. 45 Congreso AEEH. 2020. Madrid, Spain. Oral communication.

- **Torrens L**, Montironi C, Haber P, Kuchuk O, Akers N, Simon-Coma M, et al. Identification of IGF2 as Genomic Driver and Therapeutic Target in Hepatoblastoma. AASLD: The Liver Meeting 2019. Boston, USA. Poster presentation.

- **Torrens L**, Montironi C, Haber P, Kuchuk O, Akers N, Simon-Coma M, et al. Identification of IGF2 as Genomic Driver and Therapeutic Target in Hepatoblastoma. ILCA 13th Annual Conference. 2019. Chicago, USA. Poster presentation.

3. Grants and Awards

Fellowship grants

- Research Stay Fellowship for International Doctorates. University of Barcelona 2018

Scientific awards

- Top-rated basic-translational poster session. ILCA 2020

- Best Poster Presentation – HUNTER Liver Workshop. Cancer Research UK, Beatson Institute, Glasgow (UK) 2020.

Competitive project grants (co-investigator)

- National Health Institute, Spain. I+D Program (Grant number: PID2019-105378RB-100). "Molecular characterization of obesity / diabetes / NASH-related hepatocellular carcinoma". PI: JM Llovet. 2020 – 2023.

- CRUK, AECC, AIRC, Accelerator Award (C9380/A26813). "HUNTER - Hepatocellular Carcinoma Expediter Network". PI: JM Llovet. 2019 – Present.

- Horizon 2020 – European Commission (Call H2020-PHC-2015, number 667273). Title: "HEP-CAR - Mechanisms underlying hepatocellular carcinoma pathogenesis and impact of co-morbidities". P: JM Llovet. 2016 – 2019.

- National Health Institute, Spain. I+D Program (Grant number: SAF2016-76390-R). "Mechanisms of resistance to TKIs in hepatocellular carcinoma". PI: JM Llovet. Team member. 2016 – 2019.

Investigator initiated sponsored studies (co-investigator)

- Bayer Pharmaceuticals. "Discovery of biomarkers predictors of response and/or resistance to anti-PD1-based immune checkpoint inhibitors in advanced HCC". PI: JM Llovet. 2018 - 2021

- Boehringer-Ingelheim. "Role of Xentuzumab for the treatment of hepatoblastoma overexpressing IGF2". PI: JM Llovet. 2019 – 2020.

3. Grants and Awards

ANNEX A

- Eisai Inc. "Impact of lenvatinib alone or in combination with anti-PD1 on the immune system in HCC and assessment of the synergistic anti-tumoral effect in experimental models". PI: JM Llovet. 2018 - 2019.

- Ipsen. "Impact of cabozantinib alone or in combination with anti-PD1 on the immune system in HCC and assessment of the synergistic anti-tumoral effect in experimental models". PI: JM Llovet. 2018 – 2019.

- Eisai Inc. "Impact of lenvatinib alone or in combination with anti-PD1 on the immune

ANNEX B

Supplementary Data Study 1

SUPPLEMENTARY MATERIALS AND METHODS

Clinical and histological data

All samples in the study were fresh-frozen. Tissue samples were coded previous to storage using consecutive numbering. The code did not include any patient identifier, and the research team at Mount Sinai received already de-identified samples. The diagnosis of HCC was confirmed after a first evaluation made by 3 independently working expert pathologists in Mount Sinai (WQL, CM, and ST), and those samples with >50% necrotic tissue (n=18), tumors other than HCC (n=7), or repeated (n=2) were excluded. Thus, a final number of 192 patients were included for further evaluation.

Baseline clinico-pathological characteristics were collected for both cohorts (Table 1). All histological evaluations were performed by 2 expert pathologists, blinded to clinical data. Fibrosis stage was scored according to the METAVIR Scale [1]. All the above-mentioned variables were also collected for the Western cohort except for BMI, alanine aminotransferase (ALT) values, tumor size and number, region of origin, and presence of steatosis/steatohepatitis. All data were stored in a database containing de-identified information, and electronic files were stored according to Mount Sinai IRB protocols with encryption and password protection.

Viral hepatitis evaluation

The presence of viral infections (HBV, HCV and HDV) was assessed in the non-tumor tissue of all Mongolian samples, and was compared to the data obtained from Mongolian charts (HBV surface antigen -HBsAg- and HCV antibodies; HDV was not routinely evaluated in Mongolia).

Intrahepatic HBV and HDV status were assessed by quantitative PCR (qPCR). HBV-DNA was assessed by Taqman qPCR (ID Pa03453406 s1, ABI, Thermo Fisher) using the ViiA7 Real Time PCR System (ABI) as previously described [2]. The calibration curve was prepared using ten-fold serial dilutions of a plasmid containing an HBV monomer (pHBV-EcoR1). Total HDV-RNA was determined by one-step RT-qPCR as previously reported [3,4]. For absolute quantification, serial dilution of an HDV-RNA standard (WHO 1st International Standard, Paul-Ehrlich-Institut) was included in each assay [5]. All samples positive for HDV were considered HBV/HDV positive.

HCV status was determined by conventional PCR. Specifically, HCV RNA was retrotranscribed to cDNA with EcoDry Premix (Double Primed) (Takara cat# 639549) and HCV-specific sequences were amplified under standard conditions using the following primer pair: Fw CACGCAGAAAGCGTCTAG, HCV; Rv TTGATCCAAGAAAGGACCC [6]. PCR products were run on an agarose gel, purified using PureLink Quick Gel Extraction Kit (Invitrogen cat# K210012) and sequenced by Sanger (Macrogen, USA).

HBV and HDV genotyping

HBV and HDV genotypes were determined by direct sequencing and phylogenetic analysis of a 1100 bp fragment of the HBV retrotranscriptase [7] and a fragment of 370bp encompassing approximately 85% of the large HDV antigen (HDAg) [8], respectively. Multiple alignments were performed with ClustalW [9] and maximum likelihood trees were obtained with MEGA X software [10].

Analysis of HBV mutations

HBV mutations associated with HCC development were assessed by nested PCR (GoTaq Flexi DNA Polymerase - Promega). Specifically, precore region was screened for nucleotide substitution G1896A and the basal core promoter (BCP) region was checked for the presence of 2 nucleotides substitutions (A1762T and G1764A). A DNA segment composing of the BCP, precore, and partial C regions was amplified by nested PCR and analyzed by direct sequencing [11].

ANNEX B

Whole exome sequencing mutational variant calling in in-house cohorts

Mutational variant calling was performed following the Tigris pipeline (v2.0.1). BWA 0.7.17 was used for alignment, followed by base quality score recalibration via BQSR, read deduplication via Picard MarkDuplicates, germline molecular variant (SNV and small indel) calling via HaplotypeCaller, and somatic molecular variant calling via Mutect2, which calls variants using local de novo assembly and then does a two-pass filter using heuristics (further details can be found in the MuTect2 whitepaper from its GitHub repo at https://github.com/broadinstitute/gatk/tree/master/docs/mutect) . After applying these filters in MuTect2, the twice-filtered MuTect2 output was then filtered for 'PASS' variants only with allele frequency >= 5% for downstream analysis. Tigris computes depth-based and other NGS library QC metrics using GATK3 DepthOfCoverage and CallableLoci, as well as Picard. Lastly, somatic copy number variants (sCNV) were called using tumor/normal SAAS-CNV (v0.3.4) workflow that models allele balance to determine balanced versus unbalanced somatic gains and losses, as well as determine somatic copy-neutral loss of heterozygosity [12]. SAAS-CNV output were further processed using GISTIC2.0 for somatic CNV analysis.

Analysis of gene mutations and filtering in previously published cohorts

We used whole exome sequenxing (WES data to assess the mutation profile in the European [13], Korean [14], TCGA [15] and Mongolian NCI [16] HCC cohorts. In the TCGA cohort, only variants with filter PASS were considered and "3_prime_UTR_variant", "5_prime_UTR_variant", "intron_variant"," synonymous_variant" were filtered from the cohort. For the Korean and European cohorts, only were accepted the following types of mutations, filtering the rest of the annotated subtypes: Missense_Mutation, "Nonsense_Mutation", "Splice_Site", "Translation_Start_Site", "Frame_Shift_Ins", "In_Frame_Ins", "Frame_Shift_Del", "In_Frame_Del", "3'Flank", "5'Flank" and "Nonstop_Mutation". For all cohorts, only VAF ≥ 0.05 was accepted for further analysis. Tumor mutational burden (TMB) from these external cohorts was calculated as previously indicated. Other TMB calculation approaches are provided in Supplementary Table 7 for comparison.

Somatic copy number variations (SNVs) analysis

HaplotypeCaller [17] was used to generate germline VCF files as input for SAAS-CNV (v0.3.4) [12], which in turn generated segmentation file as input for GISTIC 2.0 run [18]. The "log2ratio.Median.adj" column from saasCNV output was used for GISTIC 2.0 run, with the following parameter flags -genegistic 1 -smallmem 1 -broad 1 -brlen 0.98 -conf 0.99 -armpeel 0 -savegene 1 -gcm extreme -qvt 0.1 -cap 2.0 -ta 0.85 -td 0.74.

Identification of potential driver genes

OncodriveCLUSTL and dN/dScv algorithms were used to identify genes harboring significantly more mutations than expected by chance [19,20] among the genes significantly more mutated in the Mongolian cohort compared to the Western cohort. Genes predicted to have an enrichment for damaging alterations by OncodriveCLUSTL or dN/dScv were selected (q<0.05). The selected genes were filtered for cancer-related genes according to the OncoKB Cancer Gene List or previously reported studies in HCC [13,21].

TERT promoter mutations detection

The promoter region of TERT in Mongolian samples was amplified by PCR and sequenced using Sanger sequencing as previously described [22]. The number of TERT promoter mutations was compared to the reported percentages in Western cohorts (55-60%) [13].

Identification of de novo mutational signatures in Mongolian tumors

R package MutationalPatterns [23] was used to perform de novo mutational signature extraction. Extracted signatures were mapped against COSMICv3. De novo signatures were mapped to single signatures and linear combinations of two if the cosine similarity was > 0.9. One novel signature "SBS

Mongolia" was revealed with cosine similarity below the threshold for all comparisons (maximum observed cosine similarity of 0.818).

Mutational signature fitting was performed using the quadprog R package [24], using HCC specific COSMICv3 mutational signatures plus SBS Mongolia. To select HCC specific signatures, COSMICv3 signatures were assessed in 493 HCC samples from the Mongolian (n=151), Western (n=112) and TCGA (n=230) cohorts. Signatures occurring in ≥ 40 HCC samples (Supplementary Table 17) or signatures that were revealed via de novo mutational signature extraction and able to be mapped to COSMICv3 reference were selected (i.e., SBS1, SBS4, SBS5, SBS6, SBS12SBS16, SBS18, SBS22, SBS26, SBS29, SBS40).

In order to assess the confidence of signature assignment across our samples in signature fitting, a previously reported bootstrap approach was adopted [25]. At each bootstrap, we randomly selected the same number of mutations with replacement from the original observed mutational profile of a given tumor sample (classified by the 96 trinucleotide mutation types) and performed signature fitting to estimate signature weight (quadprog R package), resulting in a distribution of signature weights for each signature from all bootstraps (N = 500) in a given tumor. Based on the signature weight distribution, for any given sample, we were able to estimate confidence level. At p value = 0.1 (one sided), the 10% quantile of signature weights would mean we were 90% confident that the signature weight was above that 10% quantile value. Finally, samples were considered positive for a mutational signature when the bootstrap exposure cutoff was ≥ 0.1.

Identification of de novo mutational signatures in Mongolian non-tumoral liver samples

Mutational signature analysis was used to assess signatures in the adjacent non-tumoral samples. First, for variant calling in the adjacent non-tumoral liver tissue, we subtracted the mutations in tumors from the mutation in non-tumoral tissue using MuTect2. Next, only samples with total SNV count ≥ 10 (for variants in exome region only at allelic frequency cutoff of 0.05) were selected for subsequent mutational signature analysis, resulting in a total of 78 samples (64 Mongolian cohort plus 14 Western cohort). Due to the small number of unique SNVs in adjacent non-tumoral samples, the analysis was performed on pooled variants from each cohort. The mutational signature fitting analysis was performed using all HCC specific COSMICv3 signatures (Supplementary Table 17) plus SBS Mongolia.

Analysis of environmental signatures

Signature fitting analysis was performed using signatures from the Compendium of Mutational Signatures of Environmental Agent [26]. Specifically, all the 52 signatures included in the Compendium from agents generating significantly different substitution profiles compared to untreated controls were used [26]. The weights of each mutational signature contributing to an individual tumor sample were obtained using the deconstructSigs R package (https://github.com/raerose01/deconstructSigs). The trinucleotide count for each sample was normalized by multiplying it by a ratio of its occurence in the genome to its occurence in the exome (exome2genome method), following recommendations for WES data. Signature contributions with a weight <0.25 were discarded from the analysis. A signature was considered present in an individual tumor sample when the weight threshold was ≥ 0.1. (Supplementary Table 18).

SUPPLEMENTARY REFERENCES

1. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology. 1996;24:289–93.
2. Malmström S, Larsson SB, Hannoun C, et al. Hepatitis B viral DNA decline at loss of HBeAg is mainly explained by reduced cccdna load - down-regulated transcription of PgRNA has limited impact. PLoS One. 2012;7.
3. Ferns RB, Nastouli E, Garson JA. Quantitation of hepatitis delta virus using a single-step internally controlled real-time RT-qPCR and a full-length genomic RNA calibration standard. J Virol Methods. 2012;179:189–94.

4.   Giersch K, Homs M, Volz T, et al. Both interferon alpha and lambda can reduce all intrahepatic HDV infection markers in HBV/HDV infected humanized mice. Sci Rep. 2017;7.

5.   Le Gal F, Brichler S, Sahli R, et al. First international external quality assessment for hepatitis delta virus RNA quantification in plasma. Hepatology. 2016;64:1483–94.

6.   Balart LA, Perrillo R, Roddenberry J, et al. Hepatitis C RNA in liver of chronic hepatitis C patients before and after interferon alfa treatment. Gastroenterology. 1993;104:1472–7.

7.   Tong Y, Liu B, Liu H, et al. New universal primers for genotyping and resistance detection of low HBV DNA levels. Med. 2016;95:e4618.

8.   Le Gal F, Brichler S, Drugan T, et al. Genetic diversity and worldwide distribution of the deltavirus genus: A study of 2,152 clinical strains. Hepatology. 2017;66:1826–41.

9.   Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.

10. Kumar S, Stecher G, Li M, et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.

11. Chen CH, Changchien CS, Lee CM, et al. Combined mutations in pre-S/surface and core promoter/precore regions of hepatitis B virus increase the risk of hepatocellular carcinoma: A case-control study. J Infect Dis. 2008;198:1634–42.

12. Zhang Z, Hao K. SAAS-CNV: A Joint Segmentation Approach on Aggregated and Allele Specific Signals for the Identification of Somatic Copy Number Alterations with Next-Generation Sequencing Data. Wang E, editor. PLOS Comput Biol. 2015;11:e1004618.

13. Schulze K, Imbeaud S, Letouzé E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet. 2015;47:505–11.

14. Ahn S-M, Jang SJ, Shim JH, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology. 2014;60:1972–82.

15. Ally A, Balasundaram M, Carlsen R, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327-1341.e23.

16. Candia J, Bayarsaikhan E, Tandon M, et al. The genomic landscape of Mongolian hepatocellular carcinoma. Nat Commun. 2020;11:4383.

17. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Biorxiv 201178. 2018;

18. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12:R41.

19. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

20. Arnedo-Pac C, Mularoni L, Muiños F, et al. OncodriveCLUSTL: A sequence-based clustering method to identify cancer drivers. Bioinformatics. 2019;35:4788–90.

21. Chakravarty D, Gao J, Phillips S, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. 2017;2017:1–16.

22. Nault JC, Mallet M, Pilati C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat Commun. 2013;4:2218.

23. Blokzijl F, Janssen R, van Boxtel R, et al. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10.

24. Lynch AG. Decomposition of mutational context signatures using quadratic programming methods. F1000Research. 2016;5:1253.

25. Huang X, Wojtowicz D, Przytycka TM. Detecting presence of mutational signatures in cancer with confidence. Bioinformatics. 2018;34:330–7.

26. Kucab JE, Zou X, Morganella S, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019;177:821-836.e16.

SUPPLEMENTARY FIGURES
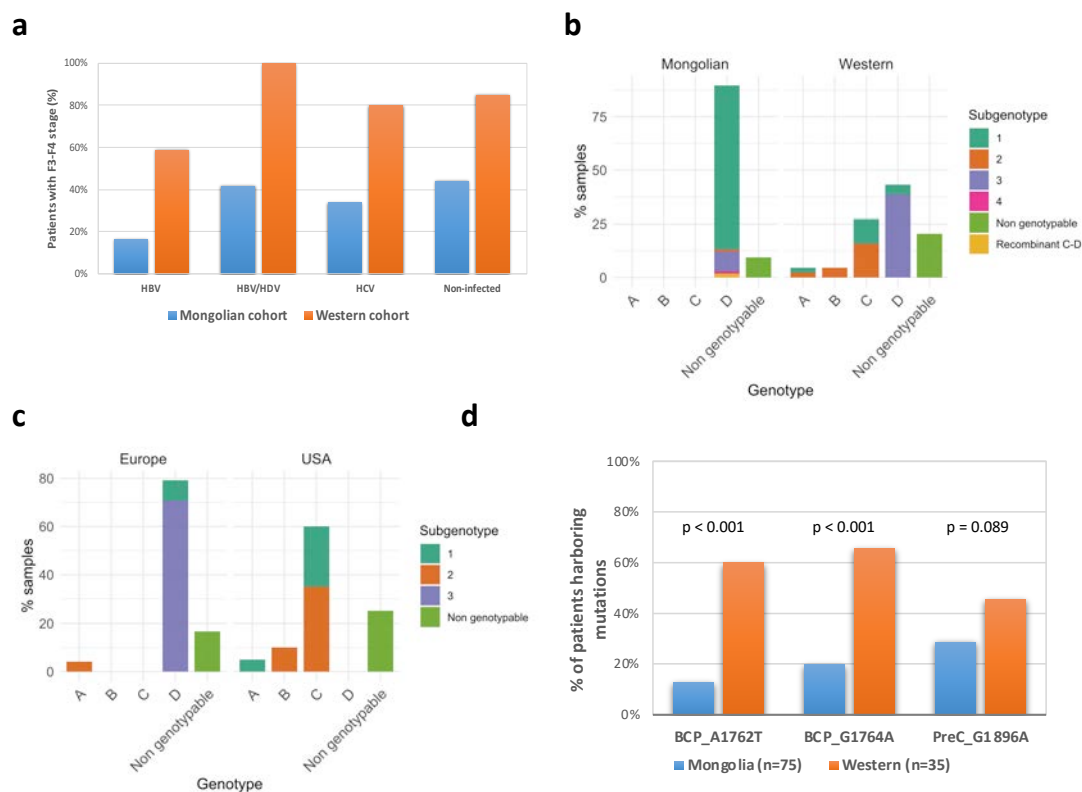
Supplementary Figure 1. **Clinical and viral characterization of the study cohort.** **a** Rate of patients within each fibrosis grade in the Mongolian and Western cohorts. p values show differences in the rate of F3-4 patients in each subgroup (Fisher exact test). **b-c** Distribution of HBV genotypes **(b)** in the Mongolian (n = 106) and Western (n = 44) cohorts, and **(c)** in the Western cohort divided by USA (n = 20) and Europe (n = 24). Sub-genotypes are indicated in color. **d** Rate of HBV basal core promoter and pre-core mutations in the Mongolian and Western cohorts.

a

HBV DNA load in Mongolian samples
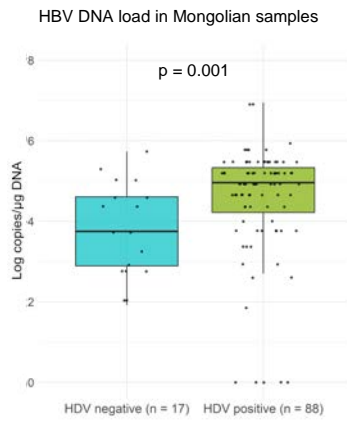
b
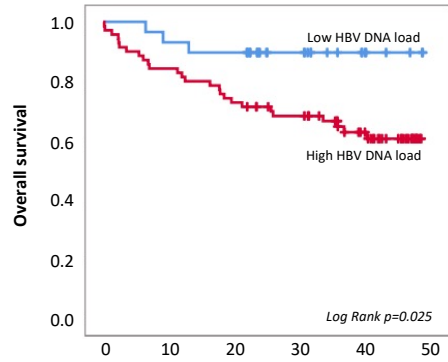
Patients at risk	Follow-up (months)					
Low HBV-DNA load	29	27	26	15	4	0
High HBV-DNA load	70	59	51	44	29	0

c

HBV DNA load in HBV positive samples
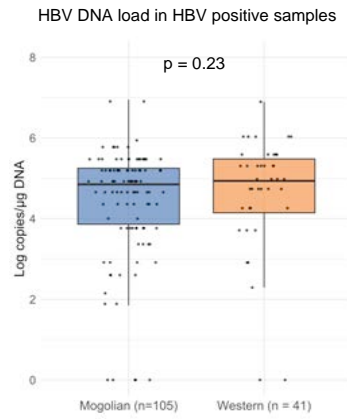
Supplementary Figure 2. HBV DNA load in liver samples. a HBV-DNA levels in liver samples from HDV positive and negative patients in the Mongolian cohort (log copies/ug total DNA). **b** Comparison of survival according to levels of HBV-DNA in the Mongolian cohort. The thresholds of HBV-DNA are established according to the percentile 25 to define low and high HBV-DNA load (log copies/µg total DNA). **c** HBV DNA levels in (**c**) all HBV-positive samples from the Mongolian and Western cohorts. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers).
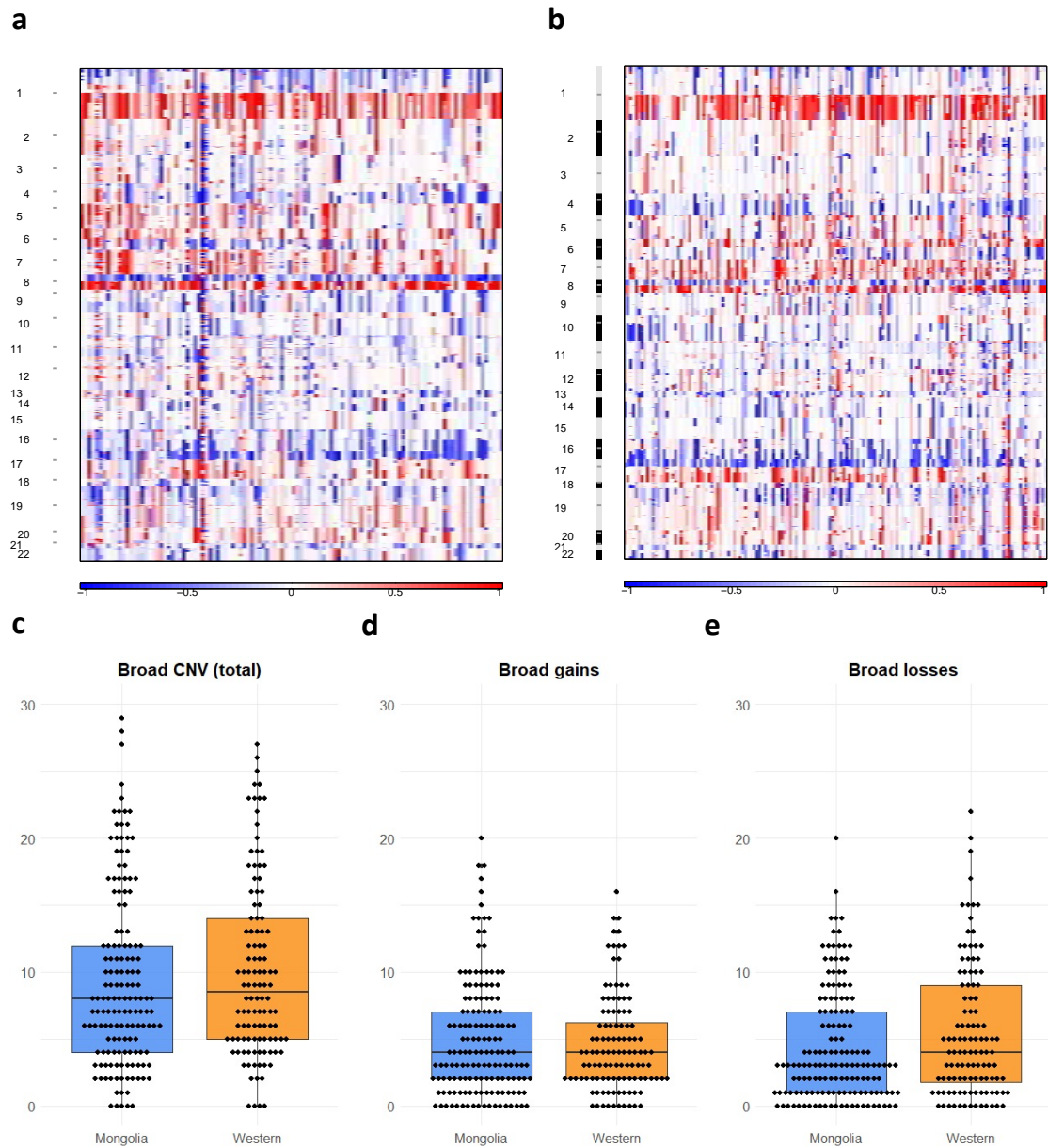
Supplementary Figure 3. Broad copy number variaton profiles in Mongolian and Western HCCs. a Frequency of broad CNV along the genome in the Mongolian and **b** Western cohorts. The right panel indicates chromosomal regions and the bottom axis indicates the frequency of broad changes (gains and losses). **c-d** Total broad CNV (**c**), broad gains (**d**), and broad losses (**e**) in Mongolian and Western HCC. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers).
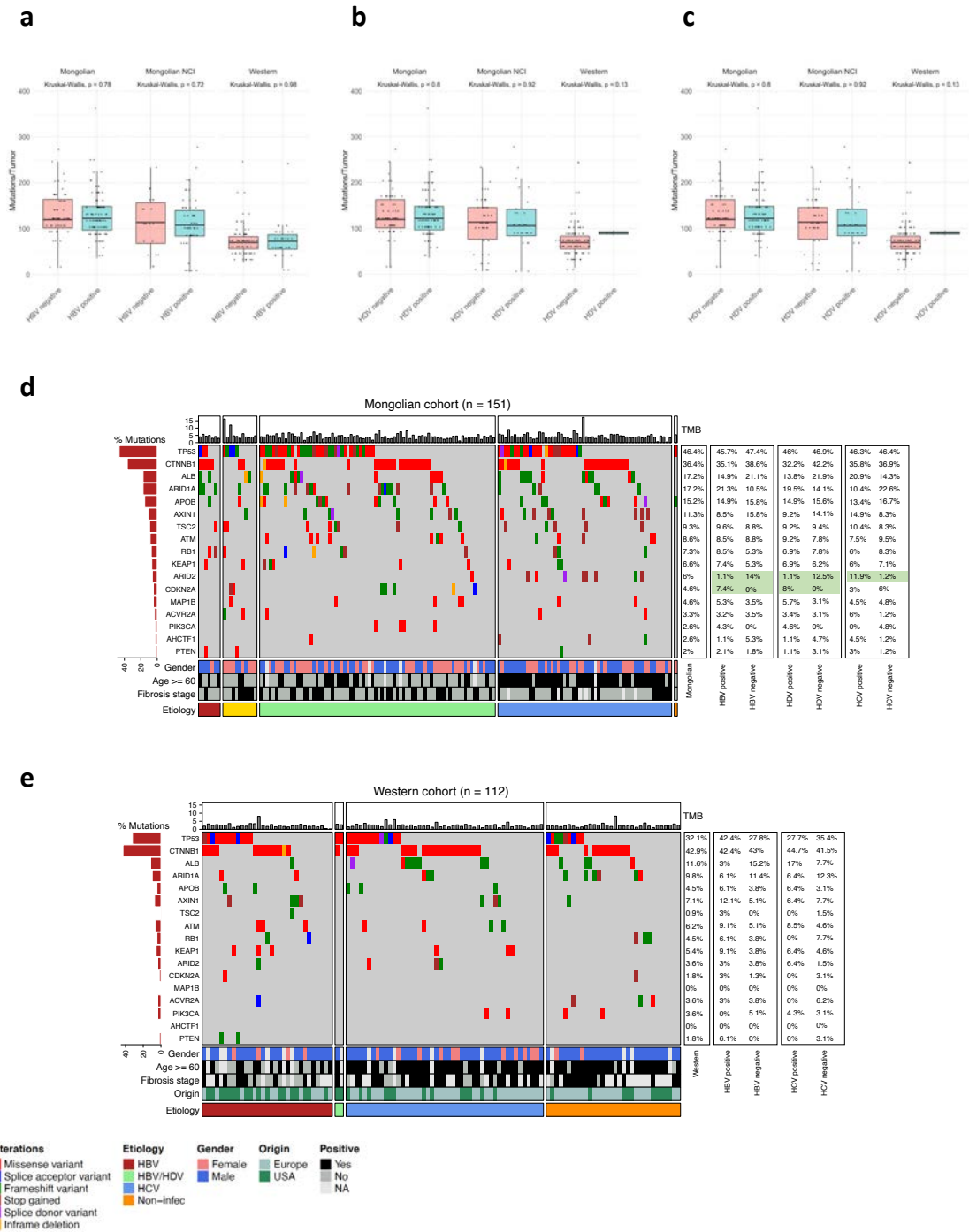
Supplementary Figure 4. Mutational profile of Mongolian and Western HCC by etiology. a-c Mutations per tumor in HCC samples from patients positive and negative for HBV (**a**), HDV (**b**), and HCV (**c**) infections. Data from the in-house Mongolian (n = 151), Mongolian NCI (n = 71) and Western (n = 112) cohorts are shown. Y axis was cut at 300 mutations/tumor to facilitate data interpretation. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers). **d-e** Mutational landscape in the Mongolian (**d**; n=151) and Western cohorts (**e**; n=112) sorted by viral status. Genes with significant differences between Mongolian and Western cohorts are shown in bold green (Fisher p < 0.05). Top panel shows tumor mutational burden (TMB, mutations/Mb) per sample. Middle panel indicates the presence of mutations per sample (right) and overall percentage (left) in the most frequently mutated HCC drivers. Percentage of mutations by viral status is also shown, and genes with significant differences are highlighted in green. Bottom panel details viral etiology, gender, and age.
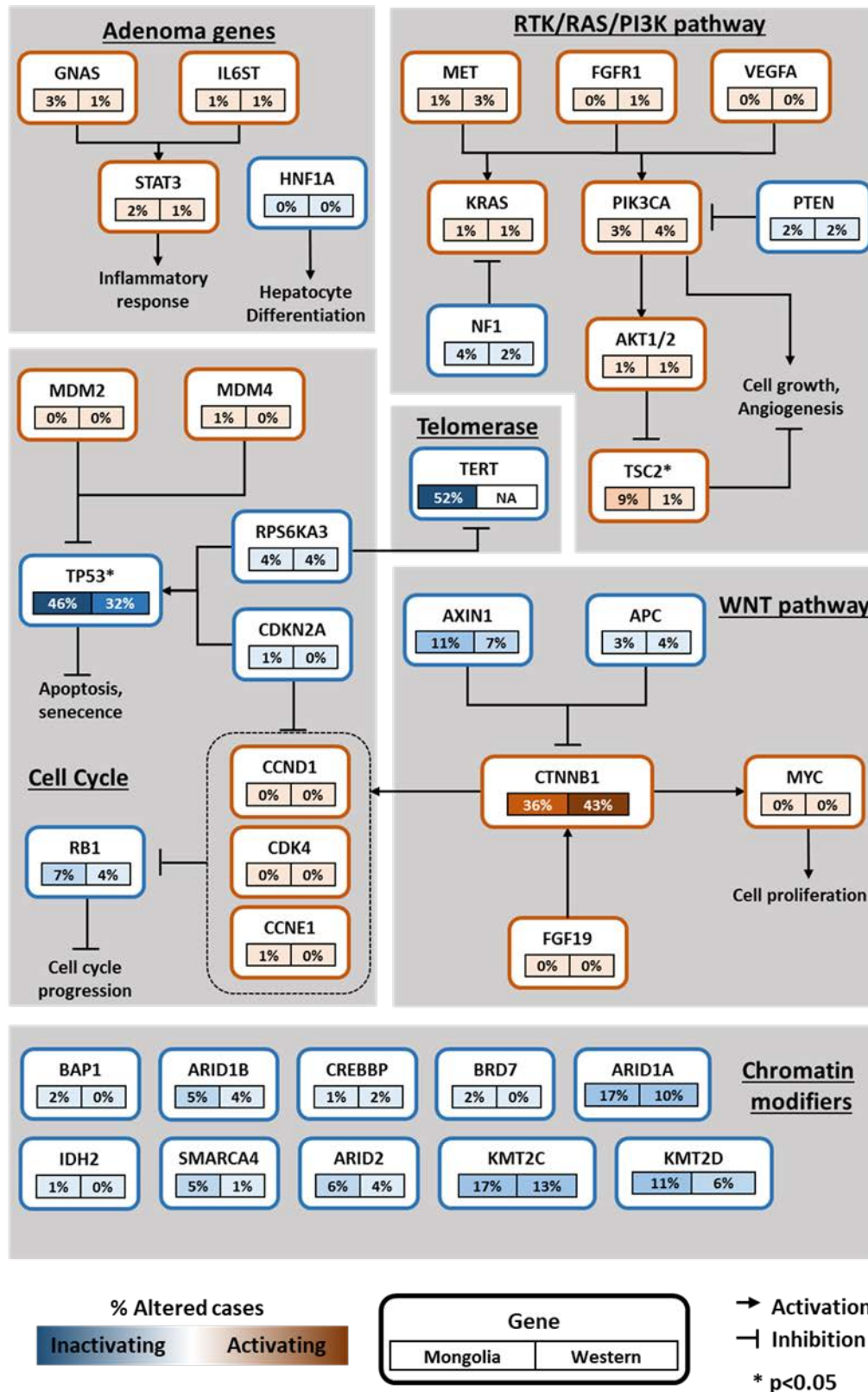
Supplementary Figure 5. Somatic mutation in HCC-driving signaling pathways. Known mutated genes in HCC grouped by signaling pathway. Molecular interactions between them are represented. The percentage of mutations in the Mongolian (left box; n=151) and Western (right box; n=112) cohorts is indicated for each gene. * p < 0.05 in Mongolian vs Western HCC.
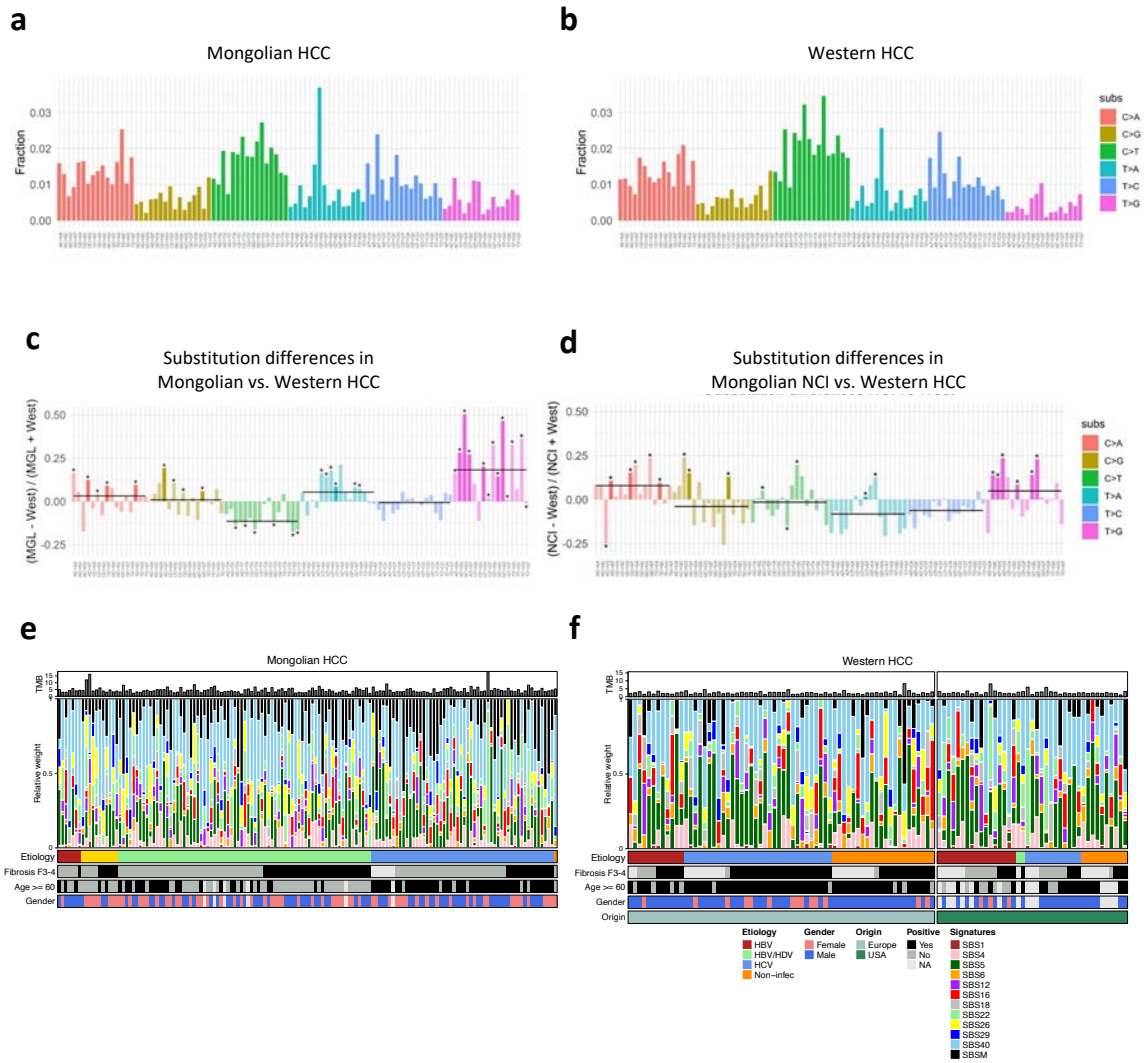
Supplementary Figure 6. Substitution profile of Mongolian and Western HCCs. a-b Trinucleotide substitution frequency in Mongolian (**a**) and Western (**b**) HCC. **c-d** Differences in trinucleotide substitution frequency between the in-house Mongolian and Western cohorts (**c**) and Mongolian NCI and Western cohorts (**d**). Bars indicate the median values for each substitution group. Significant substitutions differences in both in-house and NCI Mongolian cohorts are highlighted in brighter colors. * p < 0.05 (Kruskal-Wallis test). **e-f** Signature fitting results in Mongolian (**e**, n = 151) and Western HCC (**f**, n = 112) using HCC-specific COSMIC signatures and SBS Mongolia. Middle panel indicates the proportion of SNVs assigned to each signature per sample (relative weight). Upper panel indicates the TMB (Mutations/Mb) for each sample, and lower panels represent clinical variables.

a

b
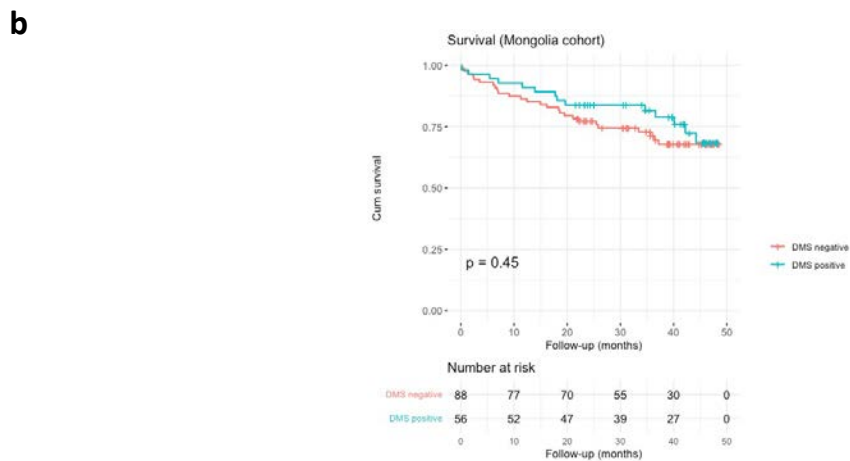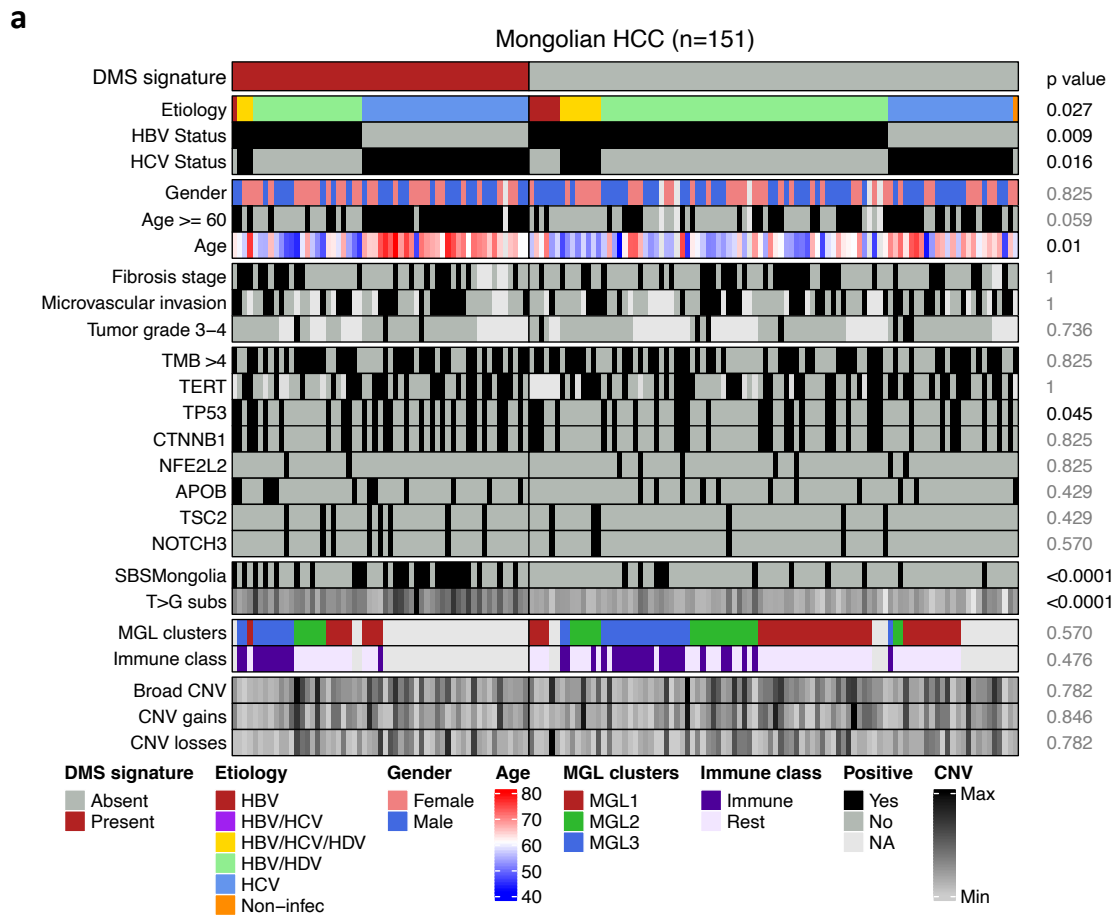
Supplementary Figure 7. Characterization of Mongolian samples presenting the DMS signature. a Clinico-pathological and molecular features of Mongolian HCC samples positive and negative for the DMS signature. P values refer to FDR-adjusted Fisher tests (categorical) and Kruskal-Wallis test (continuous). **b** Kaplan-Meier estimates of overall survival in Mongolian HCC patients according to the presence of the DMS signature.
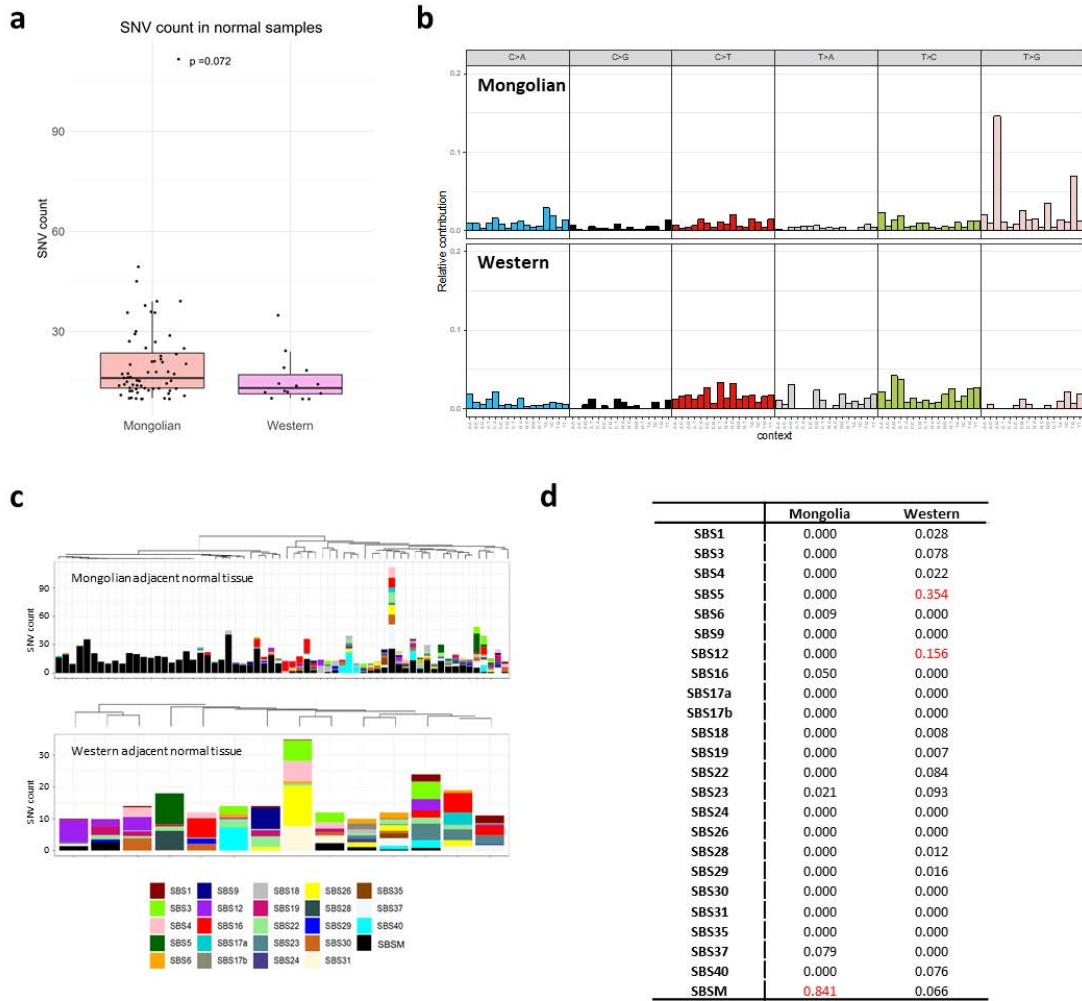
Supplementary Figure 8. Single-base substitution signature analysis in adjacent non-tumor samples. a Adjacent non-tumor samples selected for SBS analysis in the Mongolian (n=64) and Western (n=14) cohorts (exome-region SNV count >= 10 at allelic frequency cutoff of 0.05). **b** Trinucleotide mutational profiles for pooled Mongolian (total variants = 1283) and Western (total variants =215) adjacent non-tumor samples. **c** Absolute signature fitting results for individual adjacent non-tumor samples. **d** Mutational signature fitting results of pooled non-tumor samples from the Western and Mongolian cohort. Values in red indicate signatures with weight ≥ 0.1.
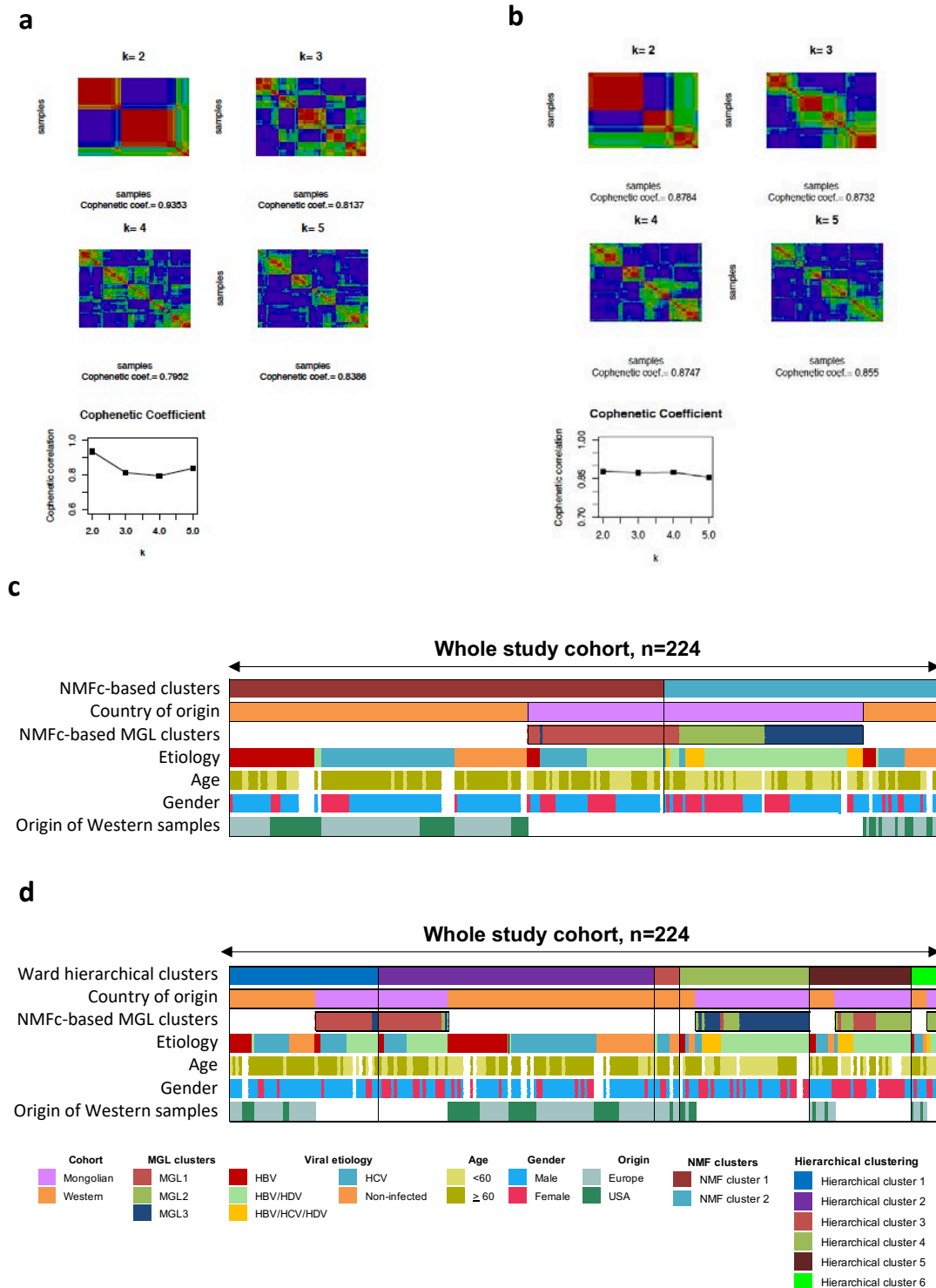
Supplementary Figure 9. a Non-negative Matrix Factorization of the whole study cohort and **b** Mongolian cohort. Bottom panels indicate cophenetic coefficients of each clusterization. **c** Heatmap integrating the whole-cohort and Mongolian clusters and clinic-pathological chatacteristics of the samples. **d** Hierarchical clustering of the whole study cohort using Euclidean distance and Ward's agglomerative procedure.
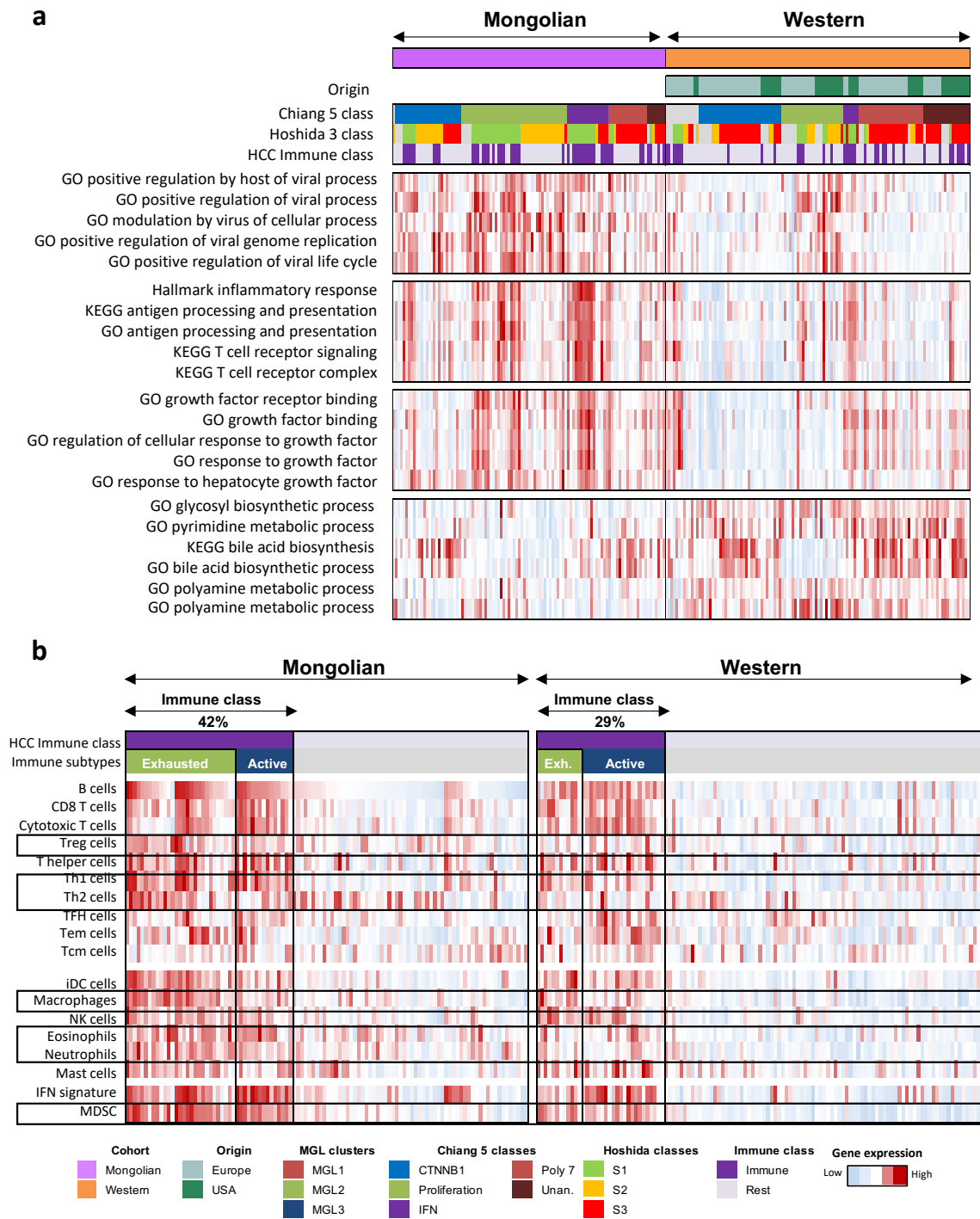
Supplementary Figure 10. Transcriptomic profile of Mongolian and Western samples. a Nearest Template Prediction (NTP) and single sample gene set enrichment analysis (ssGSEA) of Mongolian (n = 106) and Western (n = 118) HCC samples delineating the main molecular differences between cohorts. **b** Single sample gene set enrichment analysis of Mongolian and Western HCC samples indicating distinct immune cell populations and the HCC immune class assessed by Nearest Template Prediction analysis.
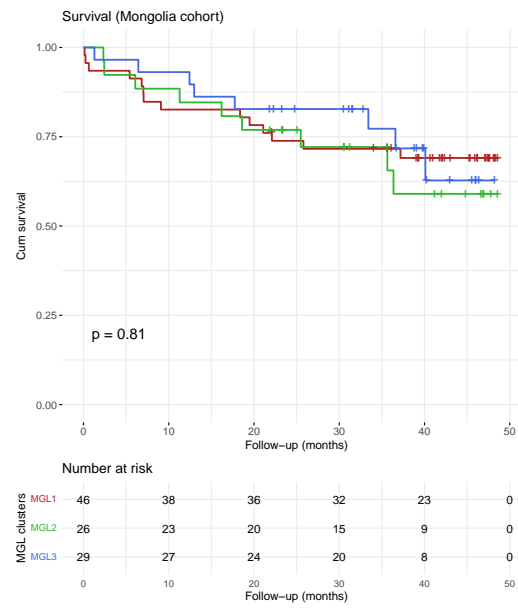
Supplementary Figure 11. Survival of Mongolian HCC patients according to the presence of MGL clusters
Kaplan-Meier estimates of overall survival in the Mongolian cohort.

a

b
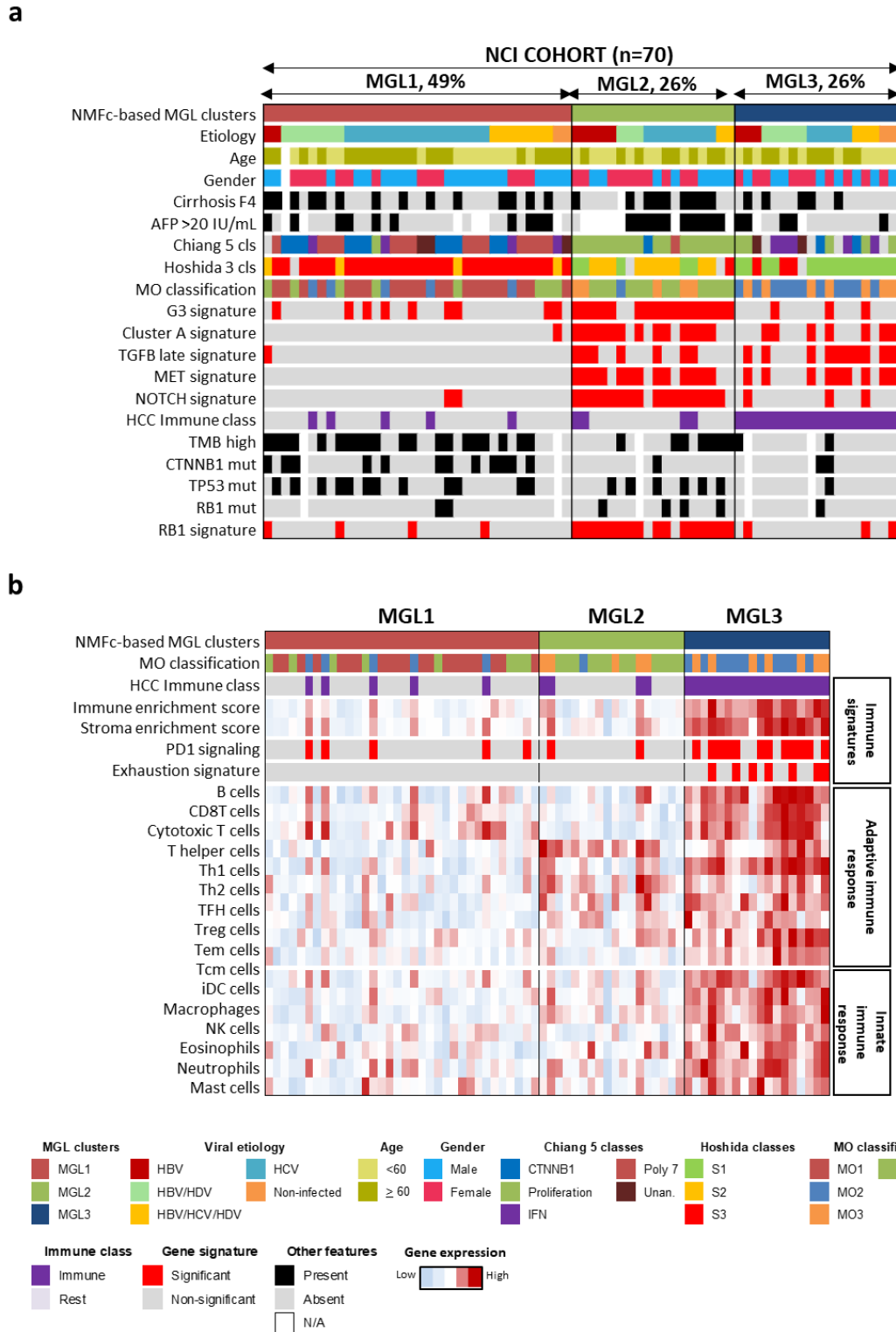
Supplementary Figure 12. **Molecular classification of Mongolian HCC in the NCI cohort. a** Consensus-clustered classification of Mongolian HCC samples (n = 70) using Non-negative Matrix Factorization. Clinico-pathological characteristics, Nearest Template Prediction (NTP) and single-sample gene set enrichment analyses (ssGSEA) are shown. **b** Characterization of inflammatory profile in the MGL clusters assessed by ESTIMATE analysis and ssGSEA capturing distinct immune populations.

a

b

		NCI classification				
		MO1	MO2	MO3	MO4	Not classified
MGL clusters	MGL1 - Metabolic	19	1	1	10	16
	MGL2 - Proliferative	0	2	3	18	4
	MGL3 - Inflamed	2	22	5	0	3
Fisher p-value = 0.0005						
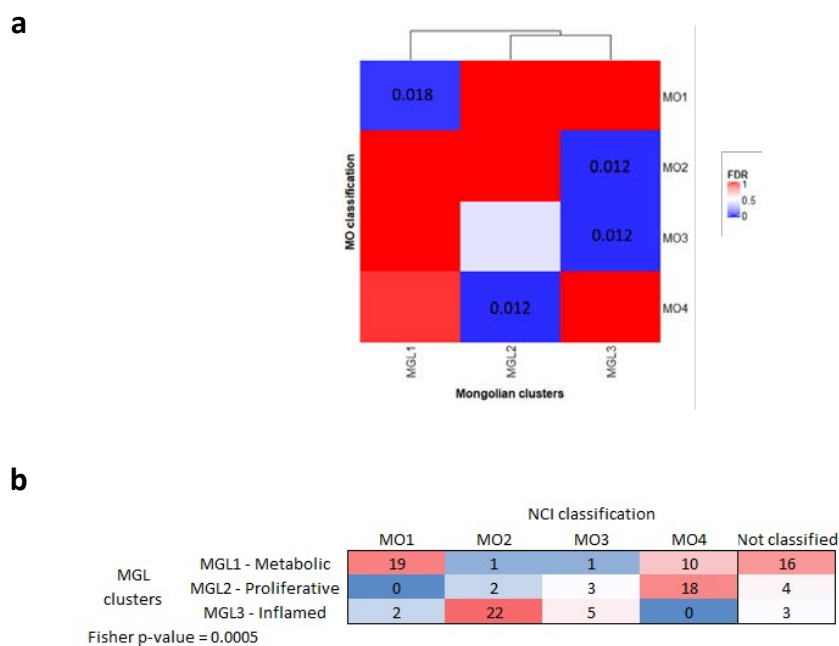
Supplementary Figure 13. Comparison between MGL and MO classifications of Mongolian HCC. a Subclass mapping analysis comparing the MGL clusters in the Mongolian cohort (n = 106) and the MO classification in the Mongolian NCI cohort (n = 70). FDR-adjusted p-values are indicated. **b** HCC samples from the in-house Mongolian cohort classified according to the MGL clusters by non-negative matrix factorization (NMFc) and to the MO classification by nearest template prediction (NTP). The number of samples in each category is indicated.

SUPPLEMENTARY TABLES

Supplementary Table 1. Baseline characteristics of the Western cohort by sample origin.

	Western Europe (n=117)	Western USA (n=70)*	P value
Age (years)	67 (40-83)	65 (29-91)	ns
< 60 years (n, %)	19 (16)	13 (24)	ns
Gender (male, %)	92 (79)	45 (82)	ns
Etiology:			
HBV+ (n, %)	23 (19)	18 (26)	
HBV/HDV+ (n, %)	1 (1)	2 (3)	0.001
HCV+ (n, %)	56 (48)	13 (19)	
Non-infected (n, %)	37 (32)	37 (53)	
Bilirubin (mg/dL)	1 (0.4-3.2)	0.7 (0.3-3.8)	ns
Albumin (g/L)	40 (25-54)	42 (22-49)	ns
Platelets (10^9/L)	161 (29-493)	154 (27-460)	ns
< 150x10^9/L (n, %)	55 (47)	26 (47)	ns
AFP > 400 IU/mL (n, %)	18 (16)	8 (17)	ns
Tumor size (cm)	4.2 (1.5-20)	4.4 (1-19)	ns
> 5 cm (n, %)	45 (41)	22 (42)	ns
BCLC stage (0-A, %)	87 (80)	35 (66)	ns
Multinodular disease (n, %)	29 (26)	13 (25)	ns
Advanced liver fibrosis (F3-4, %)	72 (88)	34 (64)	0.002
Cirrhosis (F4, %)	57 (70)	24 (45)	0.007
Microvascular invasion (yes, %)	44 (38)	36 (66)	0.001
Tumor grade (G3-4, %)	24 (27)	17 (31)	ns

*Baseline characteristics, except etiology, are missing for 15 (21%) patients in the Western USA subcohort

Supplementary Table 2. Viral genotypes in the Mongolian and Western cohorts.

	Mongolian Cohort (N=106)	Western Cohort (N=44)	p value
HBV genotype			
Genotype A (n, %)	0 (0)	2 (4.5)	
Genotype B (n, %)	0 (0)	2 (4.5)	
Genotype C (n, %)	0 (0)	12 (27.3)	<0.001
Genotype D (n, %)‡	95 (89.6)	19 (43.2)	
Non-genotypable* (n, %)	11 (10.4)	9 (20.5)	
HDV genotype			
Genotype 1 (n, %)	85 (95.5)	1 (33.3)	
Genotype 2 (n, %)	0 (0)	1 (33.3)	<0.001
Non-genotypable* (n, %)	4 (4.5)	1 (33.4)	
HBV mutations			
BCP A1762T (yes, %)†	10 (13.7)	21 (60)	<0.001
BCP G1764A (yes, %)†	15 (20.5)	23 (65.7)	<0.001
Precore G1896A (yes, %)†	21 (28.8)	16 (45.7)	ns

HBV, hepatitis B virus; HDV, hepatitis delta virus; BCP, basal pre-core
‡2 patients showed recombinant forms of the C and D genotypes
* non-genotypable due to technical failure
†HBV mutations were evaluated in 73 (69%) patients in the Mongolian cohort and 35 (80%) patients in the Western cohort

Supplementary Table 3. Baseline characteristics of HBV-infected Mongolian patients according to the 25th quartile of HBV-DNA load.

	Low HBV-DNA (<4 log copies/µg total DNA) (n=30)	High HBV-DNA (≥4 log copies/µg total DNA) (n=75)	p value
Age (years)	57.7 (18-71.3)	56,2 (41.1- 75.8)	ns
Gender (male, %)[‡]	15 (51.7)	37 (52.1)	ns
Etiology			
HBV (n, %)	8 (26.7)	7 (9.3)	
HBV/HDV (n, %)	20 (66.7)	56 (74.7)	ns
HBV/HCV/HDV (n, %)	1 (3.3)	11 (14.7)	
HBV/HCV (n, %)	1 (3.3)	1 (1.3)	
HDV+ (n, %)	21 (70)	67 (89.3)	0.021
Region			
Western (n, %)	3 (11.5)	17 (25.8)	
Central (n, %)	12 (46.2)	25 (37.9)	ns
Eastern (n, %)	4 (15.4)	4 (6.1)	
Ulaanbaatar (n, %)	7 (26.9)	20 (30.3)	
Liver fibrosis (F3-4, %)*	6 (20.7)	34 (47.2)	0.015
Tumor size (cm)	6 (3-14.9)	6.5 (1.6-20)	ns
Multinodular (yes, %)	0 (0)	12 (17.4)	0.032
BCLC stage (0-A, %)	24 (96)	50 (72.5)	0.02
BCP A1762T (yes, %)[†]	2 (11.1)	8 (14,5)	ns
BCP G1764A (yes, %)[†]	2 (11.1)	13 (23.6)	ns
Precore G1896A (yes, %)[†]	7 (38.9)	14 (25,5)	ns

HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; BCLC, Barcelona Clinic Liver Cancer; BCP, basal pre-core

‡Gender information was missing in 9 individuals in the Mongolia cohort

*Fibrosis stage was evaluated in 168 (88%) in the Mongolia cohort

†HBV mutations were evaluated in 73 (69%) patients in the Mongolian cohort (18 HBV-DNA low and 55 HBV-DNA high)

Supplementary Table 4. Baseline characteristics of HDV-infected Mongolian patients according to the 25th quartile of HDV-RNA load.

	Low HDV-RNA (<2.4 log IU/ng total RNA)(n=23)	High HDV-RNA (≥2.4 log IU/ng total RNA) (n=66)	p value
Age (years)	57 (41.1-72.6)	56 (44.2-75.6)	ns
Gender (male, %)	8 (36.4)	33 (53.2)	ns
Etiology			
HBV/HDV (n, %)	17 (73.9)	60 (90.9)	ns
HBV/HCV/HDV (n, %)	6 (26.1)	6 (9.1)	
Region			
Western (n, %)	7 (35)	7 (12.3)	
Central (n, %)	5 (25)	28 (49.1)	ns
Eastern (n, %)	3 (15)	5 (8.8)	
Ulaanbaatar (n, %)	5 (25)	17 (29.8)	
Liver fibrosis (F3-4, %)*	8 (34.8)	30 (45.5)	ns
Multinodular (yes, %)	1 (4.8)	12 (20.3)	ns
BCLC stage (0-A, %)	18 (85.7)	43 (72.9)	ns
AFP > 400 IU/mL (n, %)	7 (35)	9 (19.6)	ns
ALT (IU/L)	49.5 (14-426)	72.3 (10-452)	0.013
BCP A1762T (yes, %)	2 (14.3)	6 (13.3)	ns
BCP G1764A (yes, %)	3 (21.4)	8 (17.8)	ns
Precore G1896A (yes, %)	4 (28.6)	8 (17.8)	ns

HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; BCLC, Barcelona Clinic Liver Cancer; AFP, alfa-fetoprotein; ALT, alanine aminotransferase;BCP, basal pre-core
†‡Gender information was missing in 9 individuals in the Mongolia cohort
*Fibrosis stage was evaluated in 168 (88%) in the Mongolia cohort
†HBV mutations were evaluated in 68 (64%) patients in the Mongolia cohort

Supplementary Table 5. Baseline characteristics of HBV-infected Western patients classified according to the median HBV-DNA load.

	Low HBV-DNA (<5 log copies/μg total DNA) (n=20)	High HBV-DNA (≥5 log copies/μg total DNA) (n=21)	p value
Age (years)	63 (29-87)	62 (41-78)	ns
Gender (male, %)‡	12 (70.5)	17 (100%)	0.016
HDV+ (n, %)	1 (5)	2 (10)	ns
Liver fibrosis (F3-4, %)*	5 (41.7)	13 (76.5)	ns
Tumor size (cm)	5 (1.8-16)	5.5 (2-18)	ns
Multinodular (yes, %)	4 (23.5)	4 (21.1)	ns
BCLC stage (0-A, %)	11 (64.7)	16 (84.2)	ns
BCP A1762T (yes, %)†	8 (50)	13 (72.2)	ns
BCP G1764A (yes, %)†	9 (56.3)	14 (77.8)	ns
Precore G1896A (yes, %)†	7 (43.8)	9 (50)	ns

HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; BCLC, Barcelona Clinic Liver Cancer; BCP, basal pre-core
‡Gender information was missing in 15 individuals in the Western cohort
*Fibrosis stage was evaluated in 135 (72%) in the Western cohort
†HBV mutations were evaluated in 35 (80%) patients in the Western cohort

Supplementary Table 6. Focal copy number alterations. P values correspond to Fisher test comparing Mongolian and Western cohorts. There was no difference in overall CNV burden between cohorts.

Chromosome arm	Mongolia (n=151)	Western (n=112)	Total (n=263)	p value	Mongolia (n=151)	Western (n=112)	Total (n=263)	p value
8p	18 (12%)	11 (10%)	29 (11%)	ns	59 (39%)	66 (59%)	125 (48%)	**0.002**
9q	11 (7%)	2 (2%)	13 (5%)	**0.047**	14 (9%)	24 (21%)	38 (14%)	**0.007**
1q	72 (48%)	55 (49%)	127 (48%)	ns	2 (1%)	6 (5%)	8 (3%)	**0.01**
1p	29 (19%)	9 (8%)	38 (14%)	**0.013**	8 (5%)	17 (15%)	25 (10%)	ns
2p	17 (11%)	10 (9%)	27 (10%)	ns	7 (5%)	8 (7%)	15 (6%)	ns
2q	15 (10%)	8 (7%)	23 (9%)	ns	9 (6%)	8 (7%)	17 (6%)	ns
3p	8 (5%)	5 (4%)	13 (5%)	ns	7 (5%)	8 (7%)	15 (6%)	ns
3q	13 (9%)	5 (4%)	18 (7%)	ns	3 (2%)	8 (7%)	11 (4%)	ns
4p	6 (4%)	5 (4%)	11 (4%)	ns	26 (17%)	20 (18%)	46 (17%)	ns
4q	2 (1%)	3 (3%)	5 (2%)	ns	37 (25%)	29 (26%)	66 (25%)	ns
5p	35 (23%)	31 (28%)	66 (25%)	ns	5 (3%)	5 (4%)	10 (4%)	ns
5q	24 (16%)	27 (24%)	51 (19%)	ns	12 (8%)	5 (4%)	17 (6%)	ns
6p	40 (26%)	28 (25%)	68 (26%)	ns	4 (3%)	4 (4%)	8 (3%)	ns
6q	19 (13%)	11 (10%)	30 (11%)	ns	22 (15%)	17 (15%)	39 (15%)	ns
7p	53 (35%)	30 (27%)	83 (32%)	ns	1 (1%)	3 (3%)	4 (2%)	ns
7q	55 (36%)	30 (27%)	85 (32%)	ns	1 (1%)	3 (3%)	4 (2%)	ns
8q	66 (44%)	61 (54%)	127(48%)	ns	8 (5%)	5 (4%)	13 (5%)	ns
9p	14(9%)	4 (4%)	18 (7%)	ns	23 (15%)	25 (22%)	48 (18%)	ns
10p	12 (8%)	10 (9%)	22 (8%)	ns	11 (7%)	10 (9%)	21 (8%)	ns
10q	6 (4%)	4 (4%)	10 (4%)	ns	26 (17%)	19 (17%)	45 (17%)	ns
11p	7 (5%)	4 (4%)	11 (4%)	ns	11 (7%)	14	25 (10%)	ns
11q	7 (5%)	4 (4%)	11 (4%)	ns	12 (8%)	15	27 (10%)	ns
12p	14 (9%)	9 (8%)	23 (9%)	ns	13 (9%)	14	27 (10%)	ns
12q	14 (9%)	9 (8%)	23 (9%)	ns	9 (6%)	11	20 (8%)	ns
13q	2 (1%)	5 (4%)	7 (3%)	ns	26(17%)	26	52 (20%)	ns
14q	7 (5%)	6 (5%)	13 (5%)	ns	16 (11%)	13	29 (11%)	ns
15q	6 (4%)	2 (2%)	8 (3%)	ns	18 (12%)	10	28 (11%)	ns
16p	5 3%)	6 (5%)	11 (4%)	ns	35 (23%)	28	63 (24%)	ns
16q	2 (1%)	6 (5%)	8 (3%)	ns	47 (31%)	41	88 (33%)	ns
17p	9 (6%)	6 (5%)	15 (6%)	ns	36 (24%)	23	59 (22%)	ns
17q	21(14%)	18 (16%)	39 (15%)	ns	7 (5%)	9	16 (6%)	ns
18p	11 (7%)	3 (3%)	14 (5%)	ns	20 (13%)	13	33 (13%)	ns
18q	8 (5%)	2 (2%)	10 (4%)	ns	23 (15%)	16	39 (15%)	ns
19p	19 (13%)	8 (7%)	27 (10%)	ns	18 (12%)	19	37 (14%)	ns
19q	20 (13%)	12 (11%)	32 (12%)	ns	13 (9%)	16	29 (11%)	ns
20p	33 (22%)	22 (20%)	55 (21%)	ns	5 (3%)	6	11 (4%)	ns
20q	37 (25%)	23 (21%)	60 (23%)	ns	4 (3%)	2	6 (2%)	ns
21q	5 (3%)	10 (9%)	15 (6%)	ns	39 (26%)	24	63 (24%)	ns
22q	12 (8%)	8 (7%)	20 (8%)	ns	23 (15%)	21	44(17%)	ns

Supplementary Table 7. Protein-coding mutations and tumor mutational burden. (TMB) in the in-house and external cohorts. TMB is shown as mutations/30 MB (Alexandrov, Nature 2013) and mutations/50 MB (Schulze, Nat Gen 2015) for comparison with previously published data.

	In-house cohorts		External cohorts			
	Mongolian	Western	Mongolian NCI	TCGA	European (Schulze)	Korean (Ahn)
Mutations	121	70	111	76	61	63
TMB (Mutations/30 Mb)	4.0	2.3	3.7	2.5	2.0	2.1
TMB (Mutations/50 Mb)	2.4	1.4	2.2	1.5	1.2	1.3

Supplementary Table 8. Mutations in DNA damage repair (DDR) genes in the Mongolian and Western cohorts.

Gene	Western cohort		Mongolian cohort	
	Patients harboring mutations (n)	Patients harboring mutations (%)	Patients harboring mutations (n)	Patients harboring mutations (%)
TP53	36	32.14	70	46.36
ATM	7	6.25	13	8.61
BRCA2	4	3.57	3	1.99
ATR	3	2.68	4	2.65
HERC2	3	2.68	3	1.99
POLE	3	2.68	2	1.32
ATRX	2	1.79	9	5.96
POLD1	2	1.79	4	2.65
REV3L	2	1.79	4	2.65
TP53BP1	2	1.79	4	2.65
PTEN	2	1.79	3	1.99
CUL3	2	1.79	2	1.32
HELQ	2	1.79	1	0.66
PER1	2	1.79	1	0.66
SMARCA4	1	0.89	7	4.64
SHPRH	1	0.89	6	3.97
HFM1	1	0.89	5	3.31
RIF1	1	0.89	5	3.31
FANCA	1	0.89	3	1.99
POLA1	1	0.89	3	1.99
SLX4	1	0.89	3	1.99
SMARCAD1	1	0.89	2	1.32
MSH6	1	0.89	1	0.66
PARP4	1	0.89	1	0.66
SMC5	1	0.89	1	0.66
SMC6	1	0.89	1	0.66
WRN	1	0.89	1	0.66
FANCD2	0	0.00	6	3.97
ASCC3	0	0.00	5	3.31
FANCM	0	0.00	5	3.31
POLQ	0	0.00	5	3.31
BLM	0	0.00	4	2.65
MDC1	0	0.00	3	1.99
PALB2	0	0.00	2	1.32
RAD50	0	0.00	2	1.32
DDB1	0	0.00	1	0.66
MLH3	0	0.00	1	0.66
RFC1	0	0.00	1	0.66
TOPB1	0	0.00	0	0.00
LIGA4	0	0.00	0	0.00

ANNEX B

Supplementary Table 9. 100 genes with statistical differences between Mongolian and Western cohorts.

Gene	% mutations Mongolian	% mutations Western	Odds ratio Mongolian vs Western	P value Mongolian vs Western	% mutations Europe	% mutations USA	P value Europe vs USA
TP53	46.4%	32.1%	0.548	0.022	26.1%	41.9%	0.098
TTN	36.4%	17.0%	0.357	0.001	20.3%	11.6%	0.305
RYR2	15.9%	7.1%	0.407	0.036	8.7%	4.7%	0.708
APOB	15.2%	4.5%	0.26	0.005	1.4%	9.3%	0.071
HMCN1	15.2%	6.3%	0.371	0.03	1.4%	14.0%	0.013
SYNE1	13.2%	5.4%	0.371	0.038	4.3%	7.0%	0.674
LAMA1	11.9%	2.7%	0.203	0.006	2.9%	2.3%	1
FLG	10.6%	0.9%	0.076	0.001	1.4%	0.0%	1
ABCA13	10.6%	2.7%	0.232	0.015	4.3%	0.0%	0.284
KMT2A	10.6%	2.7%	0.232	0.015	2.9%	2.3%	1
DNAH7	10.6%	3.6%	0.313	0.036	5.8%	0.0%	0.296
NOTCH3	6.6%	0.9%	7.828	0.027	1.4%	0.0%	1
KRT7	9.3%	0.0%	0	0	0.0%	0.0%	NA
DNAH8	9.3%	0.9%	0.088	0.003	0.0%	2.3%	0.384
TSC2	9.3%	0.9%	0.088	0.003	0.0%	2.3%	0.384
AHNAK2	9.3%	0.9%	0.088	0.003	0.0%	2.3%	0.384
PTPN13	9.3%	1.8%	0.178	0.016	2.9%	0.0%	0.523
DNAH9	9.3%	2.7%	0.269	0.041	4.3%	0.0%	0.284
COL6A3	8.6%	1.8%	0.193	0.028	1.4%	2.3%	1
LRBA	7.9%	0.9%	0.104	0.009	0.0%	2.3%	0.384
ANK3	7.9%	1.8%	0.211	0.029	1.4%	2.3%	1
KIAA1109	7.9%	1.8%	0.211	0.029	1.4%	2.3%	1
CMYA5	7.3%	0.0%	0	0.003	0.0%	0.0%	NA
RNF213	7.3%	0.9%	0.115	0.015	0.0%	2.3%	0.384
PLXNA4	7.3%	0.9%	0.115	0.015	0.0%	2.3%	0.384
ALMS1	7.3%	1.8%	0.231	0.047	1.4%	2.3%	1
MYO15A	7.3%	1.8%	0.231	0.047	1.4%	2.3%	1
SLC7A8	6.6%	0.0%	0	0.006	0.0%	0.0%	NA
MYH13	6.6%	0.9%	0.127	0.027	1.4%	0.0%	1
MGAM	6.6%	0.9%	0.127	0.027	0.0%	2.3%	0.384
GPR112	6.6%	0.9%	0.127	0.027	1.4%	0.0%	1
SDK2	6.0%	0.0%	0	0.011	0.0%	0.0%	NA
NAV3	6.0%	0.9%	0.142	0.047	1.4%	0.0%	1
MKI67	6.0%	0.9%	0.142	0.047	1.4%	0.0%	1
NFE2L2	6.0%	0.9%	0.142	0.047	1.4%	0.0%	1
OTOG	6.0%	0.9%	0.142	0.047	1.4%	0.0%	1
LAMC3	6.0%	0.9%	0.142	0.047	1.4%	0.0%	1
PXDNL	6.0%	0.9%	0.142	0.047	0.0%	2.3%	0.384
GTF3C1	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
NOS1	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
KRT6A	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
JMY	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
PTPRS	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
HCN1	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
MAP2	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
DLEC1	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
CHD6	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
PDZD2	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
PLXNA3	5.3%	0.0%	0	0.023	0.0%	0.0%	NA

TELO2	5.3%	0.0%	0	0.023	0.0%	0.0%	NA
TMEM132D	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
PCDHA6	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
DMXL1	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
PRRC2A	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
ZFP36L1	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
SLIT1	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
MAP1B	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
SYCP2	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
CCDC30	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
IPO9	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
ABCA1	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
TNN	4.6%	0.0%	0	0.022	0.0%	0.0%	NA
LMTK3	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
PDGFRA	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
CD1C	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
FAM184B	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
SLC23A1	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
TTLL5	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
LILRA2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
PCSK5	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
ASTN1	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
ZRSR2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
DLGAP3	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
ABHD17A	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
BAZ2B	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
FAR2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
KCNT1	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
GABRB2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
HEPACAM2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
SLC44A5	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
ANKRD31	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
PHACTR4	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
FANCD2	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
COL16A1	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
CYP2A13	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
SPEN	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
PABPC5	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
STK31	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
CCDC146	4.0%	0.0%	0	0.04	0.0%	0.0%	NA
IGSF9B	0.7%	8.0%	13.107	0.002	0.0%	0.0%	NA
FLNB	0.7%	6.3%	10	0.012	0.0%	0.0%	NA
PTPN21	0.7%	5.4%	8.491	0.044	0.0%	0.0%	NA
FRG1B	0.0%	5.4%	inf	0.006	0.0%	0.0%	NA
KAT6A	0.0%	4.5%	inf	0.013	0.0%	0.0%	NA
TDO2	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA
TAF1A	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA
ASB14	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA
KIF20A	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA
SFSWAP	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA
IGSF3	0.0%	3.6%	inf	0.032	0.0%	0.0%	NA

Supplementary Table 10. Genes more frequently mutated in Mongolia versus other cohorts. Genes significantly mutated in 1 or more external non-Mongolian cohorts are shown.

Gene	% Mongolian in-house cohort	Number of cohorts with Significant diff	% Western in-house cohort	P VALUE Mongolian vs Western	% Korean	P VALUE Mongolian vs Korean	% European	P VALUE Mongolian vs European	% TCGA	P VALUE Mongolian vs TCGA	% Mongolian NCI	P VALUE in-house Mongolian vs Mongolian NCI
TP53	46	4	32	0.022	31	0.00	22	0.000	28	0.000	30	0.020
LAMA1	12	4	3	0.006	6	0.03	3	0.001	4	0.003	8	0.496
KRT7	9	4	0	0.000	0	0.00	1	0.000	1	0.000	0	0.006
PTPN13	9	4	2	0.016	1	0.00	1	0.000	4	0.015	3	0.099
COL6A3	9	4	2	0.028	3	0.04	3	0.036	4	0.026	8	1.000
GPR112	7	4	1	0.027	0	0.00	1	0.007	0	0.000	0	0.033
ALMS1	7	4	2	0.047	3	0.04	2	0.006	3	0.026	3	0.233
MYO15A	7	3	2	0.047	3	0.04	2	0.006	1	0.001	4	0.556
PRRC2A	5	3	0	0.022	0	0.00	1	0.0497	0	0.001	0	0.100
DLEC1	5	4	0	0.023	1	0.02	1	0.016	1	0.025	4	1.000
JMY	5	4	0	0.023	0	0.00	0	0.003	0	0.000	0	0.057
KRT6A	5	4	0	0.023	1	0.03	0	0.000	1	0.008	0	0.057
TELO2	5	4	0	0.023	0	0.00	1	0.016	1	0.001	3	0.508
PLXNA3	5	4	0	0.023	0	0.00	1	0.016	1	0.025	3	0.508
ARID1A	17	3	10	0.107	3	0.00	10	0.043	7	0.001	4	0.002
KMT2C	9	3	3	0.062	3	0.01	0	0.000	3	0.008	6	0.591
CSMD3	15	3	13	0.718	8	0.04	6	0.008	8	0.034	13	0.674
AGRN	4	3	1	0.244	0	0.02	0	0.015	0	0.001	6	0.730
CUBN	13	3	7	0.217	6	0.04	3	0.000	5	0.005	10	0.496
AKAP9	8	3	4	0.193	3	0.02	2	0.023	1	0.000	0	0.011
PRDM5	5	3	3	0.524	0	0.01	1	0.031	1	0.018	1	0.441
PIK3R4	5	3	1	0.143	1	0.03	1	0.031	1	0.009	3	0.722
CELSR1	7	3	3	0.163	3	0.04	1	0.001	2	0.009	1	0.109
MUC6	7	3	5	0.797	2	0.02	1	0.002	2	0.012	1	0.181
ARAP3	5	3	2	0.197	0	0.00	0	0.003	2	0.047	3	0.508
BTAF1	5	3	1	0.083	1	0.02	1	0.016	2	0.033	6	1.000
GABRA1	5	3	1	0.083	0	0.00	0	0.003	2	0.033	0	0.057
KIAA1731	5	3	1	0.083	1	0.03	1	0.026	0	0.000	0	0.057
BAI1	5	3	1	0.143	0	0.01	1	0.031	0	0.000	0	0.100
BEND5	3	3	0	0.074	0	0.01	0	0.033	1	0.025	0	0.180
C10orf118	3	3	0	0.074	0	0.04	0	0.033	0	0.002	0	0.180
CCDC147	3	3	0	0.074	0	0.04	0	0.033	0	0.002	0	0.180
CCT8	3	3	0	0.074	0	0.04	0	0.033	0	0.002	1	0.667
CDC42EP4	3	3	0	0.074	0	0.04	0	0.008	0	0.002	1	0.667
FAM83E	3	3	1	0.244	0	0.04	0	0.033	0	0.010	1	0.667
FERD3L	3	3	1	0.244	0	0.04	0	0.033	0	0.010	0	0.180
GART	3	3	1	0.244	0	0.04	0	0.008	1	0.025	0	0.180
HOMER3	3	3	0	0.074	0	0.04	0	0.033	0	0.002	1	0.667
LMNA	3	3	1	0.244	0	0.04	0	0.008	1	0.025	1	0.667
LRRC36	3	3	0	0.074	0	0.01	0	0.033	0	0.010	3	1.000
MAP6	3	3	0	0.074	0	0.01	0	0.033	0	0.002	1	0.667
AADAT	3	3	1	0.398	0	0.02	0	0.021	0	0.028	1	1.000
ARFIP1	3	3	0	0.139	0	0.02	0	0.021	0	0.028	1	1.000
ATP6AP1	3	3	0	0.139	0	0.02	0	0.021	0	0.028	0	0.309
PRKAR1B	3	3	1	0.398	0	0.02	0	0.021	0	0.007	0	0.309
PNPLA2	3	3	0	0.139	0	0.02	0	0.021	0	0.007	0	0.309
OSMR	3	3	0	0.139	0	0.02	0	0.021	0	0.007	1	1.000
CSMD1	0	3	0	1.000	4	0.01	8	0.000	6	0.002	10	0.001
DMD	0	3	0	1.000	6	0.00	3	0.047	4	0.008	8	0.003
DNAH17	0	3	0	1.000	5	0.00	3	0.026	4	0.013	4	0.032

DST	0	3	0	1.000	10	0.00	5	0.004	4	0.013	3	0.101
FRAS1	0	3	0	1.000	4	0.01	5	0.002	5	0.003	6	0.010
MUC16	0	3	0	1.000	23	0.00	12	0.000	16	0.000	30	0.000
MUC2	0	3	0	1.000	3	0.045	3	0.047	5	0.002	0	1.000
PCLO	0	3	0	1.000	15	0.000	9	0.000	11	0.000	15	0.000
NLRP8	5	3	3	0.363	1	0.03	1	0.016	1	0.001	0	0.057
PHACTR4	4	3	0	0.040	1	0.06	0	0.015	1	0.022	0	0.180
HMCN1	15	3	6	0.030	13	0.55	5	0.000	7	0.008	11	0.302
SYNE1	13	3	5	0.038	10	0.32	2	0.000	4	0.000	13	1.000
LRBA	8	3	1	0.009	5	0.29	1	0.002	1	0.000	4	0.397
KMT2A	11	3	3	0.015	5	0.07	2	0.001	3	0.001	6	0.316
CMYA5	7	3	0	0.003	3	0.08	1	0.003	2	0.009	1	0.109
RNF213	7	3	1	0.015	4	0.25	2	0.006	2	0.009	1	0.109
MGAM	7	3	1	0.027	6	0.83	2	0.030	2	0.018	4	0.558
LAMC3	6	3	1	0.047	2	0.09	2	0.038	2	0.017	0	0.061
CHD6	5	3	0	0.023	3	0.18	1	0.026	2	0.047	0	0.057
CCDC30	5	3	0	0.022	1	0.06	0	0.001	1	0.009	1	0.441
PCDHA6	5	3	0	0.022	1	0.06	0	0.006	1	0.047	0	0.100
ABHD17A	4	3	0	0.040	1	0.06	0	0.003	0	0.001	0	0.180
CD1C	4	3	0	0.040	1	0.06	0	0.015	0	0.003	0	0.180
CYP2A13	4	3	0	0.040	1	0.06	0	0.003	1	0.010	1	0.435
FAM184B	4	3	0	0.040	1	0.16	0	0.015	1	0.022	1	0.435
FANCD2	4	3	0	0.040	2	0.20	0	0.003	1	0.022	1	0.435
ZRSR2	4	3	0	0.040	1	0.16	0	0.015	0	0.003	1	0.435
CCDC146	4	3	0	0.040	0	0.00	0	0.003	1	0.071	4	1.000
NOTCH3	7	3	1	0.027	0	0.00	2	0.064	3	0.047	1	0.181
SLC23A1	4	3	0	0.040	0	0.00	1	0.059	0	0.003	0	0.180
OTOG	6	3	1	0.047	2	0.04	2	0.105	1	0.006	6	1.000
HCN1	5	3	0	0.023	0	0.00	2	0.067	2	0.033	0	0.057
MYH13	7	3	1	0.027	2	0.02	2	0.064	1	0.003	4	0.558
TSC2	9	3	1	0.003	3	0.01	5	0.088	3	0.008	7	0.798
FAR2	4	3	0	0.040	0	0.00	1	0.059	0	0.003	1	0.435
HEPACAM2	4	3	0	0.040	0	0.02	1	0.059	1	0.022	0	0.180
TTLL5	4	3	0	0.040	0	0.00	1	0.059	0	0.003	3	1.000

Supplementary Table 11. Assignment of the four de novo extracted HCC signatures to all single and linear combinations of two COSMIC v3 signatures. Signature mapping highlighted in yellow indicated the assignment used in our analysis, selected based on cosine similarity (Cos sim).

de novo signature 1			de novo signature 2			de novo signature 3			de novo signature 4 (SBSM)		
Reference signature	Weight	Cos sim	Reference signature	Weight	Cos sim	Reference signature	Weight	Cos sim	Reference signature	Weight	Cos sim
SBS22	SBS22=1	0.9735	SBS6 + SBS40	SBS6=0.12; SBS40=0.88	0.91406	SBS16 + SBS26	SBS16=0.45; SBS26=0.55	0.92379	SBS28 + SBS40	SBS28=0.09; SBS40=0.91	0.81766
SBS3 + SBS22	SBS3=0.12; SBS22=0.88	0.97496	SBS15 + SBS40	SBS15=0.09; SBS40=0.91	0.90401	SBS12 + SBS16	SBS12=0.55; SBS16=0.45	0.91218	SBS17b + SBS40	SBS17b=0.07; SBS40=0.93	0.81288
SBS22 + SBS25	SBS22=0.86; SBS25=0.14	0.97488	SBS30 + SBS40	SBS30=0.13; SBS40=0.87	0.90294	SBS16 + SBS54	SBS16=0.67; SBS54=0.33	0.89886	SBS9 + SBS40	SBS9=0.3; SBS40=0.7	0.80978
SBS8 + SBS22	SBS8=0.09; SBS22=0.91	0.97469	SBS23 + SBS40	SBS23=0.1; SBS40=0.9	0.9012	SBS16 + SBS46	SBS16=0.59; SBS46=0.41	0.88026	SBS40 + SBS55	SBS40=0.9; SBS55=0.1	0.8047
SBS4 + SBS22	SBS4=0.07; SBS22=0.93	0.97468	SBS19 + SBS40	SBS19=0.09; SBS40=0.91	0.8984	SBS16 + SBS33	SBS16=0.76; SBS33=0.24	0.87846	SBS40 + SBS43	SBS40=0.93; SBS43=0.07	0.78171
SBS22 + SBS40	SBS22=0.9; SBS40=0.1	0.97462	SBS40 + SBS42	SBS40=0.85; SBS42=0.15	0.89789	SBS16 + SBS21	SBS16=0.74; SBS21=0.26	0.86621	SBS40 + SBS60	SBS40=0.98; SBS60=0.02	0.77619
SBS22 + SBS46	SBS22=0.95; SBS46=0.05	0.97415	SBS40 + SBS84	SBS40=0.89; SBS84=0.11	0.89751	SBS5 + SBS16	SBS5=0.59; SBS16=0.41	0.86563	SBS40 + SBS54	SBS40=0.97; SBS54=0.03	0.77316
SBS18 + SBS22	SBS18=0.04; SBS22=0.96	0.97412	SBS32 + SBS40	SBS32=0.11; SBS40=0.89	0.89374	SBS5 + SBS26	SBS5=0.55; SBS26=0.45	0.85408	SBS40 + SBS41	SBS40=0.93; SBS41=0.07	0.77314
SBS22 + SBS45	SBS22=0.97; SBS45=0.03	0.97409	SBS1 + SBS40	SBS1=0.05; SBS40=0.95	0.89348	SBS16 + SBS25	SBS16=0.56; SBS25=0.44	0.85369	SBS37 + SBS40	SBS37=0.06; SBS40=0.94	0.77238
SBS22 + SBS39	SBS22=0.94; SBS39=0.06	0.97407	SBS40 + SBS44	SBS40=0.88; SBS44=0.12	0.89218	SBS16 + SBS37	SBS16=0.57; SBS37=0.43	0.85156	SBS40 + SBS51	SBS40=0.97; SBS51=0.03	0.77162
SBS22 + SBS36	SBS22=0.97; SBS36=0.03	0.97406	SBS24 + SBS40	SBS24=0.13; SBS40=0.87	0.89124	SBS3 + SBS16	SBS3=0.46; SBS16=0.54	0.84754	SBS40 + SBS57	SBS40=0.98; SBS57=0.02	0.77099
SBS22 + SBS38	SBS22=0.98; SBS38=0.02	0.97392	SBS11 + SBS40	SBS11=0.08; SBS40=0.92	0.89086	SBS5 + SBS12	SBS5=0.55; SBS12=0.45	0.84635	SBS1 + SBS40	SBS1=0; SBS40=1	0.7703
SBS17a + SBS22	SBS17a=0.02; SBS22=0.98	0.97391	SBS7b + SBS40	SBS7b=0.07; SBS40=0.93	0.88954	SBS16 + SBS44	SBS16=0.74; SBS44=0.26	0.84225	SBS2 + SBS40	SBS2=0; SBS40=1	0.7703
SBS22 + SBS35	SBS22=0.96; SBS35=0.04	0.97391	SBS7a + SBS40	SBS7a=0.05; SBS40=0.95	0.88584	SBS16 + SBS40	SBS16=0.62; SBS40=0.38	0.83552	SBS3 + SBS40	SBS3=0; SBS40=1	0.7703
SBS22 + SBS56	SBS22=0.98; SBS56=0.02	0.9739	SBS29 + SBS40	SBS29=0.1; SBS40=0.9	0.88549	SBS16 + SBS17a	SBS16=0.89; SBS17a=0.11	0.83266	SBS4 + SBS40	SBS4=0; SBS40=1	0.7703
SBS5 + SBS22	SBS5=0.06; SBS22=0.94	0.97389	SBS5 + SBS40	SBS5=0.22; SBS40=0.78	0.88366	SBS26 + SBS40	SBS26=0.62; SBS40=0.38	0.83076	SBS5 + SBS40	SBS5=0; SBS40=1	0.7703
SBS22 + SBS54	SBS22=0.98; SBS54=0.02	0.97388	SBS31 + SBS40	SBS31=0.07; SBS40=0.93	0.88193	SBS9 + SBS16	SBS9=0.27; SBS16=0.73	0.83046	SBS6 + SBS40	SBS6=0; SBS40=1	0.7703
SBS22 + SBS29	SBS22=0.97; SBS29=0.03	0.97387	SBS18 + SBS40	SBS18=0.07; SBS40=0.93	0.88079	SBS8 + SBS26	SBS8=0.28; SBS26=0.72	0.8304	SBS7a + SBS40	SBS7a=0; SBS40=1	0.7703
SBS22 + SBS24	SBS22=0.97; SBS24=0.03	0.97383	SBS20 + SBS40	SBS20=0.05; SBS40=0.95	0.88025	SBS4 + SBS16	SBS4=0.23; SBS16=0.77	0.8302	SBS7b + SBS40	SBS7b=0; SBS40=1	0.7703
SBS22 + SBS53	SBS22=0.98; SBS53=0.02	0.97381	SBS2 + SBS40	SBS2=0.02; SBS40=0.98	0.88024	SBS26 + SBS58	SBS26=0.78; SBS58=0.22	0.82948	SBS7c + SBS40	SBS7c=0; SBS40=1	0.7703
SBS13 + SBS22	SBS13=0.01; SBS22=0.99	0.97381	SBS21 + SBS40	SBS21=0.04; SBS40=0.96	0.8797	SBS7d + SBS16	SBS7d=0.13; SBS16=0.87	0.82906	SBS7d + SBS40	SBS7d=0; SBS40=1	0.7703
SBS20 + SBS22	SBS20=0.02; SBS22=0.98	0.97378	SBS36 + SBS40	SBS36=0.04; SBS40=0.96	0.87903	SBS3 + SBS26	SBS3=0.38; SBS26=0.62	0.82839	SBS8 + SBS40	SBS8=0; SBS40=1	0.7703
SBS10a + SBS22	SBS10a=0.01; SBS22=0.99	0.97378	SBS4 + SBS40	SBS4=0.07; SBS40=0.93	0.87889	SBS8 + SBS16	SBS8=0.24; SBS16=0.76	0.82801	SBS10a + SBS40	SBS10a=0; SBS40=1	0.7703
SBS9 + SBS22	SBS9=0.04; SBS22=0.96	0.97375	SBS40 + SBS52	SBS40=0.98; SBS52=0.02	0.87874	SBS16 + SBS20	SBS16=0.86; SBS20=0.14	0.82791	SBS10b + SBS40	SBS10b=0; SBS40=1	0.7703
SBS22 + SBS57	SBS22=0.98; SBS57=0.02	0.97371	SBS7d + SBS40	SBS7d=0.03; SBS40=0.97	0.87816	SBS4 + SBS26	SBS4=0.24; SBS26=0.76	0.82751	SBS11 + SBS40	SBS11=0; SBS40=1	0.7703
SBS22 + SBS52	SBS22=0.99; SBS52=0.01	0.97369	SBS40 + SBS50	SBS40=0.96; SBS50=0.04	0.87811	SBS16 + SBS42	SBS16=0.82; SBS42=0.18	0.82731	SBS12 + SBS40	SBS12=0; SBS40=1	0.7703
SBS22 + SBS33	SBS22=0.99; SBS33=0.01	0.97369	SBS35 + SBS40	SBS35=0.05; SBS40=0.95	0.87799	SBS16 + SBS35	SBS16=0.81; SBS35=0.19	0.82729	SBS13 + SBS40	SBS13=0; SBS40=1	0.7703
SBS22 + SBS44	SBS22=0.98; SBS44=0.02	0.97368	SBS14 + SBS40	SBS14=0.03; SBS40=0.97	0.87768	SBS25 + SBS26	SBS25=0.33; SBS26=0.67	0.82666	SBS14 + SBS40	SBS14=0; SBS40=1	0.7703
SBS22 + SBS41	SBS22=0.97; SBS41=0.03	0.97368	SBS10b + SBS40	SBS10b=0.02; SBS40=0.98	0.87748	SBS16 + SBS18	SBS16=0.84; SBS18=0.16	0.82605	SBS15 + SBS40	SBS15=0; SBS40=1	0.7703
SBS22 + SBS55	SBS22=0.99; SBS55=0.01	0.97366	SBS40 + SBS46	SBS40=0.96; SBS46=0.04	0.87704	SBS14 + SBS16	SBS14=0.11; SBS16=0.89	0.82554	SBS16 + SBS40	SBS16=0; SBS40=1	0.7703
SBS22 + SBS49	SBS22=0.99; SBS49=0.01	0.97365	SBS40 + SBS45	SBS40=0.98; SBS45=0.02	0.87686	SBS16 + SBS24	SBS16=0.83; SBS24=0.17	0.82548	SBS17a + SBS40	SBS17a=0; SBS40=1	0.7703
SBS22 + SBS31	SBS22=0.98; SBS31=0.02	0.97365	SBS40 + SBS48	SBS40=0.99; SBS48=0.01	0.87654	SBS16 + SBS31	SBS16=0.85; SBS31=0.15	0.82525	SBS18 + SBS40	SBS18=0; SBS40=1	0.7703
SBS14 + SBS22	SBS14=0.01; SBS22=0.99	0.97364	SBS40 + SBS49	SBS40=0.99; SBS49=0.01	0.87642	SBS16 + SBS29	SBS16=0.85; SBS29=0.15	0.82461	SBS19 + SBS40	SBS19=0; SBS40=1	0.7703
SBS22 + SBS42	SBS22=0.98; SBS42=0.02	0.97363	SBS40 + SBS59	SBS40=0.99; SBS59=0.01	0.8761	SBS16 + SBS36	SBS16=0.88; SBS36=0.12	0.8245	SBS20 + SBS40	SBS20=0; SBS40=1	0.7703

SBS22 + SBS26	SBS22=0.98; SBS26=0.02	0.97362	SBS40 + SBS56	SBS40=0.99; SBS56=0.01	0.87605	SBS16 + SBS43	SBS16=0.89; SBS43=0.11	0.82397	SBS21 + SBS40	SBS21=0; SBS40=1	0.7703
SBS7d + SBS22	SBS7d=0.01; SBS22=0.99	0.97362	SBS40 + SBS53	SBS40=0.99; SBS53=0.01	0.87588	SBS16 + SBS57	SBS16=0.87; SBS57=0.13	0.82355	SBS22 + SBS40	SBS22=0; SBS40=1	0.7703
SBS21 + SBS22	SBS21=0.01; SBS22=0.99	0.9736	SBS3 + SBS40	SBS3=0.05; SBS40=0.95	0.87583	SBS16 + SBS41	SBS16=0.84; SBS41=0.16	0.82332	SBS23 + SBS40	SBS23=0; SBS40=1	0.7703
SBS22 + SBS30	SBS22=0.99; SBS30=0.01	0.97359	SBS38 + SBS40	SBS38=0.01; SBS40=0.99	0.87583	SBS16 + SBS22	SBS16=0.89; SBS22=0.11	0.82315	SBS24 + SBS40	SBS24=0; SBS40=1	0.7703
SBS22 + SBS28	SBS22=0.99; SBS28=0.01	0.97359	SBS25 + SBS40	SBS25=0.01; SBS40=0.99	0.87559	SBS16 + SBS58	SBS16=0.87; SBS58=0.13	0.82201	SBS25 + SBS40	SBS25=0; SBS40=1	0.7703
SBS7b + SBS22	SBS7b=0.01; SBS22=0.99	0.97358	SBS33 + SBS40	SBS33=0.01; SBS40=0.99	0.87558	SBS7c + SBS16	SBS7c=0.09; SBS16=0.91	0.82177	SBS26 + SBS40	SBS26=0; SBS40=1	0.7703
SBS15 + SBS22	SBS15=0.01; SBS22=0.99	0.97358	SBS7c + SBS40	SBS7c=0; SBS40=1	0.87558	SBS16 + SBS39	SBS16=0.82; SBS39=0.18	0.82168	SBS27 + SBS40	SBS27=0; SBS40=1	0.7703
SBS22 + SBS37	SBS22=0.98; SBS37=0.02	0.97358	SBS8 + SBS40	SBS8=0; SBS40=1	0.87558	SBS16 + SBS45	SBS16=0.93; SBS45=0.07	0.82148	SBS29 + SBS40	SBS29=0; SBS40=1	0.7703
SBS22 + SBS50	SBS22=0.99; SBS50=0.01	0.97357	SBS9 + SBS40	SBS9=0; SBS40=1	0.87558	SBS16 + SBS50	SBS16=0.9; SBS50=0.1	0.82125	SBS30 + SBS40	SBS30=0; SBS40=1	0.7703
SBS22 + SBS23	SBS22=0.99; SBS23=0.01	0.97356	SBS10a + SBS40	SBS10a=0; SBS40=1	0.87558	SBS16 + SBS17b	SBS16=0.95; SBS17b=0.05	0.82121	SBS31 + SBS40	SBS31=0; SBS40=1	0.7703
SBS12 + SBS22	SBS12=0.02; SBS22=0.98	0.97356	SBS12 + SBS40	SBS12=0; SBS40=1	0.87558	SBS12 + SBS58	SBS12=0.79; SBS58=0.21	0.82118	SBS32 + SBS40	SBS32=0; SBS40=1	0.7703
SBS22 + SBS32	SBS22=0.99; SBS32=0.01	0.97355	SBS13 + SBS40	SBS13=0; SBS40=1	0.87558	SBS26 + SBS35	SBS26=0.8; SBS35=0.2	0.82115	SBS33 + SBS40	SBS33=0; SBS40=1	0.7703
SBS6 + SBS22	SBS6=0.01; SBS22=0.99	0.97353	SBS16 + SBS40	SBS16=0; SBS40=1	0.87558	SBS16 + SBS56	SBS16=0.95; SBS56=0.05	0.82099	SBS34 + SBS40	SBS34=0; SBS40=1	0.7703
SBS22 + SBS84	SBS22=0.99; SBS84=0.01	0.97353	SBS17a + SBS40	SBS17a=0; SBS40=1	0.87558	SBS10a + SBS16	SBS10a=0.04; SBS16=0.96	0.82091	SBS35 + SBS40	SBS35=0; SBS40=1	0.7703
SBS19 + SBS22	SBS19=0.01; SBS22=0.99	0.97353	SBS17b + SBS40	SBS17b=0; SBS40=1	0.87558	SBS16 + SBS53	SBS16=0.93; SBS53=0.07	0.82086	SBS36 + SBS40	SBS36=0; SBS40=1	0.7703
SBS22 + SBS43	SBS22=0.99; SBS43=0.01	0.97353	SBS22 + SBS40	SBS22=0; SBS40=1	0.87558	SBS7b + SBS16	SBS7b=0.07; SBS16=0.93	0.82053	SBS38 + SBS40	SBS38=0; SBS40=1	0.7703
SBS11 + SBS22	SBS11=0.01; SBS22=0.99	0.9735	SBS26 + SBS40	SBS26=0; SBS40=1	0.87558	SBS16 + SBS38	SBS16=0.95; SBS38=0.05	0.8205	SBS39 + SBS40	SBS39=0; SBS40=1	0.7703
SBS1 + SBS22	SBS1=0; SBS22=1	0.9735	SBS27 + SBS40	SBS27=0; SBS40=1	0.87558	SBS16 + SBS30	SBS16=0.92; SBS30=0.08	0.82029	SBS40 + SBS42	SBS40=1; SBS42=0	0.7703
SBS2 + SBS22	SBS2=0; SBS22=1	0.9735	SBS28 + SBS40	SBS28=0; SBS40=1	0.87558	SBS16 + SBS51	SBS16=0.91; SBS51=0.09	0.82005	SBS40 + SBS44	SBS40=1; SBS44=0	0.7703
SBS7a + SBS22	SBS7a=0; SBS22=1	0.9735	SBS34 + SBS40	SBS34=0; SBS40=1	0.87558	SBS15 + SBS16	SBS15=0.06; SBS16=0.94	0.81994	SBS40 + SBS45	SBS40=1; SBS45=0	0.7703
SBS7c + SBS22	SBS7c=0; SBS22=1	0.9735	SBS37 + SBS40	SBS37=0; SBS40=1	0.87558	SBS16 + SBS23	SBS16=0.94; SBS23=0.06	0.81989	SBS40 + SBS46	SBS40=1; SBS46=0	0.7703
SBS10b + SBS22	SBS10b=0; SBS22=1	0.9735	SBS39 + SBS40	SBS39=0; SBS40=1	0.87558	SBS16 + SBS32	SBS16=0.92; SBS32=0.08	0.81977	SBS40 + SBS47	SBS40=1; SBS47=0	0.7703
SBS16 + SBS22	SBS16=0; SBS22=1	0.9735	SBS40 + SBS41	SBS40=1; SBS41=0	0.87558	SBS18 + SBS26	SBS18=0.17; SBS26=0.83	0.81968	SBS40 + SBS48	SBS40=1; SBS48=0	0.7703
SBS17b + SBS22	SBS17b=0; SBS22=1	0.9735	SBS40 + SBS43	SBS40=1; SBS43=0	0.87558	SBS11 + SBS16	SBS11=0.05; SBS16=0.95	0.81942	SBS40 + SBS49	SBS40=1; SBS49=0	0.7703
SBS22 + SBS27	SBS22=1; SBS27=0	0.9735	SBS40 + SBS47	SBS40=1; SBS47=0	0.87558	SBS12 + SBS40	SBS12=0.65; SBS40=0.35	0.81938	SBS40 + SBS50	SBS40=1; SBS50=0	0.7703
SBS22 + SBS34	SBS22=1; SBS34=0	0.9735	SBS40 + SBS51	SBS40=1; SBS51=0	0.87558	SBS6 + SBS16	SBS6=0.05; SBS16=0.95	0.81914	SBS40 + SBS52	SBS40=1; SBS52=0	0.7703
SBS22 + SBS47	SBS22=1; SBS47=0	0.9735	SBS40 + SBS54	SBS40=1; SBS54=0	0.87558	SBS16 + SBS84	SBS16=0.94; SBS84=0.06	0.81912	SBS40 + SBS53	SBS40=1; SBS53=0	0.7703
SBS22 + SBS48	SBS22=1; SBS48=0	0.9735	SBS40 + SBS55	SBS40=1; SBS55=0	0.87558	SBS12 + SBS25	SBS12=0.67; SBS25=0.33	0.8191	SBS40 + SBS56	SBS40=1; SBS56=0	0.7703
SBS22 + SBS51	SBS22=1; SBS51=0	0.9735	SBS40 + SBS57	SBS40=1; SBS57=0	0.87558	SBS16 + SBS55	SBS16=0.96; SBS55=0.04	0.81907	SBS40 + SBS58	SBS40=1; SBS58=0	0.7703
SBS22 + SBS58	SBS22=1; SBS58=0	0.9735	SBS40 + SBS58	SBS40=1; SBS58=0	0.87558	SBS16 + SBS85	SBS16=0.93; SBS85=0.07	0.81886	SBS40 + SBS59	SBS40=1; SBS59=0	0.7703
SBS22 + SBS59	SBS22=1; SBS59=0	0.9735	SBS40 + SBS60	SBS40=1; SBS60=0	0.87558	SBS16 + SBS19	SBS16=0.96; SBS19=0.04	0.81872	SBS40 + SBS84	SBS40=1; SBS84=0	0.7703
SBS22 + SBS60	SBS22=1; SBS60=0	0.9735	SBS40 + SBS85	SBS40=1; SBS85=0	0.87558	SBS16 + SBS59	SBS16=0.97; SBS59=0.03	0.81868	SBS40 + SBS85	SBS40=1; SBS85=0	0.7703
SBS22 + SBS85	SBS22=1; SBS85=0	0.9735	SBS40	SBS40=1	0.87558	SBS7a + SBS16	SBS7a=0.03; SBS16=0.97	0.81845	SBS40	SBS40=1	0.7703
SBS25 + SBS27	SBS25=0.9; SBS27=0.1	0.7411	SBS3 + SBS30	SBS3=0.8; SBS30=0.2	0.86242	SBS16 + SBS52	SBS16=0.98; SBS52=0.02	0.81845	SBS3 + SBS9	SBS3=0.51; SBS9=0.49	0.74308
SBS25 + SBS34	SBS25=0.96; SBS34=0.04	0.72535	SBS5 + SBS29	SBS5=0.77; SBS29=0.23	0.85806	SBS22 + SBS26	SBS22=0.14; SBS26=0.86	0.81825	SBS3 + SBS28	SBS3=0.86; SBS28=0.14	0.74092
SBS25 + SBS35	SBS25=0.91; SBS35=0.09	0.72431	SBS5 + SBS18	SBS5=0.77; SBS18=0.23	0.85756	SBS10b + SBS16	SBS10b=0.02; SBS16=0.98	0.81815	SBS4 + SBS9	SBS4=0.27; SBS9=0.73	0.72368
SBS8 + SBS25	SBS8=0.04; SBS25=0.96	0.72021	SBS4 + SBS5	SBS4=0.27; SBS5=0.73	0.85722	SBS16 + SBS49	SBS16=0.98; SBS49=0.02	0.81808	SBS9 + SBS55	SBS9=0.85; SBS55=0.15	0.72333
SBS25 + SBS47	SBS25=0.98; SBS47=0.02	0.72016	SBS5 + SBS36	SBS5=0.84; SBS36=0.16	0.85037	SBS16 + SBS60	SBS16=0.99; SBS60=0.01	0.81805	SBS9 + SBS39	SBS9=0.67; SBS39=0.33	0.71522
SBS1 + SBS25	SBS1=0; SBS25=1	0.71986	SBS5 + SBS24	SBS5=0.76; SBS24=0.24	0.84979	SBS4 + SBS12	SBS4=0.23; SBS12=0.77	0.8179	SBS9 + SBS18	SBS9=0.78; SBS18=0.22	0.71505

ANNEX B

SBS2 + SBS25	SBS2=0; SBS25=1	0.71986	SBS3 + SBS7a	SBS3=0.89; SBS7a=0.11	0.84274	SBS16 + SBS47	SBS16=0.98; SBS47=0.02	0.81789	SBS8 + SBS9	SBS8=0.29; SBS9=0.71	0.71393
SBS3 + SBS25	SBS3=0; SBS25=1	0.71986	SBS2 + SBS3	SBS2=0.07; SBS3=0.93	0.83827	SBS1 + SBS16	SBS1=0; SBS16=1	0.81776	SBS9 + SBS29	SBS9=0.79; SBS29=0.21	0.71238
SBS4 + SBS25	SBS4=0; SBS25=1	0.71986	SBS3 + SBS32	SBS3=0.82; SBS32=0.18	0.83301	SBS2 + SBS16	SBS2=0; SBS16=1	0.81776	SBS5 + SBS9	SBS5=0.4; SBS9=0.6	0.71213
SBS5 + SBS25	SBS5=0; SBS25=1	0.71986	SBS3 + SBS5	SBS3=0.49; SBS5=0.51	0.83275	SBS13 + SBS16	SBS13=0; SBS16=1	0.81776	SBS9 + SBS50	SBS9=0.82; SBS50=0.18	0.71142
SBS6 + SBS25	SBS6=0; SBS25=1	0.71986	SBS5 + SBS52	SBS5=0.92; SBS52=0.08	0.83155	SBS16 + SBS27	SBS16=1; SBS27=0	0.81776	SBS9 + SBS51	SBS9=0.8; SBS51=0.2	0.71029
SBS7a + SBS25	SBS7a=0; SBS25=1	0.71986	SBS3 + SBS6	SBS3=0.87; SBS6=0.13	0.83149	SBS16 + SBS28	SBS16=1; SBS28=0	0.81776	SBS9 + SBS24	SBS9=0.79; SBS24=0.21	0.70762
SBS7b + SBS25	SBS7b=0; SBS25=1	0.71986	SBS5 + SBS8	SBS5=0.75; SBS8=0.25	0.83072	SBS16 + SBS34	SBS16=1; SBS34=0	0.81776	SBS3 + SBS17b	SBS3=0.9; SBS17b=0.1	0.7051
SBS7c + SBS25	SBS7c=0; SBS25=1	0.71986	SBS5 + SBS30	SBS5=0.83; SBS30=0.17	0.82942	SBS16 + SBS48	SBS16=1; SBS48=0	0.81776	SBS9 + SBS36	SBS9=0.85; SBS36=0.15	0.70343
SBS7d + SBS25	SBS7d=0; SBS25=1	0.71986	SBS5 + SBS45	SBS5=0.9; SBS45=0.1	0.82939	SBS16	SBS16=1	0.81776	SBS5 + SBS28	SBS5=0.83; SBS28=0.17	0.70299
SBS9 + SBS25	SBS9=0; SBS25=1	0.71986	SBS3 + SBS19	SBS3=0.87; SBS19=0.13	0.82841	SBS26 + SBS36	SBS26=0.87; SBS36=0.13	0.81735	SBS2 + SBS9	SBS2=0.07; SBS9=0.93	0.70113
SBS10a + SBS25	SBS10a=0; SBS25=1	0.71986	SBS3 + SBS18	SBS3=0.81; SBS18=0.19	0.82689	SBS12 + SBS54	SBS12=0.82; SBS54=0.18	0.81686	SBS9 + SBS30	SBS9=0.86; SBS30=0.14	0.6986
SBS10b + SBS25	SBS10b=0; SBS25=1	0.71986	SBS5 + SBS35	SBS5=0.81; SBS35=0.19	0.82575	SBS8 + SBS12	SBS8=0.25; SBS12=0.75	0.81658	SBS9 + SBS25	SBS9=0.73; SBS25=0.27	0.69554
SBS11 + SBS25	SBS11=0; SBS25=1	0.71986	SBS5 + SBS42	SBS5=0.8; SBS42=0.2	0.82529	SBS26 + SBS29	SBS26=0.84; SBS29=0.16	0.81623	SBS9 + SBS42	SBS9=0.84; SBS42=0.16	0.69521
SBS12 + SBS25	SBS12=0; SBS25=1	0.71986	SBS5 + SBS50	SBS5=0.85; SBS50=0.15	0.82463	SBS3 + SBS12	SBS3=0.35; SBS12=0.65	0.81582	SBS9 + SBS35	SBS9=0.82; SBS35=0.18	0.69516
SBS13 + SBS25	SBS13=0; SBS25=1	0.71986	SBS5 + SBS19	SBS5=0.88; SBS19=0.12	0.82431	SBS24 + SBS26	SBS24=0.17; SBS26=0.83	0.81564	SBS9 + SBS32	SBS9=0.86; SBS32=0.14	0.69422
SBS14 + SBS25	SBS14=0; SBS25=1	0.71986	SBS3 + SBS11	SBS3=0.88; SBS11=0.12	0.82371	SBS26 + SBS42	SBS26=0.84; SBS42=0.16	0.81558	SBS9 + SBS45	SBS9=0.9; SBS45=0.1	0.69358
SBS15 + SBS25	SBS15=0; SBS25=1	0.71986	SBS3 + SBS36	SBS3=0.86; SBS36=0.14	0.82323	SBS26 + SBS31	SBS26=0.85; SBS31=0.15	0.81503	SBS7a + SBS9	SBS7a=0.08; SBS9=0.92	0.69292
SBS16 + SBS25	SBS16=0; SBS25=1	0.71986	SBS3 + SBS23	SBS3=0.88; SBS23=0.12	0.82169	SBS12 + SBS35	SBS12=0.8; SBS35=0.2	0.815	SBS9 + SBS43	SBS9=0.89; SBS43=0.11	0.69124
SBS17a + SBS25	SBS17a=0; SBS25=1	0.71986	SBS8 + SBS30	SBS8=0.61; SBS30=0.39	0.82138	SBS12 + SBS26	SBS12=0.45; SBS26=0.55	0.81435	SBS9 + SBS13	SBS9=0.93; SBS13=0.07	0.69049
SBS17b + SBS25	SBS17b=0; SBS25=1	0.71986	SBS4 + SBS30	SBS4=0.59; SBS30=0.41	0.82111	SBS26 + SBS56	SBS26=0.93; SBS56=0.07	0.81104	SBS9 + SBS28	SBS9=0.92; SBS28=0.08	0.68905
SBS18 + SBS25	SBS18=0; SBS25=1	0.71986	SBS3 + SBS15	SBS3=0.89; SBS15=0.11	0.82079	SBS12 + SBS18	SBS12=0.84; SBS18=0.16	0.8105	SBS9 + SBS38	SBS9=0.94; SBS38=0.06	0.68465
SBS19 + SBS25	SBS19=0; SBS25=1	0.71986	SBS3 + SBS7b	SBS3=0.89; SBS7b=0.11	0.82037	SBS12 + SBS42	SBS12=0.84; SBS42=0.16	0.80975	SBS9 + SBS52	SBS9=0.95; SBS52=0.05	0.68463
SBS20 + SBS25	SBS20=0; SBS25=1	0.71986	SBS3 + SBS42	SBS3=0.8; SBS42=0.2	0.82002	SBS26 + SBS39	SBS26=0.82; SBS39=0.18	0.80972	SBS9 + SBS58	SBS9=0.89; SBS58=0.11	0.68384
SBS21 + SBS25	SBS21=0; SBS25=1	0.71986	SBS3 + SBS24	SBS3=0.8; SBS24=0.2	0.81946	SBS26 + SBS50	SBS26=0.89; SBS50=0.11	0.80971	SBS9 + SBS54	SBS9=0.93; SBS54=0.07	0.68267
SBS23 + SBS25	SBS23=0; SBS25=1	0.71986	SBS5 + SBS7a	SBS5=0.92; SBS7a=0.08	0.81882	SBS26 + SBS45	SBS26=0.93; SBS45=0.07	0.80947	SBS9 + SBS53	SBS9=0.94; SBS53=0.06	0.68235
SBS24 + SBS25	SBS24=0; SBS25=1	0.71986	SBS29 + SBS30	SBS29=0.53; SBS30=0.47	0.81861	SBS14 + SBS26	SBS14=0.08; SBS26=0.92	0.80904	SBS9 + SBS60	SBS9=0.97; SBS60=0.03	0.6821
SBS25 + SBS26	SBS25=1; SBS26=0	0.71986	SBS5 + SBS56	SBS5=0.93; SBS56=0.07	0.81833	SBS12 + SBS22	SBS12=0.87; SBS22=0.13	0.80897	SBS9 + SBS11	SBS9=0.94; SBS11=0.06	0.68133

Supplementary Table 12. The 96 trinucleotide frequency for SBS Mongolia.

trinucleotide change	Relative contribution (SBS Mongolia)
A[C>A]A	0.035955461
A[C>A]C	0.013847095
A[C>A]G	0.001721218
A[C>A]T	0.004252817
C[C>A]A	0.014587099
C[C>A]C	0.012362044
C[C>A]G	0.00118355
C[C>A]T	0.003425482
G[C>A]A	0.014750146
G[C>A]C	0.004370825
G[C>A]G	0.001760215
G[C>A]T	0.008717436
T[C>A]A	0.01448695
T[C>A]C	0.035744186
T[C>A]G	0.005123383
T[C>A]T	0.018882905
A[C>G]A	0.006779861
A[C>G]C	0.002203666
A[C>G]G	0.001205332
A[C>G]T	0.017434403
C[C>G]A	0.006238753
C[C>G]C	0.004975743
C[C>G]G	0.004117009
C[C>G]T	0.007098419
G[C>G]A	0.001594505
G[C>G]C	4.75875E-20
G[C>G]G	0.000473898
G[C>G]T	0.003988173
T[C>G]A	0.003795465
T[C>G]C	0.004487134
T[C>G]G	0.001583128
T[C>G]T	0.023982606
A[C>T]A	0.014945584
A[C>T]C	0.008389485
A[C>T]G	0.004551594
A[C>T]T	0.011651419
C[C>T]A	0.009402314
C[C>T]C	0.006496308
C[C>T]G	0.002564394
C[C>T]T	0.009626158
G[C>T]A	0.010533685
G[C>T]C	0.009922544
G[C>T]G	0.002524279
G[C>T]T	0.001925169
T[C>T]A	0.028736617
T[C>T]C	0.011600356
T[C>T]G	0.004565619
T[C>T]T	0.011004112
A[T>A]A	0.006363384

A[T>A]C	4.75875E-20
A[T>A]G	0.011583259
A[T>A]T	0.001046362
C[T>A]A	4.75875E-20
C[T>A]C	0.008078025
C[T>A]G	0.009598124
C[T>A]T	0.013844478
G[T>A]A	0.004196896
G[T>A]C	0.000210043
G[T>A]G	0.0048592
G[T>A]T	0.000784631
T[T>A]A	0.007414714
T[T>A]C	0.005257449
T[T>A]G	4.75875E-20
T[T>A]T	0.012157859
A[T>C]A	4.75875E-20
A[T>C]C	0.004967699
A[T>C]G	0.030530339
A[T>C]T	6.78691E-16
C[T>C]A	0.004469143
C[T>C]C	0.011343202
C[T>C]G	0.009519282
C[T>C]T	0.004018189
G[T>C]A	0.000606475
G[T>C]C	1.49375E-11
G[T>C]G	0.002786154
G[T>C]T	0.004888446
T[T>C]A	0.009366417
T[T>C]C	0.000592396
T[T>C]G	0.001751501
T[T>C]T	0.014951023
A[T>G]A	0.027779096
A[T>G]C	0.020987142
A[T>G]G	0.056400253
A[T>G]T	0.045994861
C[T>G]A	0.007120867
C[T>G]C	0.00877218
C[T>G]G	0.01775835
C[T>G]T	0.037815563
G[T>G]A	0.00780353
G[T>G]C	0.003159935
G[T>G]G	0.014371422
G[T>G]T	0.012311793
T[T>G]A	0.019668356
T[T>G]C	0.014263082
T[T>G]G	0.037929245
T[T>G]T	0.053111095

Supplementary Table 13. Number of signature occurrences above bootstrap exposure cutoff of 0.1 (at P = 0.1, one sided) in the Mongolian and Western cohorts. Differences between Western samples from different origins (Europe and USA) are also indicated. Statistical differences between cohorts were assessed by Fisher test corrected by FDR.

	Mongolian	% Mongolian	Western (all)	% Western (all)	adj. p val	Europe	% Europe	USA	% USA	adj. p val
SBS1	0	0.0%	0	0.0%	1.000	0	0.0%	0	0.0%	NA
SBS4	2	1.3%	2	1.8%	1.000	2	2.9%	0	0.0%	0.834
SBS5	10	6.6%	13	11.6%	0.515	9	13.0%	4	9.3%	0.834
SBS6	1	0.7%	0	0.0%	1.000	0	0.0%	0	0.0%	NA
SBS12	2	1.3%	0	0.0%	0.800	0	0.0%	0	0.0%	NA
SBS16	0	0.0%	2	1.8%	0.515	2	2.9%	0	0.0%	0.834
SBS18	0	0.0%	1	0.9%	0.781	0	0.0%	1	2.3%	0.834
SBS22	21	13.9%	6	5.4%	0.137	0	0.0%	6	14.0%	0.018
SBS26	3	2.0%	0	0.0%	0.580	0	0.0%	0	0.0%	NA
SBS29	1	0.7%	0	0.0%	1.000	0	0.0%	0	0.0%	NA
SBS40	49	32.5%	33	29.5%	0.944	21	30.4%	12	27.9%	0.834
SBSM	38	25.2%	5	4.5%	0.00003	4	5.8%	1	2.3%	0.834

Supplementary Table 14. Number of environmental signature occurrences above weight cutoff of 0.1 in the Mongolian and Western cohorts. Statistical differences between cohorts were assessed by Fisher test corrected by FDR.

	Mongolian	% Mongolian	Western	% Western	adj. p val
MNU_350_uM	61	40.4%	41	36.6%	0.79922
DMS_0_078_mM	57	37.7%	21	18.8%	0.041921
AFB1_0_25_uM_plus_S9	46	30.5%	24	21.4%	0.461537
DES_0_938_mM	42	27.8%	33	29.5%	0.940622
DMH_11_6_mM_plus_S9	36	23.8%	32	28.6%	0.715437
ENU_400_uM	36	23.8%	23	20.5%	0.781518
Formaldehyde_120_uM	29	19.2%	32	28.6%	0.412997
S_6_Nitrochrysene_12_5_uM_plus_S9	29	19.2%	31	27.7%	0.46342
DBP_0_0039_uM	27	17.9%	12	10.7%	0.461537
AZD7762_1_625_uM	23	15.2%	13	11.6%	0.758845
Mechlorethamine_0_3_uM	22	14.6%	21	18.8%	0.715437
Methyleugenol_1_25_mM	21	13.9%	14	12.5%	0.99728
S_1_8_DNP_8_uM	20	13.2%	10	8.9%	0.691591
AAII_37_5_uM	18	11.9%	18	16.1%	0.715437
Temozolomide_200_uM	11	7.3%	16	14.3%	0.460788
S_6_Nitrochrysene_50_uM_plus_S9	10	6.6%	2	1.8%	0.412997
S_5_Methylchrysene_1_6_uM_plus_S9	9	6.0%	5	4.5%	0.940622
Furan_100_mM_plus_S9	7	4.6%	3	2.7%	0.781518
DBPDE_0_000625_uM	7	4.6%	1	0.9%	0.46342
MX_7_uM_plus_S9	5	3.3%	11	9.8%	0.372154
OTA_0_08_uM_plus_S9	5	3.3%	6	5.4%	0.781518
AAI_1_25_uM	4	2.6%	4	3.6%	0.924609
S_1_8_DNP_0_125_uM	2	1.3%	1	0.9%	1
SSR_1_25_J	2	1.3%	10	8.9%	0.108054
S_3_NBA_0_1_uM	2	1.3%	2	1.8%	1
DBPDE_0_000156_uM	1	0.7%	0	0.0%	1
DBP_0_0313_uM_plus_S9	1	0.7%	0	0.0%	1
Benzidine_200_uM	1	0.7%	2	1.8%	0.781518
Cyclophosphamide_18_75_uM_plus_S9	1	0.7%	3	2.7%	0.691591
Potassium_bromate_260_uM	1	0.7%	3	2.7%	0.691591
S_6_Nitrochrysene_0_78_uM	1	0.7%	0	0.0%	1
S_1_6_DNP_0_09_uM	1	0.7%	2	1.8%	0.781518
N_Nitrosopyrrolidine_50_mM	1	0.7%	0	0.0%	1
Cisplatin_3_125_uM	1	0.7%	3	2.7%	0.691591
Carboplatin_5_uM	1	0.7%	6	5.4%	0.372154
S_6_Nitrochrysene_50_uM	1	0.7%	4	3.6%	0.473602
DBADE_0_109_uM	0	0.0%	0	0.0%	1
Potassium_bromate_875_uM	0	0.0%	0	0.0%	1
S_4_ABP_300_uM_plus_S9	0	0.0%	5	4.5%	0.186098
Semustine_150_uM	0	0.0%	2	1.8%	0.473602
DBADE_0_0313_uM	0	0.0%	0	0.0%	1
BPDE_0_125_uM	0	0.0%	0	0.0%	1
BaP_0_39_uM_plus_S9	0	0.0%	0	0.0%	1
Ellipticine_0_375_uM_plus_S9	0	0.0%	3	2.7%	0.412997
DBA_75_uM_plus_S9	0	0.0%	0	0.0%	1
PhIP_3_uM_plus_S9	0	0.0%	0	0.0%	1
S_3_NBA_0_025_uM	0	0.0%	0	0.0%	1
Propylene_oxide_10_mM	0	0.0%	2	1.8%	0.473602
DBAC_5_uM_plus_S9	0	0.0%	0	0.0%	1
Cisplatin_12_5_uM	0	0.0%	1	0.9%	0.715437
PhIP_4_uM_plus_S9	0	0.0%	1	0.9%	0.715437
BaP_2_uM_plus_S9	0	0.0%	0	0.0%	1

Supplementary Table 15. Clinico-pathological characteristics of the Mongolian gene expression-based clusters in the Mongolian cohort.

	MGL1 (n=47, 44%)	MGL2 (n=27, 26%)	MGL3 (n=32, 30%)	p value
Age (years) <60 years (n, %)	62 (43-76) 18 (39.1)	56 (44-72) 20 (76.9)	56 (41-75) 19 (63.3)	0.028 0.005
Gender (male, %)	29 (63)	8 (30.8)	19 (63.3)	0.017
BMI > 25 kg/m^2 (n, %)	26 (56.5)	12 (46.2)	11 (39.3)	ns
Etiology:				
• HBV (n, %)	4 (8.5)	0 (0)	0 (0)	
• HBV/HDV (n, %)	27 (57.4)	19 (70.4)	26 (81.3)	<0.001
• HBV/HCV/HDV (n, %)	1 (2.1)	6 (22.2)	5 (15.6)	
• HCV (n, %)	15 (31.9)	2 (7.4)	1 (3.1)	
Region				
• Western (n, %)	9 (20.5)	6 (26.1)	4 (14.8)	
• Central (n, %)	18 (40.9)	9 (39.1)	13 (48.1)	ns
• Eastern (n, %)	1 (2.3)	3 (13)	3 (11.1)	
• Ulaanbaatar (n, %)	16 (36.4)	5 (21.7)	7 (25.9)	
BCLC stage (0-A, %)	35 (77.8)	17 (68)	24 (85.7)	ns
AFP >400 IU/mL (n, %)	5 (13.5)	9 (42.9)	4 (16.7)	0.026
Liver fibrosis (F3-4, %)	17 (36.2)	12 (44.4)	12 (37.5)	ns
Microvascular invasion (yes, %)	16 (41)	13 (65)	9 (34.6)	0.099
Tumor grade (G3-4, %)	4 (10)	4 (26.7)	0 (0)	0.045

BMI, body mass index; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; BCLC, Barcelona Clinic Liver Cancer, AFP, alfa-fetoprotein

Some variables have missing values for MGL1, MGL2 and MGL3, respectively:

Age, gender: 1, 1, and 2 patients. BMI: 1, 1, and 4 patients. Region: 3, 4, and 5 patients. BCLC stage: 2, 2, and 4 patients. AFP: 10, 6, and 8 patients. Microvascular invasion: 8, 7, and 8 patients. Tumor grade: 7, 12, and 16 patients.

Supplementary Table 16. Clinico-pathological characteristics of the Mongolian gene expression-based clusters in the Mongolian NCI cohort.

	MGL1 (n=34, 49%)	MGL2 (n=18, 26%)	MGL3 (n=18, 26%)	p value
Age (years) <60 years old (n, %)	62 (41-76) 13 (38.2)	56.5 (45-77) 10 (55.6)	59 (23-68) 9 (50.0)	0.19 0.546
Gender (male, %)	21 (61.7)	8 (44.4)	8 (44.4)	0.45
BMI > 25 kg/m2 (n, %)	13 (38.2)	6 (33.3)	7 (38.9)	0.83
Viral infection:				
HBV + (n, %)	22 (64.7)	14 (77.8)	13 (72.2)	0.804
HDV + (n, %)	14 (41.2)	5 (27.8)	8 (44.4)	0.626
HCV + (n, %)	23 (67.6)	10 (55.6)	8 (44.4)	0.473
AFP >20 IU/ml (n, %)	10 (29.4)	10 (55.6)	5 (27.8)	0.023
Liver cirrhosis (F4, %)	14 (41.2)	9 (50.0)	5 (27.8)	0.135

BMI, body mass index; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; AFP, alfa-fetoprotein

Supplementary Table 17. HCC specific COSMIC v3 signatures, along with its prevalence across a total of 493 HCC samples and its cosine similarity against COSMIC v2 version. Note that SBS16 changes greatly from v2 to v3.

COSMIC v3 signature	HCC Sample Number	Cosine Similarity against COSMIC v2
SBS1	303	0.95
SBS3	35	0.96
SBS4	88	0.94
SBS5	491	0.96
SBS6	5	0.95
SBS9	4	0.98
SBS12	214	0.94
SBS16	83	0.79
SBS17a	4	NA
SBS17b	6	NA
SBS18	45	0.99
SBS19	5	0.89
SBS22	51	0.96
SBS23	3	1
SBS24	19	0.94
SBS26	2	0.92
SBS28	2	0.92
SBS29	139	0.97
SBS30	3	0.96
SBS31	1	NA
SBS35	19	NA
SBS37	18	NA
SBS40	45	NA

Supplementary Table 18. Publicly available gene sets and signatures used in the study.

Name	Reference
HCC classification	
Sia HCC immune class	Sia D, et al. Gastroenterology 2017;153:812-826
Chiang classification	Chiang D, et al. Cancer Res 2008;68:6779-6788
Hoshida classification	Hoshida Y, et al. Cancer Res 2009;69:7385-7392
Cluster A signature	Lee JS, et al. Hepatology 2004;40:667–76
Molecular pathways	
TGFB late signature	Coulouarn C, et al. Hepatology 2008: 47 2059-2067
MET signature	Kaposi-Novak P, et al. J Clin Invest 2006;116:1582-95
NOTCH signature	Villanueva A, et al. Gastroenterology 2012;143: 1660-1669
RB1 signature	Bollard J, et al. Gut 2017;66: 1286-1296
Antitumor immune response	
PD1 signaling signature	Quigley M, et al. Nat Med 2010;16:1147-51
Exhaustion signature	Quigley M, et al. Nat Med 2010;16:1147-51
Stromal enrichment score	Yoshihara K, et al. Nat Commun 2013;4:2612
Immune enrichment score	Yoshihara K, et al. Nat Commun 2013;4:2612
IFN signature	Ayers M, et al. J Clin Invest 2017; 127:2930–40
Immune cell infiltrate gene sets	Bindea G, et al. Immunity 2013;39:782-95

Supplementary Data Study 2

Figure S1
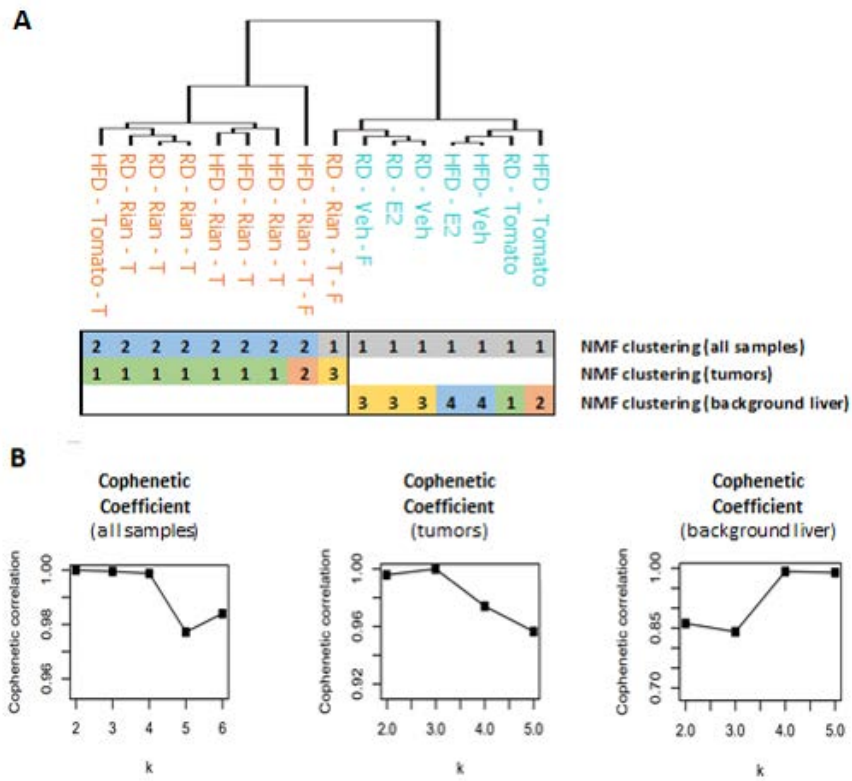
Figure S1. Unsupervised clustering analysis. Tumor samples and background liver samples clustered together but in separate branches. Female tumors clustered with the background liver even though they had the Rian signature. **(A)** Hierarchical clustering and non-negative matrix factorization (NMF) from all the samples and the samples separated by tumor or adjacent tissue. **(B)** Cophenetic coefficient of the NFM clustering from all samples and tumor or background liver samples alone.

Figure S2
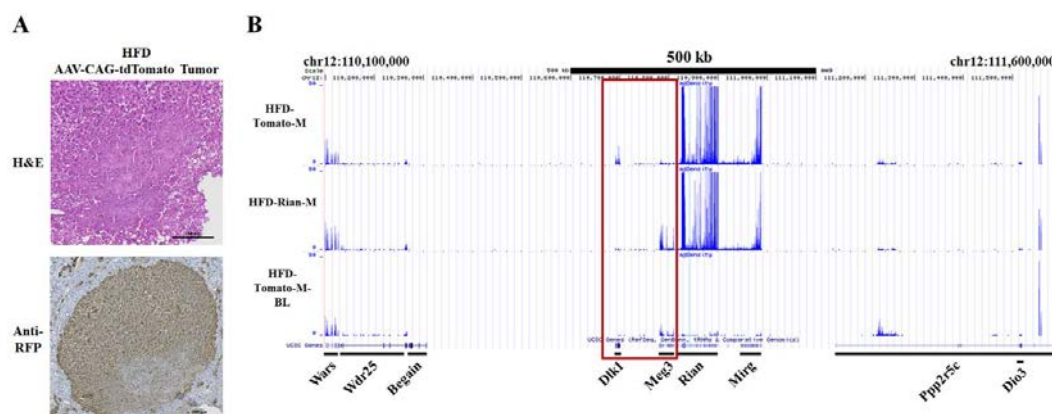
Figure S2. Tomato positive tumor exhibit *Rian* signature. (A) Histology of tumor section

from a mouse on HFD that received the AAV-tdTomato virus. Brown stains in the anti-RFP

panel shows tdTomato positive hepatocytes. Scale bars = 200 µM. **(B)** Expression levels of

mRNA transcripts near the vector insertion site. Chromosome position is shown the x-axis

chr12:110,100,000-111,600,000 and tag density is on the y-axis. TdTomato tumor has a slightly

different expression pattern near the integration site suggesting random integration in the Rian

locus.

Table S1: Summary of different treatment groups used in the neonatal experimental paradigm. Neonatal mice were infected with rAAV, started on HFD or RD at 3 weeks of age and euthanized at 6 months of age.

Vector	Diet	# of Males	# of Females	Total
AAV-CAG-tdTomato	HFD	4	5	9
AAV-CAG-tdTomato	RD	3	3	6
AAV-Rian-CMV	HFD	14	5	19
AAV-Rian-CMV	RD	7	8	15
No AAV	HFD	3	3	6

Table S2: Summary of different treatment groups used in the adult experimental paradigm. 3 weeks old mice were started on HFD or RD, injected with rAAV at 10 weeks of age, and euthanized at 9 months.

Vector	Diet	# of mice (Male)
AAV-CAG-tdTomato	HFD	10
AAV-CAG-tdTomato	RD	5
AAV-Rian-CMV	HFD	10
AAV-Rian-CMV	RD	20
AAV-Rian-CMV + 2/3 PH	RD	5
No AAV	HFD	5

Table S3. Summary of the samples used for RNAseq.

	HCC Tumor samples							
Virus inoculated	AAV-Rian				AAV-Tomato (control)			
Gender	Male		Female		Male		Female	
Diet	RD	HFD	RD	HFD	RD	HFD	RD	HFD
n of samples	3	3	1	1	-	1	-	-
Short name used in the figures	RD-Rian-T	HFD-Rian-T	RD-Rian-T-F	HFD-Rian-T-F		HFD-Tomato-T		

	Background liver samples															
Virus inoculated	N/A								AAV-Tomato (control)							
Treatment	Estrogen				Vehicle				N/A				N/A			
Gender	Male		Female		Male		Female		Male		Female		Male		Female	
Diet	RD	HFD	RD	HFD	RD	HFD	RD	HFD	RD	HFD	RD	HFD	RD	HFD	RD	HFD
n of samples	1	1	-	-	1	1	1	-	1	1	-	-	-	-	-	-
Short name used in the figures	RD-E2	HFD-E2			RD-Veh	HFD-Veh	RD-Veh-F		RD-Tomato	HFD-Tomato						

Supplemental materials and methods:

Murine samples:

Neonatal male and female mice were infected with an AAV targeting the Rian locus (AAV-Rian) or a control AAV (AAV-Tomato), fed with RD or HFD for six months, and three weeks old mice on fed HFD or RD and treated with estrogen or vehicle for one month. Table below summarizes the samples used for RNAseq. All background liver samples correspond to a pooled mix of 3 samples sequenced together. Tumor samples were sequenced individually.

Unsupervised cluster analysis:

RNA-seq data were filtered to remove all the genes that were not expressed in any of the samples (FPKM=0). Unsupervised clustering of the samples was performed using non-negative matrix factorization (NMF) and hierarchical clustering GenePattern modules.[1, 2] For the NMF, the k with greater cophenetic coefficient was used for each analysis.

Gene expression profile analysis and generation of an AAV-Rian gene signature:

Murine gene expression data were humanized using the mouse-human orthologues extracted from the BioMart database through BiomaRt R package.[3] Next, class prediction using previously published HCC molecular classifications was conducted by Nearest Template Prediction (NTP) analysis (Gene Pattern modules).[4] The evaluation of the enrichment of distinct molecular pathways and gene expression signatures was performed using single-sample Gene Set Enrichment Analysis (ssGSEA). To this end, Molecular Signature Database gene sets (MSigDB, www.broadinstitute.org/msigdb) and previously reported gene-expression signatures representing different states of inflammation and liver function were tested (Table S3).[5]

To evaluate the similarity to the tumors developed in Wang et al. 2012, a gene signature was generated using the top differentially expressed genes in Wang et al. 2012 tumors compared to healthy liver (AAV-Rian).[6] The generated gene signature (from now on called AAV-Rian

signature) was composed of 199 up-regulated genes (AAV-Rian-UP) and 100 down-regulated genes (AAV-Rian-DOWN). ssGSEA and NTP were used to assess similarity between tumors reported in this report and AAV-Rian signature.

In addition, to further evaluate the expression of the genes located in the Rian Locus, the "Rian-Locus" geneset was generated including all the genes in the murine genome region chr12:108860000-1104180000, where the Rian Locus is located (genome of reference GRCm38) (Table S2). The Rian-Locus geneset and the Wang gene signature (AAV-Rian) were evaluated in non-humanized gene expression data. Differentially expressed genes between groups of samples were identified using the Limma R package[11] (FDR<0.05 and Fold-change [FC]$\geq$ 2 or $\leq$0.5).[7]

Supplemental references:

1. Brunet, JP, Tamayo, P, Golub, TR, and Mesirov, JP (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* **101**: 4164-4169.
2. Eisen, MB, Spellman, PT, Brown, PO, and Botstein, D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**: 14863-14868.
3. Smedley, D, Haider, S, Ballester, B, Holland, R, London, D, Thorisson, G, *et al.* (2009). BioMart--biological queries made easy. *BMC Genomics* **10**: 22.
4. Reich, M, Liefeld, T, Gould, J, Lerner, J, Tamayo, P, and Mesirov, JP (2006). GenePattern 2.0. *Nat Genet* **38**: 500-501.
5. Llovet, JM, Montal, R, Sia, D, and Finn, RS (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* **15**: 599-616.
6. Wang, PR, Xu, M, Toffanin, S, Li, Y, Llovet, JM, and Russell, DW (2012). Induction of hepatocellular carcinoma by in vivo gene targeting. *Proc Natl Acad Sci U S A* **109**: 11264-11269.
7. Smyth, G, Gentleman, R, Carey, V, Dudoit, S, Irizarry, R, and Huber, W (2005). Limma: linear models for microarray data, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. *Springer*: 397-420.

Supplementary Data Study 3

SUPPLEMENTARY MATERIALS AND METHODS

Treatment strategy and sample collection

Lenvatinib (Eisai, Ibaraki, Japan) and anti-PD1 (anti-murine PD-1 mab clone J43 BioXCell BE0033-2) treatments were administered in accordance to current bibliography and provider's recommendations [1–4]. Lenvatinib was diluted in distilled water, while anti-PD1 and IgG were diluted in InVivoPure dilution buffer (BioXCell IP0065) in agreement with manufacturer's recommendations. Lenvatinib (10 mg/kg) was administered daily by oral gavage [1,2], and anti-PD1 (10 mg/kg) was intraperitoneally administered every 3 days for a total of 5 doses [4]. Sample size calculation was based on power analysis and on our previous studies using similar models [5].

In order to analyze the mechanisms of action of the drugs, 5 mice per arm from the Hepa1-6 model were euthanized 1 day after the last anti-PD1 dose (day 13 post-randomization, early timepoint) and 2 hours after the last lenvatinib administration, in accordance with pharmacodynamic studies [6–9]. Tumors were weighed and samples were collected for molecular analyses and immune population characterization. Blood samples were collected by cardiac puncture from mice under deep terminal anesthesia. The remaining mice were monitored until a tumor volume of 1000 mm3 was reached or until study termination at day 125 post-randomization after sacrificing the last placebo animal (late timepoint). Tumor samples were collected, and tumor growth and survival, defined as time to reach 1000 mm3 tumor volume, were measured. Response to the treatment was measured according to the percentage change thresholds of the RECIST criteria adapted for murine models [10–12]. Time to objective response was defined as time to achieve a 30% decrease in tumor volume [11]. Animals developing tumor ulcers were censored from the survival and response analysis at the time of sacrifice (placebo n=1, combination n=1). Potential treatment-related toxicity was evaluated by monitoring body weight.

The Hep53.4 models were used as validation and all animals were sacrificed at day 13 post-randomization. Tumor samples were harvested and response to treatment was evaluated as described above.

Assessment of potential gender effect in the syngeneic model

The syngeneic mouse models of HCC were used for this study in accordance with current bibliography supporting the use of this model to investigate the effect of immunotherapy and combination treatments [13]. Additional experiments in both male and female mice were conducted to determine whether there are gender differences in the subcutaneous HCC model. Specifically, the Hepa1-6 model was generated in male and female mice (n=30) by injecting 5x106 Hepa1-6 cells in 100ml PBS in the right flank of in 5-6-week old C57BL/6J mice (Charles River Laboratories). Animals were sacrificed once a tumor volume of 1000 mm3 was reached or at study termination (40 days post-injection). No differences were observed in terms of tumor penetrance, time to tumor onset, tumor growth or survival (Supplementary Figure 1A-D). Histological analysis in haematoxylin and eosin slides did not reveal differences in histopathological features or immune infiltrate (Supplementary Figure 1E). Considering this, we discard any effect of mice gender in the animal model presented in the study.

Sample processing for flow cytometry analysis

Tumor and blood samples from the Hepa1-6 model were collected to perform flow cytometry analysis. Tumor samples were minced and digested using Accumax cell detachment solution (Innovative Cell Technologies) for 45 min at 37°C and cells were filtered using a 70 µm nylon cell strainer. Blood samples were collected in EDTA-coated tubes and peripheral blood leukocytes were isolated after red blood cell lysis using RBC lysis buffer (eBioscience). Cells from tumor and blood samples were then washed, filtered and stained according to standard flow cytometer protocols (Supplementary Table 1,2).

ANNEX B

Immunohistochemistry staining of tumor samples

Only tumors with sufficient tumoral material were processed for histological analysis. From the Hepa1-6 model, 14 and 17 samples were analyzed at the early and late timepoints, respectively. Samples from the early timepoint included 4 placebo, 4 lenvatinib, 3 anti-PD1 and 3 combination. Samples from the late timepoint included 8 placebo, 4 lenvatinib, 2 anti-PD1 and 3 combination. 6 tumors from the Hep53.4 subcutaneous model were used for validation.

The primary antibodies used were CD3 (Abcam ab16669), CD4 (Abcam ab183685), CD8 (Abcam ab203035), FOXP3 (Abcam ab215206), CTLA4 (Abcam ab237712), PD1 (Sino Biological INC 50124-RP02), PDL1 (Novus Biologicals MAB90782), CD31 (Abcam ab28364), CD163 (Abcam ab182422) and CD68 (Abcam ab125212) (Supplementary Table 2, Supplementary Table 3). All immunohistochemical (IHC) stainings were carried out on 3 μm-thick FFPE tissue sections after heat-induced antigen retrieval.

CTLA4 and FOXP3 staining was quantified by the digital pathology imaging software QuPath (version 0.2.0) [14]. To ensure representative sampling of the entire tumor, 3 to 5 regions of interest (ROI) were analyzed. Each ROI had a magnification of 200X and each slide contained an average of 2700 cells in total. The number of positive cells per total cells was determined for each ROI.

FOXP3 was used to assess the presence and distribution of Regulatory T-cells (T_{reg}). Positivity for the immune checkpoints PD1, PDL1 and CTLA4 was also analyzed. To detect angiogenesis, CD31 positivity and presence of vessels encapsulating tumor clusters (VETC) were analyzed [15]. Interaction between T_{reg} cells and tumor vasculature was assessed by double staining of FOXP3 and CD31. the interaction of FOXP3 and CD31 was evaluated following the ensuing criteria: a) when FOXP3 positive cells were localized close to a well-formed vessel the interaction was considered marked; b) when FOXP3 was localized close to vessel sprouts the interaction was considered moderate; c) when FOXP3 positive cells were neither close to a well-formed vessel nor to a vessel sprout, the interaction was considered minimal; and d) when there was no FOXP3 staining, the interaction was considered absent.

Finally, the M2 per total macrophage ratio was assessed by measuring CD68 and CD163 markers in an automatized IHC stainer (Autostainer 48 AS48030, Agilent).

RNA preparation and transcriptome analysis

Tumor samples from each treatment arm in the Hepa1-6 model collected at the early timepoint or from mice reaching the survival endpoint at day 13 post-randomization were included (n=21). Samples were cut, collected in RNAlater solution and then stored at -80ºC. Total RNA was extracted from 30 mg of tissue using Trizol reagent (Invitrogen) and purified with RNeasy columns (Qiagen, Valencia, CA). RNA sample concentration and quality were assessed by NanoDrop ND-1000 spectrometer (NanoDrop, Wilmington, DE) and bioanalyzer (Agilent, Palo Alto, CA), respectively. Gene expression microarray studies were conducted using the Clariom S Mouse Array (Affymetrix, Santa Clara, CA). The raw CEL files were background corrected and normalized using the Robust Multiarray Averaging (RMA) procedure using the oligo R package [16]. To analyze the gene expression profile of the tumors, intensity values were log transformed and mice genes were humanized. Briefly, mouse-human orthologues were obtained from the BioMart database through the BiomaRt R package [17] and expression values for each human gene were calculated using the CollapseDataset module from GenePattern.

The gene expression profile was analyzed using the NTP, GSEA and ssGSEA modules from GenePattern [18] (gene sets available in MSigDB or previously reported [19,20], Supplementary Table 10). For GSEA analysis, only gene sets with enrichment score >1.5 were analyzed. Differentially expressed genes between treatment arms were identified (Bonferroni p<0.05, fold change [FC] >1.5) and pathway analysis was performed using the DAVID functional annotation tool. The ESTIMATE score of stromal infiltration and relative tumor purity was assessed using the ESTIMATE R package [21].

Identification of potential responders to combination treatment in independent human cohorts

Gene expression data from a cohort of 228 surgically resected fresh-frozen HCC samples and 169 paired non-tumoral samples (Heptromic dataset, GSE63898) was analyzed. Full descriptions of the cohort and RNA profiling data are available in previous publications [22,23]. The upregulation and downregulation of the combination rescue signature and HCC immune signature [22] were assessed in tumor samples by NTP to classify patients as combination-only responder class, potential responders to ICI or rest. The gene expression profile of each group was assessed using the NTP, GSEA and ssGSEA modules from GenePattern. NTP predictions have been classified as significant using nominal p-value <0.05 and FDR <0.1 for the combination rescue signature and using FDR<0.05 for the remaining signatures.

The relative fraction of immune cells in the tumor tissue was estimated using CIBERSORT tool [24]. Similarities between murine and human HCC tumors were assessed by principal component analysis (PCA) and sub-map analysis.

SUPPLEMENTARY REFERENCES

1. Kato Y, Tabata K, Kimura T, et al. Lenvatinib plus anti-PD-1 antibody combination treatment activates CD8 + T cells through reduction of tumor-associated macrophage and activation of the interferon pathway. PLoS One. 2019;14:e0212513.
2. Kimura T, Kato Y, Ozawa Y, et al. Immunomodulatory activity of lenvatinib contributes to antitumor activity in the Hepa1-6 hepatocellular carcinoma model. Cancer Sci. 2018;109:3993–4002.
3. European Medicines Agency. Keytruda Assessment Report. Procedure No. EMEA/H/C/003820/0000. 2015.
4. Böttcher JP, Reis e Sousa C. The Role of Type 1 Conventional Dendritic Cells in Cancer Immunity. Trends in Cancer. 2018;4:784–92.
5. Martinez-Quetglas I, Pinyol R, Dauch D, et al. IGF2 Is Up-regulated by Epigenetic Mechanisms in Hepatocellular Carcinomas and Is an Actionable Oncogene Product in Experimental Models. Gastroenterology. 2016;151:1192–205.
6. Lindauer A, Valiathan CR, Mehta K, et al. Translational Pharmacokinetic/Pharmacodynamic Modeling of Tumor Growth Inhibition Supports Dose-Range Selection of the Anti-PD-1 Antibody Pembrolizumab. CPT pharmacometrics Syst Pharmacol. 2017;6:11–20.
7. Okamoto K, Kodama K, Takase K, et al. Antitumor activities of the targeted multi-tyrosine kinase inhibitor lenvatinib (E7080) against RET gene fusion-driven tumor models. Cancer Lett. 2013;340:97–103.
8. Kato Y, Bao X, Macgrath S, et al. Lenvatinib mesilate (LEN) enhanced antitumor activity of a PD-1 blockade agent by potentiating Th1 immune response. Ann Oncol. 2016;27:(suppl 6).
9. Kato Y, Tabata K, Hori Y, et al. Effects of lenvatinib on tumor-associated macrophages enhance antitumor activity of PD-1 signal Inhibitors. 2015;14:(Suppl 2).
10. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228–47.
11. Gao H, Korn JM, Ferretti S, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med. 2015;21:1318–25.
12. Byrne AT, Alférez DG, Amant F, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. Nat Rev Cancer. 2017;17:254–68.
13. Brown ZJ, Heinrich B, Greten TF. Mouse models of hepatocellular carcinoma: an overview and highlights for immunotherapy research. Nat Rev Gastroenterol Hepatol. 2018;15:536–54.
14. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7:16878.
15. Itoh S, Yoshizumi T, Yugawa K, et al. Impact of Immune Response on Outcomes in Hepatocellular Carcinoma: Association with Vascular Formation. Hepatology. 2020;
16. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249–64.

17. Smedley D, Haider S, Ballester B, et al. BioMart - Biological queries made easy. BMC Genomics. 2009;10:22.
18. Reich M, Liefeld T, Gould J, et al. GenePattern 2.0. Nat Genet. 2006;38:500–1.
19. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013;39:782–95.
20. Jerby-Arnon L, Shah P, Cuoco MS, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell. 2018;4:984–97.
21. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
22. Sia D, Jiao Y, Martinez-Quetglas I, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology. 2017;153:812–26.
23. Villanueva A, Portela A, Sayols S, et al. DNA Methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61:1945–56.
24. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.

A

Time to tumor onset

n.s.

B

Tumor growth

Gender — Male — Female

C

	p value (Mann-Whitney)	p corrected (bonferroni)
Day 1	0,6830	1
Day 2	0,6516	1
Day 3	0,5907	1
Day 4	0,3536	1
Day 5	0,4773	1
Day 6	0,5907	1
Day 7	0,8132	1
Day 8	0,9145	1
Day 9	0,9486	1
Day 10	0,7800	1
Day 11	0,8467	1
Day 12	0,6516	1
Day 13	0,6209	1
Day 14	0,2898	1
Day 15	0,1861	1

D

Survival (Time to 1000 mm^3)

$p = 0.95$

Gender + Female + Male

E
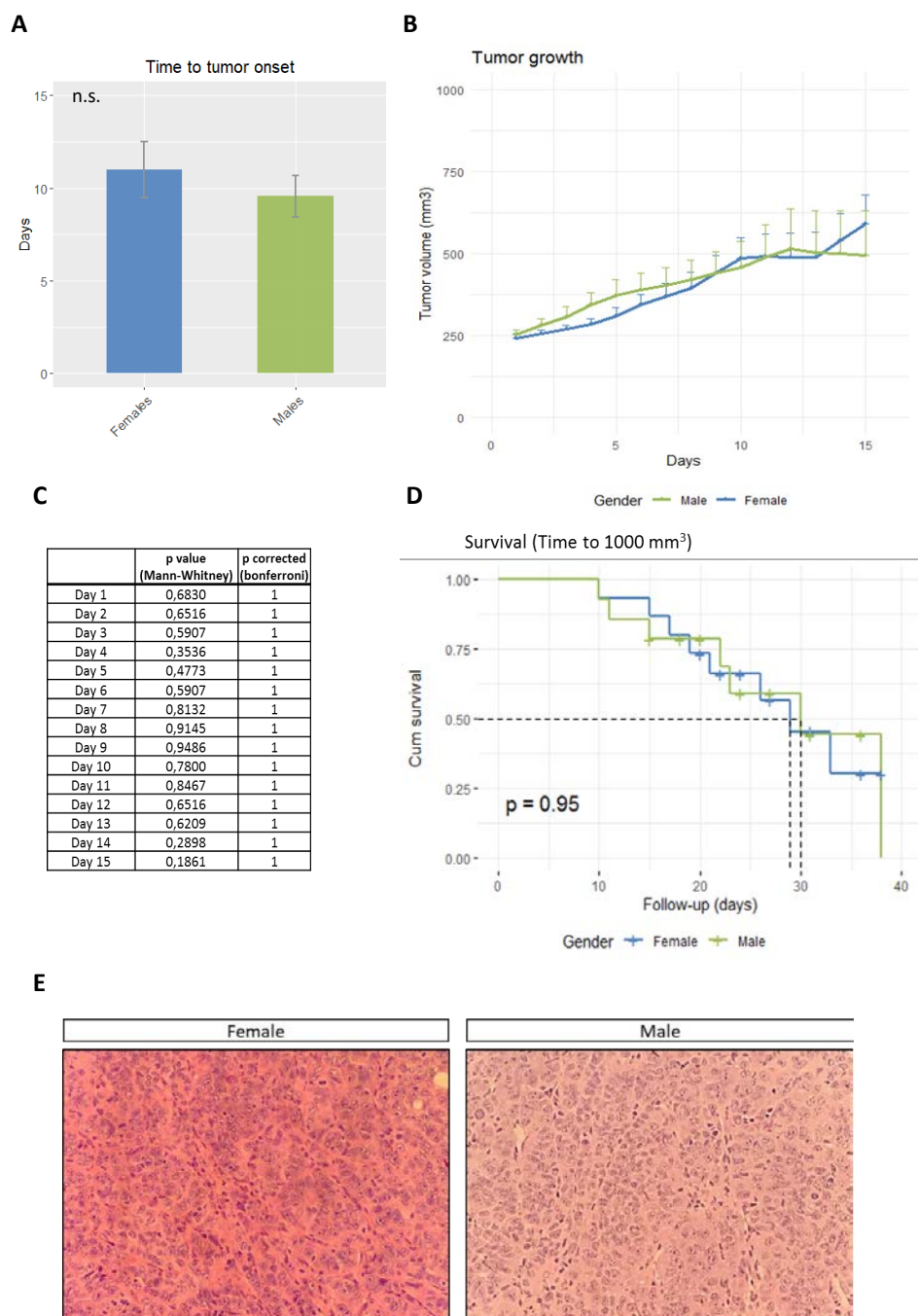
Female | Male

Supplementary Fig. 1. Comparison of male and female subcutaneous syngeneic HCC models. (A) Time to randomization (time to reach 200 mm^3) after injection of the cells. Plots represent mean values plus standard error. **(B)** Mean tumor growth after randomization in males vs females and **(C)** p-values as determined by Mann-Whitney, with Bonferroni corrected p-values also displayed. **(D)** Kaplan-Meier curve showing survival (time to reach 1000 mm3) between male and female mice. Mice sacrificed at study termination (40 days post tumor injection) are represented as censored. **(E)** Histopathological characteristics of male and female syngeneic murine tumors. Representative images of H&E stained tumors captured with 40X magnification.
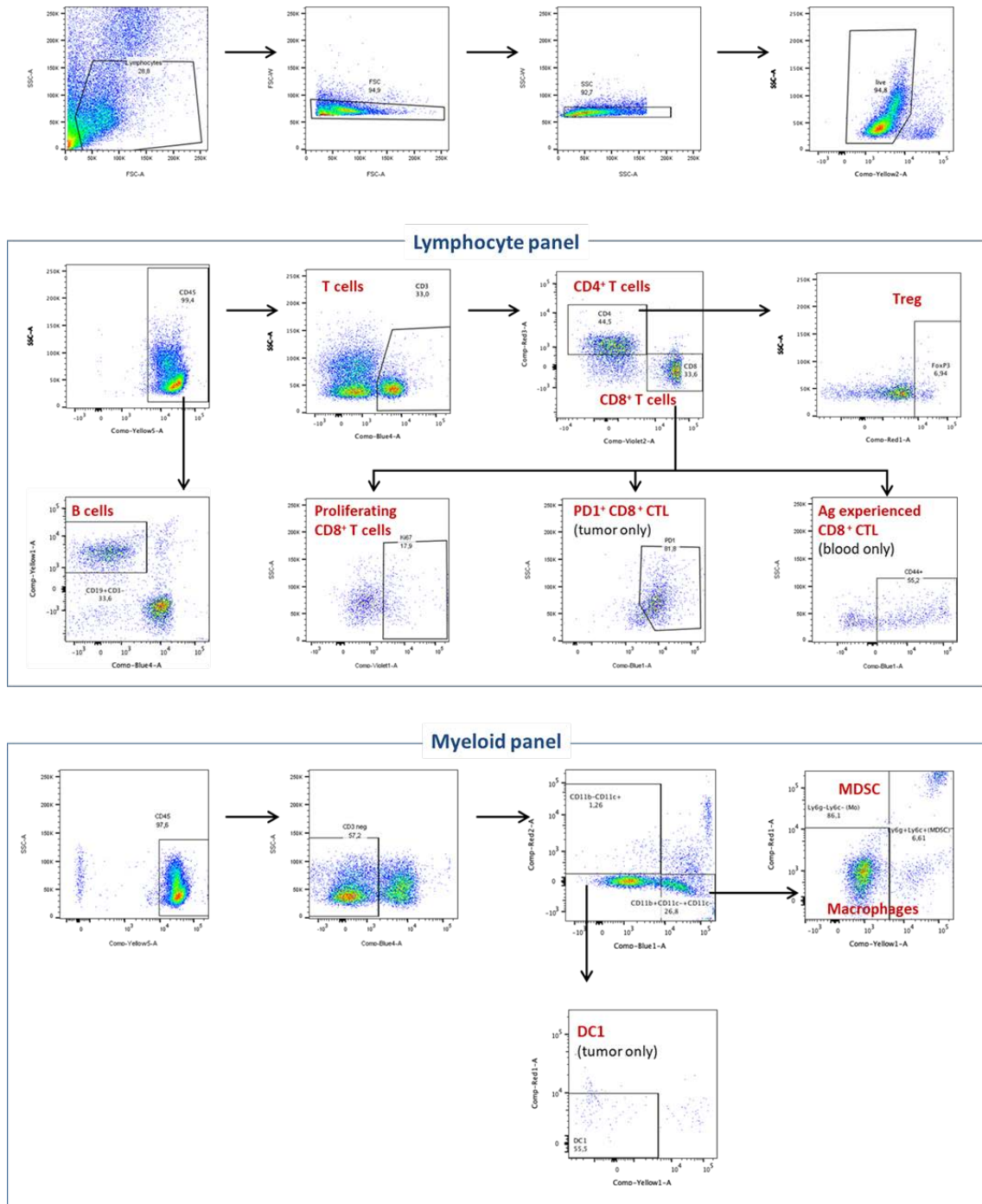
Supplementary Fig. 2. Gating strategy for the flow cytometric analysis in myeloid and T cell panels.
Immune cell populations analyzed by flow cytometry in the lymphoid and myeloid antibody panels.
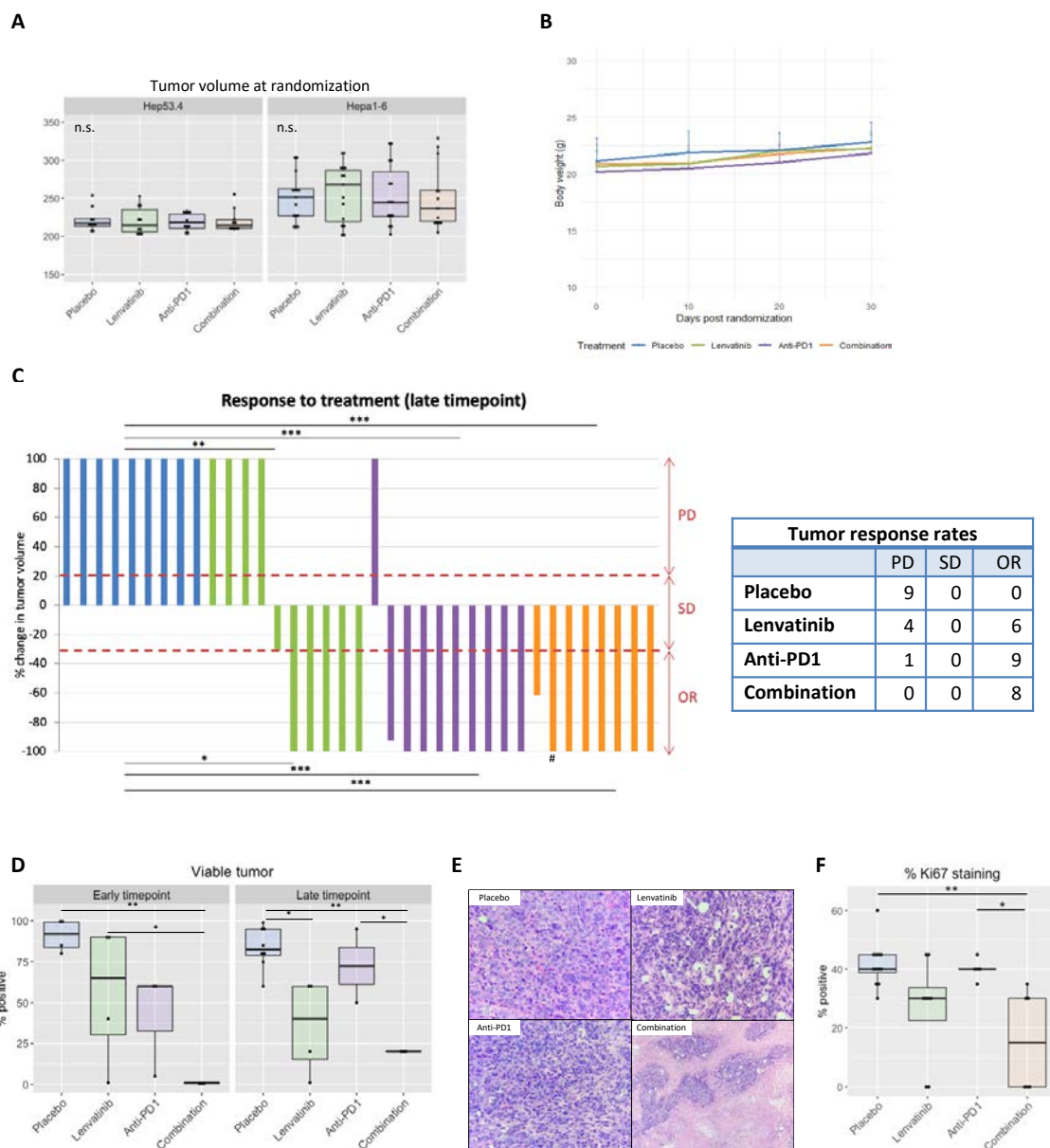
Supplementary Fig. 3. Antitumoral effect of lenvatinib plus anti-PD1. (A) Tumor volume at randomization in the subcutaneous models. **(B)** Body weight monitoring in treated mice from the Hepa1-6 model. (**C**) Response to treatment at late timepoint (n=37). Upper part indicates differences in progressive disease rate and lower part, differences in objective response rate. Right table shows number of mice per group and response. # indicates animal with no measurable tumor but with tumor tissue found in the necropsy. **(D)** Tumor viability assessed in H&E slides at the early and late timepoints. **(E)** Representative images of tumor viability in each treatment arm. Images were captured with 20X magnification. **(F)** Percentage of positive staining for Ki67 in tumors from animals in each treatment arm. Boxplots indicate median and quartiles. *p<0.05, **p<0.01, ***p<0.001 vs placebo. PD, progressive disease, SD, stable disease, OR, objective response.

A

Response to treatment
Hep53.4 subcutaneous model

Tumor response rates			
	PD	SD	OR
Placebo	10	0	0
Lenvatinib	0	8	2
Anti-PD1	0	6	4
Combination	0	0	10

B

Change in bioluminescence
Hep53.4 orthotopic model
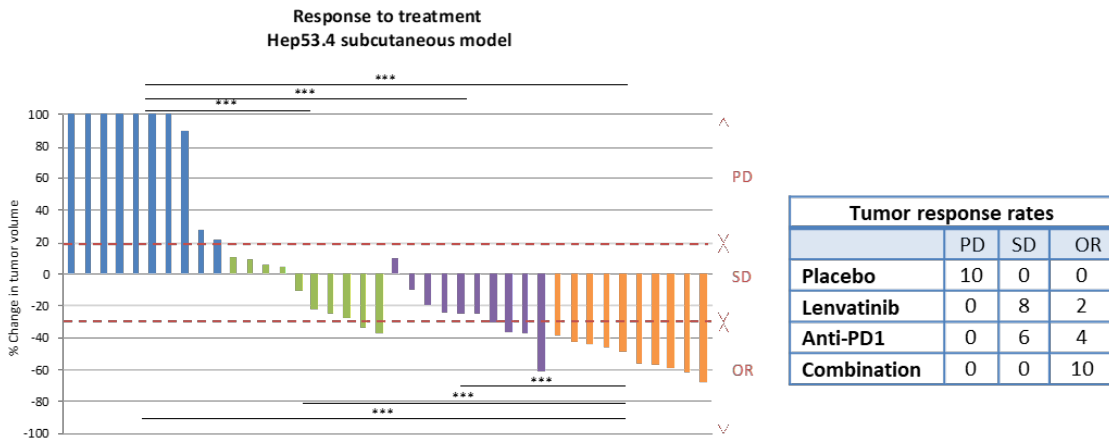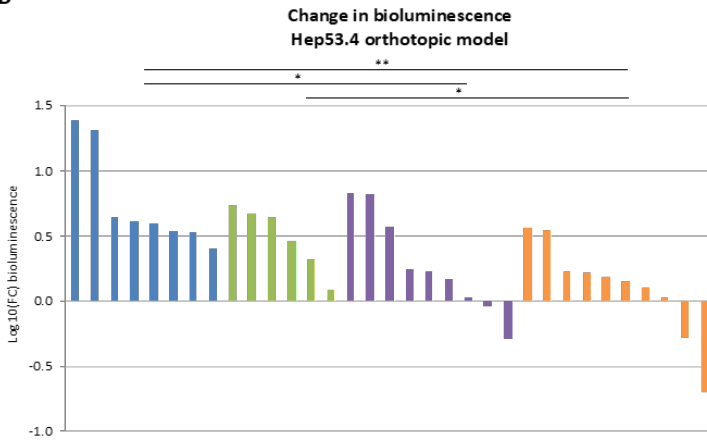
Supplementary Fig. 4. Response to treatment in the Hep53.4 models. (A) Response to treatment in the Hep53.4 subcutaneous model (n = 40). Upper part indicates differences in progressive disease rate and lower part, differences in objective response rate. Right table shows the number of mice per group and response. **(B)** Change in bioluminescence in the orthotopic model at day 12 post-randomization. * $p<0.05$, **$p<0.01$, ***$p < 0.001$. PD, progressive disease, SD, stable disease, OR, objective response.

A

B
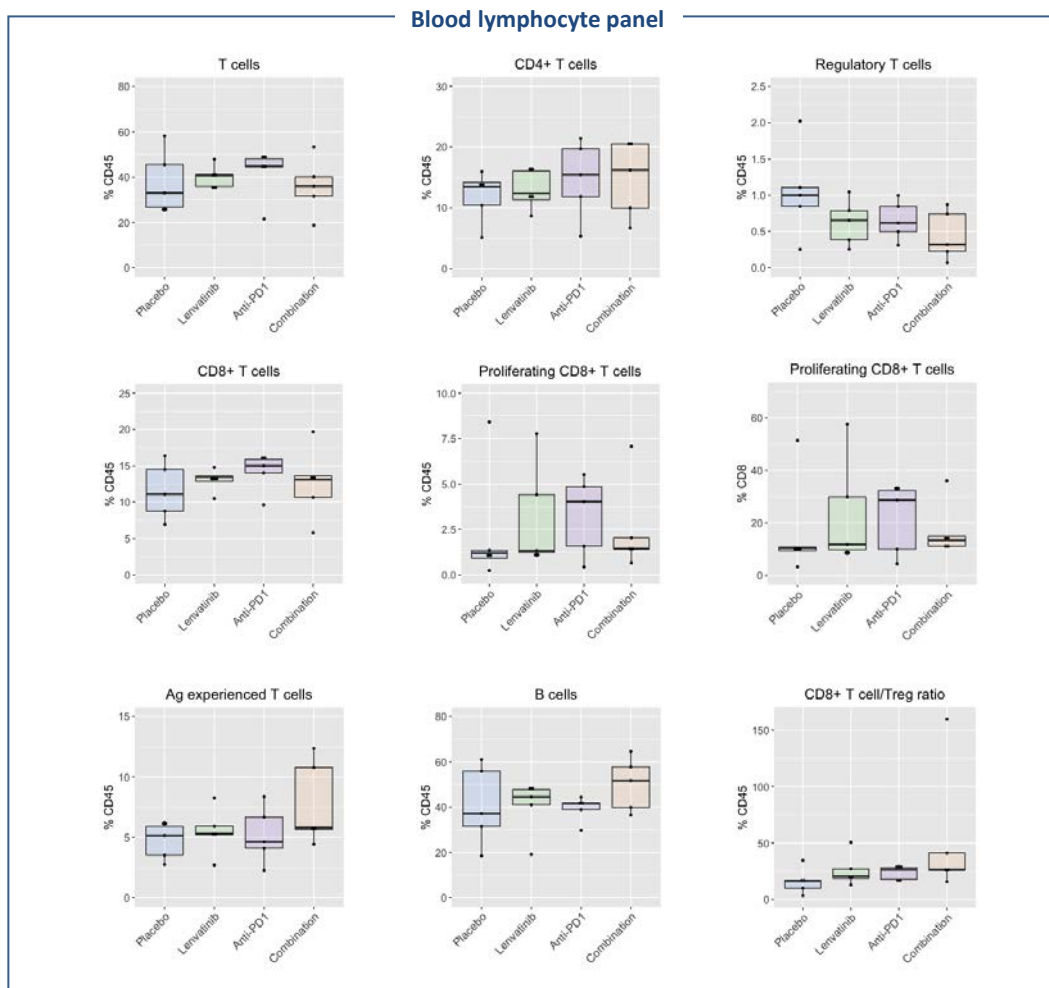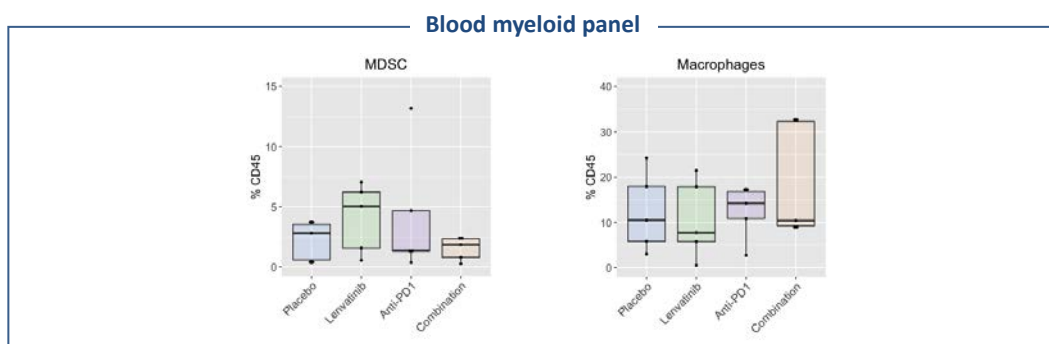
Supplementary Fig. 5. Immune cell populations in blood samples detected by flow cytometry analysis. **(A)** Lymphoid and **(B)** myeloid immune cell populations from blood samples collected at the early timepoint. Results for each treatment arm are shown (n=5 samples per arm). Boxplots indicate median and quartiles. *p<0.05.
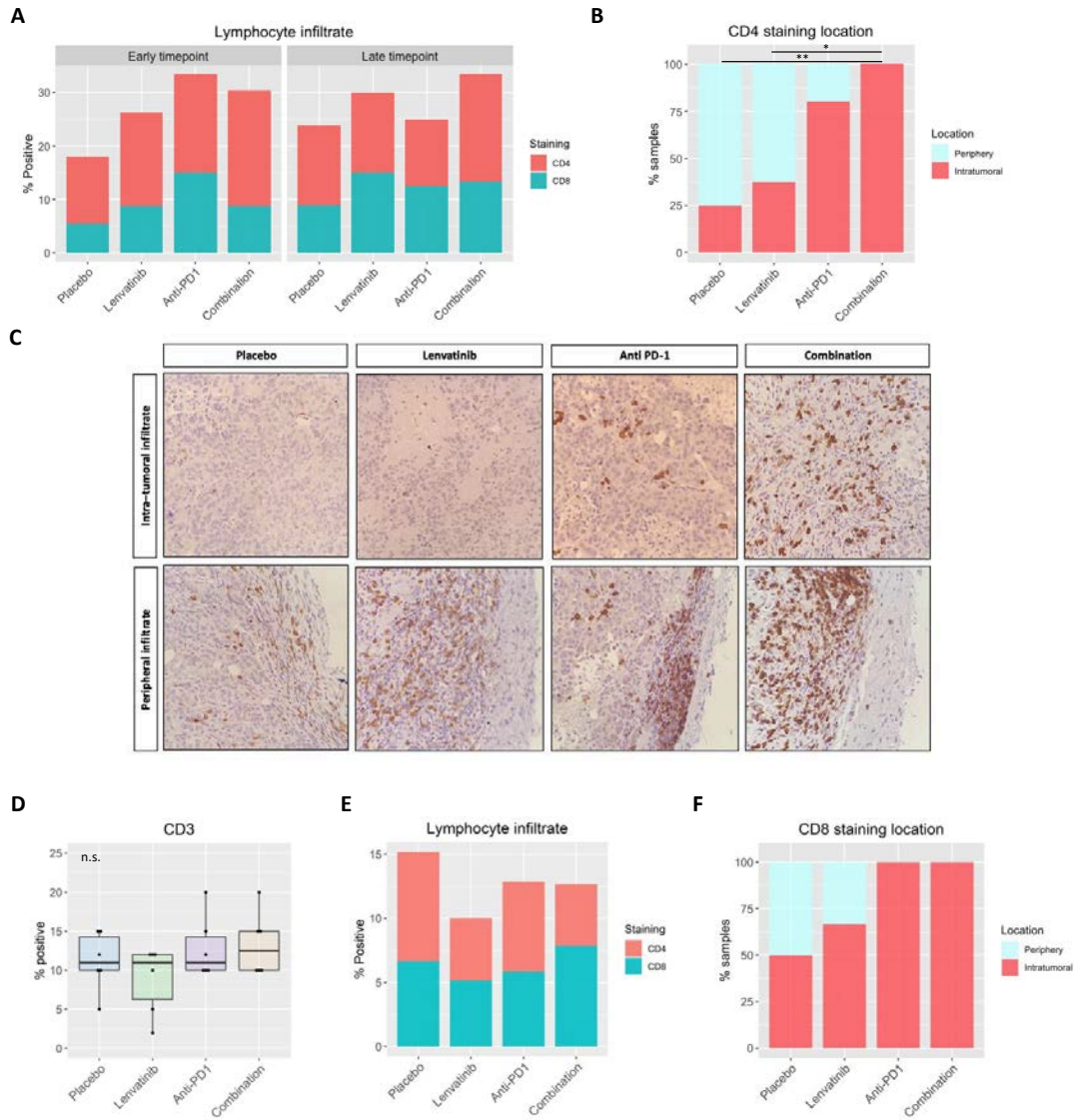
Supplementary Fig. 6. Histological analysis of lymphocytic tumor infiltrate. (A) Mean percentage of CD4 and CD8 staining in tumors from animals in each treatment arm in the Hepa1-6 model. **(B)** Percentage of samples with intra-tumoral or peripheral CD4 staining location in the Hepa1-6 model and **(C)** representative images. Images were captured with 40X. **(D)** Percentage of positive cells for CD3 and **(E)** mean percentage of CD4 and CD8 staining in the Hep53.4 subcutaneous model. **(F)** Percentage of samples with intra-tumoral or peripheral CD8 staining in the Hep53.4 model. * $p<0.05$, ** $p<0.01$.
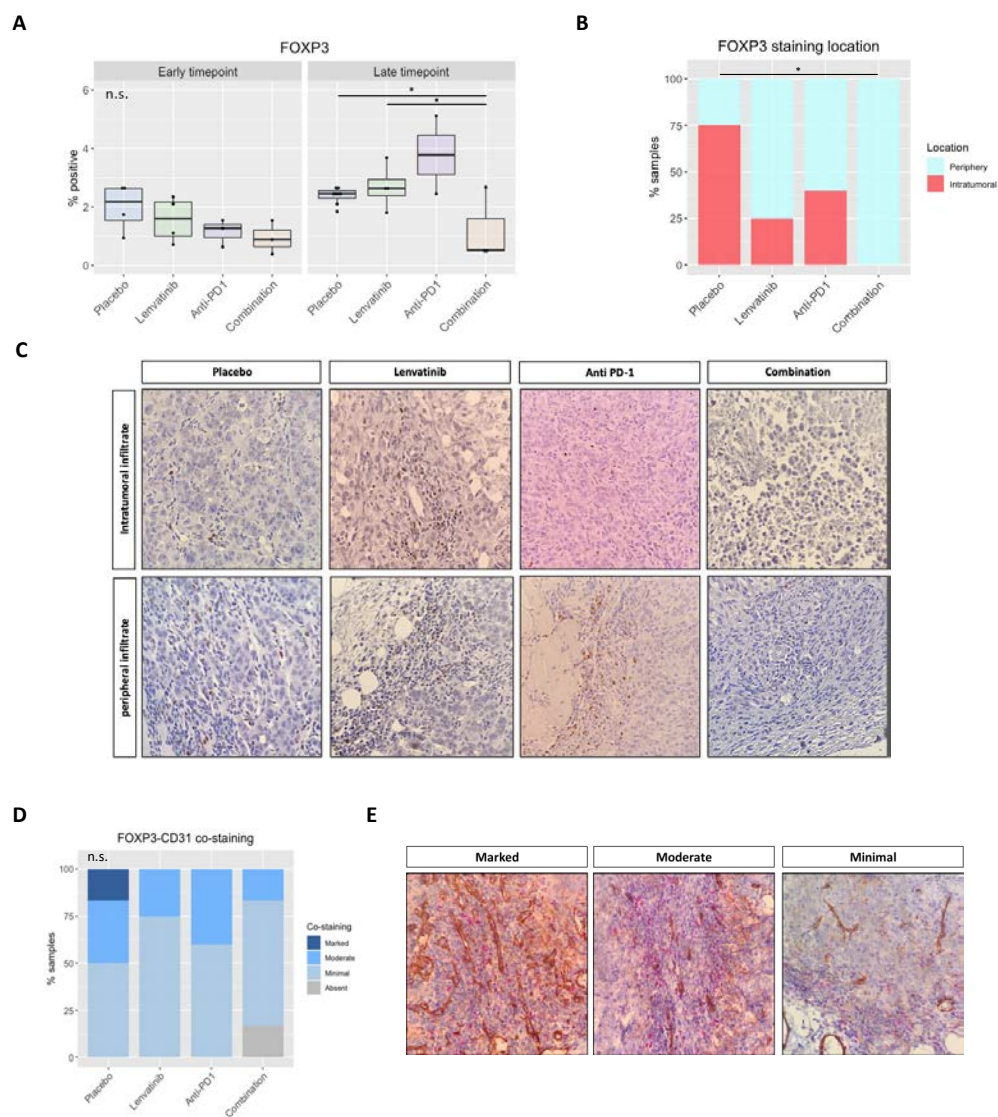
Supplementary Fig. 7. Histological analysis of T_reg cell tumor infiltrate and interaction with tumor vasculature. (A) Percentage of positive cells for FOXP3 staining in tumor samples from the Hepa1-6 model. Boxplots indicate median and quartiles. **(B)** Percentage of samples with intra-tumoral or peripheral FOXP3 staining and **(C)** representative images. **(D)** Percentage of samples with marked, moderate, minimal T_reg and tumor vasculature co-localization, or absence of T_reg, assessed by FOXP3 and CD31 co-staining. **(E)** Representative images of FOXP3 (pink) and CD31 (brown) co-staining. Images were captured with 40X. * Images were captured with 40X. * p<0.05, ** p<0.01.
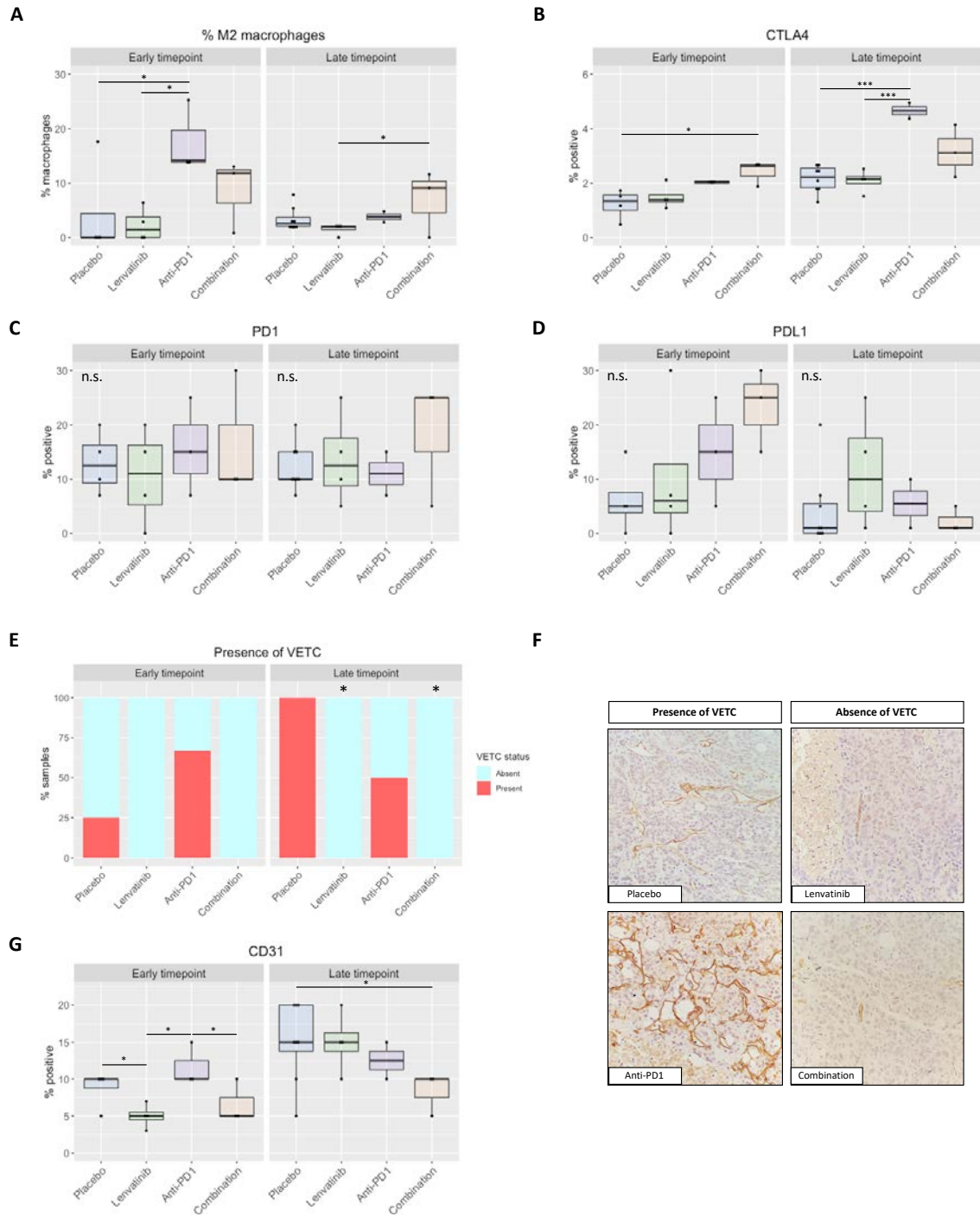
Supplementary Fig. 8. Histological analysis of tumor microenvironment at the early and late timepoints.
(A) Percentage of cells double-stained with CD68 and CD163 (M2 macrophages) out of total CD68-stained samples (total macrophages) in tumor samples from treated animals. **(B)** Percentage of positive cells for CTLA4 and positive area for **(C)** PD1, and **(D)** PDL1 staining in tumor samples from treated animals. **(E)** Percentage of samples with vessels encapsulating tumor clusters in tumor samples (VETC) from treated animals. **(F)** Representative images of CD31 staining and **(G)** percentage of positive area in each treatment arm. Images were captured with 40X. Boxplots indicate median and quartiles. * p<0.05, *** p<0.001.
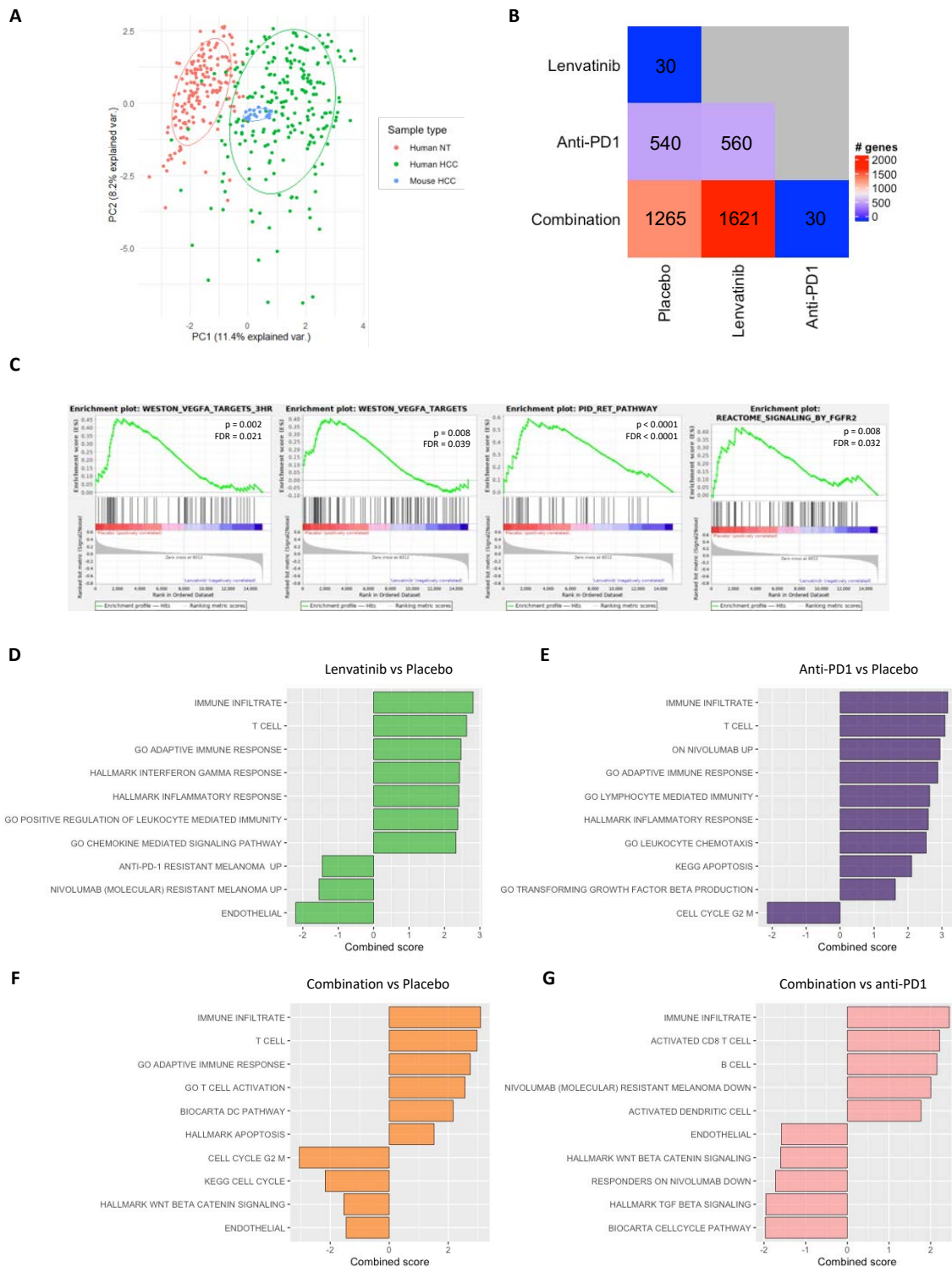
Supplementary Fig. 9. Transcriptomic differences between treatment groups. (A) Principal component analysis including samples from murine HCC, human HCC tumors and adjacent non-tumoral cirrhotic tissue (Heptromic cohort). **(B)** Number of differentially expressed genes among treatment groups (Bonferroni p<0.05, FC>1.5 or <0.67). **(C)** GSEA analysis of gene sets associated with pathways downstream of Lenvatinib targets in tumor samples from the lenvatinib group compared to placebo. **(D-F)** Representative differentially expressed gene sets between lenvatinib (D), anti-PD1 (E) or combination (F) and placebo assessed by GSEA. **(G)** Top differentially expressed gene sets between combination and anti-PD1 assessed by GSEA (Bonferroni test, all p<0.05).
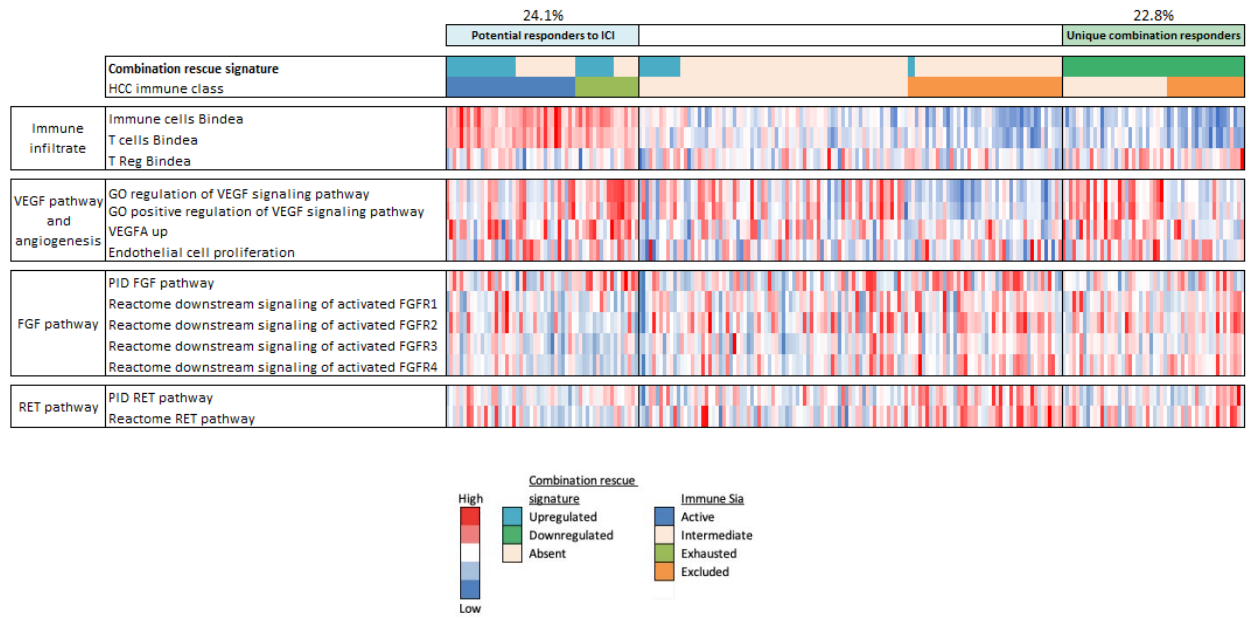
Supplementary Fig. 10. Enrichment of gene sets associated with lenvatinib target pathway activation in human HCC. Transcriptomic profile of HCC samples classified as HCC Immune class (potential responders to ICI) or combination-only responder class. The heatmap represents the enrichment score of VEGF, FGF, RET, or immune-related gene sets.

SUPPLEMENTARY TABLES

Supplementary Table 1. Cell populations assessed by flow cytometry.

Lymphoid cell populations	Markers
Leukocytes	CD45$^+$
T cells	CD45$^+$ CD3$^+$
Ag-experienced T cells*	CD45$^+$ CD3$^+$ CD44$^+$
CD8 T cells	CD45$^+$ CD3$^+$ CD8$^+$
PD1$^+$ CD8 T cells$^\#$	CD45$^+$ CD3$^+$ CD8$^+$ PD1$^+$
Proliferating CD8 T cells	CD45$^+$ CD3$^+$ CD8$^+$ Ki67$^+$
CD4 T cells	CD45$^+$ CD3$^+$ CD4$^+$
T regulatory cells	CD45$^+$ CD3$^+$ CD4$^+$ Foxp3$^+$
B cells	CD45$^+$ CD3$^-$ CD19$^+$
Myeloid cell populations	**Markers**
Type 1 dendritic cells$^\#$	CD45$^+$ CD3$^-$ CD11b$^-$ CD11c$^+$ Ly6c$^-$Ly6g$^-$
Macrophages	CD45$^+$ CD3$^-$ CD11b$^+$ Ly6c$^-$ Ly6g^{Low}
MDSC	CD45$^+$ CD3$^-$ CD11b$^+$ Ly6c$^+$ Ly6g$^+$

* Population analyzed in blood only. $^\#$ Population analyzed in tumor tissue only.

Supplementary Table 2. Antibody panels for flow cytometry analysis.

Antibody	Fluorochrome	Color	Clone	Test (μg in 100 ul)	Company
Lymphocyte panel (tumor)					
CD45	PE-Cy7	Yel 5	30F11	0.125	Biolegend
CD3	PerCP/Cy5.5	Blue 4	145-2C11	0.25	Biolegend
CD19	PE	Yel 1	1D3/CD19	0.03125	Biolegend
CD4	APC/Fire 750	Red 3	GK 1.5	0.125	Biolegend
CD8	Brilliant Violet 510	Viol 2	53–6.7	0.5	Biolegend
FoxP3	APC	Red 1	FJK-16s	1	eBioscience
PD1	FITC	Blue 1	J43	1	eBioscience
Ki67	eFluor 450	Viol 1	SolA15	0.1	eBioscience
Live Dead Red	-	Yel 2	-	1 ul per 10^6 cells	ThermoFisher
Lymphocyte panel (blood)					
CD45	PE-Cy7	Yel 5	30F11	0.125	Biolegend
CD3	PerCP/Cy5.5	Blue 4	145-2C11	0.25	Biolegend
CD19	PE	Yel 1	1D3/CD19	0.03125	Biolegend
CD4	APC/Fire 750	Red 3	GK 1.5	0.125	Biolegend
CD8	Brilliant Violet 510	Viol 2	53–6.7	0.5	Biolegend
FoxP3	APC	Red 1	FJK-16s	1	eBioscience
CD44	FITC	Blue 1	IM7	0.5	Biolegend
Ki67	eFluor 450	Viol 1	SolA15	0.1	eBioscience
Live Dead Red	-	Yel 2	-	1 ul per 10^6 cells	ThermoFisher
Myeloid panel					
CD45	PE-Cy7	Yel 5	30F11	0.125	Biolegend
CD3	PerCP/Cy5.5	Blue 4	145-2C11	0.25	Biolegend
CD11b	FITC	Blue 1	M1/70	0.25	Biolegend
CD11c	Alexa fluor 700	Red 2	N418	0.5	Biolegend
Ly6C	PE	Yel 1	HK1.4	0.25	Biolegend
Ly6G	APC	Red 1	1A8	0.06	Biolegend
Zombie Aqua stain	-	Viol 2	-	1 ul per 10^6 cells	Biolegend

Supplementary Table 3. List of antibodies used for flow cytometry and immunohistochemical analyses.

Antibody	Supplier	Cat no.	Clone no.
CD45 (PE/Cyanine7)	Biolegend	103114	30F11
CD3ε (PerCP/Cyanine5.5)	Biolegend	100327	145-2C11
CD19 (PE)	Biolegend	152407	1D3/CD19
CD4 (APC/Fire™ 750)	Biolegend	100459	GK 1.5
CD8a (Brilliant Violet 510™)	Biolegend	100752	53–6.7
FOXP3 (APC)	eBioscience	17-5773-82	FJK-16s
PD1 (FITC)	eBioscience	11-9985-81	J43
Ki67 (eFluor450)	eBioscience	48-5698-80	SolA15
CD44 (FITC)	Biolegend	103021	IM7
CD11b (FITC)	Biolegend	101205	M1/70
CD11c (Alexa Fluor 700)	Biolegend	117319	N418
Ly-6C (PE)	Biolegend	128007	HK1.4
Ly-6G (APC)	Biolegend	127613	1A8
CD3	Abcam	ab16669	SP7
CD4	Abcam	ab183685	EPR19514
CD8	Abcam	ab203035	Polyclonal
FOXP3	Abcam	ab215206	EPR22102-37
CTLA4	Abcam	ab237712	CAL49
PD1	Sino Biological	50124-RP02	Polyclonal
PDL1	Novus Biologicals	MAB90782	2096A
CD31	Abcam	ab28364	Polyclonal
CD163	Abcam	ab182422	EPR19518
CD68	Abcam	ab125212	Polyclonal
Anti-rabbit IgG (HRP)	Dako	P0448	Polyclonal
Ki67	Abcam	Ab16667	SP6

Supplementary Table 4. Immune cell population in tumor samples from treated mice. Tumor immune infiltrate in tumor samples from treated mice in the Hepa1-6 model measured by flow cytometry analysis represented as percentage of CD45+ events. Values indicate mean ± standard deviation.

	Placebo (n=5)	Lenvatinib (n=5)	Anti-PD1 (n=5)	Combination (n=5)
Tumor				
T cells	50.74 ± 9.37	52.96 ± 19.67	79.01 ± 8.44	75.80 ± 14.84
CD4 T cells	14.24 ± 8.43	4.84 ± 1.74	11.45 ± 5.28	8.08 ± 2.81
Regulatory T cells	3.32 ± 0.98	1.67 ± 1.64	2.49 ± 1.04	0.48 ± 0.25
CD8 T cells	26.54 ± 5.12	19.22 ± 5.01	31.82 ± 11.72	27.28 ± 17.34
Proliferating CD8 T cells	5.92 ± 4.97	6.91 ± 3.12	11.07 ± 3.45	13.83 ± 9.86
PD1$^+$ CD8 T cells	11.27 ± 8.10	10.45 ± 2.75	2.20 ± 1.09	1.85 ± 1.22
B cells	15.19 ± 12.39	10.05 ± 8.01	5.11 ± 6.27	5.93 ± 4.74
Type 1 dendritic cells	0.60 ± 0.52	0.71 ± 0.41	2.45 ± 1.13	2.89 ± 1.17
Macrophages	3.07 ± 3.44	2.78 ± 1.92	1.63 ± 1.06	2.95 ± 3.49
MDSC	2.23 ± 2.60	9.58 ± 7.57	2.58 ± 2.25	3.50 ± 3.95
Blood				
T cells	37.59 ± 14.07	40.14 ± 5.17	41.69 ± 11.47	35.92 ± 12.55
Antigen experienced T cells	4.27 ± 1.55	5.48 ± 1.98	5.20 ± 2.36	7.81 ± 3.52
CD4 T cells	11.82 ± 4.25	13.00 ± 3.35	14.75 ± 6.46	14.78 ± 6.25
Regulatory T cells	1.05 ± 3.91	0.63 ± 0.32	0.65 ± 0.27	0.44 ± 0.34
CD8 T cells	11.57 ± 3.91	13.03 ± 1.58	14.16 ± 2.69	12.57 ± 5.03
Proliferating CD8 T cells	2.42 ± 3.38	3.14 ± 2.94	3.28 ± 2.18	2.53 ± 2.59
B cells	40.80 ± 17.57	40.23 ± 12.19	39.31 ± 5.68	50.06 ± 11.85
Macrophages	12.29 ± 8.73	10.67 ± 8.72	12.46 ± 6.04	18.72 ± 12.77
MDSC	2.21 ± 1.70	4.08 ± 2.87	4.18 ± 5.28	1.59 ± 0.96

Supplementary Table 5. Top 30 dfferentially expressed genes in tumors from each treatment arm (Bonferroni p <0.05, FC>1.5 or <0.67).

Lenvatinib vs Placebo			Anti-PD1 vs Placebo			Combination vs Placebo			Combination vs anti-PD1		
gene	FC	p-val	gene	FC	p-val	gene	FC	p-val	gene	FC	p-val
TSC22D3	2.891	0.040	CLEC12A	10.070	0.004	JCHAIN	14.257	0.004	ABHD1	1.627	0.008
SLC16A3	2.143	0.005	JCHAIN	9.016	0.049	CLEC12A	12.422	0.001	ACVR1	0.652	0.050
LAP3	1.782	0.038	ATP6V0D2	7.975	0.004	CLEC4D	10.589	0.001	BIN1	1.829	0.028
AMPD3	1.779	0.018	SLAMF7	7.136	0.012	ATP6V0D2	9.702	0.003	BMP2	0.587	0.015
GAL3ST4	1.749	0.027	EMB	6.618	0.007	IL7R	8.912	0.002	CALD1	0.619	0.038
GADD45B	1.748	0.018	OGN	6.542	0.036	SLAMF7	8.560	0.005	CHST10	1.562	0.013
S1PR3	1.746	0.020	LUM	6.533	0.043	TREM2	7.679	0.001	CMC2	1.529	0.043
CRELD2	1.579	0.028	IL7R	6.367	0.009	ADAM8	6.776	0.001	COG4	0.628	0.050
ISG15	1.573	0.042	CLEC4D	5.732	0.024	SAMSN1	6.629	0.002	CORO6	1.621	0.024
SGK1	1.538	0.043	PTPN22	5.650	0.004	EMB	6.455	0.010	ENDOU	1.762	0.032
HERPUD1	1.514	0.028	NOS2	5.529	0.029	MZB1	6.425	0.003	FMN2	0.662	0.018
ITIH4	1.512	0.041	SAMSN1	5.493	0.008	RGS1	6.326	0.002	FST	0.444	0.038
MAB21L3	1.508	0.039	HGF	5.489	0.019	PLA2G7	6.181	0.002	GCNT4	0.617	0.021
CDON	0.640	0.018	ADAM8	5.488	0.008	PTPN22	6.110	0.002	HERC6	0.489	0.043
PCID2	0.637	0.015	CD226	5.408	0.046	CD69	6.055	0.001	INPP5J	0.583	0.011
GPN1	0.629	0.006	PLA2G7	5.056	0.009	CD226	5.948	0.013	KRT10	0.604	0.028
PCDH12	0.624	0.022	TREM2	4.998	0.020	CD84	5.868	0.001	MAGIX	1.651	0.009
PLSCR4	0.600	0.029	ARHGAP15	4.952	0.006	TLR7	5.732	0.001	MYOM1	1.957	0.043
HLCS	0.597	0.012	CD84	4.803	0.005	ITK	5.598	0.001	NDUFS7	1.533	0.032
WDFY3	0.595	0.042	RGS1	4.766	0.017	MEF2C	5.570	0.046	NR4A1	1.697	0.038
CROCC	0.581	0.046	ATP8B4	4.740	0.017	NCEH1	5.542	0.004	PAF1	1.677	0.001
ANKRD16	0.563	0.042	ITGAX	4.657	0.005	MMP3	5.409	0.047	RMDN2	0.658	0.013
TRIT1	0.561	0.023	ICOS	4.490	0.012	PLA2G2D	5.407	0.003	SLC22A2	1.590	0.011
EPHB3	0.546	0.030	MPEG1	4.428	0.012	ATP8B4	5.397	0.003	STAB2	1.627	0.008
FBXO36	0.542	0.034	IL2RG	4.421	0.002	LY9	5.372	0.001	STK38L	0.524	0.028
GPIHBP1	0.524	0.045	ITK	4.335	0.005	ARHGEF6	5.317	0.019	TDRP	1.504	0.038
ZMYM4	0.512	0.042	KCNN4	4.270	0.035	ARHGAP15	5.265	0.010	TGM2	0.586	0.028
GJA1	0.477	0.044	DPEP2	4.253	0.008	KCNN4	5.191	0.007	TRIM7	1.791	0.028
HMGCS1	0.467	0.042	CD3G	4.152	0.016	MERTK	5.190	0.002	TXLNB	2.230	0.050
MEST	0.399	0.008	KLRD1	4.119	0.008	TM6SF1	5.149	0.003	USP14	0.640	0.032

Supplementary Table 6. Functional annotation of overexpressed genes only in the combination group compared to placebo.

Category	Term	Count	Nominal p value	Bonferroni p	Fold change
KEGG pathway	hsa04662:B cell receptor signaling pathway	11	2.21E-05	0.0049	5.57
KEGG pathway	hsa04750:Inflammatory mediator regulation of TRP channels	12	9.75E-05	0.0215	4.28
KEGG pathway	hsa04660:T cell receptor signaling pathway	12	0.0001	0.0258	4.19
KEGG pathway	hsa04142:Lysosome	13	0.0002	0.0345	3.75
KEGG pathway	hsa04062:Chemokine signaling pathway	17	7.23E-05	0.0160	3.19
UP keywords	Lysosome	18	5.74E-05	0.0177	3.19
GO term CC direct	GO:0005887~integral component of plasma membrane	60	9.05E-06	0.0030	1.81
GO term CC direct	GO:0005886~plasma membrane	132	4.74E-05	0.0157	1.37
UP keywords	Cytoplasm	141	3.21E-05	0.0099	1.36
UP keywords	Membrane	215	9.62E-08	2.99E-05	1.34

Supplementary Table 7. GSEA results. Top pathways enriched in lenvatinib vs placebo (**A**), anti-PD1 vs placebo (**B**), anti-PD1 vs placebo (**C**) and combination vs anti-PD1 (**D**).

A. Lenvatinib vs Placebo					
NAME	SIZE	ES	NES	p-val	FDR
ON NIVOLUMAB UP	131	0.657	2.964	0.000	0.000
HALLMARK ALLOGRAFT REJECTION	174	0.620	2.860	0.000	0.000
IMMUNE INFILTRATE	91	0.660	2.808	0.000	0.000
T CELL	95	0.619	2.626	0.000	0.000
KEGG PRIMARY IMMUNODEFICIENCY	31	0.722	2.478	0.000	0.000
GO ADAPTIVE IMMUNE RESPONSE	183	0.529	2.467	0.000	0.000
GO ACTIN MYOSIN FILAMENT SLIDING	38	0.685	2.466	0.000	0.000
GO CELLULAR DEFENSE RESPONSE	43	0.661	2.457	0.000	0.000
HALLMARK INTERFERON GAMMA RESPONSE	174	0.523	2.422	0.000	0.000
GO REGULATION OF ADAPTIVE IMMUNE RESPONSE	100	0.568	2.416	0.000	0.000
HALLMARK INFLAMMATORY RESPONSE	191	0.505	2.406	0.000	0.000
GO STRUCTURAL CONSTITUENT OF MUSCLE	40	0.672	2.395	0.000	0.000
GO POSITIVE REGULATION OF LEUKOCYTE MEDIATED IMMUNITY	69	0.593	2.376	0.000	0.000
T CD4	40	0.645	2.373	0.000	0.000
T CD8	43	0.638	2.370	0.000	0.000
KEGG RIBOSOME	74	0.569	2.346	0.000	0.000
GO POSITIVE REGULATION OF LYMPHOCYTE MEDIATED IMMUNITY	57	0.594	2.338	0.000	0.000
GO CHEMOKINE MEDIATED SIGNALING PATHWAY	45	0.628	2.322	0.000	0.000
BIOCARTA NO2IL12 PATHWAY	17	0.783	2.307	0.000	0.000
GO REGULATION OF LEUKOCYTE MEDIATED IMMUNITY	126	0.520	2.306	0.000	0.001
NIVOLUMAB (MOLECULAR) RESISTANT MELANOMA DN	353	0.459	2.292	0.000	0.000
GO REGULATION OF INTERFERON GAMMA PRODUCTION	80	0.548	2.282	0.000	0.001
GO REGULATION OF LYMPHOCYTE MEDIATED IMMUNITY	90	0.541	2.272	0.000	0.001
HALLMARK TNFA SIGNALING VIA NFKB	187	0.481	2.267	0.000	0.000
GO RESPONSE TO INTERFERON GAMMA	94	0.534	2.266	0.000	0.001
GO MYELOID LEUKOCYTE ACTIVATION	87	0.543	2.261	0.000	0.001
BIOCARTA CTLA4 PATHWAY	16	0.786	2.257	0.000	0.001
GO LEUKOCYTE ACTIVATION	356	0.442	2.255	0.000	0.001
HALLMARK INTERFERON ALPHA RESPONSE	83	0.543	2.255	0.000	0.000
T CD8 EXHAUSTED	68	0.560	2.243	0.000	0.000
GO MYOFILAMENT	24	0.702	2.224	0.000	0.001
GO CELLULAR RESPONSE TO INTERFERON GAMMA	74	0.545	2.221	0.000	0.001
GO POSITIVE REGULATION OF ADAPTIVE IMMUNE RESPONSE	62	0.570	2.211	0.000	0.001
GO IMMUNOLOGICAL SYNAPSE	31	0.645	2.210	0.000	0.001
GO B CELL RECEPTOR SIGNALING PATHWAY	31	0.652	2.203	0.000	0.001
BIOCARTA NKT PATHWAY	27	0.656	2.196	0.000	0.001
GO INNATE IMMUNE RESPONSE	390	0.427	2.189	0.000	0.001
ENDOTHELIAL	179	-0.525	-2.200	0.000	0.000
STROMA	89	-0.552	-2.111	0.000	0.000
CAF	189	-0.356	-1.514	0.001	0.029
KEGG VALINE LEUCINE AND ISOLEUCINE DEGRADATION	41	-0.697	-2.318	0.000	0.000
KEGG PROPANOATE METABOLISM	30	-0.675	-2.101	0.000	0.000
KEGG BUTANOATE METABOLISM	31	-0.622	-1.921	0.000	0.005
KEGG PEROXISOME	77	-0.476	-1.782	0.000	0.030
KEGG GLYCINE SERINE AND THREONINE METABOLISM	29	-0.577	-1.770	0.007	0.027
KEGG BETA ALANINE METABOLISM	22	-0.614	-1.743	0.002	0.032
KEGG HISTIDINE METABOLISM	26	-0.572	-1.730	0.003	0.032
KEGG CITRATE CYCLE TCA CYCLE	28	-0.561	-1.696	0.005	0.039
NIVOLUMAB (MOLECULAR) RESISTANT MELANOMA UP	178	-0.367	-1.541	0.000	0.012
ANTI-PD-1 RESISTANT MELANOMA UP	421	-0.319	-1.451	0.000	0.019

B. Anti-PD1 vs Placebo					
NAME	SIZE	ES	NES	p-val	FDR
IMMUNE INFILTRATE	91	0.842	3.168	0.000	0.000
T CELL	95	0.816	3.087	0.000	0.000
ON NIVOLUMAB UP	131	0.749	2.942	0.000	0.000
GO ADAPTIVE IMMUNE RESPONSE	294	0.682	2.876	0.000	0.000
NIVOLUMAB (MOLECULAR) RESISTANT MELANOMA DN	353	0.658	2.823	0.000	0.000
GO B CELL RECEPTOR SIGNALING PATHWAY	49	0.809	2.731	0.000	0.000
GO T CELL ACTIVATION	389	0.626	2.714	0.000	0.000
HALLMARK INTERFERON GAMMA RESPONSE	175	0.674	2.704	0.000	0.000
GO REGULATION OF ANTIGEN RECEPTOR MEDIATED SIGNALING PATHWAY	53	0.788	2.668	0.000	0.000
GO POSITIVE REGULATION OF CELL ACTIVATION	270	0.635	2.659	0.000	0.000
GO ANTIGEN RECEPTOR MEDIATED SIGNALING PATHWAY	203	0.647	2.650	0.000	0.000
GO POSITIVE REGULATION OF LEUKOCYTE CELL CELL ADHESION	183	0.653	2.641	0.000	0.000
GO LYMPHOCYTE MEDIATED IMMUNITY	182	0.646	2.636	0.000	0.000
GO LEUKOCYTE CELL CELL ADHESION	286	0.621	2.621	0.000	0.000
GO IMMUNE RESPONSE REGULATING CELL SURFACE RECEPTOR SIGNALING PATHWAY	322	0.612	2.614	0.000	0.000
GO REGULATION OF LEUKOCYTE MEDIATED IMMUNITY	147	0.655	2.604	0.000	0.000
GO T CELL DIFFERENTIATION	213	0.626	2.601	0.000	0.000
GO LEUKOCYTE MEDIATED CYTOTOXICITY	71	0.737	2.601	0.000	0.000
GO IMMUNE RESPONSE REGULATING SIGNALING PATHWAY	468	0.598	2.600	0.000	0.000
HALLMARK INFLAMMATORY RESPONSE	192	0.633	2.595	0.000	0.000
GO INTERFERON GAMMA PRODUCTION	94	0.689	2.593	0.000	0.000
GO LYMPHOCYTE DIFFERENTIATION	306	0.611	2.592	0.000	0.000
GO ADAPTIVE IMMUNE RESPONSE BASED ON SOMATIC RECOMBINATION OF IMMUNE RECEPTORS BUILT FROM IMMUNOGLOBULIN SUPERFAMILY DOMAINS	195	0.629	2.579	0.000	0.000
GO LEUKOCYTE PROLIFERATION	235	0.618	2.578	0.000	0.000
GO B CELL ACTIVATION	203	0.627	2.567	0.000	0.000
GO CELL KILLING	92	0.685	2.561	0.000	0.000
GO REGULATION OF LYMPHOCYTE MEDIATED IMMUNITY	104	0.673	2.561	0.000	0.000
GO REGULATION OF T CELL ACTIVATION	267	0.612	2.555	0.000	0.000
GO CELL CHEMOTAXIS	235	0.611	2.552	0.000	0.000
GO POSITIVE REGULATION OF LYMPHOCYTE ACTIVATION	222	0.618	2.548	0.000	0.000
HALLMARK INTERFERON ALPHA RESPONSE	85	0.692	2.546	0.000	0.000
GO REGULATION OF IMMUNE EFFECTOR PROCESS	312	0.599	2.545	0.000	0.000
KEGG T CELL RECEPTOR SIGNALING PATHWAY	107	0.664	2.539	0.000	0.000
GO LEUKOCYTE MIGRATION	356	0.592	2.537	0.000	0.000
GO LEUKOCYTE CHEMOTAXIS	167	0.634	2.536	0.000	0.000
GO POSITIVE REGULATION OF CYTOKINE PRODUCTION	387	0.584	2.535	0.000	0.000
GO REGULATION OF B CELL RECEPTOR SIGNALING PATHWAY	24	0.852	2.529	0.000	0.000
GO T CELL ACTIVATION INVOLVED IN IMMUNE RESPONSE	76	0.700	2.526	0.000	0.000
GO T CELL DIFFERENTIATION INVOLVED IN IMMUNE RESPONSE	60	0.690	2.402	0.000	0.000
GO INTERLEUKIN 4 PRODUCTION	32	0.772	2.402	0.000	0.000
GO ALPHA BETA T CELL PROLIFERATION	28	0.696	2.112	0.000	0.000
KEGG APOPTOSIS	77	0.576	2.107	0.000	0.000
GO TRANSFORMING GROWTH FACTOR BETA PRODUCTION	35	0.515	1.627	0.014	0.037
HALLMARK P53 PATHWAY	191	0.376	1.529	0.002	0.014
HALLMARK APICAL JUNCTION	190	0.365	1.473	0.004	0.026
HALLMARK COAGULATION	127	0.362	1.406	0.034	0.049
HALLMARK MYC TARGETS V1	192	-0.307	-1.491	0.000	0.037

NAME	SIZE	ES	NES	p-val	FDR
HALLMARK MYOGENESIS	196	-0.332	-1.640	0.000	0.010
HALLMARK E2F TARGETS	191	-0.377	-1.824	0.000	0.002
CELL CYCLE G2 M	52	-0.542	-2.144	0.000	0.000
C. Combination vs Placebo					
NAME	**SIZE**	**ES**	**NES**	**p-val**	**FDR**
IMMUNE INFILTRATE	91	0.845	3.097	0.000	0.000
T CELL	95	0.804	2.980	0.000	0.000
MACROPHAGE	350	0.668	2.926	0.000	0.000
NIVOLUMAB (MOLECULAR) RESISTANT MELANOMA DN	353	0.672	2.926	0.000	0.000
ON NIVOLUMAB UP	131	0.726	2.796	0.000	0.000
GO ADAPTIVE IMMUNE RESPONSE	294	0.646	2.751	0.000	0.000
GO B CELL RECEPTOR SIGNALING PATHWAY	49	0.822	2.696	0.000	0.000
GO ANTIGEN RECEPTOR MEDIATED SIGNALING PATHWAY	203	0.653	2.679	0.000	0.000
GO POSITIVE REGULATION OF CELL ACTIVATION	270	0.634	2.672	0.000	0.000
HALLMARK ALLOGRAFT REJECTION	175	0.656	2.648	0.000	0.000
GO B CELL ACTIVATION	203	0.638	2.617	0.000	0.000
GO T CELL ACTIVATION	389	0.590	2.570	0.000	0.000
GO POSITIVE REGULATION OF LEUKOCYTE CELL CELL ADHESION	183	0.629	2.567	0.000	0.000
GO REGULATION OF CELL ACTIVATION	464	0.579	2.564	0.000	0.000
KEGG NATURAL KILLER CELL MEDIATED CYTOTOXICITY	88	0.696	2.562	0.000	0.000
GO LYMPHOCYTE DIFFERENTIATION	306	0.601	2.554	0.000	0.000
GO B CELL DIFFERENTIATION	106	0.677	2.535	0.000	0.000
GO REGULATION OF ANTIGEN RECEPTOR MEDIATED SIGNALING PATHWAY	53	0.745	2.529	0.000	0.000
GO POSITIVE REGULATION OF LYMPHOCYTE ACTIVATION	222	0.607	2.524	0.000	0.000
GO REGULATION OF LYMPHOCYTE ACTIVATION	354	0.577	2.521	0.000	0.000
GO IMMUNE RESPONSE REGULATING CELL SURFACE RECEPTOR SIGNALING PATHWAY	322	0.587	2.520	0.000	0.000
KEGG B CELL RECEPTOR SIGNALING PATHWAY	72	0.716	2.518	0.000	0.000
GO B CELL PROLIFERATION	75	0.703	2.504	0.000	0.000
GO SPECIFIC GRANULE MEMBRANE	85	0.682	2.488	0.000	0.000
GO T CELL DIFFERENTIATION	213	0.600	2.479	0.000	0.000
KEGG T CELL RECEPTOR SIGNALING PATHWAY	107	0.652	2.471	0.000	0.000
GO IMMUNE RESPONSE REGULATING SIGNALING PATHWAY	468	0.556	2.460	0.000	0.000
GO REGULATION OF B CELL DIFFERENTIATION	26	0.824	2.458	0.000	0.000
BIOCARTA DC PATHWAY	16	0.846	2.169	0.000	0.000
HALLMARK APOPTOSIS	153	0.381	1.522	0.004	0.014
ENDOTHELIAL	179	-0.329	-1.465	0.000	0.037
HALLMARK WNT BETA CATENIN SIGNALING	40	-0.441	-1.540	0.024	0.023
GO CELL CYCLE G2 M PHASE TRANSITION	249	-0.422	-1.962	0.000	0.008
KEGG CELL CYCLE	121	-0.519	-2.169	0.000	0.000
GO CHROMOSOMAL REGION	305	-0.488	-2.293	0.000	0.000
KEGG MATURITY ONSET DIABETES OF THE YOUNG	23	-0.763	-2.299	0.000	0.000
GO CHROMOSOME SEPARATION	83	-0.581	-2.306	0.000	0.000
GO REGULATION OF CHROMOSOME SEPARATION	58	-0.629	-2.327	0.000	0.000
GO NUCLEAR CHROMOSOME SEGREGATION	224	-0.515	-2.346	0.000	0.000
GO REGULATION OF UBIQUITIN PROTEIN LIGASE ACTIVITY	21	-0.812	-2.358	0.000	0.000
GO METAPHASE ANAPHASE TRANSITION OF CELL CYCLE	54	-0.643	-2.366	0.000	0.000
GO CHROMATIN REMODELING AT CENTROMERE	29	-0.746	-2.371	0.000	0.000
GO REGULATION OF CHROMOSOME SEGREGATION	97	-0.596	-2.395	0.000	0.000
GO NEGATIVE REGULATION OF CHROMOSOME SEGREGATION	42	-0.688	-2.401	0.000	0.000
GO SISTER CHROMATID SEGREGATION	165	-0.568	-2.480	0.000	0.000
GO MITOTIC NUCLEAR DIVISION	259	-0.533	-2.504	0.000	0.000
GO CHROMOSOME CENTROMERIC REGION	177	-0.566	-2.506	0.000	0.000
GO MITOTIC SISTER CHROMATID SEGREGATION	135	-0.614	-2.597	0.000	0.000

NAME	SIZE	ES	NES	p-val	FDR
GO CONDENSED CHROMOSOME CENTROMERIC REGION	102	-0.653	-2.692	0.000	0.000
HALLMARK G2M CHECKPOINT	184	-0.638	-2.850	0.000	0.000
CELL CYCLE G2 M	52	-0.852	-3.053	0.000	0.000
Combination vs Anti-PD1					
NAME	**SIZE**	**ES**	**NES**	**p-val**	**FDR**
IMMUNE INFILTRATE	91	0.610	2.446	0.000	0.000
KEGG CARDIAC MUSCLE CONTRACTION	72	0.560	2.171	0.000	0.000
B CELL	64	0.570	2.152	0.000	0.000
HALLMARK OXIDATIVE PHOSPHORYLATION	174	0.480	2.108	0.000	0.000
B CELLS	23	0.666	2.009	0.000	0.003
NIVOLUMAB (MOLECULAR) RESISTANT MELANOMA DN	353	0.420	2.005	0.000	0.000
KEGG PRIMARY IMMUNODEFICIENCY	31	0.595	1.915	0.000	0.014
KEGG DILATED CARDIOMYOPATHY	88	0.482	1.903	0.000	0.011
KEGG HYPERTROPHIC CARDIOMYOPATHY HCM	82	0.481	1.897	0.000	0.010
KEGG B CELL RECEPTOR SIGNALING PATHWAY	71	0.468	1.773	0.002	0.034
KEGG VIRAL MYOCARDITIS	49	0.491	1.764	0.002	0.032
T CELL	95	0.377	1.514	0.009	0.058
T CD4	40	0.426	1.478	0.037	0.058
TFH CELLS	25	0.374	1.150	0.276	0.792
CD8 T CELLS	24	0.365	1.128	0.272	0.573
NK CELLS	20	0.331	0.950	0.534	0.794
T HELPER CELLS	19	0.290	0.843	0.673	0.710
TH1 CELLS	24	-0.300	-0.874	0.650	0.729
NEUTROPHILS	17	-0.351	-0.957	0.518	0.775
TCM CELLS	27	-0.436	-1.293	0.150	0.309
IDC	22	-0.458	-1.305	0.127	0.391
HALLMARK APOPTOSIS	152	-0.359	-1.484	0.002	0.021
HALLMARK TNFA SIGNALING VIA NFKB	187	-0.358	-1.526	0.002	0.016
ENDOTHELIAL	179	-0.372	-1.584	0.002	0.024
HALLMARK ESTROGEN RESPONSE EARLY	192	-0.373	-1.586	0.000	0.011
TREG CELLS	147	-0.386	-1.594	0.000	0.093
STROMA	89	-0.416	-1.599	0.005	0.028
HALLMARK WNT BETA CATENIN SIGNALING	40	-0.489	-1.599	0.015	0.010
HALLMARK KRAS SIGNALING UP	183	-0.387	-1.634	0.000	0.008
HALLMARK MYC TARGETS V2	58	-0.478	-1.699	0.002	0.004
KEGG ECM RECEPTOR INTERACTION	81	-0.453	-1.705	0.000	0.034
IFN SIGNATURE	26	-0.572	-1.719	0.004	0.059
RESPONDERS ON NIVOLUMAB DN	437	-0.370	-1.725	0.000	0.000
KEGG ADHERENS JUNCTION	68	-0.475	-1.739	0.000	0.028
HALLMARK APICAL SURFACE	42	-0.531	-1.759	0.007	0.002
HALLMARK ESTROGEN RESPONSE LATE	189	-0.433	-1.822	0.000	0.001
KEGG AXON GUIDANCE	125	-0.457	-1.840	0.000	0.008
KEGG PATHWAYS IN CANCER	309	-0.411	-1.844	0.000	0.008
HALLMARK TGF BETA SIGNALING	54	-0.557	-1.955	0.000	0.000
KEGG P53 SIGNALING PATHWAY	58	-0.552	-1.961	0.002	0.003
BIOCARTA CELLCYCLE PATHWAY	23	-0.681	-1.969	0.000	0.019
HALLMARK INTERFERON GAMMA RESPONSE	174	-0.480	-2.022	0.000	0.000
KEGG TGF BETA SIGNALING PATHWAY	83	-0.546	-2.069	0.000	0.001
KEGG CELL CYCLE	121	-0.573	-2.276	0.000	0.000
HALLMARK INTERFERON ALPHA RESPONSE	83	-0.606	-2.315	0.000	0.000
HALLMARK MITOTIC SPINDLE	193	-0.588	-2.494	0.000	0.000
CELL CYCLE G1 S	42	-0.745	-2.501	0.000	0.000
HALLMARK E2F TARGETS	190	-0.647	-2.752	0.000	0.000
CELL CYCLE G2 M	52	-0.820	-2.870	0.000	0.000
HALLMARK G2M CHECKPOINT	187	-0.681	-2.879	0.000	0.000

Supplementary Table 8. Combination rescue gene signature.

Gene	Group	p-value (Bonferroni)	FC	Score (Log2FC)
MZB1	Upregulation by combination	0.003	6.425	2.684
MEF2C	Upregulation by combination	0.046	5.570	2.478
NCEH1	Upregulation by combination	0.004	5.542	2.470
MMP3	Upregulation by combination	0.047	5.409	2.435
ARHGEF6	Upregulation by combination	0.019	5.317	2.411
TM6SF1	Upregulation by combination	0.003	5.149	2.364
TSC22D3	Upregulation by combination	0.000	5.005	2.323
GPNMB	Upregulation by combination	0.007	4.718	2.238
DNASE1L1	Upregulation by combination	0.002	4.481	2.164
FOLR2	Upregulation by combination	0.011	4.372	2.128
CD79B	Upregulation by combination	0.014	4.290	2.101
RGS2	Upregulation by combination	0.036	4.244	2.085
MYLIP	Upregulation by combination	0.014	4.192	2.068
GPR65	Upregulation by combination	0.001	4.087	2.031
B3GALNT1	Upregulation by combination	0.019	4.002	2.001
HVCN1	Upregulation by combination	0.004	3.928	1.974
HCST	Upregulation by combination	0.002	3.846	1.943
PRRG1	Upregulation by combination	0.044	3.775	1.916
SLPI	Upregulation by combination	0.011	3.770	1.915
SCARB2	Upregulation by combination	0.016	3.747	1.906
GDE1	Upregulation by combination	0.026	3.671	1.876
TRPV2	Upregulation by combination	0.006	3.578	1.839
RAB29	Upregulation by combination	0.003	3.556	1.830
GNGT2	Upregulation by combination	0.012	3.520	1.816
LY6K	Upregulation by combination	0.006	3.508	1.811
SLC9A9	Upregulation by combination	0.002	3.494	1.805
FGD4	Upregulation by combination	0.008	3.477	1.798
ADSSL1	Upregulation by combination	0.018	3.472	1.796
HMOX1	Upregulation by combination	0.003	3.449	1.786
FLI1	Upregulation by combination	0.009	3.432	1.779
LST1	Upregulation by combination	0.035	3.415	1.772
CD8A	Upregulation by combination	0.030	3.403	1.767
PTPN6	Upregulation by combination	0.027	3.351	1.744
RASSF4	Upregulation by combination	0.002	3.347	1.743
TNFAIP2	Upregulation by combination	0.030	3.343	1.741
DOCK10	Upregulation by combination	0.020	3.321	1.731
SLC43A2	Upregulation by combination	0.010	3.298	1.722
MMP27	Upregulation by combination	0.011	3.283	1.715
KBTBD11	Upregulation by combination	0.010	3.245	1.698
GPM6B	Upregulation by combination	0.049	3.196	1.676
PIP4K2A	Upregulation by combination	0.007	3.179	1.669
CD38	Upregulation by combination	0.006	3.171	1.665
RINL	Upregulation by combination	0.008	3.160	1.660

Gene	Regulation	p		
CELF2	Upregulation by combination	0.012	3.106	1.635
IKZF3	Upregulation by combination	0.007	3.105	1.635
SNX29	Upregulation by combination	0.014	3.079	1.623
MAFB	Upregulation by combination	0.010	3.078	1.622
SYT11	Upregulation by combination	0.003	3.075	1.620
CNR2	Upregulation by combination	0.002	3.071	1.619
LIPA	Upregulation by combination	0.011	3.045	1.606
RSPO3	Upregulation by combination	0.004	3.042	1.605
HFE	Upregulation by combination	0.029	3.027	1.598
GTPBP10	Upregulation by combination	0.038	3.015	1.592
EPHA7	Downregulation by combination	0.029	0.121	-3.048
LECT2	Downregulation by combination	0.043	0.121	-3.047
AMBP	Downregulation by combination	0.011	0.126	-2.986
CPN1	Downregulation by combination	0.045	0.163	-2.614
SLC16A4	Downregulation by combination	0.021	0.185	-2.433
TM4SF20	Downregulation by combination	0.028	0.185	-2.432
TTR	Downregulation by combination	0.041	0.186	-2.425
ABCC2	Downregulation by combination	0.033	0.194	-2.368
LMO7	Downregulation by combination	0.005	0.195	-2.360
THSD4	Downregulation by combination	0.005	0.206	-2.277
ELOVL7	Downregulation by combination	0.002	0.214	-2.223
PLCH1	Downregulation by combination	0.021	0.217	-2.206
EPB41L4B	Downregulation by combination	0.023	0.221	-2.177
SERPIND1	Downregulation by combination	0.044	0.222	-2.173
ITGA3	Downregulation by combination	0.031	0.224	-2.156
ROBO1	Downregulation by combination	0.022	0.228	-2.134
PKHD1	Downregulation by combination	0.042	0.229	-2.124
ZIC5	Downregulation by combination	0.016	0.231	-2.116
TSPAN8	Downregulation by combination	0.041	0.236	-2.082
CATSPERD	Downregulation by combination	0.009	0.239	-2.067
ALDOB	Downregulation by combination	0.029	0.247	-2.020
ACSL3	Downregulation by combination	0.045	0.251	-1.997
RAMP3	Downregulation by combination	0.029	0.251	-1.993
TROAP	Downregulation by combination	0.015	0.252	-1.988
TMED6	Downregulation by combination	0.025	0.257	-1.962
CCDC136	Downregulation by combination	0.007	0.258	-1.953
SQLE	Downregulation by combination	0.044	0.262	-1.930
AXIN2	Downregulation by combination	0.037	0.268	-1.900
RALGAPA2	Downregulation by combination	0.033	0.269	-1.894
KHDRBS3	Downregulation by combination	0.025	0.270	-1.889
SEMA6A	Downregulation by combination	0.011	0.272	-1.877
ALAD	Downregulation by combination	0.008	0.273	-1.872
EFNA1	Downregulation by combination	0.011	0.274	-1.866
SOX9	Downregulation by combination	0.010	0.275	-1.861
ENPP2	Downregulation by combination	0.050	0.276	-1.858
IRS1	Downregulation by combination	0.011	0.278	-1.844

SLC39A8	Downregulation by combination	0.042	0.281	-1.834
TSPAN12	Downregulation by combination	0.027	0.283	-1.819
LAMB3	Downregulation by combination	0.049	0.284	-1.819
TPX2	Downregulation by combination	0.037	0.286	-1.807
GRB7	Downregulation by combination	0.016	0.289	-1.789
STRA6	Downregulation by combination	0.035	0.291	-1.782
BMP4	Downregulation by combination	0.008	0.293	-1.773
PTPN3	Downregulation by combination	0.009	0.300	-1.735
CLDN4	Downregulation by combination	0.009	0.305	-1.712
PRKG2	Downregulation by combination	0.025	0.306	-1.711
FAM222A	Downregulation by combination	0.040	0.307	-1.705
DIAPH3	Downregulation by combination	0.036	0.309	-1.694
BAIAP2L1	Downregulation by combination	0.018	0.310	-1.691
PSRC1	Downregulation by combination	0.022	0.310	-1.690
DHCR24	Downregulation by combination	0.015	0.311	-1.686
RCAN2	Downregulation by combination	0.018	0.311	-1.684
NR5A2	Downregulation by combination	0.006	0.314	-1.671
TNNT2	Downregulation by combination	0.027	0.315	-1.668
DSP	Downregulation by combination	0.033	0.316	-1.663
PLCD3	Downregulation by combination	0.027	0.317	-1.659
PALMD	Downregulation by combination	0.002	0.317	-1.656
MYBL2	Downregulation by combination	0.046	0.319	-1.648
ZC3H12C	Downregulation by combination	0.047	0.320	-1.646
PTPRF	Downregulation by combination	0.024	0.320	-1.645
BCAR1	Downregulation by combination	0.018	0.323	-1.630
UBE2C	Downregulation by combination	0.042	0.325	-1.624
PAX2	Downregulation by combination	0.022	0.325	-1.620
DUSP4	Downregulation by combination	0.027	0.326	-1.618
GPR87	Downregulation by combination	0.029	0.326	-1.616
SULT1C2	Downregulation by combination	0.047	0.328	-1.606
MIA2	Downregulation by combination	0.014	0.329	-1.605
CHSY3	Downregulation by combination	0.029	0.330	-1.601
CHRNB1	Downregulation by combination	0.037	0.330	-1.599
RASSF6	Downregulation by combination	0.009	0.330	-1.598
PHGDH	Downregulation by combination	0.032	0.333	-1.589
CCND1	Downregulation by combination	0.043	0.333	-1.586

Supplementary Table 9. Correlation between classification according to the combination rescue signature and clinic-pathological parameters.

	Immune class (n=55)	Combination responder class (n=52)	Rest (n=121)	p-value
Median age (IQR)	67 (62-73)	67 (63-71)	65 (61-71)	0.25
Gender, male (%)	42 (76)	37 (71)	101 (84)	0.12
Etiology (%)				
Hepatitis C	25 (46)	26 (52)	52 (44)	0.65
Hepatitis B	14 (26)	7 (14)	27 (23)	0.32
Alcohol	7 (13)	4 (8)	22 (19)	0.21
Others	8 (15)	13 (26)	17 (14)	0.19
Tumor size, median (IQR)	3.5 (2.5-6.5)	3.5 (2.8-5)	3.5 (2.5-7)	0.97
Multiple nodules (%)				
Absent	43 (80)	37 (71)	89 (74)	0.61
Present	11 (20)	15 (29)	31 (26)	
Vascular invasion (%)				
Absent	40 (74)	33 (65)	73 (61)	0.27
Present (micro or macro)	14 (26)	18 (35)	46 (39)	
Satellites (%)				
Absent	42 (78)	35 (67)	87 (72)	0.48
Present	12 (22)	17 (33)	34 (28)	
Degree of tumor differentiation (%)				
Well	10 (24)	4 (9)	19 (19)	0.14
Moderate/poor	31 (76)	41 (91)	79 (81)	
Bilirubin, >1 mg/dL (%)	29 (54)	23 (45)	62 (52)	0.63
Albumin, <3.5 g/L (%)	5 (9)	6 (12)	14 (12	0.92
Platelet count, <100,000/mm^3 (%)	10 (19)	9 (18)	23 (19)	1.00
AFP, >100 mg/dL (%)	19 (35)	9 (18)	26 (22)	0.09
Events (%)				
Recurrence	37 (70)	35 (67)	81 (70)	0.96
Death	28 (51)	34 (65)	71 (60)	0.30

Supplementary Table 10. Publicly available gene sets and signatures used in the study.

Name	Reference
HCC classification	
Sia HCC immune class	Sia D, et al. Gastroenterology 2017;153: 812-826
Chiang classification	Chiang D, et al. Cancer Res 2008;68:6779–6788
Hoshida classification	Hoshida Y, et al. Cancer Res 2009;69:7385–7392
Boyault classification	Boyault S, et al. Hepatology 2007;45:42–52
TGFß pathway	
Fibroblast response to TGFß	Calon A, et al. Cancer Cell 2012;22:571-84
T cell response to TGFß	Calon A, et al. Cancer Cell 2012;22:571-84
Tumor composition	
Immune cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
T cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
Th1 cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
Th2 cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
Regulatory T cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
iDC Bindea	Bindea G, et al. Immunity 2013;39:782-95
B cells Bindea	Bindea G, et al. Immunity 2013;39:782-95
Activated dendritic cells Charoentong	Charoentong P, et al. Cell Rep 2017;18:248-262
Cytotoxic T cells Jerby Arnon	Jerby-Arnon L, et al. Cell 2018;175:984-997
Endothelial cells Jerby Arnon	Jerby-Arnon L, et al. Cell 2018;175:984-997
M1/M2 Coates	Coates PJ, et al. Cancer Res 2008;68:450-6
Response to ICI	
IFN signature Ribas	Ribas A, et al. J Clin Oncol 33, 2015 (suppl; abstr 3001)
Anti-PD1 resistant melanoma	Riaz N, et al. Cell 2017;171:934-949
Nivolumab (molecular) resistant melanoma	Riaz N, et al. Cell 2017;171:934-949
On nivolumab	Riaz N, et al. Cell 2017;171:934-949
Responders on nivolumab	Riaz N, et al. Cell 2017;171:934-949