Origine et conséquences fonctionnelles des variant structuraux dans la tumorigénèse hépatique

Quentin Bayard

▶ **To cite this version:**

Quentin Bayard. Origine et conséquences fonctionnelles des variant structuraux dans la tumorigénèse hépatique. Human health and pathology. Université Paris Cité, 2021. English. NNT : 2021UNIP7302 . tel-04083183

HAL Id: tel-04083183
https://theses.hal.science/tel-04083183

Submitted on 27 Apr 2023

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

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

Université de Paris

Ecole doctorale Hématologie Oncogenèse et Biothérapies (ED561)

INSERM UMRS-1138, Centre de Recherche des Cordeliers (CRC)

Équipe 28 : Functional Genomics of Solid Tumors (FunGeST)

Origin and functional consequences of structural variations in liver tumorigenesis

Quentin Bayard

Thèse de doctorat d'Oncogenèse

Dirigée par Éric Letouzé

Présentée et soutenue publiquement le 15 Mars 2021

Devant un jury composé de :

Dr. Chunlong Chen, Chargé de recherche, Institut Curie	Rapporteur
Dr. Pascal Pineau, Directeur de recherche, Institut Pasteur	Rapporteur
Dr. Sabine Colnot, Directeur de recherche, CRC	Examinateur
Dr. Tatiana Popova, Chargé de recherche, Institut Curie	Examinateur
Dr. Olivier Delattre, Directeur de recherche, Institut Curie	Examinateur

Except where otherwise noted, this is work licensed under
https://creativecommons.org/licenses/by-nc-nd/3.0/fr/

Titre : Origine et conséquences fonctionnelles des variant structuraux dans la tumorigénèse hépatique.

Résumé :

Le cancer du foie est très hétérogène, avec un large spectre d'étiologies, de caractéristiques histologiques et de voies biologiques dérégulées. Les projets de séquençage à grande échelle réalisés durant la dernière décennie ont révélé de nombreuses mutations *driver* et des changements du nombre de copies. Cependant, le rôle des variant structuraux (VS) dans la tumorigénèse du foie est encore mal compris.

Au cours de mon projet de doctorat, j'ai développé des approches informatiques innovantes intégrant des données de séquençage du génome entier et de l'ARN pour identifier de nouveaux variant structuraux dans les cancers du foie. J'ai mis au jour des fusions récurrentes impliquant des oncogènes *ROS1* et *FRK*, conduisant à une activation constitutive de la voie JAK/STAT et au développement d'adénomes hépatocellulaires inflammatoires (AHC). J'ai également décrit un mécanisme oncogène original impliquant la perte d'éléments régulateurs post-transcriptionnels, conduisant à une surexpression massive d'*IL6* dans un cas clinique d'AHC inflammatoire associé à l'amylose. Enfin, j'ai identifié les VS *driver* affectant les éléments régulateurs dans les carcinomes hépatocellulaires (CHC), y compris les événements d'*enhancer hijacking* activant les oncogènes *CCNE1* ou *TERT*.

Afin de mieux comprendre l'origine des VS, j'ai mis en place un cadre d'analyse de signature prenant en compte la nature et la taille des variation structuraux identifiées dans chaque tumeur. Cette approche m'a permis de mettre en évidence les sous-groupes de HCC présentant des signatures de VS particulières. En particulier, j'ai découvert un nouveau sous-groupe de CHC clinique-moléculaire qui est caractérisé par l'activation de *CCNA2* ou *CCNE1* par des mécanismes *drivers*. Ces tumeurs (CCN-HCC) représentent 7% des carcinomes hépatocellulaire et présentent une signature spécifique de réarrangements (signature RS1) liés au stress réplicatif, avec des centaines de duplications focales et des « Template Insertion Cycles ». Les CCN-HCC sont des tumeurs agressives mais peuvent être ciblées par des inhibiteurs de la réponse ATR au stress réplicatif.

Dans l'ensemble, ces travaux contribuent à affiner la caractérisation moléculaire des tumeurs hépatiques bénignes et malignes en révélant de nouveaux gènes *drivers* et mécanismes moléculaires. Ces résultats ont des implications cliniques, car la définition de sous-groupes homogènes de cancers est essentielle pour développer des thérapies efficaces et adaptées.

Mots clefs :
Adénome hépatocellulaire, Carcinome hépatocellulaire, Séquençage génome complet, Variant Structuraux, Gènes drivers, Signature de réarrangements

Title: Origin and functional consequences of structural variations in liver tumorigenesis.

Abstract:

Liver cancers are extremely heterogeneous, with a wide spectrum of etiologies, histological characteristics and deregulated biological pathways. Large-scale sequencing projects in the last decade revealed numerous driver mutations and copy-number changes. However, the role of structural variations (SVs) in liver tumorigenesis is still poorly understood.

During my PhD project, I developed innovative computational approaches integrating whole genome and RNA sequencing data to identify new driver SVs in liver cancers. I unraveled recurrent fusions involving *ROS1* and *FRK* oncogenes, leading to a constitutive activation of the JAK/STAT pathway and to the development of inflammatory hepatocellular adenomas (HCA). I also described an original oncogenic mechanism involving the loss of post-transcriptional regulatory elements leading to a massive overexpression of *IL6* in a clinical case of inflammatory HCA associated with amyloidosis. Finally, I identified driver SVs affecting regulatory elements in hepatocellular carcinomas (HCC), including enhancer hijacking events activating *CCNE1* or *TERT* oncogenes.

To better understand the origin of SVs, I implemented a signature analysis framework considering the nature and size of SVs identified in each tumor. This approach allowed me to highlight HCC subgroups with particular SV signatures. In particular, I discovered a new clinico-molecular HCC subgroup driven by the activation of *CCNA2* or *CCNE1* by various mechanisms. These tumors (CCN-HCC) represent 7% of HCC and display a specific pattern of rearrangements (signature RS1) related to replication stress, with hundreds of focal duplications and templated insertion cycles. CCN-HCC are aggressive tumors but may be targetable by inhibitors of ATR response to replication stress.

Altogether, this work helps to refine the molecular characterization of both benign and malignant liver tumor by revealing new driver genes and molecular mechanisms. Those findings have clinical implications, as the definition of homogeneous cancer subgroups is essential to develop efficient tailored therapies.

Keywords:

Hepatocellular Adenoma, Hepatocellular Carcinoma, WGS, Structural Variations, Driver genes, Rearrangement Signature

Remerciements

Avant tout, je tiens à remercier mon jury de thèse pour avoir accepté d'examiner mes travaux de doctorat. C'est un honneur pour moi d'avoir l'opportunité de présenter l'aboutissement de mes travaux de thèse à des scientifiques experts dans leurs domaines. L'apport de votre expertise fera de ma soutenance de thèse une expérience des plus enrichissantes pour moi sur le plan scientifique.

Quatre années se sont écoulées depuis mon arrivée dans le laboratoire FunGeST, et j'en garderai un excellent souvenir, que ce soit sur le plan professionnel mais aussi personnel.

Pour cela, je remercie Jessica Zucman-Rossi, pour m'avoir accueilli dans son laboratoire et offert la possibilité de travailler sur la collection de données de séquençage NGS de HCC, tout à fait impressionnante. Par ailleurs, la richesse des raisonnements scientifique que tu apportes en lab-meeting m'a grandement aidé dans mes propres réflexions.

Je tiens à remercier aussi Éric Letouzé, mon directeur de thèse. J'ai appris énormément pendant ces quelque année et en grande partie grâce à ta pédagogie et ton pragmatisme. Tes conseils pertinents et les discussions qu'on a pu avoir m'ont permis de clarifier à de nombreuses reprises mes objectifs et ouvrir le chemin à de nouvelles réflexions prometteuses. Tu es un directeur de thèse bienveillant, de bonne humeur et toujours pertinent dans les raisonnements scientifiques. Et puis ce n'est pas n'importe quel directeur de thèse qui accompagne ses étudiants à l'Oktoberfest ! Batman approuve !

Je remercie chaleureusement tous mes collègues qui ont rendu si agréable cette expérience. Que ce soit en meetings, en congrès, mais aussi en pause déjeuner, pause-café ou after-work, j'ai passé des moments vraiment agréables avec vous tous et j'ose espérer retrouver une ambiance similaire dans mes prochaines aventures professionnelles.

Je remercie particulièrement les gens avec qui j'ai travaillé étroitement : Merci Sandrine pour toute les questions auxquelles tu m'as aidé à répondre ainsi que les pauses café du matin, merci Stefano pour ta 'jeunesse' éternelle et tes connaissances en stat, merci Théo pour ton énergie débordante et ta curiosité scientifique, merci Léa pour ton humour et le soutien mutuel inhérent à deux thésards contemporains. Merci à Eric T., Jayendra, Benedict, Raphaël, Amélie, Laura, Lisa, Camille, Anna-Line, Ana, Hélène, Clément, Jie, Jill, Massih, JC, JB, Guillaume, Julie, Maya, Elvire, Michalina, Alix, Pan Long, Gabrielle, Barkha, Samantha, Sandra, Didier et François…. Vous êtes nombreux et pourtant chacun d'entre vous a rendu ce 'voyage' exceptionnel.

Merci à ma famille, je regrette que je vous ne puissiez pas venir à ma soutenance à cause de la situation sanitaire. J'ai hâte de vous retrouver !

Merci à mes amis les plus proches, la compote complète (Tanguy, Kelbeu et Palmito) ainsi qu'Alexandre et Mika avec qui j'ai partagé des moments d'exception depuis 4 années et qui m'ont beaucoup apporté sur le plan personnel.

Enfin je remercie Tiziana, qui m'a grandement soutenu pendant cette dernière ligne droite et sans qui la dernière année covidesque aurait été bien déprimante...

ABBREVIATIONS

AA	Amyloid A
AAV2	Adeno-associated Virus Type 2
ATR	Ataxia Telangiectasia and Rad3-related
BCLC	Barcelona Clinic Liver Cancer
bex3HCA	Betacatenin exon 3 HCA
bex7,8HCA	Betacatenin exon 7/8 HCA
BFB	Breakage-Fusion-Bridge
BIR	Break-Induced Replication
BWA	Burrow-Wheeler Aligner
CCN	Cyclin
cDNA	Complementary DNA
CFS	Common Fragile Sites
CGC	Cancer Gene Census
ChIP-Seq	Chromatin Immunoprecipitation sequencing
CNA	Copy-number alterations
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRP	C-Reactive Protein
CTCF	CCCTC-binding factor
DNA	DesoxyriboNucleic Acide
DNase	DeoxyriboNuclease
DRR	DNA Damage Response
DSBs	Double Strand Breaks
ENCODE	Encyclopedia of DNA Elements
FDA	Food and Drug Administration
FGFR	Fibroblast Growth Factor Receptors
FLC	Fibrolamellar Carcinoma
HBV	Hepatitis B virus
HCC	HépatoCellular Carcinoma
HCV	Hepatitis C virus
HHCA	HNF1A activated HCA
HI-C	HIgh-throughput Chromatin conformation capture
HPV	Human Papilloma Virus
HR	Homologous recombination
ICGC	International Cancer Genome Consortium
ID	small Insertions and Deletions signature
IHCA	Inflammatory HCA
LICA-CN	Liver Cancer - CN
LICA-FR	Liver Cancer - FR
LIHC-US	Liver Hepatocellular carcinoma - TCGA, US
LINC-JP	Liver Cancer - NCC, JP
LIRI-JP	Liver Cancer - RIKEN, JP
MCC	Merkel Cell Carcinoma
MCPyV	Merkel Cell Polyomavirus
MODY	maturity-onset diabetes of the young
mRNA	messenger RNA

NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic SteatoHepatitis
NER	Nucleotide Excision Repair
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NK-cells	Natural Killer Cells
NMF	Non-negative Matrix Factorization
OC	Oral Contraception
pancan	pan-cancer
PARP	poly ADP ribose polymerase
PCAWG	Pan-cancer Analysis of Whole Genomes
PD-1	Programmed cell death-1
PDGFR	Platelet-derived growth factor receptor
PEI	Percutaneous Ethanol Injection
qPCR	quantitative Polymerase Chain Reaction
RFA	Percutaneous radiofrequency ablation
RNA	RiboNucleic Acide
RNA-Seq	RNA Sequencing
ROS	Reactive Oxygen Species
RS	Rearrangement Signature
SAA	Serum Amyloid A
SBS	Single Base Substitution signature
SH2	Src Homology 2
SH3	Src Homology 3
shHCA	sonic hedgehog HCA
SMC	Structural Maintenance of Chromosomes
SNVs	Single Nucleotide Variations
SVs	Structural Variations
TACE	Trans-Arterial ChemoEmbolization
TADs	Topologically Associated Domains
TCGA	The Cancer Genome Atlas
TC-NER	transcription-coupled nucleotide excision repair
TERTp	TERT promoter
TIC	Template Insertion Cycle
TSG	Tumor Suppressor Gene
UHCA	Unclassified HCA
UTRs	UnTranslated Regions
UV	Ultra-Violet
VEGFR	vascular endothelial growth factors receptors
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WNT	Wingless/Integrated

INDEX

I INTRODUCTION

1. Cancer is a disease of the genome

Cancer is the second leading cause of death worldwide after cardiovascular diseases and correspond to a major public health concern, with 9.6 million deaths from cancer worldwide in 2018 (Ferlay *et al.*, 2019). First description of cancer was written approximately 2600BC, found in the Edwin Smith Papyrus, where Egyptian physician Imhotep describes breast cancer as a "bulging mass in the breast" that was resistant to any known treatment available at that time (Mukherjee, 2010). Medicine development during modern era led in the 17th and 18th centuries to various observations that directed strong links between chemical exposures and cancer development. Of note, in 1777, Percival Pott from London described an enrichment of scrotal cancers in chimney sweeps, which was caused by soot that was collected in the skin folds of the scrotum (Brown and Thornton, 1957). However, our understanding of cancer at the cellular and molecular level was initiated by its first scientific description using modern microscopes (Lin, 1983). Then, in the early 20th century, David Von Hansemann and Theodor Boveri's work led suggested that all cancers arise from a single cell undergoing genetic alterations that lead to abnormal behavior and uncontrolled cell division (Boveri, 1914). In the 1950's, Watson and Crick's findings on the double helix structure of DNA drastically accelerated genome comprehension (Watson and Crick, 1953). Three key studies paved the way for modern cancer genomics, as they described for the first time that cancers originate from somatic alterations in the human genome:

- In 1973, the Philadelphia chromosome (translocation between chromosomes 9 and 22) in chronic myeloid leukemia was the first identified genomic alteration related to cancer (Rowley, 1973).
- Ten years later, a single nucleotide base substitution (G>T) in the *HRAS* gene was shown to induce neoplastic cell transformation (Reddy *et al.*, 1982).
- Another study showed the requirement of src gene in Rous sarcoma virus to form neoplasm into chicken cells (Bishop, 1983).

Cancer cases dramatically increased in recent decades; in 2018 there were 18.1 million cases and 9.6 million deaths reported worldwide (Ferlay *et al.*, 2019), justifying the global research effort invested every year. The identification of molecular alterations at the origin of cancer has led to the development of effective targeted treatments in some specific tumor types. However, molecular studies revealed tremendous inter- and intra-organ heterogeneity. An exhaustive characterization of cancer subtypes and the molecular mechanisms driving them is essential for the development of effective treatments and diagnosis tools.

1.1. Hallmarks of cancer

The year 2000 marked the first sketch of important cellular capabilities fostering cancer development (Hanahan and Weinberg, 2000). This study highlighted six essential alterations in cell physiology that collectively dictate malignant growth (Figure 1):

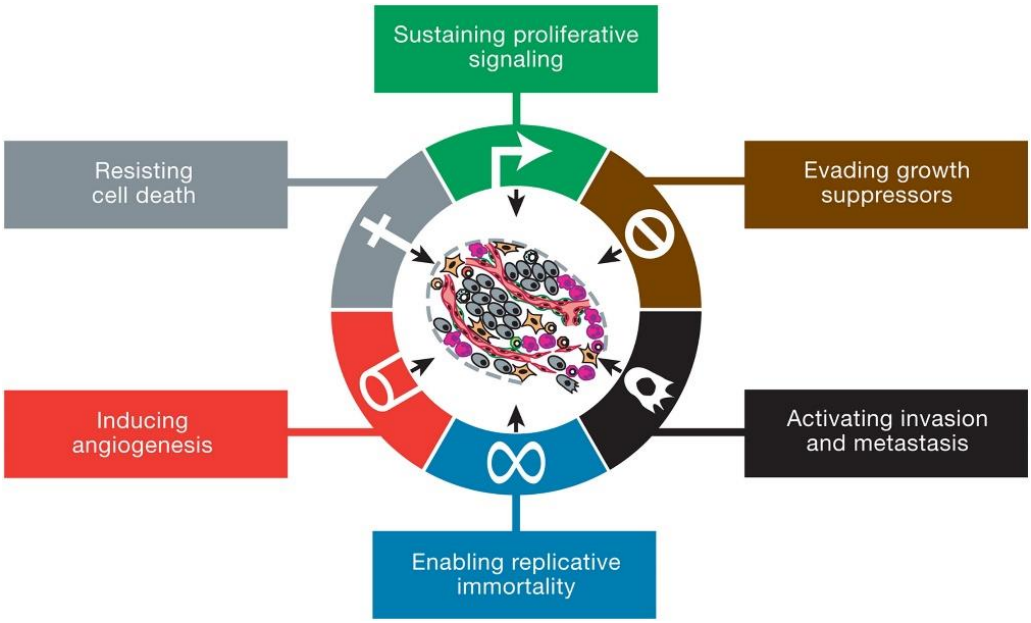


Figure 1: Representation of six functional capabilities acquired by most if not all cancers during their development. Those biological features can be acquired through various mechanisms (Hanahan and Weinberg, 2000).

- self-sufficiency in proliferative signaling;
- insensitivity to growth suppressors (antigrowth) signals;
- evasion of programmed cell death (i.e.: apoptosis);
- limitless replicative potential;
- sustained angiogenesis;
- tissue invasion and metastasis.

Some capabilities are required from the first steps of tumor growth, including self-sufficiency in proliferative signaling, insensitivity to growth suppressor signals, evading apoptosis and a limitless replicative potential. Other, like angiogenesis and metastasis potential, become necessary as the tumor develops. Additional knowledge gained from genomic and functional studies led Hanahan and Weinberg to update their landscape of tumor cell capabilities with two additional hallmarks (Hanahan and Weinberg, 2011). Firstly, tumor cells have the capability to modify or reprogram cell metabolism to accelerate neoplastic proliferation. Secondly, cancer cells are able to evade immunological destruction, in particular from T and B-lymphocytes, macrophages and NK-cells. In this update, authors also described two additional "enabling characteristics" of cancer cells that facilitate the acquisition of the hallmarks: genomic instability that accelerates the acquisition of driver events, and inflammation that contributes multiple hallmarks by supplying growth factors, pro-angiogenic factors, and extracellular matrix-modifying enzymes to the tumor microenvironment, sustaining proliferation, angiogenesis, invasion, and metastasis (Hanahan and Weinberg, 2011).

The alteration of functional capabilities mentioned above results from genomic alterations that can be inherited from parent genetic material or accumulated along cell divisions due to intrinsic mutational processes or environmental exposures. However, not all genomic alterations contribute to malignant transformation or cancer progression. Genomic alterations that provide a selective growth advantage and contribute to cancer development are called "driver" mutations; those that do not are termed "passenger" mutations (Stratton, Campbell and Futreal, 2009). Thus, the accumulation of driver events leads to more and more aggressive tumor cells. In some cases, disruption of DNA repair pathways can induce mutator phenotypes, accelerating the acquisition of additional drivers (Figure 2).

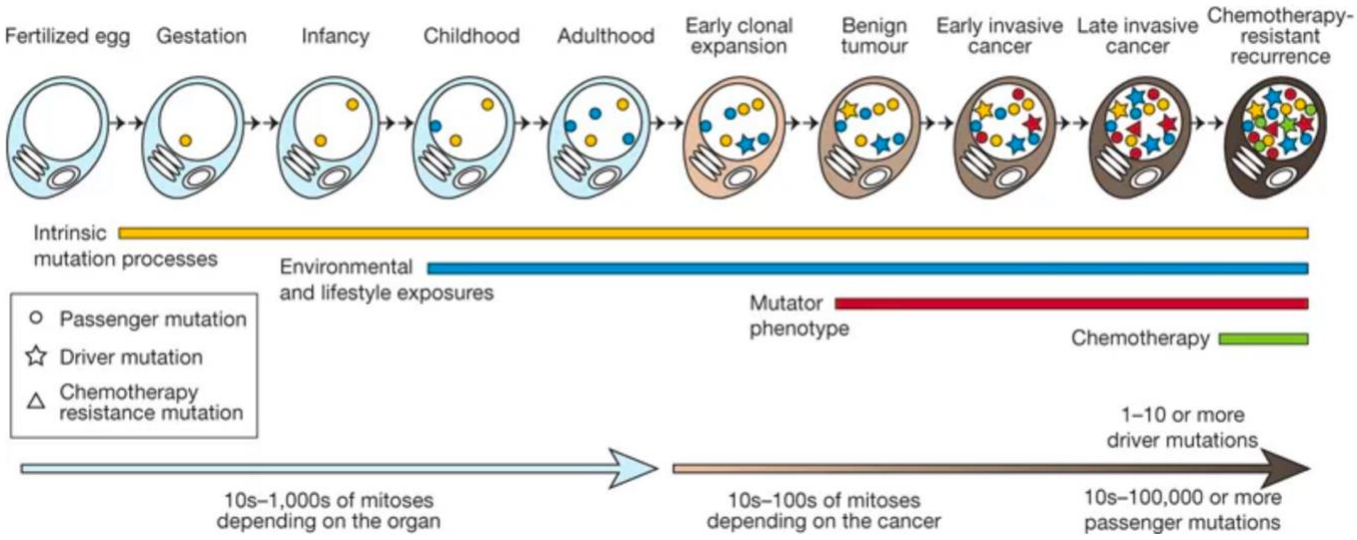


Figure 2: Model of somatic mutation accumulation in a cell lineage, contributing to cancer progression, through intrinsic mutational processed and environmental exposures. Those alterations are selected by giving cell a selective advantage, leading to clonal expansion cancer development. (Stratton, Campbell and Futreal, 2009)

The age distribution of cancer mortality rates let the statisticians Armitage and Doll to predict that 5 to 8 driver events were necessary for cancer development (Armitage and Doll, 1954). In agreement with this prediction, the Pan-Cancer Whole Genome (PCAWG) analysis from the International Cancer Genome Consortium (ICGC) recently demonstrated that cancer genomes harbor on average 4-5 driver mutations (Campbell *et al.*, 2020). Finding these drivers is like looking for a *needle in a haystack* as thousands of passenger mutation accumulate from the zygote to the final tumor cell. Genes that have been identified as drivers in at least one cancer type are described as cancer genes and are categorized as oncogenes or tumor suppressor genes (TSG). Oncogenes are activated by mechanisms increasing their expression (e.g. chromosome amplifications) or missense mutations at specific hotspot regions increasing the activity of the protein. Tumor suppressors are inactivated by chromosome rearrangements (e.g. deletions), inactivating missense mutations or nonsense/frameshift mutations dispersed across the gene (Kern *et al.*, 2006; Winter, Brody and Kern, 2006). Thus, cancer is a disease rooted in our genes, triggered by specific alterations in DNA, changing the biological equilibrium of a single cell and leading to clonal expansion. Recent evolution in sequencing technology lead to a discipline known as cancer genomics, aiming at unraveling genomic alterations to better understand cancer and develop more efficient treatments.

1.2. Next-generation sequencing of cancer genomes

Pioneer works by Frederik Sanger on the development of the first generation of automated DNA sequencers (Sanger sequencing) and further technological improvements allowed the Human Genome Project to produce the first draft of the human genome in early 21th century (Craig Venter *et al.*, 2001; Lander *et al.*, 2001). This project paved the way for modern high throughput sequencing method, termed Next Generation Sequencing (NGS). Three methods are extensively used in cancer genomics: whole exome, whole genome and RNA sequencing (Figure 3).

Whole exome sequencing (WES) involves the capture and sequencing of the coding regions (around 1%) of the genome and is massively used to identify coding cancer driver mutations. Two main protocols can be used for the capture step:

- In amplicon-based capture, hybridization probes are designed based on sequences of annotated exons available in transcriptome libraries, such as Ensembl (Cunningham *et al.*, 2019), and fixed to a microarray. After extracting and fragmenting the hybridized DNA, sequencing adapters are attached to the double-stranded DNA fragment in order to produce short sequences, which correspond to both ends of the fragment.
- In-solution hybridization capture methods involve fragmenting double-stranded DNA and ligating adapters to those fragments. Then, target sequences are labeled using a pool of custom oligonucleotides designed to capture annotated exons, as in previous method. Those oligonucleotides are labeled with beads, allowing the separation of sequences of interest.

Finally sequencing step will produce short-read reads covering the exons of the genome.

Exponential decrease of sequencing costs allowed the democratization of ***whole genome sequencing (WGS)*** in cancer genomics. The protocol consists in extracting and fragmenting DNA of a tissue. Then, adapters are added to each end of the fragments and the DNA library is sequenced. Contrary to WES, that only allows the analysis of exonic mutations and copy-number alterations, WGS reveals mutations and structural variations in non-coding RNA genes and regulatory regions (enhancers, transcription-factor binding sites, insulators) that can profoundly alter the expression of cancer genes.

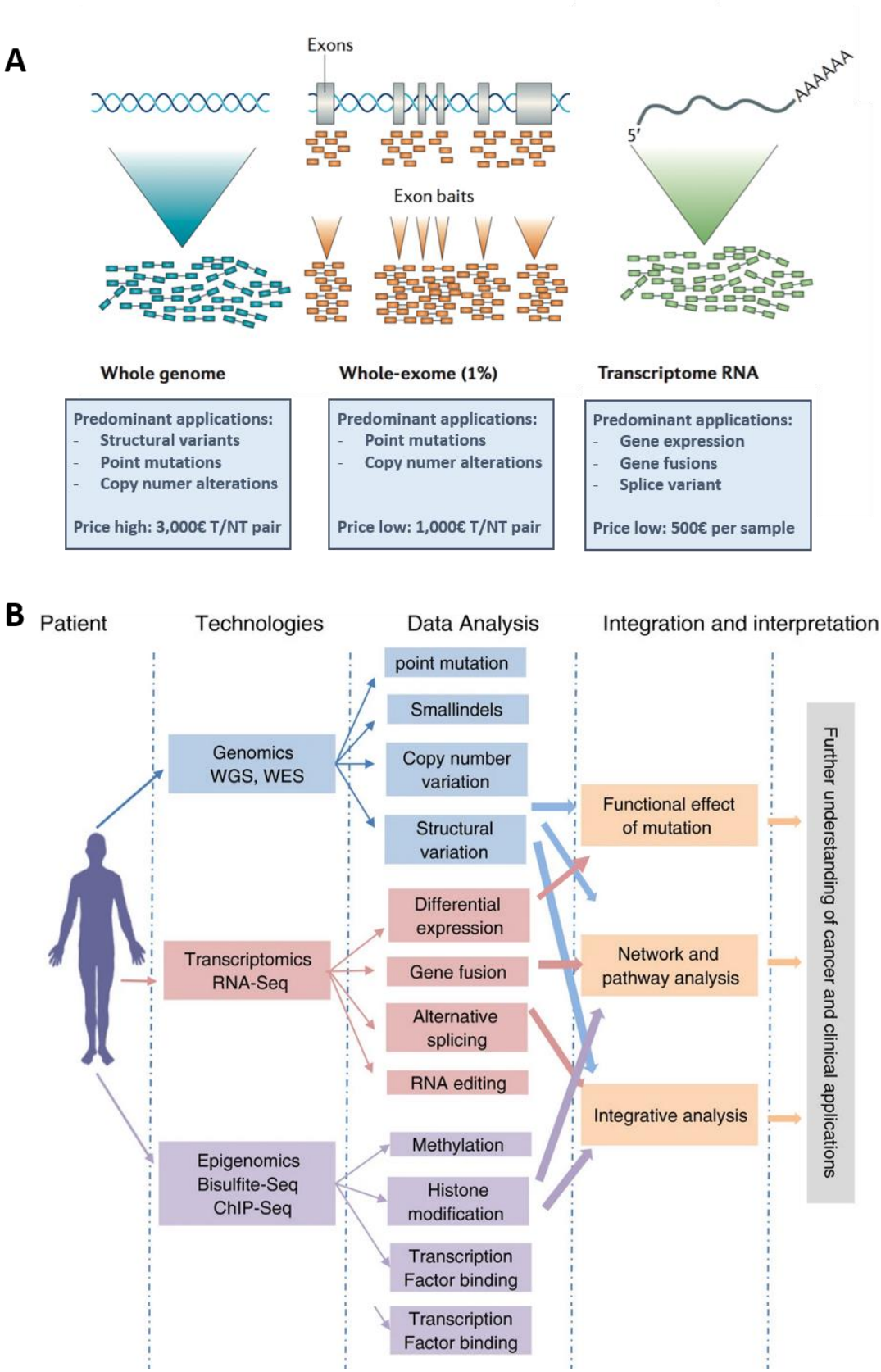


Figure 3: Next Generation Sequencing in cancer genomics. A: Scheme representing the material sequenced in WGS, WES and RNAseq. WGS gives access to the whole genome but is the most expensive technique. WES is a cheaper alternative, but only targets the coding part of the genome. Finally, RNAseq purpose is to sequence the transcriptome, giving a different information that can be integrated with genomic data to identify functional alterations. B: The 3 above-mentioned methods, additionally with innovative implementation of epigenetics sequencing (ChIP-Seq, HI-C, Bisulfite-Seq) must be integrated to identify genomic/epigenetic alterations and their functional consequences to better characterize cancer deregulations and foster the development of tailored therapies. (Liu et al, 2013).

The 3rd method, ***RNA sequencing (RNAseq)***, gives access to the transcriptome. Two protocols exist for RNAseq. The first protocol involves isolating RNA using a deoxyriboluclease (DNase) and sequencing the whole RNA material in a sample. The second method consists in capturing messenger RNAs with 3'polyadeylated (poly(A)) tail by mixing RNA with polyT oligomers covalently attached to magnetic beads. Then, RNA material from both methods can be converted into double-stranded cDNA fragments and sequenced using classical NGS workflow. This method is complementary with WGS and WES as it gives access to gene expression changes and abnormal transcripts such as fusion genes and alternative splicing events.

Besides the information provided by each NGS technology, integration of WGS with RNAseq is a powerful way to reveal driver events affecting the expression of genes or transcript structure. For example, mutations in gene promoters can both increase or decrease the expression level of genes, by changing the affinity of cis regulatory elements (Nault and Zucman-Rossi, 2016). Splice site mutations can alter the splicing process of the pre mRNA and modify the structure of the mRNA.

Finally, recent development of alternative sequencing approaches expands the possibilities of cancer genomics. These sequencing technologies allow identifying epigenetic alterations, corresponding to modifications in the genome structure or regulation, without direct alteration of the sequence. Recent studies have exploited those types of data to better understand the causes and consequences of cancer driver alterations:

- ChipSeq (chromatin immunoprecipitation coupled with sequencing) consists in sequencing the DNA regions bound by a specific protein. This method allows identification of transcription factor binding sites, histone modifications and more (Raha, Hong and Snyder, 2010). This method is largely used to understand epigenetic deregulations, at the origin of driver gene expression changes that could cause tumorigenesis (Bender *et al.*, 2013)
- Hi-C (chromatin conformation capture) data provide contact probabilities between genomic regions at high resolution. This information is fundamental to understand some alterations activating distant region of the genome by changing genome 3D conformation (Lieberman-Aiden *et al.*, 2009).

- RepliSeq data give a high resolution view of replication timing along the genome, using a differential marking of nascent DNA during replication (Hansen *et al.*, 2010). Replication timing is indeed known to influence spatial organization of the genome, and thus shape the mutational landscapes of cancer genomes (De and Michor, 2011).

1.3. Unraveling driver events

Computational analysis of NGS data involves two main steps to unravel the driver events leading to tumorigenesis:

- First, the identification of somatic alterations in each tumor by comparing sequencing data of tumors versus its paired healthy tissue or a pool of non-tumor data.
- Second, the identification of functional events, supported by either a significant recurrence or a functional proof that the alteration of a gene favors tumor development.

1.3.1. Identifying somatic alterations

Once the raw sequencing data has been aligned to a human genome reference with algorithms like BWA mem (Li *et al.*, 2009), various tools are applied to unravel different types of somatic alterations:

- Somatic mutations can be called using algorithms that identify mismatches between aligned reads and the reference genome and compare the distribution of variant reads between the tumor and matched normal sample. The main challenge is to distinguish real variants from sequencing errors. Tumor heterogeneity and contamination of clinical samples with normal cells makes this task challenging. MuTect2, developed by the Broad Institute, is one of the most popular algorithm to identify somatic point mutations and small insertions and deletions (indels) (Benjamin *et al.*, 2019). It uses two Bayesian classifiers: the first one detects whether the tumor is non-reference at a given genomic position and, for those sites that are found as non-reference, the second classifier makes sure the non-tumor sample does not carry this specific variant. Post-filtering steps are also included to

remove recurrent artifact mutations like variants found in a panel of non-tumor samples or occurring in a region with many clustered variants (Figure 4).

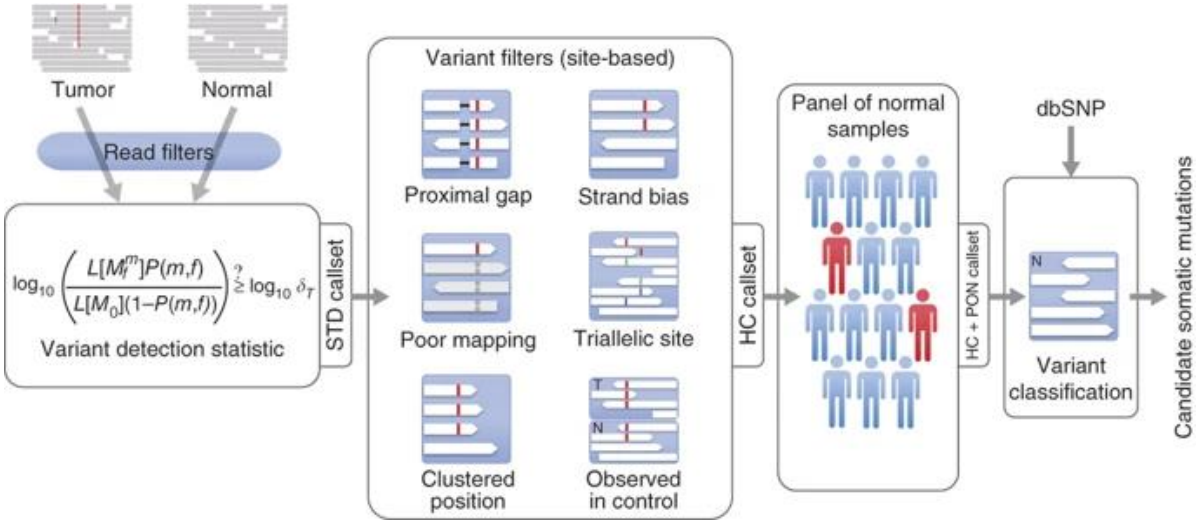


Figure 4: Reads of sequencing generated from tumor and non-tumor sample are differentially aligned to the reference genome. After initial calling of variant by the Bayesian classifier, post filtering is used to remove first recurrent artefactual results, but also putative somatic mutations that are compared to a panel of normal samples to remove residual false-positives, caused by rare errors (Cibulskis et al,. 2013).

- Copy-number alterations (CNA) analysis is based on the depth of sequencing. The genome is split in equal-size regions, and the median coverage in the tumor and non-tumor sample (WGS or WES) is calculated in each window. The log-ratio of tumor versus non-tumor depth is calculated for each bin i :

$$LRR_i = log_2\left(\frac{Depth_{T_i}}{Depth_{NT_i}}\right)$$

LRR can then be processed by simple segmentation algorithm, that compare the LRR points distribution along the chromosomes and extract segments with a similar LRR distribution. Then amplifications, gains, deletions and homozygous deletions can be estimated genome wide. Additionally some methods, such as the Genome Alteration Print (Popova *et al.*, 2009), were developed to exploit both LRR changes and allelic imbalance at germline polymorphisms (B Allele Frequency, BAF). In other words, BAF is the normalized proportion of arbitrary established allele B in a two allele mixture (Figure 5). This information can be very useful in case of copy-neutral loss of heterozygosity (cn-LOH), where a parental allele is lost and the other duplicated, that can foster tumorigenesis for example by deregulating imprinted regions (O'Keefe, McDevitt and Maciejewski, 2010).

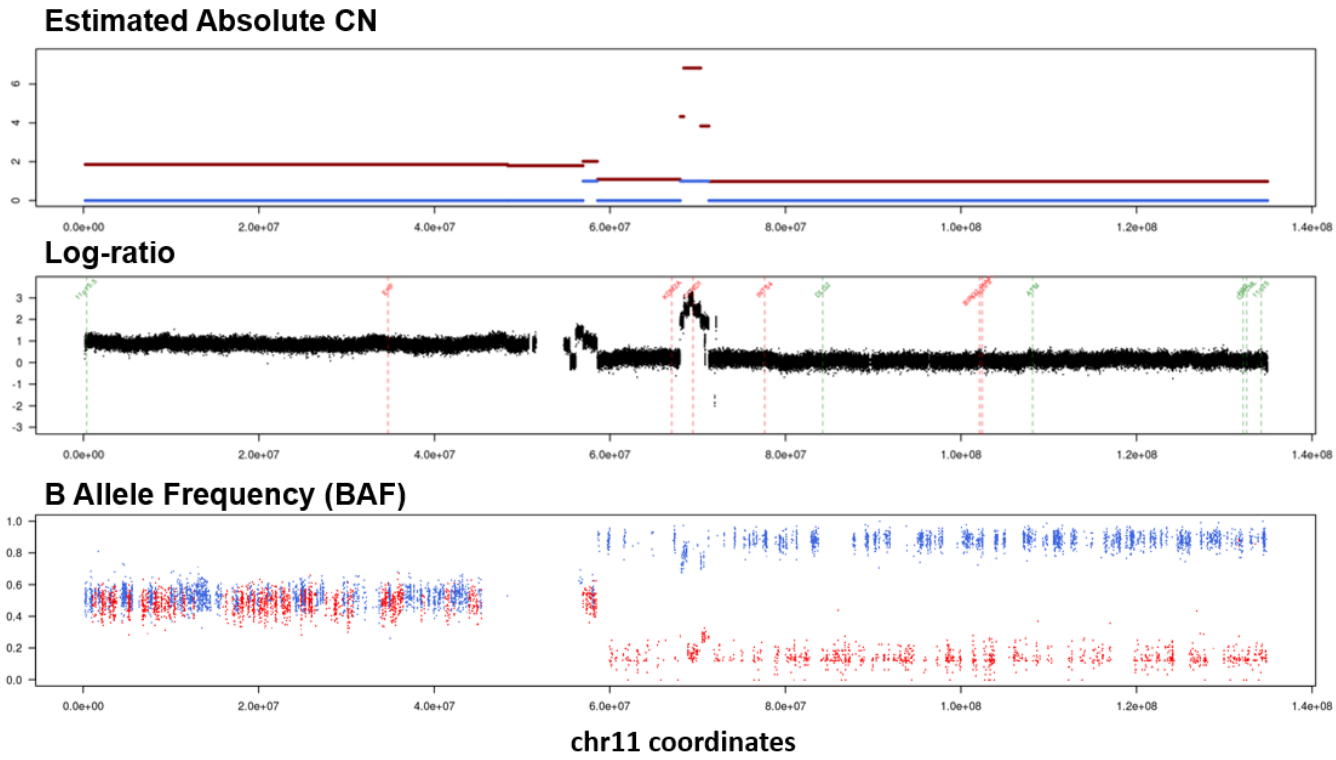


Figure 5: CCND1 amplification identified in an HCC sample by WGS. *Top: estimated absolute copy number. Red line corresponds to the total number of copy at each position, while the blue correspond to the minority allele. Middle,* log-*ratio estimated from tumor and non-tumor sequencing coverage. Bottom: B Allele Frequency (BAF). Blue dots represent the fraction of major allele while red dots represent the minor allele fraction. This track allows to identify copy neutral LOH, when a parental allele is removed and replaced by the other parental allele. This can lead to deregulation of parentally imprinted loci.*

- Finally, Structural Variations (SVs) are more and more described to be involved in cancer development. Manta (Chen *et al.*, 2016) is a tool developed by Illumina for SV calling, that operates in two phases: first a graph of all distant junction of the genome is built where edges represent putative somatic abnormal junction between 2 distant genomic regions. In this step, the entire reads mapped on the genome are scanned to fine evidence of i) multi-mapped individual reads, ii) mate-reads aligned on two distant regions, supporting a long range association or iii) reads whose mate is not mapped at all (Figure 6A-B). In the second step, Manta analyses these individual graph edges and groups highly connected edges to unravel and score SVs of high confidence (Figure 6C). Robust identification of SV somatic events can be more challenging than detection of SNVs. Indeed, recent benchmark of somatic variant callers by the ICGC consortium showed that the sensitivity of identification is way lower for SVs than for SNVs on simulated data (Alioto *et al.*, 2015). This results in an under-representation of SVs among driver alterations in cancer (Huddleston and Eichler, 2016; Audano *et al.*, 2019).

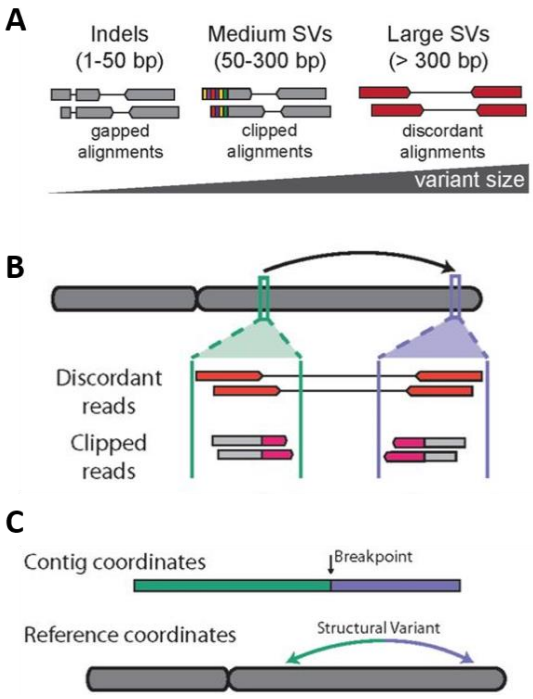


Figure 6: Calling of SV from short-read WGS. A: Representation of the different read pairs that can be informative to find evidences of SV. On the left, gapped alignment are the consequence of small indels, whose size is of the same order of magnitude than read length. For medium sized SV, most informative reads are soft-clipped alignment, where one of the two mate pair is half aligned on both distant contig corresponding to the abnormal junction. Also this kind of mapping is used to obtain the precise coordinate of the breakpoint. Finally, for large SVs, mate-pair reads, aligning on very distant regions of the genome are informative of abnormal junction. B: Chromosomal representation of a distant abnormal junction between green and purple loci. Different supporting reads are schematized below. C: Reconstruction of the rearranged molecule from supportive reads. Modified from Jeremiah Wala et al 2018.

1.3.2. Recurrence and genomic covariates

Using the abovementioned methods, we can obtain the catalogue of somatic alterations in a series of tumors. The next step is to discriminate driver from passenger alterations. The most widely used approach is to look for significantly recurrent events, as drivers should be selected in multiple tumors whereas passengers occur by chance and should be different across tumors. Several studies, from the lab and others, have used this approach to define the landscape of driver genes in liver cancers from recurrent mutations and focal CNAs (Schulze, Imbeaud, Letouzé, *et al.*, 2015; Fujimoto *et al.*, 2016).

Importantly, several genomic covariates impact the mutation rate along the genome such as chromatin organization, DNA accessibility, local base composition, gene expression and DNA replication timing (Lawrence *et al.*, 2013; Polak *et al.*, 2014; Morganella *et al.*, 2016; Glodzik *et al.*, 2017). Key genomics covariates are detailed below:

- **Chromatin accessibility.** Chromatin is the complex of DNA and proteins corresponding to chromosomes. Nucleosomes are the protein complexes that regulate the 3D architecture of chromatin, in order to maintain cell phenotype (Woodcock and Ghosh, 2010; Gibcus and Dekker, 2013). Nucleosomes are involved in differential DNA packaging, by maintaining two different chromatin

states: relaxed (euchromatin) and condensed (heterochromatin). These differential 3D states have a key role in gene regulation, as open chromatin is required for active gene expression. Besides regulation, 3D conformation also has an impact on mutation rate. Indeed, heterochromatin and repressive histone marks like H3K9me3 are associated with elevated mutation rates, whereas euchromatin and active marks like H3K27ac have lower mutation rates (Schuster-Böckler and Lehner, 2012; Polak *et al.*, 2014, 2015) (Figure 7A). This may be due to restricted activity of DNA repair enzymes to dense chromatin regions.

- **Replication.** DNA replication processes segments of ~400-800kb in a specific order during the S phase of the cell cycle. Thus, the genome of a specific cell type can be divided into early and late replication domains. Late-replicating regions are known to have higher mutation rates (Lawrence *et al.*, 2013; Sima and Gilbert, 2014) (Figure 7B). This change in mutation rate was shown to be a consequence of mutations occurring after the inactivation of the DNA mismatch repair (MMR) mechanism in late replicated regions (Chen *et al.*, 2010; Supek and Lehner, 2015). Moreover, some SVs tends to be enriched in late replicated regions, congaing Common Fragile Sites (CFS) (Glover, Wilson and Arlt, 2017; Gómez-González and Aguilera, 2019), while some SVs such as duplications are shown to be enriched in early replicated regions (Li *et al.*, 2020).

- **Transcription.** Mutation rate is inversely correlated with gene expression level (Lawrence *et al.*, 2013) (Figure 7C). The transcription-coupled nucleotide excision repair (TC-NER) can efficiently remove mutations occurring on the transcribed strand. This mechanism reduces the overall mutation burden in expressed genes and results in a mutational asymmetry between the two DNA strands aptly known as transcriptional strand bias (Haradhvala *et al.*, 2016). Finally, a strong enrichment of mutations at transcription factor binding sites was demonstrated in cell lines, as a consequence of decreased nucleotide excision repair (NER) activity (Sabarinathan *et al.*, 2016) (Figure 7D).

Interestingly, recent studies have shown that different mutational processes can have different interactions with genomic covariates (Morganella *et al.*, 2016). For example, mutational processes related to bulky DNA adducts (tobacco, aflatoxin B1) are strongly repressed in highly expressed genes due to the removal of adducts by TC-NER, whereas a mutational signature related to alcohol consumption is activated in the same regions (Letouzé *et al.*, 2017).

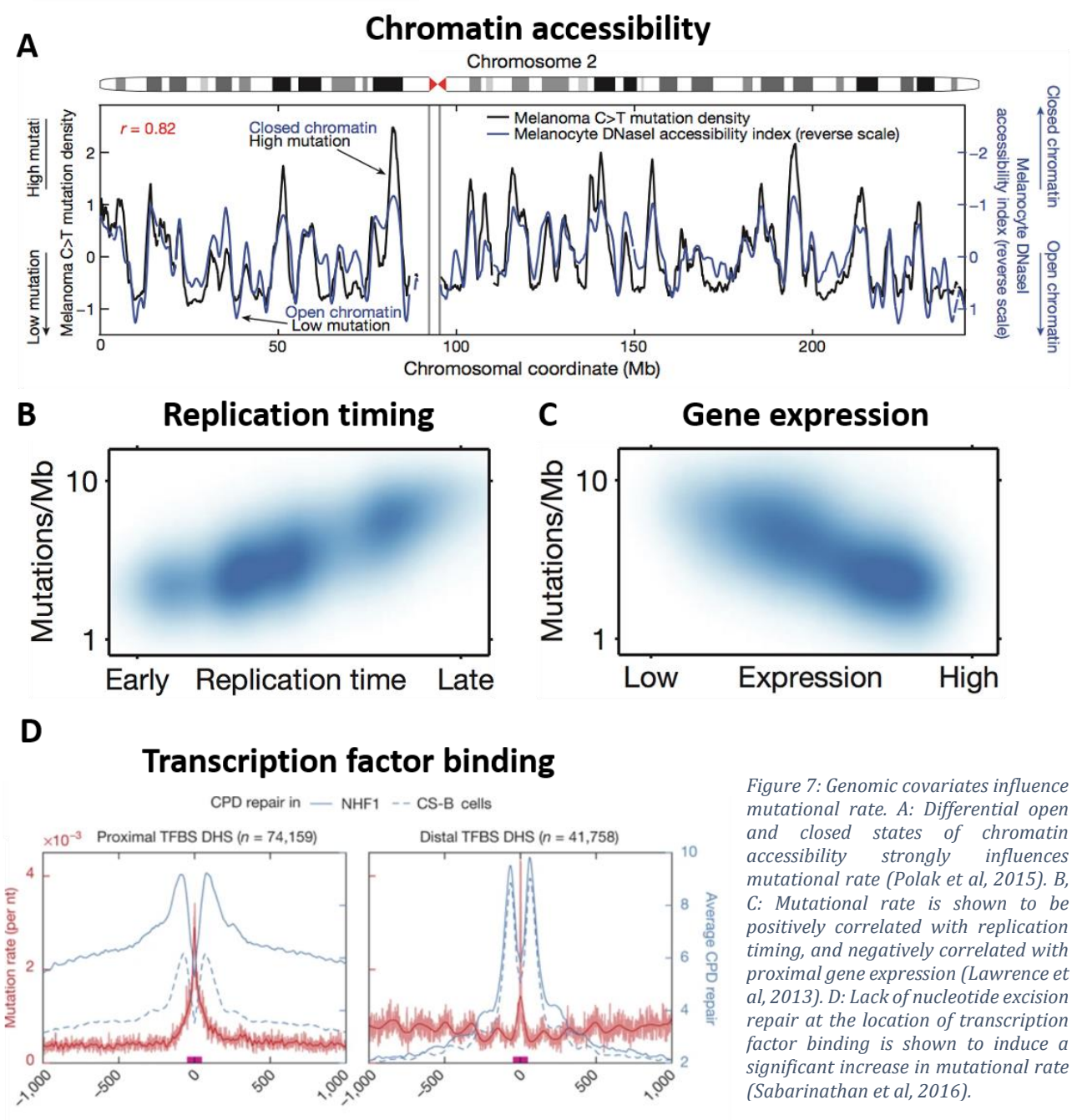


Figure 7: Genomic covariates influence mutational rate. A: Differential open and closed states of chromatin accessibility strongly influences mutational rate (Polak et al, 2015). B, C: Mutational rate is shown to be positively correlated with replication timing, and negatively correlated with proximal gene expression (Lawrence et al, 2013). D: Lack of nucleotide excision repair at the location of transcription factor binding is shown to induce a significant increase in mutational rate (Sabarinathan et al, 2016).

Those genomic covariates can bias the analysis of mutation recurrence by generating hotspots due to high local mutability rather than functional selection of alterations. Tools like MutSigSV (Lawrence *et al.*, 2013), developed by the Broad institute, take into account

the abovementioned parameters and the background distribution of mutations in the tumor series to identify significantly recurrent mutations while controlling for genomic covariates. Similarly, a regression approach taking into account various (epi)genomic features was recently proposed to identify significantly recurrent SVs not explained by local genomic context in breast cancer genomes (Glodzik *et al.*, 2017). Other methods have been developed to identify recurrent CNAs, like GISTIC2.0 (Mermel *et al.*, 2011) that takes into account both the frequency and amplitude of CNAs to define hotspots of gains and deletions.

1.4. The Pan Cancer Atlas of driver alterations

In the last decade, many sequencing projects were conducted by individual laboratories or international consortia like TCGA (The Cancer Genome Atlas; https://www.cancer.gov/tcga) and ICGC (International Cancer Genome Consortium; https://icgc.org/), producing WES, WGS and/or RNAseq data from thousands of tumors spanning most cancer types. This scientific effort allowed to establish the landscape of driver genes involved in each cancer.

1.4.1. Cancer Gene Census

The Cancer Gene Census (CGC) is an ongoing effort to catalogue those genes, which contain mutations causally implicated in cancer. Regularly updated by the COSMIC (Sondka *et al.*, 2018) literature curation team, 719 cancer-driving genes are now described with a variety of alteration mechanisms across all human cancer types. This pan-cancer list of driver genes is a useful resource to annotate new data with known drivers. However, it is important to keep in mind that rare alterations remain to be identified, especially in rare tumor entities.

1.4.2. PCAWG (Pan Cancer Analysis of Whole Genomes)

The most comprehensive pan-cancer genome analysis to date is probably the ICGC-PCAWG (https://dcc.icgc.org/pcawg) published in Nature in early 2020. The PCAWG consortium analyzed whole genome sequences from 2,600 primary cancers and their matching normal tissues across 38 distinct tumor types (Campbell *et al.*, 2020), to define an exhaustive landscape of driver genes and reconstruct the natural history of cancers.

Across the PCAWG tumors, 43,778,859 somatic SNVs, 2,418,247 somatic indels and 288,416 somatic SVs were identified and used to define an integrated set of driver events.

91% of tumors had at least one identified driver mutation, with an average of 4.6 drivers per tumor and 2.6 coding driver mutations per tumor, consistent with previous TCGA publication (Martincorena *et al.*, 2017). 25% of PCAWG tumors harbored at least one putative non-coding driver mutation, with 1/3 of them affecting *TERT* promoter (9% of PCAWG tumors). This underlines the benefit of WGS over WES to define the exhaustive set of driver events in a tumor genome (Figure 8A). This work also confirmed that many driver mutations affecting TSG are two-hit inactivation events: on the 954 tumors in the cohort with driver mutations in *TP53*, 736 (77%) had both alleles mutated, 96% of which (707 out of 736) combined a somatic point mutation that affected one allele with somatic deletion of the other allele (Figure 8C).

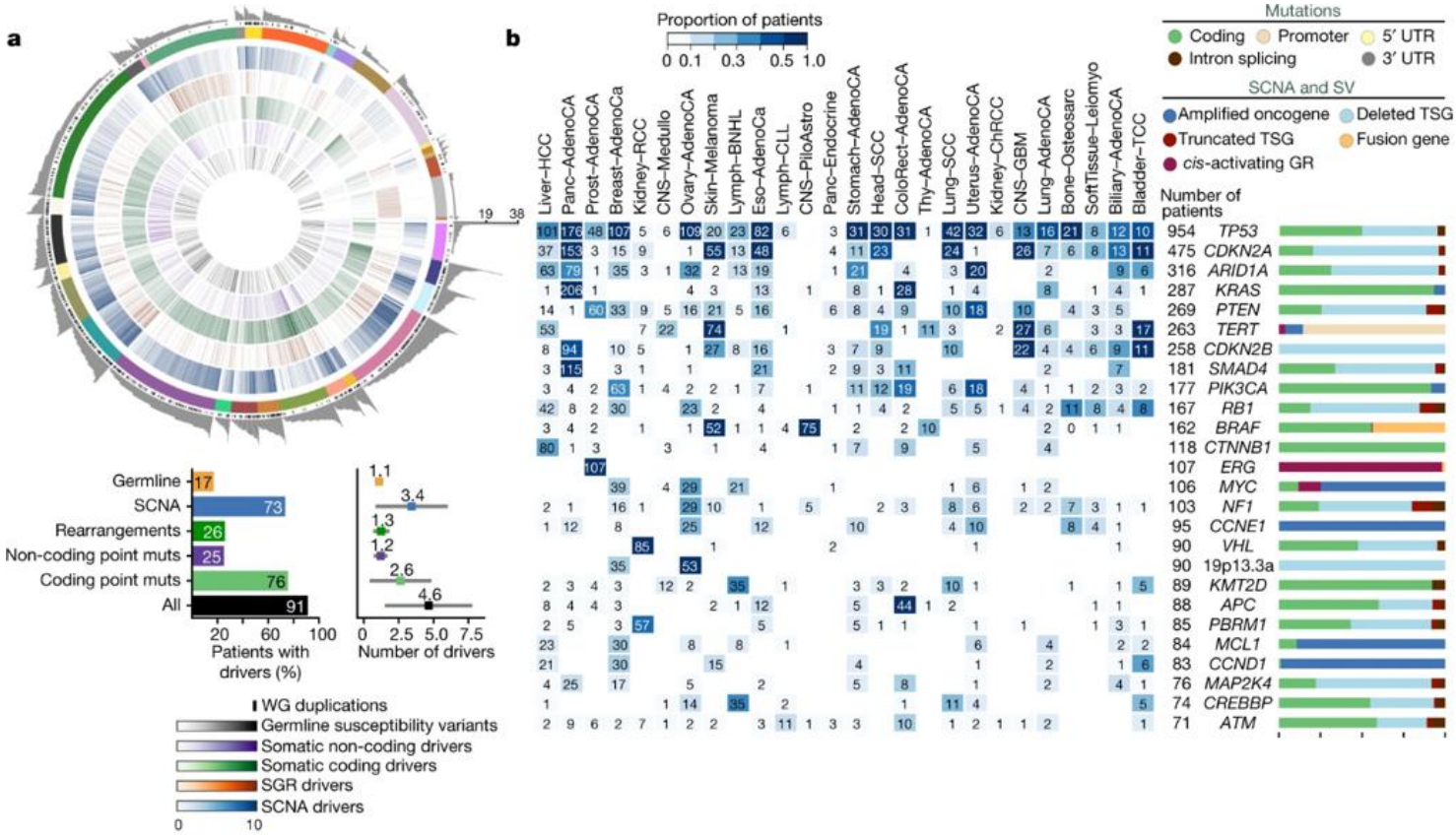


Figure 8: Landscape of PCAWG pan-cancer driver analysis (Campbell et al, 2020). A: The circos plot represents putative driver mutations identified in the PCAWG project. Each colored sector represents a tumor type. From the periphery to the center of the plot the concentric rings represent: (1) the total number of driver alterations; (2) the presence of whole-genome (WG) duplication; (3) the tumor type; (4) the number of driver CNAs (SCNA); (5) the number of driver genomic rearrangements (SGR); (6) driver coding point mutations; (7) driver non-coding point mutations; and (8) pathogenic germline variants. On the bottom, bar plot (left) represents the proportion of patients with different types of drivers. The dot plot (right) represents the mean number of each type of driver mutation across tumor cohort with at least one event (the square dot) and the standard deviation (grey whiskers), based on n = *2,583 patients. B, Driver genes targeted by different types of mutations in the cohort. Both germline and somatic variants are included. Left, the heatmap shows the recurrence of alterations across cancer types. The color indicates the proportion of mutated tumors and the number indicates the absolute count of mutated tumors. Right, the proportion of each type of alteration that affects each gene. C: Tumor-suppressor genes with biallelic inactivation in 10 or more patients. The values included under the gene labels represent the proportions of patients who have biallelic mutations in the gene out of all patients with a somatic mutation in that gene.*

PCAWG efforts led to a complete landscape of driver alteration associated to each cancer type, with more than 90% of PCAWG cases with at least one identified driver. However, in 181 tumors no driver alteration was identified. This can be due to either low quality or contaminated samples but also statistical power limitation associated with cohort size, leading to failures in the bioinformatic algorithms to identify new driver genes (Campbell *et al.*, 2020). Notably, structural variations in non-coding regions are more and more shown to be involved in tumorigenesis and there is evidences that this type of alteration is under-represented among driver alteration of cancer (Huddleston and Eichler, 2016; Audano *et al.*, 2019).

2. Functional consequences and molecular origin of SVs

2.1. Functional impact of structural variants

Structural Variations (SV) represent all the rearrangements of chromosome structure and include 4 fundamental types: deletions, duplications, inversions and translocations (Figure 9).

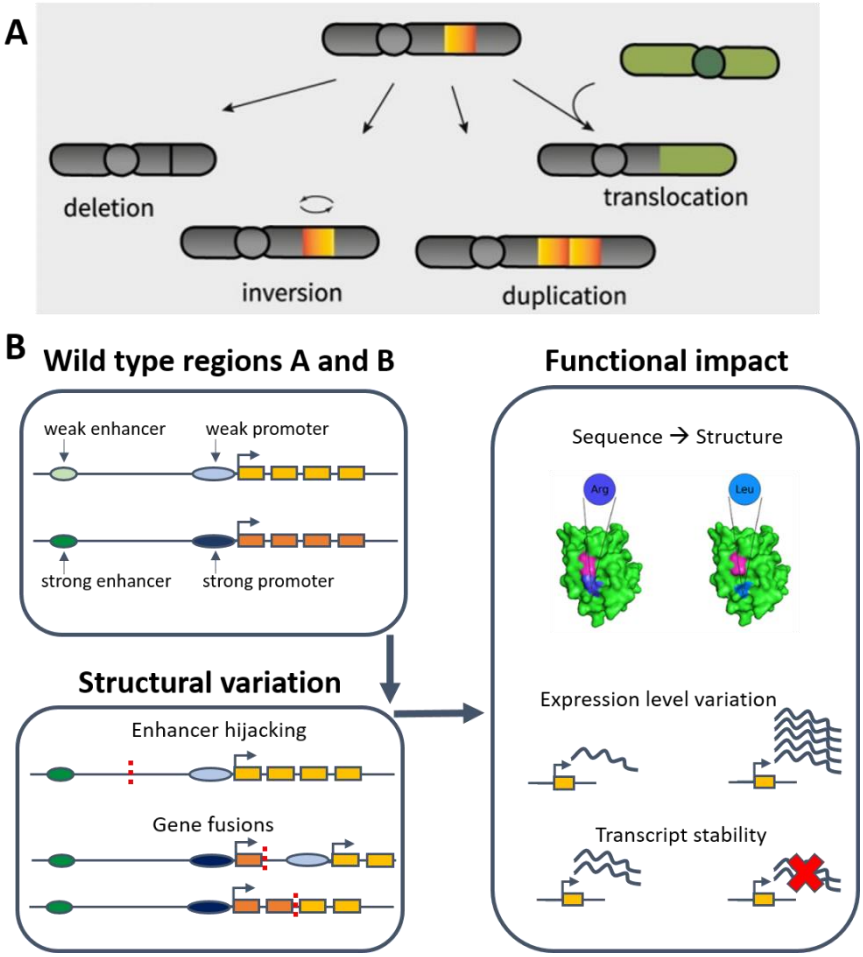


Figure 9: A: The 4 fundamental types of structural variation of the genome. Deletions, inversions, duplications and translocations are known to be recurrent molecular mechanisms at the origin of cancer development (Yi & Ju 2018). B: Functional consequences of abnormal junctions (red dotted line) induced by structural variations. Intergenic breakpoint can modulate gene expression by altering regulatory regions such as enhancer, promoters or post-transcriptional regulatory domains. Moreover, protein structure can be impaired following intragenic alteration (chimeric protein following gene fusion, truncated protein, post-translational regulatory domains alteration)

SVs can induce copy-number changes that alter the expression of target genes. Besides, they induce abnormal junctions between distant regions of the genome. Those junctions can have functional consequences by i) joining exons of two different genes, producing a chimeric transcript with a tumorigenic potential (Hutchinson *et al.*, 2013), ii) altering the gene structure and induce loss of negative regulatory regions (Mudduluru *et al.*, 2011), iii) placing the gene under the influence of a strong enhancer, called enhancer hijacking (Northcott *et al.*, 2014), and iv) placing the gene in a region of open chromatin and more accessible to transcription factors (Weischenfeldt *et al.*, 2017).

2.1.1. Copy-number alterations

CNAs include gains, amplifications (focal regions with a high number of extra copies), heterozygous and homozygous deletions of chromosome regions. Last decade, WES was extensively used to identify driver CNAs and revealed deletion hotspots involving many tumor suppressor genes (*CDKN2A*, *PTEN* or *RB1*…). In contrast, hotspots of genomic amplification contribute to the overexpression of oncogenes (*CCND1/FGF19*, *MYC*, *CCNE1*…). A recent pan-cancer study established that 70 genes are recurrently amplified and 70 genes are recurrently deleted across all their pan-cancer cohort including almost 5,000 tumors from 11 cancer-types (Zhang *et al.*, 2018). However, WES does not allow to define the mechanisms at the origin of CNAs, as the abnormal junction are usually located outside coding regions covered by this approach. Thus, WGS and RNAseq are necessary to identify other types of driver SVs.

2.1.2. Gene fusions

Gene fusions occur when a structural rearrangement brings together the 5' and 3' ends of two different genes. Philadelphia chromosome, discovered in 1960, induces the *BCR-ABL1* fusion in myeloid leukemia. This fusion leads to a constitutively active tyrosine kinase domain, inducing downstream activation of PI3K and MAPK pathways (Cilloni and Saglio, 2012; Hantschel, 2012; Sinclair, Latif and Holyoake, 2013). Since this early discovery, many oncogenic fusions have been identified. For example, a frequent translocation between chromosomes 11 and 22 creates the *EWSR1-FLI1* fusion, characteristic of Ewing sarcoma. This fusion combines the 5′ transcriptional activation domain of *EWSR1* with the 3′ ETS DNA-binding domain of *FLI1* to yield a potent oncogene through transcription factor activation (Delattre *et al.*, 1992, 1994). Some fusions are cancer specific, like the *PRKACA-DNAJB1* fusion specific of fibrolamellar carcinomas that is used to differentiate this specific cancer type from other liver tumors (Honeyman *et al.*,

2014; Dinh *et al.*, 2017). On the contrary some fusions are seen in a large variety of cancers, like the *FGFR3-TACC3* in-frame activating kinase fusion found in glioblastomas, lung adenocacrinomas and gliomas (Singh *et al.*, 2012; Collisson *et al.*, 2014; Lasorella, Sanson and Iavarone, 2017). Some studies used public RNAseq data from thousands of tumor samples to perform pan-cancer analyses of fusions and develop comprehensive databases such as ChimerDB (Lee *et al.*, 2017; Jang *et al.*, 2020). The TCGA consortium analyzed gene fusions in almost 10,000 tumors from 33 cancer types. They concluded that fusions frequently involve a kinase domain and drive tumorigenesis in 16.5% of cancer cases. Importantly, 1% of cancer cases are characterized by a single gene fusion as driver event fueling tumorigenesis (Gao *et al.*, 2018) (Figure 10). Gene fusions are thus interesting targets for precision medicine, exemplified by the success of inhibitors targeting the kinase activity of *RET*, *ROS1* and *ALK* fusions in lung cancer (Takeuchi *et al.*, 2012; Kohno *et al.*, 2015).

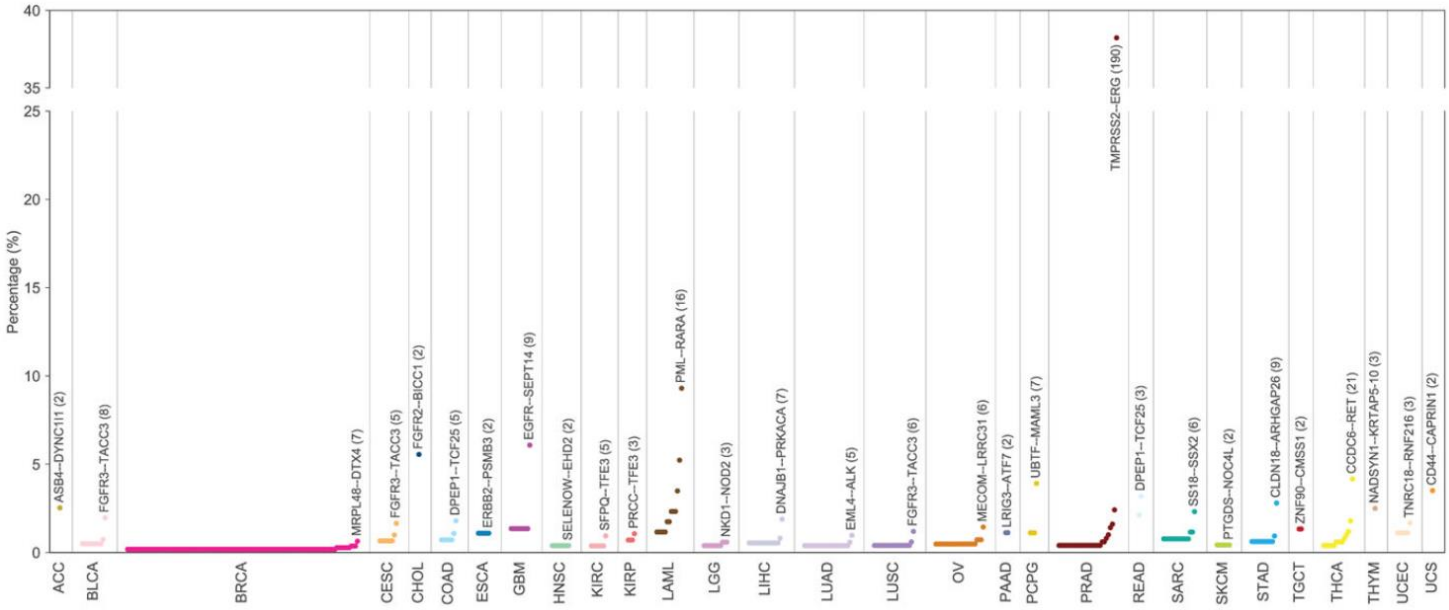


Figure 10: Frequency of gene fusions in 29 cancer types. The dot plot shows the frequency of recurrent fusions found in each cancer type. The most recurrent fusion in each cancer type is labeled. Cancer types without recurrent fusions are not shown. (Gao et al, 2018).

2.1.3. Enhancer Hijacking

Structural variations can also alter the regulatory regions and thus the transcription of cancer genes. Direct alteration of cis-regulatory regions was extensively studied. In hepatocellular carcinomas (HCC), *TERT* promoter mutation is required as an early event to trigger *TERT* expression and activate the telomere maintenance pathway (Nault *et al.*, 2013). Apart from *TERT* promoter, WGS studies revealed few non-coding driver mutations, but more frequent examples of SVs altering the regulatory regions of

driver genes (Fredriksson *et al.*, 2014; Weinhold *et al.*, 2014; Rheinbay *et al.*, 2020). A well-known mechanism named "*enhancer hijacking*" corresponds to the displacement of a strong enhancer upstream an oncogene. Enhancer hijacking was first described in medulloblastma (Northcott *et al.*, 2014). In this work, the authors identified a cluster of SVs leading to the juxtaposition of *GFI1B* and *GFI1* to local or distal DNA elements. Using ChIPseq, they confirmed the presence of enhancer chromatin marks in this region, with peaks of H3K27ac and H3K9ac immediately adjacent to the SV breakpoints, leading to the activation of *GFI1B* or *GFI1* (Figure 11).

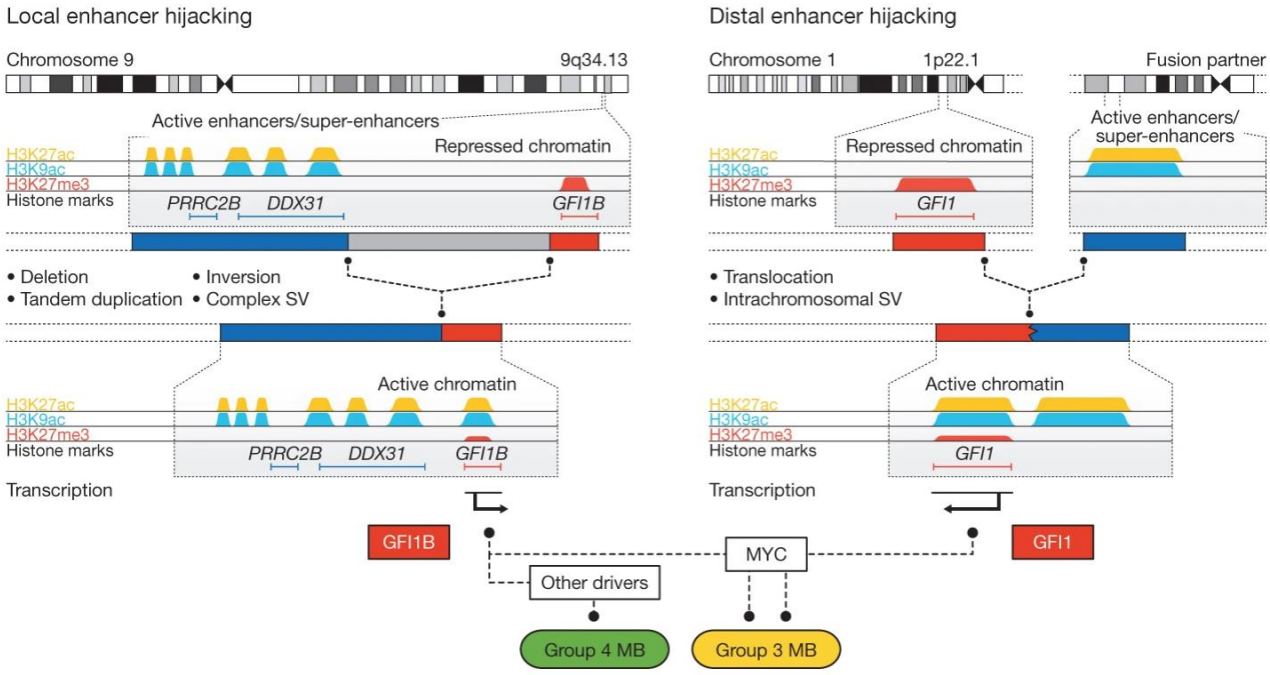


Figure 11: Mechanisms of SV leading to enhancer hijacking observed in GFI1/GFI1B-activated medulloblastomas. Activation of GFI1 and GFI1B are respectively induced by intra (left) and inter-chromosomal (right) rearrangements. Both oncogenes can cooperate with MYC to promote medulloblastoma pathogenesis. (Northcott et al., 2014)

A related mechanism, called promoter hijacking, was recently descried in the lab in a subgroup of hepatocellular adenomas. In these tumors, a focal deletion juxtaposes the promoter and exon 1 of the highly expressed *INHBE* gene upstream the *GLI1* oncogene. This promoter hijacking induces a massive over-expression of *GLI1* and the development of adenomas activating sonic hedgehog pathway (shHCA)(Nault *et al.*, 2017).

2.1.4. Local alteration of the 3D architecture

Eukaryotic genome has a highly organized three-dimensional (3D) structure, with an important role in gene regulation. Different levels of organization were described, using methods like HI-C (Lieberman-Aiden *et al.*, 2009) (Figure 12):

- **Nuclear center versus periphery**. This first level of organization involves two distinct nuclear compartments. Expressed genes tend to be located in euchromatin, in the nuclear center, while repressed genes are located in heterochromatic regions, at the periphery of the nucleus (Bickmore, 2013).
- **Chromosomes territories**. Chromosomes occupy distinct sub-nuclear territories, with transcriptionally active loci located at the surface of those structures (Cremer and Cremer, 2001, 2010) (Figure 12A).
- **Compartments A and B**. Chromosomes segregate into regions corresponding to long-range interactions, forming two distinct compartments names "A" and "B". The A compartments are characterized by active chromatin, enriched in highly expressed genes, whereas "B" compartment correspond to inactive chromatin, enriched in repressed genes (Lieberman-Aiden *et al.*, 2009). Thus A and B compartments are tens to hundreds of kilo-bases regions, characterized by intra-domain interactions but a lack of inter-domain interaction. These intra-domains contacts are commonly named Topologically Associated Domains (Nora *et al.*, 2012) (Figure 12A).
- **Topologically Associated Domains (TADs)**. In mammals, TADs are strongly conserved features bordered by protein complexes composed of the CCCTC-binding factor (CTCF) and the structural maintenance of chromosomes (SMC) cohesin complex (Dixon *et al.*, 2012). CTCF and cohesion protein complexes also define intra-TAD chromatin loops. These shorter length structures enable more subtle regulatory control of enhancer-promoter interactions (Matthews and Waxman, 2018).

Genomic alterations disrupting 3D chromatin conformation can trigger tumorigenesis by altering gene regulation. For example, a pan-cancer study revealed recurrent overexpression of *TERT*, *IRS4* and *IGF2* oncogenes in different cancers due to SVs disrupting TADs boundaries (Weischenfeldt *et al.*, 2017) (Figure 12B).

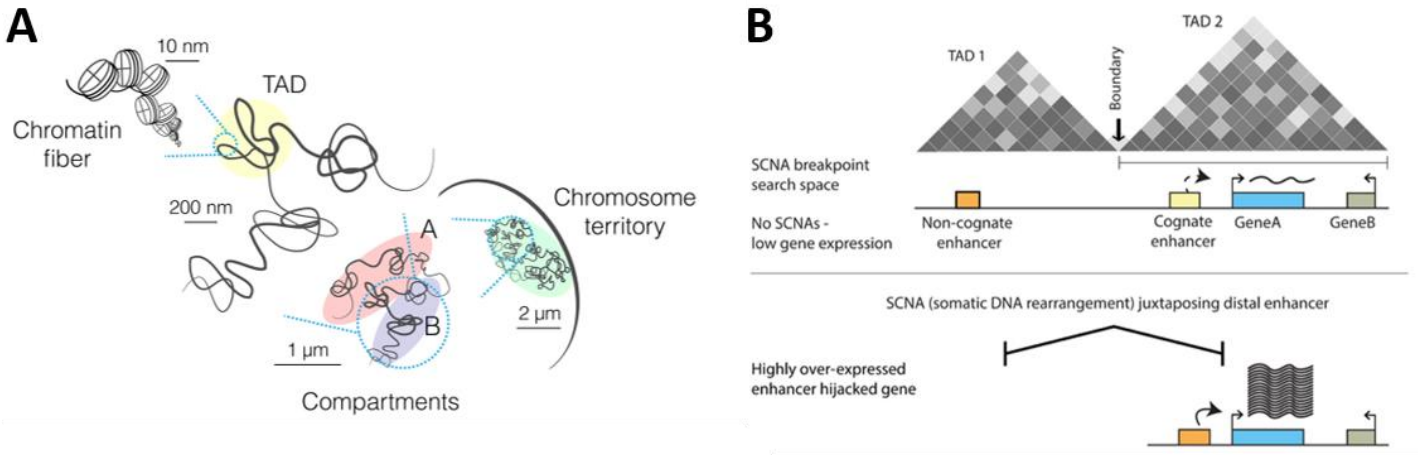


Figure 12: A: Representation of chromatin structure, from chromosome territories to Topologically Associated Domains (TADs) (Szabo et al., 2019). B: Model of enhancer hijacking following TAD boundaries disrupting. Heatmap shade of greys indicate contact probability between pair of genomic regions. In this scheme, the orange enhancer, associated to TAD1, will activate blue gene transcription following a TAD disrupting through somatic structural rearrangement. (Weischenfeldt et al. 2017)

2.1.5. Alteration of post-transcriptional regulatory regions

After transcription, untranslated regions (UTRs) and polyA tails control the stability of the transcripts. Alterations of those post-transcriptional regulatory regions can activate oncogenes or inactivate TSG. For example, recurrent *CSF1* fusions have been described in tenosynovial giant cell tumors (upstream *COL6A3* in 33% of cases), with a junction always downstream exon 5 of *CFS1.* This fusion leads to an overexpression of the transcript and the first interpretation was that *COL6A3* or other partner genes may bring active regulatory domains leading to increased transcription (Möller *et al.*, 2008). However, recent work showed that deletion of *CSF1* 3' UTR triggers tumorigenesis even in absence of gene fusion, suggesting that the loss of regulatory elements in 3'UTR explains the increased expression (Ho *et al.*, 2020).

2.1.6. Viral insertions

Viral DNA can integrate into the human genome and induce trigger tumorigenesis by placing a potent viral enhancer upstream an oncogene a mechanism called "insertional mutagenesis" (Bishop, 1987). Only 4 viruses are known to be able to induce insertional mutagenesis:

- **Human Papilloma Virus (HPV).** HPV is frequently integrated in the host cell genome and associated with the partial or complete loss of the E1 and E2 viral

genes, which regulate the activity of viral oncoproteins E6 and E7. However, hotspots of HPV integrations in the human genome have been described, including *MYC*, *POU5F1B*, *FHIT, KLF12, KLF5, LRP1B* and *LEPREL1* (Gudlevičiene *et al.*, 2015; Hu *et al.*, 2015). As those genes are involved in modulating cell proliferation, development and carcinogenesis both oncoproteins and insertional mutagenesis may have a tumorigenic effect

- **Merkel Cell Polyomavirus (MCPyV).** MCPyV DNA is detected in around 80% of biopsies of Merkel Cell Carcinoma (MCC), a rare and aggressive skin cancer. Polyomaviruses depend on the host cell to replicate and use two DNA-binding proteins to trigger replication: the large T antigen physically interacts with and inhibits tumor suppressors pRB and p53, allowing the cell to progress through the G_1-to-S checkpoint, and small T antigen induces transcription of E2F, necessary for turning on essential S phase genes. In MCCs, the MCPyV genome random integrations contain the large T antigen gene, but systematically mutated in helicase and DNA binding domains, inducing an inhibition of viral replication and preventing lytic viral replication that could be lethal for cancer cells. However, the produced chimeric protein is still able to interfere with pRB and drive cell proliferation, leading to a pro-oncogenic state that may lead to carcinomas (Erstad and Cusack, 2014).
- **Hepatitis B Virus (HBV).** This virus has a tropism for liver and is a major cause of hepatocellular carcinomas (HCC). Most frequently the virus acts through indirect mechanism by the role of its oncoproteins or by inducing liver cirrhosis, an inflammatory environment prone to HCC development. Conversely, HBV can directly activate driver genes following HBV insertional mutagenesis. *TERT*, *CCNE1*, *KMT2B*, *SENP5* and *ROCK1* are well-known oncogenes whose expression is recurrently activated by HBV in liver cancer (Sung *et al.*, 2012).
- **Adeno-associated Virus Type 2(AAV2).** AAV2 was identified in the lab as a new virus able to generate insertional mutagenesis. In 2015, recurrent AAV2 integrations were described in HCC affecting the well-known oncogenes *TERT, CCNA2, CCNE1* and *TNFSF10* (Nault *et al.*, 2015). In 2019, another oncogene, *GLI1*, was identified to be activated through AAV2 insertions in HCC developed from benign hepatocellular adenomas (HCA) (La Bella *et al.*, 2020). The insertional mutagenesis of both HBV and AAV2 in *TERT* promoter suggests that viral

insertions are early events in the tumorigenesis process, as *TERT* promoter mutations were recurrently found in pre-neoplastic lesions (Nault *et al.*, 2013) (Figure 2).

2.2. Mechanisms at the origin of structural variants

Whole genome sequencing of thousands of tumors, notably by the ICGC-PCAWG consortium (Li *et al.*, 2020), revealed considerable variability in the number and types of structural variations both across and within cancer types (Figure 13). Some tumors display large numbers of SVs belonging to a same category, defining deletor or duplicator phenotypes that can be further subdivided by the size of rearrangements (Degasperi *et al.*, 2020). Other tumors are dominated by clustered or non-clustered inter-chromosomal translocations. These distinct phenotypes suggest the existence of various mechanisms of genomic instability operative with different strength across tumors.

In addition to the 4 simple SV categories (deletions, duplications, inversions and translocations), complex rearrangements involving many abnormal junctions and copy-number changes acquired through a single event have been identified:

- **Breakage-fusion-bridge (BFB) cycle** is a complex pattern of rearrangement that was first described almost a century ago (McClintock, 1939) and later shown to be associated with the amplification of *RUNX1* driving acute lymphoblastic leukemia, (Papaemmanuil *et al.*, 2014). In a situation of telomere crisis, at the anaphase step of mitosis, telomeres of sister chromatids can fuse together before being pulled apart in opposite directions. This will lead to a dicentric chromosome and the formation of an anaphase bridge, visible in microscopy (Figure 14A bottom). Eventually the two fused chromatids will break, but not necessarily at the fusion point, resulting in a daughter cell with an arm loss when the other harbors a fold-back inversion and gain of genomic material (Murnane, 2012). Since resulting chimeric chromosomes still lack telomeres, this process can repeat and produce recognizable patterns fold-back inversion junctions and stair-like copy-number increase at the extremity of the altered chromosome arm (Figure 14A top).

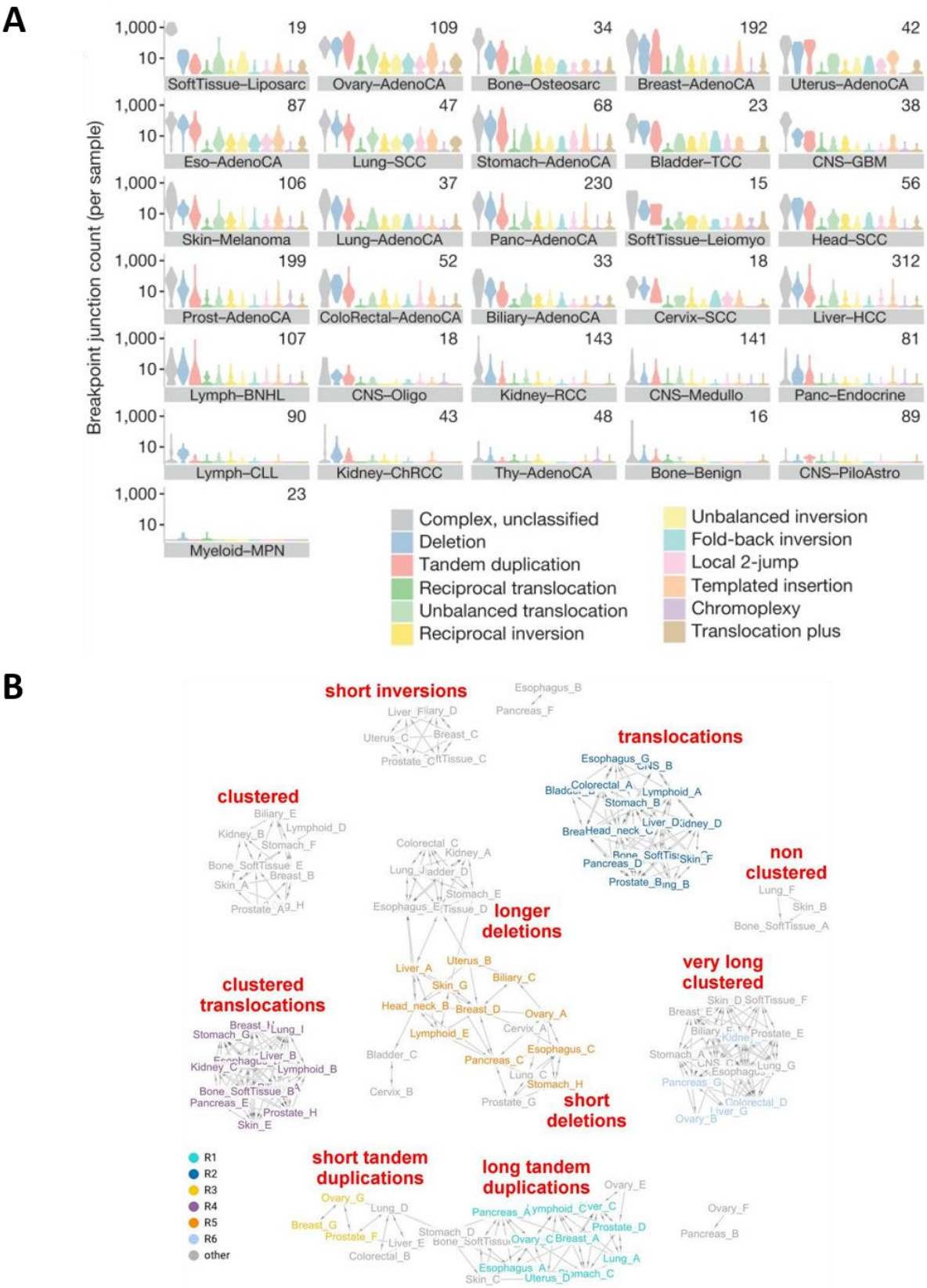


Figure 13: A: Violin plots of types of SVs across different cancer types. In each panel, the number of patients is indicated at the top right. AdenoCA, adenocarcinoma; BNHL, B-cell non-Hodgkin lymphoma; ChRCC, chromophobe renal cell carcinoma; CLL, chronic lymphocytic leukaemia; CNS, central nervous system; GBM, glioblastoma; HCC, hepatocellular carcinoma; leiomyo, leiomyosarcoma; medullo, medulloblastoma; MPN, myeloproliferative neoplasm; eso, oesophageal; oligo, oligodendrocytic; panc, pancreatic; piloastro, pilocytic astrocytoma; prost, prostate; RCC, renal cell carcinoma; sarc, sarcoma; SCC, squamous cell carcinoma; TCC, transitional cell carcinoma; thy, thyroid. (Li et al., 2020) B: Network connecting highly similar rearrangement signatures across different organs. Each circle represents a signature and particular phenotypes of deletions, duplications, etc. are annotated (Degasperi et al. 2020).

- **Chromothripsis** (from the greek *thripsis* which means "shattering into pieces") is a complex rearrangement first described in 2011 (Stephens *et al.*, 2011). Chromothripsis induces hundreds of DNA breaks simultaneously on one or a few chromosomes. The chromosome fragments are then randomly stitched together or lost, presumably by the non-homologous end-joining (NHEJ) repair mechanism (Stephens *et al.*, 2011). Chromotripsis results in chimeric chromosomes with numerous copy-number changes (oscillations between normal and deleted states) and abnormal junctions (Figure 14B top), both of which can induce gene expression changes. For example, chromothripsis has been shown to induce recurrent loss of *SMAD4*, APC, *PTEN* or *CDKN2A.* On the contrary, abnormal junctions can activate oncogenes like *CCND1*, *CDK4* or *MDM2* (Cortes-Ciriano *et al.*, 2014). The most frequent chromothripsis-associated driver gene is *TP53*, probably because inactivation of genes maintaining the genome stability favors chromothripsis occurrence and selection. Two main mechanisms linked to mis-segregated chromosomes have been proposed to be at the origin of chromothripsis (Ly and Cleveland, 2017). The first one is a telomere crisis leading to chromosome or sister chromatids end-to-end fusions, followed by formation of chromatin bridges (Garsed *et al.*, 2014). The second proposed mechanism is a micronuclei formation during mitosis (Kato and Sandberg, 1968; Johnson and Rao, 1970). Molecular processes in micronuclei are known to be error prone, and it was proposed that isolated genomic materials can be massively broken into pieces and reassembled.

- **Chromoplexy** (from the greek *pleko*, meaning to weave) is another complex rearrangement pattern. It was first described in prostate cancers harboring *ETS* gene fusions, then in other solid tumors like non-small cell lung, head and neck and melanomas cancers (Baca *et al.*, 2013). In contrast to chromothripsis, this phenomenon is mostly characterized by inter-chromosomal translocations, involving up to 8 chromosomes, whereas chromothripsis is restricted to 1 or 2 chromosomes. In addition, the altered regions are focal on each chromosome, whereas chromothripsis alterations can be chromosome wide. Finally, Baca and collaborators defined chromoplexy as a closed chain of rearrangements, occasionally combined with focal "deletion bridges" in the vicinity of abnormal

junction. Except those deletions bridges, there is no loss of genomic regions characteristic of the 2 levels of copy number seen in chromothripsis. Recent pan-cancer studies (Agrawal *et al.*, 2014; Cortés-Ciriano *et al.*, 2020) described recurrent oncogenic events induced by chromoplexy, such as *RET*, *BRAF* and *NTRK3* fusions, or *IGF2BP3* enhancer hijacking. The genome-wide distribution of DSBs in chromoplexy is enriched in actively transcribed and open chromatin regions (Marnef, Cohen and Legube, 2017), suggesting that these catastrophic events may occur in nuclear transcription hubs where many co-regulated genomic regions from different chromosomes are spatially aggregated (Baca *et al.*, 2013) (Figure 14C). Thus, the proposed mechanism is a multiple breakage of DNA in a restricted transcription hub, followed by the mis-reparation of the different blunt-end (Figure 14D).

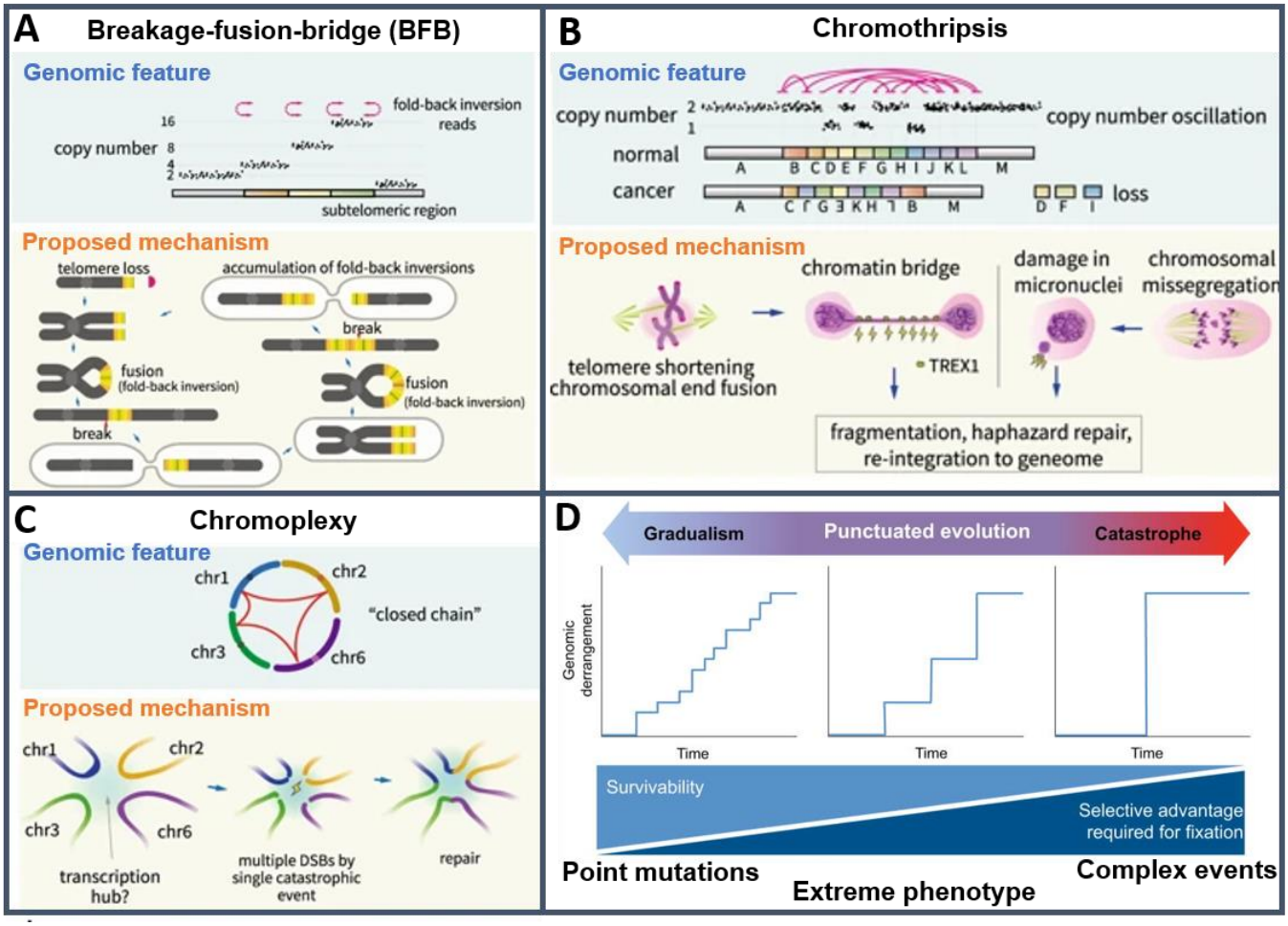


Figure 14: A-C: Complex genomic alteration recurrently seen in cancer, and their associated proposed mechanisms. (modified from Yi et al., 2018). D: Proposed model of punctuated evolution of cancer, where tumorigenesis can progressively be triggered by i) gradual accumulation of simple mutations, giving progressively cell survivability, ii) Rapid accumulation of alteration induced by a particular biological defect of environmental exposure, or iii) a single catastrophic event, giving a significant selective advantage to become a clonal feature. (modified from Baca et al., 2013)

In addition to these 3 well-documented events, Li *et al.* recently described new categories of complex SVs (Li *et al.*, 2020) . Thus, we are only beginning to understand the diversity of structural rearrangements in cancers.

Chromothripsis and chromoplexy are complex events occurring at a single time of cell life (Stephens *et al.*, 2011; Baca *et al.*, 2013). Thus, there is a heterogeneous dynamic in cancer development that can be discretized in three states (Figure 14D).

- First cancer can slowly develop, by progressively accumulating mutations such as stochastic "clock-like" mutations described by L. Alexandrov and collaborators, induced by stochastic error and repair mechanisms (Alexandrov *et al.*, 2015).
- Secondly, rapid accumulation of alteration in a short period can be the result of particular biological defect, as the extreme phenotypes described by Degasperi and collaborators, triggered by homologous recombination deficiency (Degasperi *et al.*, 2020). Accelerated evolution of cancer can also be induced by specific exposures such as Aflatoxin B1 mycotoxin produced by *Aspergillus flavus* and *Aspergillus parasiticus* which grow in soil or decaying vegetation. Exposure to this mycotoxin was shown to be associated with specific pattern of mutations in hepatocellular carcinoma inducing an increase of mutational burden (Letouzé *et al.*, 2017).
- Finally, singe catastrophic events are probably lethal for the cell and may tend to require co-occurring oncogenic lesions to become fixed in a tumor by giving a strong selective advantage (Baca *et al.*, 2013) (Figure 14D).

In summary, different patterns of SVs were described in cancer, reflecting the diverse mechanisms at their origin, more or less active in different tumors. SVs may be due to i) stochastic biological errors such as mis-segregation of chromosomes, ii) abnormal cell state such as telomere crisis leading to inherent errors of NHEJ, triggered by an abnormal biological context, or iii) a specific gene defect leading to inhibition of a repair mechanism, such as *BRCA1*-mutated breast cancer harboring a HR defect. However, the molecular cause of most SV phenotypes remains to be unraveled. Defining signatures of SVs related to specific processes is important for basic research, but also to develop efficient personalized therapies. For example, breast cancers patients harboring a *BRCA1* or *BRCA2*

mutations and the associated signature of duplications are good responders to PARP inhibitors (Sikov *et al.*, 2015).

2.3. Signatures of processes generating somatic alterations

Cancer cells accumulate somatic alterations through various intrinsic and extrinsic processes. Those processes leave characteristic imprints on the genome, termed mutational signatures (Alexandrov, Nik-Zainal, Wedge, Campbell, *et al.*, 2013). Mutational signatures are more or less active in each tumor depending on intrinsic biological processes (e.g. DNA repair) or environmental factors, and the contribution of each process to the tumor mutation burden is called its *exposure*. Mutational processes can also evolve between the early and late steps of tumorigenesis. The combination of processes a cancer cell has been exposed to results in its final mutational portrait (Figure 15). For example, one of the first mutational patterns related to a cancer risk factor was the abundance of C>T / G>A substitutions and CC>TT / GG>AA dinucleotide substitutions associated to UV (Ultra-Violet) *exposure*. This observation soon led to the first description of mutational processes signatures (Howard and Tessman, 1964; Witkin, 1969).

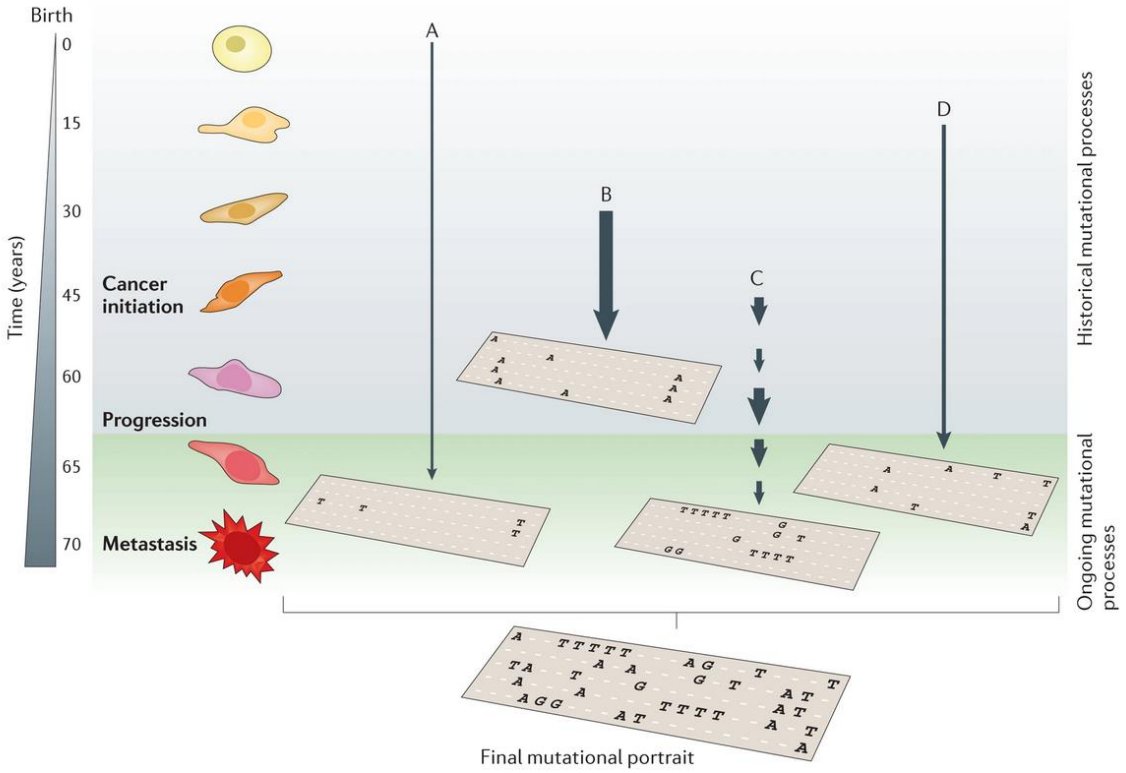


Figure 15: Clock-like processes, DNA repair defects and environmental exposures can leave characteristic imprints, termed signatures, in cancer genome. Arrows indicate both duration and intensity of mutational process exposure. In this model, the final mutational portrait is the sum of the 4 mutational processed involved (A, B, C and D) and their identification is useful to understand tumor biology and develop targeted therapies. (Helleday et al., 2014)

2.3.1. NMF implementation for mutational signatures

With the advent of NGS, and in particular whole genome sequencing, the characterization of mutational signatures has greatly improved in the last decade. Ludmil Alexandrov, at the Wellcome Trust Sanger Institute, developed the first mathematical method to extract signatures of single nucleotide variations (SNVs) (Alexandrov, Nik-Zainal, Wedge, Campbell, *et al.*, 2013). First, SNVs are classified into 96 categories based on the type of substitution and the 3' and 5' bases around the mutation. Using the pyrimidine (C or A) of base pair as reference, 6 substitutions can be defined: 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. The 3' and 5' bases around the mutation define 16 different trinucleotide contexts, resulting in 96 (6x16) SNV categories that can be used to describe the mutation catalogue of a tumor (Figure 16, left).

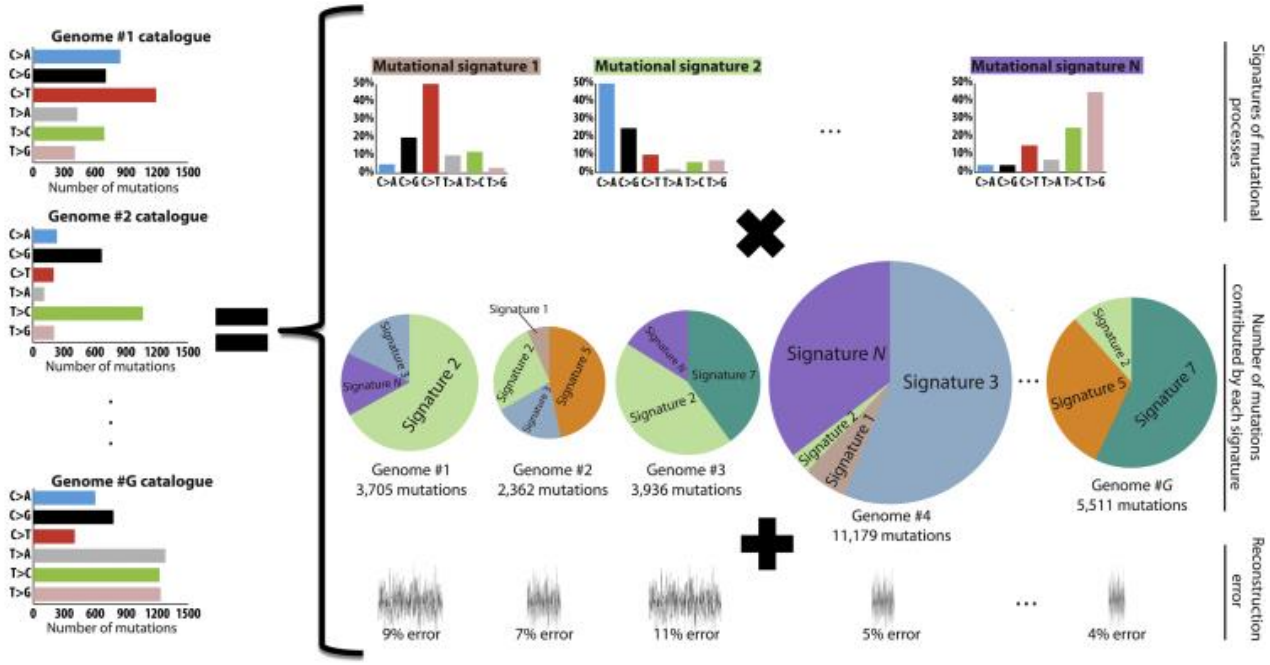


Figure 16 : Scheme illustrating mutational processes operative in a set of G *cancer genomes catalogues of mutations. This data can be used to deconvolute the signatures of* N *mutational processes as well as the number of mutations caused by each of the processes in each of the genomes. As the inverse process corresponding to reconstruction of the original from signatures and respective exposure lead to slight variations with the original catalogue. This is the result of genome-specific reconstruction error. (Alexandrov et al., 2013a).*

Alexandrov *et al.* also introduced a statistical framework to extract, from the mutation catalogues of a series of tumors, the number of operative mutational processes, their signatures and their *exposure* in each tumor (Figure 16, right). This is a classic case of "cocktail party" problem (Alexandrov, Nik-Zainal, Wedge, Campbell, *et al.*, 2013). Imagine a cocktail party room with tens of guest groups speaking about various subjects, resulting in an incomprehensible mixture of sounds. This room is equipped with several record devices, located at different spots in the room, recording the mixture of conversations,

whose respective intensities depend on the distance of guest groups from each microphone. By integrating the different records, their location in the room and the differential intensity of conversations, it can be possible to decompose each distinct conversation from the mixture of sounds and locate them in the room. In this analogy, distinct conversations are mutational signatures, individual records are the catalogue of mutational signature in one tumor and the intensity of a recorded conversation is the *exposure*. By integrating the series of mutations catalogues, it is possible to decipher the set of mutational signatures at the origin of alterations and their respective *exposure* in each cancer. From a mathematical point of view, the mutational catalogue of a cancer genome is the linear combination of a set of signatures whose exposures can be very heterogeneous among the different cancers of the cohort. Moreover, in the case of NGS data, catalogues of alterations are expected to contain noise due to stochastic sequencing/analysis errors. We can express the mutational catalogues over the tumor series as a matrix M, with M_{ij} enumerating over mutation categories i and cancer samples j. The mutation matrix M can be decomposed into 2 matrices of smaller size (Alexandrov, Nik-Zainal, Wedge, Campbell, *et al.*, 2013):

$$M_{ij} = \sum_{k=1}^{N} S_{ik} E_{kj} + \varepsilon_{ij},$$

where N represents the number of active signatures in the cohort, and ε the minimized residual matrix. The columns of matrix S describe the composition of the signature in terms of mutation categories, with S_{ik} the frequency of mutation category i in signature k. The second matrix E is the exposure matrix of each signature k in row, associated to each sample j in column.

In a more generalized definition, the mutational catalogue is approximately the sum of mutational signatures multiplied by their respective exposures:

$$M \approx SE \qquad \text{(Eq.1)}$$

In order to deconvolute signal and extract S and E matrices, the mutational signature extraction computational workflow is based on the Non-Negative Matrix Factorization (NMF). This core step corresponds to the Brunet et al. implementation of the multiplicative update algorithm, shown to extract biologically meaningful components from complex biological data (Lee and Seung, 1999; Brunet *et al.*, 2004). Formally, the NMF extracts N signatures, approximately solve the Equation 1, by finding non-negative matrices S and E that minimize the Frobenius norm:

$$\min_{S\geq 0, E\geq 0} ||M - SE||_F^2$$

This framework paved the way to the association of etiologies and specific mutational patterns at the origin of tumorigenesis. Ludmil Alexandrov et al. applied this innovative framework to analyze 4,938,362 mutations from 7,042 cancers, subdivided in 30 cancer types and extracted more than 20 distinct mutational signatures (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013). More recently, pan-cancer analysis of 4,645 cancer genomes and 19,184 cancer exomes by the PCAWG group revealed an extended set of 49 mutational signatures (Alexandrov et al., 2020). These signatures can be consulted in the COSMIC c.3 database [https://cancer.sanger.ac.uk/cosmic/signatures]. Single Base Substitution (SBS) signatures identified so far include (Figure 17):

- Clock-like signatures SBS1 (deamination of methylated cytosines) and SBS5 (unexplained) that accumulate with age in all cell types.
- Signatures related to specific environmental exposures like SBS4 (tobacco), SBS7 (UV light) or SBS24 (aflatoxin B1 exposure)
- Signatures of DNA repair defects like SBS3 (BRCA1/2 mutations inducing homologous recombination deficiency) or SBS6 (mismatch repair deficiency).
- Finally, some signatures are induced by genotoxic treatments. For example, SBS11 and SBS31/35 are respectively associated with temozolomide and cisplatin treatment.

The repertoire of mutational signature (Figure 17) is heterogeneous in terms of etiologies, but also in term of organs, some signatures being ubiquitous (SBS1), or on the contrary restricted to a few cancer types (SBS24). Finally, the mutational burden induce by each signature is extremely variable.

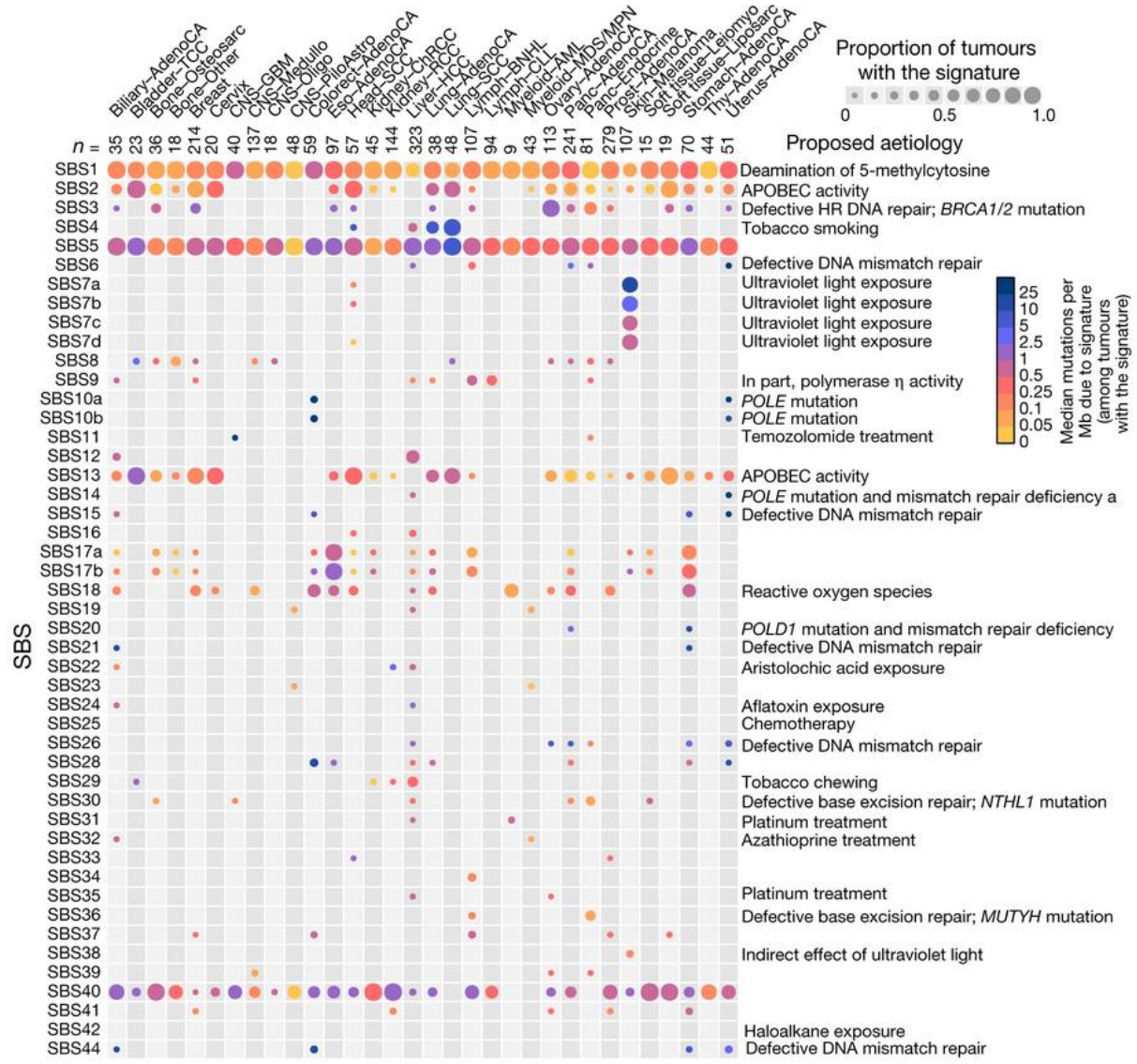


Figure 17: Number of mutations associated to each SBS mutational signature in PCAWG tumors.. Dot size represents the proportion of samples of each tumor type that harbor the mutational signature. The color of each dot represents the median mutation burden of the signature in exposed samples. There is a large heterogeneity in signature exposure across the different tissue types. Moreover, the mutational burden induced by signature is also heterogeneous with some signature inducing large number of mutations compared to others. (Alexandrov et al., 2020).

2.3.2. Extension of signatures to structural rearrangements

The signature analysis framework, based on NMF algorithm, was first developed for point mutations but it can be extended to any type of alteration. For example, the recent work by the PCAWG group identified, in addition to the 49 single-base-substitution signatures, 11 doublet-base-substitution and 17 small insertions and deletions (indel) signatures (Alexandrov *et al.*, 2020). The definition of categories needs to be adapted to each type of alteration, but the core step of signature deconvolution remains unchanged. This approach has also been applied to structural variations (SVs) to identify tumor phenotypes characterized by an enrichment of specific types of rearrangements. Nik-Zainal and collaborators described the first SV signature analysis framework to unravel

rearrangement signatures in 560 breast cancer genomes. In this framework, SVs were divided in 32 categories based on 3 criteria:

- **The type of alteration**, within the 4 fundamental types of SVs: inversions, duplications, deletions and translocations.
- **The size of the event** (for inversions, deletions and duplications only), i.e. the distance between the two breakpoints: 1-10kb; 10kb-100kb; 100kb-1Mb; 1Mb-10Mb and more than 10Mb.
- **Rearrangement clusters**. Clustered rearrangements and non-clustered rearrangements are considered separately.

This classification allowed the authors to extract 6 distinct rearrangement signatures operative in breast cancers (Nik-Zainal *et al.*, 2016) (Figure 18) including a signature of focal tandem duplications (<10 kb) related to *BRCA1* inactivation, a signature of large tandem duplications (>100 kb), and a signature of focal deletions associated with *BRCA2* deficiency.

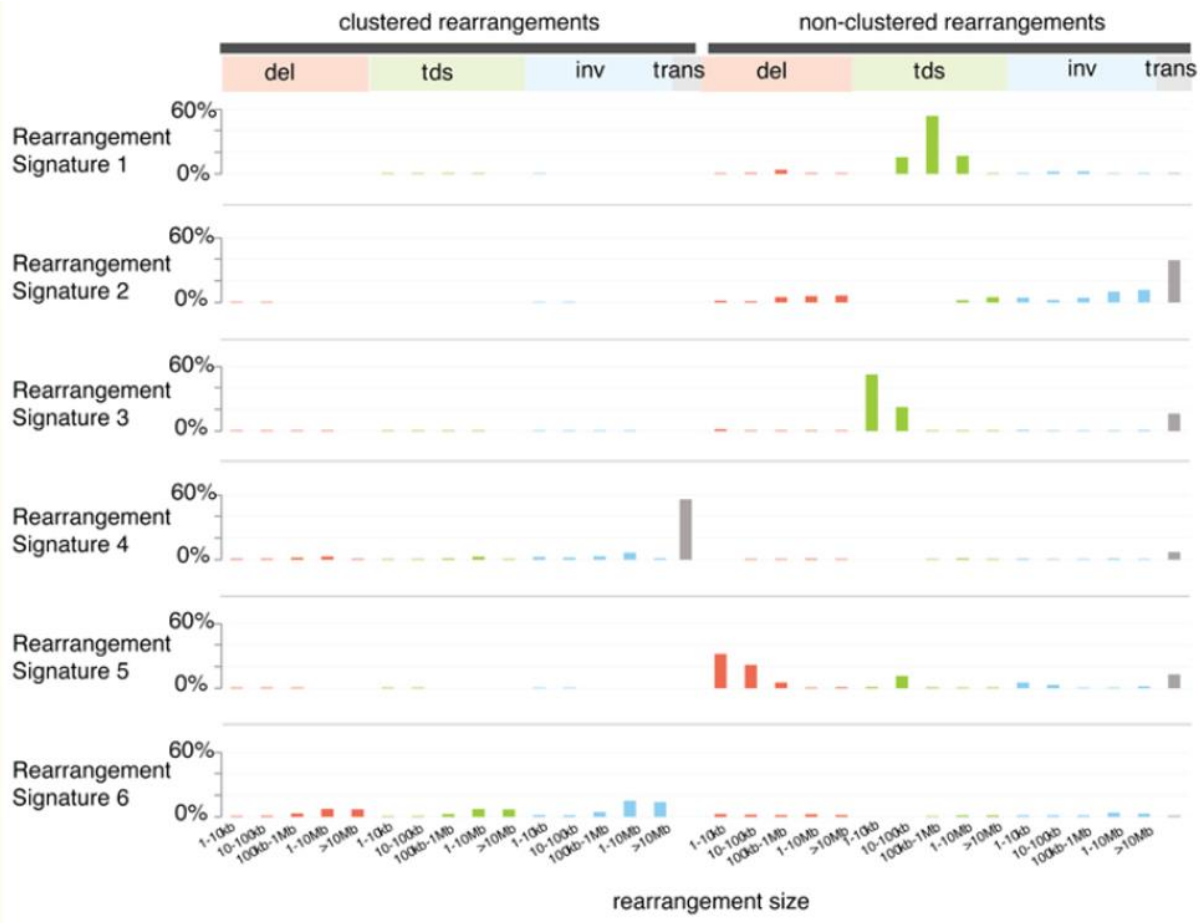


Figure 18: Six rearrangement signatures extracted using Non-Negative Matrix Factorization. Probability of rearrangement element on y-axis. Rearrangement size on x-axis. Del= deletion, tds = tandem duplication, inv = inversion, trans = translocation. (Nik-Zainal et al., 2016a)

3. Liver cancer

Liver cancer is the second leading cause of cancer-related deaths worldwide. With an incidence of approximately 850,000 new cases and 810,000 deaths per year (overall ratio of incidence to mortality of 0.95) this cancer has a very poor prognosis (Llovet *et al.*, 2016; Bray *et al.*, 2018). Liver neoplasms include benign and malignant lesions. Hepatocellular adenoma (HCA) is the most studied benign tumor because of its clinical management associated with tumor bleeding and malignant transformation into hepatocellular carcinoma (HCC). HCC is a very aggressive malignant tumor representing approximately 90% of all cases of primary liver cancer. Both HCA and HCC are developed from hepatocytes, the main liver cell type managing metabolic and detoxification processes.

3.1. Hepatocellular adenomas

3.1.1. Epidemiology and risk factors for HCA

HCA are rare benign tumors, with an overall prevalence around one case per million. Various associated risk factors have been identified:

- **Hormonal exposure.** Oral contraception (OC) administration and in general androgen and estrogens are the major risk factor for HCA development. Indeed, this cancer mostly arises in young women, with a 30-fold increased incidence associated to the use of oral contraceptives (OC), and a significantly increased risk after 5 years of treatment (Rooks *et al.*, 1977). This results in an higher incidence of HCA in western countries where the use of oral contraception is more common compared to Asia or Africa (Matsumoto *et al.*, 2011). Moreover, androgen administration for therapeutic use (ie: Fanconi anemia) is associated with HCA development (Touraine *et al.*, 1993). Similarly, anabolic androgenic steroids abuse for bodybuilding purpose is also a risk factor of HCA (Sánchez-Osorio *et al.*, 2008).
- **Alcohol and obesity.** Alcohol consumption and obesity were shown to be risk factors for HCA development (Bedossa *et al.*, 2007; Guichard *et al.*, 2012). This results presumably from the direct toxic role of alcohol or the production of cytokines during both long-term alcohol consumption and chronic liver inflammation associated with obesity. Interestingly those two risk factor are restricted to an inflammatory subgroup of HCA (Nault *et al.*, 2017).
- **Genetic background**. Finally, several rare genetic syndromes are strongly associated with HCA occurrence. MODY (maturity-onset diabetes of the young)

type 3 is a type of non-insulin dependent diabetes mellitus with autosomal dominant inheritance, occurring usually before the age of 25. This pathology is associated with a germline monoallelic *HNF1A* mutation leading to the development of liver adenomatosis, characterized by more than ten hepatic adenomas (Bacq *et al.*, 2003; Reznik *et al.*, 2004). In MODY3 patients, HCA development follows a "2-hit" alteration model of TSG where the germline alteration is followed by a somatic inactivation of the second wild type allele of *HNF1A* (Nault *et al.*, 2017). Another rare genetic disease associated with HCA development is glycogenosis type 1a. This is a rare recessive metabolic syndrome occurring in very young patients and characterized by a germline mutation inactivating *G6PC* (glucose-6-phosphatase). G6PC inactivation leads to the accumulation of glycogen and fat in the liver and to a 25 to 75% risk of HCA development depending on the studies (Labrune *et al.*, 1997; Froissart *et al.*, 2011). Finally, McCune-Albright syndrome, characterized by "café au lait" skin macula and fibrous dysplasia of bones, is also strongly associates with HCA development. This syndrome is caused by a somatic mosaic mutation of *GNAS* gene, located on an imprinted locus, that leads to aberrant activation of adelylate cyclase (Weinstein and Shenker, 1993). Interestingly, *GNAS* gene is recurrently mutated in the inflammatory subgroup of HCAs (Nault *et al.*, 2017).

Risk factors associated to HCA are various, resulting in multiple subgroups of HCA, characterized by specific molecular alterations and clinical presentation, such as the inflammatory adenomas above-mentioned.

3.1.2. Molecular diversity of HCA

Before NGS studies, targeted sequencing of candidate loci had revealed 4 HCA driver genes (*HNF1A*, *CTNNB1*, *IL6ST* and *GNAS*) but a large part of tumors remained without a molecular driver alteration (Calderaro *et al.*, 2013). Next generation sequencing allowed non supervised multi-omics analysis that helped refine the heterogeneous landscape of HCA sub-groups. The current classification, based on molecular and transcriptomic features is constituted of 6 HCA subgroups, associated with specific risks factors, clinical behavior and tumor histology (Nault *et al.*, 2017) (Figure19):

- Inflammatory HCA (**IHCA**) is the most frequent HCA subgroup, accounting for 34 to 44% of all HCAs. IHCA is characterized by constitutive activation of the

IL6/JAK/STAT signaling pathway due to activating mutations of various members of the pathway. IHCA driver genes include *IL6ST* encoding gp130 receptor (77%), *FRK* encoding a src kinase (9%), *STAT3* encoding a transcription factor of JAK/STAT pathway (4%), *GNAS* encoding stimulatory G-protein alpha subunit (3%) and *JAK1* encoding a tyrosine kinase protein of the JAK/STAT pathway (1%). 6% of IHCAs remain without identified molecular alteration (Nault *et al.*, 2017). Those different alterations lead to the translocation of STAT3 in the nucleus, triggering the uncontrolled activation of the inflammatory pathway, characterized by an over-expression of acute phase inflammatory proteins CRP (C-Reactive Protein) and SAA (Serum Amyloid A).

- HCAs harboring a *CTNNB1* exon 3 or exon 7,8 represent respectively 13% and 7% of all HCAs. **b^{ex3}HCA** harbor mutations or deletions in *CTNNB1* exon 3 that inhibit β-catenin protein phosphorylation by APC/GSK3B/AXIN complex, preventing proteasome degradation. This unregulated β-catenin will be translocated to the nucleus, inducing the strong activation of WNT/β-catenin target genes, such as *GLUL*, *ZNRF3* or *LGR5* (Rebouissou *et al.*, 2016). The smaller **b^{ex7,8}HCA** subgroup is defined by *CTNNB1* mutations in two hotspots in exon 7 or 8, leading to a weaker activation of the WNT/β-catenin pathway (Pilati *et al.*, 2014). Interestingly, half of *CTNNB1*-mutated HCAs (either b^{ex3}HCA and b^{ex7,8}HCA) also display an inflammatory phenotype through CRP and SAA protein over-expression following *JAK/STAT* pathway activation. This leads to two mixed subtypes of HCA: **b^{ex3}IHCA** and **b^{ex7,8}IHCA**. (Nault *et al.*, 2017) (Figure 19).
- *HNF1A* inactivated HCAs (**HHCA**) are the second most frequent subtype of HCA. This sub-group is defined by a bi-allelic inactivation of *HNF1A*, a key transcription factor involved in hepatocyte differentiation (Bluteau *et al.*, 2002). Its inactivation leads to glycolysis and aberrant fatty acid production that can induce tumor steatosis, lipogenesis and mammalian target of rapamycin (mTOR) activation (Merle *et al.*, 2019).
- The last identified subgroup is characterized by the activation of sonic hedgehog pathway (**shHCA**) and represents 4% of HCAs. The molecular mechanism at their origin is a small deletion leading to the *INHBE-GLI1* gene fusion. *INHBE* is highly expressed in the liver. *GLI1* is the key transcription factor of sonic hedgehog pathway and is normally barely expressed in the liver. The fusion leads to a

chimeric transcript, producing an over-expressed functional *GLI1* protein. shHCA are associated with a high risk of bleeding (Nault *et al.*, 2017).

- Finally, around 7% of HCAs (**UHCAs**) tumors are still unclassified. These tumors do not harbor any mutation in known driver genes nor particular transcriptomic signature (Nault *et al.*, 2017).

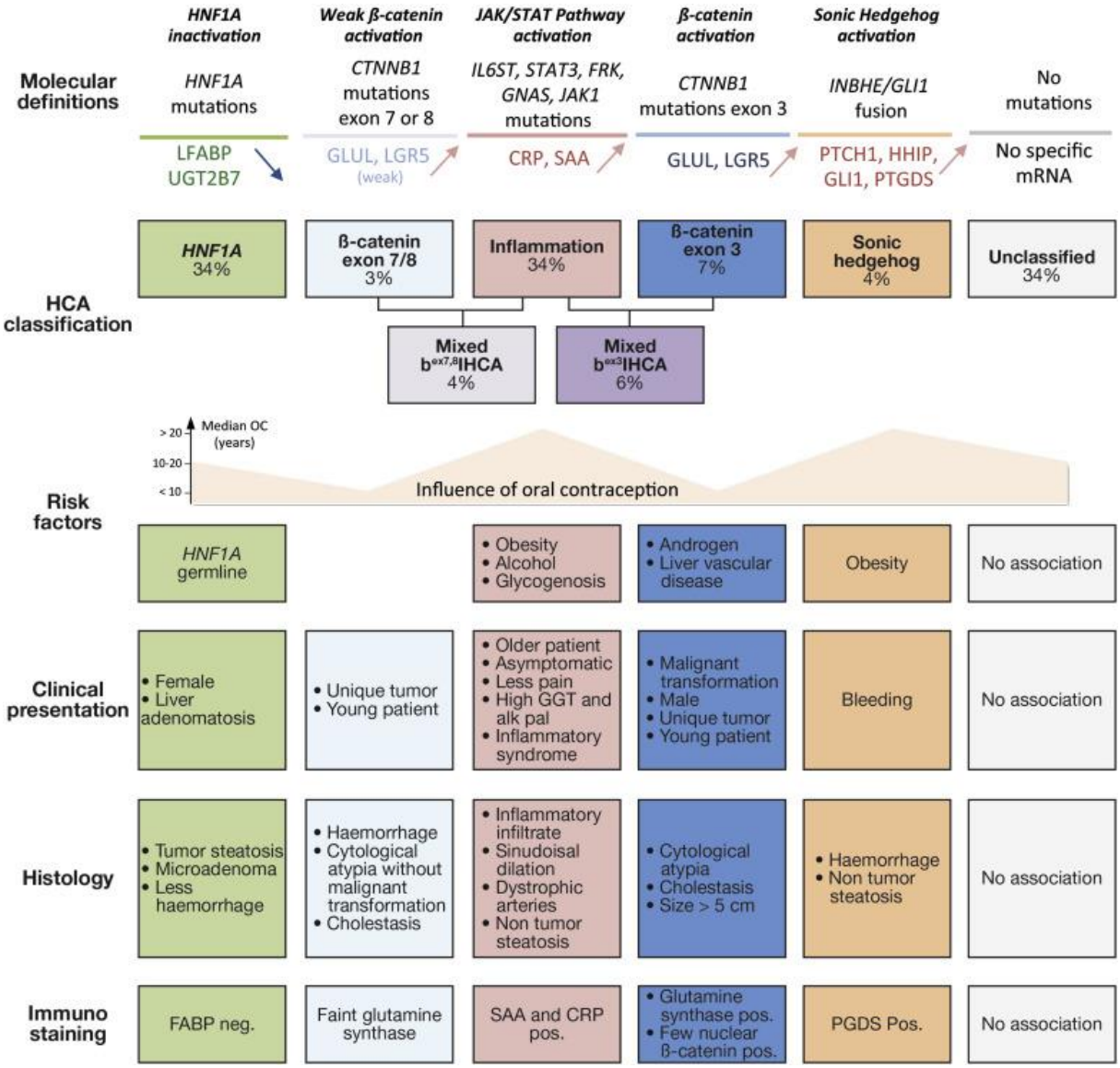


Figure 19: Molecular classification of HCA into 6 subgroups and 2 mixed subgroups. Associated risks factors, clinical and histological features of tumors are detailed above each subgroup (Nault et al., 2017).

3.1.3. Clinical features

Hepatocellular adenomas require clinical management to avoid the main complications that are tumor bleeding (in shHCA) and malignant transformation to hepatocellular carcinoma (HCC) (Nault *et al.*, 2017). Mutations in *CTNNB1* exon 3 are

known to be associated with a high risk of malignant transformation, making b^{ex3}HCA and b^{ex3I}HCA subgroups the most at risk of progression into HCC. By contrast, bex^{7,8}HCA and bex^{7,8I}HCA subgroups are not associated with HCC development, suggesting that a strong WNT/β-catenin activation is necessary to trigger malignant transformation. (Zucman-Rossi *et al.*, 2006; Pilati *et al.*, 2014; Rebouissou *et al.*, 2016; Nault *et al.*, 2017). *TERT* promoter mutations activating telomerase maintenance pathway were shown to be a recurrent second hit in adenoma to carcinoma transition, following *CTNNB1* exon 3 mutation (Pilati *et al.*, 2014).

3.2. Hepatocellular carcinoma

3.2.1. Epidemiology and risks factors for HCC

Hepatocellular carcinoma is the most frequent primary liver cancer (Yang and Roberts, 2010) and is responsible for more than 500,000 deaths yearly, making it the 3rd deadliest cancer worldwide (Bray *et al.*, 2018). HCC incidence depends on the geographical location, with around 80% of cases occurring in southeast Asia and sub-Saharan Africa (McGlynn *et al.*, 2001). Interestingly, etiology is also highly associated with geography (Figure 20A).

In Africa and Asia, HCC are most of the time associated with Hepatitis B Virus (HBV) chronic infection, whereas Hepatitis C Virus (HCV) is the main cause of HCC in some countries like Japan or Egypt (Wansbrough-Jones *et al.*, 1998; Lok *et al.*, 2009). HBV and HCV viruses induce acute of chronic hepatitis and ultimately cirrhosis, favoring HCC development (Baron, 1996). Indeed, chronic liver damage leads to a cycle of cell death, regeneration and fibrosis during which HCC precursor cells undergo malignant transformation. Moreover, insertional mutagenesis of HBV but also AAV2 are shown to drive tumorigenesis by activating oncogenes (Nault *et al.*, 2015; La Bella *et al.*, 2020). Aflatoxin B1 exposure is another major risk factor for HCC in those countries. Aflatoxin B1 is produced by *Aspergillus flavus*, a fungus that infects maize kernels in hot and wet countries, inducing subsequent accumulation of this toxin, one of the most potent carcinogens produced in nature (McGregor *et al.*, 1998), in growing food. Moreover, individuals exposed to both high levels of aflatoxin B1 and chronic hepatitis B virus infection have an increased risk of liver cancer (Liu and Wu, 2010). Another example of toxin as risk factor of HCC is aristolochic acid, a toxin produced plants from the genus

Aristolochia. This compound largely affects Southeast Asia region due to its use in traditional herbal medicine. Metabolites of aristolochic acid form DNA adducts on adenine residues leading to a specific mutational signature of T>A substitutions (Nault and Letouzé, 2019).

In western countries, the main risk factors are completely different. In France, HCC development is related to alcohol abuse in 50% of the cases (Rosa *et al.*, 2010). In the United States, obesity is the major risk factor of HCC as accumulation of fat in the liver induces chronic liver inflammation (Ascha *et al.*, 2010). Some rare diseases such as hemochromatosis that lead to iron deposit in liver can lead to cirrhosis (Kew, 2014). Finally, Nonalcoholic steatohepatitis (NASH), corresponding to liver steatosis and strong necro-inflammation, is the most common cause of chronic liver disease in developed countries and its incidence is rapidly increasing worldwide. As schematized on figure 20B, Nonalcoholic fatty liver diseases (NAFLD) are associated to risks factor such as obesity and diabetes and eventually lead to cirrhosis and NASH, that are both highly predisposing factor for HCC development (Fujii *et al.*, 2013; Anstee *et al.*, 2019) (Figure 20B).

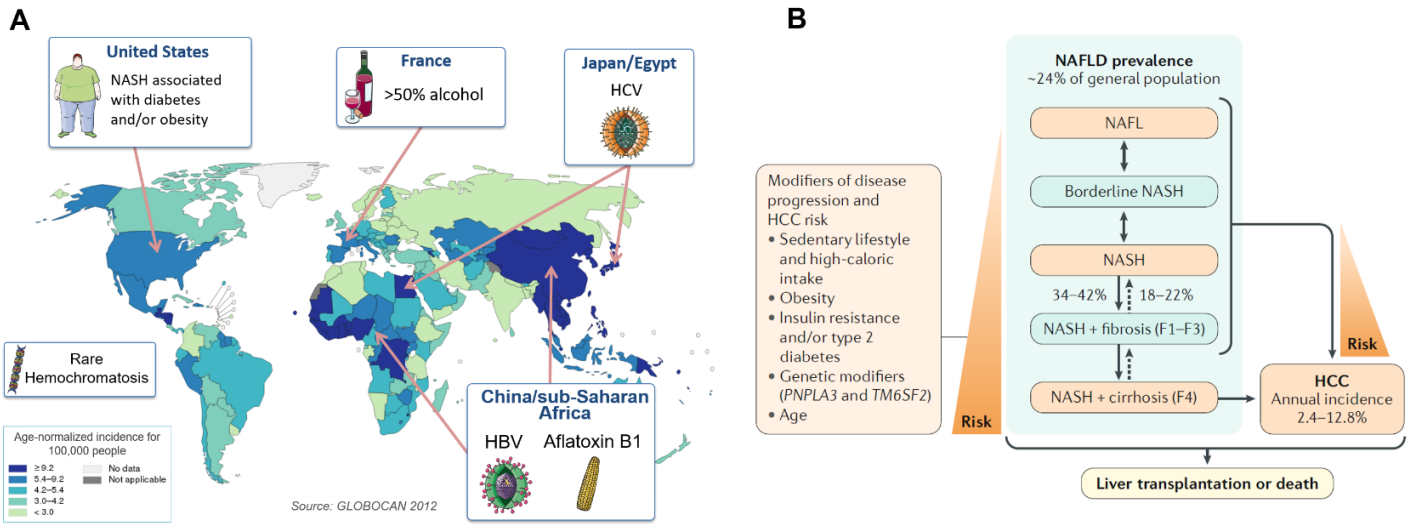


Figure 20: A: Estimated incidence rate of liver cancer worldwide. Major risk factor associated with specific regions are also represented. Data from GLOBOCAN 2012. B: Relation between nonalcoholic fatty liver disease (NAFLD) and HCC development initiated by various risks factors such as obesity or diabetes types I/II. NAFLD can lead to and further to NASH (steatosis and strong necro-inflammation), fibrosis, cirrhosis and eventually hepatocellular carcinoma (HCC). The HCC incidence vary from 2.4 to 12.8% depending on the disease state. (Anstee et al., 2019)

HCC develops most of the time on a diseased liver, with underlying cirrhosis in 80-90% of cases (Figure 21). Liver cirrhosis occurs after years of chronic liver inflammation induced by abovementioned exposures or diseases. Chronic inflammation induces a stress leading to excess of extracellular matrix and permanent fibrotic scars in the liver

(Wynn, 2004). The Metavir score indicates different stages of fibrosis, from F0 (normal liver) to F4 (cirrhosis). This liver background induces architecture changes leading to the formation of structurally abnormal nodules. Cells in these lesions accumulate somatic alterations such as mutations associated to high risk of developing HCC and hepatic function alteration (Ginès *et al.*, 2016; Brunner *et al.*, 2019). For example, *TERT* promoter mutation is a well-known early-acquired genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis, with more than 60% of early HCC harboring such alteration. (Nault *et al.*, 2013) (Figure 21).

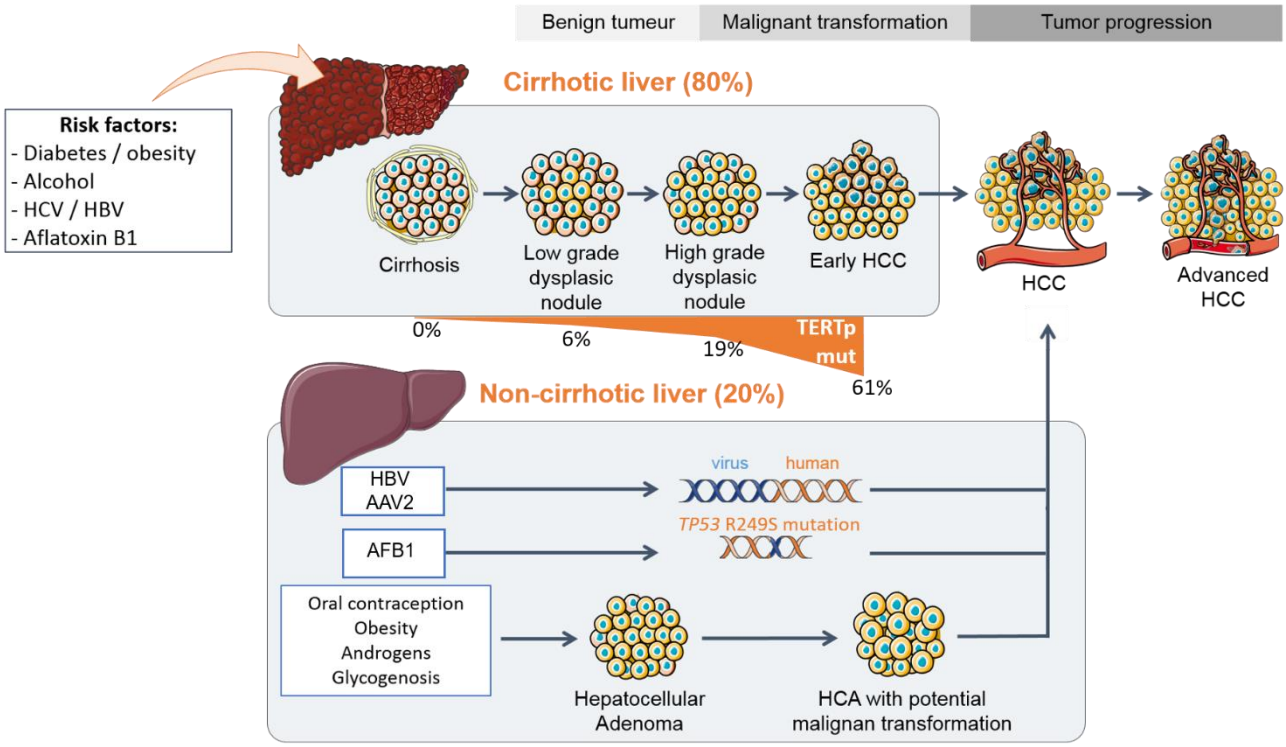


Figure 21 : HCC develops most of the time on a cirrhosis liver background associated to specific risks factors. The transition from cirrhotic liver to early HCC is characterized by a recurrent early alteration of TERT promoter, leading to tumor progress. However, HCC can also be developed from HCA malignant transformation of specific tumorigenic process such as insertional mutagenesis of HBV and AAV2 or exposure of Aflatoxin B1, known to induce R249S TP53 hotspot mutation. Modified from Zucman-Rossi J et al. 2015; Nault et al. 2015.

HCC also develops in a healthy liver in around 10-20% of cases (Figure 21). Except HCA transformation already discussed in the previous chapter, insertional mutagenesis by HBV or AAV2 viruses, or specific mutagenic exposures can trigger HCC tumorigenesis without long term damage induced by cirrhosis (Figure 21).

3.2.2. Major biological pathways altered in HCC

Since the advent of NGS, large cohorts of HCC were analyzed using WES, RNAseq and WGS (Schulze, Imbeaud, Letouzé, *et al.*, 2015; Fujimoto *et al.*, 2016), allowing to elucidate the landscape of driver genes and biological pathways recurrently altered in HCC (Zucman-Rossi *et al.*, 2015) (Figure 22).

- **Telomere maintenance** is the most frequently altered pathway with 90% of HCC harboring an increased telomerase expression (Farazi *et al.*, 2003; Plentz *et al.*, 2007). *TERT* gene encodes the catalytic sub-unit of the telomerase complex that is crucial to maintain telomere length and genome stability. This gene is not expressed in normal liver but activated in the early stages of hepato-carcinogenesis. Molecular mechanisms at the origin of *TERT* activation are i) TERT promoter mutation in 54 to 60% of cases (Nault *et al.*, 2013), ii) gene amplification or other structural rearrangements in 5 to 6% of cases (Totoki *et al.*, 2014) and iii) insertional mutagenesis in *TERT* promoter by HBV and AAV2 in 10 to 15% of cases (Nault *et al.*, 2015).
- **WNT/β-catenin pathway** is altered in 17 to 44% of HCCs (depending of geographical and etiological background) characterized by the transcriptional activation of *CTNNB1* target genes, such as *GLUL* and *LGR5* (Boyault *et al.*, 2007). This pathway activation is frequently associated with *TERT* promoter, suggesting a cooperation between the two pathways (Nault *et al.*, 2013; Totoki *et al.*, 2014; Schulze, Imbeaud, Letouzé *et al.*, 2015). This pathway is crucial for liver function, controlling important mechanisms such as cell differentiation in hepatocytes (Touboul *et al.*, 2016). The most common driver events in this pathway are *CTNNB1* mutations in exon 3 or in frame deletions in the exon 3. In addition, mutations in *AXIN1*, *ZNRF3* or *APC* tumor suppressor genes are found in respectively 10%, 3% and 1-2% of HCCs, also triggering WNT/β-catenin pathway activation (Yanai *et al.*, 2000; Basham *et al.*, 2019).
- **Cell cycle pathway** is altered in at least half of HCC, with *TP53* mutations in 12 to 48% of cases (Bressac *et al.*, 1991; Hsu *et al.*, 1991; Guichard *et al.*, 2012; Schulze, Imbeaud, Letouzé *et al.*, 2015; Fujimoto *et al.*, 2016). As already described for TSG, *TP53* mutations are spread along all the gene, except for the R249S hotspot characteristic of aflatoxin B1 exposure (Guichard *et al.*, 2012; Fujimoto *et al.*, 2016). *TP53* is the most frequently altered gene in cancer, and is involved in

multiple biological processes such as G1/S cell cycle transition, DNA repair and apoptosis (Aubrey *et al.*, 2018). The retinoblastoma pathway, that controls the G1/S phase transition, is also recurrently altered in HCC, following homozygous deletions of *CDKN2A* (2-12%) or *RB1* (3-8%) (Guichard *et al.*, 2012; Ahn *et al.*, 2014; Totoki *et al.*, 2014). Finally, *CCNE1* gene and *CCND1/FGF19* locus, both involved in cell cycle progression, are recurrently altered by insertional mutagenesis or amplifications in 5% and 5-14% of cases, respectively (Sawey *et al.*, 2011; Sung *et al.*, 2012; Kan *et al.*, 2013). Interestingly, *TP53* and *CTNNB1* mutations are exclusive (Friemel *et al.*, 2019).

- **Oxidative stress pathway** alterations are seen in 5-15% of HCCs and are recurrent in a context of chronic liver inflammation or long term exposure to toxic compounds such as alcohol (Llovet *et al.*, 2016). These liver insults lead to the accumulation of Reactive Oxygen Species (ROS) and subsequent selection of drivers events activating the detoxification program controlled by NRF2 (Denicola *et al.*, 2011; Bryan *et al.*, 2013). These driver events include activating mutations of *NFE2L2* (nuclear factor erythroid 2-related factor 2) encoding NRF2, or inactivating *KEAP1* (Kelch-like ECH-associated protein 1) mutations preventing NRF2 from proteasome degradation (Guichard *et al.*, 2012; Sporn and Liby, 2012; Cleary *et al.*, 2013).
- **PI3K/AKT/MTOR and RAS/RAF/mitogen-activated protein kinase pathways** are also frequently altered in HCC. PI3K/AKT/MTOR is involved in metabolism and cell proliferation (Laplante and Sabatini, 2012) while RAS/RAF/MAPK is involved in cell proliferation, differentiation and survival (Schubbert, Shannon and Bollag, 2007). PI3K/AKT/MTOR is altered by homozygous deletions or mutations of *PTEN* (1-3%), inactivating mutation of *TSC1*, *TSC2* (2-8%) (Totoki *et al.*, 2014) and *FGF3*, *FGF4*, *FGF19* amplifications (5%) that could induce FGFR receptor over-expression (Babina and Turner, 2017). RAS/RAF/MAPK is altered through *RPS6KA3* inactivating mutation in 2-9% of cases, leading to a constitutive activation of the pathway (Zucman-Rossi *et al.*, 2015).
- **Epigenetic modifiers** are the last identified pathway recurrently altered in HCC through chromatin remodeling complexes or histone methyl-transferase alterations (Schulze, Imbeaud, Letouzé *et al.*, 2015) (Figure 13). Recurrent inactivating mutations affect *ARID1A/B* (4-17% of HCC) and *ARID2* (3-18%),

involved in the SWI/SNF chromatin remodeling complexes BAF and PBAF. Histone methyl transferase complexes alterations include inactivating mutations in *KMT2A* (3-4%), *KMT2B* (3-4%), KMT2C (3-6%) and *KMT2D* (2-3%), or HBV insertional mutagenesis in *KMT2B* (10%) (Sung *et al.*, 2012; Cleary *et al.*, 2013; Schulze, Imbeaud, Letouzé, Ludmil B. Alexandrov, *et al.*, 2015).

3.2.3. Treatments

The BCLC (Barcelona Clinic Liver Cancer) stage of HCC development is the major metric to guide the management of patients with HCC. This stage depends on various physio-pathological parameters, the number and size of tumor nodules (Figure 23).

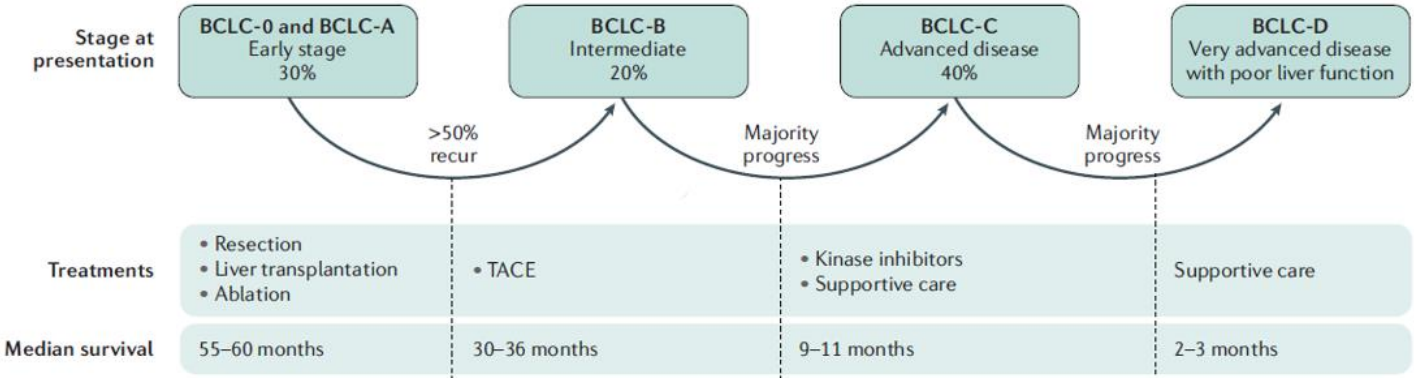


Figure 22: Percentage of patients with HCC within the different BCLC stages. Within the first 2 years after liver resection, around 50% of patients relapse with liver cancer. Modified from (Anstee et al. 2019).

For BCLC-0 and BCLC-A stages, transplantation is the best therapeutic option as it resolves both tumor and cirrhosis. However, the number of donors is obviously limited. Resection and percutaneous ablation through Percutaneous Ethanol Injection (PEI) or Percutaneous radiofrequency ablation (RFA) are alternative curative treatments, but require a preserved hepatic function of the patient, that is altered by both cancer developments and cirrhosis. Unfortunately, patients are frequently diagnosed at advances stages of HCC development, limiting curative treatments. Altogether, only 15% of all hepatocellular carcinomas are amenable to operative treatment (Bruix and Llovet, 2002; Chedid, 2017).

For patients not eligible to curative treatment, two categories of non-curative treatments are used according to the BCLC stage. For BCLC-B (intermediate HCC), intra-arterial chemoembolization preventing blood supplying to the tumor and intra-arterial

chemotherapy are the two options. Those types of treatment increase patient survival but can have important side effects (Galuppo *et al.*, 2014).

For patients with BCLC-C (advanced HCC), 4 multi-kinase inhibitors targeting PI3K/AKT/MTOR or RAS/RAF/MAPK pathways are approved. Sorafenib was for a long time the only therapeutic option. This is an inhibitor of VEGFR-1/2/3 (vascular endothelial growth factors receptors), BRaR/CRaF (serine threonine kinase) and PDGFR-α/β (platelet-derived growth factor receptors). This treatment was shown to increase the median survival by 3 months (Llovet *et al.*, 2008). Regorafenib is a multi-kinase inhibitor more efficient against VEFGR kinases, with a wider activity than sorafenib, also targeting TIE2, KIT and RET kinases involved in ontogenesis. This treatment showed positive results in a cohort of patients progressing under sorafenib (Bruix *et al.*, 2017). Recently, new compounds were approved for the treatment of advanced HCC such as lenvatinib (multi-kinase inhibitor of VEGFR 1-3, FGFR 1-4, PDGFRα and RET), Cabozantinib in second line, an inhibitor of the tyrosine kinases c-Met and VEGFR2, and the tyrosine kinase receptors AXL and RET (Abou-Alfa *et al.*, 2018). Moreover, Ramucirumab, a monoclonal anti-body targeting VEGFR-2 shown promising result in clinical phase III (Zhu *et al.*, 2019).

Immune checkpoint inhibitors are emerging in HCC because of their considerable efficacy in other cancers, particularly in patients with melanoma, lung, renal, and bladder cancer (Smyth *et al.*, 2016). Nivolumab and pembrolizumab, two monoclonal antibodies against PD-1 (Programmed cell death-1), have just received accelerated approval by the FDA based on promising results from two Phase 2 studies showing a reduction in tumor mass in 20% and 17% of patients previously treated with sorafenib, respectively (El-Khoueiry *et al.*, 2017; Guo, Zhang and Chen, 2017; Zhu *et al.*, 2018). Thus, immunotherapy is promising in HCC treatment, and phase III studies are ongoing, including a phase III trial comparing sorafenib to nivolumab in first line (CheckMate 459; NCT02576509). However, treatment response is heterogeneous and, in a small proportion of patients, these treatments can cause severe and possibly permanent auto-immune adverse effect such as diarrhea, skin reactions, fatigue and hypertension (Weber *et al.*, 2015). Thus, there is a huge need to identify candidate bio-markers to target patients who are likely to benefit from immunotherapies. So far only PD-L1 expression has been approved as bio-marker but does not seems to be associated with nivolumab response in HCC (El-Khoueiry

et al., 2017). A recent study defined an "immune exclusion phenotype" subgroup of HCC, characterized by *CTNNB1* mutations, lower immune infiltration (based on immune-specific gene signature) and overexpression of *PTK2,* an oncogenic pathway associated with low T-cell infiltrate in the tumor (Luke *et al.*, 2019).

Despite the recent increase in available therapies, the prognosis of HCC patients is still very bad, with a median survival of less than 1 year. Furthermore, patient response to available therapies is heterogeneous, making the discovering of new therapeutic options crucial. The recent advances in cancer genomics lead to a new era in cancer research consisting in identifying molecular and transcriptomic subgroups of tumors harboring similar molecular features, likely to respond well to the same targeted therapy. Unraveling the diversity of molecular alterations and transcriptomic deregulation is thus crucial to i) understand the tumor biology and deregulated pathways in order to develop efficient patient tailored therapy and ii) identify biomarkers of homogeneous molecular subgroups. For example, MEK inhibitors already used in breast cancer, melanoma and lung cancer displayed efficacy in HCC patients characterized by RAS pathway activation (Das *et al.*, 2018).

4. Objectives of the PhD

Liver cancers are extremely heterogeneous, with a wide spectrum of etiologies, tumor histological characteristics, deregulated biological pathways and molecular alterations. Our understanding of the mechanisms driving tumor initiation and development increased a lot in the last decade, both in terms or biological pathways and altered oncogenes or TSG. However, some tumors remain poorly understood, and even if the majority of them are well characterized in terms of biological phenotype, the underlying driver alterations are sometimes missing. The aim of my PhD project was to better understand the molecular mechanisms at the origin of HCAs and HCCs. More precisely, I focused on structural variations (SVs) that had not been analyzed extensively in liver cancers. Indeed, when I arrived in the laboratory, the only known driver SV was the *INHBE-GLI1* fusion, characteristic of the shHCA cancer subgroup. Additionally, SV

signature analysis helped to better understand biological mechanisms at the origin of specific genomic rearrangement.

In this project, I developed innovative computational strategies applied to WGS and RNA-seq data of HCA and HCC tumor collections of the UMRS 1138 INSERM laboratory, obtained from collaborations with various French hospitals. I also used public data sets to increase statistical power and/or validate results, including TCGA (334 HCC samples with multi-omic data) and ICGC series (257 HCC samples with WGS). Finally, I used epigenetic data generated by the ENCODE and ROADMAP consortia, such as replication timing, chromatin opening and histone marks to better understand both the origin and functional consequences of structural rearrangements.

Overall, during my PhD, I contributed to shape the landscape of both HCA and HCC cancers subgroups:

- First, by integrating WGS and RNAseq data, I characterized 3 new driver alterations in inflammatory HCA. These include recurrent fusions activating *FRK* and *ROS1* genes, and an original mechanism leading to *IL6* gene overexpression following the loss of regulatory 3'UTR regions. All these alterations lead to a constitutive activation of the JAK/STAT pathway characteristic of inflammatory HCA.
- Secondly, by analyzing SV signatures in HCC, I discovered a unique rearrangement signature of tandem duplications and templated insertion cycles, induced by replication stress. This signature allowed me to define a new HCC subgroup driven by various oncogenic mechanisms activating *CCNA2* or *CCNE1*.

II MATERIAL AND METHODS

1. Data used in the project

1.1. Tissue samples

Tumor samples from both hepatocellular carcinoma (HCC) or hepatocellular adenoma (HCA) samples and their non-tumor counterparts were collected from patients surgically treated in four French hospitals located in Bordeaux and Paris region. Institutional review board committees (CCPRB Paris Saint-Louis, 1997, 2004, and 2010, approval number 01–037; Bordeaux, 2010-A00498–31) approved the studies. Written informed consent was obtained in accordance with French legislation. All samples were immediately frozen in liquid nitrogen and stored at −80 °C. HCC were enriched in cases developed on a non-cirrhotic liver with an over-representation of tumors developed in non-fibrotic (METAVIR F0-F1) compared to tumors developed in chronic hepatitis (F2–F3) and in cirrhotic liver (F4). Clinicopathological data were available for all cases (detailed tables can be found in respective articles). A diversity of risk factors were represented in the cohorts mentioned in this manuscript, including alcohol, metabolic syndrome, HBV, HCV infections, hemochromatosis and without particular etiologies.

1.2. Sequencing methods

1.2.1. Target sequencing

Target sequencing was performed on panels of genes including *HNF1A* (exon 1–10), *IL6ST* (exons 6 and 10), *CTNNB1* (exons 2, 3, 4, 7 and 8), *FRK* (exon 6), *STAT3* (exons 3, 6, 17 and 21), *GNAS* (exons 7, 8 and 9) and *JAK1* (all exons). Those genes were sequenced by either Sanger sequencing or Miseq Illumina PCR-based sequencing. Somatic mutations were confirmed by sequencing of the tumor and non-tumor counterpart.

1.2.2. Whole Exome Sequencing (WES)

Whole Exome Sequencing data was performed by IntegraGen (Evry, France). Agilent in-solution enrichment was used with the manufacturer's biotinylated oligonucleotide probe library SureSelect Human All-Exon kit V.4 70Mb, or SureSelect Human All-Exon kit V.5 +UTRs, or Twist Human Core Exome Enrichment System. Genomic DNA was sonicated and purified to yield fragments of 150–200bp. Adaptor oligo-nucleotides were ligated on A-tailed fragments and enriched by four to six PCR cycles. Purified libraries (500ng) were hybridized to the library for 24 hours. The eluted enriched DNA sample was sequenced

on an Illumina HiSeq 2000 or HiSeq 4000 as paired-end 75bp reads.

1.2.3. RNA-seq

RNA samples were enriched for polyadenylated RNA from 5µg of total RNA, and the enriched samples were used to generate sequencing libraries with the NEBNext Ultra II Directional RNA or Illumina TruSeq Stranded mRNA kits and associated protocol as provided by the manufacturer. Libraries were sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 2000 or 4000 as paired-end 75 or 100 bp reads.

1.2.4. Whole Genome Sequencing (WGS)

Whole Genome Sequencing of samples was realized by the Centre National de Génotypage (Evry, France) or were sequenced at Integragen (Evry, France). Samples were sequenced on an Illumina HiSeq X Five as paired-end 75, or 101 base pair (bp) reads.

1.2.5. Table of sequencing data by project

Project	Sample type	Gene panel sequencing	WGS	WES	RNA-seq	Other Public Data
Article1: Systemic AA amyloidosis	HCA / NT	NA	N=1	NA	N=2	
Article2: Inflammatory HCA alterations	HCA	N=657 (533 already published)	NA	N=19 (4 already published)	N=22 (1 already published)	
Article3: Palimpsest	Method paper, not using any specific data set					
Article4: Cyclin A2/E1 activated HCC	HCC	NA	N=45 (35 already published)	N=156 (96 already published)	N=160	- TCGA series of HCC (n=334 RNAseq; n=334 WES; n=48WGS) -ICGC-JP series of HCCs (n=257 WGS/RNAseq) -PCAWG SVs from 2606 tumors WGS -Encode and Roadmap annotations

2. Bioinformatics Pipelines

2.1. RNAseq alignement and expression matrix generation:

Full Fastq files were aligned to the reference human genome GRCh38 using TopHat2 V.2.0.14. Read mapping from multiple locations was removed and HTSeq was used to obtain the number of reads associated with each gene in the Gencode V.27 database, restricting to protein-coding genes, pseudogenes, antisense and lincRNAs. Bioconductor

DESeq2 package was used to import raw HTSeq counts for each sample into R statistical software and apply variance stabilising transformation to the raw count matrix. FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalizing the count matrix for the library size estimated with DESeq2 package and the coding length of each gene. In order to handle batch effect, we used an in-house method. Briefly, area under the ROC curve (AUC) was used to identify and remove genes with a significant batch effect (AUC >0.95 between one sequencing project and others).

2.2. RNAseq Gene Set Enrichment Analysis (GSEA):

To perform gene set enrichment analysis (GSEA) analysis, genes expression from previous pipeline were sorted according to fold-change from Limma analysis (genes with positive fold-changes first, then negative fold-changes) using an input of 16387 genes (Hallmarks gene set; MSigDB V.6 database used). This gene list was obtained from Genecode database; then genes with a variance equal 0 or with a total number of reads from samples inferior of 50 were removed.

2.3. RNAseq fusion calling

Fusions detected by TopHat2 (--fusion-search; --fusion-min-dist 2000; --fusion-anchor-length 13; --fusion-ignore-chromosomes chrM) were filtered using the TopHatFusion pipeline. Fusions validated by BLAST and with at least 10 split-reads or pairs of reads spanning the fusion event were retained, whereas any fusion identified at least twice in a cohort of normal liver samples was removed.

2.4. WES / WGS alignment and mutation calling

Sequences were aligned to the hg19 version of the human genome using BWA version 0.7.12. and variant calling was processed using Mutect2 by comparing each tumor sample with its matched non-tumor counterpart and a panel of normal (PON) file, following the genome analysis toolkit (GATK) best practices.

Mutations belonging to the ENCODE Data Analysis Consortium blacklisted regions (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz) were excluded and regions covered by < 6reads in the tumor or normal sample. We then selected only single nucleotide variants (SNVs) with a MuTect2 flag among “PASS”, “clustered_events”, “t_lod_fstar”, “alt_allele_in_normal” or “homologous_mapping_event” and small insertions

and deletions (indels) with a MuTect2 flag among "PASS", "cluster-ed_events" or "str_contraction". To improve specificity in the calling of mutations with low variant allele frequency (VAF), we quantified the number of high quality variant reads in the tumor (mapping quality≥20, base quality≥20) and the number of variant reads in the non-tumor sample with no quality threshold using bamreadcount (https://github.com/genome/bam-readcount). Only variants matching the following criteria were finally retained: VAF≥2% in the tumor with≥3 variant reads, VAF≤5% in the non-tumor samples with≤2 variant reads, and a VAF ratio≥5 between the tumor and non-tumor sample.

2.5. WES / WGS copy number analysis and SV calling

The copy number profiles were defined using an in-house pipeline. Briefly, the logR ratio was processed using the coverage of each tumor and its non-tumor counterpart in each 1kb genomic windows; then a segmentation was done in order to identify copy number alterations.

MANTA software was used to identify somatic structural rearrangements in WGS data. To keep only the most reliable events, we selected only rearrangements supported by≥10 reads and with a variant allele fraction ≥ 5%.

2.6. Signature analysis and statistical tools

Some biological processes linked to cancer development was shown to be at the origin of particular pattern of alterations, as explain in Introduction section. Those patterns of alterations associated to cancer are called mutational signatures and the combination of those processes in a cancer cell result in a complex mutational portrait. This portrait can be deconvoluted in order to extract the different biological sources of alteration in each tumor. Non-negative Matrix Factorization (NMF) mathematical method was shown to be very effective in the deconvolution of complex biological signal. This method is implemented in the Palimpsest R package (https://github.com/FunGeST/Palimpsest), corresponding to the 3rd Article of this manuscript. All function related to signature analysis was realized using this package.

Concerning in house developed methods, both R (version from 3.4.4 to 4.0.2) and Python (versions 2.7 and 3.6) were used in the bioinformatics analysis, including graphical representation and statistical analysis through hypothesis testing.

2.7. Table of pipeline used by project

Project	Target sequencing	RNAseq	WGS	WES
Article1: Systemic AA amyloidosis	NA	-Gene expression	-Structural variant calling	NA
Article2: Inflammatory HCA alterations	-Mutation calling -Expression of gene panel	-Gene expression -Gene fusions -GSEA	-Structural variant calling	-Mutation calling -Copy number alterations
Article3: Palimpsest	Method paper, not using any processed data from specific dataset			
Article4: Cyclin A2/E1 activated HCC	NA	-Gene expression -Gene fusions -GSEA	-Structural variation calling -Mutation calling -Viral insertion calling	-Mutation calling -Copy number alterations

III RESULTS

1. Identification of oncogenic SVs in HCA

The first goal of my thesis was to identify new SVs driving hepatocellular adenoma development. As explained in the introduction, when I started my PhD, the molecular landscape of HCA was divided in 6 groups (Figure 10) associated with specific molecular alterations: HHCA with *HNF1A* inactivation, b^{ex3}HCA and b^{ex7,8}HCA with *CTNNB1* activation, shHCA with *GLI1* activation and IHCA with *IL6ST*, *FRK*, *STAT1*, *GNAS*, *JAK1* mutations and UHCA without known driver. However, this landscape was still incomplete. Firstly, because the driver genes of UHCA are still missing. Secondly because a large part of IHCA samples, characterized by SAA and CRP protein over-production, had no identified molecular driver mechanism. Except for the recently described *INHBE-GLI1* fusion characteristic of shHCA, all driver alterations in HCA were point mutations. The aim of this project was to unravel new alterations such as gene fusions or regulatory region alterations that could trigger tumorigenesis of HCA.

To that aim, we leveraged the large HCA tumor collection from the lab, including extensive clinical and molecular features:

- Clinical data and history of patients
- Histological images of tumor tissue
- Gene panel sequencing (from Illumina's MiSeq technology)
- Gene panel expression (from Fluidigm technology)

This large cohort was explored to select samples belonging to IHCA subgroup (based on gene expression) without identified point mutations in known driver genes. Selected samples were sequenced using RNA-seq and/or WGS, and analyzed to identify the missing drivers. This project lead to two original articles describing new alterations in IHCA subgroup.

Article 1 describes a rare clinical case associating inflammatory HCA with amyloid A (AA) systemic amyloidosis, a complication of chronic inflammatory diseases induced by deposits of insoluble SAA deposits. The patient, a 49-years-old woman, was admitted to emergency for diarrhea, rectal bleeding and lower limbs edema. Her clinical history was also characterized by an unexplained chronic inflammation for at least 10 years. Medical analysis revealed inflammation, renal failure, and SAA amyloid deposits were found in

kidneys, intestines and rectum. Liver biopsy led to HCA diagnosis that was further explored using both RNA-seq and WGS. WGS data revealed an inversion in chromosome 7 leading to the loss of *IL6* gene 3'UTR that contains post-transcriptional regulatory sequences involved in the degradation of this unstable transcript. RNAseq data indicated a massive over-expression of the *IL6* transcript. This mechanism of IHCA development is very different from those already known, where *STAT3* is activated through activation of a major actor of signal transduction of the IL6 axis, such as *IL6ST*, *JAK1*, *GNAS*, *FRK* or *STAT3* itself. In addition to IHCA development, the autocrine interleukine-6 secretion by the tumor triggers SAA production not only in the IHCA nodule but also in the surrounding liver tissue. Interestingly, amyloidosis symptoms improved following HCA resection. Thus, this single genomic alteration allowed us to make a link between IHCA and AA-amyloidosis, recurrently associated in literature regarding the very low prevalence of both diseases.

Article 2 describes the identification of new driver gene fusions leading to JAK/STAT pathway activation in previously unexplained IHCAs. The first recurrent fusions involve *FRK*, already known to be mutated in 8% of IHCA (Nault *et al.*, 2017). *FRK* fusions downstream different partner genes were identified in 5 patients (2% of the IHCA cohort), leading to chimeric transcripts only including exons 3 to 8 of FRK. The resulting proteins have lost the SH2 and SH3 auto-inhibitor elements of FRK, leading to constitutive activation of the tyrosine protein kinase domain. In addition, we identified recurrent gene fusions involving *ROS1* in 10 patients (3% of the IHCA cohort). The chimeric transcripts only retain exons 33 to 43, 34 to 43 or 35 to 43 of *ROS1*. In these samples, the upstream partner gene induces a massive overexpression of the truncated *ROS1* that contains kinase domain. *FRK* and *ROS1* fusions constitute new driver events activating the JAK/STAT pathway and leading to the development of IHCA.

Article 1: Systemic AA Amyloidosis Caused by Inflammatory Hepatocellular Adenoma

Systemic AA Amyloidosis Caused by Inflammatory Hepatocellular Adenoma

TO THE EDITOR: Amyloid A (AA) systemic amyloidosis is a complication of chronic inflammatory diseases that is caused by the deposition of insoluble aggregates of cleaved N-terminal fragments of serum amyloid A (SAA) protein in tissues and organs throughout the body.[1] Under physiologic conditions, SAA protein is produced by hepatocytes during the acute inflammatory phase in response to various cytokines such as interleukin-6. SAA is also overexpressed by neoplastic hepatocytes in inflammatory hepatocellular adenomas, a specific molecular subtype of benign liver tumors.[2,3]

Here, we describe a 49-year-old female patient who presented with diarrhea, rectal bleeding, and leg edema and received a diagnosis of systemic AA amyloidosis. Clinical characteristics of the patient are provided in Table S1 in the Supplementary Appendix, available with the full text of this letter at NEJM.org. An inflammatory hepatocellular adenoma was identified and resected, resulting in the improvement of amyloidosis-related symptoms, with progressive normalization of the serum C-reactive protein (Fig. 1A) and SAA levels. Unfortunately, the patient's kidney function did not improve, and she died 18 months after surgery from fulminant septic shock.

Whole-genome sequencing of the inflammatory hepatocellular adenoma revealed a complex structural rearrangement on chromosome 7 with an inversion leading to the truncation of the *IL6* 3′ untranslated region (3′UTR), which comprises key sequences involved in the degradation of this unstable transcript (Fig. 1, and Tables S2 and S3

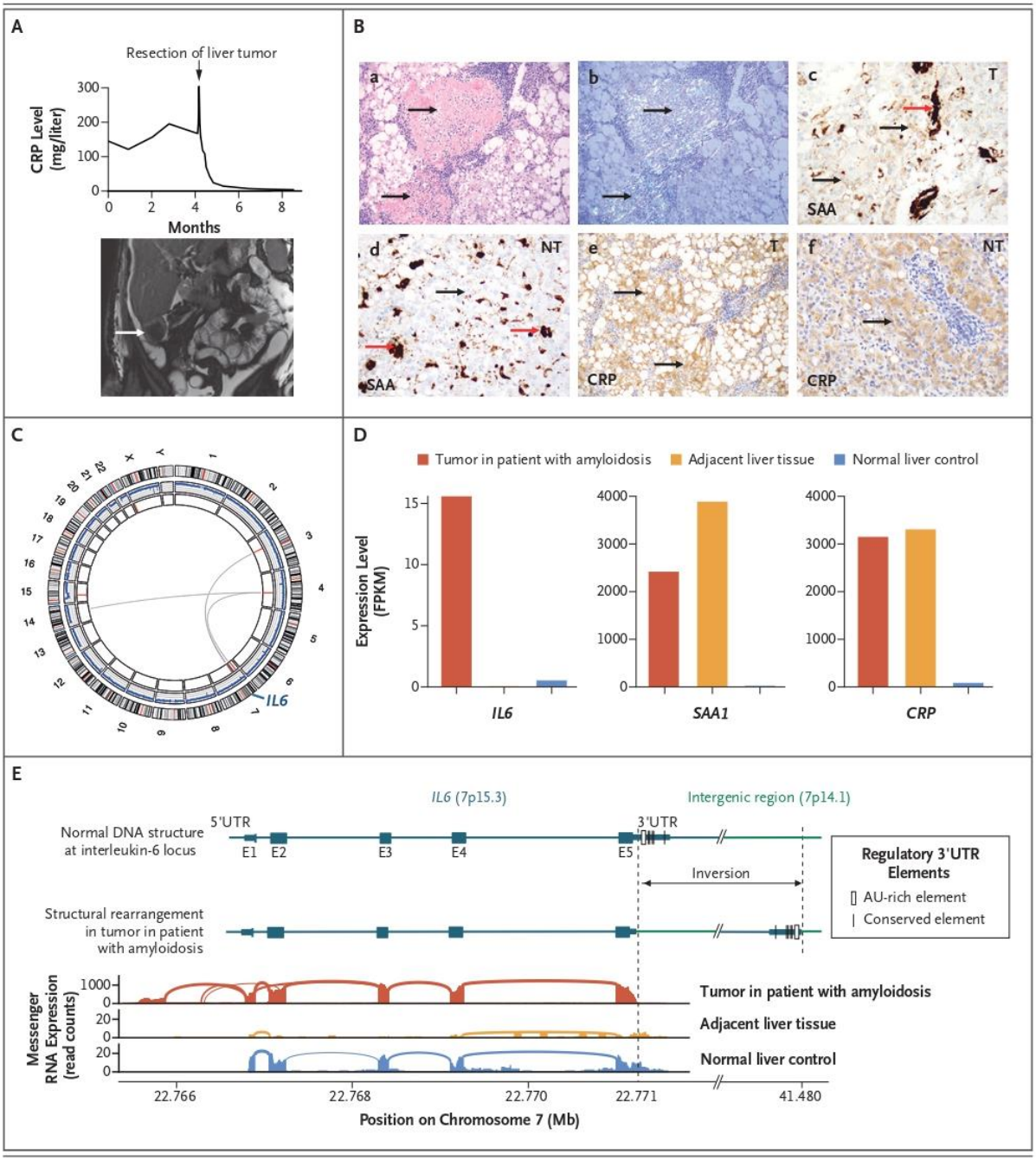
Figure 1 (facing page). Biologic, Imaging, Pathological, and Molecular Findings in Our Patient.

As shown in Panel A, liver resection was followed by a clinically significant decrease in serum levels of C-reactive protein (CRP) (top), and magnetic resonance imaging showed a heterogeneous, 5-cm liver mass at the tip of segment VI on T_2-weighted sequences (bottom, arrow). As shown in Panel B, microscopic examination of the resected specimen revealed a well-differentiated tumor with massive amyloid deposits (arrows) (subpanel a, Congo red staining, low magnification) and typical yellow-green birefringence under polarized light (subpanel b, low magnification). Immunohistochemical tests showed positive staining for serum amyloid A (SAA) protein in neoplastic hepatocytes (black arrows) and amyloid deposits (red arrows) in the tumor (T) sample (subpanel c, high magnification); SAA expression was also observed in the adjacent nontumorous (NT) parenchyma (subpanel d, high magnification). CRP was expressed in the tumor (T) (subpanel e, low magnification) (neoplastic hepatocytes, black arrows) and the adjacent NT liver tissue (subpanel f, high magnification) (non-neoplastic hepatocytes, black arrows). As shown in Panel C, whole-genome sequencing of the tumor sample obtained from the patient revealed a cluster of somatic structural rearrangements at the *IL6* locus on chromosome 7. The blue line in the outer circle indicates copy number, and structural rearrangements are indicated in the inner circle. Red indicates deletion, blue inversion, gray interchromosomal translocation, and black classic chromatin staining of the cytobands. As shown in Panel D, although overexpression of *IL6* was specific to the tumor sample, the genes *SAA1* and *CRP* were also massively overexpressed in the adjacent liver tissue, as compared with normal liver tissue. FPKM denotes fragments per kilobase of exon per 1 million reads in the RNA sequencing experiment. As shown in Panel E, chromosomal inversion with one breakpoint located in the 3′ untranslated region (UTR) of the gene *IL6* and the other breakpoint in an intergenic region at 7p14.1 led to the massive overexpression of an interleukin-6 transcript lacking regulatory 3′UTR elements in the patient's tumor (top, structural rearrangement identified by means of whole-genome sequencing; bottom, gene expression and transcript structure identified by means of RNA sequencing). AU denotes adenylate–uridylate.

N ENGL J MED 379;12 NEJM.ORG SEPTEMBER 20, 2018

The New England Journal of Medicine
Downloaded from nejm.org at INSERM DISC DOC on May 6, 2020. For personal use only. No other uses without permission.
Copyright © 2018 Massachusetts Medical Society. All rights reserved.

in the Supplementary Appendix). RNA sequencing consistently showed a massive overexpression of an interleukin-6 transcript lacking all 3′ regulatory elements; this overexpression was limited to the tumor area. However, in contrast to classic inflammatory hepatocellular adenoma, we observed overexpression of SAA in both the tumor and nontumorous liver. Amyloid deposits were also identified in the adenoma and the adjacent parenchyma (Fig. 1). These results support an autonomous production of interleukin-6 by the inflammatory hepatocellular adenoma with activation of a paracrine Janus kinase–signal transducers and activators of transcription 3



The New England Journal of Medicine
Downloaded from nejm.org at INSERM DISC DOC on May 6, 2020. For personal use only. No other uses without permission.
Copyright © 2018 Massachusetts Medical Society. All rights reserved.

(JAK–STAT3) pathway in the liver (Fig. S1 in the Supplementary Appendix).

In the literature, we identified eight additional patients with both AA amyloidosis and hepatocellular adenomas. In at least six of these patients, amyloid-related symptoms also improved after surgical resection of the liver tumor (Table S4 in the Supplementary Appendix). We reviewed the liver histologic findings of two patients from the literature and identified features that were similar to those of our patient; inflammatory hepatocellular adenoma with SAA expression by the tumor and the adjacent parenchyma was indeed identified in both cases from the literature (Fig. S2 in the Supplementary Appendix).[4,5] Birefringent, SAA-positive amyloid deposits were also identified in the adenomas and adjacent liver tissue.

In conclusion, rare cases of inflammatory hepatocellular adenoma can induce massive SAA production by the tumor and nontumorous liver, leading to systemic AA amyloidosis. Molecular analyses identified a somatic rearrangement of the *IL6* 3′UTR leading to an autocrine interleukin-6 secretion by the tumor. We speculate that this subset of AA amyloidosis may be treatable by surgical resection of the liver tumor.

Julien Calderaro, M.D., Ph.D.
Henri Mondor University Hospital
Créteil, France

Eric Letouzé, Ph.D.
Quentin Bayard, Ph.D.
Anais Boulai
INSERM Unité 1162
Paris, France

Victor Renault
Jean-François Deleuze, Ph.D.
Centre National de Recherche en Génomique Humaine
Évry, France

Oriol Bestard, M.D., Ph.D.
Bellvitge University Hospital
Barcelona, Spain

Dominique Franco, M.D., Ph.D.
Antoine Béclère University Hospital
Clamart, France

Elie-Serge Zafrani, M.D.
Henri Mondor University Hospital
Créteil, France

Jean-Charles Nault, M.D., Ph.D.
INSERM Unité 1162
Paris, France

Michel Moutschen, M.D., Ph.D.
Liege University Hospital
Liege, Belgium

Jessica Zucman-Rossi, M.D., Ph.D.
INSERM Unité 1162
Paris, France
jessica.zucman-rossi@inserm.fr

Drs. Calderaro, Letouzé, and Bayard contributed equally to this letter.

Supported by the Institut National du Cancer with the International Cancer Genome Consortium (ICGC Liver Cancer–France project), INSERM with the "Cancer et Environnement" (plan Cancer) and Heterogeneity of Colorectal and Liver Cancer (HETCOLI) projects (Tumor Heterogeneity and Ecosystem program), the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology Investissement d'Avenir, Coup d'Elan de la Fondation Bettencourt–Schueller, the Site de Recherche Intégrée sur le Cancer–Cancer Research and Personalized Medicine (SIRIC CARPEM), Fondation Mérieux, Cancéropôle Ile de France (exhauTrans project), and Instituto de Salud Carlos III (Spanish competitive grants FIS PI16/01321 and INT15/00112, to Dr. Bestard) through the European Regional Development Fund for research.

Disclosure forms provided by the authors are available with the full text of this letter at NEJM.org.

1. Wechalekar AD, Gillmore JD, Hawkins PN. Systemic amyloidosis. Lancet 2016;387:2641-54.
2. Nault JC, Couchy G, Balabaud C, et al. Molecular classification of hepatocellular adenoma associates with risk factors, bleeding, and malignant transformation. Gastroenterology 2017;152(4):880-894.e6.
3. Rebouissou S, Amessou M, Couchy G, et al. Frequent in-frame somatic deletions activate gp130 in inflammatory hepatocellular tumours. Nature 2009;457:200-4.
4. Bestard Matamoros O, Poveda Monje R, Ibernon Vilaró M, Carrera Plans M, Grinyó Boira JM. Systemic AA amyloidosis induced by benign neoplasms. Nefrologia 2008;28:93-8. (In Spanish.)
5. Fievet P, Sevestre H, Boudjelal M, et al. Systemic AA amyloidosis induced by liver cell adenoma. Gut 1990;31:361-3.

DOI: 10.1056/NEJMc1805673

Supplementary Appendix

Table of contents

Additional clinical, biological and molecular data

Additional clinical and biological data

The 49-year-old female patient was admitted to the emergency department at CHU de Liège (Belgium) for diarrhea, rectal bleeding and lower limbs edema. She had been suffering from chronic inflammation for at least 10 years, a condition that remained unexplained despite multiple explorations performed in another hospital. Her medical history revealed pregnancy-associated thyrotoxicosis treated with propylthiouracil, she was chronically treated with bisoprolol and had no particular familial history.

Physical examination showed a body mass index of 32, a blood pressure of 175/110 mm Hg, a heart rate of 120 bpm and a body temperature of 36.7 °C. Severe pitting edema was present at the lower limbs level. Blood and urine analyses revealed major inflammation, renal failure, and nephrotic syndrome with normal procalcitonin (Table S1). Colonoscopy showed diffuse erythema and petechiae. Abdominal Computerized Tomography scan demonstrated a 5 cm hypodense lesion of segment VI of the liver, that was further confirmed by magnetic resonance imaging. Kidney biopsy showed nodular mesangial thickening with strong positivity for SAA protein in the mesangium, interstitium and vessel walls. Typical SAA amyloid deposits were also present on rectal and colic biopsies. A biopsy of the liver nodule concluded to a diagnosis of HCA, and the tumor was further surgically resected. Pathological examination of the liver tumor revealed hallmark features of IHCA: neoplastic hepatocytes without cytological atypia and arranged in thin trabeculae, inflammatory infiltrates, dystrophic vessels, sinusoidal dilatation. IHCA are benign monoclonal proliferations of hepatocytes and are characterized in 80% of the cases by an activating somatic mutation of one of the genes of the IL6 axis (*IL6ST*, *GNAS*, *FRK*, *JAK1* or *STAT3*) that lead to constitutive STAT3 signaling pathway activation. These genetic alterations are responsible for the production by neoplastic cells of acute phase proteins such as SAA and C-reactive protein (CRP), regardless of the presence of ligands such as IL6.

Partial hepatectomy allowed rapid improvement of general condition and diarrhea. Serum CRP and SAA progressively normalized. Unfortunately, kidney function did not improve and the patient had to remain under dialysis. Eighteen months after her initial admission, she developed a methicillin-resistant Staphylococcus aureus bacteremia (presumably originating from the dialysis catheter) and died from a fulminant septic shock. Besides observations related to the shock, necropsy revealed multiple SAA deposits in the liver, spleen, kidneys, adrenal glands, thyroid and ovaries.

Molecular data

Whole-genome sequencing of tumor showed a relatively low mutation burden, with 3,132 somatic mutations (1.1 mutations/Mb) including 22 protein-coding mutations (Table S2), 5 copy-number alterations, and 31 structural rearrangements (Table S3). The tumor did not harbor any mutation in known IHCA driver genes involved in the IL6 signaling pathway (IL6ST, JAK1, STAT3, FRK, GNAS), but we identified a complex cluster of structural rearrangements on chromosome 7, including an inversion leading to the truncation of IL6 3'UTR sequence. IL6 mRNA is unstable and degraded with a half-life of 30 minutes.[1] Key regions for the post-transcriptional degradation of IL6 are located in its 3'UTR region and include 6 AU-rich elements and a conserved region predicted to form a stem-loop structure recognized by the RNase Zc3h12a.[2] The inversion breakpoint was located in the 5' of all these features

and we confirmed by RNA-sequencing the massive overexpression (184-fold) of an IL6 transcript lacking all 3' regulatory elements. Interestingly, the acute inflammatory response genes including SAA and CRP were overexpressed in the tumor and non-tumor liver whereas IL6 overexpression was restrained to the tumor area (Figure 1). Expression of SAA and CRP by non-neoplastic hepatocytes are usually not observed in classical IHCA except in the context of tumor hemorrhage and/or chronic inflammatory syndromes of other origin.[3] Altogether, these findings support a paracrine JAK/STAT pathway activation in the liver caused by an autonomous production of IL6 produced by the IHCA (Figure S1).

Discussion

In the present study, we showed a recurrent association between two rare diseases, inflammatory hepatocellular adenoma (prevalence lower than 1/100,000 female) and AA-amyloidosis (incidence around 1/1,000,000 person/y in Europe). This association together with improved amyloidosis related symptoms after tumor resection supports a causal relationship between the two diseases (Figure S1). The role of IHCA in AA-amyloidosis was further elucidated by the identification of a unique somatic chromosome rearrangement at the IL-6 3'UTR leading to an autocrine IL-6 secretion by the tumor. In the liver, IL-6 acts as a ligand to activate STAT3 signaling in normal hepatocytes and the synthesis of the inflammatory acute phase proteins.[4] In the index patient we identified an IL-6 production by the adenoma hepatocytes that induced paracrine effects on the surrounding non-neoplastic hepatocytes in the entire liver. We further showed the massive production of SAA by the tumor and the non-tumor liver tissues in the 2 additional cases of IHCA associated with amyloidosis analyzed by immunohistochemistry. This mechanism is different from the classical IHCAs in which STAT3 is activated in the absence of IL6 expression by a somatic mutation activating one of the major actors of signal transduction: IL6ST, JAK1, GNAS1, FRK or STAT3 itself.

In the patients with IHCA and amyloidosis, the total amount of SAA proteins produced by the whole liver is probably higher than SAA produced by a classical IHCA. The amount and duration of chronic SAA production by the tumor and non-tumor liver tissues are probably major sources of amyloidosis deposits in the body. This hypothesis is strongly supported by the clinical stabilization or resolution of the amyloidosis symptoms observed in the majority of the patients after IHCA resection. However, in our index patient, the kidneys were already severely injured at the time of surgery and renal function did not improve, underscoring the importance of an early diagnosis and a complete resection of the tumor.

In conclusion, rare cases of IHCA characterized by an autonomous IL-6 expression are able to induce massive SAA production by the liver, leading to systemic AA amyloidosis. This subset of AA-amyloidosis may be treatable by the resection of the liver tumor. Antibodies blocking IL-6, which have been approved in other conditions characterized by IL6 overproduction such as Castleman disease, may also be of interest in unresectable IHCA associated with amyloidosis.[5]

Author contributions

Study concept and design: JC, EL, JCN, MM and J.Z.R.
Acquisition of data: JC, EL, QB, AB, VR, JFD, OB, DF, ESZ, JCN, MM, and J.Z.R.
Analysis and interpretation of data: JC, EL, QB, AB, VR, JFD, ESZ, JCN, MM, and J.Z.R.
Drafting the manuscript: JC, EL, JCN and J.Z.R.
Critical revision of the manuscript: JC, EL, QB, AB, VR, JFD, OB, DF, ESZ, JCN, MM, and J.Z.R.
Statistical analysis: E.L. and QB
Obtained funding: J.Z.R.

Supplementary material and methods

Patients

The index patient has signed an inform consent for molecular and genetic research investigations. Frozen liver tumor and non-tumor tissues were analyzed at the Inserm U1162 laboratory (Paris, France), as part of a research project approved by Paris Saint-Louis and Inserm Institutional Review Board committees (Paris Saint-Louis, 2004; INSERM IRB 2010).

Histological and immunohistochemical analyses

Slides stained with Hematein-Eosin-Saffron and Congo Red were reviewed by two expert pathologists (JC and ESZ), using an Olympus BX51 microscope. Amyloid deposits were identified using Congo Red staining and examination under polarized light.
For immunohistochemistry, 4 µm thick sections were cut from paraffin embedded blocks. Slides were processed on an automated autostainer (Leica Bondmax). After deparaffinisation and rehydratation, sections were placed into a boiling target retrieval solution (Leica Epitope Retrieval Solution 1 for CRP and Leica Epitope Retrieval Solution 2 for SAA). Endogenous peroxidase was blocked with Peroxide Block (Leica). Sections were further incubated with primary anti-bodies against CRP (Abcam, Clone Y284, AB32412, dilution 1/4000) or SAA (Dako, Clone MC1, M0759, dilution 1/2000).
After incubation with a polymer system (Leica Polymer Refine Detection), and staining with 3,3' diaminobenzidine (Dako) as chromogen, slides were counterstained with Mayer's haematoxylin, dehydrated and coverslipped.

Sequencing and bioinformatic analyses

Whole genome sequencing

Whole-genome-sequencing was performed by the Centre National de Recherche en Génomique Humaine (Evry, France) on an Illumina HiSeqX5 platform as paired-end 150 bp reads. RNA-sequencing was performed by Integragen (Evry, France) on an Illumina HiSeq2000 as paired-end 100bp reads. WGS and RNAseq were analyzed as previously described.[6] More details on sequencing and bioinformatics analyses are provided in the Supplementary Appendix. Raw WGS and RNA-seq data have been deposited in the EGA (European genome-phenome archive, http://www.ebi.ac.uk/ ega/) database, accessions numbers are pending).

We extracted DNA using a salting-out procedure. Genomic DNA was loaded on a 0.8% agarose gel for quality control; only DNA >10 kb in size was selected.[7] DNA quantification was performed using Hoechst 33258 from Sigma Chemical. Whole genome sequencing (WGS) was performed by the Centre National de Recherche en Génomique Humaine (Institut de biologie François Jacob, CEA). WGS libraries were prepared using the Illumina TruSeq DNA PCR-Free Library Preparation Kit according to the manufacturer's instructions, and the tumor and matched normal liver were sequenced on an Illumina HiSeqX5 platform as paired-end 150 bp reads. Sequences were aligned to the hg19 version of the human genome using BWA version 0.7.12.[8] We used Picard tools version 1.108 (http://broadinstitute.github.io/picard/) to remove PCR duplicates and GATK version v3.5 for local indel realignment and base quality recalibration, as recommended in GATK best practices.[9,10] We obtained an average depth of 90-fold for the tumor and 60-fold for the normal sample. We used MuTect2 to call somatic mutations (single nucleotide variants and small insertions and deletions) the tumor sample with its matched non-tumor counterpart and a panel of normal samples from the lab [11].

We selected only mutations with a "PASS" MuTect2 filter and we excluded mutations belonging to the ENCODE Data Analysis Consortium blacklisted regions (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz).
We used cgpBattenberg algorithm for copy-number analysis and MANTA software to identify somatic structural variations from the tumor and normal bam files.[12,13] To keep only the most reliable events, we selected only structural variants (SVs) supported by ≥15 reads (for all SV types), representing ≥10% of reads (for inversions and interchromosomal translocations). Raw sequences have been deposited in the EGA (European genome-phenome archive - http://www.ebi.ac.uk/ega/) database (EGAS00001003025) and variant calls are available on the International Cancer Genome Consortium (ICGC) data portal (http://dcc.icgc.org/).

RNA-sequencing
Index tumor (#2615T) sample and control normal and IHCA mutated for IL6ST were analyzed by RNA-sequencing. Libraries were prepared by Integragen (Evry, France) using Illumina TruSeq Stranded mRNA kit according to the manufacturer's instructions, and sequenced as paired-end 100bp reads on an Illumina HiSeq2000. Fastq files were aligned to the reference human genome hg19/GRCh37 with TopHat2 (-p 24 -r 150 -g 2 --library-type fr-firststrand).[14] After removing reads mapping to multiple locations, we used HTSeq to obtain the number of reads associated to each gene in the GENCODE v24lift37 database (restricted to protein-coding genes, antisense and lincRNAs).[15] We used the Bioconductor DESeq package to import raw HTSeq counts for each sample into R statistical software and extract the count matrix.[16] After normalizing for library size, we normalized the count matrix by the coding length of genes to compute FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads). We used Integrative Genomics Viewer (IGV) to visualize the IL6 transcripts and generate Sashimi plots.[17] Raw sequences have been deposited in the EGA (European genome-phenome archive - http://www.ebi.ac.uk/ega/) database (accessions numebrs are pending).

Table S1. Blood and urine biological analyses of the patient.

	Patient values at diagnosis	*Normal values*
Hemoglobin	10 g/dL	11.7-15.0
Mean corpuscular volume	69.8 fL	83.7-100.8
Platelet count	533,000/mm^3	150,000-353,000
Leukocytes (total)	12,900/ mm^3	4,600-10,100
Neutrophils	10,060/ mm^3	2,200-6,100
Lymphocytes	1,650/ mm^3	1,100-3,700
Fibrinogen	8.02 g/L	2.30-4.30
CRP	145.3 mg/L	0.0-6.0
SAA	153.0 mg/L	0.0-6.4
Procalcitonin	0.43 mg/L	0.00-0.50
Urea	0.97 g/L	0.15-0.53
Creatinine	41.7 mg/L	5.5-10.2
Alkaline phophatase	493 U/L	34-117
GGT	265 U/L	05-50
ASAT	29 U/L	14-40
ALAT	41 U/L	06-40
LDH	190 U/L	<250
Protein (total)	62 g/L	66-83
Albumin	26 g/L	38-49
Protein (urine)	7,481 mg/L	0-150
Alpha-1-microglobulin (urine)	298.0 mg/L	<12
IL-2	8.96 pg/mL	0.79-21.7
IL-6	186.64 pg/mL	0.62-8.0
IL-7	1.74 pg/mL	0.40-3.2
IL-10	33.65 pg/mL	0.0-4.3
IL-17A	5.46 pg/mL	0.0-15.1
sCD154	16.39 pg/mL	10.9-286.4
IFN gamma	0 pg/mL	0.0-2.7

Table S2: Coding somatic mutations identified in the tumor of the index patient (#2615T)

Chromosome	*Position*	*Reference base*	*Variant base*	*Hugo Symbol*	*Variant Classification*	*Genome Change*	*cDNA Change*	*Protein Change*
chr1	180022207	G	T	CEP350	Missense_Mutation	g.chr1:180022207G>T	c.4895G>T	p.S1632I
chr1	236979804	A	T	MTR	Missense_Mutation	g.chr1:236979804A>T	c.725A>T	p.E242V
chr10	134659633	C	A	TTC40	Missense_Mutation	g.chr10:134659633C>A	c.6366G>T	p.Q2122H
chr11	113103452	T	C	NCAM1	Missense_Mutation	g.chr11:113103452T>C	c.1054T>C	p.S352P
chr14	77605915	A	T	ZDHHC22	Missense_Mutation	g.chr14:77605915A>T	c.167T>A	p.F56Y
chr17	5462679	T	G	NLRP1	Missense_Mutation	g.chr17:5462679T>G	c.1337A>C	p.K446T
chr17	38572320	A	G	TOP2A	Splice_Site	g.chr17:38572320A>G	c.269T>C	p.V90A
chr19	55954202	A	T	SHISA7	Missense_Mutation	g.chr19:55954202A>T	c.29T>A	p.L10Q
chr20	37174999	A	G	RALGAPB	Missense_Mutation	g.chr20:37174999A>G	c.2816A>G	p.H939R
chr3	123167254	C	CG	ADCY5	Frame_Shift_Ins	g.chr3:123167254_123167254insCG	c.139_139insCG	p.-46fs
chr4	90872904	T	A	MMRN1	Splice_Site	g.chr4:90872904T>A		
chr4	94690579	G	T	GRID2	Missense_Mutation	g.chr4:94690579G>T	c.2579G>T	p.G860V
chr6	3123976	C	A	BPHL	Missense_Mutation	g.chr6:3123976C>A	c.142C>A	p.L48I
chr6	30673359	T	G	MDC1	Missense_Mutation	g.chr6:30673359T>G	c.3601A>C	p.T1201P
chr6	32007362	G	C	CYP21A2	Missense_Mutation	g.chr6:32007362G>C	c.589G>C	p.E197Q
chr6	56472969	A	T	DST	Missense_Mutation	g.chr6:56472969A>T	c.5824T>A	p.L1942I
chr7	29918696	G	T	WIPF3	Nonsense_Mutation	g.chr7:29918696G>T	c.295G>T	p.G99*
chr7	73254374	AT	A	WBSCR27	Frame_Shift_Del	g.chr7:73254374delAT	c.465delAT	p.Q155fs
chr8	26722021	C	T	ADRA1A	Missense_Mutation	g.chr8:26722021C>T	c.466G>A	p.V156I
chr8	38205493	A	C	WHSC1L1	Missense_Mutation	g.chr8:38205493A>C	c.197T>G	p.L66R
chrX	99596960	G	A	PCDH19	Missense_Mutation	g.chrX:99596960G>A	c.2789C>T	p.A930V
chrX	142716743	C	T	SLITRK4	Missense_Mutation	g.chrX:142716743C>T	c.2182G>A	p.E728K

Table S3. Somatic structural rearrangements identified in the tumor of the index patient (#2615)

Rearrangement Type	*Breakpoint 1 Chromosome*	*Breakpoint 1 Position*	*Breakpoint 1 Category*	*Breakpoint 1 Gene*	*Breakpoint 2 Chromosome*	*Breakpoint 2 Position*	*Breakpoint 2 Category*	*Breakpoint 2 Gene*
Interchromosomal Translocation	chr1	243119768	Intergenic Region	NA	chrY	13977276	Intergenic Region	NA
Inversion	chr11	44572105	Intergenic Region	NA	chr11	116683145	RNA	AP006216.5
Interchromosomal Translocation	chr12	116911744	Intergenic Region	NA	chr18	39532046	Intergenic Region	NA
Inversion	chr15	100338220	RNA	DNM1P46	chr15	102306887	RNA	DNM1P47
Deletion	chr15	26879870	Intron	GABRB3	chr15	48647047	Intergenic Region	NA
Interchromosomal Translocation	chr16	88439347	Intergenic Region	NA	chr17	43059782	lincRNA	CTD-2534I21.9
Interchromosomal Translocation	chr17	70083524	RNA	AC005152.2	chrX	8451963	Intergenic Region	NA
Interchromosomal Translocation	chr2	18817184	Intergenic Region	NA	chr20	22308518	Intergenic Region	NA
Deletion	chr3	69767360	Intergenic Region	NA	chr3	69768630	Intergenic Region	NA
Interchromosomal Translocation	chr3	69767777	Intergenic Region	NA	chr7	36903921	Intron	ELMO1
Interchromosomal Translocation	chr3	69768559	Intergenic Region	NA	chr7	41484105	Intergenic Region	NA
Inversion	chr4	109747252	Intron	COL25A1	chr4	110927627	Intron	EGF
Interchromosomal Translocation	chr4	109747817	Intron	COL25A1	chr14	69289448	Intergenic Region	NA
Inversion	chr4	109749159	Intron	COL25A1	chr4	109749299	Intron	COL25A1
Deletion	chr4	109749231	Intron	COL25A1	chr4	110921447	Intron	EGF
Deletion	chr4	109749327	Intron	COL25A1	chr4	110960547	Intergenic Region	NA
Interchromosomal Translocation	chr4	110921542	Intron	EGF	chr14	69289392	Intergenic Region	NA
Inversion	chr4	110982418	Intron	ELOVL6	chr4	111017388	Intron	ELOVL6
Interchromosomal Translocation	chr4	110984500	Intron	ELOVL6	chr7	21930994	Intron	DNAH11
Interchromosomal Translocation	chr4	111018301	Intron	ELOVL6	chr7	22662350	RNA	AC002480.5
Deletion	chr4	111226035	Intergenic Region	NA	chr4	111226721	Intergenic Region	NA
Interchromosomal Translocation	chr4	185597507	Intron	PRIMPOL	chr22	36449794	Intergenic Region	NA
Deletion	chr4	91931666	Intron	CCSER1	chr4	91933525	Intron	CCSER1
Interchromosomal Translocation	chr6	154638263	Intron	IPCEF1	chr10	53209375	Intron	PRKG1
Deletion	chr6	32458343	Intergenic Region	NA	chr6	32504751	Intergenic Region	NA
Interchromosomal Translocation	chr7	128151688	Intergenic Region	NA	chr17	60535882	5'Flank	TLK2

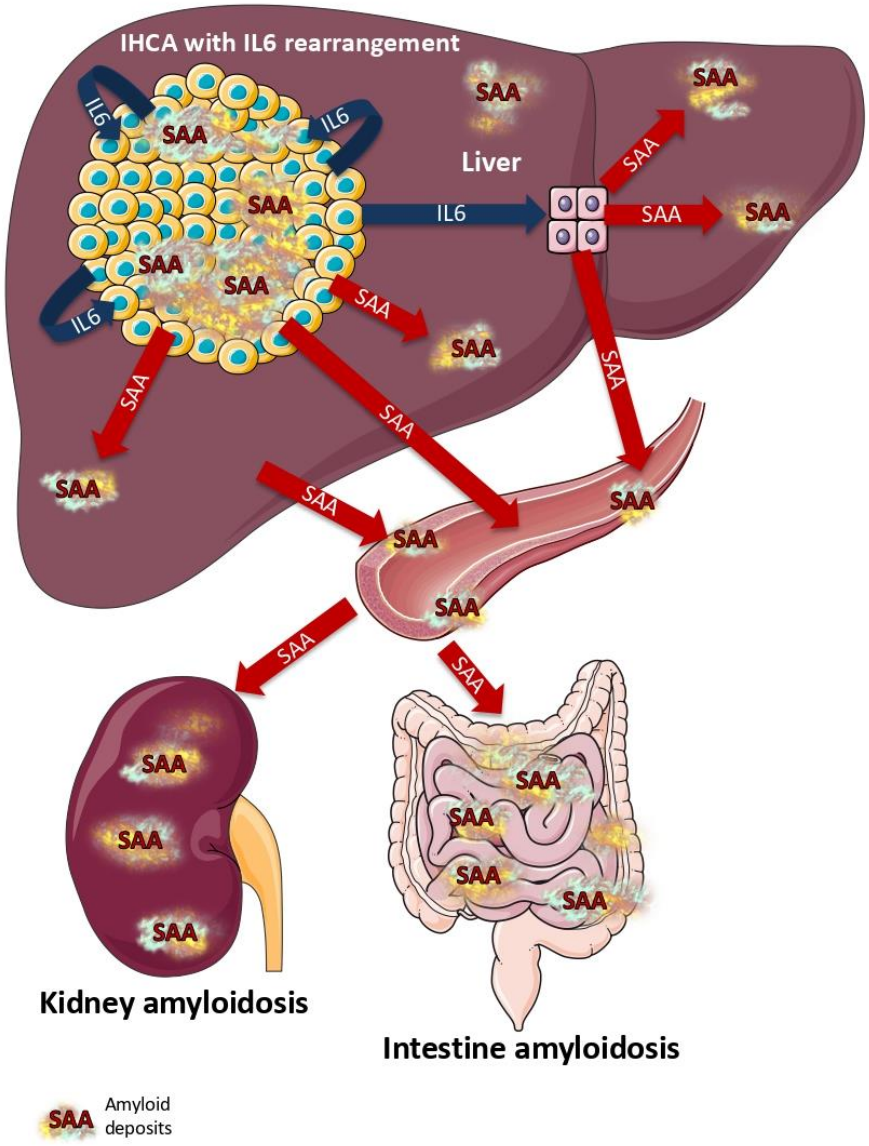
Deletion	chr7	16599884	Intron	LRRC72	chr7	16600278	Intron	LRRC72
Deletion	chr7	16626595	Intergenic Region	NA	chr7	16627762	Intergenic Region	NA
Deletion	chr7	21019896	lincRNA	AC006481.1	chr7	39390092	Intron	POU6F2
Tandem Duplication	chr7	22666192	RNA	AC002480.5	chr7	36894098	3'UTR	ELMO1
Inversion	chr7	22771257	3'UTR	IL6	chr7	41479923	Intergenic Region	NA
Tandem Duplication	chr7	22979435	Intergenic Region	FAM126A	chr7	37030083	Intron	ELMO1
Inversion	chr7	36901703	Intron	ELMO1	chr7	38573324	Intron	AMPH
Inversion	chr7	36903099	Intron	ELMO1	chr7	38573032	Intron	AMPH
Inversion	chr7	37031598	Intron	ELMO1	chr7	41480304	Intergenic Region	NA
Deletion	chr7	39389384	Intron	POU6F2	chr7	41484626	Intergenic Region	NA
Deletion	chrX	6355883	Intergenic Region	NA	chrX	6396739	Intergenic Region	NA

10

Table S4. Clinical and biological features of patients with AA amyloidosis and hepatocellular adenoma. [18-25]

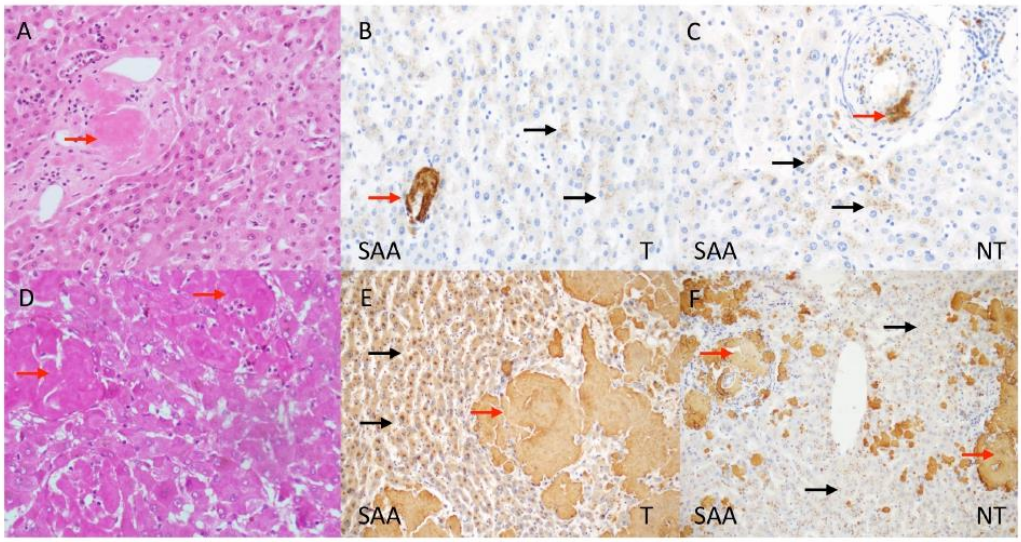
Patient ID	Clinical features of the patient	Symptoms at diagnosis	Clinical follow-up	Tumor characteristics/ Pathology	Reference
Index case #2615	49 year-old female, thyrotoxicosis, hypertension	Renal insufficiency, proteinuria, biological inflammatory syndrome, renal and digestive amyloidosis	Regression of digestive symptoms and decrease of the peripheral inflammatory syndrome after resection	5cm IHCA, Amyloid deposits and SAA expression in both the tumor and non-tumor liver tissues	Index
1	56 year-old male, hypertension, hyper-cholesterolemia, alcohol abuse	Renal insufficiency, proteinuria, biological inflammatory syndrome, hepatic, renal and digestive amyloidosis	Renal function stabilization and decrease of the peripheral inflammatory syndrome after resection	11cm IHCA, Amyloid deposits and SAA expression in both the tumor and non-tumor liver tissues	Fievet et al., Gut 1990
2	34 year-old female, oral contraception, hyperlipidemia	Nephrotic syndrome; hepatic, renal and rectal amyloidosis	Acute renal failure, favorable evolution after kidney transplantation and resection	IHCA, Amyloid deposits and SAA expression in both the tumor and non-tumor liver tissues	Bestard Matamoros et al., Nefrologia 2008
3	37 year old female, oral contraception, Spain	Nephrotic syndrome, intra-tumor bleeding	Regression of nephrotic syndrome and HCA after oral contraception withdrawal	8cm HCA, amyloid deposits in the non-tumor liver and the kidney	Cosme et al. Liver, 1995
4	32 year old male, Japan	Renal, heart and digestive amyloidosis	Regression of nephrotic syndrome and faborable outcome after surgical resection	SAA negative HCA with SAA expression in the non tumor liver, amyloid deposits in the tumor	Shibasaki et al., 1997
5	39 year old female, oral contraception, Sweden	Nephrotic syndrome	Regression of nephrotic syndrome	15cm HCA with histological features consistent with IHCA, lack of SAA expression by immunofluorescence, amyloid deposits in the tumor and non-tumor liver tissues, and kidney	Thysell et al. J Hepatol, 1986
6	28 year old female, oral contraception, France	Proteinuria and biological inflammatory syndrome	Regression of nephrotic and inflammatory syndrome after surgical resection	HCA, amyloid deposits in the HCA, non-tumor liver and the kidney.	Melin et al., Nephron, 1993
7	49 year old male	Non-resectable HCA at 32 years. Transformation in HCC associated with amyloidosis and nephrotic syndrome at 49 years	Death from HCC	HCA transformed in HCC, amyloid deposits in the liver and kidney.	Delgado et al. J LA State Med Soc, 1999
8	40 year old male, Glycogen storage disease type 1b, hypertension	Severe renal dysfunction	-	Multiple adenomas (no biopsy or resection), Amyloidosis in the liver, kidney and spleen	Dick et al., Clin Kidney J, 2012

Figure S1. Schematic representation of the biological mechanisms leading to systemic AA amyloidosis in the index patient.



12

Figure S2. Histological findings of two additional cases of IHCA associated with amyloidosis.



Microscopic examination of case n°2 (Bestard Matamoros et al., Nefrologia, 2008) showed a well-differentiated tumor with amyloid deposits (Panel A, red arrows, Hematein-Eosin-Saffron, X400). Immunohistochemical experiments demonstrated SAA expression in both neoplastic and non neoplastic hepatocytes, along with stained amyloid deposits (Panel B: tumor, Panel C: adjacent parenchyma, black arrows: hepatocytes, red arrows: amyloid deposits, X400). Case n°3 (Fievet et al., Gut 1990) displayed massive amyloid deposits (Panel D, red arrows, Hematein-Eosin-Saffron, X400). As observed for case n°2, SAA was expressed in the tumor and the non-tumoral liver (Panel E: tumor, Panel F: adjacent parenchyma, black arrows: hepatocytes, red arrows: amyloid deposits, X200). A strong staining was observed in the deposits (Panel E and F, red arrows, X200). T: tumor; NT: non-tumor liver.

References

1. Paschoud S, Dogar AM, Kuntz C, Grisoni-Neupert B, Richman L, Kuhn LC. Destabilization of interleukin-6 mRNA requires a putative RNA stem-loop structure, an AU-rich element, and the RNA-binding protein AUF1. Molecular and cellular biology 2006;26:8228-41.
2. Matsushita K, Takeuchi O, Standley DM, et al. Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. Nature 2009;458:1185-90.
3. Sa Cunha A, Blanc JF, Lazaro E, et al. Inflammatory syndrome with liver adenomatosis: the beneficial effects of surgical management. Gut 2007;56:307-9.
4. Heinrich PC, Castell JV, Andus T. Interleukin-6 and the acute phase response. The Biochemical journal 1990;265:621-36.
5. Garbers C, Heink S, Korn T, Rose-John S. Interleukin-6: designing specific therapeutics for a complex cytokine. Nat Rev Drug Discov 2018.
6. Letouze E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nature communications 2017;8:1315.
7. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic acids research 1988;16:1215.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
9. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 2010;20:1297-303.
10. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics 2011;43:491-8.
11. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology 2013;31:213-9.
12. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. Cell 2012;149:994-1007.
13. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32:1220-2.
14. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 2013;14:R36.
15. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-9.
16. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology 2010;11:R106.
17. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nature biotechnology 2011;29:24-6.
18. Fievet P, Sevestre H, Boudjelal M, et al. Systemic AA amyloidosis induced by liver cell adenoma. Gut 1990;31:361-3.
19. Bestard Matamoros O, Poveda Monje R, Ibernon Vilaro M, Carrera Plans M, Grinyo Boira JM. [Systemic AA amyloidosis induced by benign neoplasms]. Nefrologia 2008;28:93-8.
20. Cosme A, Horcajada JP, Vidaur F, Ojeda E, Torrado J, Arenas JI. Systemic AA amyloidosis induced by oral contraceptive-associated hepatocellular adenoma: a 13-year follow up. Liver 1995;15:164-7.
21. Shibasaki T, Matsumoto H, Watabe K, et al. A case of renal amyloidosis associated with hepatic adenoma: the pathogenetic role of tumor necrosis factor-alpha. Nephron 1997;75:350-3.
22. Thysell H, Ingvar C, Gustafson T, Holmin T. Systemic reactive amyloidosis caused by hepatocellular adenoma. A case report. Journal of hepatology 1986;2:450-7.
23. Delgado MA, Liu JY, Meleg-Smith S. Systemic amyloidosis associated with hepatocellular carcinoma. Case report and literature review. J La State Med Soc 1999;151:474-8.

24. Dick J, Kumar N, Horsfield C, Jayawardene S. AA Amyloidosis in a patient with glycogen storage disorder and progressive chronic kidney disease. Clinical kidney journal 2012;5:559-61.
25. Melin JP, Lecrique T, Wynckel A, et al. Remission of nephrotic syndrome related to AA amyloidosis after hepatic adenomectomy. Nephron 1993;64:157-8.

Article 2: Recurrent chromosomal rearrangements of ROS1, FRK and IL6 activating JAK/STAT pathway in inflammatory hepatocellular adenomas.

Hepatology

ORIGINAL RESEARCH

Recurrent chromosomal rearrangements of *ROS1*, *FRK* and *IL6* activating JAK/STAT pathway in inflammatory hepatocellular adenomas

Quentin Bayard,[1] Stefano Caruso,[1] Gabrielle Couchy,[1] Sandra Rebouissou,[1] Paulette Bioulac Sage,[2,3] Charles Balabaud,[3] Valerie Paradis,[4,5] Nathalie Sturm,[6] Anne de Muret,[7] Catherine Guettier,[8] Benjamin Bonsang,[9] Christiane Copie,[9] Eric Letouzé,[1] Julien Calderaro,[9] Sandrine Imbeaud ,[1] Jean-Charles Nault ,[1,10,11] Jessica Zucman-Rossi[1,12]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2019-319790).

For numbered affiliations see end of article.

Correspondence to
Dr Jean-Charles Nault, INSERM UMR 1338, INSERM, Paris 93140, France;
naultjc@gmail.com
Professor Jessica Zucman-Rossi;
jessica.zucman-rossi@inserm.fr

J-CN and JZ-R are joint senior authors.

Received 4 September 2019
Revised 27 November 2019
Accepted 14 December 2019

Check for updates

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Bayard Q, Caruso S, Couchy G, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2019-319790

ABSTRACT

Background Inflammatory hepatocellular adenomas (IHCAs) are benign liver tumours characterised by an activation of the janus kinase (JAK)/signal transducers and activators of transcription (STAT) pathway caused by oncogenic activating mutations. However, a subset of IHCA lacks of identified mutation explaining the inflammatory phenotype.

Methods 657 hepatocellular adenomas developed in 504 patients were analysed for gene expression of 17 genes and for mutations in seven genes by sequencing. 22 non-mutated IHCAs were analysed by whole-exome and/or RNA sequencing.

Results We identified 296 IHCA (45%), 81% of them were mutated in either *IL6ST* (61%), *FRK* (8%), *STAT3* (5%), *GNAS* (3%) or *JAK1* (2%). Among non-mutated IHCA, RNA sequencing identified recurrent chromosome rearrangement involving *ROS1*, *FRK* or *IL6* genes. *ROS1* fusions were identified in 8 IHCA, involving C-terminal part of genes highly expressed in the liver (*PLG*, *RBP4*, *APOB*) fused with exon 33–35 to 43 of *ROS1* including the tyrosine kinase domain. In two cases a truncated *ROS1* transcript from exon 36 to 43 was identified. *ROS1* rearrangements were validated by fluorescence in situ hybridisation (FISH) and led to *ROS1* overexpression. Among the 5 IHCA with *FRK* rearrangements, 5 different partners were identified (*MIA3*, *MIA2*, *LMO7*, *PLEKHA5*, *SEC16B*) fused to a common region in *FRK* that included exon 3–8. No overexpression of *FRK* transcript was detected but the predicted chimeric proteins lacked the auto-inhibitory SH2–SH3 domains. In two IHCA, we identified truncated 3'UTR of *IL6* associated with overexpression of the transcript.

Conclusion Recurrent chromosomal alterations involving *ROS1*, *FRK* or *IL6* genes lead to activation of the JAK/STAT pathway in IHCAs.

Significance of this study

What is already known on this subject?
- Inflammatory hepatocellular adenomas (IHCAs) are benign liver tumours characterised by an activation of the JAK/STAT pathway caused by oncogenic activating mutations.
- A subset of IHCA lacks of identified mutation explaining the inflammatory phenotype.

What are the new findings?
- We identified recurrent chromosomal alterations involving *ROS1*, *FRK* or *IL6* in IHCAs.
- *ROS1* activation is a novel mechanism of JAK/STAT pathway activation in human benign liver tumours.

How might it impact on clinical practice in the foreseeable future?
- These new genetic alterations could be potential diagnostic biomarkers of a subtype of hepatocellular adenoma.
- Targeted therapies adapted to genetic alterations could be tested in selected patients with multiple or unresectable IHCAs.

INTRODUCTION

Hepatocellular adenomas (HCAs) are rare benign tumours mainly fostered by exposure to oestrogen, particularly in women taking oral contraceptives.[1] Management of hepatocellular adenoma requires assessment of the risk of complications mainly related to tumour bleeding and malignant transformation to hepatocellular carcinoma (HCC).[2]

Genotype/phenotype classification of HCA has dissected adenomas in homogeneous subgroups characterised by somatic genetic alterations dysregulating signalling pathways and linked with specific risk factors, clinical behaviours, histological and imaging features.[3] The first subgroups of HCA (HHCA, 30%–40%) are defined by bi-allelic inactivating mutations of *HNF1A* and are associated with tumour steatosis, lipogenesis and mammalian target of rapamycin (mTOR) activation.[4 5] Another group of HCA is defined by mutations in *CTNNB1*, coding for ß-catenin, either in exon 3 (b^{ex3}HCA, around 15% of HCA) or in exon 7 or 8 ($b^{ex7,8}$HCA, around 10% of HCA).[3 6] Mutations in *CTNNB1* exon 3 are associated with a high risk of malignant transformation in HCC.[3] A new subgroup associated

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

with a high risk of tumour bleeding was recently described in 4%–5% of the patients (sonic hedgegog hepatocellular adenoma (shHCA)); it is defined by an activation of the sonic hedgehog pathway due to a *GLI1* overexpression caused by a somatic fusion with *INHBE*, its neighbouring gene.[7]

The most prevalent subgroup of HCA are inflammatory tumours (inflammatory hepatocellular adenomas (IHCAs), around 50% of HCA), characterised by the constitutive and uncontrolled activation of interleukin 6 (IL-6)/JAK/STAT pathway easily detectable by the overexpression of serum amyloid A (SAA) and C-reactive protein (CRP) at the transcriptomic level, associated with inflammatory infiltrate, sinusoidal dilatation and dystrophic arteries at histological level.[8] Overexpression of SAA and CRP in the tumour on immunohistochemistry helps to identify this subtype in routine. Interestingly, a subset of IHCA is mixed, harbouring mutation of *CTNNB1* either in exon 3 or exon 7/8 (bex^{3}IHCA, bex^{7,8}IHCA). Risk factors for development of IHCA are high alcohol consumption and obesity.[9] Several genetic alterations are responsible for the constitutive activation of JAK/STAT pathway in IHCA. The most frequent mutation are in *IL6ST* (coding for the gp130 receptor), *STAT3* (coding for a transcription factor of the JAK/STAT pathway), *FRK* (coding for a src kinase), *JAK1* (coding for one of the protein tyrosine kinase of the JAK/STAT pathway) and *GNAS* (coding stimulatory G-protein alpha subunit).[6 10 11] Inflammatory syndrome, anaemia or fever could be observed in patients with IHCA; these symptoms regress after tumour resection. Recently, we described an IHCA with a rearrangement of the 3'UTR of the *IL6* gene leading to a secondary amyloidosis through an autocrine/paracrine mechanism of activation of JAK/STAT pathway by IL-6 in the whole liver.[12] However, a subset of IHCA has currently no known genetic alterations explaining the constitutive activation of JAK/STAT pathway.

The aim of our study was to unravel new mechanisms of activation of inflammatory pathway in inflammatory adenomas without any mutation in known driver genes to better understand the mechanism of regulation of IL-6/JAK/STAT pathway in hepatocytes and its role in benign liver tumourigenesis.

MATERIAL AND METHODS

Selection of tumours and patients

Frozen samples from 657 HCA developed in 504 patients were analysed. This series included 124 new HCA cases apart from previously analysed HCA cases by Nault *et al*[7] . Samples were collected in 28 centres in France from 2000 to 2017. All patients gave their informed consent according to French law. All samples were frozen in nitrogen immediately after resection or biopsy and conserved at −80°C.

DNA sequencing

HNF1A (exon 1–10), *IL6ST* (exons 6 and 10), *CTNNB1* (exons 2, 3, 4, 7 and 8), *FRK* (exon 6), *STAT3* (exons 3, 6, 17 and 21), *GNAS* (exons 7, 8 and 9) and *JAK1* (all exons) were sequenced by either Sanger sequencing or Miseq Illumina PCR-based sequencing (for detailed protocol see [6 13 14]). Somatic mutations were confirmed by sequencing of the tumour and non-tumour counterpart.

Quantitative RT-PCR

Gene expression of a selection of 17 genes was assessed by quantitative reverse transcriptase PCR (RT-PCR) using Fluidigm technology with Taqman probes detailed in online supplementary table 1, as previously described.[7 13] Data were normalised using RNA ribosomal 18S as calibrator. We used the 2 delta delta CT method with the level of the corresponding gene expression in tumour tissues expressed as an n-fold ratio compared with the mean level of expression in five normal livers.

As previously described,[7] HHCAs were defined by *HNF1A* bi-allelic mutations and/or downregulation of *LFABP* and *UGT2B7* (T/N<0.2); IHCAs were defined by SAA and/or CRP overexpression T/N>5 and/or activating mutations in *IL6ST*, *FRK*, *JAK1*, *STAT3*, *GNAS*. b^{ex3}HCAs were defined by mutations or in frame deletion of *CTNNB1* exon 3, b^{ex7,8}HCAs were defined by mutations of *CTNNB1* in exon 7 or 8, bHCAs were defined by overexpression of Wnt/catenin pathway target genes *GLUL* or *LGR5* (T/N>5) without identified gene mutation. b^{ex3}IHCAs were identified by the combination of b^{ex3}HCA and IHCA criteria, bIHCA by the combination of bHCA and IHCA criteria and b^{ex7,8}IHCA by the combination of b^{ex7,8}HCA and IHCA criteria. shHCA was defined by overexpression of the following target genes of sonic hedgehog pathway: *GLI1*, *TNNC1*, *FCRLA*, *GPR97*, *HHIP*, *PTCH1* and *PTGDS*.

Whole exome sequencing

Whole exome data were analysed in 19 HCA cases which included 4 previously published and 15 new cases. Sequence capture, enrichment and elution of genomic DNA samples from the tumour/normal pairs were performed by IntegraGen (Evry). Agilent in-solution enrichment was used with the manufacturer's biotinylated oligonucleotide probe library SureSelect Human All-Exon kit V.4 70Mb (n=4), SureSelect Human All-Exon kit V.5 +UTRs (n=8), or Twist Human Core Exome Enrichment System (n=7). Briefly, 3 µg of each genomic DNA was sonicated and purified to yield fragments of 150–200 bp. Adaptor oligonucleotides were ligated on A-tailed fragments and enriched by four to six PCR cycles. Purified libraries (500 ng) were hybridised to the SureSelect library for 24 hours. The eluted enriched DNA sample was sequenced on an Illumina HiSeq 2000 (n=12) or HiSeq 4000 (n=7) as paired-end 75 bp reads. Sequencing details for each sample are indicated in online supplementary table 2. Variant calling was processed using Mutect2 by comparing each tumour sample with its matched non-tumour counterpart, following the genone analysis toolkit (GATK) best practices.[15 16] The copy number profiles were defined using an in-house pipeline. Briefly, the logR ratio was processed using the coverage of each tumour and its non-tumour counterpart in each 1 kb genomic windows; then a segmentation was done in order to identify copy number alterations.

RNA sequencing

RNA samples from the 22 HCA tumours were sequenced comprising 1 previously published[12] and 21 new cases. RNA samples were enriched for polyadenylated RNA from 5 µg of total RNA, and the enriched samples were used to generate sequencing libraries with the NEBNext Ultra II Directional RNA (n=13) or Illumina TruSeq Stranded messenger RNA (mRNA) (n=9) kits and associated protocol as provided by the manufacturer. Libraries were sequenced by IntegraGen on an Illumina HiSeq 2000 or 4000 as paired-end 101 bp or 75 bp reads (n=8 and n=14, respectively). Full Fastq files were aligned to the reference human genome GRCh38 using TopHat2 V.2.0.14.[17] Sequencing details for each sample and the parameters used for TopHat2 are indicated in online supplementary table 2. Read mapping from multiple locations was removed and HTSeq[18] was used to obtain the number of reads associated with each gene in the Gencode V.27 database, restricting to protein-coding genes,

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

pseudogenes, antisense and lincRNAs (n=39947). Bioconductor DESeq2 package[19] was used to import raw HTSeq counts for each sample into R statistical software and apply variance stabilising transformation to the raw count matrix. FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalising the count matrix for the library size estimated with DESeq2 package and the coding length of each gene. In order to handle batch effect, we used an in-house method, according to our previous published method.[20] Briefly, area under the ROC curve (AUC) was used to identify and remove 2696 genes with a significant batch effect (AUC >0.95 between one sequencing project and others).

To perform gene set enrichment analysis (GSEA) analysis, genes from RNA-seq data were sorted according to fold-change from Limma analysis (genes with positive fold-changes first, then negative fold-changes) using an input of 16387 genes (Hallmarks gene set; MSigDB V.6 database used). This gene list was obtained from Genecode database; then genes with a variance equal 0 or with a total number of reads from samples inferior of 50 were removed.

Gene fusion detection

Fusions detected by TopHat2 (--fusion-search --fusion-min-dist 2000 --fusion-anchor-length 13 --fusion-ignore-chromosomes chrM) were filtered using the TopHatFusion[20] pipeline. Fusions validated by BLAST and with at least 10 split-reads or pairs of reads spanning the fusion event were retained, whereas any fusion identified at least twice in a cohort of 36 normal liver samples was removed. For each fusions of interest, the sequence corresponding to chimaera was generated and used as a reference to realign RNAseq reads. Visual inspection of chimaera alignments on Integrative Genome Viewer (broad institute) was used to refine junction positions and determine in phase chimeric transcripts. Oncofuse V.1.0.9b2 was finally used to annotate fusions of interest.[21]

FISH experiment

Whole-tissue sections from formalin-fixed, paraffin-embeded surgical samples of HCA were deparaffinised in xylene and washed in ethanol. After rehydratation, they were subjected to heat pre-treatment (Pre-Treatment Solution, DakoCytomation, 10 min, 75°C). After digestion with pepsin (10 min, DakoCytomation) and airdrying, the SPEC ROS1 dual colour break apart probe (Zytovision Z-2144) was applied to the tissue sections. Probe mixture and tissue sections were further codenatured and hybridised overnight (Dako Hybridizer). After washing, dehydration and drying, tissue sections were finally counterstained using fluorescence mounting media containing de-amidino-phenyl-indole.

Statistical analysis

Statistical analysis was performed using the GraphPad Prism and the R V.3.4.2 software. Comparison between qualitative data was performed using Fisher's exact test. Quantitative data in two different group were compared using the non-parametric Mann-Whitney test. All tests were two-tailed and a p value<0.05 was considered as significant.

Data availability

The aligned sequencing data reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (RNA-seq accession EGAS00001003025 and EGAS00001003685; WES accessions EGAS00001000679 and EGAS00001003686).

Table 1 Description of the tumours

Variables	Number (percentages) n=657
HCA	605 (92%)
Borderline lesions (HCA/HCC)	39 (6%)
HCC on HCA	13 (2%)
H HCA	226 (34%)
I HCA	222 (34%)
bex ^{7,8} IHCA	21 (3%)
bex^{3} IHCA	44 (7%)
b IHCA	9 (1%)
bex ^{7,8} HCA	19 (3%)
bex^{3} HCA	52 (8%)
b HCA	4 (1%)
Sh HCA	34 (5%)
U HCA	26 (4%)

b IHCA, IHCA with activation of B-catenin pathway without mutations. b HCA, HCA with activation of B-catenin pathway without mutations ;bex^{3} IHCA, IHCA with mutations of *CTNNB1* in exon 3; bex^{7,8} IHCA, IHCA with mutations of *CTNNB1* in exon 7 or 8;bex^{3} HCA, HCA with mutations of *CTNNB1* in exon 3;bex^{7,8} HCA, HCA with mutations of *CTNNB1* in exon 7 or 8; H HCA, *HNF1A* inactivated HCA;HCA, hepatocellular adenoma; HCC on HCA, hepatocellular carcinoma developed on hepatocellular adenoma;HCC, hepatocellular carcinoma; IHCA, inflammatory hepatocellular adenoma; Sh HCA, sonic hedgehog HCA; U HCA, unclassified HCA;

RESULTS

Clinical features of the cohort of patients

657 HCA developed in 504 patients with a median age of 38 years and comprising of 85% of women with 87% intake of oral contraceptive pills were included in the study (see flow chart, online supplementary figure 1).

Forty-four patients harboured liver adenomatosis (features of the series are described in table 1). 296 HCA (45%) were IHCA with an overexpression of *SAA* and *CRP* at the transcriptomic level and/or a mutation in the JAK/STAT signalling pathway (see material and methods). 76% of patients with IHCA showed an increase in serum CRP. Patients with IHCA were older (median age 40 years old vs 36 years old, p value=0.039, Mann-Whitney test), with a longer cumulative uses of oral contraception (median 19 years vs 11 years, p value<0.0001, Mann-Whitney test), a higher body mass index (median 25 kg/m^2 vs 22 kg/m^2, p value=0.003, Mann-Whitney test) and more frequent chronic alcohol intake (24% vs 9%, value=0.002, Fisher's exact test) compared with patients harbouring other subgroups of HCA.

Spectrum of mutations in known oncogenes in IHCA

Among the 296 IHCA, 74 were mixed tumours with 21 b^{ex7,8}IHCA (3% of the whole series, defined by mutations of *CTNNB1* in exon 7 or 8) and 53 b^{ex3}IHCA/bIHCA (8% of the whole series) including 44 b^{ex3}IHCA (defined by a mutation of *CTNNB1* in exon 3, 7%) and 9 bIHCA (defined by an overexpression of Wnt/catenin pathway target genes without mutation of *CTNNB1*, 1%). Borderlines lesions between HCA and HCC or HCC developed on HCA were identified in 6% of all IHCA (n=18). Nine borderlines lesions between HCA and HCC or HCC developed on HCA were identified among 53 b^{ex3}IHCA/bIHCA (17%) versus 9 among 243 IHCA without activation of ß-catenin pathway (4%, p value Fisher's exact=0.0013) confirming that activation of ß-catenin pathway is associated

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

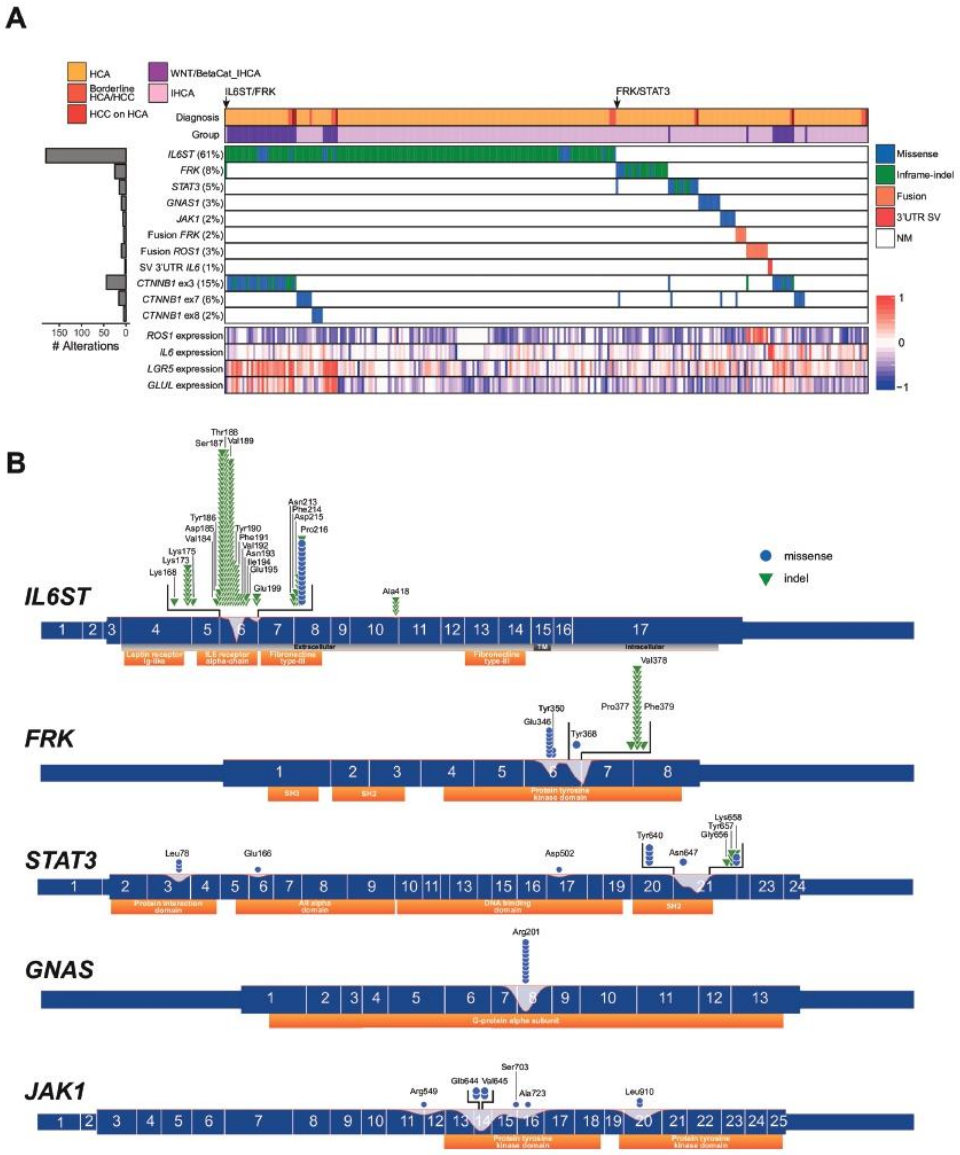


Figure 1 Distribution of mutations in inflammatory hepatocellular adenomas (IHCAs). (A) We represented a heatmap of the somatic mutations and chromosomal rearrangements in driver genes (*IL6ST, FRK, STAT3, GNAS, JAK1, ROS1, IL6* in line) in 296 IHCA (in column) (top). We included the expression level in log2 of *ROS1*, *IL6, GLUL, LGR5,* assessed by quantitative RT-PCR (bottom). (B) Mutational spectrum of *IL6ST*, *FRK*, *STAT3*, *GNAS* and *JAK1*. Red curves represent the density of mutation among genes. Orange boxes are protein domains annotations from the pfam database. HCA, hepatocellular adenoma; IHCA, inflammatoryhepatocellular adenoma; HCC on HCA, hepatocellular carcinoma developed on hepatocellular adenoma; indel, inframe deletion.

with malignant transformation even in IHCA. In contrast, *HNF1A* inactivation (n=226) and sonic hedgehog activation (n=34) were almost exclusive from the inflammatory phenotype since only one mixed IHCA with bi-allelic inactivating mutations of *HNF1A* (#1355) was identified in our series.

Among the 296 IHCA, we identified an activating mutation of a gene belonging to the *JAK/STAT* pathway in 81% of the cases (figure 1A and B, online supplementary table 3):

- Mutations in *IL6ST,* identified in 180 cases, were almost all located in exon 6 in the IL-6 receptor alpha-chain domain and were mostly small in-frame deletions and rarely point mutations. Nevertheless, the remaining four in-frame deletions clustered in exon 10 corresponding to the D4 domain of gp130 delineating this second hotspot of mutations in *IL6ST*.
- *FRK* mutations, identified in 25 cases, clustered in the protein tyrosine kinase domain (exon 6) and were mainly small in-frame deletions. Interestingly, three samples displayed two *FRK* mutations on the same allele in exon 6 including two samples with p.Glu346Gly/p.Tyr350Cys mutations and one sample with p.Glu346Gly/p.Tyr368Cys mutations.
- *STAT3* mutations, identified in 15 cases, were either small in frame deletion or single nucleotide variation located in exons 3, 6, 17 and 21. Most of the mutations clustered in exons 3

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

and 21 coding for the protein interaction and SH2 domains, respectively. Interestingly, one sample with a mutation in *STAT3* exon 21 was also mutated in exon 17 (p.Lys658Tyr/p.Asp502Tyr) and another was mutated both in exons 21 and 3 (p.Tyr640Phe/p.Leu78Arg). Finally, one IHCA harboured a single mutation in exon 6 (p.Glu166Gln).

- All mutations in *GNAS*, identified in 10 cases, were missense identified at the Arginine 201, the known exon 8 hotspot activating the protein Gs-alpha subunit.
- *JAK1* mutations were identified in seven cases (2% of IHCA) in exons 11, 14, 15, 16 and 20 all leading to amino acid substitution activating *JAK1*. Interestingly, two samples were double mutated in *JAK1* exon 14 on the same allele at residues 644 and 645.

All oncogenic mutations identified in IHCA were mutually exclusive except for two cases: an IHCA harboured both *STAT3* (p.Leu78Arg) and *FRK* (p.Glu346Gly) mutations, another tumour showed an *IL6ST* in-frame deletion (p.Tyr186_Tyr190del) together with a *FRK* mutation (p.Val378_Lys380delinsGlu) (figure 1A). As previously described,[11] the expression level of *SAA2* and *CRP* was lower in *GNAS* activating mutated IHCA compared with IHCA with other genetic alterations (online supplementary figure 2)

Overall, 61 IHCA (21%) harboured no mutations in driver genes known to activate the JAK/STAT pathway.

Identification of recurrent *ROS1* fusion genes in IHCA

In order to search for new oncogenes in IHCA, we performed whole exome sequencing (WES) in 19 IHCA without known mutations. Comparing tumours and matched non-tumour liver samples, a median number of 2 synonymous mutations and 14 non-synonymous mutations per tumour were identified (online supplementary table 4). None of these samples harboured recurrent genetic alterations or recurrent copy number variation and no genes belonging to the *JAK/STAT* pathway was altered (online supplementary tables 4,5).

In contrast, using RNA sequencing in 22 IHCA without a known driver gene mutation, recurrent *ROS1* transcript alterations were identified in 10 cases (figure 2, online supplementary table 6); among these, 8 IHCA showed a chimeric transcript between the 3' part of *ROS1* and the 5' part of three different partner genes which are highly expressed in normal hepatocytes. *PLG-ROS1* fusions were identified in five cases as the result of a chromosome translocation t(6,10) (q22.1;q23.33). *RBP4-ROS1* fusions were identified in two cases as the result from a chromosome inversion at chromosome 6, Inv6 (q22.1-q25.3). Finally, *APOB-ROS1* fusion was identified in one case resulting from a chromosome translocation t(2,6) (q22.1;q24.1). Breakpoints in *ROS1* and the partner genes were variable leading in each case to at least one in phase transcript . However, all fusions included at least the last 9 exons of *ROS1* (exon 33–43 in one case, exon 34–43 in six cases and exon 35–43 in one case) fused in frame with the 5' of the various partner genes. One tumour with a *ROS1* fusion was also mutated for *CTNNB1* in exon 3 (figure 1A, online supplementary table 2).

Finally, in the last two cases, we identified a truncated mRNA of *ROS1* with an overexpression of exons 36–43 (#4089T) or 37–43 (#1896T) and, in these cases, we did not identify any fusion partner in RNA sequencing data; however, RNA-seq quality of sample #1896T was poor, due to a degraded RNA. In three cases, we performed a FISH experiment that confirmed the *ROS1* break in each tumour, including the two samples with only a truncated transcript identified (#4089T and #1896T) (figure 2D). GSEA analyse of the transcriptome of HCA with ROS1 fusions confirmed an enrichment in JAK/STAT pathway (HALLMARK IL6 JAK STAT3 SIGNALING; online supplementary figure 3).

In all the cases, the tyrosine kinase domain of *ROS1* (located in exons 36–42) was retained in the rearranged genes and the *ROS1* transcript was highly overexpressed (figure 2B, median FPKM=20.4) compared with HCA without *ROS1* fusion (median FPKM=7.5×10^{-3}, p value=8.6×10^{-5}) and compared with non-tumour liver (n=4, median FPKM=5.1×10^{-2}, p value=7.0×10^{-3}).

Identification of recurrent *FRK* fusion genes in IHCA

By analysing the 22 IHCA with RNAseq, five tumours with a chimeric transcript involving *FRK* were identified (figure 3A, online supplementary table 6), with one of them also mutated for *CTNNB1* in exon 7 (online supplementary table 2). All of them harboured an enrichment in genes belonging to JAK/STAT signalling at the transcriptomic level using GSEA analysis (online supplementary figure 3). *FRK* fusions resulted from various translocations linking exons 1–25 of *MIA3* (chromosome 1q41), or exons 1–20 of *MIA2* (chromosome 14q21.1) or exons 1–25 of *LMO7* (chromosome 13q22.2) or exons 1–19 of *PLEKHA5* (chromosome 12p12.3) with exons 3–8 of *FRK* (chromosome 6q21.1). The last *FRK* fusion linked exons 1–20 of *SEC16B* (chromosome 1q25.2) to exon 1–8 of *FRK* resulting in both a chimeric transcript including all exons of *FRK* and a truncated transcript starting with an alternative TSS in the intron 2 (figure 3C). All the fusions were in phase at the nucleotide level. In all the chimeric transcript, the tyrosine kinase domain of *FRK* (exons 4–8) was retained whereas the auto-inhibitory SH2 and SH3 domains were absent in the predicted aberrant protein. The expression level of *FRK* was not increased in IHCA with *FRK* fusions compared with other HCA (figure 3B).

Recurrent *IL6* gene rearrangements in IHCA

Finally, we identified two IHCA with abnormal 3' untranslated region (UTR) of *IL6* on RNA sequencing (figure 4A) and characterised by activation of JAK/STAT pathway (GSEA analysis, online supplementary figure 3). One case was previously published (#2615T with inflammatory SAA amyloidosis); the tumour showed a deletion of 3'UTR of *IL6* due to an inversion of 18.7 megabases validated by whole genome sequencing.[12] We identified a second IHCA, #1215T, with a truncation of the 3'UTR regulatory region of *IL6* due to a eight megabases deletion identified using RNA sequencing. In contrast to the first case, #1215T was not associated with SAA amyloidosis symptoms. In these two IHCA, *IL6* was strongly overexpressed (figure 4B, median FPKM=56.4) compared with others IHCA (median FPKM=2.01×10^{-1}, p value=2.6×10^{-2}).

Finally, six IHCAs analysed by WES and/or RNA sequencing remained without identifiable driver genetic alterations. Interestingly, on pathological reviewing, 5 among these 6 IHCA (83%) showed haemorrhage on histology (vs 58% in the whole series of IHCA) with a minimal expression of *CRP* and *SAA* at the transcriptomic level.

DISCUSSION

IHCA serves as a unique model to study the inflammatory pathways in mature hepatocytes and to identify new drivers of oncogene-induced inflammation in benign liver tumourigenesis. Based on the analysis of a large number of IHCA (see flow chart,

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

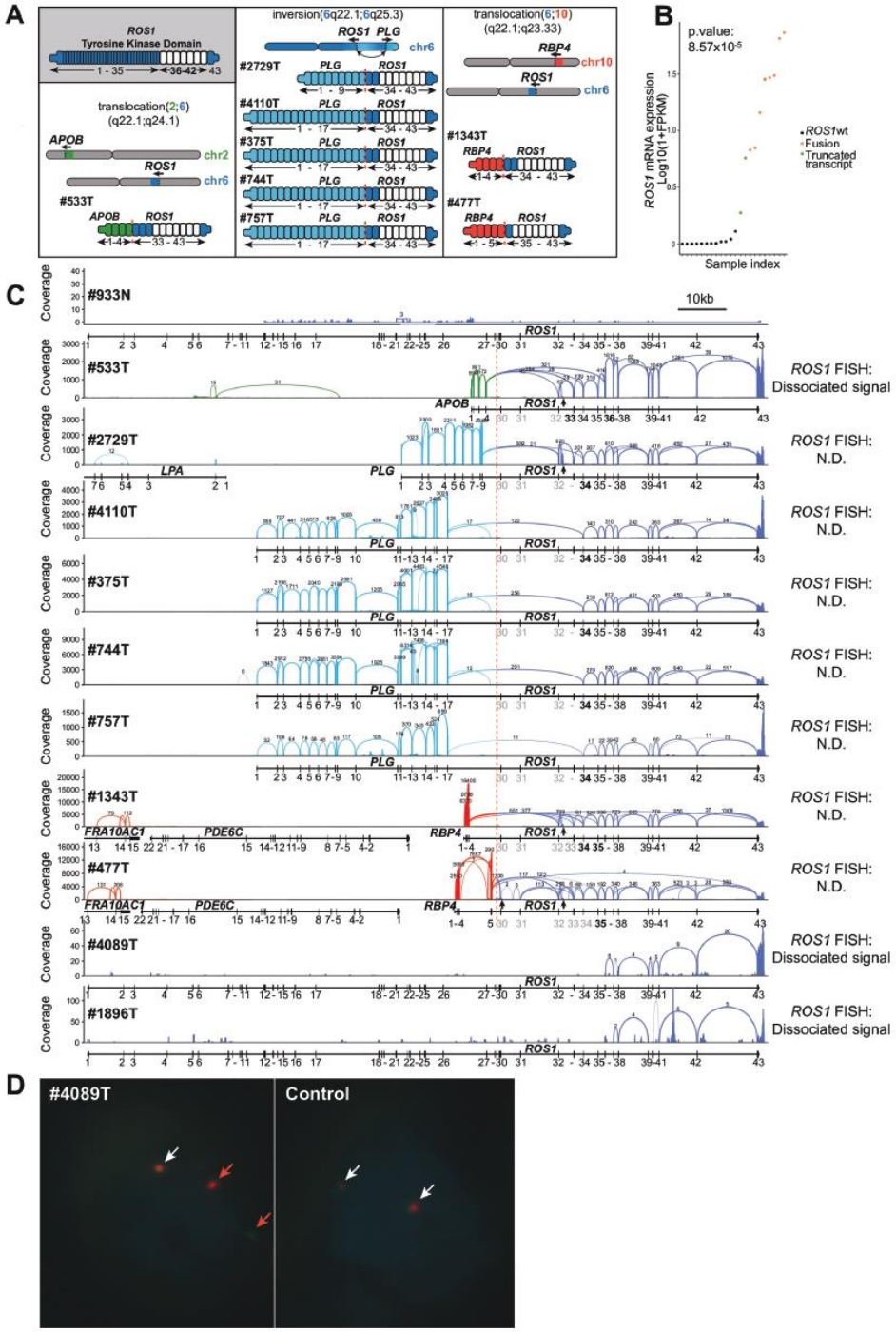


Figure 2 Recurrent *ROS1* fusions in inflammatory hepatocellular adenomas (IHCAs). (A) Representation of the *ROS1* fusion in 8 IHCAs whereas 2 IHCAs with *ROS1* truncated transcript were not represented. IHCA, inflammatory hepatocellular adenoma. (B) Expression level of *ROS1* in FPKM (log 10) in 22 hepatocellular adenoma analysed by RNA sequencing (10 IHCA with *ROS1* fusion/truncated transcripts and 12 IHCA with other genetic alterations). The p value compare the level of expression of *ROS1* between IHCA with *ROS1* alterations versus other IHCA using the Mann-Whitney non-parametric test. (C) Sashimi plots showing RNA-seq alignments for the wild-type *ROS1* in a normal liver (#933N) (top) and 10 cases of IHCA with *ROS1* fusion (bottom). Splice junction reads are represented as arcs connecting a pair of exons. Arc width is proportional to the number of reads aligning to the junction. Grey exon number shows exons not involved in chimeric transcripts and bold exon number represents the *ROS1* exons involved in the in-frame junction of the chimeric transcript. Black arrow indicates stop of transcription following abnormal junction. (D) FISH using *ROS1* dual-colour break-apart probe in 1 IHCA with known *ROS1* fusion and one control. White arrows indicate wild-type *ROS1* with combined red and green signals. Red arrows indicate dissociated signals validating *ROS1* rearrangement.

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

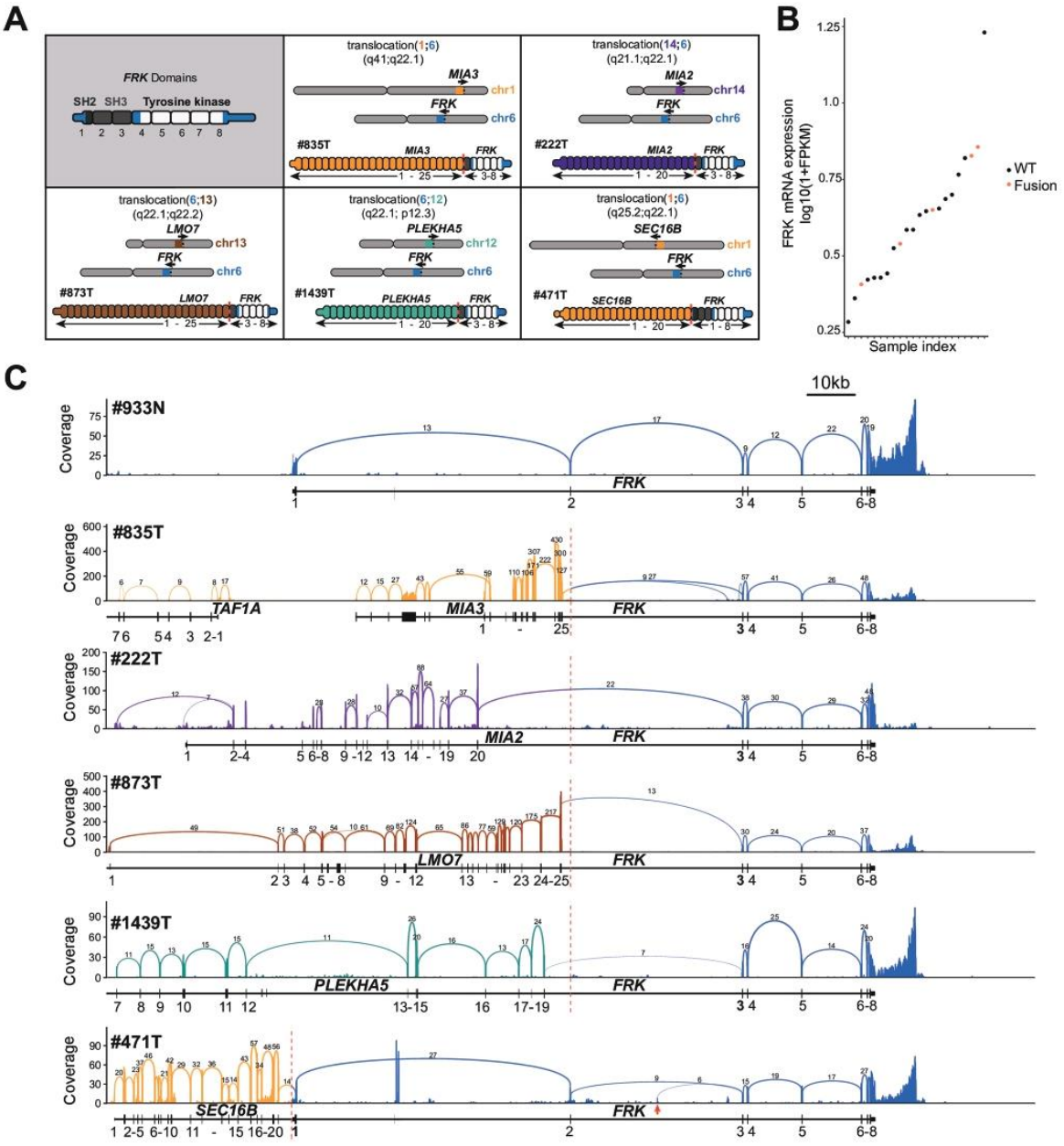


Figure 3 Recurrent *FRK* fusions in inflammatory hepatocellular adenomas (IHCAs). (A) Representation of the *FRK* fusions in five IHCAs. (B) Expression level of *FRK* in FPKM (log 10) in 22 HCA analysed by RNA sequencing. (C) Sashimi plots showing RNA-seq alignments for the wild-type *FRK* in a normal liver (#933N) (top) and five cases of IHCA with *FRK* fusion (bottom). Splice junction reads are represented as arcs connecting a pair of exons. Arc width is proportional to the number of reads aligning to the junction. Bold exon number represents the *ROS1* exons involved in the in-frame junction of the chimeric transcript. Red arrow indicates alternative transcription start site in the *FRK-SEC16B* fusion.

online supplementary figure 1), we identified novel recurrent somatic chromosomal rearrangements of *ROS1*, *FRK* and *IL6*.

In our study, IHCA is the most frequent molecular subgroup with 45% of HCA harbouring an inflammatory phenotype. Overall, we now described in IHCA a total of seven genes that are able to activate the JAK/STAT pathway in hepatocytes when mutated and therefore are associated with benign liver tumourigenesis.[22] Some of these oncogenes belong directly to the JAK/STAT pathway such as *IL6ST* (61% of mutations), *STAT3* (5% of mutations), *JAK1* (2% of mutations) or *IL6* (1% of alterations).[6 10] However, the remaining three genes, *FRK* (8% of mutations, 2% of fusions, mutually exclusive), *GNAS* (3% of mutations) and *ROS1* (3% of fusions), belong to other signalling pathways but are also known to activate the IL-6/JAK/STAT pathway in cellulo (figures 1A and 5).[6 11 23] Though in most of the IHCA, gene mutations were exclusive, in two cases, we observed a concomitant mutation of *STAT3* L78R with either a classical *FRK* or *STAT3* mutations. We had previously shown,

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

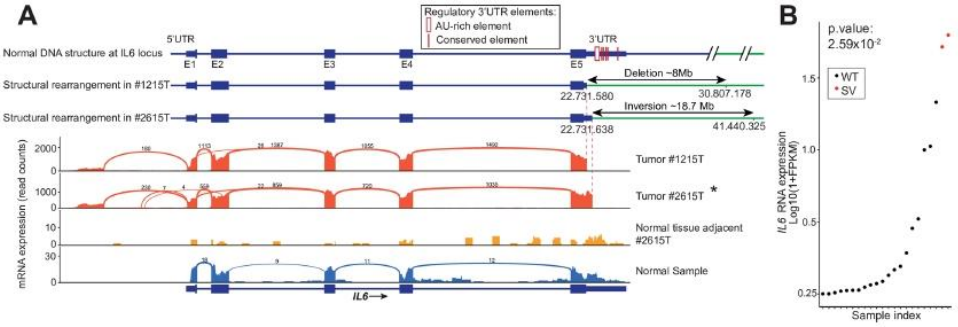


Figure 4 *IL6* rearrangements in inflammatory hepatocellular adenomas (IHCAs). (A) Sashimi plots showing RNA-seq alignments for the wild-type *IL6* in a normal liver and a corresponding non-tumour liver (bottom) and two cases of IHCA with truncation of 3'UTR (top). Splice junction reads are represented as arcs connecting a pair of exons. Arc width is proportional to the number of reads aligning to the junction. *already published[12]. (B) Expression level of *IL6* in FPKM (log 10) in 22 HCA analysed by RNA sequencing. The p value compared the level of expression of *IL6* between IHCA with *IL6* alterations versus other IHCA using the Mann-Whitney non-parametric test.Coordinates in GRCh38. FPKM, fragments per kilobase of exon model per million reads mapped; IHCA, inflammatory hepatocellular adenoma; WT, wild type; SV, structural variants.

in cellulo, that *STAT3* L78R mutations alone can only weakly activate the JAK/STAT pathway.[24] Another study suggested that the L78R mutation, located in the N-terminal domain, could prolong the activation of *STAT3*.[25] We could hypothesise that *STAT3* L78R mutation could cooperate with additional mutation in the JAK/STAT pathway to induce an inflammatory response in mature hepatocytes.

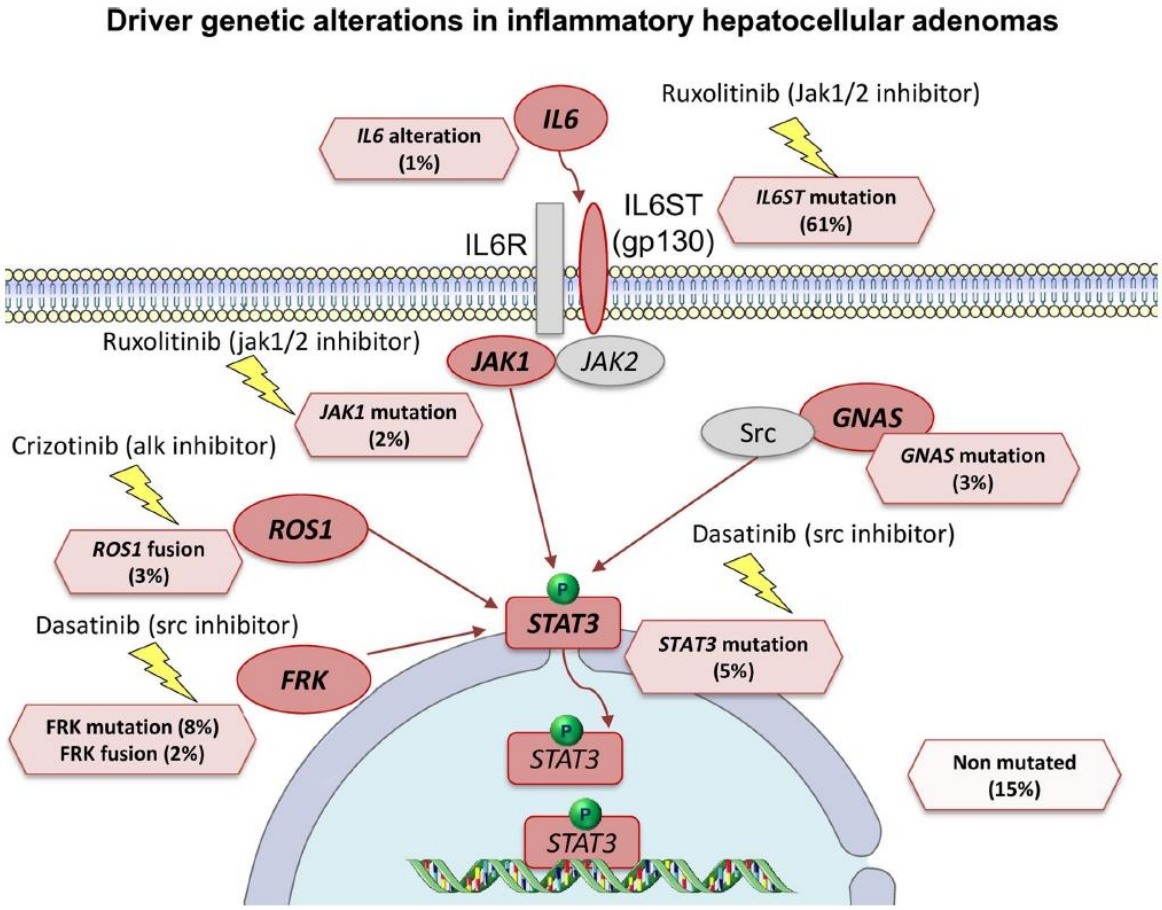


Figure 5 Mutations in oncogenes in inflammatory hepatocellular adenomas (IHCAs). Frequencies of somatic genetic alterations in driver genes observed in IHCA are indicated. Targeted therapies potentially active against each genetic alteration are indicated by a yellow flag. IHCA, inflammatory hepatocellular adenoma.

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

In this work, we identified recurrent *ROS1* rearrangements in 10 IHCA without other mutations in driver genes. *ROS1* is a tyrosine kinase receptor composed of a N-terminal extracellular domain, a transmembrane domain and a C-terminal intracellular tyrosine kinase domain, known to activate several downstream pathways such as the MAP kinase and AKT/mTOR pathways.[26 27] Recurrent *ROS1* rearrangements are identified in 2% of non-small cell lung carcinoma and in other cancer types including cholangiocarcinoma, lymphoma, glioblastoma, colorectal and gastric carcinoma.[26] In anaplastic lymphoma and in inflammatory myofibroblastic tumours, *ROS1* fusions induce a strong activation of the JAK/STAT pathway.[23 28] To our knowledge IHCA is the second type of benign tumour showing *ROS1* fusion after Spitz nevi.[29] This observation reinforces the hypothesis that *ROS1* fusion is an early event during the mechanism of carcinogenesis in solid tumours but it is not sufficient alone to promote a full malignant transformation of hepatocyte or melanocyte.

In IHCA, exons 33–35 of *ROS1* are fused in-phase with different partners including *PLG*, *APOB* and *RBP4*. All these partners are highly expressed in adult liver whereas *ROS1* is almost not expressed in normal hepatocytes. Resulting from the gene fusion, *ROS1* is under the control of the promoter of the different partner genes and consequently overexpressed specifically in all IHCA harbouring the *ROS1* rearrangement. Interestingly, all *ROS1* fusions retained the intracellular tyrosine kinase domain (exon 36–42) and lost the transmembrane domain. These findings are similar to all the *ROS1* fusion proteins already identified in lung cancer, lymphoma or cholangiocarcinoma that all retained the tyrosine kinase domain, essential for its oncogenic activity.[23 30 31] In cellulo model we have also demonstrated that *ROS1* fusions were able to activate *STAT3* in haematological cancer.[23] Moreover, most of *ROS1* fusions described in cancer lack their transmembrane domain leading to the relocalsation of the fusion proteins in the cytoplasm that could interact with different proteins and modified downstream signalling pathway.[27 32] In two of the IHCA cases, though we identified a truncated overexpressed *ROS1* mRNA, its partner genes could not be identified owing to the poor quality of RNA. However, *ROS1* breaks were confirmed by FISH in both the cases.

We previously identified in IHCA recurrent mutations of *FRK* able to activate JAK/STAT pathway in cellulo.[6] In this study, recurrent *FRK* fusions were identified in five cases of IHCA, all showed an in-frame fusion gene. The fusion transcript was predicted to code for a chimeric protein retaining the tyrosine kinase domain, encoded by exons 4–8, but lacking the SH2 and SH3 auto-inhibitory domains, encoded by exons 1–3.[33] Interestingly, *FRK* fusions have been described in acute myeloid leukaemia with *ETV6-FRK* and in anaplastic large cell lymphoma with *CAPRIN-FRK*, all with a similar junctions at exon 3 of *FRK*[34].[35] *In cellulo* model showed that *ETV6-FRK* fusion protein has a constitutive autophosphorylation on its tyrosine residues with an elevated kinase activity.[35] All these data suggested that in hepatocyte like in haematological diseases, the deletion of the SH2/SH3 domain together with an intact tyrosine kinase domain is critical for the oncogenic mechanism of *FRK* fusions.

We also described two cases of somatic alterations of *IL6* with truncation of the 3'UTR due to large chromosome inversion or deletion at the genomic level. These two IHCA harboured a strong overexpression of *IL6* suggesting that the lacks of the 3'UTR regulatory elements induced an overexpression of the transcript. Interestingly, the IL-6 mRNA has a short half-life of around 30 min that is regulated by the 3'UTR.[36] Consequently, the truncation of the 3'UTR in these two IHCA could lead to an increased stability and accumulation of the transcript. We previously described one of the cases that was associated with occurrence of inflammatory amyloidosis,[12] but the second case did not.

In our series, we identified *IL6* alterations, *ROS1* and *FRK* fusions in 3%, 2% and 1% of all HCA, respectively. In all these HCA, no other driver mutations were identified except one *CTNNB1* mutation exon 3 in one case and one *CTNNB1* mutation exon 7 in another case. As *ROS1* and *IL6* rearrangements induced an mRNA overexpression, we did not identify an additional candidate IHCA for these genetic alterations after screening of the whole series of IHCA by quantitative RT-PCR. In contrast, the exact frequency of *FRK* rearrangement could be under-estimated.

In six IHCAs, no functional genetic alterations were identified using next generation sequencing. However, in 5 of these 6 cases, we observed haemorrhage at histological level which is known to induce inflammation and thus a slight increase in SAA and CRP expression in these adenomas without a genomic driven activation of the *IL6/JAK/STAT3* pathway.

Finally, IHCA harboured several driver genes that could be used as a target for drug therapy in the future in clinical practice. *ROS1* fusion could be targeted by crizotinib, a small molecule currently approved by Food and Drug Administration (FDA) for the treatment of lung cancer with *ROS1* fusion.[37] Moreover, we previously showed that dasatinib, a src inhibitor, shutdown JAK/STAT activation due to *FRK* mutations *in cellulo*.[6] We can predict that the same results could be obtained in IHCA with *FRK* fusions. These data suggested that targeted therapies adapted to genetic alterations (ruxolitinib in *IL6ST* mutation, dasatinib for *FRK* mutations/fusions and crizotinib for *ROS1* fusion) could be tested in selected patients with multiple or unresectable IHCA (figure 5).[38]

In conclusion, we described recurrent chromosomal alterations of *ROS1*, *FRK* and *IL6* as new mechanisms of JAK/STAT pathway activation driving benign tumourigenesis in IHCAs that could be tested as therapeutic targets in future clinical trials.

Short summary

We identified recurrent chromosomal alterations involving *ROS1*, *FRK* or *IL6* in IHCAs. *ROS1* activation is a novel mechanism of JAK/STAT pathway activation in human benign liver tumours.

Author affiliations
[1]Centre de Recherche des Cordeliers, Sorbonne Université, Inserm,Université de Paris, Université Paris 13, Functional Genomics of Solid Tumors laboratory, F-75006 Paris, France
[2]Service de Pathologie, Hôpital Pellegrin, CHU de Bordeaux, F 33076 Bordeaux, France
[3]Université Bordeaux, UMR1053 Bordeaux Research in Translational Oncology, BaRITOn, F-33076 Bordeaux, France
[4]Service d'anatomopathologie, Hôpital Beaujon, Assistance-Publique Hôpitaux de Paris, Clichy, France
[5]INSERM U1149, Clichy, France
[6]Service d'anatomopathologie, CHU, Grenoble, France
[7]Service d'anatomopathologie, CHU, Tours, France
[8]Service d'anatomopathologie, CHU Bicêtre, Assistance-Publique Hôpitaux de Paris, Bicêtre, France, Bicêtre, France
[9]Service d'anatomopathologie, Hôpital Henri Mondor; Université Paris Est, Inserm U955, Team 18, Institut Mondor de Recherche Biomédicale, France, Créteil, France
[10]Service d'hépatologie, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bondy, France
[11]Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris 13, Communauté d'Universités et Etablissements Sorbonne Paris Cité, Paris, France, Paris, France
[12]Hôpital Européen Georges Pompidou, F-75015, Assistance Publique-Hôpitaux de Paris, Paris, France

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

Twitter Jean-Charles Nault @naultjc and Jessica Zucman-Rossi @Zucmanrossi

Acknowledgements This study was previously presented as an abstract at the American Association for the Study of Liver Diseases liver meeting 2019.

Contributors Author contributions to manuscript—study design: J-CN, JZ-R; generation of experimental data: QB, GC, SC, J-CN, BB, CB, JZ-R; analysis and interpretation of data: QB, GC, SC, J-CN, SI, EL, SR, JZ-R; collection of samples and related histological and clinical data: PBS, CB, VP, NS, AdM, CG, J-CN; drafting of the manuscript: QB, SC, EL, J-CN, JZ-R; revision of the manuscript and approval of the final version of the manuscript: all the authors. J-CN, JZ-R: these authors shared the last authorship.

Funding This work was supported by Institut National du Cancer (INCa) with the International Cancer Genome Consortium (ICGC LICA-FR project), INSERM with the 'Cancer et Environnement' (plan Cancer) and MUTHEC projects (INCa). The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir), grant IREB, Coup d'Elan de la Fondation Bettencourt-Schueller, the SIRIC CARPEM, Raymond Rosen Award from the Fondation pour le Recherche Médicale, Prix René and Andrée Duquesne Comité de Paris Ligue Contre le Cancer and Fondation Mérieux.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Paris Saint-Louis Institutional Review Board committee approved this study (Paris Saint-Louis, 2004; INSERM IRB 2010; the French Liver Biobanks network—AFAQ NF S96-900 and Hepatobio bank)

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

ORCID iDs
Sandrine Imbeaud http://orcid.org/0000-0001-8439-6732
Jean-Charles Nault http://orcid.org/0000-0002-4875-9353

REFERENCES

1 European Association for the Study of the Liver (EASL). EASL clinical practice guidelines on the management of benign liver tumours. *J Hepatol* 2016;65:386–98.
2 Nault J-C, Paradis V, Cherqui D, *et al*. Molecular classification of hepatocellular adenoma in clinical practice. *J Hepatol* 2017;67:1074–83.
3 Zucman-Rossi J, Jeannot E, Nhieu JTV, *et al*. Genotype-Phenotype correlation in hepatocellular adenoma: new classification and relationship with HCC. *Hepatology* 2006;43:515–24.
4 Bluteau O, Jeannot E, Bioulac-Sage P, *et al*. Bi-Allelic inactivation of TCF1 in hepatic adenomas. *Nat Genet* 2002;32:312–5.
5 Pelletier L, Rebouissou S, Paris A, *et al*. Loss of hepatocyte nuclear factor 1alpha function in human hepatocellular adenomas leads to aberrant activation of signaling pathways involved in tumorigenesis. *Hepatology* 2010;51:557–66.
6 Pilati C, Letouzé E, Nault J-C, *et al*. Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation. *Cancer Cell* 2014;25:428–41.
7 Nault J-C, Couchy G, Balabaud C, *et al*. Molecular Classification of Hepatocellular Adenoma Associates With Risk Factors, Bleeding, and Malignant Transformation. *Gastroenterology* 2017;152:880–94.
8 Bioulac-Sage P, Rebouissou S, Thomas C, *et al*. Hepatocellular adenoma subtype classification using molecular markers and immunohistochemistry. *Hepatology* 2007;46:740–8.
9 Paradis V, Champault A, Ronot M, *et al*. Telangiectatic adenoma: an entity associated with increased body mass index and inflammation. *Hepatology* 2007;46:140–6.
10 Rebouissou S, Amessou M, Couchy G, *et al*. Frequent in-frame somatic deletions activate gp130 in inflammatory hepatocellular tumours. *Nature* 2009;457:200–4.
11 Nault JC, Fabre M, Couchy G, *et al*. GNAS-activating mutations define a rare subgroup of inflammatory liver tumors characterized by STAT3 activation. *J Hepatol* 2012;56:184–91.
12 Calderaro J, Letouzé E, Bayard Q, *et al*. Systemic AA amyloidosis caused by inflammatory hepatocellular adenoma. *N Engl J Med* 2018;379:1178–80. 20.
13 Rebouissou S, Franconi A, Calderaro J, *et al*. Genotype-Phenotype correlation of CTNNB1 mutations reveals different ß-catenin activity associated with liver tumor progression. *Hepatology* 2016;64:2047-2061.
14 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.
15 DePristo MA, Banks E, Poplin R, *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
16 Van der Auwera GA, Carneiro MO, Hartl C, *et al*. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43.
17 Kim D, Pertea G, Trapnell C, *et al*. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.
18 Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166–9.
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 Bayard Q, Meunier L, Peneau C, *et al*. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun* 2018;9:5235.
21 Shugay M, Ortiz de Mendíbil I, Vizmanos JL, *et al*. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* 2013;29:2539–46.
22 Bromberg JF, Wrzeszczynska MH, Devgan G, *et al*. Stat3 as an oncogene. *Cell* 1999;98:295–303.
23 Crescenzo R, Abate F, Lasorsa E, *et al*. Convergent mutations and kinase fusions lead to oncogenic STAT3 activation in anaplastic large cell lymphoma. *Cancer Cell* 2015;27:516–32.
24 Pilati C, Amessou M, Bihl MP, *et al*. Somatic mutations activating STAT3 in human inflammatory hepatocellular adenomas. *J Exp Med* 2011;208:1359–66.
25 Domoszlai T, Martincuks A, Fahrenkamp D, *et al*. Consequences of the disease-related L78R mutation for dimerization and activity of STAT3. *J Cell Sci* 2014;127:1899–910.
26 Davies KD, Doebele RC. Molecular pathways: ROS1 fusion proteins in cancer. *Clin Cancer Res* 2013;19:4040–5.
27 Uguen A, De Braekeleer M. Ros1 fusions in cancer: a review. *Future Oncol* 2016;12:1911–28.
28 Lovly CM, Gupta A, Lipson D, *et al*. Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 2014;4:889–95.
29 Wiesner T, He J, Yelensky R, *et al*. Kinase fusions are frequent in spitz tumours and spitzoid melanomas. *Nat Commun* 2014;5:3116.
30 Takeuchi K, Soda M, Togashi Y, *et al*. Ret, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012;18:378–81.
31 Gu T-L, Deng X, Huang F, *et al*. Survey of tyrosine kinase signaling reveals ROS kinase fusions in human cholangiocarcinoma. *PLoS One* 2011;6:e15640.
32 Pal P, Khan Z. ROS1 [corrected]. *J Clin Pathol* 2017;70:1001–9.
33 Goel RK, Lukong KE. Understanding the cellular roles of fyn-related kinase (FRK): implications in cancer biology. *Cancer Metastasis Rev* 2016;35:179–99.
34 Hu G, Dasari S, Asmann YW, *et al*. Targetable fusions of the FRK tyrosine kinase in ALK-negative anaplastic large cell lymphoma. *Leukemia* 2018;32:565–9.
35 Hosoya N, Qiao Y, Hangaishi A, *et al*. Identification of a SRC-like tyrosine kinase gene, FRK, fused with ETV6 in a patient with acute myelogenous leukemia carrying a t(6;12)(q21;p13) translocation. *Genes Chromosomes Cancer* 2005;42:269–79.
36 Paschoud S, Dogar AM, Kuntz C, *et al*. Destabilization of interleukin-6 mRNA requires a putative RNA stem-loop structure, an AU-rich element, and the RNA-binding protein AUF1. *Mol Cell Biol* 2006;26:8228–41.
37 Shaw AT, Kim D-W, Nakagawa K, *et al*. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013;368:2385–94.
38 Poussin K, Pilati C, Couchy G, *et al*. Biochemical and functional analyses of gp130 mutants unveil JAK1 as a novel therapeutic target in human inflammatory hepatocellular adenoma. *Oncoimmunology* 2013;2:e27090.

Gut: first published as 10.1136/gutjnl-2019-319790 on 6 January 2020. Downloaded from http://gut.bmj.com/ on May 7, 2020 at INSERM Consortia. Protected by copyright.

2. Structural rearrangement signatures reveal a new molecular entity of HCC

The second objective of my PhD was to explore the causes and consequences of structural variations in hepatocellular carcinomas. As for HCA, I exploited RNAseq data but I also implemented an innovative approach to explore SV signatures from WGS data. This approach allowed me to extract subgroups of tumors sharing similar patterns of structural variations, induced by a shared biological defect.

During my first year of PhD, I participated to the development of Palimpsest (https://github.com/FunGeST/Palimpsest), an R package dedicated to the analysis of mutational signatures from NGS data. **Article 3** presents Palimpsest's functionalities, including *de novo* or supervised analysis of mutation and SV signatures, but also clonality reconstruction to unravel the mutational history of each tumor. I contributed to this work by implementing the SV signature analysis workflow.

Article 4 presents the application of Palimpsest's rearrangement signature framework to a large cohort of HCC samples. In this work, I identified six rearrangement signatures, characterized by preferential types and sizes of SVs, operative with different intensities in each tumor. In particular, rearrangement signature 1 (RS1), characterized by focal duplications (10s to 100s of kb) and complex rearrangements called Template Insertion Cycles (T.I.C.), revealed a homogenous HCC subgroup with a widespread duplicator phenotype. All tumors of this subgroup (named CCN-HCC) displayed an activation of *CCNA2* or *CCNE1* gene by various mechanisms. *CCNA2* activating events included new gene fusions and insertional mutagenesis by HBV or AAV2 virus, all leading to the production of a truncated cyclin A2 protein lacking regulatory domains while keeping functional domains. *CCNE1* activation resulted from insertional mutagenesis by HBV and AAV2, but also SVs leading to enhancer hijacking or gene amplifications.
Cyclins E1 and A2 are involved in S phase entry and progression, and cyclin E1 activation is known to induce replicative stress and genomic instability in various models. Consistently, CCN-HCC displayed an overexpression of ATR pathway implicated in the response to replication stress. These data, and the fact that duplications and TIC can be induced when collapsed replication forks are processed by the break-induced replication

(BIR) mechanism, suggest that RS1 is a signature of replication stress. Integrative analysis of HCC genomes with epigenetic data from Encode and ROADMAP consortia revealed an enrichment of RS1 breakpoints in early replicated, highly expressed active chromatin domains. Besides, RS1 events frequently associate active chromatin with quiescent regions, which may induce secondary oncogenic events like *TERT* promoter activation by enhancer hijacking, recurrent in CCN-HCC.

Overall, SV signature analysis allowed us to identify a new HCC entity (CCN-HCC), with homogenous molecular and clinical features. These tumors usually develop in non-cirrhotic liver tissue, in absence of classical risk factors. They have a very poor prognosis but may benefit from targeted treatment against replication stress response.

Article 3: Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer

Bioinformatics, 34(19), 2018, 3380–3381
doi: 10.1093/bioinformatics/bty388
Advance Access Publication Date: 16 May 2018
Applications Note

OXFORD

Genome analysis

Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer

Jayendra Shinde[1,*], Quentin Bayard[1], Sandrine Imbeaud[1], Théo Z. Hirsch[1], Feng Liu[2], Victor Renault[2], Jessica Zucman-Rossi[1,3] and Eric Letouzé[1,*]

[1]INSERM, UMR-1162, Génomique Fonctionnelle des Tumeurs Solides, Equipe Labellisée Ligue Contre le Cancer, Institut Universitaire d'Hématologie; Faculté de Médecine, Université Paris Descartes, Labex Immuno-Oncology, Sorbonne Paris Cité; Unité de Formation et de Recherche Santé, Médecine, Biologie Humaine, Université Paris 13, Sorbonne Paris Cité, Bobigny, France; Université Paris Diderot, Paris, France, [2]Laboratory for Bioinformatics, Fondation Jean Dausset – CEPH, Paris F-75010, France and [3]Assistance Publique-Hôpitaux de Paris, Hopital Europeen Georges Pompidou, Paris F- 75015, France

*To whom correspondence should be addressed.

Associate Editor: Russell Schwartz

Received on March 26, 2018; revised on May 2, 2018; editorial decision on May 3, 2018; accepted on May 12, 2018

Abstract

Summary: Cancer genomes are altered by various mutational processes and, like palimpsests, bear the signatures of these different processes. The *Palimpsest* R package provides a complete workflow for the characterization and visualization of mutational signatures and their evolution along tumor development. The package covers a wide range of functions for extracting both base substitution and structural variant signatures, inferring the clonality of each alteration and analyzing the evolution of mutational processes between early clonal and late subclonal events. *Palimpsest* also estimates the probability of each mutation being due to each process to predict the mechanisms at the origin of driver events. *Palimpsest* is an easy-to-use toolset for reconstructing the natural history of a tumor using whole exome or whole genome sequencing data.

Availability and implementation: *Palimpsest* is freely available at www.github.com/FunGEST/Palimpsest.

Contact: jayendra.shinde91@gmail.com and eric.letouze@inserm.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Mutational signature analysis is a powerful approach to understand the origin of somatic mutations in cancer (Alexandrov *et al.*, 2013). This approach can be extended to any kind of genomic alteration including structural variants (Helleday *et al.*, 2014). Mutational processes may evolve over time, giving different mutational signatures in early clonal and late subclonal mutations (Nik-Zainal *et al.*, 2012; Rubanova *et al.*, 2018). Next-generation sequencing data also allow to reconstruct the timing of copy-number alterations (CNAs) by distinguishing clonal/subclonal events and using the proportion of duplicated mutations to time chromosome duplications (Greenman *et al.*, 2012). Finally, different mutational processes have different propensities to target specific loci depending on the local chromatin state (Polak *et al.*, 2015; Sabarinathan *et al.*, 2016) and sequence context (Letouzé *et al.*, 2017). Understanding how mutational processes evolve during tumorigenesis and alter specific driver genes is critical to better understand the natural history of cancers and eventually predict and influence their clinical behavior.

© The Author(s) 2018. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/34/19/3380/4996591 by guest on 07 May 2020

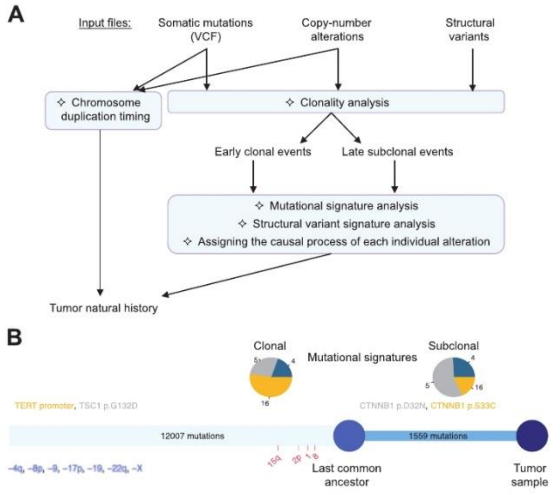


Fig. 1. **(A)** Workflow of a typical analysis with *Palimpsest*. **(B)** Oncogenic timeline of a tumor representing the number of clonal/subclonal mutations, the contribution of mutational processes to each step and the timing of driver mutations (colored according to their most likely causal mutational process) and CNAs

To the best of our knowledge, no comprehensive workflow integrates mutational signature and clonality analyses to reconstruct the natural history of a tumor. In addition, no automated tools for extracting structural variant signatures, timing chromosome duplications and estimating the process at the origin of each mutation are currently available. The *Palimpsest* R package provides an extensive toolset to explore and visualize the diverse processes generating somatic alterations in a tumor, their evolution along tumorigenesis and their interaction with driver genes.

2 Features of *palimpsest*

Palimpsest takes as input somatic mutations and structural variants (optional), copy-number data and a minimal sample annotation file indicating gender and tumor purity. Figure 1A and Supplementary Figure S1 illustrate a typical Palimpsest analysis, as described step by step below.

Mutational signature analysis. *Palimpsest* allows both *de novo* extraction of novel mutational signatures or quantification of previously described signatures using non-negative matrix factorization (NMF), as implemented in the *NMF* R package (Gaujoux and Seoighe, 2010). Utilities are provided to visualize signatures, transcriptional strand biases and inter-mutation distance, and to compare signatures using cosine similarity.

Structural variant signature analysis. *Palimpsest* implements an adaptation of the mutational signature analysis framework for structural variants (SVs). SVs are first classified into 38 categories according to the type (deletion, tandem duplication, inversion and interchromosomal translocation) and size of rearrangements (Nik-Zainal *et al.*, 2016). NMF then allows extracting new or known structural variant signatures and their contribution to each tumor genome.

Predicting the process at the origin of each individual mutation. Once the mutation catalogue of a tumor has been deconvoluted as the addition of several mutational processes, *Palimpsest* estimates the probability of each individual mutation being due to each process using simple Bayesian statistics (Letouzé *et al.*, 2017). This key feature, available for both point mutations and structural variants, allows predicting the processes most likely at the origin of driver events in each tumor.

Clonality analysis and evolution of mutational signatures. *Palimpsest* integrates variant allele fraction, tumor purity and local copy number to estimate the proportion of tumor cells harboring each mutation and classify them as clonal or subclonal. Then, the package extracts and compares mutational signatures operative in clonal and subclonal mutations to explore the evolution of mutational processes along tumorigenesis.

Timing CNAs. Several tools have been described to differentiate clonal from subclonal CNAs, which can be directly specified in the CNA input file. In addition, *Palimpsest* will estimate the molecular timing of duplications using the proportion of duplicated/non-duplicated somatic mutations.

Oncogenic timelines. Finally, *Palimpsest* integrates the results of clonality and mutational signature analyses to generate a comprehensive representation of the natural history of the tumor (Fig. 1B).

3 Conclusions

Unraveling what mutational processes generate driver mutations and fuel the initiation and progression of tumor clones is crucial to better understand the natural history of cancers and optimize personalized patient care. *Palimpsest* integrates a broad range of functions within a single convenient workflow to provide a complete picture of tumorigenesis. The minimum basic input data make the package easy-to-use in any cancer genomic study downstream classical variant calling and copy number analyses.

Acknowledgements

We thank the CIT program from the French Ligue Nationale Contre le Cancer for providing useful R codes (cit.x functions).

Funding

This work was supported by INSERM ("Cancer et environnement" & HTE program), cancéropôle Ile-de-France (exhauTrans project), Région Ile-de-France, BPI France (ICE project), CARPEM and ICGC. The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology Investissement d'Avenir and Fondation Mérieux.

Conflict of Interest: none declared.

References

Alexandrov,L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.

Gaujoux,R. and Seoighe,C. (2010) A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, **11**, 367.

Greenman,C.D. *et al.* (2012) Estimation of rearrangement phylogeny for cancer genomes. *Genome Res.*, **22**, 346–361.

Helleday,T. *et al.* (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.*, **15**, 585–598.

Letouzé,E. *et al.* (2017) Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.*, **8**, 1315.

Nik-Zainal,S. *et al.* (2012) The life history of 21 breast cancers. *Cell*, **149**, 994–1007.

Nik-Zainal,S. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, **534**, 47–54.

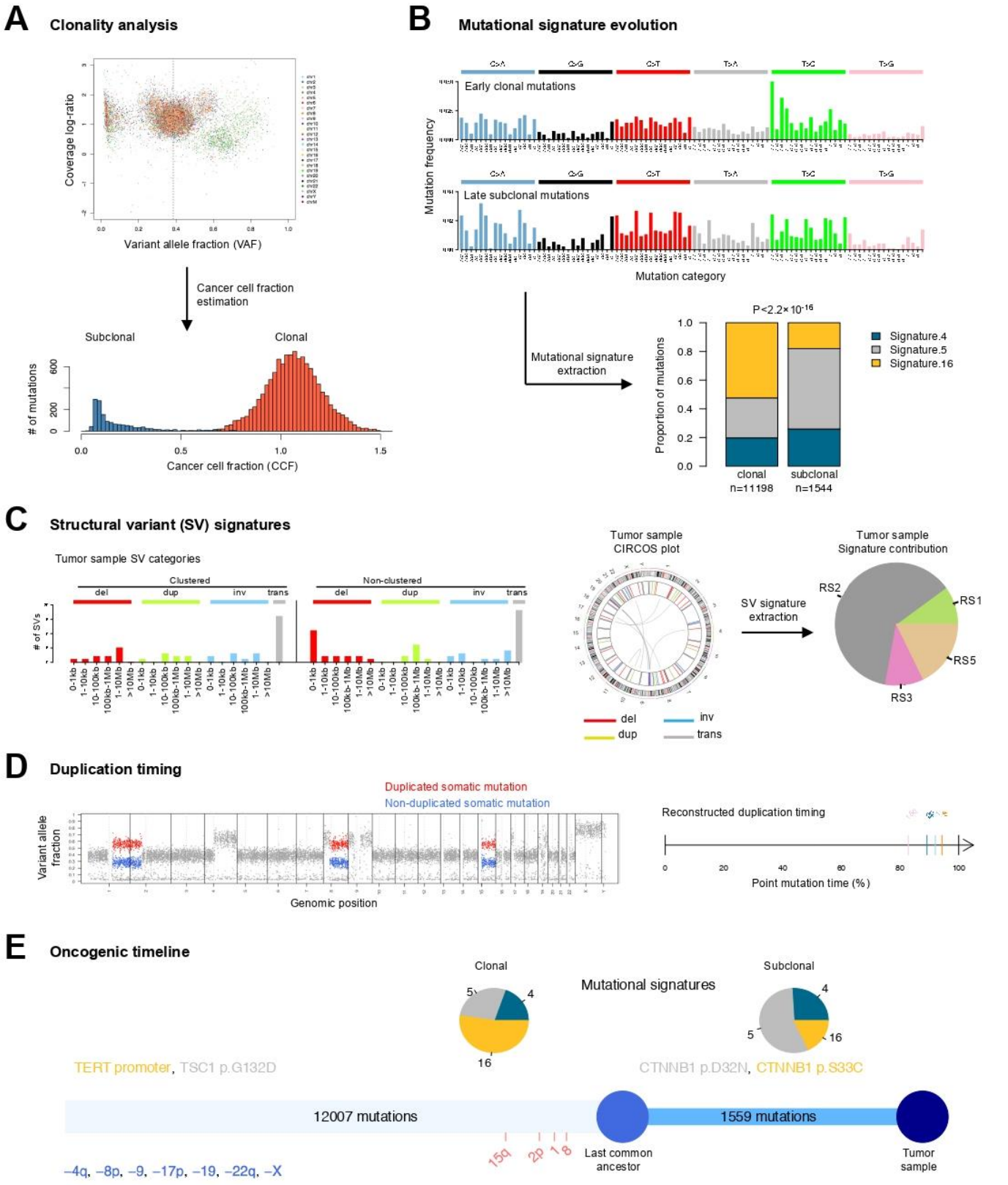
Polak,P. *et al.* (2015) Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*, **518**, 360–364.

Rubanova,Y. *et al.* (2018) TrackSig: reconstructing evolutionary trajectories of mutation signature exposure. *bioRxiv*, 260471.

Sabarinathan,R. *et al.* (2016) Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature*, **532**, 264–267.

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/34/19/3380/4996591 by guest on 07 May 2020

Supplementary figure:



Supplementary Fig.1. Typical outputs of a *Palimpsest* analysis. (A) *Palimpsest* estimates the cancer cell fraction (proportion of tumor cells harboring a somatic variant) and classifies each variant as clonal or subclonal. **(B)** Mutational signatures analyzed separately in early clonal and late subclonal mutations highlight the evolution of mutational processes along tumorigenesis. **(C)** *Palimpsest* can also extract structural variant (SV) signatures considering the type and size of SVs identified in the tumor. Del: deletion; dup: duplication; inv: inversion; trans: translocation. **(D)** The ratio of duplicated/non-duplicated somatic mutations allows estimating the timing of each chromosome duplication in molecular time. **(E)** Finally, an oncogenic timeline of the tumor is generated, with the number of clonal and subclonal mutations, their distribution per mutation signature, the driver alterations (colored according to the most likely causal mutational process) and CNA timing. This example comes from a previously analyzed tumor (Letouzé *et al.*, 2017).

Supplementary note:

Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer.

Jayendra Shinde, Quentin Bayard, Sandrine Imbeaud, Théo Z Hirsch, Feng Liu, Victor Renault, Jessica Zucman-Rossi and Eric Letouzé*

*Corresponding author. E-mail: eric.letouze@inserm.fr

In this supplementary note, we describe in more details the methodological details underlying the main functions of the *Palimpsest* package.

S1. Mutational signature analysis

Palimpsest allows both *de novo* analysis of mutational signatures and quantification of previously described signatures. In both cases, a set of base substitutions from a tumor series is first imported from a VCF file and each mutation is assigned to one of 96 mutation categories, defined by the 6 substitution types multiplied by 16 possible trinucleotide contexts (Alexandrov *et al.*, 2013). We can then represent the mutational catalog of the *n* tumors as a mutation matrix $M = \begin{pmatrix} m_1^1 & \dots & m_n^1 \\ \vdots & \ddots & \vdots \\ m_1^{96} & \dots & m_n^{96} \end{pmatrix}$ where m_i^j is the number of mutations of category *j* in tumor *i*. The goal of mutational signature analysis is to deconvolute this matrix *M* as the product of a matrix of mutational processes $P = \begin{pmatrix} p_1^1 & \dots & p_K^1 \\ \vdots & \ddots & \vdots \\ p_1^{96} & \dots & p_K^{96} \end{pmatrix}$ where p_i^j is the probability of the process *i* to cause a mutation of category *j*, and a matrix of exposures $E = \begin{pmatrix} e_1^1 & \dots & e_n^1 \\ \vdots & \ddots & \vdots \\ e_1^K & \dots & e_n^K \end{pmatrix}$ where e_i^j is the number of mutations attributed to process *i* in tumor *j*. To solve this problem, we use non-negative matrix factorization, as implemented in the *NMF* package (Gaujoux and Seoighe, 2010). For a *de novo* analysis, the number of processes *K* can be manually defined by the user or estimated automatically considering the cophenetic correlation coefficients and residual sum of squares (RSS) for each number of signatures. NMF then identifies the matrices *P* and *E* that verifies $M \approx P \times E$ and minimizes a Frobenius norm while maintaining non-negativity. For an analysis of previously described signatures, the user provides the matrix of processes P, and NMF is used only to reconstruct the exposure matrix *E*. The *P* matrix for the 30 signatures currently referenced in the COSMIC database is provided with the *Palimpsest* package.

S2. Association of mutational signatures with driver genes

Once a cancer genome has been deconvoluted as a combination of several mutational processes, we can estimate the probability of each somatic mutation being generated by each

mutational process using simple Bayes statistics (Letouzé *et al.*, 2017). This key feature has been introduced in Palimpsest as follows:

Consider a mutation category c out of the 96 mutation categories, the number of mutations of category *c* in a tumor *t* can be expressed as:

$$m_t^c = \sum_{s=1}^{10} p_s^c \times e_t^s$$

where the product $p_s^c \times e_t^s$ represents the number of mutations of category *c* attributed to signature *s* in tumor *t*. The probability *P(m,s)* of a mutation *m* of category *c* in tumor *t* being due to signature *s* can then be estimated as:

$$P(m,s) = \frac{p_s^c \times e_t^s}{\sum_{s=1}^{10} p_s^c \times e_t^s}$$

This important feature of Palimpsest can be used to predict the mutational processes (point mutations or structural variants) at the origin of driver alterations in a cancer genome.

S3. Structural rearrangement signature analysis

Palimpsest performs structural rearrangement signature analysis by applying the same statistical framework used for mutational signature analysis. Somatic structural rearrangements from a series of tumors are first classified into 38 categories considering the type (deletion, tandem duplication, inversion, interchromosomal translocation) and size (<1kb, 1-10kb, 10-100kb, 100kb-1Mb, 1-10Mb, >10Mb) of rearrangements, as previously described (Nik-Zainal *et al.*, 2016). Clustered events, defined by the presence of ≥10 breakpoints within a 1Mb window, are identified using the *bedr* package (Haider *et al.*, 2016) and considered separately from non-clustered events. We then used non-negative matrix factorization, as implemented in the *NMF* package, to extract rearrangement signatures and their exposure in each tumor. Like mutational signature analysis, structural rearrangement signature analysis can be performed *de novo* or using a pre-defined set of known signatures to estimate the exposure matrix.

S4. Estimating the cancer cell fraction and clonality

Using the variant allele fraction (VAF), tumor cell content and absolute copy-number estimates, *Palimpsest* estimates the cancer cell fraction (CCF), i.e. the proportion of tumor cells harboring each mutation:

$$CCF = VAF \times \frac{\rho N_t + (1-\rho) N_n}{\rho n_{chr}}$$

where ρ is the tumor cell content, N_t and N_n the copy-number at the locus in tumor and normal cells, and n_{chr} the number of chromosomal copies harboring the mutation in tumor cells (also

called multiplicity of the mutation). ρ and N_t can be estimated from copy-number data using various algorithms. The multiplicity of each mutation (n_{chr}) is set to the integer value closest to:

$$max\left(1,\ VAF\times\frac{\rho N_t+(1-\rho)N_n}{\rho}\right)$$

Finally, Palimpsest determines the 95% confidence of VAF using a binomial test and converts this interval to obtain the 95% confidence interval of CCF using the above formula. A mutation is then classified subclonal if the upper boundary of the 95% confidence interval is below a threshold set by the user (default 0.95), and clonal otherwise.

S5. Timing chromosomal duplications

When a chromosome is duplicated, mutations harbored by the chromosome are also duplicated and their VAF is increased as compared to mutations present on the other chromosome copy, or acquired after the duplication. The ratio of duplicated/non-duplicated mutations can thus be used to time the chromosome duplication event, early events having a low amount of duplicated mutations as compared to late events (Nik-Zainal *et al.*, 2012). *Palimpsest* uses previously described formulas (Letouzé *et al.*, 2017) to estimate the point mutation time of each chromosome duplication with a minimum of 30 mutations located on the duplicated segment.

Let us consider the simple case of a chromosome with absolute copy-number N_t=3. The molecular time at which the extra copy of the chromosome was gained can be estimated as:

$$T=\frac{N_{dup}}{N_{dup}+\frac{N_{ndup}-N_{dup}}{3}}\times 100$$

where N_{dup} and N_{ndup} are the number of duplicated and non-duplicated mutations, respectively.

We extrapolated this formula to chromosomes with N_t≥4. In this case, we timed the first duplication event using:

$$T=\frac{N_{dup}}{N_{dup}+\frac{N_{ndup}-N_{dup}}{(3+N_t)/2}}\times 100$$

where N_{dup} is the number of mutations at the maximal level of multiplicity and N_{ndup} the number of mutations at intermediate levels of multiplicity or non-duplicated.

For cases where the two parental chromosome copies were duplicated, e.g. N_t=4 with 2 copies of each chromosome copy, we adapted the formula as follows:

$$T=\frac{N_{dup}/2}{N_{dup}/2+\frac{N_{ndup}}{(3+N_t)/2}}\times 100$$

S6. Limitations of the package – number of mutations needed

Palimpsest can be applied to both whole genome and whole exome sequencing data. However, whole genome data are preferable for structural rearrangement signature analysis and timing chromosome duplications. Factors influencing extraction of mutational signatures have been extensively analyzed elsewhere (Alexandrov *et al.*, Cell Rep 2013) using simulated data with varying numbers of signatures, similarities between signatures, numbers of tumors and numbers of mutations per tumor. The authors concluded that, although both whole genome and whole exome sequencing data are suitable to accurately extract mutational signatures, the number of required mutations/samples increases with the number of signatures, and it is preferable to have more mutations/sample (e.g. whole genome sequences) than more samples with less mutations (e.g. large whole exome series). However, giving precise limitations regarding a required number of mutations per sample is delicate as it also depends on the type of mutational signatures analyzed. In our experience, very specific signatures caracterized by only a few mutation categories mutated at high frequency will be easily identified with relatively few mutations, whereas signatures with widespread mutation spectra will by harder to quantify.

References

Alexandrov,L.B. *et al.* (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.*, **3**, 246–259.

Gaujoux,R. and Seoighe,C. (2010) A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, **11**, 367.

Haider,S. *et al.* (2016) A bedr way of genomic interval processing. *Source Code Biol. Med.*, **11**, 14.

Letouzé,E. *et al.* (2017) Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.*, **8**, 1315.

Nik-Zainal,S. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, **534**, 47–54.

Nik-Zainal,S. *et al.* (2012) The life history of 21 breast cancers. *Cell*, **149**, 994–1007.

Article 4: Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-07552-9 OPEN

Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress

Quentin Bayard[1,2,3,4], Léa Meunier[1,2,3,4], Camille Peneau[1,2,3,4], Victor Renault[5], Jayendra Shinde[1,2,3,4], Jean-Charles Nault[1,2,3,4,6,7], Iadh Mami[1,2,3,4], Gabrielle Couchy[1,2,3,4], Giuliana Amaddeo[8,9], Emmanuel Tubacher[5], Delphine Bacq[10], Vincent Meyer[10], Tiziana La Bella[1,2,3,4], Audrey Debaillon-Vesque[11], Paulette Bioulac-Sage[12,13], Olivier Seror[1,14], Jean-Frédéric Blanc[11,12], Julien Calderaro[8,15], Jean-François Deleuze[5,10], Sandrine Imbeaud[1,2,3,4], Jessica Zucman-Rossi[1,2,3,4,16] & Eric Letouzé[1,2,3,4]

Cyclins A2 and E1 regulate the cell cycle by promoting S phase entry and progression. Here, we identify a hepatocellular carcinoma (HCC) subgroup exhibiting cyclin activation through various mechanisms including hepatitis B virus (HBV) and adeno-associated virus type 2 (AAV2) insertions, enhancer hijacking and recurrent *CCNA2* fusions. Cyclin A2 or E1 alterations define a homogenous entity of aggressive HCC, mostly developed in non-cirrhotic patients, characterized by a transcriptional activation of E2F and ATR pathways and a high frequency of *RB1* and *PTEN* inactivation. Cyclin-driven HCC display a unique signature of structural rearrangements with hundreds of tandem duplications and templated insertions frequently activating *TERT* promoter. These rearrangements, strongly enriched in early-replicated active chromatin regions, are consistent with a break-induced replication mechanism. Pan-cancer analysis reveals a similar signature in *BRCA1*-mutated breast and ovarian cancers. Together, this analysis reveals a new poor prognosis HCC entity and a rearrangement signature related to replication stress.

[1] INSERM, UMR-1162, Génomique Fonctionnelle des Tumeurs Solides, Equipe Labellisée Ligue Contre le Cancer, Institut Universitaire d'Hématologie, Paris 75010, France. [2] Université Paris Descartes, Labex Immuno-Oncology, Sorbonne Paris Cité, Faculté de Médecine, Paris 75006, France. [3] Université Paris 13, Sorbonne Paris Cité, Unité de Formation et de Recherche Santé, Médecine, Biologie Humaine, Bobigny 93017, France. [4] Université Paris Diderot, Sorbonne Paris Cité, Paris 75013, France. [5] Laboratory for Bioinformatics, Fondation Jean Dausset – CEPH, Paris 75010, France. [6] Liver unit, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, APHP, Bondy 93140, France. [7] Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris 13, Communauté d'Universités et Etablissements Sorbonne Paris Cité, Bobigny 93017, France. [8] Inserm, U955, Team 18, Université Paris-Est Créteil, Faculté de Médecine, Créteil 94010, France. [9] Assistance Publique-Hôpitaux de Paris, Service d'Hépatologie, CHU Henri Mondor, Créteil 94010, France. [10] Centre National de Recherche en Génomique Humaine, CEA, Evry 91000, France. [11] Service Hépato-Gastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, Centre Hospitalier Universitaire de Bordeaux, Bordeaux 33076, France. [12] Université Bordeaux, Bordeaux Research in Translational Oncology, Bordeaux 33076, France. [13] Service de Pathologie, Hôpital Pellegrin, Centre Hospitalier Universitaire de Bordeaux, Bordeaux 33000, France. [14] Radiology Department, Jean Verdier Hospital, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, Bondy 93140, France. [15] Assistance Publique-Hôpitaux de Paris, Département de Pathologie, Hôpital Henri Mondor, Créteil 94010, France. [16] Assistance Publique-Hôpitaux de Paris, Hopital Européen Georges Pompidou, 75015 Paris, France. These authors contributed equally: Léa Meunier, Camille Peneau. These authors jointly supervised this work: Jessica Zucman-Rossi, Eric Letouzé. Correspondence and requests for materials should be addressed to J.Z-R. (email: jessica.zucman-rossi@inserm.fr) or to E.Lé. (email: eric.letouze@inserm.fr)

Hepatocellular carcinoma (HCC) is the third leading cause of cancer death worldwide. Only 30% of cases are diagnosed at an early stage and are amenable to curative treatment by tumor resection or liver transplantation[1]. The multikinase inhibitors sorafenib[2] and regorafenib[3] are currently the only drugs approved for advanced HCC cases, but the median life expectancy of patients with HCC on sorafenib is only 1 year. All phase III clinical trials involving targeted molecular therapies have failed so far for various reasons including liver toxicity, lack of antitumoral potency, and the molecular heterogeneity of the disease[4]. Identifying homogeneous HCC subgroups sharing similar driving mechanisms and vulnerabilities is thus crucial to design successful patient-tailored clinical trials.

Most HCC develop in a cirrhotic liver, associated with various etiologies including hepatitis B virus (HBV) and hepatitis C virus (HCV) infections, alcohol abuse, metabolic disease, and exposure to carcinogenic compounds like aflatoxin B1[5]. The natural history of HCC in cirrhosis follows a well-established sequence with the successive development of dysplastic nodules that can transform into early stage and advanced HCC. *TERT* promoter mutations are the initial oncogenic events already detected in dysplastic nodules[6] whereas alterations in other HCC drivers[7–11] involved in cell cycle control (*TP53*, *RB1*, *CCND1*, *CDKN2A*), Wnt/ß-catenin signaling (*CTNNB1*, *AXIN1*), oxidative stress response (*NFE2L2*, *KEAP1*) epigenetic regulation (*ARID1A*, *ARID2*) and the AKT/mTOR and MAP kinase pathway (*RPS6KA3*, *TSC1*, *TSC2*, *PTEN*) only occur in progressed HCC[12].

In 20% of the cases, HCC develops in absence of cirrhosis. These patients usually maintain adequate liver functions and, being less subject to liver toxicity, may be eligible for more treatment options. The etiology of HCC in absence of cirrhosis is largely unknown, but one mechanism of transformation involves insertional mutagenesis by the HBV virus. The first oncogenic HBV insertion was identified in cyclin A2 gene (*CCNA2*)[13]. Since then, recurrent HBV insertions were mapped in several oncogenes including *CCNE1*, *KMT2B* and *TERT*[14,15]. Recently, we identified adeno-associated virus type 2 (AAV2) insertions as a new etiology for HCC developed in absence of cirrhosis, with recurrent insertions in *CCNA2* and *CCNE1* genes[16]. However, the molecular consequences of viral insertions in cyclin genes and their precise role in HCC development remain poorly understood.

Here, we report the systematic screening of *CCNA2* and *CCNE1* alterations in 751 HCC. We identify new mechanisms of cyclin A2/E1 activation, and we explore the clinical and molecular characteristics of this tumor subgroup.

Results

Viral insertions and gene fusions activate cyclin A2. To identify the exhaustive landscape of *CCNA2* and *CCNE1* alterations in HCC, we analyzed 751 HCC comprising an in-house series of 160 tumors (LICA-FR) analyzed by RNA sequencing (RNAseq, $n = 160$), whole exome (WES, $n = 156$) and whole genome sequencing (WGS, $n = 45$) (Supplementary Data 1), the TCGA[17] series (334 HCC with RNA-seq and WES, 48 or which also analyzed by WGS) and the ICGC-JP[11] series (257 HCC with WGS data, Supplementary Data 2).

We first screened the LICA-FR series of 160 tumors to characterize the exhaustive mechanisms activating *CCNA2* and *CCNE1* in HCC. We identified one HBV and 5 AAV2 insertions (four previously described in the ref. [16]) in *CCNA2* gene (Supplementary Data 3), all but one located within *CCNA2* intron 2 (Fig. 1a). Viral insertions were associated with *CCNA2* mRNA over-expression ($P = 8.2 \times 10^{-9}$, fold-change = 5.6, Fig. 1b), but also altered the transcript and protein structure. AAV2 and HBV insertions induced the expression of various abnormal transcripts (Supplementary Fig. 1), predicted to generate a truncated cyclin A2 protein starting at methionine 148 or 158 with occasionally a few amino acids translated from the viral genome (Fig. 1c).

In addition we identified novel gene fusions in 4 tumors (Supplementary Data 4), all involving the C-terminal part of *CCNA2* (exons 3–8) at chromosome 4q27 downstream 3 different partner genes: *GSTCD* at 4q24, *SNX29* at 16p13.13 and *TET2* (×2) at 4q24 (Fig. 1a, d). In the *TET2-CCNA2* and *GSTCD-CCNA2* fusion transcripts, the first untranslated exons of *TET2* and *GSTCD* were linked with *CCNA2* exons 3–8. The *SNX29-CCNA2* fusion revealed an alternative transcription start site (TSS) in *SNX29* intron 14 generating a 448-nucleotide sequence spliced with *CCNA2* exon 3. In all fusions, the predicted translation initiation site of the fused RNA was located at methionine 158 in *CCNA2* exon 3, predicted to generate a truncated cyclin A2 protein of 275 amino acids (32 Kda), lacking the destruction box[18] and ubiquitination targeting sequences[19] but retaining the functional cyclin box, without any protein fragment from the partner genes (Fig. 1e).

Western blot analysis of 9 tumors with viral insertion or gene fusion confirmed the over-expression, as predicted, of a truncated 32 KDa protein (Fig. 1f). Thus, gene fusions and viral insertions in *CCNA2* both lead to the production of a stable protein lacking the N-terminal regulatory domains.

In the TCGA series, we identified 7 *CCNA2* fusions with 5 different partner genes (*FAM160A1*, *KIAA1109* × 3, *LIPC*, *UBA6* and *TDO2*, Fig. 1a, d), all of which involved the first untranslated exon(s) of the partner gene linked with exons 3–8 of *CCNA2*. WGS revealed in another tumor a focal deletion starting in the 5′ UTR region and ending in *CCNA2* intron 2 (Supplementary Fig. 2). All these events were predicted to generate the same 32 KDa truncated cyclin A2 protein lacking N-terminal regulatory domains. We also identified one tumor with HBV insertion and 3 tumors with AAV2 insertions in *CCNA2*. Finally, 6 tumors strongly overexpressed *CCNA2* (FPKM > 15), 3 of which displayed 23–48 Mb intra-chromosomal deletions linking the intergenic region downstream *CCNA2* with the highly expressed *ALB*, *AFP*, and *ADH6* genes (Supplementary Fig. 2). The ICGC-JP cohort comprised one HBV insertion in *CCNA2* intron 2 and one fusion between the first untranslated exon of *ANXA5* and exons 3–8 of *CCNA2* (Fig. 1a, d).

In total, we identified 10 HCC with *CCNA2* activation events in the LICA-FR series (6.2%), 2 in the ICGC-JP series (0.8%) and 18 in the TCGA series (5.4%), associated with a significant increase of *CCNA2* mRNA expression, but also generating a truncated cyclin A2 protein lacking the N-terminal destruction box and the ubiquitination site.

Viral insertions and enhancer hijacking activate cyclin E1. In our series of 160 HCC, we identified 5 AAV2 insertions (three previously described in the ref. [16]) and one HBV insertion in the 5′ region or upstream the transcription start site (TSS) of *CCNE1* (Fig. 2a, Supplementary Data 3). These viral insertions induced a massive overexpression of the full-length *CCNE1* gene (Fig. 2b), confirmed by western-blot analysis (Supplementary Fig. 3). Interestingly, one case with AAV2 insertion (FR2141T) also displayed an amplification of *CCNE1* locus including the viral sequence (Supplementary Fig. 3), suggesting a two-step selection of *CCNE1* activation in the natural history of this tumor. Four other tumors overexpressed *CCNE1* (FPKM > 6), explained by high-level amplification in one case. In the 3 remaining cases, whole genome sequencing revealed interchromosomal translocation breakpoints in the regulatory region of *CCNE1* (Fig. 2a). Tumor FR2048T

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07552-9

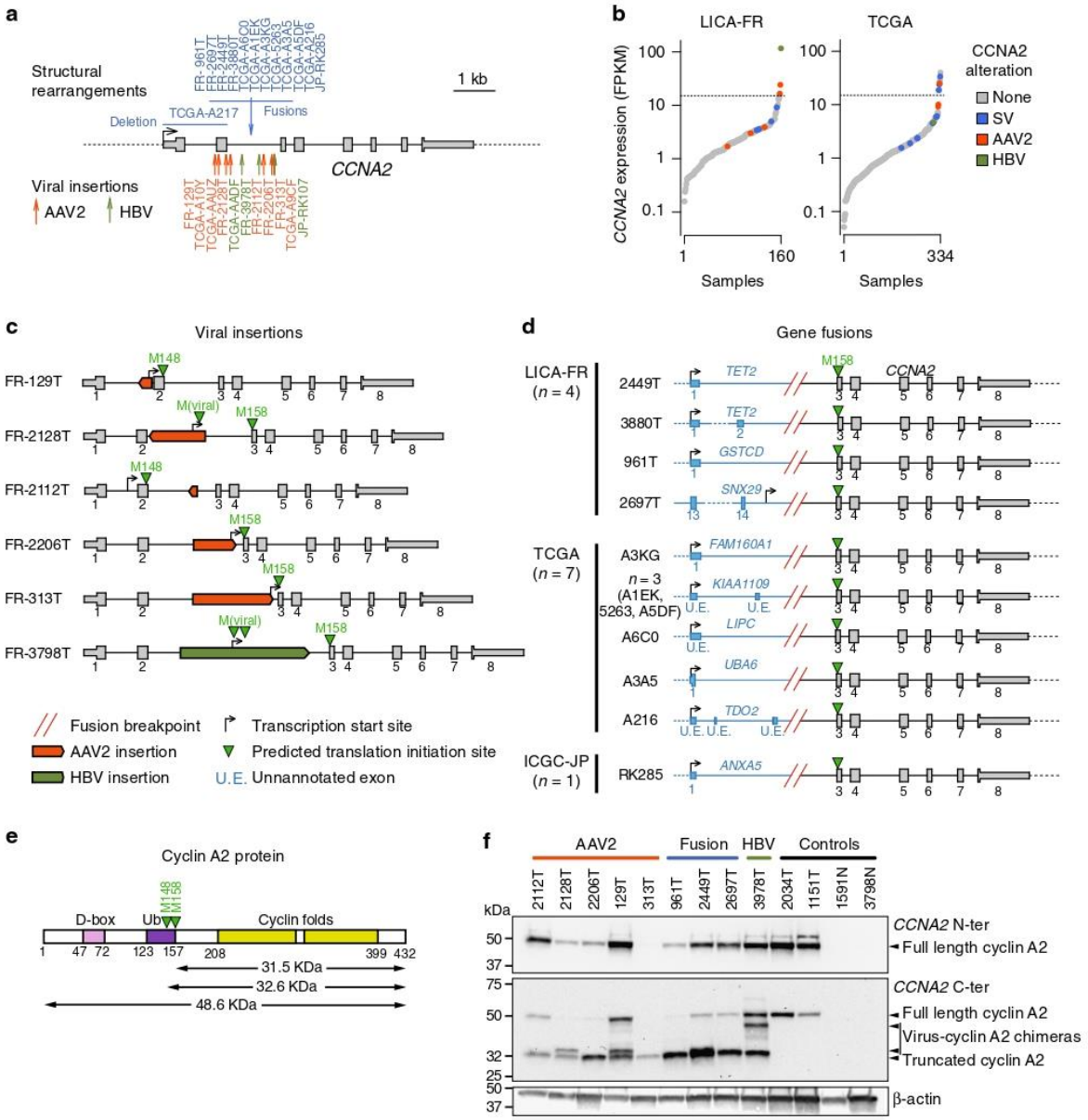


Fig. 1 Diverse mechanisms leading to *CCNA2* activation in HCC. **a** Summary of structural rearrangements (top) and viral insertions (bottom) affecting *CCNA2* gene identified in 751 HCC from the LICA-FR, TCGA and ICGC-JP cohorts. **b** Sorted *CCNA2* expression (log scale) in the LICA-FR and TCGA cohorts. Gene expression was obtained from RNA-seq data and is given in fragments per kilobase of exons per million reads (FPKM). Samples harboring structural variants (SV) or viral insertions are indicated with a color code. **c** Functional consequences of AAV2 and HBV insertions in *CCNA2*. Viral insertions identified in the LICA-FR cohort were precisely mapped using WGS or viral capture data, and RNA-seq reads were aligned on the reconstructed chimeric DNA to identify the transcription start sites and predicted translation initiation sites of abnormal transcripts. **d** *CCNA2* fusions identified in the LICA-FR, TCGA and ICGC-JP cohorts. The transcription start site of the fusion transcript is represented together with the predicted translation initiation site. Fusions with *KIAA1109*, *LIPC* and *TDO2* involve 5′ exons not annotated in transcript databases but expressed in normal liver. **e** Schematic representation of cyclin A2 protein with functional domains. D-box Destruction box; Ub, Ubiquitination targeting sequences. **f** Western blot analysis of cyclin A2 using antibodies targeting the N-terminal (top) or C-terminal (middle) domains. Tumors with viral insertions or gene fusions are compared with tumors without *CCNA2* alteration and non-tumoral liver controls

displayed a translocation placing *CCNE1* downstream the first untranslated exon of the highly expressed *ERRFI1* gene, leading to a highly expressed *ERRFI1-CCNE1* fusion. The two other translocations lead to juxtapose *CCNE1* promoter with enhancer-rich chromatin areas located close to the highly expressed genes *RAPH1* and *CYB5A* (Fig. 2c). Thus, both viral insertions and structural rearrangements can activate *CCNE1* expression by bringing viral or distal human enhancers in the regulatory region of the gene.

In the TCGA series, 10 tumors overexpressed *CCNE1* (Fig. 2b), including 2 cases with HBV insertion, one with HBV insertion

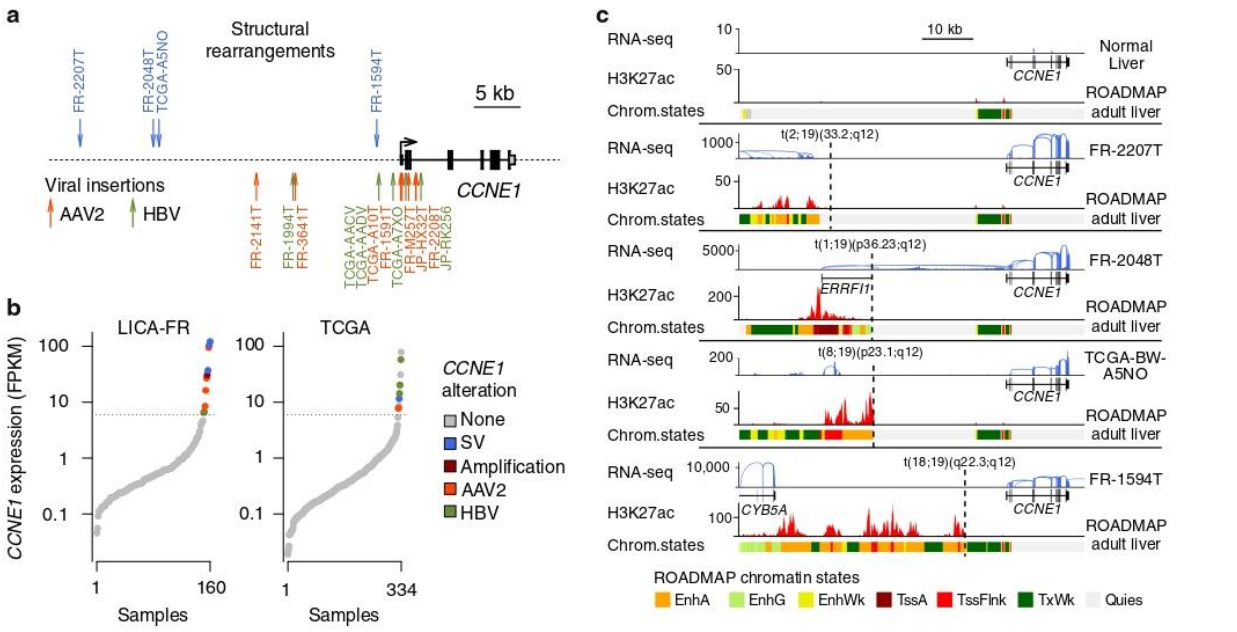


Fig. 2 Viral and non-viral mechanisms of *CCNE1* activation in HCC. **a** Summary of structural rearrangements (top) and viral insertions (bottom) affecting *CCNE1* gene identified in 751 HCC from the LICA-FR, TCGA and ICGC-JP cohorts. **b** Sorted *CCNE1* expression (log scale) in the LICA-FR and TCGA cohorts. Gene expression was obtained from RNA-seq data and is given in fragments per kilobase of exons per million reads (FPKM). Samples harboring structural variants (SV), focal amplifications and viral insertions are indicated with a color code. **c** Functional consequences of structural rearrangements affecting *CCNE1* regulatory region. RNA-seq read counts along *CCNE1* locus are represented in normal liver (top) and in 4 tumors harboring structural rearrangements upstream *CCNE1* trancription start site (TSS). H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) signal and chromatin states in adult liver were obtained from the ROADMAP consortium and are depicted below each reconstruted DNA sequence. EnhA: active enhancer; EngG: genic enhancer; EnhWk: weak enhancer; TssA: active TSS; TssFlnk: flanking TSS; TxWk: weak transcription; Quies: quiescent chromatin

plus high-level amplification, one with AAV2 insertion and one with a translocation between *CCNE1* regulatory region and an enhancer-rich region on chromosome 5 (Fig. 2c). In the 5 remaining cases, the mechanism leading to *CCNE1* overexpression remained unexplained in absence of WGS data. In the ICGC-JP cohort, we identified one AAV2 and one HBV insertion associated with *CCNE1* overexpression. In total, we identified 10 HCC with *CCNE1* activation events in the LICA-FR cohort (6.2%), two in the ICGC series (0.8%) and 10 in the TCGA series (3.0%).

Across the three data sets, 52/751 tumors (6.9%) displayed an activation of cyclin A2 ($n = 30$) or E1 ($n = 22$) due to viral insertions or structural rearrangements. These are later referred to as CCN-HCC. The proportion of CCN-HCC varied between the cohorts (12.5% in our series, 8.4% in TCGA and 1.6% in ICGC-JP) due to differences in etiological backgrounds (Supplementary Data 2). It was particularly high in our series enriched in cancers developed in a non-fibrotic liver, and low in the IGGC-Japan series dominated by HCV-related cases.

Cyclin A2 or E1 activation defines a homogenous HCC subgroup. We next explored the molecular and clinical characteristics of CCN-HCC. Gene expression analysis of the LICA-FR and TCGA showed that CCN-HCC defined homogeneous transcriptional clusters (Fig. 3a). They were characterized by an overexpression of cell cycle genes, in particular E2F targets, and an activation of the ATR pathway in response to replication stress (Fig. 3b, Supplementary Data 5). The most significant downregulated pathways were oxidative phosphorylation, suggesting a metabolic switch to aerobic glycolysis (Warburg effect), and *MYC* targets. We also compared the alteration frequencies of known liver cancer driver genes[10] between CCN-HCC and others. *CCNA2* and *CCNE1* activation events were remarkably exclusive from *CTNNB1* and *TERT* promoter mutations, but frequently associated with *PTEN* and *RB1* inactivation in both the LICA-FR and TCGA series (Fig. 3b, Supplementary Data 6). *RB1* inactivation may allow cells to overcome oncogene-induced senescence[20] in these tumors, whereas *PTEN* inactivation might favor the oncogenic metabolic switch that we observed at the transcriptional level[21]. Compared to the other tumors in the LICA-FR series, CCN-HCC were enriched in large tumors (median largest nodule diameter = 115 vs. 60 mm, $P = 0.0033$), of poor prognosis (median overall survival = 21 vs. 69 months, $P = 0.0072$, Fig. 3c), developed in younger patients (median age = 57 vs. 67 years old, $P = 0.050$) with a non-fibrotic liver (fibrosis stage F0-F1 80 vs. 42%, $P = 0.0011$). Thus, CCN-HCC define a homogenous HCC entity with characteristic clinical and molecular features.

CCN-HCC display a unique structural rearrangement signature. To identify mutational signatures associated with CCN-HCC, we analyzed the whole genome sequences of 45 of our 160 HCC (35 were previously published[22], 10 new), including 13 CCN-HCC. With a median of 12,463 mutations, CCN-HCC were rather less mutated than others (median = 16,397 mutations, $P = 0.065$). Mutational signatures 4, 5, and 16 (COSMIC nomenclature), ubiquitous in liver cancers[22], accounted for most mutations in CCN-HCC, with a slight increase of signature 5 (53 vs. 33%, $P = 0.036$) and decrease of signature 16 (23 vs. 32%, $P = 0.05$) as compared with other HCC (Supplementary Fig. 4).

In contrast, CCN-HCC displayed > 3 times more structural variants (median = 415 vs. 126, $P = 1.1 \times 10^{-4}$). We identified 6 rearrangement signatures, termed RS1 to RS6, characterized by

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07552-9 ARTICLE

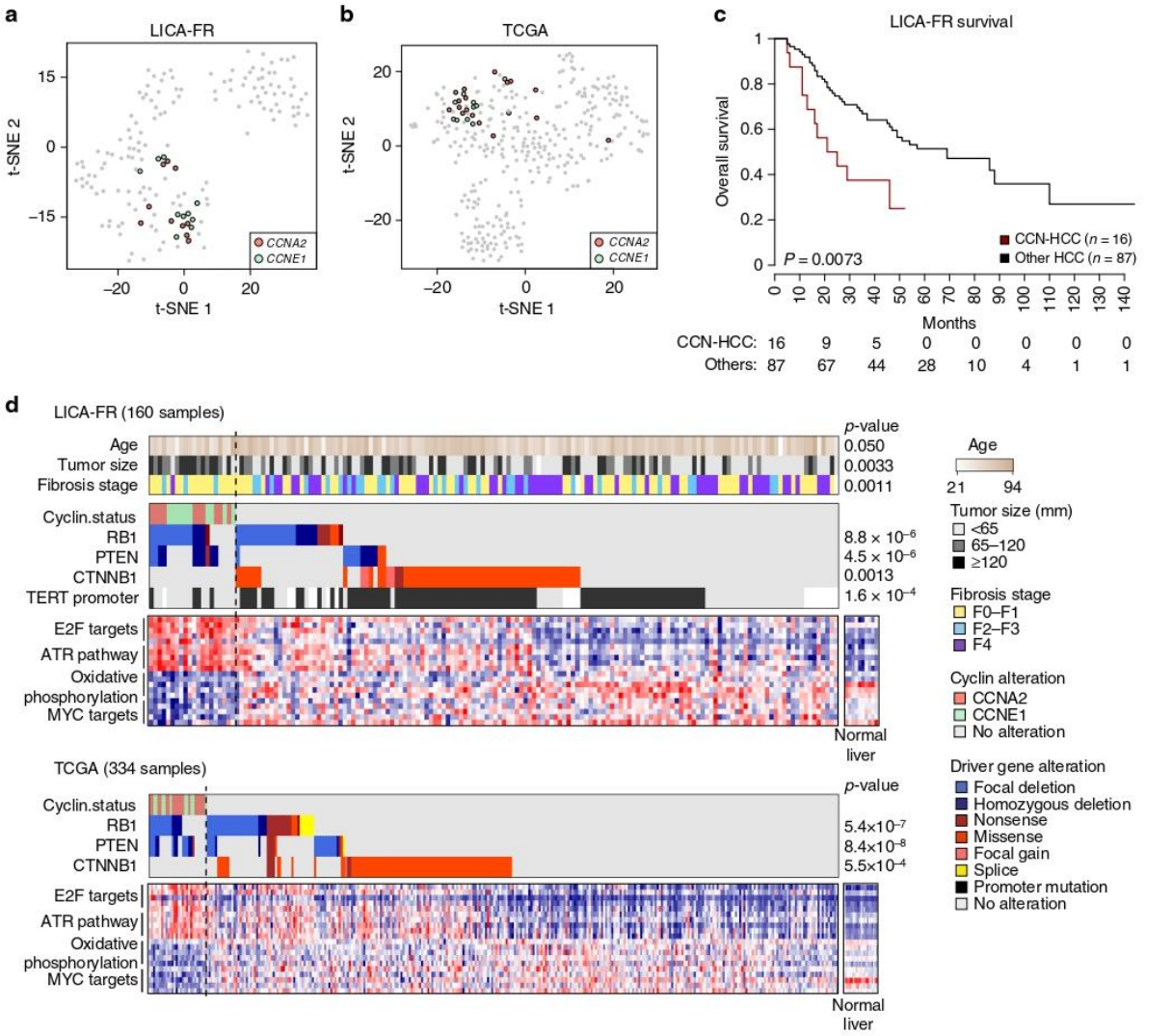


Fig. 3 Clinical and molecular features of cyclin-activated HCC. **a** t-SNE plots depicting the classification of HCC from the LICA-FR and TCGA cohorts based on their transcriptional profiles. Tumors harboring *CCNA2* or *CCNE1* activating alterations are indicated with a color code. **b** Clinical characteristics, driver genes and deregulated pathways associated with CCN-HCC in the LICA-FR (top) and TCGA (bottom) cohorts. **c** Overall survival in CCN-HCC as compared with other HCC in the LICA-FR cohort. Only HCC with curative resection (R0) were included

different combinations of rearrangement categories defined according to the type, size, and clustered nature of rearrangements (Fig. 4a). Strikingly, a high number of rearrangements attributed to signature RS1 (≥50 events) was specifically encountered in a cluster of 13 tumors corresponding exactly to CCN-HCC ($P = 1.4 \times 10^{-11}$, Fig. 4b). We validated this association using WGS data from the ICGC-JP series and a subset of 48 samples from the TCGA series (Fig. 4c, Supplementary Data 7). In absence of WGS data for the rest of the TCGA series, we used SNP array data to estimate the number of focal gains (<200 kb) in each tumor as a surrogate marker of the RS1 signature. With a median of 120 events, CCN-HCC displayed a significant increase of focal gains as compared with other HCC in the TCGA series (median = 6, $P < 2.2 \times 10^{-16}$, Supplementary Fig. 5). Thus, CCN-HCC have a relatively low mutation burden but a large number of structural rearrangements with a specific signature.

RS1 features suggest a replication stress-induced mechanism. Almost all rearrangements in CCN-HCC belonged to signature RS1, characterized by a combination of small tandem duplications (<100 kb) and inter-chromosomal translocations (Fig. 4d). CCN-HCC also displayed a typical copy-number profile showing hundreds of focal gains, usually one copy above surrounding chromosome segments (Supplementary Fig. 6). Surprisingly, overlaying structural rearrangement breakpoints with copy-number profiles revealed that only 68% of these gains were due to tandem duplications, other gains being frequently surrounded by translocation or inversion breakpoints (Fig. 4e, Supplementary Fig. 6). A recurrent feature consisted of several chromosome segments, usually between 10 and 100 kb, stung together and with the same duplication level relative to their source chromosomes. Most of these events involved segments from two (Fig. 4f) or more (Supplementary Fig. 7) different chromosomes, a feature recently described as templated

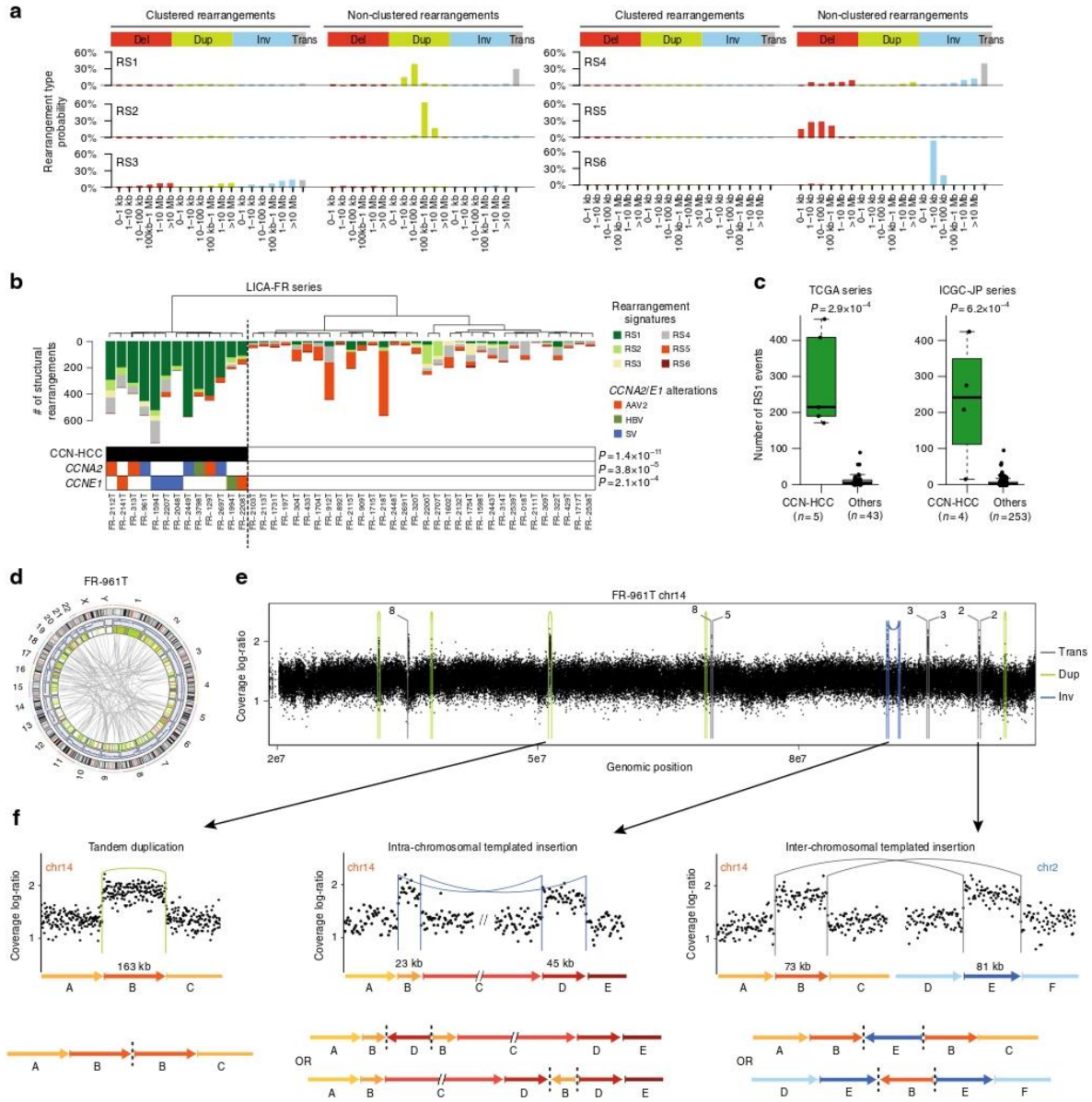


Fig. 4 Cyclin-activated HCC display a specific signature or structural rearrangements. **a** Six rearrangement signatures identified across 350 HCC genomes by non-negative matrix factorization. Structural rearrangements were classified in 38 categories considering their type (del: deletion, dup: tandem duplication, inv: inversion, trans: inter-chromosomal translocation) and size, and distinguishing clustered from non-clustered events. The probability of each rearrangement category in each signature is represented, with rearrangement types indicated above and rearrangement sizes below. **b** Unsupervised classification of 45 HCC from the LICA-FR cohort based on the contribution of rearrangement signatures in each tumor. Significant molecular alterations associated with the cluster of tumors having a high contribution of signature RS1 are represented below. *P*-values were obtained using Fisher's exact tests. **c** Validation of the association between the RS1 signature and CCN-HCC in the TCGA and ICGC-JP series. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range. **d** CIRCOS plot representing the structural rearrangement profile of a representative CCN-HCC (FR-961T, harboring a *GSTCD-CCNA2* fusion). **e** Copy-number profile showing the accumulation of focal gains along chromosome 14 in tumor FR-961T. Structural rearrangements are overlaid on the copy-number profile with a color code indicating the type of event. trans: inter-chromosomal translocation; dup: tandem duplication; inv: inversion. **f** Three types of rearrangements leading to focal chromosome gains in CCN-HCC. A representative example of each type of event is shown with a copy-number plot above and a schematic representation of the rearranged chromosome below. Structural rearrangements are represented with the same color code as in **e**. Dashed lines on schematic chromosome reconstructions represent the abnormal junctions detected in WGS data

insertion cycle[23]. Inter-chromosomal templated insertions accounted for 11% of focal gains in CCN-HCC. Other events, which we call intra-chromosomal templated insertions, involved distal segments of a same chromosome and appeared as couples of inversions (Fig. 4f) or duplication and deletion (Supplementary Fig. 7), depending on the orientation of the junctions. Intra-chromosomal templated insertions accounted for 7% of focal gains in CCN-HCC. All these events are consistent with a replication-based mechanism in which a DNA polymerase at a stalled replication fork would switch template, replicate one or more other DNA regions and switch back to the original template strand behind the point of departure, generating a duplication on the host chromosome[23–26]. Such mechanism could be particularly active in CCN-HCC due to replication stress induced by premature S phase entry.

Structural rearrangements activate *TERT* promoter in CCN-HCC. To better understand the functional consequences of the rearrangement phenotype observed in CCN-HCC, we examined the location of 8466 breakpoints attributed to the RS1 signature among the 350 liver cancer genomes from the LICA-FR, TCGA and ICGC cohorts. RS1 breakpoints were not distributed evenly along the genome but formed clusters located almost exclusively within active topologically associated domains (TADs, Fig. 5a) characterized by early replication, high gene expression and active chromatin states in normal liver (Fig. 5b). In particular, RS1 breakpoint hotspots were frequently observed at loci encoding very highly expressed liver enzymes exemplified by the albumin (*ALB*), alcohol dehydrogenase (*ADH*) and hydroxysteroid 17-Beta dehydrogenases (*HSD17B*) loci on chromosome 4 (Fig. 5a, Supplementary Fig. 8). Among the 18 chromatin states defined by the ROADMAP consortium in normal adult liver, active transcription start sites (TSS) and enhancer regions were the most strongly enriched in RS1 breakpoints (fold-change > 3), whereas quiescent and heterochromatin domains were the most depleted (Fig. 5c). TSS and enhancer regions were also enriched, to a lesser extent, in breakpoints related to signature RS2 characterized by large tandem duplications. By contrast, breakpoints related to signature RS6, dominated by inversions < 10 kb, were predominantly observed in heterochromatin and ZNF repeats.

We then used binomial regression[27] to model the density of rearrangement breakpoints along the genome considering an extensive set of genomic features (Supplementary Fig. 9) and to identify hotspots harboring more breakpoints than expected by chance from the background model, which may indicate positive selection in CCN-HCC. We identified a single significant locus corresponding to *TERT* promoter region (q=0.0029, Fig. 5d). Although *TERT* promoter mutations were rare in CCN-HCC (9 vs. 55% in others, $P = 2.4 \times 10^{-5}$), *TERT* promoter rearrangements were highly enriched (82 vs. 7%, $P = 1.8 \times 10^{-15}$, Fig. 5e) and involved regions of active chromatin in normal liver, in the vicinity of highly expressed liver enzymes (*ALB*, *FGG*, *SEP15*, *SLC12A7* and *BAAT*) or transcription factors (*HNF4A*, *CEBPA*, and *CEBPB*) (Supplementary Data 8, Supplementary Fig. 10). *TERT* promoter rearrangements induced an overexpression of *TERT*, stronger than promoter mutations but lower than HBV insertions (Supplementary Fig. 11). Of the 18 *TERT* promoter rearrangements identified in CCN-HCC, 16 could be associated with signature RS1 with a probability ≥ 0.5 (Fig. 5f). By contrast, most *TERT* promoter rearrangements in other HCC were related to signature RS4. Thus, structural rearrangements induced by replication stress are enriched at active chromatin regions and can promote CCN-HCC development by activating oncogenes like *TERT*.

CCN-HCC share a similar signature with *BRCA1*-altered cancers. To investigate the prevalence of the RS1 signature in other cancer types, we applied our method to 2606 tumors from the ICGC PanCancer Analysis of Whole Genomes (PCAWG) dataset[23,28,29]. In this pan-cancer series, we identified 9 rearrangement signatures (Supplementary Fig. 12), including one signature (RS1-pancan) highly similar to the RS1 signature that we identified in liver cancers (cosine similarity = 0.91). The RS1-pancan signature was detected at low frequency in several cancer types (e.g. bladder, lung, esophageal and gastric cancers), and was highly active in breast (18% of samples with ≥ 50 RS1 events) and ovarian (33%) cancers. However, this signature was associated with *CCNA2/E1* rearrangements only in liver cancer (Fig. 6a, Supplementary Data 9). Thus, the relationship between cyclin A2/E1 activation and signature RS1 is specific to liver cancer, and the molecular cause of this signature in other cancer types remains to be elucidated. In ovarian and breast cancer, RS1 signature was not associated with *CCNE1* amplifications but with *BRCA1* inactivation (Fig. 6b, c), consistent with previous reports[30,31]. Despite sharing a common signature of short tandem duplications and templated insertions, *CCNA2*, *CCNE1* and *BRCA1*-altered tumors displayed slightly different characteristics. First, the number of RS1 rearrangements was higher in *CCNA2*-activated HCC (median = 269) than in *CCNE1*-activated HCC (137) and *BRCA1*-altered breast (132) and ovarian (159) cancers (Fig. 6d). Second, tandem duplications were larger in *CCNE1*-activated HCC (median = 39 kb) than in *CCNA2*-activated HCC (22 kb), and smaller in *BRCA1*-altered breast (9 kb) and ovarian (10 kb) cancers (Fig. 6e). Finally, duplication and translocation breakpoints were strongly enriched in early-replicated regions in CCN-HCC as compared with other HCC, but not in *BRCA1*-altered as compared with other breast and ovarian cancers (Fig. 6f). Cyclin E1 activation was recently shown to induce replication stress by firing novel replication origins located within highly transcribed genes and prone to collapse[32]. BRCA1 is implicated in the response to replication stress[33,34] and its inactivation leads to tandem duplication formation at stalled forks by a replication restart-bypass mechanism[35]. Cyclin A2/E1 activation in HCC and *BRCA1* inactivation in breast and ovarian cancers may thus converge towards a similar rearrangement signature, with specificities reflecting the different ways by which these genetic alterations induce replication stress or modulate response to it (Fig. 6g).

Discussion

Here, we report the characterization of a homogeneous HCC subgroup driven by the activation of *CCNA2* or *CCNE1* gene. CCN-HCC represent 7% of HCC across the 3 data sets analyzed here, but up to 14% of HCC developed in a non-fibrotic liver. These patients often have atypical clinical presentation, without any history of primary risk factors, and can be remarkably young, exemplified by tumor FR-3880T developed in a 32 year-old woman without any risk factor, due to a *TET2-CCNA2* fusion. CCN-HCC are usually large tumors of poor prognosis but share molecular characteristics, in particular high proliferation and replication stress, that could provide therapeutic opportunities[36]. First, conventional chemotherapies mainly affect actively dividing cells by generating DNA damage or blocking DNA replication, and the tandem duplicator phenotype was identified as a marker for chemotherapeutic response in breast cancer cell lines and patient-derived xenografts[37]. Transarterial chemoembolization (TACE) with doxorubicin, cisplatin or epirubicin, usually recommended for patients with intermediate HCC not eligible for surgery, may thus be an interesting option for CCN-HCC. Poly (ADP-ribose) polymerase (PARP) inhibitors, the first clinically

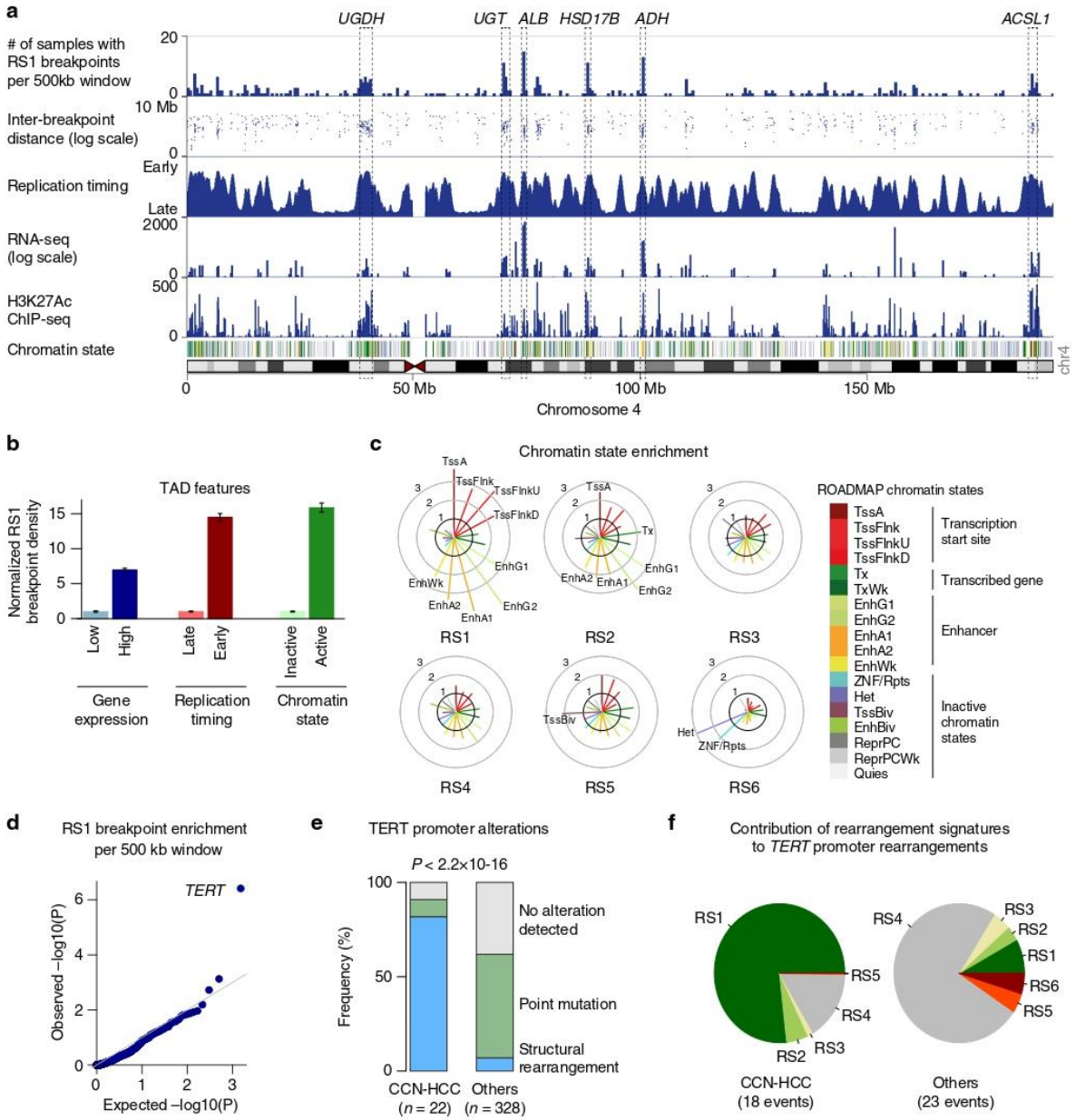
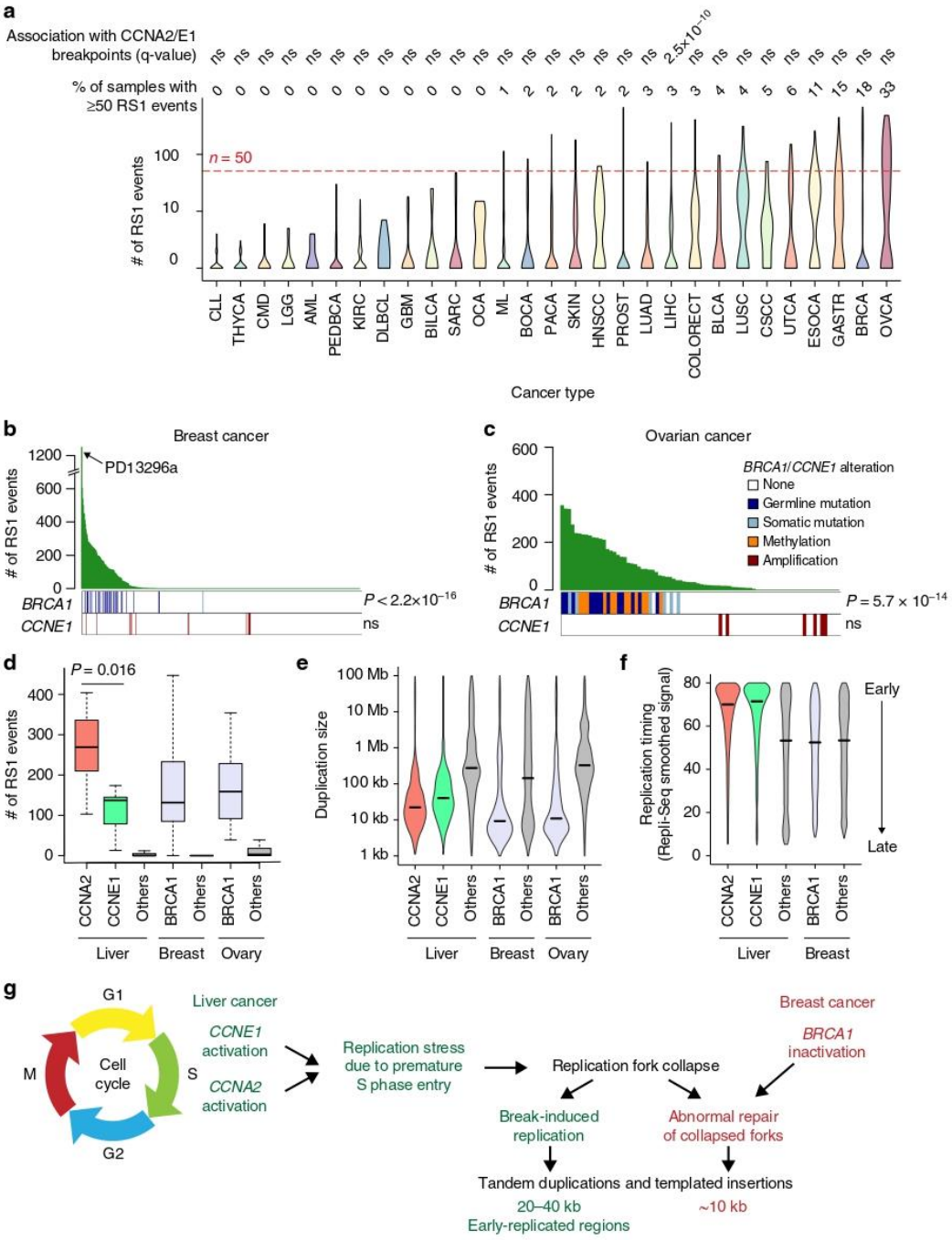


Fig. 5 Hotspot analysis of rearrangement signature 1 (RS1) breakpoints. **a** The density of RS1 breakpoints along chromosome 4 is displayed above replication timing, RNA-seq expression, H3K27Ac ChIP-seq profile and chromatin state. Replication timing was determined using Repli-Seq data from the liver cancer cell line HepG2. RNA-seq profile was generated from a normal liver sample. H3K27Ac and chromatin states in normal adult liver were obtained from the ROADMAP consortium. The legend for chromatin state color codes is displayed in **c**. Hotspots corresponding to highly expressed liver enzymes are annotated (UGDH, UDP-glucose 6-dehydrogenase; UGT, UDP glucuronosyltransferase family cluster; ALB, albumin; HSD17B, hydroxysteroid 17-Beta dehydrogenases 11 and 13, ADH, alcohol dehydrogenase cluster; ACSL1, acyl-CoA synthetase long chain family member 1). **b** RS1 breakpoint density in topologically associated domains (TADs). TADs were defined in human embryonic stem cells (H1) and classified based on gene expression in normal liver, replication timing and chromatin state. For each comparison, breakpoint density was normalized to be 1 in the group with the lowest density. Error bars indicate the 95% confidence interval. **c** Enrichment of rearrangement breakpoints in ROADMAP chromatin states for the 6 rearrangement signatures identified in HCC. For each signature, the fold-change between the observed and expected number of breakpoints falling within each chromatin state is represented, and chromatin states with a >2-fold enrichment are annotated. **d** Quantile-quantile plot of RS1 breakpoint enrichment p-values across 500 kb windows. **e** Proportion of *TERT* promoter alterations in CCN-HCC and other HCC analyzed by WGS. **f** Contribution of the 6 rearrangement signatures to *TERT* promoter rearrangements in CCN-HCC and other HCC

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07552-9 ARTICLE



approved drugs designed to exploit synthetic lethality, have demonstrated benefit for patients carrying *BRCA1* mutations[38]. CCN-HCC do not harbor a DNA repair defect but share with *BRCA1*-altered tumors a signature of genomic instability that could conceivably confer these tumors sensitivity to PARP inhibitors. Finally, there are currently several compounds in phase I and II trials targeting the replication stress response pathway members *ATR*, *CHK1* and *WEE1*[39]. If brought to the clinic, such compounds would be promising for CCN-HCC treatment, given that the ATR pathway is strongly upregulated in CCN-HCC and overexpression of *CCNE1* has been shown to confer increased sensitivity to ATR inhibition[40].

We describe for the first time recurrent fusions involving *CCNA2* gene and recurrent rearrangements of *CCNE1* promoter region. *CCNA2* fusions are only the second recurrent fusion event identified in hepatocellular carcinoma, after the *PRKACA-DNAJB1* fusion characteristic of the rare fibrolamellar carcinoma subtype[41]. These fusions always involve the untranslated 5′ region of different partner genes upstream exons 3–8 of *CCNA2*, which constitutes an original mechanism leading to oncogene activation

Fig. 6 Pan-cancer analysis of the RS1 signature **a** Violin plots representing the number of rearrangements attributed to signature RS1 across patients within each cancer type in the ICGC PCAWG data set. For each cancer type, we assessed the association between tumors with ≥ 50 RS1 events and tumors with a rearrangement breakpoint < 80 kb from *CCNA2* or *CCNE1* gene using Fisher's exact tests. ns: not significant. The definition of cancer codes and number of samples per cancer type are available in Supplementary Data 9. **b** Number of RS1 events across 524 breast cancer genomes[30] and association with *BRCA1* alterations and *CCNE1* amplifications. PD13296a, the only tumor with both *BRCA1* mutation and *CCNE1* amplification, has the highest number of RS1 events in the series. **c** Number of RS1 events across 80 ovarian cancer genomes[75] and association with *BRCA1* alterations and *CCNE1* amplifications. *P*-values were obtained using one-sided Wilcoxon rank-sum tests. **d** Number of RS1 events in liver, breast and ovarian cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range. **e**, Violin plots representing the distribution of tandem duplication sizes across liver, breast and ovarian cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. **f** Violin plots representing the replication timing of duplication and inter-chromosomal translocation breakpoint loci in liver and breast cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. Replication timing was determined using Repli-Seq data from the HepG2 cell line for liver cancer and from the MCF-7 cell line for breast cancer. **g** Proposed connexion beween rearrangement signatures in CCN-HCC and in *BRCA1*-inactivated breast and ovarian cancers

by truncating a regulatory N-terminal domain. Apart from liver cancers, none of the 2606 tumor genomes from the ICGC PCAWG dataset displayed a rearrangement breakpoint in *CCNA2* intron 2. Consistently, a recent RNA-seq analysis of 9,624 TCGA samples from 33 cancer types[42] did not reveal any *CCNA2* fusion in other cancer types. *CCNA2* fusions thus appear to be specific of liver cancers. Rearrangements affecting *CCNE1* promoter region result in the overexpression of cyclin E1 by bringing active enhancer regions upstream the transcription start site, mirroring the effect of viral enhancers. This mechanism was more frequent than *CCNE1* amplification in the liver cancer series we analyzed. Although HBV and AAV2 insertions were previously identified in *CCNA2* and *CCNE1*[14,16], the functional consequences of these insertions were unknown. By integrating WGS and RNA-seq data, we demonstrate here that viral insertions in *CCNA2*, like gene fusions, induce abnormal transcripts leading to truncated proteins lacking N-terminal regulatory domains. By contrast, viral insertions in *CCNE1* region lead to the overexpression of a full-length transcript and protein.

CCN-HCC display a characteristic transcriptional program, with a strong overexpression of E2F targets. Activation of the E2F pathway is expected in *RB1*-altered tumors and was already described in HCC[43]. However, E2F pathway is also activated in CCN-HCC without *RB1* inactivation event and may be partly explained by the ability of cyclin E/Cdk2 complexes to phosphorylate Rb. Interestingly, E2F-1 overexpression in the liver causes dysplasia and tumors in mice[43], and E2F1 was shown to inhibit c-Myc-driven apoptosis by activating PIK3CA/Akt/mTOR and c-Myb/COX-2 pathways[44].

A striking feature of CCN-HCC is the accumulation of hundreds of tandem duplications and templated insertion cycles. A recent study showed that *CCNE1* activation in U2OS cell lines leads to shortened G1 phase, early S phase entry and firing of normally silenced replication origins in highly expressed genes, prone to collapse and associated with DNA double-strand breaks formation[32]. Double-strand breaks formed following replication fork breakdown are primarily repaired by break-induced replication (BIR)[45]. In a cyclin E overexpression model of DNA replication stress, BIR was shown to be required for cell cycle progression and to induce duplications < 200 kb[46]. In addition, template switching may occur during BIR and generate complex chromosome rearrangements[24,25,47]. Thus, the nature of rearrangements identified in CCN-HCC and the enrichment of breakpoints in early-replicated, actively transcribed regions are consistent with a BIR mechanism induced by replication stress. However, future studies addressing the precise molecular mechanism generating templated insertions will be crucial to fully understand the relationship between replication stress and the RS1 rearrangement signature. The mechanism of tandem duplication formation in *BRCA1*-mutant cells was recently identified[35]. It involves abnormal repair of collapsed replication forks by a "replication restart bypass" mechanism with extension of the stalled leading strand by a migration bubble mechanism similar to BIR[48], terminated by end joining or by microhomology-mediated template switching. Thus, structural rearrangements induced by cyclin activation and *BRCA1* deficiency are initiated by replication fork collapse and processed by different repair mechanisms leading to a similar rearrangement signature with subtle differences regarding the size of rearrangements and breakpoint location. Interestingly, *BRCA1* inactivation and *CCNE1* amplification are mutually exclusive in ovarian cancers[49], and have been shown to be synthetically lethal[50]. The single breast tumor that we identified with both *BRCA1* mutation and *CCNE1* amplification (PD13296a) had the highest number of rearrangements related to the RS1 signature ($n = 1221$) across all the tumors we analyzed.

Contrary to *CCNA2* alterations that seem to be specific of liver cancers, *CCNE1* activation by high-level amplification is frequent across human cancers, in particular in gynecologic cancers[51]. Yet, *CCNE1* amplification in breast and ovarian cancers does not lead to the rearrangement phenotype that we observed in CCN-HCC. Several reasons may explain this discrepancy. First, adult hepatocytes are quiescent, rarely divide, and may thus be particularly sensitive to replication stress. Second, *CCNE1* is mostly activated by viral insertions and structural rearrangements of regulatory regions in HCC, rather than chromosome amplifications. These alterations may not have exactly the same functional consequence. Finally, we believe that viral insertions and structural rearrangements activating *CCNA2* or *CCNE1* are early events triggering hepatocarcinogenesis because they occur in patients without cirrhosis and in absence of other oncogenic event like *CTNNB1* mutations. *CCNE1* amplifications may occur later in breast and ovarian tumors, not leaving enough time for rearrangements to accumulate. Fujimoto et al. reported a positive correlation between the number of structural rearrangements and HBV insertion sites, suggesting that double-strand breaks generated by structural rearrangements may provide opportunities for HBV integration[11]. Here we describe the reciprocal relationship where viral insertions in cyclin genes lead to structural rearrangement formation due to replication stress.

The propensity of signature RS1 breakpoints to occur in enhancer-rich regions makes these rearrangements likely to activate oncogenes in trans. In this limited series of 22 CCN-HCC analyzed by WGS, we identified a single significantly recurrent hotspot at *TERT* promoter. However, the power to identify recurrent somatic rearrangement hotspots increases sharply with sample size[27], and future studies of larger CCN-HCC series may uncover additional sites under positive selection in CCN-HCC.

In conclusion, viral insertions and structural rearrangements activating *CCNA2* and *CCNE1* define a homogeneous subgroup of aggressive HCC developed in non-cirrhotic liver, sharing similar transcriptional profiles and frequent inactivation of *RB1*

and *PTEN*. These tumors display a specific rearrangement signature induced by replication stress that sustains tumor growth by activating *TERT* but may constitute a targetable vulnerability.

Methods

Description of the LICA-FR cohort. A series of 160 hepatocellular carcinoma (HCC) samples and their non-tumor counterparts were collected from patients surgically treated in four French hospitals located in Bordeaux and Paris region. The study was approved by institutional review board committees (CCPRB Paris Saint-Louis, 1997, 2004, and 2010, approval number 01–037; Bordeaux, 2010-A00498–31). Written informed consent was obtained in accordance with French legislation. All samples were immediately frozen in liquid nitrogen and stored at −80 °C. HCC were enriched in cases developed on a non-cirrhotic liver (107/160, 67%): 75 tumors developed in non-fibrotic (METAVIR F0-F1), 32 in chronic hepatitis (F2–F3) and 53 in cirrhotic liver (F4). Clinicopathological data were available for all cases. A diversity of risk factors were represented in our series, including alcohol ($n = 63$), metabolic syndrome ($n = 37$), HBV ($n = 30$), and HCV infection ($n = 30$). Twenty-nine patients had none of the above risk factors. These 160 samples were analyzed by RNA sequencing, 156 were analyzed by whole exome sequencing (including 96 were previously published[10]) and 45 by whole genome sequencing (35 were previously published[22]). Detailed clinical characteristics and sequencing details for each sample are provided in Supplementary Data 1.

Whole genome sequencing. Whole genome data from 45 tumors of the LICA-FR series were analyzed in this study, comprising 35 previously published[22] and 10 new cases. The whole genomes of 10 new tumor/normal pairs were sequenced for this project at the Center National de Recherche en Génomique Humaine (CNRGH, Evry, France) on an Illumina HiSeq X Five as paired-end 151 bp reads. Sequences were aligned to the hg19 version of the human genome using BWA[52] version 0.7.12. We used Picard tools version 1.108 (http://broadinstitute.github.io/picard/) to remove PCR duplicates and GATK[53] version v3.5 for local indel realignment and base quality recalibration, as recommended in GATK best practices[54]. We obtained an average depth of 119-fold for tumors (range 104–126) and 41-fold for matched non-tumor liver samples (range 38–43).

Whole exome sequencing. Whole exome data from 156 tumors of the LICA-FR series were analyzed in this study, comprising 96 previously published[10] and 60 new cases. Sequence capture, enrichment and elution of genomic DNA samples from the 60 new tumor/normal pairs was performed by IntegraGen (Evry, France). Agilent in-solution enrichment was used with the manufacturer's biotinylated oligonucleotide probe library SureSelect Human All-Exon kit v5 + UTRs ($n = 39$) or SureSelect Clinical Research Exome V2 ($n = 21$) according to the manufacturer's instructions. The eluted enriched DNA sample was sequenced on an Illumina HiSeq 2000 ($n = 39$) or HiSeq 4000 ($n = 21$) as paired-end 75 bp reads. Sequencing details for each sample are indicated in Supplementary Data 1.

Somatic mutation calling. We used MuTect2 to call somatic mutations from WES and WGS data by comparing each tumor sample with its matched non-tumor counterpart and a panel of normals (PON) file. We excluded mutations belonging to the ENCODE Data Analysis Consortium blacklisted regions (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz) and regions covered by < 6 reads in the tumor or normal sample. We then selected only single nucleotide variants (SNVs) with a MuTect2 flag among "PASS", "clustered_events", "t_lod_fstar", "alt_allele_in_normal" or "homologous_mapping_event" and small insertions and deletions (indels) with a MuTect2 flag among "PASS", "clustered_events" or "str_contraction". To improve specificity in the calling of mutations with low variant allele frequency (VAF), we quantified the number of high quality variant reads in the tumor (mapping quality ≥ 20, base quality ≥ 20) and the number of variant reads in the non-tumor sample with no quality threshold using bamreadcount (https://github.com/genome/bam-readcount). Only variants matching the following criteria were finally retained: VAF ≥ 2% in the tumor with ≥ 3 variant reads, VAF ≤ 5% in the non-tumor samples with ≤ 2 variant reads, and a VAF ratio ≥ 5 between the tumor and non-tumor sample.

Copy-number and structural rearrangement analysis. We used MANTA[55] software to identify somatic structural rearrangements in WGS data. To keep only the most reliable events, we selected only rearrangements supported by ≥ 10 reads and with a variant allele fraction ≥ 5%. We used cgpBattenberg[56] algorithm to reconstruct copy-number profiles from WGS data. We used the circular binary segmentation algorithm implemented in the Bioconductor package DNAcopy[57] to reconstruct copy-number profiles from WES data.

RNA sequencing. RNA samples from the 160 tumors of the LICA-FR series were sequenced in several batches with slightly different protocols. RNA samples were enriched for polyadenylated RNA from 5 µg of total RNA, and the enriched samples were used to generate sequencing libraries with the Illumina TruSeq or Illumina TruSeq Stranded mRNA kit and associated protocol as provided by the manufacturer. Libraries were sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 2000 or 4000 as paired-end 75 or 100 bp reads. Full Fastq files were aligned to the reference human genome hg19 using TopHat2[58]. Sequencing details for each sample and the parameters used for TopHat2 are indicated in Supplementary Data 1. We removed reads mapping to multiple locations, and we used HTSeq[59] to obtain the number of reads associated to each gene in the Gencode v19 database, restricting to protein-coding genes, pseudogenes, antisense and lincRNAs ($n = 42540$). We used the Bioconductor DESeq2 package[60] to import raw HTSeq counts for each sample into R statistical software and apply variance stabilizing transformation (VST) to the raw count matrix. FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalizing the count matrix for the library size and the coding length of each gene. We used the area under the ROC curve (AUC) to identify and remove 2724 genes with a significant batch effect (AUC > 0.95 between one sequencing project and others).

Gene fusion detection. Fusions detected by TopHat2 (--fusion-search --fusion-min-dist 2000 --fusion-anchor-length 13 --fusion-ignore-chromosomes chrM) were filtered using the TopHatFusion-post algorithm. We kept only fusions validated by BLAST and with at least 10 split-reads or pairs of reads spanning the fusion event, and we removed fusions identified at least twice in a cohort of 36 normal liver samples.

Gene expression analysis. We used t-distributed stochastic neighbor embedding (t-SNE) to classify HCC based on their gene expression profiles. We selected the 1000 most variably expressed genes, and we used 1 minus the weighted Pearson correlation coefficient as the distance measure. Pairwise Pearson correlation was calculated using the wtd.cors function of the *weights* R package. We used standard deviation substracted by 0.2 as the weight, giving more variable genes greater influence. The resulting distance matrix was used to perform the t-SNE analysis using the R package *Rtsne*[61] with default parameters except the following: theta = 0, is_distance = T, pca = F, max_iter = 2000. We used the Bioconductor *limma* package[62] to test for differential expression between CCN-HCC and other HCC of all genes expressed in at least five samples (FPKM > 0). We applied a *q*-value threshold of ≤ 0.05 to define differentially expressed genes. We used an in-house adaptation of the GSEA method[63] to identify gene sets from the MSigDB v6 database overrepresented among upregulated and downregulated genes.

Viral insertion screening. AAV2 insertions had previously been screened by viral capture and whole exome sequencing in 83 tumors from the LICA-FR cohort[16]. We extended this screen to AAV2 and HBV insertions in all HCC from the LICA-FR cohort using RNA-seq and WES data. In the ICGC-JP cohort, AAV2 and HBV insertions had already been screened using WGS data and were provided by Fujimoto et al.[11] In the TCGA cohort, we screened AAV2 and HBV insertions using RNA-seq data from all tumors and WES data from 37 tumors showing viral reads or overexpression of *CCNA2* or *CCNE1* in RNA-seq data. For each tumor and matched normal sample, the sequence reads were mapped to the AAV2 (AF043303.1) and HBV (X02763, renumbered using the EcoR1 restriction site as the +1) reference genomes using BWA[52]. Read pairs with at least one read aligned on the virus were extracted using samtools[64], and aligned to a custom reference genome including human chromosomes and virus fasta sequences as pseudo-chromosomes. Tumors with ≥ 6 chimeric reads or read pairs aligned on both the human and viral genomes were further analyzed. All viral insertions were validated by visual inspection on IGV[65]. We used chimeric reads to identify insertion breakpoints at base resolution by mapping sequences on both sides of the junctions. Of the 12 LICA-FR tumors with viral insertions detected in *CCNA2* or *CCNE1*, 7 were previously analyzed by viral capture sequencing[16] and 3 were analyzed by whole genome sequencing. For these 10 tumors, we were able to extract reads covering the full length of the inserted viral genome and to reconstruct the complete human-virus-human chimeric sequence.

Consequences of cyclin A2 alterations on protein structure. All tumors from the LICA-FR series harboring AAV2 or HBV insertions in *CCNA2* were analyzed by WGS or viral capture[16] to determine the precise boundaries of viral insertion breakpoints. RNA-seq reads were then aligned on the reconstructed chimeric sequence with TopHat2[58], and we used Cufflinks v2.2.1[66] to identify and quantify the different transcripts. We used ElemeNT[67] to predict transcription initiation sites and Alamut Visual software (Interactive Biosoftware) to identify splicing signals on the chimeric DNA sequence. We used ATGpr[68] to identify translation initiation sites on abnormal transcripts resulting from viral insertion or gene fusions.

Western blot analysis of cyclin A2 and cyclin E1 proteins. Cell protein extracts were prepared using hot Laemmli buffer (50 mM Tris, pH = 6.8, 2% SDS, 5% glycerol, 2 mM DTT, 2.5 mM EDTA, 2.5 mM EGTA, Protease inhibitor cocktail complete MINI EDTA-free (Roche Applied Science), 1× HALT Phosphatase inhibitor (Perbio), 2 mM Na3VO4 and 10 mM NaF). Protein concentration was assessed using the BCA Protein Assay Kit (Pierce). Western blot analyses were

conducted using the following primary antibodies: CCNA2 N-ter (#211735, Abcam); CCNA2 C-ter (#32386, Abcam), CCNE1 (#33911, Abcam), and β-actin (#4967, Cell Signaling Technology) used as loading control. Proteins of interest were detected using an anti-rabbit IgG horseradish peroxidase–linked secondary antibody (#7074, Cell Signaling Technology) and the ECL Chemiluminescence Western Blotting Detection Kit (GE Healthcare), according to the provided protocol. Signal detection was performed using the ChemiDoc XRS system and the Image Lab software (Bio-Rad). All antibodies were used at 1:1000 dilution except secondary antibody, which was used at 1:2000.

Mutational and rearrangement signature analysis. We used the *Palimpsest* R package[69] to extract mutational and rearrangement signatures from WGS data. For point mutations, we quantified the contribution of the 10 mutational signatures referenced on the COSMIC website (https://cancer.sanger.ac.uk/cosmic/signatures) and described as operative in liver cancers (signatures 1, 4, 5, 6, 12, 16, 17, 22, 23, 24)[22] to each tumor genome. For structural rearrangements, we performed a de novo signature analysis across the 350 HCC genomes from the LICA-FR, TCGA and ICGC-JP datasets. We identified 6 rearrangement signatures that were very similar to the 6 signatures we previously obtained on a smaller dataset[22], except that the two initially described deletion signatures were now merged into signature RS5, and that a new signature emerged (RS6, dominated by inversions < 10 kb). We used *Palimpsest* to quantify the contribution of each signature to each tumor genome and to estimate the probability of each structural rearrangement being due to each process.

Identification of rearrangement hotspots. We identified 8466 breakpoints attributed to signature RS1 (probability > 0.5) across the 350 HCC genomes from the LICA-FR, TCGA and ICGC-JP datasets. To account for the uneven distribution of rearrangements in the genome, we then modeled the background distribution of breakpoints considering various genomic features as described by Glodzik et al.[27], with some modifications. In short, we divided the genome into 500 kb bins, and we characterized for each bin 17 genomic features likely to influence the density of rearrangements: replication timing in HepG2 cell line (ENCODE[70]), highly expressed (top 25%) and low-expressed (remaining 75%) genes in normal liver, average copy-number in the cohort, repetitive sequences (segmental duplications, ALU elements and other repeats), number of N bases in the reference genome, known fragile sites[71], chromatin staining, DNAse hyper-sensitive sites and 6 histone marks (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K27ac) in adult liver (ROADMAP[72]). All features were normalized to a mean of 0 and standard deviation of 1 across the bins. The total number of RS1 breakpoints were counted for each bin, and we used negative binomial regression to model the distribution of breakpoints according to the 17 normalized features. The model was trained across 4993 bins after removing bins containing validated cancer genes from the Cancer Gene Census[73] (https://cancer.sanger.ac.uk/census). For signature RS1, the most predictive features of a high breakpoint density were DNAse accessibility, H3K27 acetylation and early replication timing. We then used this model to estimate the expected number of breakpoints across 761 bins containing cancer genes, and we compared the number of observed breakpoints to the number of expected breakpoints using a one-sided binomial test. Finally, p-values were corrected for multiple testing using Benjamini-Hochberg procedure.

Chromatin state analysis. We used various genomic features to correlate with structural rearrangement density and to better understand the functional consequences of rearrangements. We used replication sequencing (Repli-seq) wavelet-smoothed signals downloaded generated by the ENCODE[70] consortium for the liver cancer cell line HepG2 to define early and late-replicating regions. We used ChIP-seq data for various histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K27ac) and chromatin states derived from these modifications in normal adult liver by the ROADMAP consortium[72]. Topologically associated domain (TAD) boundaries in human embryonic stem cells (H1) were provided by Tsirigos et al.[74]

Pan-cancer analysis of structural rearrangement signatures. Somatic structural rearrangements called by a uniform pipeline over 2,606 tumor genomes were downloaded from the ICGC PanCancer Analysis of Whole Genomes (PCAWG) project[23,28,29]. Using *Palimpsest*[69], we identified 9 rearrangement signatures in this data set, including one (RS1-pancan) very similar to the RS1 signature identified in CCN-HCC, and we quantified the contribution of each signature to each tumor genome. In each cancer type, we tested if the presence of ≥ 50 rearrangements attributed to signature RS1-pancan was associated with the presence of rearrangement breakpoints < 80 kb from *CCNA2* or *CCNE1* gene using Fisher's exact test. We analyzed two additional series of breast ($n = 524$)[30] and ovarian ($n = 80$)[75] cancer genomes to correlate the amount of RS1-pancan events with *CCNE1* amplifications and *BRCA1* alterations.

Clinical associations. We tested the association of CCN-HCC in the LICA-FR cohort with gender, age, etiology, liver fibrosis, Edmonson grade, and vascular invasion using Wilcoxon rank sum test for continuous variables, Fisher's exact test for binary variables and Chi square test for trend for categorical variables. We used log-rank test and Kaplan–Meier method to compare overall survival between CCN-HCC and others, considering only HCC with curative resection (R0) and excluding patients who died within 3 months after surgery.

Computing codes. The functions used to perform the signatures analysis and associated figures are available as an open-source R package, Palimpsest, available on Github: https://github.com/FunGeST/Palimpsest.

URLs. ICGC data portal, https://dcc.icgc.org/; COSMIC database, https://cancer.sanger.ac.uk/cosmic; ENCODE project, https://www.encodeproject.org; GENCODE v19, http://www.gencodegenes.org/releases/19.html; ROADMAP project, http://www.roadmapepigenomics.org; NCI GDC data portal, https://portal.gdc.cancer.gov.

Data availability

The sequencing data reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (RNA-seq accession [EGAS00001002879]; WES accessions [EGAS00001000217], [EGAS00001001002] and [EGAS00001003063]; WGS accessions [EGAS00001002408], [EGAS00001000706] and [EGAS00001002888]) and the Inter- national Cancer Genome Consortium (ICGC) data portal (http://dcc.icgc.org/; release 27, April 2018).

Received: 23 July 2018 Accepted: 8 November 2018
Published online: 07 December 2018

References

1. 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.* **56**, 908–943 (2012).
2. Llovet, J. M. et al. Sorafenib in advanced hepatocellular carcinoma. *New Engl. J. Med.* **359**, 378–390 (2008).
3. Bruix, J. et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **389**, 56–66 (2017).
4. Llovet, J. M. & Hernandez-Gea, V. Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. *Clin. Cancer Res.* **20**, 2072–2079 (2014).
5. Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* **2**, 16018 (2016).
6. Nault, J. C. et al. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* **60**, 1983–1992 (2014).
7. Guichard, C. et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
8. Fujimoto, A. et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
9. Ahn, S.-M. et al. Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* **60**, 1972–1982 (2014).
10. Schulze, K. et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
11. Fujimoto, A. et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
12. Zucman-Rossi, J., Villanueva, A., Nault, J.-C. & Llovet, J. M. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226–1239.e4 (2015).
13. Wang, J., Chenivesse, X., Henglein, B. & Bréchot, C. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. *Nature* **343**, 555–557 (1990).
14. Sung, W.-K. et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
15. Ding, D. et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet.* **8**, e1003065 (2012).
16. Nault, J.-C. et al. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187–1193 (2015).
17. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu & Cancer Genome Atlas Research Network. Comprehensive and

integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341.e23 (2017).
18. Geley, S. et al. Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J. Cell. Biol.* **153**, 137–148 (2001).
19. Fung, T. K., Yam, C. H. & Poon, R. Y. C. The N-terminal regulatory domain of cyclin A contains redundant ubiquitination targeting sequences and acceptor sites. *Cell Cycle* **4**, 1411–1420 (2005).
20. Sage, J., Miller, A. L., Pérez-Mancera, P. A., Wysocki, J. M. & Jacks, T. Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. *Nature* **424**, 223–228 (2003).
21. Garcia-Cao, I. et al. Systemic elevation of PTEN induces a tumor suppressive metabolic state. *Cell* **149**, 49–62 (2012).
22. Letouzé, E. et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
23. Li, Y. et al. Patterns of structural variation in human cancer. Preprint at https://www.biorxiv.org/content/early/2017/08/27/181339, https://doi.org/10.1101/181339 (2017).
24. Lee, J. A., Carvalho, C. M. B. & Lupski, J. R. A. DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
25. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, e1000327 (2009).
26. Carvalho, C. M. B. et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.* **43**, 1074–1081 (2011).
27. Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* **49**, 341–348 (2017).
28. Campbell, P. J., Getz, G., Stuart, J. M., Korbel, J. O. & Stein, L. D. Pan-cancer analysis of whole genomes. Preprint at https://www.biorxiv.org/content/early/2017/07/12/162784, https://doi.org/10.1101/162784 (2017).
29. Wala, J. A. et al. Selective and mechanistic sources of recurrent rearrangements across the cancer genome. Preprint at https://www.biorxiv.org/content/early/2017/09/14/187609, https://doi.org/10.1101/187609 (2017).
30. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
31. Waszak, S. M. et al. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. Preprint at https://www.biorxiv.org/content/early/2017/11/01/208330, https://doi.org/10.1101/208330 (2017).
32. Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112–116 (2018).
33. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
34. Pathania, S. et al. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **5**, 5496 (2014).
35. Willis, N. A. et al. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**, 590–595 (2017).
36. Forment, J. V. & O'Connor, M. J. Targeting the replication stress response in cancer. *Pharmacol. Ther.* https://doi.org/10.1016/j.pharmthera.2018.03.005 (2018).
37. Menghi, F. et al. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl Acad. Sci. USA* **113**, E2373–E2382 (2016).
38. Lord, C. J. & Ashworth, A. PARP inhibitors: synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
39. O'Connor, M. J. Targeting the DNA damage response in cancer. *Mol. Cell* **60**, 547–560 (2015).
40. Toledo, L. I. et al. A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations. *Nat. Struct. Mol. Biol.* **18**, 721–727 (2011).
41. Honeyman, J. N. et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science* **343**, 1010–1014 (2014).
42. Gao, Q. et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* **23**, 227–238.e3 (2018).
43. Conner, E. A. et al. Dual functions of E2F-1 in a transgenic mouse model of liver carcinogenesis. *Oncogene* **19**, 5054–5062 (2000).
44. Ladu, S. et al. E2F1 inhibits c-Myc-driven apoptosis via PIK3CA/Akt/mTOR and COX-2 in a mouse model of human liver cancer. *Gastroenterology* **135**, 1322–1332 (2008).
45. Kramara, J., Osia, B. & Malkova, A. Break-induced replication: the where, the why, and the how. *Trends Genet.* **34**, 518–531 (2018).
46. Costantino, L. et al. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* **343**, 88–91 (2014).
47. Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007).
48. Saini, N. et al. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389–392 (2013).
49. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
50. Etemadmoghadam, D. et al. Synthetic lethality between CCNE1 amplification and loss of BRCA1. *Proc. Natl Acad. Sci. USA* **110**, 19489–19494 (2013).
51. Sanchez-Vega, F. et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
54. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
55. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
56. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
57. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
58. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
59. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
61. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
62. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
63. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
64. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
66. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
67. Sloutskin, A. et al. ElemeNT: a computational tool for detecting core promoter elements. *Transcription* **6**, 41–50 (2015).
68. Nishikawa, T., Ota, T. & Isogai, T. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics* **16**, 960–967 (2000).
69. Shinde, J. et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty388 (2018).
70. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
71. Bignell, G. R. et al. Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
72. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
73. Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
74. Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
75. Patch, A.-M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).

Acknowledgements

We thank Hidewaki Nakagawa for fruitful discussions and providing *TERT* promoter mutation data for the ICGC-JP series. We thank Rameen Beroukhim and Joachim Weischenfeldt for helping us access ICGC structural variant tables, and Aristotelis Tsirigos for providing topologically associated domain boundaries. We thank Tatiana Popova and Céline Vallot for critical discussion of the results. We thank the principal investigators of the liver cancer TCGA (Lewis Roberts, David Wheeler) and ICGC-JP (Tatsuhiro Shibata, Hidewaki Nakagawa) projects, and the ICGC consortium as a whole for providing the high quality data sets used in this study. We thank all the clinician surgeons and pathologists who have participated to this work: Jean Saric, Christophe Laurent, Laurence Chiche, Brigitte Le Bail, Claire Castain (CHU Bordeaux), Alexis Laurent, Daniel Cherqui, Daniel Azoulay (CHU Henri Mondor, Créteil,

APHP), Marianne Ziol, Nathalie Ganne-Carrié and Pierre Nahon (Jean Verdier Hospital, Bondy, APHP). We also thank the Réseau national CRB Foie (BB-0033-0085), the tumor banks of CHU Bordeaux (BB-0033-00036), Jean Verdier Hospital (APHP) and CHU Henri Mondor (APHP) for contributing to the tissue collection. This work was supported by INCa within the ICGC project, MUTHEC project (INCa translationnel PRTK2014), France Génomique, Cancéropole Ile de France (ExhauTrans project), ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan ("HTE program-HetColi network" and "Cancer et environnement program"), BPI France (ICE project), ANRS and the French Liver Biobanks network – INCa, BB-0033-00085, Hepatobio bank. The group is supported by the Ligue Nationale Contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir), Coup d'Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM and Fondation Mérieux. QB and LM are supported by a fellowship from the HOB doctoral school and the ministry of Education and Research, TLB is supported by an "Attractivite IDEX" fellowship from IUH and CP is supported by a doctoral fellowship funded by ANRS.

Author contributions

J.Z-R. and E.L. conceived and directed the research. Q.B., L.M., C.P., S.I., J.Z-R., and E.L. designed the study and wrote the manuscript. C.P., I.M., G.C., and T.L-B. performed the experiments. Q.B., L.M., C.P., I.M., T.L-B., S.I., J.Z-R., and E.L. analyzed and interpreted the data. D.B., V.M., and J-F.D. generated whole-genome sequencing data. Q.B., L.M., C. P., V.R., J.S., E.T., D.B., V.M. S.I., and E.L. performed bioinformatics and statistical analysis. J-C.N., G.A., A.D-V, P.B-S., O.S., J-F.B., and J.C. provided essential biological resources and collected clinical data. All authors approved the final manuscript and contributed to critical revisions to its intellectual context.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07552-9.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

IV DISCUSSION

1. Identification of new driver structural rearrangements

Using innovative approaches integrating WGS and RNAseq data, I was able to unravel several new driver alterations, both in HCA and HCC. These alterations are the result of genomic rearrangements and can be separated in two categories. First, I identified gene fusions that alter the protein structure of specific driver genes. Second, I was able to highlight "regulatory rearrangements" events, whose alteration involve a change in transcription level or a change in post-transcriptional regulation.

1.1. Gene fusions altering protein structure

During my masters 2 internship, I set up a pipeline to identify gene fusions from the large RNAseq collection of HCC and HCA tumor samples available in the lab. This work allowed me to characterize new recurrent fusions in HCA and HCC during my PhD.

First I found recurrent fusions involving *CCNA2* or *FRK* oncogenes downstream various partner genes, respectively at the origin of CCN-HCC and inflammatory HCA development. The particularity of these fusions is that they involve the loss of regulatory protein domains, leading to more stable or constitutively active proteins. In the case of *CCNA2* fusions, chimeric transcripts only include untranslated regions of partner genes upstream the exons 3 to 8 of *CCNA2*. *In silico* identification of translation starting sites indicated an alternative Kozak sequence at Methionine 158 of *CCNA2*, leading to the accumulation of a truncated *CCNA2* protein lacking negative regulatory domains, which was confirmed by Western blot. *CCNA2* protein accumulation leads to premature S phase entry and replicative stress characteristic of CCN-HCC tumors. In IHCAs, *FRK* fusions also involve different partner genes in 5'. Chimeric transcripts contain translated regions from the partner genes but lack *FRK*'s SH2 and SH3 regulatory domains that auto-inhibit the protein kinase domain of *FRK*. The oncogenic effect of those fusions is thus induced by the lack of regulatory domains of *FRK* resulting in a constitutively activated protein able to trigger JAK/STAT pathway. Consequently, expression data showed that in both *CCNA2* and *FRK* fusions, the level of the transcript is not significantly altered in tumors harboring respective gene fusions.

Gene fusions can also reveal inactivating SVs leading to the disruption of tumor suppressors. Recently, I contributed to the characterization of a new HCC subgroup harboring mixed features with fibrolamellar carcinomas (FLC), a rare histological subtype

of HCC (Hirsch *et al.*, 2020) (Article 5: Annex 3). This subgroup is similar to FLC, enriched in fibrotic tumors without underlying liver disease. However, whereas FLC tumors are triggered by the specific *PRKACA-DNAJB1* fusions, this subgroup was characterized by bi-allelic *BAP1* (BRCA1-associated protein 1) inactivation. Most *BAP1* alterations were mutations, but I also identified two tumors with gene fusions involving *BAP1*. The first tumor harbored an inversion leading to an out of frame fusion between *DAG1* (exons 1-2) and *BAP1* (exons 13-17). The second tumor harbored a complex balanced translocation involving 3 regions of chromosomes 2 and 3, resulting in i) a frame-shift fusion of *RAB3GAP1* (exons 1-23) with *BAP1* (exons 5-17) and ii) a truncated exons 1 to 3 of *BAP1*, placed upstream *LINC01460* gene (Article 5: Annex 3 Sup.Fig.6). Thus, both rearrangements lead to the inactivation of *BAP1*. Importantly, BAP1-HCC are older and with a poorer prognosis than FLC tumors, so the identification of this subgroup has clinical implications.

1.2. Gene fusions inducing the overexpression of oncogenes

In addition to protein structure, gene fusions can alter the expression of partner genes, in particular when their baseline expression is very different. During my PhD, I identified recurrent fusions involving highly expressed genes upstream oncogenes, leading to their massive overexpression.

In IHCA, *ROS1* gene is recurrently activated by this mechanism. *ROS1*, encoding a receptor tyrosine kinase activating several signaling pathways related to cell proliferation, is normally poorly expressed in hepatocytes. I identified recurrent fusions involving various highly expressed liver genes (*APOB* n=1; *PLG* n=5, *RBP4* n=2), upstream exons 33-35 of *ROS1*. These events induce the high expression of a chimera containing exons of both genes. However, as the predicted transcript is in phase, the kinase domain of ROS1 protein is still functional. In addition to the increased expression, *ROS1* is a transmembrane protein, and the relocation of the protein from the membrane to the cytoplasm could also be involved in the constitutive activation of the JAK/STAT pathway in these tumors.

I also identified one case of CCN-HCC resulting from a gene fusion between the highly expressed *ERRFI1* gene and *CCNE1*. Because the chimeric transcript only includes untranslated exon 1 of *ERRFI1*, it will be translated using the same Kozak starting site as the wild-type *CCNE1* transcript, leading to an overexpression of the normal cyclin E1

protein. This fusion is an example of *promoter hijacking*, like the recurrent *INHBE-GLI1* fusions identified earlier in the lab (Nault *et al.*, 2017), where only the promoter and 5'UTR of a highly expressed genes is fused to an oncogene, leading to the overexpression of the full length oncogenic protein, without any part of the 5' partner gene.

1.3. Structural rearrangements altering regulatory sequences

Although gene fusions identification from RNAseq is a powerful tool in cancer genomics, some structural variations affect only the regulatory sequences of cancer genes, without the existence of a chimeric transcript. By integrating WGS and RNAseq data, as well as epigenetic features from public databases, I could identify new driver events involving regulatory sequences in liver cancers.

In CCN-HCC, I identified recurrent enhancer hijacking events leading to the activation of *CCNE1*. Translocations lead to juxtapose *CCNE1* promoter with enhancer-rich chromatin areas located close to highly expressed liver genes (*RAPH1*, *CYB5A* and *ERGIC1*), leading to massive over-expression of *CCNE1* transcript. In addition, replication stress-induced structural rearrangements, in particular templated insertion cycles (T.I.C.), lead to frequent enhancer hijacking of *TERT* as secondary events in CCN-HCC. *TERT* enhancer hijacking was demonstrated using an integrated analysis of SV breakpoints from WGS, with the annotation of chromatin features such as H3K27ac chromatin immunoprecipitation sequencing (ChIP-Seq) signal and chromatin states in adult liver (Roadmap Epigenomics Consortium *et al.*, 2015) (Article 4: Sup.Fig.10). All tumors with *TERT* enhancer hijacking display a massive overexpression of the transcript (Article 4: Sup.Fig.11).

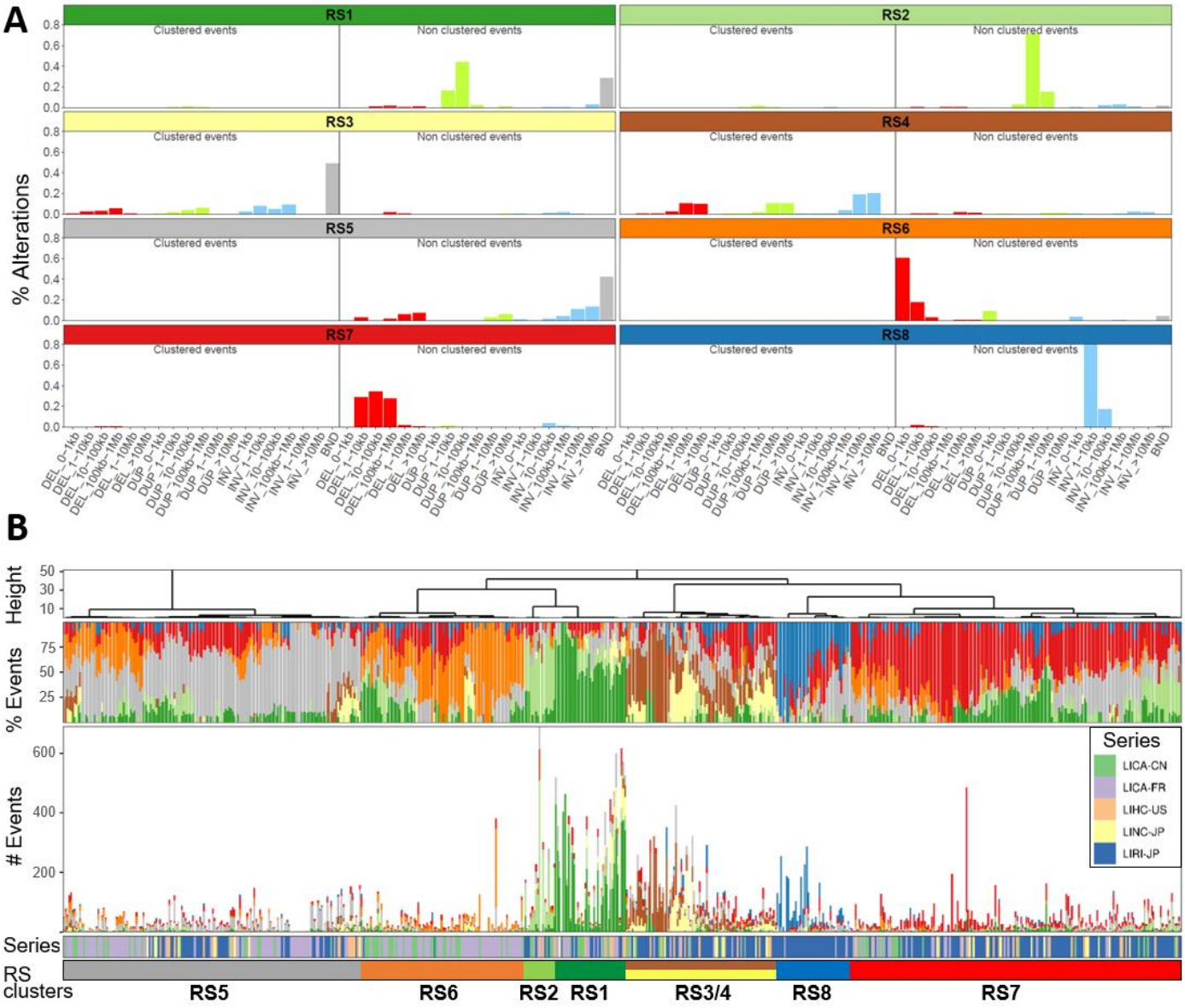
The last level of transcript expression regulation is post-transcriptional. In inflammatory adenomas, I identified a new mechanism of JAK/STAT pathway activation. In two IHCA characterized by a massive overexpression of *IL6* transcript, WGS revealed structural rearrangements (one inversion, one deletion) leading to the loss of regulatory sequences in the 3'UTR region of *IL6.* The lost 3'UTR part in the 2 tumors contain multiple mRNA-destabilizing elements, highly conserved and responsible for the short half-life of *IL6* transcripts (Paschoud *et al.*, 2006). Moreover, in murine models, those regions were shown to be targeted by Zc3h12a, an RNase responsible of immune response modulation by regulating mRNA decay (Matsushita *et al.*, 2009). Disruption of *IL6* 3'UTR region

induces a tremendous overexpression that is the consequence of impaired degradation of the transcript rather than increased transcription.

To conclude, this work allowed us to update the landscape of driver alterations in liver cancer with various new driver genes and mechanisms of activation. However, the complexity and diversity of cancer genomics is becoming evident as increasingly sophisticated sequencing technologies are developed. Consequently, others mechanisms of liver tumorigenesis remain to be explored, such as alterations impacting the 3D conformation of the chromatin.

2. Rearrangement signatures in HCC

The development of a computational tool for structural rearrangement signature analysis (Palimpsest, Article 3) allowed me to identify 6 rearrangement signatures in HCC (RS1 to RS6, Article 4). We unraveled the molecular cause of signature RS1: replication stress induced by *CCNA2/E1* activation. However, the molecular cause of the remaining signatures remains to be established. Besides, these signatures can be operative in a small proportion of the tumors (e.g. 7% for CCN-HCC displaying high RS1 contribution). Thus, it is important to continue exploring rearrangement signatures in larger HCC cohorts. Recently, I integrated SV data from 616 liver tumor samples, from my laboratory (French cohort, n=190), TCGA (North-American cohort, n=53) and ICGC (Japanese cohorts LIRI-JP, n=258; LINC-JP, n=28 and Chinese cohort LICA-CN, n=87) series. Using Palimpsest package, I was able to extract an extended set of 8 rearrangement signatures (RS) operating with different intensities in each tumor and leading to extreme SV phenotypes in subsets of HCC (Figure 25) This larger cohort revealed 2 new signatures compared to the 6 RS described in Article 4. The RS1 (short-duplications), RS2 (long-duplications), RS4 (mixed, non-clustered events) and RS6 (inversions) from Article 4 remained unchanged; however, RS3 and RS5 were subdivided into two signatures, respectively. I then examined correlations between each signature and various driver alterations.


A
RS1
RS2
RS3
RS4
RS5
RS6
RS7
RS8
Clustered events
Non clustered events
% Alterations
B
Height
% Events
# Events
Series
RS clusters
LICA-CN
LICA-FR
LIHC-US
LINC-JP
LIRI-JP
RS5
RS6
RS2
RS1
RS3/4
RS8
RS7

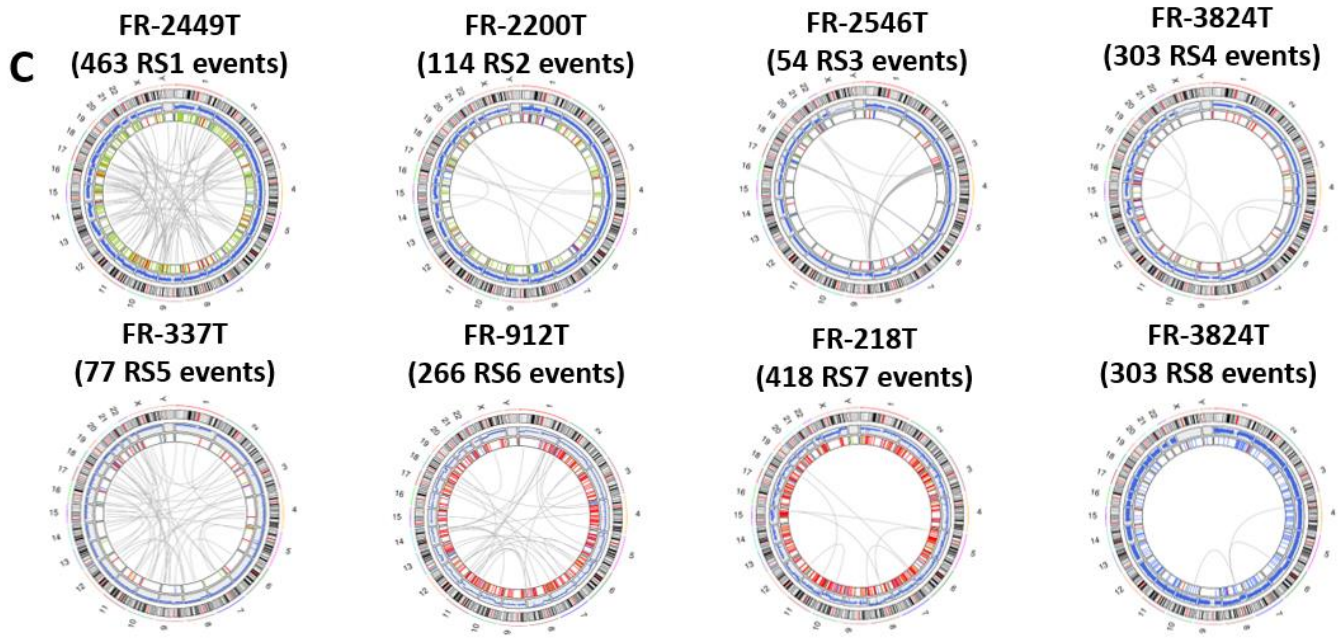

C
FR-2449T
(463 RS1 events)
FR-2200T
(114 RS2 events)
FR-2546T
(54 RS3 events)
FR-3824T
(303 RS4 events)
FR-337T
(77 RS5 events)
FR-912T
(266 RS6 events)
FR-218T
(418 RS7 events)
FR-3824T
(303 RS8 events)

Figure 25: Summary of rearrangement signature results in 616 HCC samples. A: Barplots representing the 8 rearrangement signatures extracted by NMF and characterized by specific category of rearrangements seen in subgroups of tumors. B: From top to bottom: Hierarchical clustering using cosine distance from number of SV associated to each tumor; Proportion of SV related to each RS per sample; Number of SV related to each RS per sample; Series of the sample (PCAWG project IDs are given); Manually defined cluster of signature, based on the hierarchical clustering. Signatures are operating at different intensities in each tumor with some signatures characterized by a higher number of generated rearrangements (ie: RS1). C: Circos plots of representative tumors of each RS. Circle heatmap represent duplications, deletions and inversions, respectively in green, red and blue. Grey arcs represent inter-chromosomal translocations. Blue line track represents LRR. In those examples, the corresponding signature mainly represents the global exposure of those tumors.

Signatures RS1 and RS2 are dominated by tandem duplications of respectively short (1-100Kb) and large size (100Kb-1Mb), RS1 is the signature of replication stress driven by *CCNA2/E1* activation that we described in **Article 4**. We validated the association with *CCNA2/E1* activating events in each cohort of our meta-series, including the Chinese cohort that had not been studied before. In addition to tandem duplications, RS1 displays frequent inter-chromosomal translocations corresponding to template insertion cycles, that will be described in more details in the next section. By contrast, signature RS2 is only characterized by tandem duplications. Like signature RS1, this signature is highly active in a small subset of HCC (4.4% tumors with ≥40 RS2 events). However, we have not found any significant molecular association that could indicate the molecular cause for this signature. Interestingly, phenotypes of large duplication have been seen in prostate and ovary tumors harboring a *CDK12* inactivating mutation, respectively named the tandem duplicator phenotype in prostate cancer (duplication median size = 1.3Mb) (Viswanathan *et al.*, 2018) and the tandem duplicator-plus phenotype in ovary cancer (duplication median size = 3Mb) (Popova *et al.*, 2016). However, RS2 median duplication size (around 270Kb) is not consistent with these phenotypes, additionally, *CDK12* mutation was not associated with RS2 signature in HCC.

Signatures RS3 and RS4 are characterized by clustered rearrangements. RS3 involves clustered inter-chromosomal translocations and intra-chromosomal rearrangements smaller than 1Mb. RS4 involves large (>1Mb) intra-chromosomal rearrangements without inter-chromosomal translocations (Figure 25A). Visualization of SV clusters together with CNA profiles at the chromosome level revealed that RS3 and RS4 are representative of chromoplexy and chromothripsis, respectively (Figure 26). Indeed, RS3-altered tumors displayed clusters of structural variations on various chromosomes, interconnected by inter-chromosomal translocations and associated with copy number fluctuations (Figure 26A top). RS4-exposed tumors displayed clusters of intra-chromosomal rearrangements inducing a characteristic 2 copy-number state, consistent with chromothripsis description (Figure 26A middle). Interestingly, RS3 and RS4 frequently co-occur in the same tumors (Figure 26A bottom), consistent with recent work

from the PCAWG showing that half of chromothripsis events are co-localized with other complex genomic alterations (Cortés-Ciriano *et al.*, 2020). Many evidence suggest that chromothripsis might occur before or after additional events of complex rearrangements, such as BFB cycles that have been identified as a predisposing factor for chromothripsis (Rausch *et al.*, 2012; Li *et al.*, 2014; Nones *et al.*, 2014; Maciejowski *et al.*, 2015). Both RS3 and RS4 activities were enriched in *TP53*-mutated HCC (Wilcoxon p-value: $1.0x10^{-3}$ and $1.6x10^{-12}$, respectively) (Figure 26B), suggesting that those complex alterations are the consequences of genome instability induced by *TP53* inactivation, as already described for chromothripsis in medulloblastoma (Rausch *et al.*, 2012).

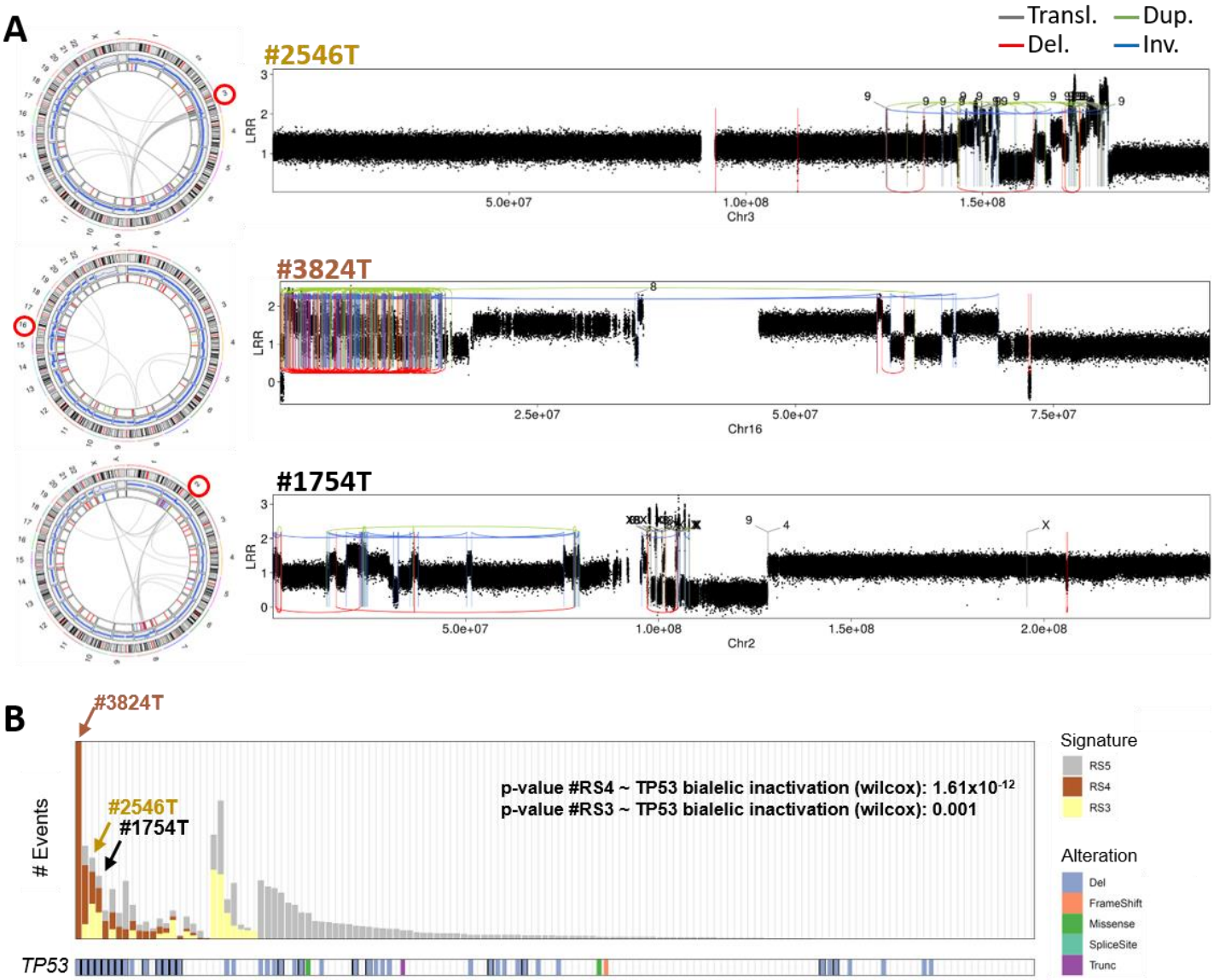


Figure 26: A: From top to bottom, circos plot and a representative chromosome (indicated with a red circle on the circos plot) are display for three tumors. The first one is characterized by lots of RS3 SV, associated to chromoplexy, which is altering chromosomes 3 and 9 of this tumor. The second tumor is highly enriched in RS4 events, and harbor an event of chromothripsis localized at the chr16p region. Finally, on the bottom is exemplified a tumor harboring, on a single chromosome, both chromothripsis and chromoplexy, inducing this presence of RS3 and RS4 in some tumors. B: Barplot representing the number of SV associated to RS3, RS4 and RS5 in each tumor of the French HCC cohort, the heatmap on the bottom indicate TP53 alterations. Black squared indicated bi-allelic inactivation of TP53. Association p-value between the number of RS3 and RS4 alterations is respectively indicated (Wilcoxon test)

In particular, tumors from African patients exposed to aflatoxin B1 displayed high contributions of RS3 and RS4. Aflatoxin B1 exposure induces a particular mutational signature (COSMIC signature 24) leading to a specific hotpot of *TP53* mutations (R249S) (Schulze, Imbeaud, Letouzé *et al.*, 2015). These tumors illustrate a complex relationship between risk factors, mutational signatures and driver genes: aflatoxin B1 exposure induces *TP53* mutations by mutational signature 24, favoring the occurrence of chromothripsis and chromoplexy events.

RS5 signature is dominated by non-clustered translocations and intra-chromosomal rearrangements of various lengths. This signature is operative at low intensity in a large proportion of the tumors and does not contribute to a particular extreme phenotype (Figure 25B, 26B). Visual inspection of characteristic rearrangements of this signature revealed either isolated structural variations or complex rearrangements not clustered enough to be associated to RS3 or RS4. Thus, RS5 may correspond to the inherent rearrangements accumulating stochastically during cell divisions.

Signatures RS6 and RS7 were associated to deletions of respectively focal (<1Kb) and large size (1Kb to 1Mb) and were highly active in a small subset of HCC (1.0% of tumors with ≥40 RS6 events, 5.8% of tumors with ≥40 RS7 events). Unfortunately, even if those signatures were extremely active in some tumors, no significant association with etiological or molecular features was identified. However, RS6 is very similar to a breast cancer signature of short deletion (Figure 18, RS5) that was found to be associated with bi-allelic *BRCA2* inactivation (Nik-Zainal *et al.*, 2016). Interestingly, the tumor FR912T harboring the most RS6 alterations (Figure 25C) has a rare *BRCA2* germline frame-shift mutation, coupled with a somatic loss of second wild-type allele. This tumor is also characterized by a high mutational burden of COSMIC SBS3 (Single Base Substitution signature 3) and ID8 (small Insertions and Deletions signature 8) signature, consistent with homologous recombination deficiency (Nik-Zainal *et al.*, 2016). Thus, RS6 may highlight rare HCC with homologous recombination deficiencies, which needs to be confirmed in larger cohorts.

Signature RS8 is new to this analysis and is characterized by a large number of focal inversions distributed all along the genome of some tumors (Figure 25C). Surprisingly, this signature was only detected in the LIRI-JP cohort from the Riken research center in Japan. Further investigations are needed to determine if it corresponds

to a true biological phenotype, e.g. related to a specific exposure to local mutagens, or to a technical artifact in this series.

Overall, driver alterations have been identified associated with signatures RS1 (*CCNA2*/*E1*), RS3 (*TP53*), RS4 (*TP53*) and possibly RS6 (*BRCA2*). The molecular cause of other signatures remains to be found.

3. In-depth characterization of the RS1 signature

3.1. RS1 as a replication stress signature in HCC

RS1 signature is characteristic of CCN-HCC, driven by *CCNA2*/*E1* activation (Figure 25B, 25C top-left), that display an accumulation of hundreds of tandem duplications and template insertion cycles. We interpreted RS1 as a signature of replication stress in HCC because i) cyclins A2 and E1 promote S phase entry and progression, and cyclin E1 activation was shown to induce replication stress, ii) CCN-HCC display a transcriptional activation of ATR response to replication stress pathway and iii) RS1 breakpoints are strongly enriched in early-replicated regions. In addition, recent works suggest that RS1 alterations may result from replication machinery defects, and more precisely break induced replication mechanism:

- First, a recent study showed that *CCNE1* activation in U2OS cells leads to a shortened G1 phase, early S phase entry and firing of normally silenced replication origins in regions containing highly expressed genes, prone to collapse and associated with DNA double-strand break (DSB) formation (Macheret and Halazonetis, 2018).
- Secondly, DSB occurring when replication fork collapses are handled and repaired by the break-induced replication (BIR) repair mechanism (Malkova, 2018).
- In a budding yeast model of replication stress induced by cyclin E overexpression, BIR was shown to be required for progression through cell cycle while inducing focal duplications (<200Kb) (Costantino *et al.*, 2014).
- Additionally, template switching of the replication machinery may occur following BIR involvement in DSB, inducing complexes chromosome rearrangements (Lee,

Carvalho and Lupski, 2007; Smith, Llorente and Symington, 2007; Hastings, Ira and Lupski, 2009).

Thus, RS1 enrichment in early-replicated, highly expressed regions is consistent with a BIR mechanism. In addition, the nature of RS1 rearrangements (tandem duplications and templated insertion cycles) is consistent with a replication-based mechanism in which the DNA polymerase at a stalled replication fork would switch template, replicate one or more other regions and move back to the original template were the replication machinery stalled in the first place, generating a duplication of the involved regions (Lee, Carvalho and Lupski, 2007; Hastings, Ira and Lupski, 2009; Carvalho *et al.*, 2011; Li *et al.*, 2020).

We also investigated the prevalence of RS1 signature in other cancers. To do so, we applied our rearrangement signature framework to the PCAWG pan-cancer cohort of WGS. This analysis revealed a signature RS1-pancan (similar to RS1), that was highly active in some liver, breast and ovarian cancers. However, the association with *CCNA2/E1* activation was specific to the liver. By contrast, RS1 was significantly associated with *BRCA1* inactivation in breast and ovarian cancers. Despite sharing this RS1-pancan signature of short tandem duplication and T.I.C., *CCNA2/E1*-altered liver tumors and *BRCA1*-altered breast and ovary tumors display slightly different features (Article 4: Fig 6b-f):

- The number of RS1-pancan rearrangements is higher in *CCNA2*-activated HCC (median = 269) than in *CCNE1*-activated HCC (137) and *BRCA1*-altered breast (132) and ovarian (159) cancers.
- Tandem duplications are larger in *CCNE1*-activated HCC (median = 39 kb) than in *CCNA2*-activated HCC (22 kb), and smaller in *BRCA1*-altered breast (9 kb) and ovarian (10 kb) cancers.
- Duplication and translocation breakpoints are strongly enriched in early-replicated regions in CCN-HCC as compared with other HCC, but not in *BRCA1*-altered as compared with other breast and ovarian cancers.

BRCA1 is known to be involved in the response to replication stress (Schlacher, Wu and Jasin, 2012; Pathania *et al.*, 2014). Its inactivation leads to tandem duplication formation at stalled forks by a replication restart-bypass mechanism, terminated by end joining or

by microhomology-mediated template switching, with extension of the stalled leading strand by a migration bubble mechanism similar to BIR (Willis *et al.*, 2017). Cyclin A2/E1 activation in HCC and *BRCA1* inactivation in breast and ovarian cancers may thus converge towards a similar rearrangement signature, with specificities reflecting the different ways by which these genetic alterations induce replication stress or modulate response to it (Article 4: Fig 6g). Indeed, rearrangements induced by cyclin activation in liver or *BRCA1* inactivation in breast and ovary cancers are initiated by replication fork collapse that will be processed by different repair mechanism, leading to similar signature of rearrangements.

Interestingly, *CCNE1* amplifications are seen in various cancer types, particularly breast and ovary cancers (Sanchez-Vega *et al.*, 2018). However, those alterations do not induce RS1 phenotype in those tumors (Article 4: Fig 6b-c). This fact could be explained by different reasons:

- First, cell types are very different, as adult hepatocytes are quiescent, rarely divide and may thus be particularly sensitive to replication stress.
- Second, *CCNE1* activation in HCC is the consequence of insertional mutagenesis or enhancer hijacking, rather than amplification as it's the case in other cancers. Those mechanism of activation could have different functional consequences.
- Finally, viral insertions and SV activating *CCNA2/E1* genes in HCC are early events triggering hepato-carcinogenesis as they occur in patients without underlying liver disease and in absence of other oncogenic early events such as *CTNNB1* or *TERTp* mutations. On the other hand, *CCNE1* amplification may occur later in breast and ovary, not leaving enough time for RS1-pancan rearrangements to accumulate.

3.2. In-depth characterization of RS1 rearrangements by long-read optical mapping

Replication-based mechanisms are more and more described as a major source of complex rearrangements in cancers (Lee, Carvalho and Lupski, 2007; Hastings, Ira and Lupski, 2009; Carvalho *et al.*, 2011; Li *et al.*, 2020). These rearrangements, like templated

insertion cycles, have been proposed based on the abnormal junctions and copy-number changes identified in WGS data, but their existence cannot be formally demonstrated with short-read sequencing. Besides, short read data do not allow to reconstruct the precise molecule resulting from complex rearrangements, as several scenarios are often equally possible [Article 4: Fig.5]. Thus, new approaches allowing the analysis of long sequences are necessary to formally demonstrate the existence of such events and their role in cancer development.

To elucidate in detail the replication stress signature RS1, we analyzed 4 tumors (2 CCN-HCC and 2 other HCC) with Bionano's optical mapping. This approach involves labeling high-molecular-weight DNA with fluorescent agents targeting specific sequences distributed on the human genome with an average inter-marker distance of 5 kb. Images of these long molecules are then captured in nanochannels by Saphyr Chip's technology, allowing simultaneous analysis of hundreds of thousands of DNA molecules. These images are processed by dedicated algorithms that recognize the "barcodes" of fluorescent markers, map each molecule to the human genome reference and identify structural variants in comparison with matched normal samples (Chan, 2018). With this approach, we obtained a genome coverage between 275X and 325X from molecules longer than 150 kb. This data allowed us to demonstrate, for the first time, the existence of T.I.C. that we had predicted from short-read data (Figure 27).

First, I developed a computational framework to extract templated insertion cycles from WGS data, by integrating abnormal junctions and copy-number gains (Figure 27A). Then, optical mapping data corresponding to T.I.C. regions was analyzed using both SV called from optical mapping technology and visual inspection of molecules of interest. Different patterns of T.I.C. were defined depending on the genomic location where the complex rearrangement is hosted (acceptor region) and the orientation of the fragments (donor regions). Intra-chromosomal T.I.C. may appear as pairs of inversions (Figure 27B) or deletion+duplication depending on fragment orientation. Inter-chromosomal T.I.C. may involve two or more (Figure 27C) regions from different chromosomes stitched together within an acceptor locus. Finally, complex events often involve a combination of

intra- and inter-chromosomal T.I.C. We are now characterizing in detail the proportion of each type of T.I.C. in our CCN-HCC series.

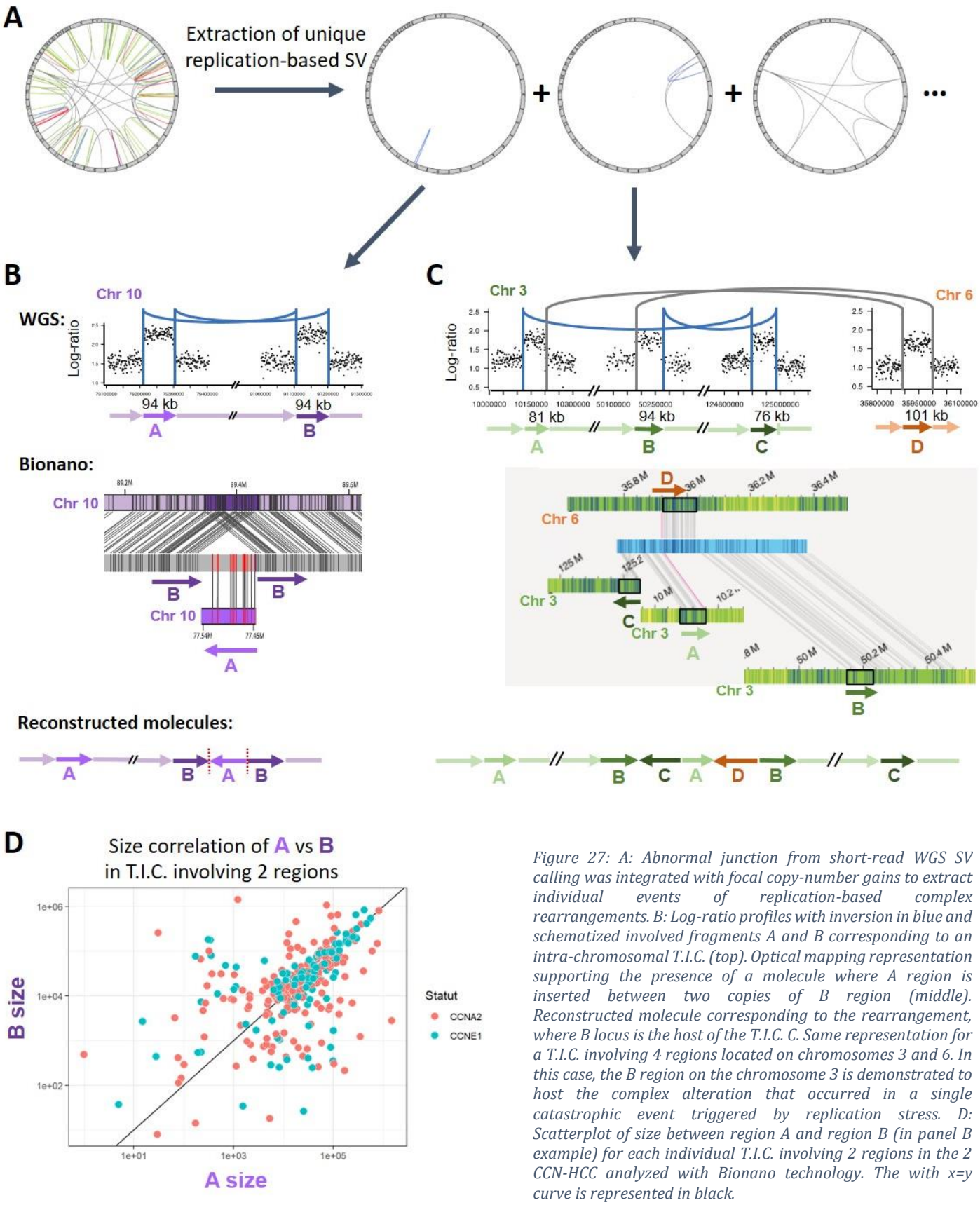


Figure 27: A: Abnormal junction from short-read WGS SV calling was integrated with focal copy-number gains to extract individual events of replication-based complex rearrangements. B: Log-ratio profiles with inversion in blue and schematized involved fragments A and B corresponding to an intra-chromosomal T.I.C. (top). Optical mapping representation supporting the presence of a molecule where A region is inserted between two copies of B region (middle). Reconstructed molecule corresponding to the rearrangement, where B locus is the host of the T.I.C. C. Same representation for a T.I.C. involving 4 regions located on chromosomes 3 and 6. In this case, the B region on the chromosome 3 is demonstrated to host the complex alteration that occurred in a single catastrophic event triggered by replication stress. D: Scatterplot of size between region A and region B (in panel B example) for each individual T.I.C. involving 2 regions in the 2 CCN-HCC analyzed with Bionano technology. The with x=y curve is represented in black.

Strikingly, the size of segments involved in a same intra- or inter-chromosomal T.I.C. is highly correlated (Figure 27D). Functional studies are required to explain this association, which may be due to the precise replication timing at which rearrangements occur, the speed of the replication machinery at the first breakpoint location, or the BIR mechanism itself.

In addition to the demonstration and characterization of T.I.C., optical mapping data, will allow me to automatically detect for each T.I.C. the host genomic region in which initial stalling occurred, and the inserted segment(s) to which template switching occurred during BIR. I will analyze these regions to better understand the properties favoring replication fork blockade as well as the properties of regions to which the replication machinery is switched in order to resolve this blockade. For example, a recent paper shows that double-strand breaks associated with replication stress tend to occur next to long A-T tracts (Tubbs *et al.*, 2018). It has also been shown, in cellular models, that activation of Cyclin E1 induces early initiation of replication at ectopic origins in the vicinity of highly expressed genes (Macheret and Halazonetis, 2018). Thus, the perspectives for this project are to analyze in detail the location of hundreds of rearrangements induced by replicative stress in CCN-HCC to identify (epi)genomic parameters (A-T sequences, ectopic initiation sites, gene expression, chromatin opening...) favoring the occurrence of RS1 breakpoints. Moreover, the analysis and comparison of sequences on each side of abnormal junctions will allow me to better understand the affinity properties (ie: micro-homologies) leading to the association of distant regions of the genome with each other following the "jumps" of the replication machinery.

4. Clinical implications of the results

4.1. Liver cancer therapeutic options

Past decade improvement of sequencing technologies made consequent contribution to the understanding of liver cancer biology landscape. However this knowledge has not been translated into clinical practices (Sia and Llovet, 2017). Indeed, the current algorithm for management of HCC only take in account the BCLC including only clinical features (Figure 23). Furthermore, only 2 drugs are used in first line and 3 in

second line, a very limited panel of therapeutic options regarding the complexity of liver cancer genomics. In order to improve HCC and HCA patient management, two axes need to be explored. Firstly, precision medicine requires predictive biomarkers to identify specific molecular subgroups of cancer. Secondly, novel targeted therapy development is required to fit each tumor subgroup biology (De Gramont *et al.*, 2015).

4.2. Identification of biomarkers of IHCA or CCN-HCC

In this project, I identified recurrent rearrangements triggering inflammatory adenomas (Articles 1 & 2) and cyclin-driven hepatocellular carcinomas (Article 4). Different biomarkers can be proposed to identify patients with IHCA or CCN-HCC. First, some alterations induce massive overexpression of driver genes. For example, *CCNE1* activation in CCN-HCC, *ROS1* and *IL6* activation in IHCA may be detected at diagnosis using RNAseq or low cost high throughput technology such as Fluidigm real-time qPCR system. RNAseq is also an interesting clinical tool in addition to WES to detect recurrent fusions in liver tumors (*FRK*, *ROS1* and *INHBE-GLI1* fusions in HCA, *CCNE1* and *CCNA2* in HCC, *PRKACA-DNAJB1* in fibrolamellar carcinoma). However, it may be difficult to obtain biological samples from advanced stage HCC. Analysis of DNA and mRNA from circulating tumor cells or cell-free plasma in liquid biopsies is thus a promising option to guide therapeutic decision in the future (Mullard, 2016; Xu *et al.*, 2017).

Finally, rearrangement signatures are emerging as useful biomarkers in cancer. For example, defective HR in *BRCA1*-mutated breast cancers induces specific signature of mutations and SVs (Nik-Zainal *et al.*, 2012; Alexandrov, Nik-Zainal, Wedge, Aparicio, *et al.*, 2013). These tumors respond well to Poly(ADP-ribose) polymerase (PARP) inhibitors (olaparib), the first clinically approved drugs designed to exploit synthetic lethality by decreasing the DNA Damage Response (DRR) in cancer cells (Crown, O'Shaughnessy and Gullo, 2012; Kanjanapan, Lheureux and Oza, 2017; Lord and Ashworth, 2017). Thus, HR deficiency mutational signatures are useful biomarkers to identify patients eligible to PARP inhibitors (Chopra *et al.*, 2020). Similarly, RS1 signature is an excellent biomarker of CCN-HCC that may be exploited for therapy.

4.3. Tailored therapy for CCN-HCC and IHCA tumors

Concerning hepatocellular carcinomas, this work allowed me to characterize a homogeneous subgroup of HCC driven by the activation of *CCNA2* or *CCNE1* genes,

representing 7% of HCCs. CCN-HCC are usually large tumors of poor prognosis. However, they share molecular characteristics, in particular high proliferation and replication stress, that could be used as therapeutic opportunities for these patients (Forment and O'Connor, 2018).

- First, chemotherapies affecting actively dividing cells may be a good option for these highly proliferative tumors with strong overexpression of E2F targets. Duplicator phenotype in breast cancer cell lines or xenografts was identified as a marker of positive chemotherapy response (Menghi *et al.*, 2018). Thus, trans-arterial chemoembolization (TACE) with doxorubicin, cisplatin or epirubisin, normally recommended with intermediate HCC nor eligible for surgery, may be an interesting therapeutic option for CCN-HCC.
- Secondly, as previously introduced, PARP inhibitors have demonstrated benefits in tumors carrying *BRCA1* mutations (Kanjanapan, Lheureux and Oza, 2017). Although CCN-HCC do nots harbor a DNA repair defect, these tumors share with *BRCA1* altered breast and ovarian tumors a similar rearrangement signature (RS1-pancan), hence PARP inhibitors may be an interesting therapeutic option for CCN-HCC.
- Finally, several compounds targeting replication stress response genes such as ATR, CHK1 and WEE1 are currently tested in phase I and II trials (Forment and O'Connor, 2018). Furthermore, *CCNE1* overexpression was shown be associated with increased sensitivity to ATR inhibition (Toledo *et al.*, 2011).

To evaluate at large scale, the clinical behavior of CCN-HCC, Barkha Gupta (anatomo-pathologist in the lab) is testing immunostaining of cyclin A2 and E1, which would allow CCN-HCC screening in the clinic from histological slides, and give us access to large retrospective series. We also started a collaboration with Amaia Lujambio (from Icahn School of Medicine at Mount Sinai, New-York) to set up a murine model of CCN-HCC. Amaia and her team have developed a mouse model of *CCNE1*-activated HCC by hydrodynamic tail injection. We have sequenced the whole genome of a few tumors and I am currently analyzing their profiles to determine if they reproduce the CCN-HCC phenotype. Indeed, accurate pre-clinical models will be crucial to test the different treatment options mentioned above and develop tailored therapy for patients with aggressive CCN-HCC.

Concerning hepatocellular adenomas, this work revealed 3 new driver alterations in the inflammatory HCA subgroup. Even if HCA is a benign tumor, it can be associated with clinical complication such as chronic inflammation or tumor bleeding. Moreover, malignant HCC can develop from a benign HCA tumor. Consequently, molecular characterization is also important for this cancer type and could be used to design drug therapy in future clinical practice. Indeed, *ROS1* fusions can be targeted by crizotinib, a small molecule currently approved by the Food and Drug Administration (FDA) for the treatment of lung cancers (Shaw *et al.*, 2013). Additionally, previous work from my laboratory showed *in cellulo* that dasatinib, a src inhibitor, inhibits JAK/STAT pathway activation due to *FRK* mutations (Pilati *et al.*, 2014). We can then predict that this compound will also be effective in IHCA harboring *FRK* gene fusion. Further analyses must be done to test the proposed targeted therapies associated to driver alteration (Article 2: Fig.5).

V REFERENCES

Abou-Alfa, G. K. *et al.* (2018) 'Cabozantinib in patients with advanced and progressing hepatocellular carcinoma', *New England Journal of Medicine*. doi: 10.1056/NEJMoa1717002.

Agrawal, N. *et al.* (2014) 'Integrated Genomic Characterization of Papillary Thyroid Carcinoma', *Cell*. doi: 10.1016/j.cell.2014.09.050.

Ahn, S. M. *et al.* (2014) 'Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification', *Hepatology*. doi: 10.1002/hep.27198.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., *et al.* (2013) 'Deciphering Signatures of Mutational Processes Operative in Human Cancer', *Cell Reports*. doi: 10.1016/j.celrep.2012.12.008.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., *et al.* (2013) 'Signatures of mutational processes in human cancer', *Nature*. doi: 10.1038/nature12477.

Alexandrov, L. B. *et al.* (2015) 'Clock-like mutational processes in human somatic cells', *Nature Genetics*. doi: 10.1038/ng.3441.

Alexandrov, L. B. *et al.* (2020) 'The repertoire of mutational signatures in human cancer', *Nature*. doi: 10.1038/s41586-020-1943-3.

Alioto, T. S. *et al.* (2015) 'A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing', *Nature Communications*. doi: 10.1038/ncomms10001.

Anstee, Q. M. *et al.* (2019) 'From NASH to HCC: current concepts and future challenges', *Nature Reviews Gastroenterology and Hepatology*. doi: 10.1038/s41575-019-0145-7.

Armitage, P. and Doll, R. (1954) 'The age distribution of cancer and a multi-stage theory of carcinogenesis', *British Journal of Cancer*. doi: 10.1038/bjc.1954.1.

Ascha, M. S. *et al.* (2010) 'The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis', *Hepatology*. doi: 10.1002/hep.23527.

Aubrey, B. J. *et al.* (2018) 'How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression?', *Cell Death and Differentiation*. doi: 10.1038/cdd.2017.169.

Audano, P. A. *et al.* (2019) 'Characterizing the Major Structural Variant Alleles of the Human Genome', *Cell*. doi: 10.1016/j.cell.2018.12.019.

Babina, I. S. and Turner, N. C. (2017) 'Advances and challenges in targeting FGFR signalling in cancer', *Nature Reviews Cancer*. doi: 10.1038/nrc.2017.8.

Baca, S. C. *et al.* (2013) 'Punctuated evolution of prostate cancer genomes', *Cell*. Elsevier Inc., 153(3), pp. 666–677. doi: 10.1016/j.cell.2013.03.021.

Bacq, Y. *et al.* (2003) 'Familial Liver Adenomatosis Associated with Hepatocyte Nuclear Factor 1α Inactivation', *Gastroenterology*. doi: 10.1016/j.gastro.2003.07.012.

Baron, S. (1996) *Medical Microbiology. 4th edition*, *University of Texas Medical Branch at Galveston*. doi: NBK8035 [bookaccession].

Basham, K. J. *et al.* (2019) 'A ZNRF3-dependent Wnt/β-catenin signaling gradient is required for adrenal homeostasis', *Genes and Development*. doi: 10.1101/gad.317412.118.

Bedossa, P. *et al.* (2007) 'Evidence for a role of nonalcoholic steatohepatitis in hepatitis C: A prospective study', *Hepatology*. doi: 10.1002/hep.21711.

La Bella, T. *et al.* (2020) 'Adeno-associated virus in the liver: Natural history and consequences in tumour development', *Gut*. doi: 10.1136/gutjnl-2019-318281.

Bender, S. *et al.* (2013) 'Reduced H3K27me3 and DNA Hypomethylation Are Major Drivers of Gene Expression in K27M Mutant Pediatric High-Grade Gliomas', *Cancer Cell*. doi: 10.1016/j.ccr.2013.10.006.

Benjamin, D. *et al.* (2019) 'Calling Somatic SNVs and Indels with Mutect2', *bioRxiv*. doi: 10.1101/861054.

Bickmore, W. A. (2013) 'The Spatial Organization of the Human Genome', *Annual Review of Genomics and Human Genetics*. doi: 10.1146/annurev-genom-091212-153515.

Bishop, J. M. (1983) 'Cellular Oncogenes and Retroviruses', *Annual Review of Biochemistry*. doi: 10.1146/annurev.bi.52.070183.001505.

Bishop, J. M. (1987) 'The molecular genetics of cancer', *Science*. doi: 10.1126/science.3541204.

Bluteau, O. *et al.* (2002) 'Bi-allelic inactivation of TCF1 in hepatic adenomas', *Nature Genetics*. doi: 10.1038/ng1001.

Boveri, T. (1914) 'Zur Frage der Entstehung Maligner Tumoren [Origin of malignant tumors]', *Gustav Fisher, Jena*. doi: 10.1016/j.yexcr.2012.02.003.

Boyault, S. *et al.* (2007) 'Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets', *Hepatology*. doi: 10.1002/hep.21467.

Bray, F. *et al.* (2018) 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: A Cancer Journal for Clinicians*. doi: 10.3322/caac.21492.

Bressac, B. *et al.* (1991) 'Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa', *Nature*. doi: 10.1038/350429a0.

BROWN, J. R. and THORNTON, J. L. (1957) 'Percivall Pott (1714-1788) and chimney sweepers' cancer of the scrotum', *British journal of industrial medicine*. doi: 10.1136/oem.14.1.68.

Bruix, J. *et al.* (2017) 'Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial', *The Lancet*. doi: 10.1016/S0140-6736(16)32453-9.

Bruix, J. and Llovet, J. M. (2002) 'Prognostic prediction and treatment strategy in hepatocellular carcinoma', *Hepatology*. doi: 10.1053/jhep.2002.32089.

Brunet, J. P. *et al.* (2004) 'Metagenes and molecular pattern discovery using matrix factorization', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0308531101.

Brunner, T. B. *et al.* (2019) 'The Role of Stereotactic Body Radiotherapy (SBRT) in Locally Advanced Hepatocellular Carcinoma Versus Transarterial Chemoembolization (TACE): A Prospective, Non-Randomized Observational Trial (HERACLES)', *International Journal of Radiation Oncology*Biology*Physics*. doi: 10.1016/j.ijrobp.2019.06.2058.

Bryan, H. K. *et al.* (2013) 'The Nrf2 cell defence pathway: Keap1-dependent and -independent mechanisms of regulation', *Biochemical Pharmacology*. doi: 10.1016/j.bcp.2012.11.016.

Calderaro, J. *et al.* (2013) 'Molecular characterization of hepatocellular adenomas developed in patients with glycogen storage disease type i', *Journal of Hepatology*. doi: 10.1016/j.jhep.2012.09.030.

Campbell, P. J. *et al.* (2020) 'Pan-cancer analysis of whole genomes', *Nature*. doi: 10.1038/s41586-020-1969-6.

Carvalho, C. M. B. *et al.* (2011) 'Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome', *Nature Genetics*. doi: 10.1038/ng.944.

Chan, S. (2018) *Copy Number Variants Methods and Protocols Methods in Molecular Biology 1833*. Available at: http://www.springer.com/series/7651.

Chedid, M. F. (2017) 'Nonalcoholic Steatohepatitis: The Second Leading Indication for Liver Transplantation in the USA', *Digestive Diseases and Sciences*. doi: 10.1007/s10620-017-4724-6.

Chen, C. L. *et al.* (2010) 'Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes', *Genome Research*. doi: 10.1101/gr.098947.109.

Chen, X. *et al.* (2016) 'Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications', *Bioinformatics*, 32(8), pp. 1220–1222. doi: 10.1093/bioinformatics/btv710.

Chopra, N. *et al.* (2020) 'Homologous recombination DNA repair deficiency and PARP inhibition activity in primary triple negative breast cancer.', *Nature communications*. Springer US, 11(1), p. 2662. doi: 10.1038/s41467-020-16142-7.

Cilloni, D. and Saglio, G. (2012) 'Molecular pathways: BCR-ABL', *Clinical Cancer Research*. doi: 10.1158/1078-0432.CCR-10-1613.

Cleary, S. P. *et al.* (2013) 'Identification of driver genes in hepatocellular carcinoma by exome sequencing', *Hepatology*. doi: 10.1002/hep.26540.

Collisson, E. A. *et al.* (2014) 'Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network', *Nature*. doi: 10.1038/nature13385.

Cortes-Ciriano, I. *et al.* (2014) 'Proteochemometric modeling in a Bayesian framework', *Journal of Cheminformatics*, 6(1), pp. 1–16. doi: 10.1186/1758-2946-6-35.

Cortés-Ciriano, I. *et al.* (2020) 'Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing', *Nature Genetics*. doi: 10.1038/s41588-019-0576-7.

Costantino, L. *et al.* (2014) 'Break-induced replication repair of damaged forks induces genomic duplications in human cells', *Science*. doi: 10.1126/science.1243211.

Craig Venter, J. *et al.* (2001) 'The sequence of the human genome', *Science*. doi: 10.1126/science.1058040.

Cremer, T. and Cremer, C. (2001) 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nature Reviews Genetics*. doi: 10.1038/35066075.

Cremer, T. and Cremer, M. (2010) 'Chromosome territories.', *Cold Spring Harbor perspectives in biology*. doi: 10.1101/cshperspect.a003889.

Crown, J., O'Shaughnessy, J. and Gullo, G. (2012) 'Emerging targeted therapies in triple-negative breast cancer', *Annals of Oncology*. doi: 10.1093/annonc/mds196.

Cunningham, F. *et al.* (2019) 'Ensembl 2019', *Nucleic Acids Research*. doi: 10.1093/nar/gky1113.

Das, K. *et al.* (2018) 'Protease-activated receptor 2 promotes actomyosin dependent transforming microvesicles generation from human breast cancer', *Molecular Carcinogenesis*. doi: 10.1002/mc.22891.

De, S. and Michor, F. (2011) 'DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes', *Nature Biotechnology*. doi: 10.1038/nbt.2030.

Degasperi, A. *et al.* (2020) 'A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies', *Nature Cancer*. doi: 10.1038/s43018-020-0027-5.

Delattre, O. *et al.* (1992) 'Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours', *Nature*. doi: 10.1038/359162a0.

Delattre, O. *et al.* (1994) 'The ewing family of tumors — a subgroup of small-round-cell tumors defined by specific chimeric transcripts', *New England Journal of Medicine*. doi: 10.1056/NEJM199408043310503.

Denicola, G. M. *et al.* (2011) 'Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis', *Nature*. doi: 10.1038/nature10189.

Dinh, T. A. *et al.* (2017) 'Comprehensive analysis of the Cancer Genome Atlas reveals a unique gene and non-coding RNA signature of fibrolamellar carcinoma', *Scientific Reports*. doi: 10.1038/srep44653.

Dixon, J. R. *et al.* (2012) 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*. doi: 10.1038/nature11082.

El-Khoueiry, A. B. *et al.* (2017) 'Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial', *The Lancet*. doi: 10.1016/S0140-6736(17)31046-2.

Erstad, D. J. and Cusack, J. C. (2014) 'Mutational analysis of merkel cell carcinoma', *Cancers*. doi: 10.3390/cancers6042116.

Farazi, P. A. *et al.* (2003) 'Differential impact of telomere dysfunction on initiation and progression of hepatocellular carcinoma', *Cancer Research*.

Ferlay, J. *et al.* (2019) 'Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods', *International Journal of Cancer*. doi: 10.1002/ijc.31937.

Forment, J. V. and O'Connor, M. J. (2018) 'Targeting the replication stress response in cancer', *Pharmacology and Therapeutics*. doi: 10.1016/j.pharmthera.2018.03.005.

Fredriksson, N. J. *et al.* (2014) 'Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types', *Nature Genetics*. doi: 10.1038/ng.3141.

Friemel, J. *et al.* (2019) 'Characterization of HCC mouse models: Towards an etiology-oriented subtyping approach', *Molecular Cancer Research*. doi: 10.1158/1541-7786.MCR-18-1045.

Froissart, R. *et al.* (2011) 'Glucose-6-phosphatase deficiency', *Orphanet Journal of Rare Diseases*. doi: 10.1186/1750-1172-6-27.

Fujii, M. *et al.* (2013) 'A murine model for non-alcoholic steatohepatitis showing evidence of association between diabetes and hepatocellular carcinoma', *Medical Molecular Morphology*. doi: 10.1007/s00795-013-0016-1.

Fujimoto, A. *et al.* (2016) 'Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer', *Nature Genetics*. Nature Publishing Group, 48(5), pp. 500–509. doi: 10.1038/ng.3547.

Galuppo, R. *et al.* (2014) 'Synergistic inhibition of HCC and liver cancer stem cell proliferation by targeting RAS/RAF/MAPK and WNT/β-Catenin pathways', *Anticancer Research*.

Gao, Q. *et al.* (2018) 'Driver Fusions and Their Implications in the Development and Treatment of Human Cancers', *Cell Reports*. doi: 10.1016/j.celrep.2018.03.050.

Garsed, D. W. *et al.* (2014) 'The Architecture and Evolution of Cancer Neochromosomes', *Cancer Cell*, 26(5), pp. 653–667. doi: 10.1016/j.ccell.2014.09.010.

Gibcus, J. H. and Dekker, J. (2013) 'The Hierarchy of the 3D Genome', *Molecular Cell*. doi: 10.1016/j.molcel.2013.02.011.

Ginès, P. *et al.* (2016) 'Screening for liver fibrosis in the general population: a call for action', *The Lancet Gastroenterology and Hepatology*. doi: 10.1016/S2468-1253(16)30081-4.

Glodzik, D. *et al.* (2017) 'A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers', *Nature Genetics*. Nature Publishing Group, 49(3), pp. 341–348. doi: 10.1038/ng.3771.

Glover, T. W., Wilson, T. E. and Arlt, M. F. (2017) 'Fragile sites in cancer: More than meets the eye', *Nature Reviews Cancer*. Nature Publishing Group, 17(8), pp. 489–501. doi: 10.1038/nrc.2017.52.

Gómez-González, B. and Aguilera, A. (2019) 'Transcription-mediated replication hindrance: A major driver of genome instability', *Genes and Development*. doi: 10.1101/gad.324517.119.

De Gramont, Armand *et al.* (2015) 'Pragmatic issues in biomarker evaluation for targeted therapies in cancer', *Nature Reviews Clinical Oncology*. doi: 10.1038/nrclinonc.2014.202.

Gudlevičiene, Ž. *et al.* (2015) 'Distribution of human papillomavirus type 16 variants in Lithuanian women with cervical cancer', *Medicina (Lithuania)*. doi: 10.1016/j.medici.2015.11.005.

Guichard, C. *et al.* (2012) 'Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma', *Nature Genetics*. doi: 10.1038/ng.2256.

Guo, L., Zhang, H. and Chen, B. (2017) 'Nivolumab as Programmed Death-1 (PD-1) Inhibitor for Targeted Immunotherapy in Tumor', *Journal of Cancer*. doi: 10.7150/jca.17144.

Hanahan, D. and Weinberg, R. A. (2000) 'The hallmarks of cancer', *Cell*, 100, pp. 57–70. doi: 10.1007/s00262-010-0968-0.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of cancer: The next generation', *Cell*. doi: 10.1016/j.cell.2011.02.013.

Hansen, R. S. *et al.* (2010) 'Sequencing newly replicated DNA reveals widespread plasticity in human replication timing', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0912402107.

Hantschel, O. (2012) 'Structure, Regulation, Signaling, and Targeting of Abl Kinases in Cancer', *Genes and Cancer*. doi: 10.1177/1947601912458584.

Haradhvala, N. J. *et al.* (2016) 'Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair', *Cell*. doi: 10.1016/j.cell.2015.12.050.

Hastings, P. J., Ira, G. and Lupski, J. R. (2009) 'A microhomology-mediated break-induced replication model for the origin of human copy number variation', *PLoS Genetics*. doi: 10.1371/journal.pgen.1000327.

Hirsch, T. Z. *et al.* (2020) 'BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA', *Journal of Hepatology*. doi: 10.1016/j.jhep.2019.12.006.

Ho, J. *et al.* (2020) 'Detection of CSF1 rearrangements deleting the 3′ UTR in tenosynovial giant cell tumors', *Genes Chromosomes and Cancer*. doi: 10.1002/gcc.22807.

Honeyman, J. N. *et al.* (2014) 'Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma.', *Science (New York, N.Y.)*, 343(6174), pp. 1010–4. doi: 10.1126/science.1249484.

Howard, B. D. and Tessman, I. (1964) 'Identification of the altered bases in mutated single-stranded DNA', *Journal of Molecular Biology*. doi: 10.1016/s0022-2836(64)80214-x.

Hsu, I. C. *et al.* (1991) 'Mutational hot spot in the p53 gene in human hepatocellular carcinomas', *Nature*. doi: 10.1038/350427a0.

Hu, Z. *et al.* (2015) 'Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism', *Nature Genetics*. doi: 10.1038/ng.3178.

Huddleston, J. and Eichler, E. E. (2016) 'An incomplete understanding of human genetic variation', *Genetics*. doi: 10.1534/genetics.115.180539.

Hutchinson, K. E. *et al.* (2013) 'BRAF fusions define a distinct molecular subset of melanomas with potential sensitivity to MEK inhibition', *Clinical Cancer Research*. doi: 10.1158/1078-0432.CCR-13-1746.

Jang, Y. E. *et al.* (2020) 'ChimerDB 4.0: an updated and expanded database of fusion genes', *Nucleic acids research*. doi: 10.1093/nar/gkz1013.

Johnson, R. T. and Rao, P. N. (1970) 'Mammalian cell fusion: Induction of premature chromosome condensation in interphase nuclei', *Nature*. doi: 10.1038/226717a0.

Kan, Z. *et al.* (2013) 'Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma', *Genome Research*. doi: 10.1101/gr.154492.113.

Kanjanapan, Y., Lheureux, S. and Oza, A. M. (2017) 'Niraparib for the treatment of ovarian cancer', *Expert Opinion on Pharmacotherapy*. doi: 10.1080/14656566.2017.1297423.

Kato, H. and Sandberg, A. A. (1968) 'Chromosome pulverization in human cells with micronuclei', *Journal of the National Cancer Institute*. doi: 10.1093/jnci/40.1.165.

Kern, J. S. *et al.* (2006) 'Expanding the COL7A1 mutation database: Novel and recurrent mutations and unusual genotype - Phenotype constellations in 41 patients with dystrophic epidermolysis bullosa', *Journal of Investigative Dermatology*. doi: 10.1038/sj.jid.5700219.

Kew, M. C. (2014) 'Hepatic iron overload and hepatocellular carcinoma', *Liver Cancer*. doi: 10.1159/000343856.

Kohno, T. *et al.* (2015) 'Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer', *Translational Lung Cancer Research*. doi: 10.3978/j.issn.2218-6751.2014.11.11.

Labrune, P. *et al.* (1997) 'Hepatocellular adenomas in glycogen storage disease type I and III: A series of 43 patients and review of the literature', *Journal of Pediatric Gastroenterology and Nutrition*. doi: 10.1097/00005176-199703000-00008.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*. doi: 10.1038/35057062.

Laplante, M. and Sabatini, D. M. (2012) 'MTOR signaling in growth control and disease', *Cell*. doi: 10.1016/j.cell.2012.03.017.

Lasorella, A., Sanson, M. and Iavarone, A. (2017) 'FGFR-TACC gene fusions in human glioma', *Neuro-Oncology*. doi: 10.1093/neuonc/now240.

Lawrence, M. S. *et al.* (2013) 'Mutational heterogeneity in cancer and the search for new cancer-associated genes', *Nature*. doi: 10.1038/nature12213.

Lee, D. D. and Seung, H. S. (1999) 'Learning the parts of objects by non-negative matrix factorization', *Nature*. doi: 10.1038/44565.

Lee, J. A., Carvalho, C. M. B. and Lupski, J. R. (2007) 'A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders', *Cell*. doi: 10.1016/j.cell.2007.11.037.

Lee, M. *et al.* (2017) 'ChimerDB 3.0: An enhanced database for fusion genes from cancer transcriptome and literature data mining', *Nucleic Acids Research*, 45(D1), pp. D784–D789. doi: 10.1093/nar/gkw1083.

Letouzé, E. *et al.* (2017) 'Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis', *Nature Communications*. doi: 10.1038/s41467-017-01358-x.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, Y. *et al.* (2014) 'Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia', *Nature*. doi: 10.1038/nature13115.

Li, Y. *et al.* (2020) 'Patterns of somatic structural variation in human cancer genomes', *Nature*. doi: 10.1038/s41586-019-1913-9.

Lieberman-Aiden, E. *et al.* (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*. doi: 10.1126/science.1181369.

Lin, J. I. (1983) 'Rudolf Virchow: Creator of Cellular Pathology', *Laboratory Medicine*. doi: 10.1093/labmed/14.12.791.

Liu, Y. and Wu, F. (2010) 'Global burden of Aflatoxin-induced hepatocellular carcinoma: A risk assessment', *Environmental Health Perspectives*. doi: 10.1289/ehp.0901388.

Llovet, J. M. *et al.* (2008) 'Sorafenib in advanced hepatocellular carcinoma.', *The New England journal of medicine*, 359(4), pp. 378–90. doi: 10.1056/NEJMoa0708857.

Llovet, J. M. *et al.* (2016) 'Hepatocellular carcinoma', *Nature Reviews Disease Primers*. doi: 10.1038/nrdp.2016.18.

Lok, A. S. *et al.* (2009) 'Incidence of Hepatocellular Carcinoma and Associated Risk Factors in Hepatitis C-Related Advanced Liver Disease', *Gastroenterology*. doi: 10.1053/j.gastro.2008.09.014.

Lord, C. J. and Ashworth, A. (2017) 'PARP inhibitors: Synthetic lethality in the clinic', *Science*. doi: 10.1126/science.aam7344.

Luke, J. J. *et al.* (2019) 'WNT/b-catenin pathway activation correlates with immune exclusion across human cancers', *Clinical Cancer Research*. doi: 10.1158/1078-0432.CCR-18-1942.

Ly, P. and Cleveland, D. W. (2017) 'Rebuilding Chromosomes After Catastrophe: Emerging Mechanisms of Chromothripsis', *Trends in Cell Biology*. Elsevier Ltd, 27(12), pp. 917–930. doi: 10.1016/j.tcb.2017.08.005.

Macheret, M. and Halazonetis, T. D. (2018) 'Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress', *Nature*. Nature Publishing Group. doi: 10.1038/nature25507.

Maciejowski, J. *et al.* (2015) 'Chromothripsis and Kataegis Induced by Telomere Crisis', *Cell*. doi: 10.1016/j.cell.2015.11.054.

Malkova, A. (2018) 'Break-Induced Replication: The Where, The Why, and The How', *Trends in Genetics*. doi: 10.1016/j.tig.2018.04.002.

Marnef, A., Cohen, S. and Legube, G. (2017) 'Transcription-Coupled DNA Double-Strand Break Repair: Active Genes Need Special Care', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2017.03.024.

Martincorena, I. *et al.* (2017) 'Universal Patterns of Selection in Cancer and Somatic Tissues', *Cell*. doi: 10.1016/j.cell.2017.09.042.

Matsumoto, Y. *et al.* (2011) 'Perception of oral contraceptives among women of reproductive age in Japan: A comparison with the USA and France', *Journal of Obstetrics and Gynaecology Research*. doi: 10.1111/j.1447-0756.2010.01461.x.

Matsushita, K. *et al.* (2009) 'Zc3h12a is an RNase essential for controlling immune responses by regulating

mRNA decay', *Nature*. doi: 10.1038/nature07924.

Matthews, B. J. and Waxman, D. J. (2018) 'Computational prediction of CTCF/ cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver', *eLife*. doi: 10.7554/eLife.34077.001.

McClintock, B. (1939) 'The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.25.8.405.

McGlynn, K. A. *et al.* (2001) 'International trends and patterns of primary liver cancer', *International Journal of Cancer*. doi: 10.1002/ijc.1456.

McGregor, D. B. *et al.* (1998) 'An IARC evaluation of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans as risk factors in human carcinogenesis', in *Environmental Health Perspectives*. doi: 10.2307/3433830.

Menghi, F. *et al.* (2018) 'The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations', *Cancer Cell*. doi: 10.1016/j.ccell.2018.06.008.

Merle, P. *et al.* (2019) 'Doxorubicin-loaded nanoparticles for patients with advanced hepatocellular carcinoma after sorafenib treatment failure (RELIVE): a phase 3 randomised controlled trial', *The Lancet Gastroenterology and Hepatology*. doi: 10.1016/S2468-1253(19)30040-8.

Mermel, C. H. *et al.* (2011) 'GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers', *Genome Biology*. doi: 10.1186/gb-2011-12-4-r41.

Möller, E. *et al.* (2008) 'Molecular identification of COL6A3-CSFI fusion transcripts in tenosynovial giant cell tumors', *Genes Chromosomes and Cancer*. doi: 10.1002/gcc.20501.

Morganella, S. *et al.* (2016) 'The topography of mutational processes in breast cancer genomes', *Nature Communications*. doi: 10.1038/ncomms11383.

Mudduluru, G. *et al.* (2011) 'Curcumin regulates miR-21 expression and inhibits invasion and metastasis in colorectal cancer', *Bioscience Reports*. doi: 10.1042/BSR20100065.

Mukherjee, S. (2010) *The emperor of all maladies : a biography of cancer*. Scribner.

Mullard, A. (2016) 'Reining in the supersized Phase I cancer trial', *Nature Reviews Drug Discovery*. doi: 10.1038/nrd.2016.110.

Murnane, J. P. (2012) 'Telomere dysfunction and chromosome instability', *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. doi: 10.1016/j.mrfmmm.2011.04.008.

Nault, J.-C. *et al.* (2017) 'Molecular Classification of Hepatocellular Adenoma Associates With Risk Factors, Bleeding, and Malignant Transformation', *Gastroenterology*, 152(4), pp. 880-894.e6. doi: 10.1053/j.gastro.2016.11.042.

Nault, J. C. *et al.* (2013) 'High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions', *Nature Communications*. doi: 10.1038/ncomms3218.

Nault, J. C. *et al.* (2015) 'Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas', *Nature Genetics*. doi: 10.1038/ng.3389.

Nault, J. C. and Letouzé, E. (2019) 'Mutational Processes in Hepatocellular Carcinoma: The Story of Aristolochic Acid', *Seminars in Liver Disease*. doi: 10.1055/s-0039-1685516.

Nault, J. C. and Zucman-Rossi, J. (2016) 'TERT promoter mutations in primary liver tumors', *Clinics and Research in Hepatology and Gastroenterology*. doi: 10.1016/j.clinre.2015.07.006.

Nik-Zainal, S. *et al.* (2012) 'Mutational processes molding the genomes of 21 breast cancers', *Cell*. doi:

10.1016/j.cell.2012.04.024.

Nik-Zainal, S. *et al.* (2016) 'Landscape of somatic mutations in 560 breast cancer whole-genome sequences', *Nature*, 534(7605), pp. 47–54. doi: 10.1038/nature17676.

Nones, K. *et al.* (2014) 'Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis', *Nature Communications*. doi: 10.1038/ncomms6224.

Nora, E. P. *et al.* (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*. doi: 10.1038/nature11049.

Northcott, P. A. *et al.* (2014) 'Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma', *Nature*. doi: 10.1038/nature13379.

O'Keefe, C., McDevitt, M. A. and Maciejewski, J. P. (2010) 'Copy neutral loss of heterozygosity: A novel chromosomal lesion in myeloid malignancies', *Blood*. doi: 10.1182/blood-2009-10-201848.

Papaemmanuil, E. *et al.* (2014) 'RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia', *Nature Genetics*. doi: 10.1038/ng.2874.

Paschoud, S. *et al.* (2006) 'Destabilization of Interleukin-6 mRNA Requires a Putative RNA Stem-Loop Structure, an AU-Rich Element, and the RNA-Binding Protein AUF1', *Molecular and Cellular Biology*. doi: 10.1128/mcb.01155-06.

Pathania, S. *et al.* (2014) 'BRCA1 haploinsufficiency for replication stress suppression in primary cells', *Nature Communications*. doi: 10.1038/ncomms6496.

Pilati, C. *et al.* (2014) 'Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation', *Cancer Cell*. doi: 10.1016/j.ccr.2014.03.005.

Plentz, R. R. *et al.* (2007) 'Telomere shortening and inactivation of cell cycle checkpoints characterize human hepatocarcinogenesis', *Hepatology*. doi: 10.1002/hep.21552.

Polak, P. *et al.* (2014) 'Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair', *Nature Biotechnology*. doi: 10.1038/nbt.2778.

Polak, P. *et al.* (2015) 'Cell-of-origin chromatin organization shapes the mutational landscape of cancer', *Nature*. doi: 10.1038/nature14221.

Popova, T. *et al.* (2009) 'Genome Alteration Print (GAP): A tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays', *Genome Biology*. doi: 10.1186/gb-2009-10-11-r128.

Popova, T. *et al.* (2016) 'Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications', *Cancer Research*. doi: 10.1158/0008-5472.CAN-15-2128.

Raha, D., Hong, M. and Snyder, M. (2010) 'ChIP-Seq: A method for global identification of regulatory elements in the genome', *Current Protocols in Molecular Biology*. doi: 10.1002/0471142727.mb2119s91.

Rausch, T. *et al.* (2012) 'Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations', *Cell*. doi: 10.1016/j.cell.2011.12.013.

Rebouissou, S. *et al.* (2016) 'Genotype-phenotype correlation of CTNNB1 mutations reveals different ß-catenin activity associated with liver tumor progression', *Hepatology*. doi: 10.1002/hep.28638.

Reddy, E. P. *et al.* (1982) 'A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene', *Nature*. doi: 10.1038/300149a0.

Reznik, Y. *et al.* (2004) 'Hepatocyte nuclear factor-1α gene inactivation: Cosegregation between liver

adenomatosis and diabetes phenotypes in two maturity-onset diabetes of the young (MODY)3 families', *Journal of Clinical Endocrinology and Metabolism*. doi: 10.1210/jc.2003-031552.

Rheinbay, E. *et al.* (2020) 'Analyses of non-coding somatic drivers in 2,658 cancer whole genomes', *Nature*. doi: 10.1038/s41586-020-1965-x.

Roadmap Epigenomics Consortium *et al.* (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature*. doi: 10.1038/nature14248.

Rooks, J. B. *et al.* (1977) 'The association between oral contraception and hepatocellular adenoma - a preliminary report', *International Journal of Gynecology and Obstetrics*. doi: 10.1002/j.1879-3479.1977.tb00664.x.

Rosa, I. *et al.* (2010) '585 A FRENCH MULTICENTRIC LONGITUDINAL DESCRIPTIVE STUDY OF HEPATOCELLULAR CARCINOMA MANAGEMENT (THE CHANGH COHORT): PRELIMINARY RESULTS', *Journal of Hepatology*. doi: 10.1016/s0168-8278(10)60587-9.

Rowley, J. D. (1973) 'A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining', *Nature*. doi: 10.1038/243290a0.

Sabarinathan, R. *et al.* (2016) 'Nucleotide excision repair is impaired by binding of transcription factors to DNA', *Nature*. doi: 10.1038/nature17661.

Sánchez-Osorio, M. *et al.* (2008) 'Anabolic-androgenic steroids and liver injury', *Liver International*. doi: 10.1111/j.1478-3231.2007.01579.x.

Sanchez-Vega, F. *et al.* (2018) 'Oncogenic Signaling Pathways in The Cancer Genome Atlas', *Cell*. doi: 10.1016/j.cell.2018.03.035.

Sawey, E. T. *et al.* (2011) 'Identification of a Therapeutic Strategy Targeting Amplified FGF19 in Liver Cancer by Oncogenomic Screening', *Cancer Cell*. doi: 10.1016/j.ccr.2011.01.040.

Schlacher, K., Wu, H. and Jasin, M. (2012) 'A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2', *Cancer Cell*. doi: 10.1016/j.ccr.2012.05.015.

Schubbert, S., Shannon, K. and Bollag, G. (2007) 'Hyperactive Ras in developmental disorders and cancer', *Nature Reviews Cancer*. doi: 10.1038/nrc2109.

Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, Ludmil B, *et al.* (2015) 'Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets', *Nature Genetics*, 47(5), pp. 505–511. doi: 10.1038/ng.3252.

Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, Ludmil B., *et al.* (2015) 'Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets', *Nature Genetics*. doi: 10.1038/ng.3252.

Schuster-Böckler, B. and Lehner, B. (2012) 'Chromatin organization is a major influence on regional mutation rates in human cancer cells', *Nature*. doi: 10.1038/nature11273.

Shaw, A. T. *et al.* (2013) 'Crizotinib versus chemotherapy in advanced ALK-positive lung cancer', *New England Journal of Medicine*. doi: 10.1056/NEJMoa1214886.

Sia, D. and Llovet, J. M. (2017) 'Liver cancer: Translating "-omics" results into precision medicine for hepatocellular carcinoma', *Nature Reviews Gastroenterology and Hepatology*. doi: 10.1038/nrgastro.2017.103.

Sikov, W. M. *et al.* (2015) 'Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (A', *Journal of Clinical Oncology*.

doi: 10.1200/JCO.2014.57.0572.

Sima, J. and Gilbert, D. M. (2014) 'Complex correlations: Replication timing and mutational landscapes during cancer and genome evolution', *Current Opinion in Genetics and Development*. doi: 10.1016/j.gde.2013.11.022.

Sinclair, A., Latif, A. L. and Holyoake, T. L. (2013) 'Targeting survival pathways in chronic myeloid leukaemia stem cells', *British Journal of Pharmacology*. doi: 10.1111/bph.12183.

Singh, D. *et al.* (2012) 'Transforming fusions of FGFR and TACC genes in human glioblastoma', *Science*. doi: 10.1126/science.1220834.

Smith, C. E., Llorente, B. and Symington, L. S. (2007) 'Template switching during break-induced replication', *Nature*. doi: 10.1038/nature05723.

Smyth, M. J. *et al.* (2016) 'Combination cancer immunotherapies tailored to the tumour microenvironment', *Nature Reviews Clinical Oncology*. doi: 10.1038/nrclinonc.2015.209.

Sondka, Z. *et al.* (2018) 'The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers', *Nature Reviews Cancer*. doi: 10.1038/s41568-018-0060-1.

Sporn, M. B. and Liby, K. T. (2012) 'NRF2 and cancer: The Good, the bad and the importance of context', *Nature Reviews Cancer*. doi: 10.1038/nrc3278.

Stephens, P. J. *et al.* (2011) 'Massive genomic rearrangement acquired in a single catastrophic event during cancer development', *Cell*. Elsevier Inc., 144(1), pp. 27–40. doi: 10.1016/j.cell.2010.11.055.

Stratton, M. R., Campbell, P. J. and Futreal, P. A. (2009) 'The cancer genome', *Nature*. doi: 10.1038/nature07943.

Sung, W. K. *et al.* (2012) 'Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma', *Nature Genetics*. doi: 10.1038/ng.2295.

Supek, F. and Lehner, B. (2015) 'Differential DNA mismatch repair underlies mutation rate variation across the human genome', *Nature*. doi: 10.1038/nature14173.

Takeuchi, K. *et al.* (2012) 'RET, ROS1 and ALK fusions in lung cancer', *Nature Medicine*. doi: 10.1038/nm.2658.

Toledo, L. I. *et al.* (2011) 'A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2076.

Totoki, Y. *et al.* (2014) 'Trans-ancestry mutational landscape of hepatocellular carcinoma genomes', *Nature Genetics*. doi: 10.1038/ng.3126.

Touboul, T. *et al.* (2016) 'Stage-specific regulation of the WNT/β-catenin pathway enhances differentiation of hESCs into hepatocytes', *Journal of Hepatology*. doi: 10.1016/j.jhep.2016.02.028.

Touraine, R. L. *et al.* (1993) 'Hepatic tumours during androgen therapy in Fanconi anaemia', *European Journal of Pediatrics*. doi: 10.1007/BF01955250.

Tubbs, A. *et al.* (2018) 'Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse', *Cell*, 174(5), pp. 1127-1142.e19. doi: 10.1016/j.cell.2018.07.011.

Viswanathan, S. R. *et al.* (2018) 'Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing', *Cell*. doi: 10.1016/j.cell.2018.05.036.

Wansbrough-Jones, M. H. *et al.* (1998) 'Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. doi:

10.1016/S0035-9203(98)90887-2.

Watson, J. D. and Crick, F. H. C. (1953) 'Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid', *Nature*. doi: 10.1038/171737a0.

Weber, J. S. *et al.* (2015) 'Toxicities of immunotherapy for the practitioner', *Journal of Clinical Oncology*. doi: 10.1200/JCO.2014.60.0379.

Weinhold, N. *et al.* (2014) 'Genome-wide analysis of noncoding regulatory mutations in cancer', *Nature Genetics*. doi: 10.1038/ng.3101.

Weinstein, L. S. and Shenker, A. (1993) 'G protein mutations in human disease', *Clinical Biochemistry*. doi: 10.1016/0009-9120(93)90109-J.

Weischenfeldt, J. *et al.* (2017) 'Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking', *Nature Genetics*, 49(1), pp. 65–74. doi: 10.1038/ng.3722.

Willis, N. A. *et al.* (2017) 'Mechanism of tandem duplication formation in BRCA1-mutant cells', *Nature*. Nature Publishing Group, 551(7682), pp. 590–595. doi: 10.1038/nature24477.

Winter, J. M., Brody, J. R. and Kern, S. E. (2006) 'Multiple-criterion evaluation of reported mutations: A proposed scoring system for the intragenic somatic mutation literature', *Cancer Biology and Therapy*. doi: 10.4161/cbt.5.4.2552.

Witkin, E. M. (1969) 'Ultraviolet-Induced Mutation and DNA Repair', *Annual Review of Microbiology*. doi: 10.1146/annurev.mi.23.100169.002415.

Woodcock, C. L. and Ghosh, R. P. (2010) 'Chromatin higher-order structure and dynamics.', *Cold Spring Harbor perspectives in biology*. doi: 10.1101/cshperspect.a000596.

Wynn, T. A. (2004) 'Fibrotic disease and the TH1/TH2 paradigm', *Nature Reviews Immunology*. doi: 10.1038/nri1412.

Xu, R. H. *et al.* (2017) 'Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma', *Nature Materials*. doi: 10.1038/NMAT4997.

Yanai, H. *et al.* (2000) 'The colorectal tumour suppressor APC is present in the NMDA-receptor-PSD-95 complex in the brain', *Genes to Cells*. doi: 10.1046/j.1365-2443.2000.00368.x.

Yang, J. D. and Roberts, L. R. (2010) 'Hepatocellular carcinoma: a global view', *Nature Reviews Gastroenterology & Hepatology*, 7(8), pp. 448–458. doi: 10.1038/nrgastro.2010.100.

Zhang, Y. *et al.* (2018) 'A Pan-Cancer Compendium of Genes Deregulated by Somatic Genomic Rearrangement across More Than 1,400 Cases', *Cell Reports*. ElsevierCompany., 24(2), pp. 515–527. doi: 10.1016/j.celrep.2018.06.025.

Zhu, A. X. *et al.* (2018) 'Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial', *The Lancet Oncology*. doi: 10.1016/S1470-2045(18)30351-6.

Zhu, A. X. *et al.* (2019) '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', *The Lancet Oncology*. doi: 10.1016/S1470-2045(18)30937-9.

Zucman-Rossi, J. *et al.* (2006) 'Genotype-phenotype correlation in hepatocellular adenoma: New classification and relationship with HCC', *Hepatology*. doi: 10.1002/hep.21068.

Zucman-Rossi, J. *et al.* (2015) 'Genetic Landscape and Biomarkers of Hepatocellular Carcinoma', *Gastroenterology*, 149(5), pp. 1226–1239. doi: 10.1053/j.gastro.2015.05.061.

VI ANNEXES

1. Annex 1: Supplementary data Article 2

Supplementary Table 1: List of taqman probes

Gene	Unabbreviated gene name	Taqman probe
RNA18S5	RNA, 18S Ribosomal 5	Hs03928990_g1
FABP1	Fatty Acid Binding Protein 1	Hs00155026_m1
UGT2B7	UDP Glucuronosyltransferase Family 2 Member B7	Hs00426592_m1
CRP	C-Reactive Protein	Hs04183452_g1
SAA2	Serum Amyloid A2	Hs00754237_s1
GLUL	Glutamate-Ammonia Ligase	Hs00374213_m1
LGR5	Leucine Rich Repeat Containing G Protein-Coupled Receptor 5	Hs00173664_m1
PTCH1	Patched 1	Hs00181117_m1
GLI1	Glioma-Associated Oncogene Homolog 1	Hs00171790_m1
HHIP	Hedgehog Interacting Protein	Hs01011015_m1
TNNC1	Troponin C1, Slow Skeletal And Cardiac Type	Hs00896999_g1
FCRLA	Fc Receptor Like A	Hs00262071_m1
PTGDS	Prostaglandin D2 Synthase	Hs00168748_m1
ADGRG3	Adhesion G Protein-Coupled Receptor G3	Hs00416887_m1
ROS1	ROS Proto-Oncogene 1	Hs01090625_m1
FRK	Fyn Related Src Family Tyrosine Kinase	Hs00176619_m1
IL6	Interleukin 6	Hs00985639_m1

Supplementary Table 2: Description of the 296 IHCA samples in the study

Sample	hological_diagn	Gender	Age	IHCA_molecular_classification	Mutation_CTNNB1_ex	Mutation_CTNNB1_ex2	Mutation_CTNNB1_ex	type_mutation_CTNNB1
#2363T	HCA	NA	NA	biHCA	M	NM	NM	missense
#477T	HCA	F	23	biHCA	M	NM	NM	indel
#1129T	HCA	F	41	biHCA	M	NM	NM	missense
#1131T	HCA	F	31	biHCA	M	NM	NM	missense
#1275T	HCA	F	40	biHCA	M	NM	NM	missense
#1287T	HCA	NA	NA	biHCA	M	NM	NM	indel
#2196T	HCA	F	40	biHCA	M	NM	NM	missense
#2809T	HCA	F	39	biHCA	M	NM	NM	missense
#2823T	HCA	F	30	biHCA	M	NM	NM	missense
#2849T	HCA	F	NA	biHCA	M	NM	NM	missense
#4058T	HCA	F	NA	biHCA	**M**	NM	NM	missense
#4098T	HCA	NA	NA	biHCA	**M**	NM	NM	indel
#4117T	HCA	F	NA	biHCA	**M**	NM	NM	missense
#536T	HCA	F	46	biHCA	M	NM	NM	missense
#1343T	HCA	F	42	IHCA	NM	NM	NM	
#819T	HCA	F	45	biHCA	M	NM	NM	missense
#823T	HCA	F	45	biHCA	M	NM	NM	missense
#1250T	HCA	F	31	biHCA	M	NM	NM	indel
#2827T	HCA	F	40	biHCA	M	NM	NM	missense
#372T	HCA	F	48	biHCA	M	NM	NM	missense
#974T	HCA	F	42	biHCA	M	NM	NM	missense
#979T	HCA	M	58	biHCA	M	NM	NM	missense
#1896T	HCA	F	45	IHCA	NM	NM	NM	
#2729T	HCA	F	25	IHCA	NM	NM	NM	
#375T	HCA	F	54	IHCA	NM	NM	NM	
#4110T	HCA	NA	NA	IHCA	NM	NM	NM	
#533T	HCA	M	32	IHCA	NM	NM	NM	
#755T	HCA	F	39	biHCA	M	NM	NM	missense
#1851T	HCA	F	40	biHCA	M	NM	NM	indel
#711T	HCA	F	26	biHCA	M	NM	NM	missense
#1940T	HCA	F	40	biHCA	M			indel
#757T	HCA	F	48	IHCA	NM	NM	NM	
#222T	HCA	M	29	biHCA	NM	M	NM	missense
#1234T	HCA	F	21	biHCA	M	NM	NM	missense
#2798T	HCA	F	37	biHCA	M	NM	NM	missense
#3074T	HCA	F	44	biHCA	M	NM	NM	missense
#3723T	HCA	F	28	biHCA	**M**	NM	NM	missense
#4116T	HCA	NA	NA	biHCA	**M**	NM	NM	indel
#1355T	HCA	F	40	IHHCA	NM	NM	NM	
#1439T	HCA	F	49	IHCA	NM	NM	NM	
#760T	HCA	F	30	biHCA	NM	NM	NM	
#471T	HCA	F	30	IHCA	NM	NM	NM	
#2561T	HCA	M	41	biHCA	NM	NM	NM	
#1231T	HCA	F	36	biHCA	NM	NM	NM	
#1496T	HCA	F	52	biHCA	NM	M	NM	missense
#835T	HCA	F	37	IHCA	NM	NM	NM	
#554T	HCA	F	40	biHCA	NM	M	NM	missense
#574T	HCA	M	54	biHCA	NM	M	NM	missense
#1342T	HCA	F	41	biHCA	NM	M	NM	missense
#1354T	HCA	F	56	biHCA	NM	M	NM	missense
#364T	HCA	F	35	biHCA	NM	M	NM	missense
#377T	HCA	F	50	biHCA	NM	M	NM	missense
#390T	HCA	M	50	biHCA	NM	M	NM	missense
#742T	HCA	M	18	biHCA	NM	M	NM	missense
#1891T	HCA	F	21	biHCA	NM	M	NM	missense
#378T	HCA	F	56	biHCA	NM	M	NM	missense
#458T	HCA	F	25	biHCA	NM	M	NM	missense
#506T	HCA	F	56	biHCA	NM	M	NM	missense
#582T	HCA	F	35	biHCA	NM	M	NM	missense
#631T	HCA	F	36	biHCA	NM	NM	M	missense
#680T	HCA	F	42	biHCA	NM	NM	M	missense
#684T	HCA	F	29	biHCA	NM	NM	M	missense
#691T	HCA	F	35	biHCA	NM	NM	M	missense
#955T	HCA	F	39	biHCA	NM	NM	M	missense
#341T	HCA	M	47	IHCA	NM	NM	NM	
#2834T	HCA	F	48	IHCA	NM	NM	NM	
#1130T	HCA	F	31	IHCA	NM	NM	NM	
#1239T	HCA	F	41	IHCA	NM	NM	NM	
#1240T	HCA	F	41	IHCA	NM	NM	NM	
#1263T	HCA	F	37	IHCA	NM	NM	NM	
#1325T	HCA	F	49	IHCA	NM	NM	NM	
#873T	HCA	M	33	IHCA	NM	NM	NM	
#2615T	HCA	F	50	IHCA	NM	NM	NM	
#1508T	HCA	F	26	IHCA	NM	NM	NM	
#2193T	HCA	F	43	IHCA	NM	NM	NM	
#2197T	HCA	F	35	IHCA	NM	NM	NM	
#2330T	HCA	F	36	IHCA	NM	NM	NM	
#2331T	HCA	F	36	IHCA	NM	NM	NM	
#2804T	HCA	F	27	IHCA	NM	NM	NM	
#2828T	HCA	F	22	IHCA	NM	NM	NM	
#2836T	HCA	F	33	IHCA	NM	NM	NM	
#3049T	HCA	F	22	IHCA	NM	NM	NM	
#4106T	HCA	NA	NA	IHCA	NM	NM	NM	
#544T	HCA	F	41	IHCA	NM	NM	NM	
#782T	HCA	F	38	IHCA	NM	NM	NM	
#970T	HCA	F	49	IHCA	NM	NM	NM	
#627T	HCA	F	30	IHCA	NM	NM	NM	
#1215T	HCA	F	33	IHCA	NM	NM	NM	
#1490T	HCA	F	34	IHCA	NM	NM	NM	
#756T	HCA	F	48	biHCA	M	NM	NM	missense
#1424T	HCA	F	29	biHCA	M	NM	NM	missense
#2334T	HCA	F	50	IHCA	NM	NM	NM	
#480T	HCA	F	34	IHCA	NM	NM	NM	
#1021T	HCA	F	39	IHCA	NM	NM	NM	
#1351T	HCA	F	39	IHCA	NM	NM	NM	
#2369T	HCA	M	28	IHCA	NM	NM	NM	
#786T	HCA/HCC	M	60	biHCA	M	NM	NM	indel
#2773T	HCA	F	28	IHCA	NM	NM	NM	
#2877T	HCA	M	NA	IHCA	NM	NM	NM	
#2879T	HCA	F	NA	IHCA	NM	NM	NM	
#379T	HCA	F	45	IHCA	NM	NM	NM	
#708T	HCA	M	54	IHCA	NM	NM	NM	
#966T	HCA	F	40	IHCA	NM	NM	NM	
#1001T	HCA	M	37	IHCA	NM	NM	NM	
#1286T	HCA	NA	NA	IHCA	NM	NM	NM	
#1519T	HCA	F	24	IHCA	NM	NM	NM	
#2810T	HCA	F	36	IHCA	NM	NM	NM	
#2811T	HCA	F	36	IHCA	NM	NM	NM	
#478T	HCA	F	33	IHCA	NM	NM	NM	
#854T	HCA	F	46	IHCA	NM	NM	NM	
#958T	HCA	F	51	IHCA	NM	NM	NM	
#1002T	HCA	F	43	IHCA	NM	NM	NM	
#1004T	HCA	F	42	IHCA	NM	NM	NM	
#1020T	HCA	F	39	IHCA	NM	NM	NM	
#343T	HCA	F	47	biHCA	M	NM	NM	missense
#1121T	HCA	F	25	IHCA	NM	NM	NM	
#1216T	HCA	F	43	IHCA	NM	NM	NM	
#1232T	HCA	F	36	IHCA	NM	NM	NM	
#1235T	HCA	F	41	IHCA	NM	NM	NM	
#1238T	HCA	F	41	IHCA	NM	NM	NM	
#1243T	HCA	F	26	IHCA	NM	NM	NM	
#1251T	HCA	F	26	IHCA	NM	NM	NM	
#1252T	HCA	F	26	IHCA	NM	NM	NM	
#1285T	HCA	M	34	IHCA	NM	NM	NM	
#1289T	HCA	F	54	IHCA	NM	NM	NM	
#1290T	HCA	F	47	IHCA	NM	NM	NM	
#1291T	HCA	F	47	IHCA	NM	NM	NM	
#1296T	HCA	F	51	IHCA	NM	NM	NM	
#1297T	HCA	F	51	IHCA	NM	NM	NM	
#1301T	HCA	F	43	IHCA	NM	NM	NM	
#1302T	HCA	F	43	IHCA	NM	NM	NM	
#1304T	HCA	F	43	IHCA	NM	NM	NM	
#1305T	HCA	F	43	IHCA	NM	NM	NM	
#1306T	HCA	F	43	IHCA	NM	NM	NM	
#1313T	HCA	M	40	IHCA	NM	NM	NM	
#1319T	HCA	F	42	IHCA	NM	NM	NM	
#2874T	HCA	F	NA	biHCA	NM	NM	NM	
#3759T	HCA	F	47	biHCA	NM	NM	NM	
#1330T	HCA	F	40	IHCA	NM	NM	NM	
#1331T	HCA	F	40	IHCA	NM	NM	NM	
#1335T	HCA	F	43	IHCA	NM	NM	NM	
#1336T	HCA	F	43	IHCA	NM	NM	NM	
#2329T	HCA/HCC	F	39	biHCA	NM	NM	NM	
#1350T	HCA	F	39	IHCA	NM	NM	NM	
#1386T	HCA	F	44	IHCA	NM	NM	NM	
#1388T	HCA	F	42	IHCA	NM	NM	NM	
#1399T	HCA	F	37	IHCA	NM	NM	NM	
#1407T	HCA	M	NA	IHCA	NM	NM	NM	
#1421T	HCA	F	37	IHCA	NM	NM	NM	
#1433T	HCA	F	38	IHCA	NM	NM	NM	
#485T	HCA/HCC	F	33	biHCA	NM	NM	NM	
#1435T	HCA	F	38	IHCA	NM	NM	NM	
#1440T	HCA	F	41	IHCA	NM	NM	NM	

#1518T	HCA	F	24	IHCA	NM	NM	NM	
#1523T	HCA	M	20	IHCA	NM	NM	NM	
#1664T	HCA	F	37	IHCA	NM	NM	NM	
#669T	HCA	F	38	IHCA	NM	NM	NM	
#1676T	HCA	F	45	IHCA	NM	NM	NM	
#1677T	HCA	F	45	IHCA	NM	NM	NM	
#2191T	HCA	F	37	IHCA	NM	NM	NM	
#2189T	HCA	M	18	IHCA	NM	NM	NM	
#2370T	HCA	M	28	IHCA	NM	NM	NM	
#754T	HCA/HCC	M	29	bIHCA	M	NM	NM	indel
#2371T	HCA	M	28	IHCA	NM	NM	NM	
#2372T	HCA	M	28	IHCA	NM	NM	NM	
#2373T	HCA	M	28	IHCA	NM	NM	NM	
#2375T	HCA	M	28	IHCA	NM	NM	NM	
#2376T	HCA	M	28	IHCA	NM	NM	NM	
#2377T	HCA	M	28	IHCA	NM	NM	NM	
#2568T	HCA	F	51	IHCA	NM	NM	NM	
#2571T	HCA	F	36	IHCA	NM	NM	NM	
#2585T	HCA	F	47	IHCA	NM	NM	NM	
#2590T	HCA	F	51	IHCA	NM	NM	NM	
#2769T	HCA	F	48	IHCA	NM	NM	NM	
#2793T	HCA	M	31	IHCA	NM	NM	NM	
#2795T	HCA	F	44	IHCA	NM	NM	NM	
#2802T	HCA	F	53	IHCA	NM	NM	NM	
#2803T	HCA	F	50	IHCA	NM	NM	NM	
#2826T	HCA	NA	NA	IHCA	NM	NM	NM	
#2829T	HCA	F	44	IHCA	NM	NM	NM	
#2830T	HCA	F	44	IHCA	NM	NM	NM	
#2835T	HCA	F	48	IHCA	NM	NM	NM	
#2837T	HCA	F	33	IHCA	NM	NM	NM	
#2856T	HCA	F	NA	IHCA	NM	NM	NM	
#750T	HCA	F	24	IHCA	NM	NM	NM	
#2900T	HCA	F	NA	IHCA	NM	NM	NM	
#2901T	HCA	F	NA	IHCA	NM	NM	NM	
#3104T	HCA	F	NA	IHCA	NM	NM	NM	
#3107T	HCA	F	NA	IHCA	NM	NM	NM	
#344T	HCA	F	40	IHCA	NM	NM	NM	
#3781T	HCA	F	28	IHCA	NM	NM	NM	
#382T	HCA	F	45	IHCA	NM	NM	NM	
#4108T	HCA	NA	NA	IHCA	NM	NM	NM	
#4109T	HCA	NA	NA	IHCA	NM	NM	NM	
#2217T	HCA/HCC	F	13	bIHCA	M	NM	NM	missense
#455T	HCA	F	27	IHCA	NM	NM	NM	
#456T	HCA	F	27	IHCA	NM	NM	NM	
#457T	HCA	F	27	IHCA	NM	NM	NM	
#466T	HCA	F	50	IHCA	NM	NM	NM	
#473T	HCA	F	38	IHCA	NM	NM	NM	
#481T	HCA	F	31	bIHCA	NM	NM	NM	
#490T	HCA	F	50	IHCA	NM	NM	NM	
#503T	HCA	F	48	IHCA	NM	NM	NM	
#1023T	HCA	M	34	bIHCA	M	NM	NM	missense
#1124T	HCA	M	50	bIHCA	M	NM	NM	missense
#552T	HCA	F	55	IHCA	NM	NM	NM	
#555T	HCA	F	40	IHCA	NM	NM	NM	
#560T	HCA	F	49	IHCA	NM	NM	NM	
#581T	HCA	F	51	IHCA	NM	NM	NM	
#594T	HCA	F	39	IHCA	NM	NM	NM	
#670T	HCA	F	38	IHCA	NM	NM	NM	
#698T	HCA	F	31	IHCA	NM	NM	NM	
#745T	HCA	F	43	IHCA	NM	NM	NM	
#751T	HCA	F	39	IHCA	NM	NM	NM	
#769T	HCA	F	34	IHCA	NM	NM	NM	
#771T	HCA	F	34	IHCA	NM	NM	NM	
#772T	HCA	F	34	IHCA	NM	NM	NM	
#821T	HCA	F	45	IHCA	NM	NM	NM	
#844T	HCA	F	40	IHCA	NM	NM	NM	
#858T	HCA	F	51	IHCA	NM	NM	NM	
#861T	HCA	F	33	IHCA	NM	NM	NM	
#864T	HCA	F	31	IHCA	NM	NM	NM	
#867T	HCA	F	38	IHCA	NM	NM	NM	
#973T	HCA	F	52	IHCA	NM	NM	NM	
#975T	HCA	F	50	IHCA	NM	NM	NM	
#1666T	HCA	F	36	bIHCA	M	NM	NM	missense
#2825T	HCA	F	23	IHCA	NM	NM	NM	
#2857T	HCA	F	NA	IHCA	NM	NM	NM	
#4094T	HCA	NA	46	IHCA	NM	NM	NM	
#468T	HCA	F	50	IHCA	NM	NM	NM	
#839T	HCA	F	46	IHCA	NM	NM	NM	
#470T	HCA	F	32	bIHCA	M	NM	NM	missense
#1916T	HCA	M	43	bIHCA	M	NM	NM	indel
#1328T	HCA	F	40	IHCA	NM	NM	NM	
#1329T	HCA	F	40	IHCA	NM	NM	NM	
#1337T	HCA	F	46	IHCA	NM	NM	NM	
#469T	HCConHCA	F	32	bIHCA	M	NM	NM	missense
#1434T	HCA	F	38	IHCA	NM	NM	NM	
#1915T	HCConHCA	M	43	bIHCA	M	NM	NM	indel
#1488T	HCA	F	35	IHCA	NM	NM	NM	
#1489T	HCA	F	35	IHCA	NM	NM	NM	
#624T	HCA	F	30	IHCA	NM	NM	NM	
#1431T	HCA	F	38	IHCA	NM	NM	NM	
#1678T	HCA	F	45	IHCA	NM	NM	NM	
#389T	HCA	F	30	IHCA	NM	NM	NM	
#4089T	HCA	F	29	IHCA	NM	NM	NM	
#744T	HCA	F	42	IHCA	NM	NM	NM	
#623T	HCA	F	30	IHCA	NM	NM	NM	
#1241T	HCA	F	26	IHCA	NM	NM	NM	
#1269T	HCA	F	30	IHCA	NM	NM	NM	
#1430T	HCA	F	49	IHCA	NM	NM	NM	
#2192T	HCA	F	50	IHCA	NM	NM	NM	
#2219T	HCA	F	14	IHCA	NM	NM	NM	
#2378T	HCA	F	48	IHCA	NM	NM	NM	
#2589T	HCA	F	32	IHCA	NM	NM	NM	
#2328T	HCConHCA	M	20	bIHCA	M	NM	NM	indel
#2591T	HCA	F	51	IHCA	NM	NM	NM	
#2796T	HCA	M	29	IHCA	NM	NM	NM	
#2902T	HCA	F	NA	IHCA	NM	NM	NM	
#3073T	HCA	F	44	IHCA	NM	NM	NM	
#331T	HCA	F	42	IHCA	NM	NM	NM	
#3410T	HCA	F	28	IHCA	NM	NM	NM	
#3766T	HCA	F	69	IHCA	NM	NM	NM	
#3780T	HCA	F	58	IHCA	NM	NM	NM	
#467T	HCA	F	50	IHCA	NM	NM	NM	
#504T	HCA	F	35	IHCA	NM	NM	NM	
#672T	HCA	F	38	IHCA	NM	NM	NM	
#748T	HCA	F	33	IHCA	NM	NM	NM	
#969T	HCA	F	19	IHCA	NM	NM	NM	
#1412T	HCA	M	37	IHCA	NM	NM	NM	
#1768T	HCA	F	41	IHCA	NM	NM	NM	
#1521T	HCA	F	26	IHCA	NM	NM	NM	
#1413T	HCA	F	58	IHCA	NM	NM	NM	
#857T	HCA	F	51	bIHCA	NM	NM	NM	
#1410T	HCA	F	41	IHCA	NM	NM	NM	
#1321T	HCA	F	41	IHCA	NM	NM		
#1356T	HCA	F	34	IHCA	NM	NM		
#1339T	HCA	F	46	IHCA	NM	NM		
#2620T	HCA	M	32	IHCA	NM			
#2573T	HCA	F	32	IHCA	NM			
#1880T	HCA	F	39	IHCA	NM			
#2574T	HCA	F	41	IHCA	NM			
#2575T	HCA	F	41	IHCA	NM			
#2576T	HCA	F	41	IHCA	NM			
#1918T	HCConHCA	M	62	bIHCA	NM	NM	NM	
#1461T	HCA	F	39	IHCA		NM	NM	
#1469T	HCA	F	34	IHCA				
#976T	HCA/HCC	M	64	bIHCA	NM	M	NM	missense
#2187T	HCA/HCC	M	43	IHCA	NM	NM	NM	
#362T	HCA/HCC	F	53	IHCA	NM	NM	NM	
#528T	HCA/HCC	M	40	IHCA	NM	NM	NM	
#2805T	HCA/HCC	M	25	IHCA	NM	NM	NM	
#545T	HCA/HCC	F	40	IHCA	NM	NM	NM	
#706T	HCA/HCC	M	53	IHCA	NM	NM	NM	
#2582T	HCConHCA	F	47	IHCA	NM	NM	NM	
#1314T	HCConHCA	M	42	IHCA	NM	NM	NM	

Sample	Mutation_IL6ST	mutation_G	type_mutation_GNAS1	Mutation_STAT3	type_mutation_STAT3	JAK1_Mutation	type_mutation_JAK1	Mutation_FRK	e_mutation_FRK
#2363T	NM	NM		M	missense	NM		NM	
#477T	NM	NM		NM		NM		NM	NM
#1129T	M	indel	NM	NM		NM		NM	
#1131T	M	indel	NM	NM		NM		NM	
#1275T	M	indel	NM	NM		NM		NM	
#1287T	M	indel	NM	NM		NM		NM	
#2136T	M	indel	NM	NM		NM		NM	
#2809T	M	indel	NM	NM		NM		NM	
#2823T	M	indel	NM	NM		NM		NM	
#2849T	M	indel	NM	NM		NM		NM	

#4058T	**M**	indel	NM		NM		NM		NM	
#4098T	M	indel	NM		NM		NM		NM	
#4117T	M	indel	NM		NM		NM		NM	
#536T	M	indel	NM		NM		NM		NM	
#1343T	NM		NM		NM		NM		NM	NM
#819T	M	indel	NM		NM		NM		NM	
#823T	M	indel	NM		NM		NM		NM	
#1250T	M	missense	NM		NM		NM		NM	
#2827T	M	missense	NM		NM		NM		NM	
#372T	M	missense	NM		NM		NM		NM	
#974T	M	missense	NM		NM		NM		NM	
#979T	M	missense	NM		NM		NM		NM	
#1896T	NM		NM		NM		NM		NM	NM
#2729T	NM		NM		NM		NM		NM	NM
#375T	NM		NM		NM		NM		NM	NM
#4110T	NM		NM		NM		NM		NM	NM
#533T	NM		NM		NM		NM		NM	NM
#755T	NM		NM		NM		NM		NM	
#1851T	M	indel	NM				NM		NM	
#711T	M	indel	NM				NM		NM	
#1940T	M	indel								
#757T	NM		NM		NM		NM		NM	NM
#222T	NM		NM		NM		NM		NM	NM
#1234T	NM		NM		NM		NM		NM	
#2798T	NM		NM		NM		NM		NM	
#3074T	NM		NM		NM		NM		NM	
#3723T	NM		NM		NM		NM		NM	
#4116T	NM		NM		NM		NM		NM	
#1355T	NM		NM		M	missense	NM		NM	
#1439T	NM		NM		NM		NM		NM	NM
#760T	M	missense	NM		NM		NM		NM	
#471T	NM		NM		NM		NM		NM	NM
#2561T	M	missense								
#1231T	NM		NM		NM		NM		NM	
#1496T	NM		NM		NM		NM		M	missense
#835T	NM		NM		NM		NM		NM	NM
#554T	NM		NM		NM		M	missense	NM	
#574T	NM		NM		M	missense	NM		NM	
#1342T	M	indel	NM		NM		NM		NM	
#1354T	M	indel	NM		NM		NM		NM	
#364T	M	indel	NM		NM		NM		NM	
#377T	M	indel	NM		NM		NM		NM	
#390T	M	indel	NM		NM		NM		NM	
#742T	M	indel	NM		NM		NM		NM	
#1891T	NM		NM		NM		NM		NM	
#378T	NM		NM		NM		NM		NM	
#458T	NM		NM		NM		NM		NM	
#506T	NM		NM		NM		NM		NM	
#582T	NM		NM		NM		NM		NM	
#631T	M	indel	NM		NM		NM		NM	
#680T	M	indel	NM		NM		NM		NM	
#684T	M	indel	NM		NM		NM		NM	
#691T	M	indel	NM		NM		NM		NM	
#955T	M	indel	NM		NM		NM		NM	
#341T	NM		NM		M	missense	NM		M	missense
#2834T	M	indel	NM		NM		NM		M	indel
#1130T	NM		NM		NM		NM		M	missense
#1239T	NM		NM		NM		NM		M	missense
#1240T	NM		NM		NM		NM		M	indel
#1263T	NM		NM		NM		NM		M	indel
#1325T	NM		NM		NM		NM		M	missense
#873T	NM		NM		NM		NM		NM	NM
#2615T	NM		NM		NM		NM		NM	NM
#1508T	NM		NM		NM		NM		M	indel
#2193T	NM		NM		NM		NM		M	indel
#2197T	NM		NM		NM		NM		M	indel
#2330T	NM		NM		NM		NM		M	missense
#2331T	NM		NM		NM		NM		M	indel
#2804T	NM		NM		NM		NM		M	indel
#2828T	NM		NM		NM		NM		M	indel
#2836T	NM		NM		NM		NM		M	indel
#3049T	NM		NM		NM		NM		**M**	indel
#4106T	NM		NM		NM		NM		**M**	missense
#544T	NM		NM		NM		NM		M	indel
#782T	NM		NM		NM		NM		M	indel
#970T	NM		NM		NM		NM		M	indel
#627T	NM		NM				NM		M	indel
#1215T	NM		NM		NM		NM		NM	NM
#1490T	NM		NM		NM		M	missense	NM	
#756T	M	indel	NM		NM		NM		NM	NM
#1424T	NM		NM		NM		NM		NM	NM
#2334T	NM		NM		NM		M	missense	NM	
#480T	NM		NM		NM		M	missense	NM	
#1021T	NM		NM		M	indel	NM		NM	
#1351T	NM		NM		M	indel	NM		NM	
#2369T	NM		NM		M	missense	NM		NM	
#786T	M	indel	NM		NM		NM		NM	
#2773T	NM		NM		M	indel	NM		NM	
#2877T	NM		NM		M	missense	NM		NM	
#2879T	NM		NM		M	indel	NM		NM	
#379T	NM		NM		M	indel	NM		NM	
#708T	NM		NM		M	missense	NM		NM	
#966T	NM		NM		M	missense	NM		NM	
#1001T	NM		M	missense	NM		NM		NM	
#1286T	NM		M	missense	NM		NM		NM	
#1519T	NM		M	missense	NM		NM		NM	
#2810T	NM		M	missense	NM		NM		NM	
#2811T	NM		M	missense	NM		NM		NM	
#478T	NM		M	missense	NM		NM		NM	
#854T	NM		M	missense	NM		NM		NM	
#958T	NM		M	missense	NM		NM		NM	
#1002T	M	indel	NM		NM		NM		NM	
#1004T	M	indel	NM		NM		NM		NM	
#1020T	M	indel	NM		NM		NM		NM	
#343T	NM		NM		NM		NM		NM	NM
#1121T	M	indel	NM		NM		NM		NM	
#1216T	M	indel	NM		NM		NM		NM	
#1232T	M	indel	NM		NM		NM		NM	
#1235T	M	indel	NM		NM		NM		NM	
#1238T	M	indel	NM		NM		NM		NM	
#1243T	M	indel	NM		NM		NM		NM	
#1251T	M	indel	NM		NM		NM		NM	
#1252T	M	indel	NM		NM		NM		NM	
#1285T	M	indel	NM		NM		NM		NM	
#1289T	M	indel	NM		NM		NM		NM	
#1290T	M	indel	NM		NM		NM		NM	
#1291T	M	indel	NM		NM		NM		NM	
#1296T	M	indel	NM		NM		NM		NM	
#1297T	M	indel	NM		NM		NM		NM	
#1301T	M	indel	NM		NM		NM		NM	
#1302T	M	indel	NM		NM		NM		NM	
#1304T	M	indel	NM		NM		NM		NM	
#1305T	M	indel	NM		NM		NM		NM	
#1306T	M	indel	NM		NM		NM		NM	
#1313T	M	indel	NM		NM		NM		NM	
#1319T	M	indel	NM		NM		NM		NM	
#2874T	M	indel	NM		NM		NM		NM	NM
#3759T	**M**	indel	NM		NM		NM		NM	NM
#1330T	M	indel	NM		NM		NM		NM	
#1331T	M	indel	NM		NM		NM		NM	
#1335T	M	indel	NM		NM		NM		NM	NM
#1336T	M	indel	NM		NM		NM		NM	NM
#2329T	M	indel	NM		NM		NM		NM	NM
#1350T	M	indel	NM		NM		NM		NM	
#1386T	M	indel	NM		NM		NM		NM	
#1388T	M	indel	NM		NM		NM		NM	
#1399T	M	indel	NM		NM		NM		NM	
#1407T	M	indel	NM		NM		NM		NM	
#1421T	M	indel	NM		NM		NM		NM	
#1433T	M	indel	NM		NM		NM		NM	
#485T	M	missense	NM		NM		NM		NM	NM
#1435T	M	indel	NM		NM		NM		NM	
#1440T	M	indel	NM		NM		NM		NM	
#1518T	M	indel	NM		NM		NM		NM	
#1523T	M	indel	NM		NM		NM		NM	
#1664T	M	indel	NM		NM		NM		NM	
#669T	NM		NM		NM		M	missense	NM	NM
#1676T	M	indel	NM		NM		NM		NM	
#1677T	M	indel	NM		NM		NM		NM	
#2191T	M	indel	NM		NM		NM		NM	
#2189T	NM		NM		NM		NM		NM	NM
#2370T	M	indel	NM		NM		NM		NM	
#754T	M	missense	NM		NM		NM		NM	
#2371T	M	indel	NM		NM		NM		NM	
#2372T	M	indel	NM		NM		NM		NM	

#2373T	M	indel	NM		NM		NM		NM	
#2375T	M	indel	NM		NM		NM		NM	
#2376T	M	indel	NM		NM		NM		NM	
#2377T	M	indel	NM		NM		NM		NM	
#2568T	M	indel	NM		NM		NM		NM	
#2571T	M	indel	NM		NM		NM		NM	
#2585T	M	indel	NM		NM		NM		NM	
#2590T	M	indel	NM		NM		NM		NM	
#2769T	M	indel	NM		NM		NM		NM	
#2793T	M	indel	NM		NM		NM		NM	
#2795T	M	indel	NM		NM		NM		NM	
#2802T	M	indel	NM		NM		NM		NM	
#2803T	M	indel	NM		NM		NM		NM	
#2826T	M	indel	NM		NM		NM		NM	
#2829T	M	indel	NM		NM		NM		NM	
#2830T	M	indel	NM		NM		NM		NM	
#2835T	M	indel	NM		NM		NM		NM	
#2837T	M	indel	NM		NM		NM		NM	
#2856T	M	indel	NM		NM		NM		NM	
#750T	NM		NM		NM		NM		NM	NM
#2900T	M	indel	NM		NM		NM		NM	
#2901T	M	indel	NM		NM		NM		NM	
#3104T	M	indel	NM		NM		NM		NM	
#3107T	M	indel	NM		NM		NM		NM	
#344T	M	indel	NM		NM		NM		NM	
#3781T	M	indel	NM		NM		NM		NM	
#382T	M	indel	NM		NM		NM		NM	
#4108T	M	indel	NM		NM		NM		NM	
#4109T	M	indel	NM		NM		NM		NM	
#2217T	NM		NM		NM		NM		NM	
#455T	M	indel	NM		NM		NM		NM	
#456T	M	indel	NM		NM		NM		NM	
#457T	M	indel	NM		NM		NM		NM	
#466T	M	indel	NM		NM		NM		NM	
#473T	M	indel	NM		NM		NM		NM	
#481T	M	indel	NM		NM		NM		NM	
#490T	M	indel	NM		NM		NM		NM	
#503T	M	indel	NM		NM		NM		NM	
#1023T	M	indel	NM		NM		NM		NM	
#1124T	M	indel	NM		NM		NM		NM	
#552T	M	indel	NM		NM		NM		NM	
#553T	M	indel	NM		NM		NM		NM	
#560T	M	indel	NM		NM		NM		NM	
#581T	M	indel	NM		NM		NM		NM	
#594T	M	indel	NM		NM		NM		NM	
#670T	M	indel	NM		NM		NM		NM	
#698T	M	indel	NM		NM		NM		NM	
#745T	M	indel	NM		NM		NM		NM	
#751T	M	indel	NM		NM		NM		NM	
#769T	M	indel	NM		NM		NM		NM	
#771T	M	indel	NM		NM		NM		NM	
#772T	M	indel	NM		NM		NM		NM	
#821T	M	indel	NM		NM		NM		NM	
#844T	M	indel	NM		NM		NM		NM	
#858T	M	indel	NM		NM		NM		NM	
#861T	M	indel	NM		NM		NM		NM	
#864T	M	indel	NM		NM		NM		NM	
#867T	M	indel	NM		NM		NM		NM	
#973T	M	indel	NM		NM		NM		NM	
#975T	M	indel	NM		NM		NM		NM	
#1666T	M	indel	NM		NM		NM		NM	
#2825T	M	missense	NM		NM		NM		NM	
#2857T	M	missense	NM		NM		NM		NM	
#4094T	M	missense	NM		NM		NM		NM	
#468T	M	missense	NM		NM		NM		NM	
#839T	M	missense	NM		NM		NM		NM	
#470T	M	indel	NM		NM		NM		NM	
#1916T	M	indel	NM		NM		NM		NM	
#1328T	M	indel	NM		NM		NM		NM	
#1329T	M	indel	NM		NM		NM		NM	
#1337T	M	indel	NM		NM		NM		NM	
#469T	M	indel	NM		NM		NM		NM	
#1434T	M	indel	NM		NM		NM		NM	
#1915T	M	indel	NM		NM		NM		NM	
#1488T	NM		NM		NM		NM		M	indel
#1489T	NM		NM		NM		NM		M	indel
#624T	NM		NM		NM		NM		M	indel
#1432T	NM		NM		NM		M	missense	NM	
#1678T	NM		NM		NM		M	missense	NM	
#389T	NM		NM		NM		NM		NM	NM
#4089T	NM		NM		NM		NM		NM	NM
#744T	NM		NM		NM		NM		NM	NM
#623T	NM		NM		NM		NM		NM	
#1241T	NM		NM		NM		NM		NM	
#1269T	NM		NM		NM		NM		NM	
#1430T	NM		NM		NM		NM		NM	
#2192T	NM		NM		NM		NM		NM	
#2219T	NM		NM		NM		NM		NM	
#2378T	NM		NM		NM		NM		NM	
#2589T	NM		NM		NM		NM		NM	
#2328T	NM		NM		NM		NM		NM	
#2591T	NM		NM		NM		NM		NM	
#2796T	NM		NM		NM		NM		NM	
#2902T	NM		NM		NM		NM		NM	
#3073T	NM		NM		NM		NM		NM	
#331T	NM		NM		NM		NM		NM	
#3410T	NM		NM		NM		NM		NM	
#3766T	NM		NM		NM		NM		NM	
#3780T	NM		NM		NM		NM		NM	
#467T	NM		NM		NM		NM		NM	
#504T	NM		NM		NM		NM		NM	
#672T	NM		NM		NM		NM		NM	
#748T	NM		NM		NM		NM		NM	
#969T	NM		NM		NM		NM		NM	
#1412T	M	indel			NM		NM		NM	
#1768T	M	indel	NM				NM		NM	
#1521T	M	indel	NM		NM				NM	
#1413T	M	indel			NM				NM	
#857T	M	indel			NM				NM	
#1410T	M	indel			NM					
#1321T	M	indel	NM		NM		NM		NM	
#1356T	M	indel			NM				NM	
#1339T	M	indel	NM		NM					
#2620T	M	indel	NM		NM		NM		NM	
#2573T	NM		NM		NM		NM		NM	
#1880T	M	indel								
#2574T	M	missense								
#2575T	M	indel								
#2576T	M	indel								
#1918T	M	indel	NM		NM		NM		NM	
#1461T	NM		M	missense	NM		NM		NM	
#1469T			M	missense						
#976T	M	indel	NM		NM		NM		NM	
#2187T	NM		NM		M	missense	NM		NM	
#362T	M	indel	NM		NM		NM		NM	
#528T	M	indel	NM		NM		NM		NM	
#2805T	M	missense	NM		NM		NM		NM	
#545T	NM		NM		NM		NM		NM	NM
#706T	NM		NM		NM		NM		NM	
#2582T	NM		NM		M	missense	NM		NM	
#1314T	NM		NM		NM		NM		NM	

Sample	Fusion_RO51	usion_FR	al_Var	RNAseq	Sequencer_RNAseq	Capture_Kit_RNAseq	Read_Size_RNAseq	brary_orientation_RNAs	EGAS_RNA-s	WES	Sequencer_E	Capture_Kit_Exome	Read_Size_E	Library_orie	EGAS_WES
#2363T															
#477T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#1129T															
#1131T															
#1275T															
#1287T															
#2196T															
#2809T															
#2823T															
#2849T															
#4058T															
#4098T															
#4117T															
#536T															
#1343T	M	NM	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#819T															
#823T															
#1250T															
#2827T															
#372T															
#974T															

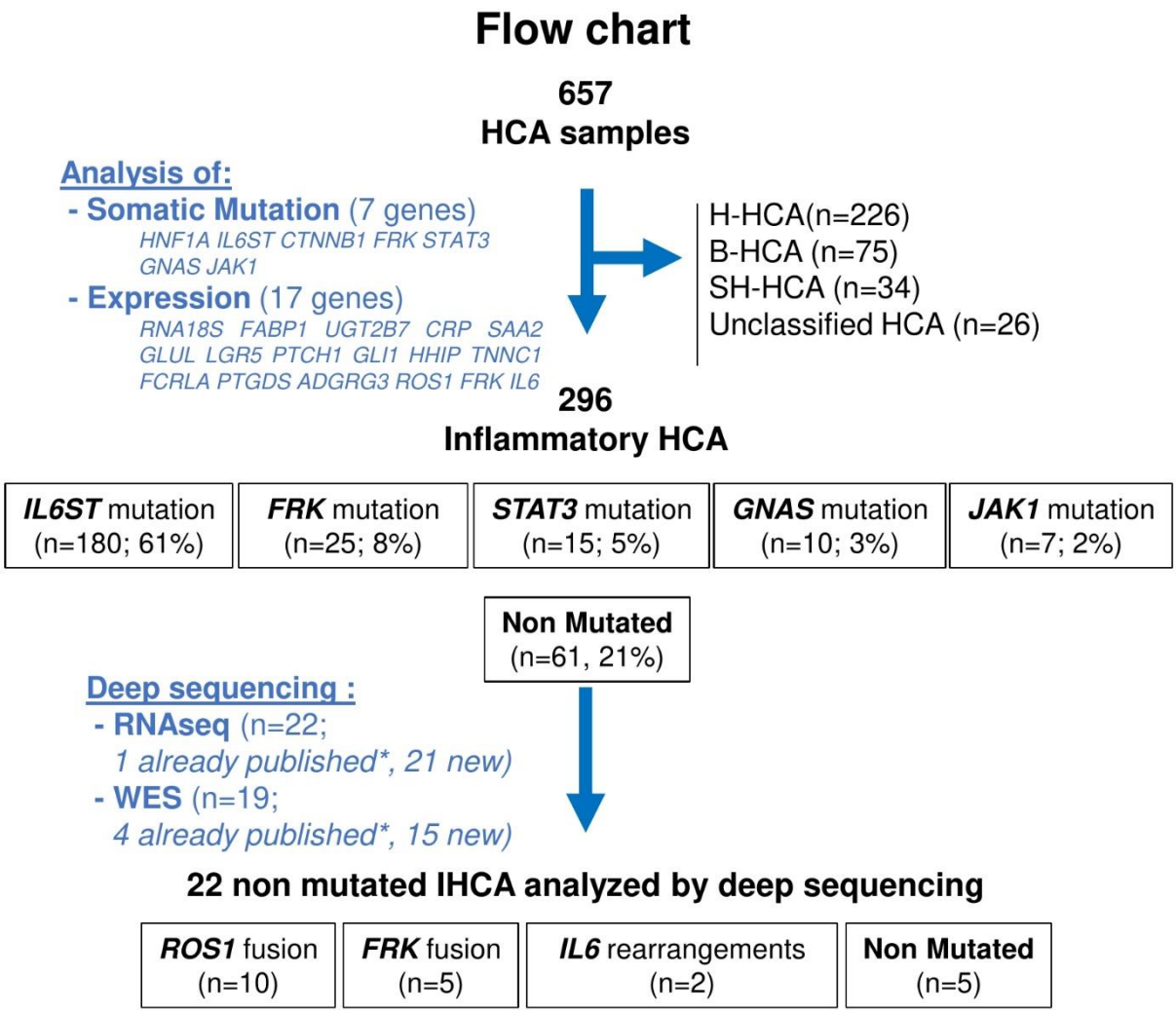
#979T															
#1896T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#2729T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#375T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#4110T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#533T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#755T															
#1851T															
#711T															
#1940T															
#757T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 4000	Twist Human Core Exome Enric	75	PE	EGAS00001003686
#222T	NM	M	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#1234T															
#2798T															
#3074T															
#3723T															
#4116T															
#1355T															
#1439T	NM	M	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001000679
#760T															
#471T	NM	M	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001000679
#2561T															
#1231T															
#1496T															
#835T	NM	M	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#554T															
#574T															
#1342T															
#1354T															
#364T															
#377T															
#390T															
#742T															
#1891T															
#378T															
#458T															
#506T															
#582T															
#631T															
#680T															
#684T															
#691T															
#955T															
#341T															
#2834T															
#1130T															
#1239T															
#1240T															
#1263T															
#1325T															
#873T	NM	M	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#2615T	NM	NM	M	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#1508T															
#2193T															
#2197T															
#2330T															
#2331T															
#2804T															
#2828T															
#2836T															
#3049T															
#4106T															
#544T															
#782T															
#970T															
#627T															
#1215T	NM	NM	M	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS0000100368S						
#1490T															
#756T	NM	NM	NM												
#1424T	NM	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001000679
#2334T															
#480T															
#1021T															
#1351T															
#2369T															
#786T															
#2773T															
#2877T															
#2879T															
#379T															
#708T															
#966T															
#1001T															
#1286T															
#1519T															
#2810T															
#2811T															
#478T															
#854T															
#958T															
#1002T															
#1004T															
#1020T															
#343T	NM	NM	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#1121T															
#1216T															
#1232T															
#1235T															
#1238T															
#1243T															
#1251T															
#1252T															
#1285T															
#1289T															
#1290T															
#1291T															
#1296T															
#1297T															
#1301T															
#1302T															
#1304T															
#1305T															
#1306T															
#1313T															
#1319T															
#2674T	NM	NM	NM												
#3759T	NM	NM	NM												
#1330T															
#1331T															
#1335T	NM	NM	NM												
#1336T	NM	NM	NM												
#2329T	NM	NM	NM												
#1350T															
#1386T															
#1388T															
#1399T															
#1407T															
#1421T															
#1433T															
#485T	NM	NM	NM												
#1435T															
#1440T															
#1518T															
#1523T															
#1664T															
#669T	NM	NM	NM												
#1676T															
#1677T															
#2191T															
#2189T	NM	NM	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS000010(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#2370T															
#754T															
#2371T															
#2372T															
#2373T															
#2375T															
#2376T															
#2377T															
#2568T															
#2571T															
#2585T															
#2590T															
#2769T															
#2793T															
#2795T															

#2802T															
#2803T															
#2826T															
#2829T															
#2830T															
#2835T															
#2837T															
#2856T															
#750T	NM	NM	NM	yes	HiSeq 2000	TruSeq Stranded mRNA	101	Stranded, PE	EGAS00001(	yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001003686
#2900T															
#2901T															
#3104T															
#3107T															
#344T															
#3781T															
#382T															
#4108T															
#4109T															
#2217T															
#455T															
#456T															
#457T															
#466T															
#473T															
#481T															
#490T															
#503T															
#1023T															
#1124T															
#552T															
#555T															
#560T															
#581T															
#594T															
#670T															
#698T															
#745T															
#751T															
#769T															
#771T															
#772T															
#821T															
#844T															
#858T															
#861T															
#864T															
#867T															
#973T															
#975T															
#1666T															
#2825T															
#2855T															
#4094T															
#468T															
#839T															
#470T															
#1916T															
#1328T															
#1329T															
#1337T															
#469T															
#1434T															
#1915T															
#1488T															
#1489T															
#624T															
#1432T															
#1678T															
#389T	NM	NM	NM	yes	HiSeq 4000	TruSeq Stranded mRNA	75	Stranded, PE	EGAS00001003685						
#4089T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS00001003685						
#744T	M	NM	NM	yes	HiSeq 4000	Next Ultra II Directional	75	Stranded, PE	EGAS00001003685						
#623T										yes	HiSeq 2000	SureSelect Human All-Exon kit v	75	PE	EGAS00001000679
#1241T															
#1269T															
#1430T															
#2192T															
#2219T															
#2378T															
#2589T															
#2328T															
#2591T															
#2796T															
#2902T															
#3073T															
#331T															
#3410T															
#3766T															
#3780T															
#467T															
#504T															
#672T															
#748T															
#969T															
#1412T															
#1768T															
#1521T															
#1413T															
#857T															
#1410T															
#1321T															
#1356T															
#1339T															
#2620T															
#2573T															
#1880T															
#2574T															
#2575T															
#2576T															
#1918T															
#1461T															
#1469T															
#976T															
#2187T															
#362T															
#528T															
#2805T															
#545T	NM	NM	NM												
#706T															
#2582T															
#1314T															

Abbreviations:

HCC=Hepatocellular Carcinoma
HCA=Hepatocellular Adenoma
IHCA=Inflammatory Hepatocellular Adenoma
bIHCA=beta-catenin activated Inflammatory Hepatocellular Adenoma
M=Mutated or Male
NM=Non-Mutated
F=Female
CTNNB1=Catenin Beta 1
IL6=Interleukin 6
ine Nucleotide Binding Protein (G Protein), Alpha Stimulating Activity F
STAT3=Signal Transducer And Activator Of Transcription 3
JAK1=Janus Kinase 1
FRK=Fyn Related Src Family Tyrosine Kinase
ROS1=ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
RNA-seq = RNA sequencing
WES=Whole Exome Sequencing

Supplementary Figure 1



Supplementary Table 3: Somatic mutations in JAK/STAT pathway genes

Sample	IL6ST_Mutation(NM_002184.3)	type_mutation_IL6ST	GNAS1_Mutation(NM_000516.4)	type_mutation_GNAS1	STAT3_Mutation(NM_139276.2)	type_mutation_STAT3	JAK1_Mutation(NM_002227.2)	type_mutation_JAK1	FRK_Mutation(NM_002031.2)	type_mutation_FRK
#480T							c.1932G>T; p.Gln644His / c.1933G>T; p.Val645Phe	missense		
#1678T							c.1932G>T; p.Gln644His / c.1933G>T; p.Val645Phe	missense		
#2334T							c.1646G>A / c.1647A>T; p.Arg549His	missense		
#222T										
#331T										
#341T					c.233T>G;p.Leu78Arg	missense			c.1037A>G;p.Glu346Gly	missense
#343T										
#344T	c.560_571del;p.Ser187_Tyr190del	indel								
#362T	c.560_571del;p.Ser187_Tyr190del	indel								
#552T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#364T	c.565_576del;p.Val189_Val192del	indel								
#372T	c.647C>A;p.Pro216His	missense								
#375T										
#377T	c.565_576del;p.Val189_Val192del	indel								
#378T										
#379T				c.1504G>T;p.Asp502Tyr / c.1972_1974delinsTAT;p.Ly		missense;indel				
#382T	c.560_571del;p.Ser187_Tyr190del	indel								
#389T										
#390T	c.557_574del;p.Tyr186_Val192delinsPhe	indel								
#455T	c.565_576del;p.Val189_Val192del	indel								
#456T	c.565_576del;p.Val189_Val192del	indel								
#457T	c.565_576del;p.Val189_Val192del	indel								
#458T										
#466T	c.560_571del;p.Ser187_Tyr190del	indel								
#467T										
#468T	c.647C>A;p.Pro216His	missense								
#469T	c.557_571del;p.Tyr186_Tyr190del	indel								
#470T	c.557_571del;p.Tyr186_Tyr190del	indel								
#471T										
#473T	c.564_575del;p.Val189_Val192del	indel								
#477T										
#478T			c.602G>A;p.Arg201His	missense						
#481T	c.615_644dup;p.Phe214_Asp215dup	indel								
#485T	c.647C>A; p.Pro216His	missense								
#490T	c.557_571del;p.Tyr186_Tyr190del	indel								
#503T	566_580delinsATG;p.Val189_Ile194delinsAspVa	indel								
#504T										
#506T										
#528T	c.557_571del;p.Tyr186_Tyr190del	indel								
#533T										
#536T	c.566_577del;p.Val189_Asn193delinsAsp	indel								
#544T									c.1132_1137del;p.Val378_Phe379de	indel
#545T										
#554T							c.2729T>C; p.Leu910Pro	missense		
#555T	c.503_580del;p.Lys168_Asn193del	indel								
#560T	c.557_571del;p.Tyr186_Tyr190del	indel								
#574T					c.496G>C;p.Glu166Gln	missense				
#581T	c.560_571del;p.Ser187_Tyr190del	indel								
#582T										
#594T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#623T										
#624T									c.1133_1138del;p.Val378_Lys380delin	indel
#627T									c.1137_1140delinsA;p.Phe379_Lys380del	indel
#631T	c.636_644del;p.Asn213_Asp215del	indel								
#669T							c.2729T>C; p.Leu910Pro	missense		
#670T	c.560_571del;p.Ser187_Tyr190del	indel								
#672T										
#680T	c.563_577del;p.Thr188_Val192del	indel								
#684T	c.565_576del;p.Val189_Val192del	indel								
#691T	c.518_532del;p.Lys173_Asp177del	indel								
#698T	c.568_579del;p.Tyr190_Asn193del	indel								
#706T										
#708T					c.1919A>T;p.Tyr640Phe	missense				
#711T	c.643_645del;p.Asp215del	indel								
#742T	c.518_532del;p.Lys173_Asp177del	indel								
#744T										
#745T	c.562_573del;p.Thr188_Phe191del	indel								
#748T										
#750T										
#751T	c.560_571del;p.Ser187_Tyr190del	indel								
#754T	c.647C>A;p.Pro216His	missense								
#755T										
#756T	c.552_566del;p.Asp185_Val189del	indel								
#757T										
#760T	c.647C>A;p.Pro216His	missense								
#769T	c.557_571del;p.Tyr186_Tyr190del	indel								
#771T	c.564_575del;p.Val189_Val192del	indel								
#772T	c.562_573del;p.Thr188_Phe191del	indel								
#782T									c.1132_1137del;p.Val378_Phe379de	indel
#786T	c.551_559del;p.Val184_Ser187delinsAla	indel								
#819T	c.560_571del;p.Ser187_Tyr190del	indel								
#821T	c.557_571del;p.Tyr186_Tyr190del	indel								
#823T	c.560_571del;p.Ser187_Tyr190del	indel								
#835T										

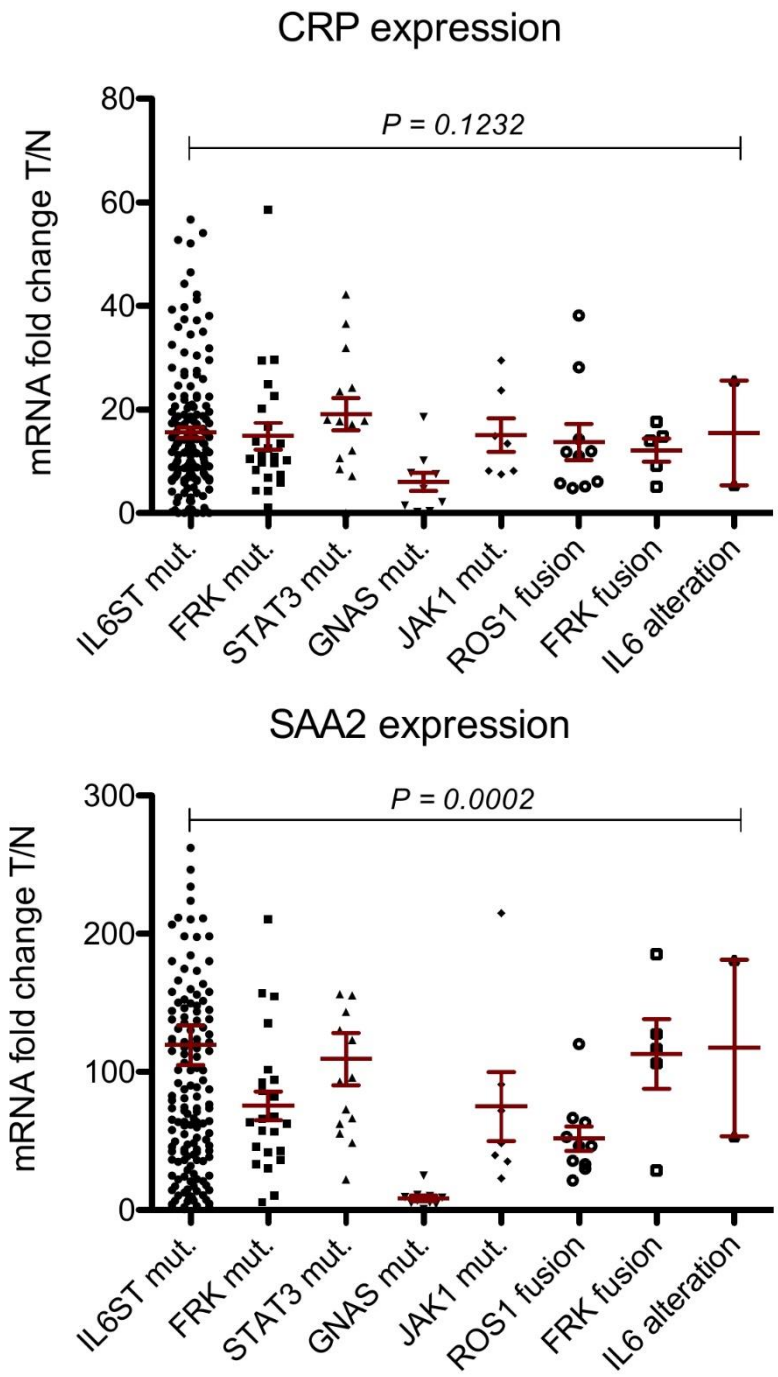
#839T	c.647C>A;p.Pro216His	missense								
#844T	c.562_573del;p.Thr188_Phe191del	indel								
#854T			c.602G>A;p.Arg201His	missense						
#857T	c.560_571del;p.Ser187_Tyr190del	indel								
#858T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#861T	c.560_571del;p.Ser187_Tyr190del	indel								
#864T	c.1252_1263del;p.Ala418_Phe421del	indel								
#867T	c.583_588del;p.Glu195_Val196del	indel								
#873T										
#955T	c.556_571delinsA;p.Tyr186_Phe191delinsIle	indel								
#958T			c.602G>A;p.Arg201His	missense						
#966T					c.1919A>T;p.Tyr640Phe	missense				
#969T										
#970T									c.1132_1137del;p.Val378_Phe379d	indel
#973T	c.569_578delinsC;p.Tyr190_Asn193delinsSer	indel								
#974T	c.647C>A;p.Pro216His	missense								
#975T	c.560_571del;p.Ser187_Tyr190del	indel								
#976T	c.570_578del;p.Phe191_Asn193del	indel								
#979T	c.646_648del;p.Pro216del	indel								
#1001T			c.601C>T;p.Arg201Cys	missense						
#1002T	c.592_618del;p.Glu199_Val207del	indel								
#1004T	c.592_618del;p.Glu199_Val207del	indel								
#1021T					c.1969_1980dup;p.Tyr657_Met660dup	indel				
#1020T	p.Tyr186_Tyr190del	indel								
#1023T	c.646_647ins33;p.Asp215_Pro216ins11	indel								
#1121T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1124T	c.571_579del;p.Phe191_Asn193del	indel								
#1129T	c.560_571del;p.Ser187_Tyr190del	indel								
#1131T	c.564_576delinsA;p.Val189_Val192del	indel								
#1130T									c.1037A>G,p.Glu346Gly/c.1049A>G,p.Tyr	missense
#1215T										
#1216T	c.1252_1263del;p.Ala418_Phe421del	indel								
#1231T										
#1232T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#1234T										
#1238T	c.573_582delinsG;p.Phe191_Ile194delinsLeu	indel								
#1239T									c.1037A>G;p.Glu346Gly	missense
#1240T									c.1132_1137del;p.Val378_Phe379d	indel
#1235T	c.565_574delinsT;p.Val189_Val192delinsPhe	indel								
#1250T	c.647C>A;p.Pro216His	missense								
#1241T										
#1243T	c.557_574del;p.Tyr186_Val192delinsPhe	indel								
#1251T	c.560_571del;p.Ser187_Tyr190del	indel								
#1252T	c.560_571del;p.Ser187_Tyr190del	indel								
#1263T									c.1132_1133delinsAAACTAAA;p.Val378delins	indel
#1269T										
#1275T	c.643_645del;p.Asp215del	indel								
#1285T	c.568_579del;p.Tyr190_Asn193del	indel								
#1286T			c.602G>A;p.Arg201His	missense						
#1287T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1289T	518_535delinsCCG;p.Lys173_Pro179delinsThrAl	indel								
#1290T	c.577_580delinsT;p.Asn193_Ile194delinsPhe	indel								
#1291T	c.564_575del;p.Val189_Val192del	indel								
#1296T	c.567_578del;p.Tyr190_Asn193del	indel								
#1297T	c.562_573del;p.Thr188_Phe191del	indel								
#1301T	c.643_645del;p.Asp215del	indel								
#1302T	c.560_571del;p.Ser187_Tyr190del	indel								
#1304T	c.565_576del;p.Val189_Val192del	indel								
#1305T	c.518_532del;p.Lys173_Asp177del	indel								
#1306T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1313T	c.560_571del;p.Ser187_Tyr190del	indel								
#1314T										
#1319T	c.560_571del;p.Ser187_Tyr190del	indel								
#1321T	c.573_582delinsG;p.Phe191_Ile194delinsLeu	indel								
#1325T									c.1037A>G,p.Glu346Gly/c.1049A>G,p.Tyr	missense
#1328T	c.567_578del;p.Tyr190_Asn193del	indel								
#1329T	c.560_571del;p.Ser187_Tyr190del	indel								
#1331T	c.567_578del;p.Tyr190_Asn193del	indel								
#1330T	c.518_532del;p.Lys173_Asp177del	indel								
#1440T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1335T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1336T	c.560_571del;p.Ser187_Tyr190del	indel								
#1337T	c.1252_1263del;p.Ala418_Phe421del	indel								
#1339T	c.564_576delinsA;p.Val189_Val192del	indel								
#1342T	c.518_532del;p.Lys173_Asp177del	indel								
#1343T										
#1351T					c.1968_1969insTTT;p.Gly656_Tyr657insPhe	indel				
#1350T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1354T	c.556_573del;p.Tyr186_Phe191del	indel								
#1355T					c.1940A>T;p.Asn647Ile	missense				
#1356T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#1386T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1388T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1407T	c.567_579delinsT;p.Tyr190_Asn193del	indel								
#1410T	c.640_642del;p.Phe214del	indel								
#1412T	c.556_573del;p.Tyr186_Phe191del	indel								

#1413T	c.564_575del;p.Val189_Val192del	indel								
#1399T	c.575_577del;p.Val192_Asn193delinsAsp	indel								
#1421T	c.565_577delinsT;p.Val189_Asn193delinsTyr	indel								
#1424T										
#1430T										
#1432T	c.1841-2A>T;p.Ala614_splice	splice					c.2108G>T;p.Ser703Ile	missense		
#1434T	c.560_571del;p.Ser187_Tyr190del	indel								
#1433T	c.618_641dup;p.Asn213_Phe214dup	indel								
#1435T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1439T										
#1461T			c.601C>T;p.Arg201Cys	missense						
#1469T			c.602G>A;p.Arg201His	missense						
#1488T									c.1132_1137del;p.Val378_Phe379d	indel
#1489T									c.1133_1138del;p.Val378_Lys380delin	indel
#1490T							c.2168C>A;p.Ala723Asp	missense		
#1496T									c.1037A>G;p.Glu346Gly	missense
#1508T									c.1132_1137del;p.Val378_Phe379d	indel
#1518T	c.556_571delinsA;p.Tyr186_Phe191delinsIle	indel								
#1519T			c.602G>A;p.Arg201His	missense						
#1521T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#1523T	c.518_532del; p.Lys173_Asp177del	indel								
#1664T	c.643_645del;p.Asp215del	indel								
#1666T	c.564_576delinsA;p.Val189_Val192del	indel								
#1676T	c.557_571del;p.Tyr186_Tyr190del	indel								
#1677T	c.560_571del;p.Ser187_Tyr190del	indel								
#1768T	c.560_571del;p.Ser187_Tyr190del	indel								
#1851T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#1880T	c.560_571del;p.Ser187_Tyr190del	indel								
#1891T										
#1896T										
#1915T	c.560_571del;p.Ser187_Tyr190del	indel								
#1916T	c.560_571del;p.Ser187_Tyr190del	indel								
#1918T	c.553_567del; p.Asp185_Val189del	indel								
#1940T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#2187T					c.1919A>T; p.Tyr640Phe / c.233T>G;p.Leu78A	missense				
#2189T										
#2191T	c.557_571del;p.Tyr186_Tyr190del	indel								
#2192T										
#2193T									c.1133_1138del;p.Val378_Lys380delin	indel
#2196T	c.522_536del;p.Lys175_Pro179del	indel								
#2197T									c.1131_1137delinsG;p.Val378_Phe37	indel
#2217T										
#2219T										
#2328T										
#2329T	c.580_585del;p.Ile194_Glu195del	indel								
#2330T									c.1037A>G; p.Glu346Gly / c.1103A>G; p.Ty	missense
#2331T									c.1133_1134insAAA; p.Val378_Phe379i	indel
#2363T					c.1972 A>T; p.Lys658Tyr	missense				
#2369T					c.1972_1974delinsTAT; p.Lys658Tyr	missense				
#2370T	c.565_579del; p.Val189_Asn193del	indel								
#2371T	c.556_573del;p.Tyr186_Phe191del	indel								
#2372T	c.565_579del;p.Val189_Asn193del	indel								
#2373T	c.569_577del;p.Phe191_Asn193del	indel								
#2375T	c.574_582del;p.Val192_Ile194del	indel								
#2376T	c.569_578delinsC;p.Tyr190_Asn193delinsSer	indel								
#2377T	c.565_579del;p.Val189_Asn193del	indel								
#2378T										
#2561T	c.647C>A;p.Pro216His	missense								
#2568T	c.560_571del;p.Ser187_Tyr190del	indel								
#2571T	c.570_578del;p.Phe191_Asn193del	indel								
#2573T										
#2574T	c.564_575del;p.Val189_Val192del	indel								
#2575T	c.557_571del;p.Tyr186_Tyr190del	indel								
#2576T	c.560_571del;p.Ser187_Tyr190del	indel								
#2582T					c.1919A>T;pTyr640Phe	missense				
#2585T	c.579_584del;p.Asn193_Glu195delinsLys	indel								
#2589T										
#2590T	c.582_587del; p.Glu195_Val196del	indel								
#2591T										
#2615T										
#2620T	c.557_571del;p.Tyr186_Tyr190del	indel								
#2729T										
#2769T	c.643_645del; p.Asp215del	indel								
#2773T					c.1969_1980dup; p.Tyr657_Met660dup	indel				
#2793T	c.568_579del;p.Tyr190_Asn193del	indel								
#2795T	c.570_578del; p.Phe191_Asn193del	indel								
#2796T										
#2798T										
#2802T	c.560_571del;p.Ser187_Tyr190del	indel								
#2803T	c.562_573del;p.Thr188_Phe191del	indel								
#2804T									c.1132_1137del ; p.Val378_Phe379d	indel
#2805T	c.647C>A;p.Pro216His	missense								
#2809T	c.642_645delinsA;p.Phe214_Asp215delinsLeu	indel								
#2810T			c.601C>T;p.Arg201Cys	missense						
#2811T			c.601C>T;p.Arg201Cys	missense						
#2823T	c.560_571del; p.Ser187_Tyr190del	indel								

#2825T	c.647C>A;p.Pro216His	missense								
#2826T	c.557_571del;p.Tyr186_Tyr190del	indel								
#2827T	c.647C>A;p.Pro216His	missense								
#2828T									c.1129insAGA ; p.insPro377	indel
#2829T	c.560_571del; p.Ser187_Tyr190del	indel								
#2830T	c.562_573del;p.Thr188_Phe191del	indel								
#2834T	c.557_571del;p.Tyr186_Tyr190del	indel							c.1133_1138del;p.Val378_Lys380delin	indel
#2835T	c.565_576del;p.Val189_Val192del	indel								
#2836T									c.1133_1139del;p.Val378_Lys380delin	indel
#2837T	c.557_571del;p.Tyr186_Tyr190del	indel								
#2849T	c.565_576del;p.Val189_Val192del	indel								
#2856T	c.565_576del;p.Val189_Val192del	indel								
#2857T	c.563_574del;p.Thr188_Val192delinsIle	indel								
#2874T	c.1252_1263del; p.Ala418_Phe421del	indel								
#2877T					c.233T>G;p.Leu78Arg	missense				
#2879T					c.1969_1980dup; p.Tyr657_Met660dup	indel				
#2900T	c.565_576del;p.Val189_Val192del	indel								
#2901T	c.565_576del;p.Val189_Val192del	indel								
#2902T										
#3049T									1132_1137del;p.Val378_Phe379de	indel
#3073T										
#3074T										
#3104T	c.518_532del; p.Lys173_Asp177del	indel								
#3107T	c.563_580del; p.Thr188_Asn193del	indel								
#3410T										
#3723T										
#3759T	c.555_566del; p.Asp185_Val189delinsGlu	indel								
#3766T										
#3780T										
#3781T	c.571_579del; p.Phe191_Asn193del	indel								
#4058T	c.565_576del;p.Val189_Val192del	indel								
#4089T										
#4094T	c.647C>A; p.Pro216His	missense								
#4098T	c.565_576del; p.Val189_Val192del	indel								
#4106T									1037A>G; p.Glu346Gly	missense
#4108T	c.556_573del; p.Tyr186_Phe191del	indel								
#4109T	c.560_571del; p.Ser187_Tyr190del	indel								
#4110T										
#4116T										
#4117T	c.565_576del; p.Val189_Val192del	indel								

Abbreviations:
IL6=Interleukin 6
GNAS=Guanine Nucleotide Binding Protein (G Protein), Alpha Stimulating Activity Polypeptide 1
STAT3=Signal Transducer And Activator Of Transcription 3
JAK1=Janus Kinase 1
FRK=Fyn Related Src Family Tyrosine Kinase

Supplementary Figure 2



Supplementary Table 4: Somatic mutations from Whole Exome Sequencing in 19 samples

Sample	Histological.Tumor.Initia	CHROM	POS_hg19	POS_hg38	REF	ALT	Hugo_Symbol	Protein_Change	Variant_Classification	Variant_Type	VAF	FILTER
#471T	HCA	chr8	144391654	143309484	A	C	TOP1MT	p.P588C	Missense_Mutation	SNV	0,231788079	PASS
#471T	HCA	chr8	144990757	143916589	G	C	PLEC	p.A4548G	Missense_Mutation	SNV	0,321100917	PASS
#471T	HCA	chr17	78223022	80249223	G	T	SLC26A11	p.G531V	Missense_Mutation	SNV	0,27173913	PASS
#471T	HCA	chr1	114354555	113811933	G	A	RSBN1	p.A160A	Silent	SNV	0,157894737	PASS
#471T	HCA	chr2	179635319	178770592	T	C	TTN	p.N2734D	Missense_Mutation	SNV	0,2875	PASS
#471T	HCA	chr6	27861316	27893538	G	C	HIST1H2BO	p.D26H	Missense_Mutation	SNV	0,343137255	PASS
#471T	HCA	chr6	109787264	109466061	C	T	ZBTB24	p.G628G	Silent	SNV	0,111111111	PASS
#471T	HCA	chr6	165723063	165309574	G	C	C6orf118	p.R5G	Missense_Mutation	SNV	0,154761905	PASS
#471T	HCA	chr22	21384369	21030080	G	A	SLC7A4	p.S418S	Silent	SNV	0,36	PASS
#471T	HCA	chr16	84449104	84415498	A	C	ATP2C2	p.K177N	Missense_Mutation	SNV	0,22	PASS
#471T	HCA	chr19	40828162	40322255	C	A	C19orf47	p.G299V	Missense_Mutation	SNV	0,251101322	PASS
#471T	HCA	chr3	45677668	45636176	G	T	LIMD1	p.V479L	Missense_Mutation	SNV	0,255813953	PASS
#471T	HCA	chr5	31299738	31299631	A	G	CDH6	p.S271G	Splice_Site	SNV	0,304347826	PASS
#471T	HCA	chr5	120021947	120686252	T	G	PRR16	p.L153R	Missense_Mutation	SNV	0,121212121	PASS
#471T	HCA	chr7	98506371	98908748	T	A	TRIP4P	p.L379Q	Missense_Mutation	SNV	0,285714286	PASS
#471T	HCA	chr7	98609721	99012098	C	T	TRIP4P	p.R377SW	Missense_Mutation	SNV	0,272727273	PASS
#471T	HCA	chr7	100681483	101038202	A	G	MUC17	p.T2262T	Silent	SNV	0,272	PASS
#471T	HCA	chr14	21831003	21362844	T	C	SUPT16H	p.I539V	Missense_Mutation	SNV	0,108108108	PASS
#471T	HCA	chr12	21459836	21306902	TTAAAATCTGGGTTCCAT	C	SLCO1A2	p.NGTQILR135fs	Frame_Shift_Del	DEL	0,107142857	PASS
#471T	HCA	chr12	31249835	31096901	G	A	DDX11	p.R558Q	Missense_Mutation	SNV	0,173913043	PASS
#471T	HCA	chr10	6557001	6515039	C	T	PRKCQ	p.V33I	Missense_Mutation	SNV	0,1875	PASS
#1343T	HCA	chr20	48140657	49524120	C	A	PTGIS	p.E26S*	Nonsense_Mutation	SNV	0,234234234	PASS
#1343T	HCA	chr20	62863566	64232213	G	A	MYT1	p.A909T	Missense_Mutation	SNV	0,224489796	PASS
#1343T	HCA	chr5	115837936	116502240	A	G	SEMA6A	p.I63T	Missense_Mutation	SNV	0,242424242	PASS
#1343T	HCA	chr5	147820738	148441175	G	A	FBXO38	p.R1109K	Missense_Mutation	SNV	0,262773723	PASS
#1343T	HCA	chr5	174938514	175511511	C	T	SFXN1	p.L165L	Silent	SNV	0,144654088	PASS
#1343T	HCA	chr1	23111399	22784906	C	T	EPHB2	p.A214V	Missense_Mutation	SNV	0,207253886	PASS
#1343T	HCA	chr1	25072111	24745620	G	A	CLIC4	p.V23I	Missense_Mutation	SNV	0,142857143	t_lod_fstar
#1343T	HCA	chr1	63788807	63323136	T	C	FOXD3	p.D26D	Silent	SNV	0,304347826	PASS
#1343T	HCA	chr1	153905222	153932746	C	G	DENND4B	p.G1219R	Missense_Mutation	SNV	0,212121212	PASS
#1343T	HCA	chr1	216465689	216292347	A	G	USH2A	p.Y556Y	Silent	SNV	0,286956522	PASS
#1343T	HCA	chrX	153927694	154699419	C	T	GAB3	p.G406E	Missense_Mutation	SNV	0,193548387	PASS
#1343T	HCA	chr19	9084834	8974158	G	T	MUC16	p.S23275	Silent	SNV	0,252873563	PASS
#1343T	HCA	chr19	13059076	12948262	C	A	RAD23A	p.P107H	Missense_Mutation	SNV	0,115384615	t_lod_fstar
#1343T	HCA	chr4	77231007	76309854	A	G	FAM47E-STBD1	p.I311V	Missense_Mutation	SNV	0,237837838	PASS
#1343T	HCA	chr16	70884418	70850515	C	A	HYDIN	p.C4195F	Missense_Mutation	SNV	0,209302326	PASS
#1343T	HCA	chr16	70884419	70850516	A	C	HYDIN	p.C4195G	Missense_Mutation	SNV	0,209302326	PASS
#1343T	HCA	chr14	68221790	67755073	T	C	ZFYVE26	p.M2322V	Missense_Mutation	SNV	0,260869565	PASS
#1343T	HCA	chr14	99925461	99459124	T	C	SETD3	p.N136S	Missense_Mutation	SNV	0,303797468	PASS
#1343T	HCA	chr14	102461583	101995246	C	T	DYNC1H1	p.I1170I	Silent	SNV	0,247191011	PASS
#1343T	HCA	chr2	106746217	106129761	G	T	UXS1	p.L159M	Missense_Mutation	SNV	0,241176471	PASS
#1343T	HCA	chr18	56202862	58535630	C	G	ALPK2	p.G1519G	Silent	SNV	0,220095694	PASS
#1343T	HCA	chr11	56143847	56376371	A	G	OR8U1	p.I250V	Missense_Mutation	SNV	0,234042553	vents;homologous_mapp
#1343T	HCA	chr3	108076930	108358083	C	T	HHLA2	p.L309F	Missense_Mutation	SNV	0,156862745	PASS
#1343T	HCA	chr3	186386829	186669040	C	T	HRG	p.P97S	Missense_Mutation	SNV	0,224137931	PASS
#1343T	HCA	chr3	197505276	197778405	G	T	FYTTD1	p.V267F	Missense_Mutation	SNV	0,301886792	PASS
#1343T	HCA	chr17	1373763	1470469	C	T	MYO1C	p.A743T	Missense_Mutation	SNV	0,142857143	t_lod_fstar
#1343T	HCA	chr17	77769355	79795556	C	T	CBX8	p.A83A	Silent	SNV	0,279069767	PASS
#1343T	HCA	chr17	79914528	81956652	C	G	NOTUM	p.R329P	Missense_Mutation	SNV	0,284722222	PASS
#1343T	HCA	chr12	124941695	124457149	G	A	NCOR2	p.A240V	Missense_Mutation	SNV	0,210526316	PASS
#1343T	HCA	chr15	31355405	31063202	G	C	TRPM1	p.P294R	Missense_Mutation	SNV	0,206896552	PASS
#1343T	HCA	chr15	44951472	44659274	A	T	SPG11	p.S158T	Missense_Mutation	SNV	0,310344828	PASS
#1343T	HCA	chr9	117071558	114309278	C	T	COL27A1	p.R1746W	Missense_Mutation	SNV	0,207207207	PASS
#1343T	HCA	chr9	135117267	132241880	C	A	NTNG2	p.N454K	Missense_Mutation	SNV	0,130434783	t_lod_fstar
#1343T	HCA	chr7	21609806	21570188	A	G	DNAH11	p.R438R	Silent	SNV	0,191176471	PASS
#1343T	HCA	chr7	83034788	83405472	A	G	SEMA3E	p.F326L	Missense_Mutation	SNV	0,235294118	PASS
#1343T	HCA	chr7	100086942	100489319	C	A	NYAP1	p.S533Y	Missense_Mutation	SNV	0,136363636	t_lod_fstar
#1424T	HCA	chrX	70443636	71223786	G	GTCATCT	GJB1	p.31_32insIF	In_Frame_Ins	INS	0,186567164	PASS
#1424T	HCA	chr16	21712217	21700896	G	T	OTOA	p.L283L	Silent	SNV	0,204301075	PASS
#1424T	HCA	chr8	21856713	21999202	AC	A	XPO7	p.Y847fs	Frame_Shift_Del	DEL	0,275862069	PASS
#1424T	HCA	chr12	21013980	20861046	A	G	SLCO1B3	p.N130S	Missense_Mutation	SNV	0,285714286	PASS
#1424T	HCA	chr2	196548462	196683738	G	T	SLC39A10	p.G350C	Missense_Mutation	SNV	0,294117647	PASS
#1424T	HCA	chr2	198571711	197706987	A	G	mars-02	p.T528A	Missense_Mutation	SNV	0,321888412	PASS
#1424T	HCA	chr3	41266136	41224645	T	C	CTNNB1	p.S45P	Missense_Mutation	SNV	0,27607362	PASS
#1424T	HCA	chr3	112718267	112999420	ATAAAAATTTGTTTTCTTA	A	GTPBP8	p.A223del	Splice_Site	DEL	0,270833333	PASS
#1424T	HCA	chr17	41004292	42852275	C	T	AOC3	p.A311V	Missense_Mutation	SNV	0,533333333	PASS
#1424T	HCA	chr17	80049332	82091456	G	A	FASN	p.L420L	Silent	SNV	0,4375	PASS
#1424T	HCA	chr9	140250762	137356310	T	G	EXD3	p.R239R	Silent	SNV	0,322580645	PASS
#1424T	HCA	chr9	140392635	137498183	G	A	PNPLA7	p.S582L	Missense_Mutation	SNV	0,211538462	PASS
#1424T	HCA	chr1	186287894	186318762	G	A	TPR	p.P2212L	Missense_Mutation	SNV	0,181818182	PASS
#1424T	HCA	chr10	48388031	47351329	A	T	RBP3	p.A940A	Silent	SNV	0,248648649	PASS
#1424T	HCA	chr10	101557010	99797253	G	T	ABCC2	p.K263N	Missense_Mutation	SNV	0,254237288	PASS
#1424T	HCA	chr19	8966741	8856065	T	C	MUC16	p.R14404R	Silent	SNV	0,395833333	PASS
#1439T	HCA	chr1	16577371	16250876	C	A	FBXO42	p.D650Y	Missense_Mutation	SNV	0,263157895	PASS
#1439T	HCA	chr1	27105955	26779464	C	T	ARID1A	p.Q1855*	Nonsense_Mutation	SNV	0,170454545	PASS
#1439T	HCA	chr1	33235860	32770259	G	T	KIAA1522	p.L301L	Silent	SNV	0,218543046	PASS
#1439T	HCA	chr1	36029439	35563838	G	T	NCDN	p.G561V	Missense_Mutation	SNV	0,261363636	PASS
#1439T	HCA	chr1	153742249	153769773	A	G	INTS3	p.Q773R	Missense_Mutation	SNV	0,257425743	PASS
#1439T	HCA	chr1	161136956	161167166	G	T	PPOX	p.A52S	Missense_Mutation	SNV	0,272727273	PASS
#1439T	HCA	chr12	15103534	14950600	A	G	ARHGDIB	p.M38T	Missense_Mutation	SNV	0,225	PASS
#1439T	HCA	chr12	25702459	25549525	T	C	IFLTD1	p.V16V	Silent	SNV	0,234042553	PASS
#1439T	HCA	chr18	65180925	67513688	C	A	DSEL	p.W317C	Missense_Mutation	SNV	0,228813559	PASS
#1439T	HCA	chr7	6189847	6150216	C	T	USP42	p.L674L	Silent	SNV	0,307692308	PASS
#1439T	HCA	chr7	47876601	47837003	G	A	PKD1L1	p.T1954M	Missense_Mutation	SNV	0,214285714	PASS
#1439T	HCA	chr7	123119855	123479801	C	A	IQUB	p.L468F	Missense_Mutation	SNV	0,333333333	PASS
#1439T	HCA	chr7	139762530	140062730	C	T	PARP12	p.E40K	Missense_Mutation	SNV	0,25	PASS
#1439T	HCA	chr7	151842334	152145249	C	T	KMT2C	p.R4693Q	Missense_Mutation	SNV	0,211764706	PASS
#1439T	HCA	chrX	53349824	53320626	C	T	IQSEC2	p.L166L	Silent	SNV	0,35	PASS
#1439T	HCA	chrX	106845648	107602418	C	A	FRMPD3	p.T1493N	Missense_Mutation	SNV	0,135416667	PASS
#1439T	HCA	chrX	139586928	140504763	G	T	SOX3	p.P100T	Missense_Mutation	SNV	0,142857143	PASS
#1439T	HCA	chr3	1394027	1352343	G	T	CNTN6	p.G462C	Missense_Mutation	SNV	0,35443038	PASS
#1439T	HCA	chr3	44683972	44642480	G	A	ZNF197	p.G450G	Silent	SNV	0,20754717	PASS
#1439T	HCA	chr3	45637381	45595889	T	A	LIMD1	p.F337Y	Missense_Mutation	SNV	0,259259259	PASS
#1439T	HCA	chr3	171969171	172251381	T	G	FNDC3B	p.P210P	Silent	SNV	0,293478261	PASS
#1439T	HCA	chr19	6312466	6312455	G	C	ACER1	p.L46L	Silent	SNV	0,236842105	PASS
#1439T	HCA	chr17	3919620	4016326	G	C	ZZEF1	p.F2714L	Missense_Mutation	SNV	0,208	PASS
#1439T	HCA	chr17	63010387	65014269	A	C	GNA13	p.L374L	Silent	SNV	0,283333333	PASS
#1439T	HCA	chr5	149748277	150368714	A	C	TCOF1	NA	Splice_Site	SNV	0,266666667	PASS
#1439T	HCA	chr5	150072470	150692908	T	A	RBM22	p.P373P	Silent	SNV	0,375	PASS
#1439T	HCA	chr5	162886910	163459904	G	T	NUDCD2	p.S49R	Missense_Mutation	SNV	0,204081633	PASS
#1439T	HCA	chr2	24222614	23999744	A	G	UBXN2A	p.T219T	Silent	SNV	0,261538462	PASS
#1439T	HCA	chr2	27464238	27241370	A	G	CAD	p.K1890E	Missense_Mutation	SNV	0,191489362	PASS
#1439T	HCA	chr2	29416436	29193570	G	A	ALK	p.T1506M	Missense_Mutation	SNV	0,266666667	PASS
#1439T	HCA	chr2	102789219	102172759	A	T	IL1R1	p.R304S	Missense_Mutation	SNV	0,264705882	PASS
#1439T	HCA	chr2	148683685	147926116	T	TA	ACVR2A	p.K435fs	Frame_Shift_Ins	INS	0,194805195	PASS
#1439T	HCA	chr2	179394748	178530021	A	G	TTN	p.Y33849Y	Silent	SNV	0,272727727	PASS
#1439T	HCA	chr10	101594208	99834451	C	A	ABCC2	p.I1110I	Silent	SNV	0,170731707	PASS
#1439T	HCA	chr10	115804707	114044948	G	A	ADRB1	p.P272P	Silent	SNV	0,181818182	PASS
#1439T	HCA	chr11	94564658	94831492	A	T	AMOTL1	p.E533D	Missense_Mutation	SNV	0,357142857	PASS
#1439T	HCA	chr21	30464877	29092556	C	G	MAP3K7CL	p.P115P	Silent	SNV	0,191489362	PASS
#1439T	HCA	chr16	67418914	67385011	C	A	LRRC36	p.S729S	Silent	SNV	0,257142857	PASS
#1439T	HCA	chr16	74949815	74915917	G	T	WDR59	p.Q393K	Missense_Mutation	SNV	0,242214533	PASS

#1439T	HCA	chr20	5157350	5176704	C	T	CDS2	p.V116V	Silent	SNV	0,260869565	PASS
#1439T	HCA	chr6	71289193	70579490	C	T	C6orf57	p.S47S	Silent	SNV	0,256097561	PASS
#1439T	HCA	chr6	116938047	116616884	G	A	RSPH4A	p.P87P	Silent	SNV	0,318518519	PASS
#1439T	HCA	chr8	22880498	23022985	C	G	TNFRSF10B	NA	Splice_Site	SNV	0,293333333	PASS
#1439T	HCA	chr8	41504023	41646504	A	T	NKX6-3	p.L118Q	Missense_Mutation	SNV	0,288888889	PASS
#1439T	HCA	chr8	125074230	124061989	C	G	FER1L6	p.P1095P	Silent	SNV	0,21875	PASS
#2189T	HCA	chr6	4049256	4049022	C	A	PRPF4B	p.R648S	Missense_Mutation	SNV	0,130434783	t_lod_fstar
#2189T	HCA	chr7	97862808	98233496	G	T	TECPR1	p.L533M	Missense_Mutation	SNV	0,130434783	t_lod_fstar
#2189T	HCA	chr19	42797003	42292851	G	T	CIC	p.S1154I	Missense_Mutation	SNV	0,166666667	t_lod_fstar
#222T	HCA	chr12	25302582	25149648	C	A	CASC1	p.A183S	Missense_Mutation	SNV	0,102564103	t_lod_fstar
#222T	HCA	chrX	20183015	20164897	A	T	RPS6KA3	NA	Splice_Site	SNV	0,253521127	PASS
#222T	HCA	chrX	20183016	20164898	C	T	RPS6KA3	NA	Splice_Site	SNV	0,27027027	PASS
#222T	HCA	chr6	47202493	47234757	C	A	TNFRSF21	p.V551L	Missense_Mutation	SNV	0,111111111	t_lod_fstar
#222T	HCA	chr6	144999583	144678447	A	C	UTRN	p.K2507N	Missense_Mutation	SNV	0,111607143	PASS
#222T	HCA	chr1	200946009	200976881	G	T	KIF21B	p.V1446V	Silent	SNV	0,107142857	t_lod_fstar
#222T	HCA	chr17	3438936	3535642	C	A	TRPV3	p.A239S	Missense_Mutation	SNV	0,142857143	t_lod_fstar
#222T	HCA	chr17	39973440	41817188	G	T	FKBP10	p.G126W	Missense_Mutation	SNV	0,125	t_lod_fstar
#222T	HCA	chr11	119217031	119346321	G	T	MFRP	p.P36P	Silent	SNV	0,217391304	PASS
#222T	HCA	chr9	2622230	2622230	C	A	VLDLR	p.A14E	Missense_Mutation	SNV	0,214285714	t_lod_fstar
#222T	HCA	chr8	10383213	10525703	C	A	PRSS55	p.R40S	Missense_Mutation	SNV	0,2	PASS
#222T	HCA	chr15	24922279	24677132	C	T	NPAP1	p.A422V	Missense_Mutation	SNV	0,124378109	PASS
#222T	HCA	chr10	120863641	119104129	CGCGGCGGCGGAAGATG	C	FAM45A	NA	Start_Codon_Del	DEL	0,15	PASS
#222T	HCA	chr3	48453957	48412548	C	A	PLXNB1	p.V1643L	Missense_Mutation	SNV	0,108095652	PASS
#222T	HCA	chr3	164712149	164994361	T	C	SI	p.S1579S	Silent	SNV	0,105263158	PASS
#222T	HCA	chr22	51065642	50627214	G	C	ARSA	p.P137P	Silent	SNV	0,133333333	PASS
#222T	HCA	chr19	36530307	36039405	T	C	THAP8	p.Q197R	Missense_Mutation	SNV	0,1	PASS
#222T	HCA	chr19	44351117	43846965	A	C	ZNF283	p.K122Q	Missense_Mutation	SNV	0,134615385	PASS
#2615T	HCA	chr10	134659633	132846129	C	A	TTC40	p.Q2122H	Missense_Mutation	SNV	0,272727273	PASS
#2615T	HCA	chr20	37174999	38546356	A	G	RALGAPB	p.H943R	Missense_Mutation	SNV	0,189393939	PASS
#2615T	HCA	chr14	77605915	77139572	A	T	ZDHHC22	p.F56Y	Missense_Mutation	SNV	0,123809524	PASS
#2615T	HCA	chr1	28223561	27897050	C	A	RPA2	p.L160L	Silent	SNV	0,125	t_lod_fstar
#2615T	HCA	chr1	180022207	180053072	G	T	CEP350	p.S1632I	Missense_Mutation	SNV	0,281818182	PASS
#2615T	HCA	chr1	236979804	236816504	A	T	MTR	p.E242V	Missense_Mutation	SNV	0,169230769	PASS
#2615T	HCA	chr8	26722021	26864504	C	T	ADRA1A	p.V156I	Missense_Mutation	SNV	0,181818182	PASS
#2615T	HCA	chr8	38205493	38347975	A	C	WHSC1L1	p.L66R	Missense_Mutation	SNV	0,172932331	PASS
#2615T	HCA	chr6	3123976	3123742	C	A	BPHL	p.L65I	Missense_Mutation	SNV	0,14084507	PASS
#2615T	HCA	chr6	56472969	56608171	A	T	DST	p.L1942I	Missense_Mutation	SNV	0,269230769	PASS
#2615T	HCA	chr11	113103452	113232730	T	C	NCAM1	p.S352P	Missense_Mutation	SNV	0,179487179	PASS
#2615T	HCA	chr4	75681110	74755900	C	T	BTC	p.G80G	Silent	SNV	0,266666667	PASS
#2615T	HCA	chr4	76878739	75957586	T	C	SDAD1	p.E567E	Silent	SNV	0,177777778	PASS
#2615T	HCA	chr4	90872904	89951753	T	A	MMRN1	NA	Splice_Site	SNV	0,112676056	PASS
#2615T	HCA	chr4	94690579	93769428	G	T	GRID2	p.G860V	Missense_Mutation	SNV	0,142857143	PASS
#2615T	HCA	chr7	29918696	29879080	G	T	WIPF3	p.G99*	Nonsense_Mutation	SNV	0,275862069	PASS
#2615T	HCA	chr7	73254374	73840044	AT	A	WBSCR27	p.H155fs	Frame_Shift_Del	DEL	0,25	PASS
#2615T	HCA	chr17	38572320	40416068	A	G	TOP2A	p.V90A	Splice_Site	SNV	0,206896552	PASS
#2615T	HCA	chrX	99596960	100341962	G	A	PCDH19	p.A930V	Missense_Mutation	SNV	0,170731707	PASS
#2615T	HCA	chrX	142716743	143628927	C	T	SLITRK4	p.E728K	Missense_Mutation	SNV	0,143478261	PASS
#2615T	HCA	chrX	153247645	153982194	C	T	TMEM187	p.P44P	Silent	SNV	0,243902439	PASS
#343T	HCA	chr3	41266137	41224646	C	A	CTNNB1	p.S45Y	Missense_Mutation	SNV	0,185430464	PASS
#343T	HCA	chr3	49053706	49016273	T	C	DALRD3	p.Y405C	Missense_Mutation	SNV	0,209302326	PASS
#343T	HCA	chr3	96945109	97226265	T	A	EPHA6	p.A372A	Splice_Site	SNV	0,3	PASS
#343T	HCA	chr12	55887053	55493269	G	A	OR6C68	p.A29T	Missense_Mutation	SNV	0,2109375	PASS
#343T	HCA	chr12	57572189	57178406	A	T	LRP1	p.E1470V	Missense_Mutation	SNV	0,191304348	PASS
#343T	HCA	chr12	117962850	117525045	G	A	KSR2	p.R676C	Missense_Mutation	SNV	0,226600985	PASS
#343T	HCA	chr9	112025368	109263088	T	C	EPB41L4B	p.Y198C	Missense_Mutation	SNV	0,285714286	PASS
#343T	HCA	chr9	115337209	112574929	G	C	KIAA1958	p.K283N	Missense_Mutation	SNV	0,29245283	PASS
#343T	HCA	chr13	86369328	85795193	T	C	SLITRK6	p.Y439C	Missense_Mutation	SNV	0,265957447	PASS
#343T	HCA	chr4	24558015	24556392	C	A	DHX15	p.M240I	Missense_Mutation	SNV	0,217391304	PASS
#343T	HCA	chr4	86916168	85995015	A	C	ARHGAP24	p.N454T	Missense_Mutation	SNV	0,18018018	PASS
#343T	HCA	chr17	2599524	2696230	C	A	CLUH	p.E735*	Nonsense_Mutation	SNV	0,115384615	t_lod_fstar
#343T	HCA	chr17	20149372	20246059	T	C	SPECC1	p.L829L	Silent	SNV	0,154471545	PASS
#343T	HCA	chr17	36490917	38335034	C	G	GPR179	p.V548V	Splice_Site	SNV	0,226415094	PASS
#343T	HCA	chr17	72436516	74440377	C	A	GPRC5C	p.P246T	Missense_Mutation	SNV	0,12244898	PASS
#343T	HCA	chr8	77617928	76705693	G	GA	ZFHX4	p.KS36fs	Frame_Shift_Ins	INS	0,203208556	PASS
#343T	HCA	chr8	127569184	126556939	T	C	FAM84B	p.I151V	Missense_Mutation	SNV	0,223140496	PASS
#343T	HCA	chr6	43188575	43220837	C	T	CUL9	p.R2172C	Missense_Mutation	SNV	0,202702703	PASS
#343T	HCA	chr10	17891740	17849736	C	A	MRC1L1	p.F407L	Missense_Mutation	SNV	0,219858156	PASS
#343T	HCA	chr2	1499905	1496133	C	T	TPO	p.D717D	Silent	SNV	0,115384615	PASS
#343T	HCA	chr2	74776576	74549449	G	A	LOXL3	p.G204G	Silent	SNV	0,119565217	PASS
#343T	HCA	chr2	103067472	102451012	C	A	IL18RAP	p.P459T	Missense_Mutation	SNV	0,198347107	PASS
#343T	HCA	chr2	128608191	127850617	G	A	POLR2D	p.A108V	Missense_Mutation	SNV	0,2	PASS
#343T	HCA	chr1	39845104	39379432	T	A	MACF1	p.A4507A	Silent	SNV	0,201117318	PASS
#343T	HCA	chr1	179528824	179559689	G	T	NPHS2	p.P175H	Missense_Mutation	SNV	0,192307692	PASS
#343T	HCA	chr19	39307142	38816502	C	A	ECH1	p.A204S	Missense_Mutation	SNV	0,235294118	PASS
#343T	HCA	chr19	51162378	50659121	G	A	C19orf81	p.A192A	Silent	SNV	0,212765957	PASS
#623T	HCA	chr4	3251225	3249498	C	A	MSANTD1	p.G92G	Silent	SNV	0,111111111	t_lod_fstar
#623T	HCA	chr3	167728157	168010369	T	C	GOLIM4	p.N664S	Missense_Mutation	SNV	0,390728477	PASS
#623T	HCA	chr2	242289487	241350072	G	T	sept-02	NA	Splice_Site	SNV	0,15	PASS
#623T	HCA	chr19	15281269	15170458	T	A	NOTCH3	p.M1663L	Missense_Mutation	SNV	0,40625	PASS
#623T	HCA	chr19	38190159	37699258	A	G	ZNF607	p.R291R	Silent	SNV	0,362745098	PASS
#623T	HCA	chr15	22955162	22917904	G	T	CYFIP1	p.C519F	Missense_Mutation	SNV	0,392156863	PASS
#623T	HCA	chr16	54966731	54932819	G	C	IRX5	p.E191Q	Missense_Mutation	SNV	0,363636364	PASS
#623T	HCA	chr16	84056452	84022847	G	A	SLC38A8	p.R245W	Missense_Mutation	SNV	0,438202247	PASS
#623T	HCA	chr9	35045155	35045158	A	G	C9orf131	p.S843S	Silent	SNV	0,312015504	PASS
#623T	HCA	chr5	1325917	1325802	C	T	CLPTM1L	p.K365K	Silent	SNV	0,218309859	PASS
#623T	HCA	chr6	88075035	87365317	A	G	C6orf163	p.H304R	Missense_Mutation	SNV	0,347107438	PASS
#623T	HCA	chr11	71941272	72230228	G	T	INPPL1	p.Q349H	Missense_Mutation	SNV	0,571428571	t_lod_fstar
#623T	HCA	chr1	52859389	52393717	C	T	ORC1	p.E270K	Missense_Mutation	SNV	0,414285714	PASS
#623T	HCA	chr1	176175807	176206671	C	A	RFWD2	p.S103I	Missense_Mutation	SNV	0,1	t_lod_fstar
#623T	HCA	chrX	3229044	3311003	A	G	MXRA5	p.D2400D	Silent	SNV	0,382352941	PASS
#623T	HCA	chrX	19568147	19550029	G	T	SH3KBP1	p.T480N	Missense_Mutation	SNV	0,298611111	PASS
#623T	HCA	chrX	22208609	22190492	C	T	PHEX	p.T545T	Silent	SNV	0,188118812	PASS
#623T	HCA	chrX	114540816	115306251	G	A	LUZP4	p.G130D	Missense_Mutation	SNV	0,405529954	PASS
#623T	HCA	chr10	79686061	77926303	G	T	DLG5	p.A73E	Missense_Mutation	SNV	0,157894737	t_lod_fstar
#623T	HCA	chr10	96064395	94304638	C	G	PLCE1	p.S1872C	Missense_Mutation	SNV	0,301075269	PASS
#623T	HCA	chr12	21015397	20862463	A	AT	SLCO1B3	p.M179fs	Frame_Shift_Ins	INS	0,378205128	PASS
#623T	HCA	chr12	31242385	31089451	A	T	DDX11	p.I281F	Missense_Mutation	SNV	0,24742268	vents;homologous_maps
#623T	HCA	chr12	105381961	104988183	G	T	C12orf45	p.L44F	Missense_Mutation	SNV	0,115384615	t_lod_fstar
#623T	HCA	chr12	121671325	121233522	G	T	P2RX4	NA	Splice_Site	SNV	0,130434783	t_lod_fstar
#750T	HCA	chr1	47403719	46938047	G	A	CYP4A11	p.R96C	Missense_Mutation	SNV	0,137254902	PASS
#750T	HCA	chr1	177908891	177939756	C	A	SEC16B	p.G717W	Missense_Mutation	SNV	0,125	PASS
#750T	HCA	chr1	247588125	247424823	C	T	NLRP3	p.H460H	Silent	SNV	0,124183007	PASS
#750T	HCA	chr6	30691805	30724028	C	T	TUBB	p.S322S	Silent	SNV	0,208333333	PASS
#750T	HCA	chr6	31730281	31762504	G	A	MSH5	p.V826V	Silent	SNV	0,102739726	PASS
#750T	HCA	chr7	72988726	73574396	G	T	TBL2	p.A83D	Missense_Mutation	SNV	0,103448276	PASS
#750T	HCA	chr17	3729501	3826207	C	A	C17orf85	p.V164F	Missense_Mutation	SNV	0,15	t_lod_fstar
#750T	HCA	chr20	4880229	4899583	T	C	SLC23A2	p.T152A	Missense_Mutation	SNV	0,106382979	PASS
#835T	HCA	chr19	1250421	1250422	C	G	MIDN	p.R42R	Silent	SNV	0,285714286	PASS
#835T	HCA	chr19	9083579	8972903	C	T	MUC16	p.V2746M	Missense_Mutation	SNV	0,214285714	PASS
#835T	HCA	chr12	48142326	47748543	C	A	RAPGEF3	NA	Splice_Site	SNV	0,1	t_lod_fstar
#835T	HCA	chr12	76742131	76348351	C	G	BBS10	p.S3T	Missense_Mutation	SNV	0,140625	PASS
#835T	HCA	chr2	21231679	21008807	G	T	APOB	p.P2687P	Silent	SNV	0,294642857	PASS
#835T	HCA	chr4	37828375	37826753	C	A	PGM2	p.S7R	Missense_Mutation	SNV	0,1	t_lod_fstar
#835T	HCA	chrX	123190018	124056168	G	A	STAG2	p.V413I	Missense_Mutation	SNV	0,162790698	PASS

#835T	HCA	chrX	153283756	154018305	C	A	IRAK1	p.V260V	Silent	SNV	0,125	t_lod_fstar
#835T	HCA	chrX	153660653	154432307	C	A	ATP6AP1	p.A135A	Silent	SNV	0,157894737	t_lod_fstar
#835T	HCA	chr18	61322961	63655727	G	T	SERPINB3	p.P368H	Missense_Mutation	SNV	0,215189873	PASS
#835T	HCA	chr11	17153468	17131921	A	G	PIK3C2A	p.D742D	Silent	SNV	0,333333333	PASS
#835T	HCA	chr11	114271424	114400702	A	G	RBM7	p.T11A	Missense_Mutation	SNV	0,23255814	PASS
#835T	HCA	chr20	60882668	62307612	G	A	ADRM1	p.A214T	Missense_Mutation	SNV	0,272727273	PASS
#835T	HCA	chr5	161322780	161895774	C	T	GABRA1	p.A322V	Missense_Mutation	SNV	0,365853659	PASS
#835T	HCA	chr8	21980357	22122844	C	A	HR	p.E651*	Nonsense_Mutation	SNV	0,230769231	PASS
#835T	HCA	chr8	95897726	94885498	C	A	CCNE2	p.E221*	Nonsense_Mutation	SNV	0,1875	t_lod_fstar
#835T	HCA	chr16	31049441	31038120	C	A	STX4	p.A165D	Missense_Mutation	SNV	0,111111111	t_lod_fstar
#835T	HCA	chr17	36486093	38330210	C	T	GPR179	p.G1120E	Missense_Mutation	SNV	0,101123596	PASS
#835T	HCA	chr17	38319019	40162766	G	T	CASC3	p.R217L	Missense_Mutation	SNV	0,2	PASS
#835T	HCA	chr17	48274018	50196657	A	G	COL1A1	p.L2735	Missense_Mutation	SNV	0,122807018	PASS
#835T	HCA	chr1	45266767	44801095	G	T	PLK3	p.S126S	Silent	SNV	0,136363636	t_lod_fstar
#835T	HCA	chr3	172480714	172762924	C	T	ECT2	p.S340S	Silent	SNV	0,15625	PASS
#835T	HCA	chr10	101558984	99799227	AAAG	A	ABCC2	p.K298del	In_Frame_Del	DEL	0,325581395	str_contraction
#835T	HCA	chr22	50502543	50064114	C	A	MLC1	p.V327L	Missense_Mutation	SNV	0,157894737	t_lod_fstar
#873T	HCA	chr1	208213340	208039995	T	C	PLXNA2	p.L1450L	Silent	SNV	0,215686275	PASS
#873T	HCA	chr19	1879979	1879980	C	A	ABHD17A	p.R156S	Missense_Mutation	SNV	0,2	PASS
#873T	HCA	chr19	18280014	18169204	C	A	PIK3R2	p.A699A	Silent	SNV	0,103448276	t_lod_fstar
#873T	HCA	chr19	53456092	52952839	A	C	ZNF816	p.S345	Silent	SNV	0,196721311	PASS
#873T	HCA	chr11	46400772	46379222	T	C	DGKZ	p.D1041D	Silent	SNV	0,25	PASS
#873T	HCA	chr6	130389487	130068342	A	G	L3MBTL3	p.E338G	Missense_Mutation	SNV	0,157894737	PASS
#873T	HCA	chr12	132561100	132076555	G	A	EP400	p.V3057M	Missense_Mutation	SNV	0,1	PASS
#873T	HCA	chr2	128379614	127622039	C	A	MYO7B	p.R1169S	Missense_Mutation	SNV	0,166666667	t_lod_fstar
#873T	HCA	chr2	220313317	219448595	G	T	SPEG	p.E479D	Missense_Mutation	SNV	0,214285714	t_lod_fstar
#873T	HCA	chr17	7125555	7222236	C	T	ACADVL	p.A271V	Missense_Mutation	SNV	0,184210526	PASS
#873T	HCA	chr16	28917451	28906130	C	A	RABEP2	p.G438W	Missense_Mutation	SNV	0,2	t_lod_fstar
#873T	HCA	chr10	27322226	27033297	C	A	ANKRD26	p.T1245T	Silent	SNV	0,172413793	PASS
#873T	HCA	chr10	122664297	120904785	A	G	WDR11	p.N1056S	Missense_Mutation	SNV	0,178082192	PASS
#873T	HCA	chr22	36705380	36309335	G	C	MYH9	p.A597G	Missense_Mutation	SNV	0,18018018	PASS
#873T	HCA	chr4	74283861	73418144	CAG	C	ALB	p.R346fs	Frame_Shift_Del	DEL	0,226415094	str_contraction
#2729T	HCA	chr1	150444429	150471953	C	A	RPRD2	p.S1002Y	Missense_Mutation	SNV	0,185185185	PASS
#2729T	HCA	chr3	23932048	23890557	C	T	UBE2E1	p.A178V	Missense_Mutation	SNV	0,129770992	PASS
#2729T	HCA	chr3	132350210	132631366	A	C	ACAD11	p.N272K	Missense_Mutation	SNV	0,31092437	PASS
#2729T	HCA	chr4	30724027	30722405	T	C	PCDH7	p.L328P	Missense_Mutation	SNV	0,311827957	PASS
#2729T	HCA	chr6	96054277	95606401	C	A	MANEA	p.S462Y	Missense_Mutation	SNV	0,147058824	PASS
#2729T	HCA	chr10	255938	209998	T	C	ZMYND11	p.S76P	Missense_Mutation	SNV	0,100775194	PASS
#2729T	HCA	chr11	65547148	65779677	G	C	AP5B1	p.I272M	Missense_Mutation	SNV	0,334975369	PASS
#2729T	HCA	chr11	93430922	93697756	T	A	KIAA1731	p.N948K	Missense_Mutation	SNV	0,4	PASS
#2729T	HCA	chr15	75014848	74722507	G	A	CYP1A1	p.V197V	Silent	SNV	0,22459893	PASS
#2729T	HCA	chr16	89351308	89284900	T	C	ANKRD11	p.T548A	Missense_Mutation	SNV	0,211009174	PASS
#2729T	HCA	chr19	36212223	35721321	C	T	KMT2B	p.H658H	Silent	SNV	0,307317073	PASS
#2729T	HCA	chr20	2411652	2431006	T	C	TGM6	p.L649P	Missense_Mutation	SNV	0,372340426	PASS
#2729T	HCA	chrX	100745781	101490794	G	C	ARMCX4	p.V735V	Silent	SNV	0,151408451	PASS
#375T	HCA	chr1	176563750	176594614	GCTT	G	PAPPA2	p.F340del	In_Frame_Del	DEL	0,253846154	str_contraction
#375T	HCA	chr1	226019594	225831893	G	C	EPHX1	p.E100Q	Missense_Mutation	SNV	0,293650794	PASS
#375T	HCA	chr1	236758942	236595642	G	A	HEATR1	p.L330F	Missense_Mutation	SNV	0,266666667	PASS
#375T	HCA	chr2	1687850	1684078	A	T	PXDN	NA	Splice_Site	SNV	0,359375	PASS
#375T	HCA	chr2	108487510	107871054	C	A	RGPD4	p.T1017N	Missense_Mutation	SNV	0,290909091	PASS
#375T	HCA	chr2	210818915	209954191	G	C	UNC80	p.A2394P	Missense_Mutation	SNV	0,215053763	PASS
#375T	HCA	chr3	130463496	130744652	T	C	PIK3R4	p.T189T	Silent	SNV	0,278195489	PASS
#375T	HCA	chr5	40747098	40746996	C	G	TTC33	p.R8T	Missense_Mutation	SNV	0,275862069	PASS
#375T	HCA	chr6	43006153	43038415	C	A	CUL7	p.R1542L	Missense_Mutation	SNV	0,285714286	PASS
#375T	HCA	chr6	146126076	145804940	T	C	FBXO30	p.Q489R	Missense_Mutation	SNV	0,226086957	PASS
#375T	HCA	chr7	150707754	151010666	C	T	NOS3	p.P919S	Missense_Mutation	SNV	0,328358209	PASS
#375T	HCA	chr8	19680842	19823331	A	G	INTS10	p.N185S	Missense_Mutation	SNV	0,20212766	PASS
#375T	HCA	chr9	15578911	15578913	G	T	CCDC171	p.R81L	Missense_Mutation	SNV	0,277777778	PASS
#375T	HCA	chr9	125612078	122849799	G	C	RC3H2	p.P1135R	Missense_Mutation	SNV	0,303571429	PASS
#375T	HCA	chr11	85722154	86011111	G	C	PICALM	p.N228K	Missense_Mutation	SNV	0,313253012	PASS
#375T	HCA	chr11	122817384	122946676	C	T	C11orf63	p.Q605*	Nonsense_Mutation	SNV	0,280991736	PASS
#375T	HCA	chr12	19282742	19129808	G	A	PLEKHA5	p.A3A	Silent	SNV	0,277777778	PASS
#375T	HCA	chr12	111800732	111362928	G	C	FAM109A	p.A167G	Missense_Mutation	SNV	0,28042328	PASS
#375T	HCA	chr13	22077064	21502925	C	T	MICU2	NA	Splice_Site	SNV	0,288135593	PASS
#375T	HCA	chr13	77742575	77168440	C	T	MYCBP2	p.V1996V	Silent	SNV	0,292307692	PASS
#375T	HCA	chr14	33015139	32545933	C	A	AKAP6	p.S427*	Nonsense_Mutation	SNV	0,295081967	PASS
#375T	HCA	chr15	24921470	24676323	G	T	NPAP1	p.W152C	Missense_Mutation	SNV	0,172727273	PASS
#375T	HCA	chr15	56032607	55740409	T	C	PRTG	p.S124G	Missense_Mutation	SNV	0,263157895	events;homologous_map
#375T	HCA	chr17	79663962	81696932	G	A	HGS	p.G606S	Missense_Mutation	SNV	0,178217822	PASS
#375T	HCA	chr19	18186562	18075752	G	A	IL12RB1	p.P233S	Missense_Mutation	SNV	0,25	PASS
#375T	HCA	chr20	44664524	46035885	A	T	SLC12A5	p.M153L	Missense_Mutation	SNV	0,221311475	PASS
#375T	HCA	chr20	61867691	63236339	C	A	BIRC7	p.F81L	Missense_Mutation	SNV	0,248	PASS
#375T	HCA	chr22	24836797	24440829	G	A	ADORA2A	p.M193I	Missense_Mutation	SNV	0,255033557	PASS
#375T	HCA	chr22	38352784	37956777	G	A	POLR2F	p.D9N	Missense_Mutation	SNV	0,272727273	PASS
#375T	HCA	chrX	70352284	71132434	C	T	MED12	p.T1437T	Silent	SNV	0,24691358	PASS
#375T	HCA	chrX	130408773	131274799	C	T	IGSF1	p.R1184Q	Missense_Mutation	SNV	0,307189542	PASS
#4110T	HCA	chr2	179593499	178728772	G	A	TTN	p.A6068V	Missense_Mutation	SNV	0,271929825	PASS
#4110T	HCA	chr2	215843564	214978840	G	A	ABCA12	p.I1647I	Silent	SNV	0,381818182	PASS
#4110T	HCA	chr3	183905767	184187979	G	T	ABCF3	p.D189Y	Missense_Mutation	SNV	0,342592593	PASS
#4110T	HCA	chr6	112575282	112254080	G	A	LAMA4	p.A24V	Missense_Mutation	SNV	0,313253012	PASS
#4110T	HCA	chr8	42716914	42861771	G	A	RNF170	p.R161W	Missense_Mutation	SNV	0,315789474	PASS
#4110T	HCA	chr10	37508256	37219328	G	C	ANKRD30A	p.V1150L	Missense_Mutation	SNV	0,324675325	PASS
#4110T	HCA	chr10	112572637	110812879	A	G	RBM20	p.T828A	Missense_Mutation	SNV	0,283464567	PASS
#4110T	HCA	chr11	551660	551660	G	A	LRRC56	p.R269H	Missense_Mutation	SNV	0,244047619	PASS
#4110T	HCA	chr11	61726990	61959518	C	G	BEST1	p.N296K	Missense_Mutation	SNV	0,223076923	PASS
#4110T	HCA	chr12	55368307	54974523	G	T	TESPA1	p.R14R	Silent	SNV	0,297297297	PASS
#4110T	HCA	chr12	98993836	98600058	G	A	SLC25A3	p.V250I	Missense_Mutation	SNV	0,120300752	PASS
#4110T	HCA	chr12	102154956	101761178	T	G	GNPTAB	p.R1028S	Missense_Mutation	SNV	0,179104478	PASS
#4110T	HCA	chr16	75467247	75433349	C	G	CFDP1	p.E2Q	Missense_Mutation	SNV	0,307017544	PASS
#4110T	HCA	chr17	55754365	57677004	G	T	MSI2	p.T321T	Silent	SNV	0,353448276	PASS
#4110T	HCA	chr17	56232786	58155425	CGATCTCCTACCA	C	OR4D1	p.I5YQ92del	In_Frame_Del	DEL	0,153846154	PASS
#4110T	HCA	chr17	67513700	69517559	CG	C	MAP2K6	p.G65fs	Frame_Shift_Del	DEL	0,324175824	PASS
#4110T	HCA	chr19	2120866	2120867	G	A	AP3D1	p.F492F	Silent	SNV	0,308411215	PASS
#4110T	HCA	chr19	11687597	11576782	G	C	ACP5	p.A108G	Missense_Mutation	SNV	0,302521008	PASS
#4110T	HCA	chr19	54377352	53874098	TC	T	MYADM	p.F190fs	Frame_Shift_Del	DEL	0,23880597	PASS
#4110T	HCA	chr20	60640497	62065441	C	A	TAF4	p.A124S	Missense_Mutation	SNV	0,115384615	t_lod_fstar
#4110T	HCA	chr22	37462918	37066878	G	A	TMPRSS6	p.T742M	Missense_Mutation	SNV	0,236486486	PASS
#4110T	HCA	chr22	39239779	38843774	C	G	NPTXR	p.A29P	Missense_Mutation	SNV	0,103896104	PASS
#4110T	HCA	chrY	43089646	42693640	A	T	AMGALT	p.T104T	Silent	SNV	0,25	PASS
#4110T	HCA	chrX	153069640	153804185	C	A	PDZD4	p.R493L	Missense_Mutation	SNV	0,178832117	PASS
#533T	HCA	chr2	152409229	151552715	G	A	NEB	p.P4897L	Missense_Mutation	SNV	0,116504854	PASS
#533T	HCA	chr2	220313027	219448305	C	A	SPEG	p.R383S	Missense_Mutation	SNV	0,176470588	PASS
#533T	HCA	chr3	155838426	156120637	C	A	KCNAB1	p.A9E	Missense_Mutation	SNV	0,12	PASS
#533T	HCA	chr4	107248649	106327492	GA	G	AIMP1	p.E51fs	Frame_Shift_Del	DEL	0,184782609	PASS
#533T	HCA	chr5	131046300	131710607	C	T	FNIP1	p.R226H	Missense_Mutation	SNV	0,133333333	PASS
#533T	HCA	chr6	36564678	36596901	G	A	SRSF3	p.G47S	Missense_Mutation	SNV	0,148387097	PASS
#533T	HCA	chr6	152806078	152484943	T	C	SYNE1	p.E359E	Silent	SNV	0,184466019	PASS
#533T	HCA	chr7	150764060	151066973	G	T	SLC4A2	p.A316S	Missense_Mutation	SNV	0,152380952	PASS
#533T	HCA	chr8	86180155	85267926	C	T	CA13	p.Q159*	Nonsense_Mutation	SNV	0,192307692	PASS
#533T	HCA	chr9	104317117	101554835	G	T	RNF20	p.A721S	Missense_Mutation	SNV	0,178571429	PASS
#533T	HCA	chr11	55135610	55368134	G	T	OR4A15	p.G84V	Missense_Mutation	SNV	0,148809524	PASS
#533T	HCA	chr15	31775516	31483313	G	T	OTUD7A	p.A921E	Missense_Mutation	SNV	0,158333333	PASS
#533T	HCA	chr15	99672706	99132501	C	A	SYNM	p.Q1380K	Missense_Mutation	SNV	0,14	clustered_events
#533T	HCA	chr15	99672716	99132511	G	C	SYNM	p.S1383T	Missense_Mutation	SNV	0,139303483	clustered_events

#533T	HCA	chr18	6942159	6942160	C	A	LAMA1	p.P3049P	Silent	SNV	0,2	PASS
#533T	HCA	chr19	8209815	8144931	C	A	FBN3	p.G163W	Missense_Mutation	SNV	0,134146341	PASS
#533T	HCA	chr20	60791598	62216542	C	G	HRH3	p.E268Q	Missense_Mutation	SNV	0,219251337	PASS
#533T	HCA	chr21	38853060	37480758	A	T	DYRK1A	p.K150*	Nonsense_Mutation	SNV	0,129496403	PASS
#533T	HCA	chrX	43628600	43769353	C	G	MAOB	p.G434A	Missense_Mutation	SNV	0,233766234	PASS
#477T	HCA	chr3	50327178	50289747	C	T	IFRD2	p.G252S	Missense_Mutation	SNV	0,354037267	PASS
#477T	HCA	chr5	90670828	91375011	C	CT	ARRDC3	p.E261fs	Frame_Shift_Ins	INS	0,29535865	PASS
#477T	HCA	chr5	179192901	179765900	A	G	MAML1	p.N297S	Missense_Mutation	SNV	0,189189189	PASS
#477T	HCA	chr7	27169940	27130321	G	A	HOXA4	p.P138L	Missense_Mutation	SNV	0,315	PASS
#477T	HCA	chr9	101818612	99056330	G	C	COL15A1	p.G1088A	Missense_Mutation	SNV	0,311111111	PASS
#477T	HCA	chr11	66360237	66592766	T	G	CCDC87	p.I84L	Missense_Mutation	SNV	0,338541667	PASS
#477T	HCA	chr12	82147918	81754139	T	C	PPFIA2	p.D28G	Missense_Mutation	SNV	0,312977099	PASS
#477T	HCA	chr14	21955825	21487666	G	T	TOX4	p.G97G	Silent	SNV	0,395061728	PASS
#477T	HCA	chr14	100759653	100293316	T	G	SLC25A29	p.K47T	Missense_Mutation	SNV	0,340425532	PASS
#477T	HCA	chr15	68582798	68290460	C	T	FEM1B	p.L368F	Missense_Mutation	SNV	0,282926829	PASS
#477T	HCA	chr17	27250972	28923954	G	C	PHF12	p.Q224E	Missense_Mutation	SNV	0,304	PASS
#477T	HCA	chr17	28887667	30560649	T	C	TBCID29	p.D37D	Silent	SNV	0,12605042	PASS
#477T	HCA	chr18	3452206	3452208	T	A	TGIF1	p.S77T	Missense_Mutation	SNV	0,207729469	PASS
#477T	HCA	chr18	51807184	54280814	G	C	POLI	p.G236A	Missense_Mutation	SNV	0,128378378	PASS
#477T	HCA	chrX	3544551	3626510	G	T	PRKX	p.P242T	Missense_Mutation	SNV	0,240310078	PASS
#477T	HCA	chrX	38158213	38298960	T	C	RPGR	p.E414G	Missense_Mutation	SNV	0,159292035	PASS
#477T	HCA	chrX	48682656	48824246	C	T	HDAC6	p.I1177I	Silent	SNV	0,142857143	PASS
#477T	HCA	chrX	123514498	124380648	C	T	TENM1	p.R2689Q	Missense_Mutation	SNV	0,196296296	PASS
#757T	HCA	chr2	17696551	17515284	G	A	RAD51AP2	p.H1044H	Silent	SNV	0,293577982	PASS
#757T	HCA	chr2	241818194	240878777	G	T	AGXT	p.A379S	Missense_Mutation	SNV	0,252941176	PASS
#757T	HCA	chr3	170102544	170384756	A	G	SKIL	p.S474G	Missense_Mutation	SNV	0,276923077	PASS
#757T	HCA	chr4	105412445	104491288	T	C	OXGC4	p.H3R	Missense_Mutation	SNV	0,243589744	PASS
#757T	HCA	chr4	144106702	143185549	A	G	USP38	p.L33L	Silent	SNV	0,25	PASS
#757T	HCA	chr5	82948586	83652767	C	T	HAPLN1	p.R53K	Missense_Mutation	SNV	0,220858896	PASS
#757T	HCA	chr6	83884141	83174422	TTTCTTATC	T	PGM3	p.DKK315fs	Frame_Shift_Del	DEL	0,169230769	PASS
#757T	HCA	chr8	128428420	127416175	G	A	POU5F1B	p.V103V	Silent	SNV	0,285223368	PASS
#757T	HCA	chr9	2029049	2029049	G	T	SMARCA2	p.A9A	Silent	SNV	0,247863248	PASS
#757T	HCA	chr9	100132979	97370697	A	C	CCDC180	p.E1458D	Missense_Mutation	SNV	0,253846154	PASS
#757T	HCA	chr9	125437820	122675541	C	T	OR1L3	p.R138C	Missense_Mutation	SNV	0,220183486	PASS
#757T	HCA	chr9	138413373	135521527	T	C	LCN1	p.L10L	Silent	SNV	0,113402062	vents;homologous_map
#757T	HCA	chr10	26446370	26157441	A	G	MYO3A	p.T975T	Silent	SNV	0,272727273	PASS
#757T	HCA	chr11	56230241	56462765	G	C	OR5M9	p.L213V	Missense_Mutation	SNV	0,130434783	vents;homologous_map
#757T	HCA	chr12	32136794	31983860	A	G	KIAA1551	p.I969V	Missense_Mutation	SNV	0,267605634	PASS
#757T	HCA	chr12	43925997	43532194	G	A	ADAMTS20	p.T152M	Splice_Site	SNV	0,268907563	PASS
#757T	HCA	chr12	101005823	100612045	G	T	GAS2L3	p.G117C	Missense_Mutation	SNV	0,268292683	PASS
#757T	HCA	chr13	37014210	36440073	T	A	CCNA1	p.L330M	Missense_Mutation	SNV	0,278145695	PASS
#757T	HCA	chr16	24564879	24553558	GTA	G	RBBP6	NA	Splice_Site	DEL	0,12345679	str_contraction
#757T	HCA	chr19	813708	813708	G	T	LPPR3	p.A340D	Missense_Mutation	SNV	0,258333333	PASS
#757T	HCA	chr19	40711106	40205199	C	G	MAP3K10	p.A364G	Missense_Mutation	SNV	0,25	PASS
#757T	HCA	chr19	53270093	52766840	C	A	ZNF600	p.E306*	Nonsense_Mutation	SNV	0,122994652	PASS
#757T	HCA	chr22	21153976	20799688	G	C	PI4KA	p.I543M	Missense_Mutation	SNV	0,293650794	PASS
#757T	HCA	chr22	26299739	25903772	G	A	MYO18B	p.A1698T	Missense_Mutation	SNV	0,368852459	PASS
#757T	HCA	chrX	16170753	16152630	C	A	GRPR	p.H380Q	Missense_Mutation	SNV	0,328571429	PASS
#757T	HCA	chrX	49142604	49286142	C	T	PPP1R3F	p.I484I	Silent	SNV	0,243781095	PASS
#1896T	HCA	chr1	22921760	22595267	C	A	EPHA8	p.C547*	Nonsense_Mutation	SNV	0,236686391	PASS
#1896T	HCA	chr1	158812095	158842305	T	C	MNDA	p.L515	Missense_Mutation	SNV	0,174193548	vents;homologous_map
#1896T	HCA	chr1	201798426	201829298	C	T	IPO9	p.T30I	Missense_Mutation	SNV	0,146666667	PASS
#1896T	HCA	chr1	204945884	204976756	C	A	NFASC	p.L598I	Missense_Mutation	SNV	0,190082645	PASS
#1896T	HCA	chr1	235969637	235806337	G	A	LYST	p.A933A	Silent	SNV	0,264516129	vents;homologous_map
#1896T	HCA	chr1	237656304	237493004	T	C	RYR2	p.R626R	Silent	SNV	0,192307692	PASS
#1896T	HCA	chr1	243809241	243645939	A	G	AKT3	p.I128T	Missense_Mutation	SNV	0,278787879	PASS
#1896T	HCA	chr2	11364587	11224461	C	T	ROCK2	NA	Splice_Site	SNV	0,23255814	PASS
#1896T	HCA	chr2	24480865	24257996	A	C	ITSN2	p.V927G	Missense_Mutation	SNV	0,227272727	PASS
#1896T	HCA	chr2	42719980	42492840	G	A	KCNG3	p.S221F	Missense_Mutation	SNV	0,179310345	PASS
#1896T	HCA	chr2	55544847	55317711	A	T	CCDC88A	p.L1152H	Missense_Mutation	SNV	0,176829268	PASS
#1896T	HCA	chr2	74043167	73816040	A	C	C2orf78	p.K606T	Missense_Mutation	SNV	0,221052632	PASS
#1896T	HCA	chr2	113482966	112725389	G	A	NT5DC4	p.S298S	Silent	SNV	0,18556701	PASS
#1896T	HCA	chr2	116066861	115309285	C	T	DPP10	p.A36V	Missense_Mutation	SNV	0,223880597	vents;homologous_map
#1896T	HCA	chr2	238296631	237387988	C	T	COL6A3	p.K302K	Silent	SNV	0,183098592	PASS
#1896T	HCA	chr3	32030572	31989080	A	G	ZNF860	p.M1V	Start_Codon_SNV	SNV	0,225165563	PASS
#1896T	HCA	chr3	33895418	33853926	T	C	PDCD6IP	p.A646A	Silent	SNV	0,217391304	PASS
#1896T	HCA	chr3	156877726	157159937	G	GT	CCNL1	p.T53fs	Frame_Shift_Ins	INS	0,177083333	PASS
#1896T	HCA	chr4	74274520	73408803	AAAGT	A	ALB	p.K161fs	Splice_Site	DEL	0,212121212	str_contraction
#1896T	HCA	chr5	38523524	38523422	T	C	LIFR	p.K186K	Silent	SNV	0,304878049	PASS
#1896T	HCA	chr5	76373371	77077546	A	T	ZBED3	p.P111P	Silent	SNV	0,315789474	PASS
#1896T	HCA	chr5	79058976	79763153	G	C	CMYA5	p.T3833T	Silent	SNV	0,100719424	t_lod_fstar
#1896T	HCA	chr5	140228923	140849338	C	A	PCDHA9	p.F281L	Missense_Mutation	SNV	0,153846154	t_lod_fstar
#1896T	HCA	chr5	169310146	169883142	C	A	FAM196B	p.E253*	Nonsense_Mutation	SNV	0,221374046	PASS
#1896T	HCA	chr6	32084264	32116487	G	T	ATF6B	p.A625A	Silent	SNV	0,18902439	PASS
#1896T	HCA	chr7	47925536	47885938	T	C	PKD1L1	p.T985A	Missense_Mutation	SNV	0,191489362	PASS
#1896T	HCA	chr7	97867838	98238526	C	T	TECPR1	p.V340M	Missense_Mutation	SNV	0,180851064	PASS
#1896T	HCA	chr9	96051511	93289229	C	T	WNK2	p.A1529V	Missense_Mutation	SNV	0,191256831	PASS
#1896T	HCA	chr10	5788653	5746690	C	A	FAM208B	p.A1090E	Missense_Mutation	SNV	0,272727273	PASS
#1896T	HCA	chr10	30918549	30629620	G	C	LYZL2	p.A29G	Missense_Mutation	SNV	0,132867133	PASS
#1896T	HCA	chr12	88487642	88093865	G	T	CEP290	p.R1072R	Silent	SNV	0,162162162	PASS
#1896T	HCA	chr15	43024013	42731815	C	A	CDAN1	p.S515I	Missense_Mutation	SNV	0,261904762	PASS
#1896T	HCA	chr15	52708476	52416279	T	C	MYO5A	p.I160V	Missense_Mutation	SNV	0,126760563	PASS
#1896T	HCA	chr16	20043580	20032258	T	G	GPR139	p.H180P	Missense_Mutation	SNV	0,196428571	PASS
#1896T	HCA	chr16	67327446	67293543	C	T	KCTD19	NA	Splice_Site	SNV	0,226804124	PASS
#1896T	HCA	chr16	84224876	84191270	C	T	ADAD2	p.R14C	Missense_Mutation	SNV	0,18699187	PASS
#1896T	HCA	chr17	75494701	77498619	C	G	sept-09	p.N574K	Missense_Mutation	SNV	0,193277311	PASS
#1896T	HCA	chr19	8807878	8697528	C	T	ACTL9	p.A392T	Missense_Mutation	SNV	0,159090909	PASS
#1896T	HCA	chr19	14066769	13955957	C	G	DCAF15	p.L138V	Missense_Mutation	SNV	0,1	PASS
#1896T	HCA	chr19	37239765	36748863	T	A	ZNF850	p.E726V	Missense_Mutation	SNV	0,244635193	PASS
#1896T	HCA	chr19	37642600	37151698	T	C	ZNF585A	p.N734S	Missense_Mutation	SNV	0,265734266	PASS
#1896T	HCA	chr19	39915329	39424689	C	G	PLEKHG2	p.L1186V	Missense_Mutation	SNV	0,163636364	PASS
#1896T	HCA	chr19	52132723	51629470	C	T	SIGLEC5	p.S196S	Silent	SNV	0,228070175	PASS
#1896T	HCA	chr19	56126241	55614875	C	G	ZNF865	p.A419A	Silent	SNV	0,169117647	PASS
#1896T	HCA	chr19	57723330	57211962	T	C	ZNF264	p.C289R	Missense_Mutation	SNV	0,210970464	PASS
#1896T	HCA	chrX	35821086	35802969	A	G	MAGEB16	p.K258R	Missense_Mutation	SNV	0,276041667	PASS

Abbreviations:
REF=reference
ALT=altered
VAF=Variant Allele Fraction
HCA=hepatocellular adenoma
SNV=Single Nucleotide Variant
DEL=deletion
INS=insertion

Supplementary Table 5: Copy Number Alterations from Whole Exome Sequencing in 19 samples

sample	chr	start_hg19	end_hg19	nMarkers	LRR	GNL	drivHCC_Schulze2015_genes	Beroukhim2010andZack201	cytoStart	cytoEnd
#1343T	chr1	861348	33116108	3655	0,3214	0	ARID1A,RPL22		1p36.33	1p35.1
#1343T	chr1	33117588	33138436	5	-0,1569	-1			1p35.1	1p35.1
#1343T	chr1	33146349	249212472	13778	0,2534	0	COL11A1		1p35.1	1q44
#1343T	chr2	40221	242842554	13479	0,2463	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#1343T	chr3	239470	14528614	837	0,2914	0			3p26.3	3p25.1
#1343T	chr3	14535253	14552745	8	0,5896	1			3p25.1	3p25.1
#1343T	chr3	14554324	197769162	10087	0,2549	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p25.1	3q29
#1343T	chr4	53291	190947044	7106	0,2312	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#1343T	chr5	140602	180687559	7964	0,246	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#1343T	chr6	203615	32292448	2171	0,2576	0			6p25.3	6p21.32
#1343T	chr6	32299746	32299958	5	-0,27	-1			6p21.32	6p21.32
#1343T	chr6	32301759	170892714	6955	0,2363	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p21.32	6q27
#1343T	chr7	195642	158937319	8245	0,2659	0	MET		7p22.3	7q36.3
#1343T	chr8	190866	146280815	6120	0,2566	0	SLURP1,MYC		8p23.3	8q24.3
#1343T	chr9	214910	141017498	6913	0,279	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#1343T	chr10	226000	105727306	5466	0,2534	0	PTEN		10p15.3	10q24.33
#1343T	chr10	105750514	105759665	5	-0,2096	-1			10q24.33	10q24.33
#1343T	chr10	105761266	135373624	1678	0,2584	0			10q24.33	10q26.3
#1343T	chr11	193118	134856548	9816	0,2861	0	HMBS		11p15.5	11q25
#1343T	chr12	208345	57937660	5329	0,2622	0	ARID2,CDKN1B		12p13.33	12q13.3
#1343T	chr12	57937922	57957439	5	-0,2369	-1			12q13.3	12q13.3
#1343T	chr12	57957942	133811687	5109	0,2492	0	HNF1A		12q13.3	12q24.33
#1343T	chr13	19600450	115091029	3185	0,2267	0	RB1		13q12.11	13q34
#1343T	chr14	20201689	107282930	5910	0,2578	0			14q11.2	14q32.33
#1343T	chr15	20170066	65677372	3585	0,2517	0			15q11.1	15q22.31
#1343T	chr15	65678941	65703532	9	0,6384	1			15q22.31	15q22.31
#1343T	chr15	65735572	102389376	2590	0,2723	0			15q22.31	15q26.3
#1343T	chr16	96492	90142284	6877	0,3149	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#1343T	chr17	6088	81052160	9708	0,3079	0	TP53		17p13.3	17q25.3
#1343T	chr18	163380	77941018	2792	0,2345	0			18p11.32	18q23
#1343T	chr19	282222	45202658	6042	0,359	0	KEAP1		19p13.3	19q13.32
#1343T	chr19	45204710	45206790	3	1,1022	1			19q13.32	19q13.32
#1343T	chr19	45207426	59083958	2468	0,3458	0	SEPW1,TMEM160		19q13.32	19q13.43
#1343T	chr20	68379	62906077	4377	0,2999	0	DNAJC5		20p13	20q13.33
#1343T	chr21	9825922	48084620	1825	0,2816	0	DYRK1A		21p11.2	21q22.3
#1343T	chr22	17060842	51219140	3567	0,33	0	ZNRF3,CLDN5		22q11.1	22q13.33
#1343T	chrX	2700138	154774846	5886	0,2544	0	RPS6KA3		Xp22.33	Xq28
#1424T	chr1	762224	249213866	16643	0,1836	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#1424T	chr2	40221	242842554	12939	0,1685	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#1424T	chr3	239468	197769154	10510	0,1737	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#1424T	chr4	53291	190862085	6995	0,1542	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#1424T	chr5	143266	180687559	7686	0,1639	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#1424T	chr6	203615	170891556	8750	0,1723	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#1424T	chr7	195642	891069	43	0,2455	0			7p22.3	7p22.3
#1424T	chr7	891633	896016	5	-0,2418	-1		7p22.3	7p22.3	7p22.3
#1424T	chr7	899892	158930780	7713	0,1806	0	MET		7p22.3	7q36.3
#1424T	chr8	183094	665934	24	0,1695	0			8p23.3	8p23.3
#1424T	chr8	676520	960314	6	0,5142	1			8p23.3	8p23.3
#1424T	chr8	1497320	146280027	5886	0,1682	0	SLURP1,MYC		8p23.3	8q24.3
#1424T	chr9	213464	141017498	6463	0,1916	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#1424T	chr10	226000	135378996	6784	0,1719	0	PTEN		10p15.3	10q26.3
#1424T	chr11	193118	134257636	9184	0,2006	0	HMBS		11p15.5	11q25
#1424T	chr12	208345	133810818	9447	0,1767	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#1424T	chr13	19752446	115091029	3118	0,1598	0	RB1		13q12.11	13q34
#1424T	chr14	20201689	107283203	5294	0,1835	0			14q11.2	14q32.33
#1424T	chr15	20450269	102389376	5542	0,1812	0			15q11.1	15q26.3
#1424T	chr16	97088	90142284	6368	0,215	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#1424T	chr17	6088	81052160	9153	0,2144	0	TP53		17p13.3	17q25.3
#1424T	chr18	163380	77960704	2674	0,155	0			18p11.32	18q23
#1424T	chr19	287612	9768755	1623	0,27	0			19p13.3	19p13.2
#1424T	chr19	9770098	9939531	12	-0,0062	0			19p13.2	19p13.2
#1424T	chr19	9959820	23445054	2105	0,2483	0	KEAP1		19p13.2	19p12
#1424T	chr19	23505878	23506923	5	0,7402	1			19p12	19p12
#1424T	chr19	23543160	23578141	5	0,1661	0			19p12	19p12
#1424T	chr19	23648920	23651150	8	0,7238	1			19p12	19p12
#1424T	chr19	23674706	59083958	4242	0,2364	0	SEPW1,TMEM160		19p12	19q13.43
#1424T	chr20	68379	62906077	4131	0,2063	0	DNAJC5		20p13	20q13.33
#1424T	chr21	9825921	48084620	1719	0,1873	0	DYRK1A		21p11.2	21q22.3
#1424T	chr22	17061701	51179116	3329	0,2244	0	ZNRF3,CLDN5		22q11.1	22q13.33
#1424T	chrX	2700138	154774846	5668	0,1723	0	RPS6KA3		Xp22.33	Xq28
#1439T	chr1	762224	115323180	8942	0,06	0	ARID1A,RPL22,COL11A1		1p36.33	1p13.2
#1439T	chr1	115399236	115453070	13	-0,3479	-1			1p13.2	1p13.2
#1439T	chr1	115454151	249213866	8201	0,0437	0			1p13.2	1q44
#1439T	chr2	40221	216981473	11188	0,0211	0	ACVR2A,NFE2L2,APOB		2p25.3	2q35
#1439T	chr2	216982491	216997078	6	-0,3037	-1			2q35	2q35
#1439T	chr2	217002856	227963462	820	0,0991	0	EPHA4		2q35	2q36.3
#1439T	chr2	227964370	227973970	10	-0,329	-1			2q36.3	2q36.3
#1439T	chr2	227974184	242842554	1323	0,0785	0			2q36.3	2q37.3
#1439T	chr3	239468	15614707	965	0,0676	0			3p26.3	3p25.1
#1439T	chr3	15616536	15631082	5	-0,4631	-1			3p25.1	3p25.1
#1439T	chr3	15633146	25666232	241	0,0278	0			3p25.1	3p24.2
#1439T	chr3	25668107	25671880	7	-0,3414	-1			3p24.2	3p24.2
#1439T	chr3	25672360	197769154	9605	0,0382	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p24.2	3q29
#1439T	chr4	53291	100257920	4235	0,0202	0	ALB,GABRA2		4p16.3	4q23
#1439T	chr4	100260802	100268952	6	-0,5787	-1			4q23	4q23
#1439T	chr4	100273822	190862085	3006	-0,0029	0	FGA,IRF2		4q23	4q35.2
#1439T	chr5	143266	180687559	7905	0,0248	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#1439T	chr6	203615	170891556	9040	0,0328	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#1439T	chr7	195642	158930780	8008	0,0482	0	MET		7p22.3	7q36.3
#1439T	chr8	183094	120768300	4723	0,015	0			8p23.3	8q24.12
#1439T	chr8	120770347	120793362	5	-0,4619	-1			8q24.12	8q24.12
#1439T	chr8	120795746	146280027	1367	0,0907	0	SLURP1,MYC		8q24.12	8q24.3
#1439T	chr9	213464	141017498	6652	0,0626	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#1439T	chr10	226000	135381590	6994	0,0329	0	PTEN		10p15.3	10q26.3
#1439T	chr11	193811	17316942	1718	0,0882	0			11p15.5	11p15.1
#1439T	chr11	17317704	17332784	5	-0,4254	-1			11p15.1	11p15.1
#1439T	chr11	17333448	134257636	7727	0,0681	0	HMBS		11p15.1	11q25
#1439T	chr12	208345	133810818	9744	0,0485	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#1439T	chr13	19752446	115091029	3259	0,011	0	RB1		13q12.11	13q34
#1439T	chr14	20201689	107283203	5490	0,0414	0			14q11.2	14q32.33
#1439T	chr15	20462622	102389376	5712	0,0447	0			15q11.1	15q26.3
#1439T	chr16	97088	90142284	6495	0,0969	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3

#1439T	chr17	6088	81052160	9366	0,0888	0	TP53		17p13.3	17q25.3
#1439T	chr18	163380	77960704	2777	0,019	0			18p11.32	18q23
#1439T	chr19	287612	55898654	7702	0,1184	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.42
#1439T	chr19	55899440	55949952	7	0,4976	1		19q13.42	19q13.42	19q13.42
#1439T	chr19	55964714	59083958	408	0,0885	0			19q13.42	19q13.43
#1439T	chr20	68379	62906077	4185	0,083	0	DNAJC5		20p13	20q13.33
#1439T	chr21	9825921	48084620	1760	0,0602	0	DYRK1A		21p11.2	21q22.3
#1439T	chr22	17056656	51179116	3376	0,1121	0	ZNRF3,CLDN5		22q11.1	22q13.33
#1439T	chrX	2700138	154774846	5805	0,038	0	RPS6KA3		Xp22.33	Xq28
#1896T	chr1	69550	179972365	15352	0,9489	0	ARID1A,RPL22,COL11A1,JAK1,NRAS		1p36.33	1q25.2
#1896T	chr1	179975645	180010328	11	1,2772	1			1q25.2	1q25.2
#1896T	chr1	180010867	249211681	4649	0,9365	0			1q25.2	1q44
#1896T	chr2	41618	242814646	14755	0,945	0	ACVR2A,NFE2L2,EPHA4,APOB,ALK		2p25.3	2q37.3
#1896T	chr3	361505	197765488	11526	0,9395	0	CTNNB1,BAP1,KLF15,SETD2,FANCD2,TNFSF10,PIK3CA		3p26.3	3q29
#1896T	chr4	53384	113242	7440	0,9403	0			4p16.3	4p16.3
#1896T	chr5	92369	58147085	2100	0,931	0	IL6ST,TERT		5p15.33	5q11.2
#1896T	chr5	58270699	58295339	10	0,5714	-1		PDE4D	5q11.2	5q11.2
#1896T	chr5	58334742	180794757	6411	0,9466	0	ATP10B,MEF2C,DOCK2,ADAMTS19,APC,MAP1B		5q11.2	5q35.3
#1896T	chr6	292550	65149150	4971	0,937	0	CDKN1A,VEGFA		6p25.3	6q12
#1896T	chr6	65300999	117203561	2022	0,5869	-1	EEF1A1,FRK		6q12	6q22.1
#1896T	chr6	117215189	117641112	27	0,908	0			6q22.1	6q22.1
#1896T	chr6	117642489	170893528	2605	0,5848	-1		PARK2	6q22.1	6q27
#1896T	chr7	193502	158937438	9099	0,9459	0	MET,HGF,EGFR,KMT2C,BRAF,NUP205		7p22.3	7q36.3
#1896T	chr8	116555	146279474	6482	0,9425	0	MYC,RSPO2		8p23.3	8q24.3
#1896T	chr9	14874	141016186	7831	0,9367	0	CDKN2A,TSC1,NOTCH1,PTPN3		9p24.3	9q34.3
#1896T	chr10	93526	135440158	7983	0,9443	0	PTEN		10p15.3	10q26.3
#1896T	chr11	193127	134257498	10910	0,9391	0	HMBS,FGF19,IGF2,CDKN1C,ATM		11p15.5	11q25
#1896T	chr12	176326	133810818	11030	0,9446	0	ARID2,HNF1A,GLI1,KMT2D,KRAS,PTPRB		12p13.33	12q24.33
#1896T	chr13	19748151	96622441	2447	0,9564	0	RB1,MYCBP2		13q12.11	13q32.1
#1896T	chr13	96624880	96675948	9	0,5276	-2			13q32.1	13q32.1
#1896T	chr13	96684184	115090537	919	0,9226	0			13q32.1	13q34
#1896T	chr14	19378084	105995651	5765	0,9435	0	DICER1		14q11.2	14q32.33
#1896T	chr15	20740212	102462803	6835	0,9496	0			15q11.2	15q26.3
#1896T	chr16	97494	90142284	8255	0,9504	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#1896T	chr17	193127	81052161	11559	0,9476	0	TP53,STAT3,ERBB2,NF1		17p13.3	17q25.3
#1896T	chr18	158706	78005195	2946	0,9477	0			18p11.32	18q23
#1896T	chr19	111138	59082559	11156	0,9554	0	KEAP1,CCNE1,JAK3,KMT2B,DNAJB1,PRKACA		19p13.3	19q13.43
#1896T	chr20	68379	62904763	4751	0,9432	0	GNAS		20p13	20q13.33
#1896T	chr21	10959769	42218566	1028	0,5849	-1	DYRK1A	ERG,DSCAM	21p11.1	21q22.2
#1896T	chr21	42540347	48084246	888	0,9134	0			21q22.2	21q22.3
#1896T	chr21	10969099	45658349	83	0,9341	0			21q22.3	21p11.1
#1896T	chr22	19378084	51220669	4045	0,9377	0	ZNRF3,NF2		22q11.21	22q13.33
#1896T	chrX	200918	155239777	6794	0,5725	-1	RPS6KA3	DMD	Xp22.33	Xq28
#1896T	chrY	150918	59342783	121	0,6128	-1			Yp11.32	Yq12
#2189T	chr1	721657	249212472	18794	0,0814	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#2189T	chr2	40221	242842554	14271	0,0887	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#2189T	chr3	239470	197769162	11532	0,0849	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#2189T	chr4	53291	190947577	7422	0,0977	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#2189T	chr5	140602	177054498	7964	0,0893	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#2189T	chr5	177163594	177309702	6	0,6198	1			5q35.3	5q35.3
#2189T	chr5	177310749	180687559	402	0,0673	0			5q35.3	5q35.3
#2189T	chr6	203615	170892714	9687	0,0893	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#2189T	chr7	195642	158937319	9014	0,0861	0	MET		7p22.3	7q36.3
#2189T	chr8	190866	145001461	6115	0,0961	0	SLURP1,MYC		8p23.3	8q24.3
#2189T	chr8	145001756	145006648	6	-0,3399	-1			8q24.3	8q24.3
#2189T	chr8	145007120	146280815	359	0,0421	0			8q24.3	8q24.3
#2189T	chr9	172880	141070232	7493	0,0739	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#2189T	chr10	226000	135439464	7604	0,0891	0	PTEN		10p15.3	10q26.3
#2189T	chr11	193118	134856548	10639	0,0777	0	HMBS		11p15.5	11q25
#2189T	chr12	208345	133811687	11105	0,0888	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#2189T	chr13	19600450	115091029	3312	0,0947	0	RB1		13q12.11	13q34
#2189T	chr14	20201689	107283203	6272	0,0871	0			14q11.2	14q32.33
#2189T	chr15	20170066	83936934	5539	0,0836	0			15q11.1	15q25.2
#2189T	chr15	83951925	84241383	5	-0,3938	-1			15q25.2	15q25.2
#2189T	chr15	84245414	102389376	1092	0,082	0			15q25.2	15q26.3
#2189T	chr16	97096	90161144	7691	0,0721	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#2189T	chr17	6088	81083508	10854	0,0702	0	TP53		17p13.3	17q25.3
#2189T	chr18	117194	9593818	489	0,0973	0			18p11.32	18p11.22
#2189T	chr18	9595083	9814052	5	-0,348	-1			18p11.22	18p11.22
#2189T	chr18	9815165	77941018	2450	0,098	0			18p11.22	18q23
#2189T	chr19	107510	59083958	10017	0,0612	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#2189T	chr20	68379	62906077	4715	0,1037	0	DNAJC5		20p13	20q13.33
#2189T	chr21	9825922	48084620	1947	0,082	0	DYRK1A		21p11.2	21q22.3
#2189T	chr22	17060842	51219140	4004	0,0691	0	ZNRF3,CLDN5		22q11.1	22q13.33
#2189T	chrX	2700138	154774846	4595	0,0774	0	RPS6KA3		Xp22.33	Xq28
#2189T	chrY	2655318	24457064	145	0,0874	0			Yp11.31	Yq11.223
#222T	chr1	861348	203012574	14887	0,1045	0	ARID1A,RPL22,COL11A1		1p36.33	1q32.1
#222T	chr1	203013543	203018848	5	-0,3601	-1			1q32.1	1q32.1
#222T	chr1	203020658	249212472	3054	0,1154	0			1q32.1	1q44
#222T	chr2	40221	242842554	13571	0,1314	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#222T	chr3	239470	197769162	11085	0,1225	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#222T	chr4	53291	190947044	7105	0,1423	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#222T	chr5	140602	180687559	8014	0,1312	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#222T	chr6	203615	170892714	9230	0,1232	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#222T	chr7	193386	100244414	4910	-0,0624	0			7p22.3	7q22.1
#222T	chr7	100244650	100253126	6	-0,6527	-1			7q22.1	7q22.1
#222T	chr7	100253464	158937319	3887	-0,056	0	MET		7q22.1	7q36.3
#222T	chr8	190866	146279982	6220	0,1286	0	SLURP1,MYC		8p23.3	8q24.3
#222T	chr9	214910	127220947	4735	0,124	0	CDKN2A,MTAP,PTPN3		9p24.3	9q33.3
#222T	chr9	127230418	127238964	7	-0,252	-1			9q33.3	9q33.3
#222T	chr9	127244400	141017498	2354	0,0475	0	TSC1		9q33.3	9q34.3
#222T	chr10	226000	135438930	7286	0,1144	0	PTEN		10p15.3	10q26.3
#222T	chr11	193118	134856548	10161	0,0992	0	HMBS		11p15.5	11q25
#222T	chr12	208345	88505552	6789	0,1226	0	ARID2,CDKN1B		12p13.33	12q21.32
#222T	chr12	88509300	88534771	8	0,5353	1			12q21.32	12q21.32
#222T	chr12	88535046	133811687	3817	0,098	0	HNF1A		12q21.32	12q24.33
#222T	chr13	19601058	115091029	3162	0,144	0	RB1		13q12.11	13q34
#222T	chr14	20201689	107282930	6019	0,1073	0			14q11.2	14q32.33
#222T	chr15	20170066	102389376	6304	0,1125	0			15q11.1	15q26.3
#222T	chr16	96492	90161144	7283	0,0655	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#222T	chr17	6088	81052160	10210	0,0736	0	TP53		17p13.3	17q25.3
#222T	chr18	117194	77941018	2818	0,1289	0			18p11.32	18q23
#222T	chr19	281462	59083958	9374	0,0153	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#222T	chr20	68379	62906077	4527	0,0896	0	DNAJC5		20p13	20q13.33

#222T	chr21	9825922	48084620	1860	0,1021	0	DYRK1A		21p11.2	21q22.3
#222T	chr22	17063472	51219140	3796	0,0617	0	ZNRF3,CLDN5		22q11.1	22q13.33
#222T	chrX	2700138	154774846	4077	0,0988	0	RPS6KA3		Xp22.33	Xq28
#222T	chrY	2655318	23548182	122	0,1171	0			Yp11.31	Yq11.223
#2615T	chr1	721657	84641485	7908	-0,1422	0	ARID1A,RPL22		1p36.33	1p31.1
#2615T	chr1	84644854	84650824	5	-0,6937	-1			1p31.1	1p31.1
#2615T	chr1	84662362	249212472	11080	-0,1442	0	COL11A1		1p31.1	1q44
#2615T	chr2	40221	242946533	14500	-0,1369	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#2615T	chr3	239470	197769162	11715	-0,1481	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#2615T	chr4	53291	190947577	7639	-0,1399	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#2615T	chr5	140602	180687559	8569	-0,1418	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#2615T	chr6	203615	170892714	9829	-0,1491	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#2615T	chr7	195642	21521641	1107	-0,142	0			7p22.3	7p15.3
#2615T	chr7	21551452	21913004	78	-0,4883	-1			7p15.3	7p15.3
#2615T	chr7	21920414	158937319	7891	-0,1479	0	MET		7p15.3	7q36.3
#2615T	chr8	190866	25286210	1236	-0,1195	0			8p23.3	8p21.2
#2615T	chr8	25287278	25298140	7	-0,4595	-1			8p21.2	8p21.2
#2615T	chr8	25301829	146280815	5333	-0,143	0	SLURP1,MYC		8p21.2	8q24.3
#2615T	chr9	172880	141017498	7493	-0,1447	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#2615T	chr10	226000	135439464	7728	-0,1406	0	PTEN		10p15.3	10q26.3
#2615T	chr11	193118	134856548	10634	-0,14	0	HMBS		11p15.5	11q25
#2615T	chr12	208345	133811687	11259	-0,1439	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#2615T	chr13	19600450	115091029	3427	-0,1414	0	RB1		13q12.11	13q34
#2615T	chr14	20201689	107282930	6363	-0,1442	0			14q11.2	14q32.33
#2615T	chr15	20170066	102389376	6697	-0,2156	0			15q11.1	15q26.3
#2615T	chr16	96492	90161144	7575	-0,142	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#2615T	chr17	6088	81052160	10743	-0,1434	0	TP53		17p13.3	17q25.3
#2615T	chr18	117194	77960704	3010	-0,1396	0			18p11.32	18q23
#2615T	chr19	107510	59083958	9643	-0,151	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#2615T	chr20	68379	62906077	4681	-0,1397	0	DNAJC5		20p13	20q13.33
#2615T	chr21	9825922	48084620	1968	-0,2622	0	DYRK1A		21p11.2	21q22.3
#2615T	chr22	17060354	51219140	3957	-0,1405	0	ZNRF3,CLDN5		22q11.1	22q13.33
#2615T	chrX	2700138	154774846	6320	-0,1478	0	RPS6KA3		Xp22.33	Xq28
#2729T	chr1	69550	110260004	10019	0,7148	0	ARID1A,RPL22,COL11A1		1p36.33	1p13.3
#2729T	chr1	110279742	110282886	8	0,4017	0			1p13.3	1p13.3
#2729T	chr1	110293018	146739131	1452	0,6656	0			1p13.3	1q21.1
#2729T	chr1	146740489	146747844	5	0,2869	-1			1q21.1	1q21.1
#2729T	chr1	146751781	153752390	1079	0,6864	0			1q21.1	1q21.3
#2729T	chr1	153782720	153791332	8	0,3793	0			1q21.3	1q21.3
#2729T	chr1	153792146	155156470	414	0,7253	0			1q21.3	1q22
#2729T	chr1	155158648	155165847	11	1,084	1			1q22	1q22
#2729T	chr1	155166918	186289497	3159	0,6621	0			1q22	1q31.1
#2729T	chr1	186291497	186300672	7	0,199	-1			1q31.1	1q31.1
#2729T	chr1	186301401	249211681	4128	0,6744	0			1q31.1	1q44
#2729T	chr2	41618	242814646	15016	0,6699	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#2729T	chr3	361505	126451993	7694	0,6925	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q21.3
#2729T	chr3	126571531	127295700	42	1,0134	1			3q21.3	3q21.3
#2729T	chr3	127295876	197765488	3970	0,6496	0			3q21.3	3q29
#2729T	chr4	53384	113242	7572	0,639	0			4p16.3	4p16.3
#2729T	chr5	92369	180018436	8550	0,6607	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#2729T	chr5	180030291	180058730	30	0,9596	0			5q35.3	5q35.3
#2729T	chr5	180076516	180794757	64	0,6987	0			5q35.3	5q35.3
#2729T	chr6	292550	170893528	9752	0,6671	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#2729T	chr7	193502	158937438	9228	0,6899	0	MET		7p22.3	7q36.3
#2729T	chr8	116555	146279474	6599	0,6813	0	SLURP1,MYC		8p23.3	8q24.3
#2729T	chr9	14874	141016186	7940	0,7144	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#2729T	chr10	93526	76854535	3792	0,6631	0			10p15.3	10q22.2
#2729T	chr10	76855443	76865526	5	0,9547	0			10q22.2	10q22.2
#2729T	chr10	76867858	135440158	4289	0,7022	0	PTEN		10q22.2	10q26.3
#2729T	chr11	193127	73078727	7167	0,7508	0			11p15.5	11q13.4
#2729T	chr11	73100204	73118632	11	1,0292	1			11q13.4	11q13.4
#2729T	chr11	73120619	126141442	3458	0,6621	0	HMBS		11q13.4	11q24.2
#2729T	chr11	126142919	126160823	10	1,0376	1			11q24.2	11q24.2
#2729T	chr11	126162692	134257498	376	0,7098	0			11q24.2	11q25
#2729T	chr12	176326	133810818	11199	0,6758	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#2729T	chr13	19748151	115090537	3441	0,6518	0	RB1		13q12.11	13q34
#2729T	chr14	19378084	105995651	5880	0,6873	0			14q11.2	14q32.33
#2729T	chr15	20740212	28625712	347	0,7131	0			15q11.2	15q13.1
#2729T	chr15	28626388	28630829	11	1,0531	1			15q13.1	15q13.1
#2729T	chr15	28632272	91354536	6127	0,698	0			15q13.1	15q26.1
#2729T	chr15	91358420	91456914	52	0,9726	1			15q26.1	15q26.1
#2729T	chr15	91459420	102462803	417	0,6783	0			15q26.1	15q26.3
#2729T	chr16	97494	74487148	7072	0,7501	0	AXIN1,TSC2,BRD7		16p13.3	16q23.1
#2729T	chr16	74490600	74683065	34	0,4966	0			16q23.1	16q23.1
#2729T	chr16	74685919	90142284	1215	0,7763	0			16q23.1	16q24.3
#2729T	chr17	193127	17716738	2969	0,7384	0	TP53		17p13.3	17p11.2
#2729T	chr17	17716983	17720361	9	1,1405	1			17p11.2	17p11.2
#2729T	chr17	17720670	19648361	418	0,7811	0			17p11.2	17p11.2
#2729T	chr17	19679691	19752700	23	0,512	0			17p11.2	17p11.2
#2729T	chr17	19753071	19769095	3	1,0394	1			17p11.2	17p11.2
#2729T	chr17	19770685	37226189	1806	0,7324	0			17p11.2	17q12
#2729T	chr17	37228721	37234233	2	1,0145	NA			17q12	17q12
#2729T	chr17	37235387	60663541	3437	0,724	0			17q12	17q23.2
#2729T	chr17	60673991	60689839	6	0,4401	0			17q23.2	17q23.2
#2729T	chr17	60705223	74303563	1760	0,7482	0			17q23.2	17q25.1
#2729T	chr17	74307702	74326689	6	0,4126	0			17q25.1	17q25.1
#2729T	chr17	74328430	79567410	808	0,8273	0			17q25.1	17q25.3
#2729T	chr17	79571664	79596791	7	0,4409	0			17q25.3	17q25.3
#2729T	chr17	79603949	81052161	457	0,8839	0			17q25.3	17q25.3
#2729T	chr18	158706	61558762	2649	0,6362	0			18p11.32	18q21.33
#2729T	chr18	61562558	61564389	2	0,9883	1			18q21.33	18q21.33
#2729T	chr18	61565020	78005195	361	0,6804	0			18q21.33	18q23
#2729T	chr19	111138	6833733	1862	0,8228	0			19p13.3	19p13.3
#2729T	chr19	6833941	6890523	11	0,3765	0			19p13.3	19p13.3
#2729T	chr19	6896481	11556246	1315	0,7317	0	KEAP1		19p13.3	19p13.2
#2729T	chr19	11557126	11560154	9	1,0457	1			19p13.2	19p13.2
#2729T	chr19	11565517	42933084	4554	0,7471	0			19p13.2	19q13.2
#2729T	chr19	42937178	43015754	6	0,4134	0			19q13.2	19q13.2
#2729T	chr19	43016566	59082559	3494	0,7608	0	SEPW1,TMEM160		19q13.2	19q13.43
#2729T	chr20	68379	62904763	4813	0,7454	0	DNAJC5		20p13	20q13.33
#2729T	chr21	45658349	48083380	2044	0,7227	0			21q22.3	21q22.3
#2729T	chr22	19378084	24724850	952	0,8068	0	CLDN5		22q11.21	22q11.23
#2729T	chr22	24726312	24765237	7	0,469	0			22q11.23	22q11.23
#2729T	chr22	24807614	44587956	2405	0,7499	0	ZNRF3		22q11.23	22q13.31

#2729T	chr22	44589692	44601673	5	1,2762	1			22q13.31	22q13.31
#2729T	chr22	44602252	50757413	520	0,8338	0			22q13.31	22q13.33
#2729T	chr22	50765329	50854553	5	0,5585	0			22q13.33	22q13.33
#2729T	chr22	50857352	51220669	209	0,9167	0			22q13.33	22q13.33
#2729T	chrX	200918	155239777	6821	0,6739	0	RPS6KA3		Xp22.33	Xq28
#2729T	chrY	150918	59342783	129	0,7476	0			Yp11.32	Yq12
#343T	chr1	762325	249212472	19104	-0,0445	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#343T	chr2	40221	228996748	13267	-0,0558	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q36.3
#343T	chr2	229034976	229889758	5	0,3338	1			2q36.3	2q36.3
#343T	chr2	230020606	242946533	1397	-0,0251	0			2q36.3	2q37.3
#343T	chr3	239470	197769162	11848	-0,05	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#343T	chr4	53291	190947044	7745	-0,0651	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#343T	chr5	140602	180687559	8641	-0,0562	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#343T	chr6	203615	170892714	9917	-0,0552	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#343T	chr7	195642	158937319	9109	-0,0482	0	MET		7p22.3	7q36.3
#343T	chr8	190866	146280815	6666	-0,0483	0	SLURP1,MYC		8p23.3	8q24.3
#343T	chr9	213465	141017498	7569	-0,0319	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#343T	chr10	226000	33093984	1706	-0,0637	0			10p15.3	10p11.22
#343T	chr10	33103339	33144303	10	-0,3936	-1			10p11.22	10p11.22
#343T	chr10	33165390	135438930	6072	-0,0485	0	PTEN		10p11.22	10q26.3
#343T	chr11	193118	134856548	10763	-0,0322	0	HMBS		11p15.5	11q25
#343T	chr12	208345	49920264	3864	-0,0599	0	ARID2,CDKN1B		12p13.33	12q13.12
#343T	chr12	49934798	49960563	27	0,1715	0			12q13.12	12q13.12
#343T	chr12	49976851	133811687	7477	-0,0485	0	HNF1A		12q13.12	12q24.33
#343T	chr13	19600450	115091029	3459	-0,0613	0	RB1		13q12.11	13q34
#343T	chr14	20201689	107282930	6439	-0,0448	0			14q11.2	14q32.33
#343T	chr15	20170066	56676189	3091	-0,0513	0			15q11.1	15q21.3
#343T	chr15	56680694	56719870	6	-0,4976	-1			15q21.3	15q21.3
#343T	chr15	56720736	102389376	3639	-0,0438	0			15q21.3	15q26.3
#343T	chr16	96492	90142284	7652	-0,0096	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#343T	chr17	6088	81052160	10797	-0,0215	0	TP53		17p13.3	17q25.3
#343T	chr18	117194	77960704	3035	-0,0638	0			18p11.32	18q23
#343T	chr19	281462	603755	58	0,0527	0			19p13.3	19p13.3
#343T	chr19	617298	619252	5	-0,4547	-1			19p13.3	19p13.3
#343T	chr19	620018	59083958	9824	0,0058	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#343T	chr20	68379	62906077	4752	-0,0231	0	DNAJC5		20p13	20q13.33
#343T	chr21	9825922	10862908	8	0,2309	0			21p11.2	21p11.2
#343T	chr21	10906821	48084620	1998	-0,0429	0	DYRK1A		21p11.1	21q22.3
#343T	chr22	17060842	51219140	4000	-0,0055	0	ZNRF3,CLDN5		22q11.1	22q13.33
#343T	chrX	2700138	38013775	1294	-0,067	0	RPS6KA3		Xp22.33	Xp11.4
#343T	chrX	38016215	38080049	7	0,2686	1			Xp11.4	Xp11.4
#343T	chrX	38080636	154774846	5066	-0,0496	0			Xp11.4	Xq28
#375T	chr1	69550	249211681	20300	0,5141	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#375T	chr2	41618	242814646	14999	0,528	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#375T	chr3	361505	197765488	11710	0,5218	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#375T	chr4	53384	113242	7579	0,5408	0			4p16.3	4p16.3
#375T	chr5	92369	180794757	8653	0,5295	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#375T	chr6	292550	170893528	9755	0,5263	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#375T	chr7	193502	158937438	9218	0,519	0	MET		7p22.3	7q36.3
#375T	chr8	116555	146279474	6615	0,5275	0	SLURP1,MYC		8p23.3	8q24.3
#375T	chr9	15115	141016186	7934	0,5081	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#375T	chr10	93526	135440158	8080	0,5253	0	PTEN		10p15.3	10q26.3
#375T	chr11	193127	134257498	11034	0,503	0	HMBS		11p15.5	11q25
#375T	chr12	176326	133810818	11185	0,52	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#375T	chr13	19748151	115090537	3449	0,5412	0	RB1		13q12.11	13q34
#375T	chr14	19378084	105995651	5872	0,5232	0			14q11.2	14q32.33
#375T	chr15	20740212	102462803	6952	0,5138	0			15q11.2	15q26.3
#375T	chr16	97494	90142284	8318	0,4943	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#375T	chr17	193127	81052161	11695	0,5022	0	TP53		17p13.3	17q25.3
#375T	chr18	158706	78005195	3008	0,5374	0			18p11.32	18q23
#375T	chr19	111138	59082559	11245	0,4781	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#375T	chr20	68379	62904763	4804	0,5005	0	DNAJC5		20p13	20q13.33
#375T	chr21	45658349	48083380	2024	0,5156	0			21q22.3	21q22.3
#375T	chr22	19378084	51220669	4098	0,4866	0	ZNRF3,CLDN5		22q11.21	22q13.33
#375T	chrX	200918	155239777	6818	0,5278	0	RPS6KA3		Xp22.33	Xq28
#375T	chrY	150918	59342783	129	0,514	0			Yp11.32	Yq12
#4110T	chr1	69550	861358	4	1,0004	0			1p36.33	1p36.33
#4110T	chr1	865626	16174604	2055	0,6521	0	RPL22		1p36.33	1p36.21
#4110T	chr1	16199471	16248794	8	1,1079	0			1p36.21	1p36.13
#4110T	chr1	16258673	20209117	636	0,7031	0			1p36.13	1p36.13
#4110T	chr1	20216952	20234151	6	1,0811	0			1p36.13	1p36.13
#4110T	chr1	20246871	21151642	118	0,6577	0			1p36.13	1p36.12
#4110T	chr1	21155703	21329221	27	1,0715	0			1p36.12	1p36.12
#4110T	chr1	21546536	21978293	68	0,6922	0			1p36.12	1p36.12
#4110T	chr1	22005921	22141103	26	0,9417	0			1p36.12	1p36.12
#4110T	chr1	22142501	22338951	116	0,5931	0			1p36.12	1p36.12
#4110T	chr1	22405024	22417965	5	1,2045	1			1p36.12	1p36.12
#4110T	chr1	22446777	22469377	5	0,5076	-1			1p36.12	1p36.12
#4110T	chr1	22816790	24666205	275	0,7384	0			1p36.12	1p36.11
#4110T	chr1	24668684	24673087	5	0,3998	-2			1p36.11	1p36.11
#4110T	chr1	24674000	52822085	3378	0,7769	0	ARID1A		1p36.11	1p32.3
#4110T	chr1	52822736	52830258	15	0,5537	-1			1p32.3	1p32.3
#4110T	chr1	52838950	92798989	2065	0,9813	0			1p32.3	1p22.1
#4110T	chr1	92801944	92948987	10	0,6942	0			1p22.1	1p22.1
#4110T	chr1	92979370	94033353	100	1,011	0			1p22.1	1p22.1
#4110T	chr1	94037313	94140328	9	0,7156	0			1p22.1	1p22.1
#4110T	chr1	94335454	94370112	12	1,0509	0			1p22.1	1p22.1
#4110T	chr1	94374656	94586568	50	0,7872	0			1p22.1	1p22.1
#4110T	chr1	94639865	109486568	618	1,0456	0	COL11A1		1p22.1	1p13.3
#4110T	chr1	109490286	110091402	167	0,7753	0			1p13.3	1p13.3
#4110T	chr1	110116380	110146071	8	1,236	1			1p13.3	1p13.3
#4110T	chr1	110146650	110170096	16	0,7241	0			1p13.3	1p13.3
#4110T	chr1	110170460	110172942	8	0,345	-2			1p13.3	1p13.3
#4110T	chr1	110173360	111439315	157	0,753	0			1p13.3	1p13.3
#4110T	chr1	111440472	111667434	10	1,2216	1			1p13.3	1p13.3
#4110T	chr1	111668882	178754717	5209	0,8644	0			1p13.3	1q25.2
#4110T	chr1	178777221	178871293	18	1,2377	1			1q25.2	1q25.2
#4110T	chr1	178875957	180080214	139	0,9776	0			1q25.2	1q25.2
#4110T	chr1	180124175	180243517	19	0,6056	0			1q25.2	1q25.2
#4110T	chr1	180257575	227504795	3210	0,9034	0			1q25.2	1q42.13
#4110T	chr1	227751411	229580705	218	0,6052	0			1q42.13	1q42.13
#4110T	chr1	229584905	249211681	1160	0,9355	0			1q42.13	1q44
#4110T	chr2	41618	42174673	2515	0,8699	0	APOB		2p25.3	2p21
#4110T	chr2	42180282	42472736	10	0,5818	0			2p21	2p21

#4110T	chr2	42483706	42950102	43	0,9732	0			2p21	2p21
#4110T	chr2	42980526	43655305	25	0,6642	0			2p21	2p21
#4110T	chr2	43657397	44047147	66	1,004	0			2p21	2p21
#4110T	chr2	44050005	44104983	23	0,6916	0			2p21	2p21
#4110T	chr2	44115767	71188102	1281	0,9719	0			2p21	2p13.3
#4110T	chr2	71188774	71216096	16	0,5383	-1			2p13.3	2p13.3
#4110T	chr2	71216912	71658508	52	0,9384	0			2p13.3	2p13.2
#4110T	chr2	71660354	72411260	69	0,6617	0			2p13.2	2p13.2
#4110T	chr2	72562113	72968515	14	1,1042	0			2p13.2	2p13.2
#4110T	chr2	73052983	73118571	4	0,7621	0			2p13.2	2p13.2
#4110T	chr2	73145241	73228681	12	0,4386	-1			2p13.2	2p13.2
#4110T	chr2	73247381	113420521	2555	0,8747	0			2p13.2	2q13
#4110T	chr2	113479431	113484285	14	0,6052	0			2q13	2q13
#4110T	chr2	113487243	127658960	455	0,9011	0			2q13	2q14.3
#4110T	chr2	127806156	127834242	19	0,5056	-1			2q14.3	2q14.3
#4110T	chr2	127864478	128065268	24	0,7882	0			2q14.3	2q14.3
#4110T	chr2	128066250	128100704	15	1,1511	0			2q14.3	2q14.3
#4110T	chr2	128177554	128244184	10	0,5838	0			2q14.3	2q14.3
#4110T	chr2	128246783	128262794	10	1,0431	0			2q14.3	2q14.3
#4110T	chr2	128281310	128466358	65	0,5826	0			2q14.3	2q14.3
#4110T	chr2	128467110	130877828	107	0,9363	0			2q14.3	2q21.1
#4110T	chr2	130897195	131130763	66	0,6237	0			2q21.1	2q21.1
#4110T	chr2	131221053	131261410	14	1,134	0			2q21.1	2q21.1
#4110T	chr2	131266548	131369311	14	0,5922	0			2q21.1	2q21.1
#4110T	chr2	131374420	131403793	12	1,1482	0			2q21.1	2q21.1
#4110T	chr2	131412554	131976236	29	0,7567	0			2q21.1	2q21.1
#4110T	chr2	131981226	132120911	13	1,0611	0			2q21.1	2q21.1
#4110T	chr2	132233774	132558487	20	0,614	0			2q21.1	2q21.2
#4110T	chr2	133066889	172844218	1866	1,0238	0	ACVR2A		2q21.2	2q31.1
#4110T	chr2	172848182	173330329	17	0,62	0			2q31.1	2q31.1
#4110T	chr2	173332248	176860299	201	0,9678	0			2q31.1	2q31.1
#4110T	chr2	176866953	177134378	23	0,6445	0			2q31.1	2q31.1
#4110T	chr2	177161612	238738003	4410	0,9301	0	NFE2L2,EPHA4		2q31.1	2q37.3
#4110T	chr2	238742713	242814646	554	0,6255	0			2q37.3	2q37.3
#4110T	chr3	361505	42642528	2296	0,8687	0	CTNNB1		3p26.3	3p22.1
#4110T	chr3	42659099	42687424	14	1,1469	0			3p22.1	3p22.1
#4110T	chr3	42700584	47456400	563	0,7767	0	SETD2		3p22.1	3p21.31
#4110T	chr3	47456629	47462153	8	0,3612	-2			3p21.31	3p21.31
#4110T	chr3	47462470	48982595	445	0,6596	0			3p21.31	3p21.31
#4110T	chr3	48999079	49011191	6	1,0004	0			3p21.31	3p21.31
#4110T	chr3	49012275	52233357	860	0,6761	0			3p21.31	3p21.2
#4110T	chr3	52236636	52242181	6	1,0347	0			3p21.2	3p21.2
#4110T	chr3	52245433	52574470	271	0,5965	0	BAP1		3p21.2	3p21.1
#4110T	chr3	52582165	129599253	3427	0,9139	0	PDE12,KLF15		3p21.1	3q22.1
#4110T	chr3	129694765	129817082	10	0,5649	0			3q22.1	3q22.1
#4110T	chr3	130092515	183862702	2464	0,9855	0			3q22.1	3q27.1
#4110T	chr3	183873504	184552484	243	0,6445	0			3q27.1	3q27.2
#4110T	chr3	184556518	195269787	558	0,9627	0			3q27.2	3q29
#4110T	chr3	195295889	195778934	51	0,6045	0			3q29	3q29
#4110T	chr3	195780359	195925703	19	0,9857	0			3q29	3q29
#4110T	chr3	195934321	195955065	10	0,4649	-1			3q29	3q29
#4110T	chr3	195955735	197765488	213	0,8803	0			3q29	3q29
#4110T	chr4	53384	493201	18	0,9049	0			4p16.3	4p16.3
#4110T	chr4	494287	2835493	414	0,5807	0			4p16.3	4p16.3
#4110T	chr4	2877740	2901074	8	1,0107	0			4p16.3	4p16.3
#4110T	chr4	2906663	8608545	518	0,6867	0			4p16.3	4p16.1
#4110T	chr4	8609073	108871490	3756	1,0076	0	ALB,GABRA2		4p16.1	4q25
#4110T	chr4	108911154	108984802	12	0,6882	0			4q25	4q25
#4110T	chr4	113242	108985516	2560	1,0053	0			4q25	4p16.3
#4110T	chr5	92369	5423071	317	0,597	0	TERT		5p15.33	5p15.32
#4110T	chr5	5440017	5454708	7	1,3268	1			5p15.32	5p15.32
#4110T	chr5	5457649	38427232	1024	0,9518	0			5p15.32	5p13.1
#4110T	chr5	38431334	38463062	8	0,5742	0			5p13.1	5p13.1
#4110T	chr5	38464011	54459964	480	1,0259	0			5p13.1	5q11.2
#4110T	chr5	54468447	54528282	11	0,5425	-1			5q11.2	5q11.2
#4110T	chr5	54529161	132094514	3093	1,0015	0	MEF2C,IL6ST,ADAMTS19		5q11.2	5q31.1
#4110T	chr5	132096589	132161344	17	0,6011	0			5q31.1	5q31.1
#4110T	chr5	132197764	160973611	2170	0,8551	0	ATP10B		5q31.1	5q34
#4110T	chr5	161113014	161530964	22	1,1868	1			5q34	5q34
#4110T	chr5	161569246	175777678	487	0,8763	0	DOCK2		5q34	5q35.2
#4110T	chr5	175779688	175906223	22	0,5508	-1			5q35.2	5q35.2
#4110T	chr5	175913423	175956775	13	0,9187	0			5q35.2	5q35.2
#4110T	chr5	175957145	178017051	401	0,6394	0			5q35.2	5q35.3
#4110T	chr5	178030688	178139841	12	1,0027	0			5q35.3	5q35.3
#4110T	chr5	178152425	179334762	160	0,7078	0			5q35.3	5q35.3
#4110T	chr5	179382662	179529107	11	0,9974	0			5q35.3	5q35.3
#4110T	chr5	179538518	180794757	151	0,7158	0			5q35.3	5q35.3
#4110T	chr6	292550	43276509	3929	0,7782	0	CDKN1A		6p25.3	6p21.1
#4110T	chr6	43306568	43320164	6	1,0521	0			6p21.1	6p21.1
#4110T	chr6	43323142	44364127	272	0,7153	0	SPATS1,VEGFA		6p21.1	6p21.1
#4110T	chr6	44371655	100010771	2020	1,0181	0	EEF1A1		6p21.1	6q16.2
#4110T	chr6	100016388	100061772	5	0,6114	0			6q16.2	6q16.2
#4110T	chr6	100368974	100895218	9	0,9945	0			6q16.2	6q16.3
#4110T	chr6	100896075	100960690	10	0,72	0			6q16.3	6q16.3
#4110T	chr6	100964130	109757355	293	1,0025	0			6q16.3	6q21
#4110T	chr6	109761726	109816592	45	0,6552	0			6q21	6q21
#4110T	chr6	109818717	170181518	2920	0,9552	0			6q21	6q27
#4110T	chr6	170591964	170599200	11	0,4704	-1			6q27	6q27
#4110T	chr6	170600153	170893528	29	0,8944	0			6q27	6q27
#4110T	chr7	193502	131060217	7103	0,8548	0	MET		7p22.3	7q32.3
#4110T	chr7	131071950	131151105	13	1,1893	1			7q32.3	7q32.3
#4110T	chr7	131155652	148544332	1067	0,8831	0			7q32.3	7q36.1
#4110T	chr7	148701094	150094392	211	0,5951	0			7q36.1	7q36.1
#4110T	chr7	150164104	150416157	12	1,0317	0			7q36.1	7q36.1
#4110T	chr7	150417574	158937438	645	0,7549	0			7q36.1	7q36.3
#4110T	chr8	116555	37635556	1779	0,9073	0			8p23.3	8p11.23
#4110T	chr8	37654920	37702370	20	0,4803	-1			8p11.23	8p11.23
#4110T	chr8	37704532	141542197	3719	0,9673	0	MYC		8p11.23	8q24.3
#4110T	chr8	141542615	141645595	18	0,5648	0			8q24.3	8q24.3
#4110T	chr8	141669671	141900739	33	0,9706	0			8q24.3	8q24.3
#4110T	chr8	142146818	146279474	867	0,5218	-1	SLURP1		8q24.3	8q24.3
#4110T	chr9	15115	117460	10	0,683	0			9p24.3	9p24.3
#4110T	chr9	121520	21818117	840	1,0015	0	MTAP		9p24.3	9p21.3
#4110T	chr9	21837958	22447398	12	0,7001	0	CDKN2A,MTAP		9p21.3	9p21.3

#4110T	chr9	22451487	33369940	216	0,9766	0			9p21.3	9p13.3
#4110T	chr9	33385146	33528782	41	0,5776	0			9p13.3	9p13.3
#4110T	chr9	33529084	95280063	1911	0,8896	0			9p13.3	9q22.31
#4110T	chr9	95285002	95526906	26	0,6284	0			9q22.31	9q22.31
#4110T	chr9	95609672	95765252	6	0,9336	0			9q22.31	9q22.31
#4110T	chr9	95766351	96214594	68	0,5902	0			9q22.31	9q22.31
#4110T	chr9	96233546	96318768	12	0,8708	0			9q22.31	9q22.31
#4110T	chr9	96320200	96846934	33	0,5548	-1			9q22.31	9q22.32
#4110T	chr9	96847618	97054686	10	0,9707	0			9q22.32	9q22.32
#4110T	chr9	97055304	97090898	9	0,5588	-1			9q22.32	9q22.32
#4110T	chr9	97137156	98638327	81	0,8572	0			9q22.32	9q22.32
#4110T	chr9	98643363	98691074	10	1,2649	1			9q22.32	9q22.32
#4110T	chr9	98703784	125946412	1762	0,8809	0	PTPN3		9q22.32	9q33.3
#4110T	chr9	126118587	126164140	16	0,5056	-1			9q33.3	9q33.3
#4110T	chr9	126165725	126692176	19	0,9365	0			9q33.3	9q33.3
#4110T	chr9	126774687	127662779	77	0,6556	0			9q33.3	9q33.3
#4110T	chr9	127670703	129246281	105	0,9246	0			9q33.3	9q33.3
#4110T	chr9	129265499	131122691	405	0,6514	0			9q33.3	9q34.11
#4110T	chr9	131133677	131140350	2	1,1149	0			9q34.11	9q34.11
#4110T	chr9	131150137	141016186	2164	0,6261	0	TSC1		9q34.11	9q34.3
#4110T	chr10	93526	43326429	1867	0,932	0			10p15.3	10q11.21
#4110T	chr10	43572743	43623638	21	0,5511	-1			10q11.21	10q11.21
#4110T	chr10	43650928	94821123	2675	0,8629	0	PTEN		10q11.21	10q23.33
#4110T	chr10	94821879	94835650	10	0,5339	-1			10q23.33	10q23.33
#4110T	chr10	94836377	131973150	3039	0,8613	0			10q23.33	10q26.3
#4110T	chr10	131973307	135440158	335	0,5854	0			10q26.3	10q26.3
#4110T	chr11	193127	2950365	788	0,5128	-1		11p15.5	11p15.5	11p15.4
#4110T	chr11	2970492	20104639	1597	0,8525	0			11p15.4	11p15.1
#4110T	chr11	20112488	20114695	3	1,1747	1			11p15.1	11p15.1
#4110T	chr11	20117266	20181687	13	0,6625	0			11p15.1	11p15.1
#4110T	chr11	20385408	44255729	747	0,9882	0			11p15.1	11p11.2
#4110T	chr11	44257884	45957239	97	0,6437	0			11p11.2	11p11.2
#4110T	chr11	45958083	46100701	14	0,9916	0			11p11.2	11p11.2
#4110T	chr11	46105743	46760886	118	0,6807	0			11p11.2	11p11.2
#4110T	chr11	46761056	46884227	45	1,0581	0			11p11.2	11p11.2
#4110T	chr11	46886043	66394035	2717	0,7122	0			11p11.2	11q13.2
#4110T	chr11	66407389	66444345	7	1,0282	0			11q13.2	11q13.2
#4110T	chr11	66453459	66512302	37	0,6111	0			11q13.2	11q13.2
#4110T	chr11	66515919	66568537	6	1,1514	0			11q13.2	11q13.2
#4110T	chr11	66571563	71529240	646	0,6322	0			11q13.2	11q13.4
#4110T	chr11	71529866	71701703	12	0,9883	0			11q13.4	11q13.4
#4110T	chr11	71705832	72539979	203	0,6765	0			11q13.4	11q13.4
#4110T	chr11	72540391	72726892	17	1,0229	0			11q13.4	11q13.4
#4110T	chr11	72794763	73371847	51	0,6482	0			11q13.4	11q13.4
#4110T	chr11	73372579	74862435	183	0,9211	0			11q13.4	11q13.4
#4110T	chr11	74873765	75526511	80	0,6361	0			11q13.4	11q13.5
#4110T	chr11	75562987	76730798	58	0,9469	0			11q13.5	11q13.5
#4110T	chr11	76731344	76938935	65	0,6136	0			11q13.5	11q13.5
#4110T	chr11	76940236	120207969	2450	0,9207	0	HMBS		11q13.5	11q23.3
#4110T	chr11	120276839	120310884	11	1,1938	1			11q23.3	11q23.3
#4110T	chr11	120312471	124761326	328	0,8679	0			11q23.3	11q24.2
#4110T	chr11	124761630	124767656	11	0,4675	-1			11q24.2	11q24.2
#4110T	chr11	124789780	133781841	441	0,8495	0			11q24.2	11q25
#4110T	chr11	133790465	133816055	17	0,5592	-1		11q25	11q25	11q25
#4110T	chr11	133826617	134257498	111	0,7966	0			11q25	11q25
#4110T	chr12	176326	112600259	8885	0,8882	0	ARID2,CDKN1B		12p13.33	12q24.13
#4110T	chr12	112600957	112610570	9	0,4597	-1			12q24.13	12q24.13
#4110T	chr12	112613602	112684792	38	0,8602	0			12q24.13	12q24.13
#4110T	chr12	112685310	112696992	10	1,1457	0			12q24.13	12q24.13
#4110T	chr12	112699171	133810818	2029	0,7472	0	HNF1A		12q24.13	12q24.33
#4110T	chr13	19748151	115090537	3346	0,9396	0	RB1		13q12.11	13q34
#4110T	chr14	19378084	71478237	3278	0,8883	0			14q11.2	14q24.2
#4110T	chr14	71479811	71495463	6	1,3736	1			14q24.2	14q24.2
#4110T	chr14	71500201	75276291	387	0,8248	0			14q24.2	14q24.3
#4110T	chr14	75276659	75302041	12	1,1989	1			14q24.3	14q24.3
#4110T	chr14	75321920	103470313	1635	0,8597	0			14q24.3	14q32.32
#4110T	chr14	103474851	103803509	32	0,5102	-1			14q32.32	14q32.32
#4110T	chr14	103804737	104518369	128	0,8092	0			14q32.32	14q32.33
#4110T	chr14	104552149	105995651	264	0,5106	-1			14q32.33	14q32.33
#4110T	chr15	20740212	73585810	4726	0,8692	0			15q11.2	15q24.1
#4110T	chr15	73590821	73660219	10	0,5673	0			15q24.1	15q24.1
#4110T	chr15	73735606	73855587	8	1,2202	1			15q24.1	15q24.1
#4110T	chr15	73862627	82431091	879	0,753	0			15q24.1	15q25.2
#4110T	chr15	82444285	82533677	14	1,1445	0			15q25.2	15q25.2
#4110T	chr15	82545074	90233835	562	0,801	0			15q25.2	15q26.1
#4110T	chr15	90245129	90258269	5	1,1258	0			15q26.1	15q26.1
#4110T	chr15	90260161	90809575	87	0,6636	0			15q26.1	15q26.1
#4110T	chr15	90814720	91147648	46	0,968	0			15q26.1	15q26.1
#4110T	chr15	91150660	91185268	9	0,6909	0			15q26.1	15q26.1
#4110T	chr15	91290672	91354536	20	1,005	0			15q26.1	15q26.1
#4110T	chr15	91358420	91504270	88	0,5901	0			15q26.1	15q26.1
#4110T	chr15	91504924	102462803	379	0,8609	0			15q26.1	15q26.3
#4110T	chr16	97494	23091379	3055	0,6967	0	AXIN1,TSC2		16p13.3	16p12.2
#4110T	chr16	23093819	23117580	9	1,1276	0			16p12.2	16p12.2
#4110T	chr16	23119435	29395491	636	0,7477	0			16p12.2	16p11.2
#4110T	chr16	29396989	29458235	9	0,9927	0			16p11.2	16p11.2
#4110T	chr16	29465003	56782249	1389	0,754	0	BRD7		16p11.2	16q13
#4110T	chr16	56792508	56878466	20	1,0301	0			16q13	16q13
#4110T	chr16	56899288	58550479	403	0,6719	0			16q13	16q21
#4110T	chr16	58550772	58616752	44	1,0048	0			16q21	16q21
#4110T	chr16	58617023	58713896	18	0,7143	0			16q21	16q21
#4110T	chr16	58742136	65038659	32	1,0445	0			16q21	16q21
#4110T	chr16	66413345	77327064	1642	0,7467	0			16q21	16q23.1
#4110T	chr16	77328922	77401479	15	1,1399	0			16q23.1	16q23.1
#4110T	chr16	77465350	90142284	1005	0,6828	0			16q23.1	16q24.3
#4110T	chr17	193127	2344818	404	0,6959	0			17p13.3	17p13.3
#4110T	chr17	2367573	2585060	15	0,9474	0			17p13.3	17p13.3
#4110T	chr17	2593980	2930301	47	0,5842	0			17p13.3	17p13.3
#4110T	chr17	2934254	3427574	32	0,9415	0			17p13.3	17p13.2
#4110T	chr17	3430184	3716456	94	0,6555	0			17p13.2	17p13.2
#4110T	chr17	3717696	3743446	10	1,01	0			17p13.2	17p13.2
#4110T	chr17	3746412	6503739	630	0,7043	0			17p13.2	17p13.1
#4110T	chr17	6510238	6526310	7	1,1509	0			17p13.1	17p13.1
#4110T	chr17	6528128	38545835	4308	0,7619	0	TP53		17p13.1	17q21.2

#4110T	chr17	38546316	38569496	26	1,0852	0			17q21.2	17q21.2
#4110T	chr17	38572288	38721656	23	0,6619	0			17q21.2	17q21.2
#4110T	chr17	38785141	39165178	103	1,0012	0			17q21.2	17q21.2
#4110T	chr17	39183145	43176827	1145	0,7077	0			17q21.2	17q21.31
#4110T	chr17	43180404	43190028	5	0,9945	0			17q21.31	17q21.31
#4110T	chr17	43190306	46926676	494	0,7525	0			17q21.31	17q21.32
#4110T	chr17	46928495	46940284	8	1,1	0			17q21.32	17q21.32
#4110T	chr17	46970799	47376030	65	0,7543	0			17q21.32	17q21.32
#4110T	chr17	47388746	47481620	5	1,0566	0			17q21.32	17q21.33
#4110T	chr17	47482460	74300551	2936	0,7803	0			17q21.33	17q25.1
#4110T	chr17	74300993	74344636	11	1,0662	0			17q25.1	17q25.1
#4110T	chr17	74349699	74621950	64	0,6243	0			17q25.1	17q25.1
#4110T	chr17	74622132	74865146	40	0,9235	0			17q25.1	17q25.2
#4110T	chr17	74868938	76046341	51	0,6844	0			17q25.2	17q25.3
#4110T	chr17	76060911	76098590	16	0,9753	0			17q25.3	17q25.3
#4110T	chr17	76099565	81052161	1098	0,6523	0			17q25.3	17q25.3
#4110T	chr18	158706	78005195	2931	0,9326	0			18p11.32	18q23
#4110T	chr19	111138	4409328	1276	0,5579	-1		STK11	19p13.3	19p13.3
#4110T	chr19	4418045	4428774	5	1,0347	0			19p13.3	19p13.3
#4110T	chr19	4429519	5594045	188	0,5992	0			19p13.3	19p13.3
#4110T	chr19	5595440	5616465	12	0,9603	0			19p13.3	19p13.3
#4110T	chr19	5621363	6680221	285	0,6298	0			19p13.3	19p13.3
#4110T	chr19	6681997	6686217	6	1,0519	0			19p13.3	19p13.3
#4110T	chr19	6686835	9449932	725	0,6572	0			19p13.3	19p13.2
#4110T	chr19	9451819	9643561	23	1,0479	0			19p13.2	19p13.2
#4110T	chr19	9644579	11373094	585	0,6516	0	KEAP1		19p13.2	19p13.2
#4110T	chr19	11407008	11436168	8	0,9701	0			19p13.2	19p13.2
#4110T	chr19	11446185	19767881	2181	0,6553	0			19p13.2	19p13.11
#4110T	chr19	19768181	29698760	143	0,9855	0			19p13.11	19q12
#4110T	chr19	29703919	40030364	1587	0,6742	0			19q12	19q13.2
#4110T	chr19	40093216	40276556	19	0,9691	0			19q13.2	19q13.2
#4110T	chr19	40316540	40487155	57	0,6538	0			19q13.2	19q13.2
#4110T	chr19	40504275	40595168	15	0,9618	0			19q13.2	19q13.2
#4110T	chr19	40698279	44306500	811	0,6579	0			19q13.2	19q13.31
#4110T	chr19	44339700	45016968	87	0,9312	0			19q13.31	19q13.31
#4110T	chr19	45017289	59082559	3222	0,6636	0	SEPW1,TMEM160		19q13.31	19q13.43
#4110T	chr20	68379	21319704	1309	0,8446	0			20p13	20p11.22
#4110T	chr20	21321424	21369944	14	1,2502	1			20p11.22	20p11.22
#4110T	chr20	21376860	30309739	230	0,7142	0			20p11.22	20q11.21
#4110T	chr20	30345333	30382278	13	1,0919	0			20q11.21	20q11.21
#4110T	chr20	30385263	62904763	3214	0,7359	0	DNAJC5		20q11.21	20q13.33
#4110T	chr21	45104483	48083380	583	0,589	0			21q22.3	21q22.3
#4110T	chr21	45103207	45658349	1428	0,8758	0			21q22.3	21q22.3
#4110T	chr22	17443694	36013257	2014	0,7133	0	ZNRF3,CLDN5		22q11.1	22q12.3
#4110T	chr22	17288796	19378084	15	1,0868	0			22q11.21	22q11.1
#4110T	chr22	36052497	36357632	22	1,0228	0			22q12.3	22q12.3
#4110T	chr22	36424380	43204834	1126	0,6885	0			22q12.3	22q13.2
#4110T	chr22	43206884	43243579	11	0,9483	0			22q13.2	22q13.2
#4110T	chr22	43253154	45741395	245	0,6615	0			22q13.2	22q13.31
#4110T	chr22	45745643	45804685	20	1,0054	0			22q13.31	22q13.31
#4110T	chr22	45809394	46068002	28	0,6905	0			22q13.31	22q13.31
#4110T	chr22	46085688	46238964	10	1,049	0			22q13.31	22q13.31
#4110T	chr22	46239550	51220669	598	0,5577	-1		22q13.32	22q13.31	22q13.33
#4110T	chrX	200918	152105010	6116	0,9113	0	RPS6KA3		Xp22.33	Xq28
#4110T	chrX	152106662	153881655	543	0,5757	0			Xq28	Xq28
#4110T	chrX	153906513	155239777	136	0,9255	0			Xq28	Xq28
#4110T	chrY	150918	59342783	128	0,6972	0			Yp11.32	Yq12
#471T	chr1	762224	249213866	17232	-0,0459	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#471T	chr2	40221	242842554	13340	-0,0279	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#471T	chr3	239468	123022972	6746	-0,0396	0	CTNNB1,BAP1,PDE12,SETD2		3p26.3	3q21.1
#471T	chr3	123036915	123071353	8	-0,3889	-1			3q21.1	3q21.1
#471T	chr3	123086820	197769154	4121	-0,0261	0	KLF15		3q21.1	3q29
#471T	chr4	53291	190862085	7182	-0,0164	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#471T	chr5	143266	180687559	7880	-0,0227	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#471T	chr6	203615	170891556	9011	-0,0285	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#471T	chr7	195642	158935187	8044	-0,0501	0	MET		7p22.3	7q36.3
#471T	chr8	183094	146280027	6109	-0,0313	0	SLURP1,MYC		8p23.3	8q24.3
#471T	chr9	213464	139888074	6397	-0,0482	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#471T	chr9	139888322	139905490	10	-0,6406	-1			9q34.3	9q34.3
#471T	chr9	139906382	141017498	287	-0,1484	0			9q34.3	9q34.3
#471T	chr10	226000	135381590	6998	-0,0327	0	PTEN		10p15.3	10q26.3
#471T	chr11	193118	47361271	3164	-0,05	0			11p15.5	11p11.2
#471T	chr11	47362742	47374186	12	-0,4033	-1			11p11.2	11p11.2
#471T	chr11	47376754	61534692	764	-0,0455	0			11p11.2	11q12.2
#471T	chr11	61536762	61543848	6	-0,7842	-1			11q12.2	11q12.2
#471T	chr11	61544836	63978199	482	-0,0794	0			11q12.2	11q13.1
#471T	chr11	63987062	63991393	8	-0,5691	-1			11q13.1	11q13.1
#471T	chr11	63992102	134257636	5131	-0,0578	0	HMBS		11q13.1	11q25
#471T	chr12	208345	133810818	9752	-0,0462	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#471T	chr13	19752446	115091029	3215	-0,0184	0	RB1		13q12.11	13q34
#471T	chr14	20201689	107282930	5490	-0,0407	0			14q11.2	14q32.33
#471T	chr15	20450269	102389376	5760	-0,0452	0			15q11.1	15q26.3
#471T	chr16	97088	90142284	6647	-0,0927	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#471T	chr17	6088	18155828	2559	-0,0873	0	TP53		17p13.3	17p11.2
#471T	chr17	18156693	18166105	9	-0,4422	-1			17p11.2	17p11.2
#471T	chr17	18166613	81052160	6927	-0,0874	0			17p11.2	17q25.3
#471T	chr18	166802	77960704	2756	-0,0174	0			18p11.32	18q23
#471T	chr19	287612	59083958	8466	-0,1424	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#471T	chr20	68379	62906077	4265	-0,0628	0	DNAJC5		20p13	20q13.33
#471T	chr21	9825921	48084620	1783	-0,0481	0	DYRK1A		21p11.2	21q22.3
#471T	chr22	17061701	51179116	3460	-0,0903	0	ZNRF3,CLDN5		22q11.1	22q13.33
#471T	chrX	2700138	154774846	5807	-0,029	0	RPS6KA3		Xp22.33	Xq28
#477T	chr1	69550	52823520	6742	0,8283	0	ARID1A,RPL22		1p36.33	1p32.3
#477T	chr1	52824068	52838950	13	1,0723	0			1p32.3	1p32.3
#477T	chr1	52840526	148252784	4219	0,7832	0	COL11A1,JAK1,NRAS		1p32.3	1q21.2
#477T	chr1	145811947	228407163	7734	0,8161	0			1q21.1	1q42.13
#477T	chr1	145816686	148253639	130	1,1186	1			1q21.2	1q21.1
#477T	chr1	228412297	228459774	11	0,4648	-1			1q42.13	1q42.13
#477T	chr1	228461111	249211681	1308	0,7817	0			1q42.13	1q44
#477T	chr2	41618	110585459	6495	0,799	0	APOB,ALK		2p25.3	2q13
#477T	chr2	110585683	110663463	11	1,13	1			2q13	2q13
#477T	chr2	110843303	242814646	8412	0,7834	0	ACVR2A,NFE2L2,EPHA4		2q13	2q37.3
#477T	chr3	361505	197765488	11634	0,8038	0	CTNNB1,BAP1,KLF15,SETD2,FANCD2,TNFSF10,PIK3CA		3p26.3	3q29
#477T	chr4	53384	113242	7511	0,7768	0			4p16.3	4p16.3

#477T	chr5	92369	180794757	8584	0,787	0	P10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT,APC,MAP1B		5p15.33	5q35.3
#477T	chr6	292550	34495187	2779	0,7471	0			6p25.3	6p21.31
#477T	chr6	34496537	34499471	6	1,2468	1			6p21.31	6p21.31
#477T	chr6	34500253	170893528	6914	0,7887	0	CDKN1A,EEF1A1,FRK,VEGFA		6p21.31	6q27
#477T	chr7	193502	158937438	9157	0,7981	0	MET,HGF,EGFR,KMT2C,BRAF,NUP205		7p22.3	7q36.3
#477T	chr8	116555	146279474	6562	0,79	0	MYC,RSPO2		8p23.3	8q24.3
#477T	chr9	14874	24925	10	1,0987	1			9p24.3	9p24.3
#477T	chr9	117460	19378439	755	0,7801	0			9p24.3	9p22.1
#477T	chr9	19378811	19446348	8	1,1304	1			9p22.1	9p22.1
#477T	chr9	19450541	141016186	7115	0,818	0	CDKN2A,TSC1,NOTCH1,PTPN3		9p22.1	9q34.3
#477T	chr10	93526	135440158	8033	0,7951	0	PTEN		10p15.3	10q26.3
#477T	chr11	193127	134257498	10963	0,812	0	HMBS,FGF19,IGF2,CDKN1C,ATM		11p15.5	11q25
#477T	chr12	176326	133810818	11121	0,7984	0	ARID2,HNF1A,GLI1,KMT2D,KRAS,PTPRB		12p13.33	12q24.33
#477T	chr13	19748151	115090537	3406	0,7849	0	RB1,MYCBP2		13q12.11	13q34
#477T	chr14	19378084	105995651	5838	0,7979	0	DICER1		14q11.2	14q32.33
#477T	chr15	20740212	102462803	6904	0,809	0			15q11.2	15q26.3
#477T	chr16	97494	90142284	8279	0,8274	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#477T	chr17	193127	36104703	4961	0,8151	0	TP53,NF1		17p13.3	17q12
#477T	chr17	36285820	42852092	1715	0,8001	0	STAT3,ERBB2		17q12	17q21.31
#477T	chr17	36285821	36287166	66	1,1272	1			17q12	17q12
#477T	chr17	42852630	42854910	8	1,222	1			17q21.31	17q21.31
#477T	chr17	42855147	81052161	4882	0,8265	0			17q21.31	17q25.3
#477T	chr18	158706	78005195	2978	0,7742	0			18p11.32	18q23
#477T	chr19	111138	59082559	11208	0,8098	0	KEAP1,CCNE1,JAK3,KMT2B,DNAJB1,PRKACA		19p13.3	19q13.43
#477T	chr20	68379	62904763	4777	0,8204	0	GNAS		20p13	20q13.33
#477T	chr21	45658349	45953687	1636	0,7892	0			21q22.3	21q22.3
#477T	chr21	45959609	46117485	17	0,5079	0			21q22.3	21q22.3
#477T	chr21	46131389	48084246	367	0,8129	0			21q22.3	21q22.3
#477T	chr22	19378084	51220669	4061	0,8273	0	ZNRF3,NF2		22q11.21	22q13.33
#477T	chrX	200918	120012716	5067	0,7113	0	RPS6KA3		Xp22.33	Xq24
#477T	chrX	120064205	120119012	24	1,0672	0			Xq24	Xq24
#477T	chrX	120182377	134928918	515	0,7015	0			Xq24	Xq26.3
#477T	chrX	134856765	134930218	28	1,0739	0			Xq26.3	Xq26.3
#477T	chrX	134947481	153421889	817	0,7248	0			Xq26.3	Xq28
#477T	chrX	153424346	153520465	23	1,0386	0			Xq28	Xq28
#477T	chrX	153524279	155239777	342	0,7585	0			Xq28	Xq28
#477T	chrY	150918	59342783	127	0,8571	0			Yp11.32	Yq12
#533T	chr1	69550	9833391	1289	0,6154	0	RPL22		1p36.33	1p36.22
#533T	chr1	9883746	10093740	21	0,88	0			1p36.22	1p36.22
#533T	chr1	10132179	249211681	19068	0,6049	0	ARID1A,COL11A1		1p36.22	1q44
#533T	chr2	41618	242814646	15102	0,5864	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#533T	chr3	361505	197765488	11764	0,5875	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#533T	chr4	53384	113242	7626	0,574	0			4p16.3	4p16.3
#533T	chr5	92369	180794757	8692	0,5912	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#533T	chr6	292550	170893528	9806	0,5894	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#533T	chr7	193502	124482939	6785	0,6152	0	MET		7p22.3	7q31.33
#533T	chr7	124487024	124493109	3	0,9194	1			7q31.33	7q31.33
#533T	chr7	124499088	158937438	2484	0,5885	0			7q31.33	7q36.3
#533T	chr8	116555	146279474	6644	0,5876	0	SLURP1,MYC		8p23.3	8q24.3
#533T	chr9	14874	75407188	2305	0,5853	0	CDKN2A,MTAP		9p24.3	9q21.13
#533T	chr9	75420362	75450878	7	0,2213	-1			9q21.13	9q21.13
#533T	chr9	75516160	141016186	5671	0,6057	0	TSC1,PTPN3		9q21.13	9q34.3
#533T	chr10	93526	18827201	1022	0,6142	0			10p15.3	10p12.31
#533T	chr10	18828406	18874950	5	1,3626	1			10p12.31	10p12.31
#533T	chr10	18885197	135440158	7110	0,5915	0	PTEN		10p12.31	10q26.3
#533T	chr11	193127	47767777	3769	0,5883	0			11p15.5	11p11.2
#533T	chr11	47772528	47808056	11	0,9216	1			11p11.2	11p11.2
#533T	chr11	47809796	51411920	98	0,5947	0			11p11.2	11p11.12
#533T	chr11	55029766	55036766	5	1,1514	1			11q11	11q11
#533T	chr11	55111170	134257498	7188	0,6045	0	HMBS		11q11	11q25
#533T	chr12	176326	133810818	11258	0,6078	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#533T	chr13	19748151	115090537	3468	0,5744	0	RB1		13q12.11	13q34
#533T	chr14	19378084	105995651	5909	0,6029	0			14q11.2	14q32.33
#533T	chr15	20740212	102462803	6998	0,5997	0			15q11.2	15q26.3
#533T	chr16	97494	90142284	8346	0,6399	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#533T	chr17	193127	6544475	1247	0,6579	0			17p13.3	17p13.1
#533T	chr17	6545114	6547883	4	0,2709	-1			17p13.1	17p13.1
#533T	chr17	6553350	36977240	3964	0,6212	0	TP53		17p13.1	17q12
#533T	chr17	36981427	37006671	6	1,0434	1			17q12	17q12
#533T	chr17	37008921	81052161	6524	0,6342	0			17q12	17q25.3
#533T	chr18	158706	43604614	1834	0,5901	0			18p11.32	18q21.1
#533T	chr18	43619963	43708066	23	0,8353	0			18q21.1	18q21.1
#533T	chr18	43796204	78005195	1174	0,5682	0			18q21.1	18q23
#533T	chr19	111138	5729919	1544	0,6878	0			19p13.3	19p13.3
#533T	chr19	5733924	5776249	17	0,9816	1			19p13.3	19p13.3
#533T	chr19	5778537	56163954	9286	0,7011	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.42
#533T	chr19	56166495	56172480	5	0,275	-1			19q13.42	19q13.42
#533T	chr19	56173926	59082559	420	0,6433	0			19q13.42	19q13.43
#533T	chr20	68379	30137895	1565	0,5771	0			20p13	20q11.21
#533T	chr20	30142591	30253821	12	0,8962	1		BCL2L1	20q11.21	20q11.21
#533T	chr20	30309739	62904763	3255	0,6219	0	DNAJC5		20q11.21	20q13.33
#533T	chr21	45658349	48084246	2054	0,6027	0			21q22.3	21q22.3
#533T	chr22	19378084	51220669	4110	0,6304	0	ZNRF3,CLDN5		22q11.21	22q13.33
#533T	chrX	200918	48830651	2036	0,5996	0	RPS6KA3		Xp22.33	Xp11.23
#533T	chrX	48831644	48844572	14	0,8847	0			Xp11.23	Xp11.23
#533T	chrX	48846072	155239777	4552	0,585	0			Xp11.23	Xq28
#533T	chrY	150918	59342783	555	0,5867	0			Yp11.32	Yq12
#623T	chr1	717420	33134894	3196	0,2244	0	ARID1A,RPL22		1p36.33	1p35.1
#623T	chr1	33135125	33147455	7	-0,1143	0			1p35.1	1p35.1
#623T	chr1	33149635	159860325	8131	0,143	0	COL11A1		1p35.1	1q23.2
#623T	chr1	159863017	159921553	17	0,443	1			1q23.2	1q23.2
#623T	chr1	159922172	249149602	5985	0,1362	0			1q23.2	1q44
#623T	chr1	249149795	249211681	7	0,449	1			1q44	1q44
#623T	chr2	41602	243037096	13813	0,122	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#623T	chr3	361504	197847544	10647	0,1384	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#623T	chr4	53383	190946966	7234	0,09	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#623T	chr5	143251	180687764	7921	0,1218	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#623T	chr6	140294	170892713	9130	0,1321	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#623T	chr7	195642	158935187	8092	0,143	0	MET		7p22.3	7q36.3
#623T	chr8	163585	146279464	5762	0,118	0	SLURP1,MYC		8p23.3	8q24.3
#623T	chr9	117444	141071146	6753	0,1642	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#623T	chr10	93566	32751620	1625	0,1216	0			10p15.3	10p11.22
#623T	chr10	32751960	32974952	15	-0,1993	-1			10p11.22	10p11.22
#623T	chr10	32978035	37438592	133	0,0735	0			10p11.22	10p11.21

#623T	chr10	37438740	37488685	15	-0,1988	-1			10p11.21	10p11.21
#623T	chr10	37490207	135491232	5505	0,1478	0	PTEN		10p11.21	10q26.3
#623T	chr11	193811	45992792	2661	0,1654	0			11p15.5	11p11.2
#623T	chr11	46001274	46105728	5	-0,2296	-1			11p11.2	11p11.2
#623T	chr11	46321599	63398727	1504	0,1816	0			11p11.2	11q12.3
#623T	chr11	63400533	63419432	7	-0,1117	0			11q13.1	11q13.1
#623T	chr11	63419979	134320599	4835	0,1759	0	HMBS		11q13.1	11q25
#623T	chr12	68386	110565902	8045	0,136	0	ARID2,CDKN1B		12p13.33	12q24.11
#623T	chr12	110566844	110646965	15	-0,1248	0			12q24.11	12q24.11
#623T	chr12	110655939	133779065	1929	0,1994	0	HNF1A		12q24.11	12q24.33
#623T	chr13	19042028	115090536	3419	0,1	0	RB1		13q11	13q34
#623T	chr14	19110319	106236157	5549	0,1365	0			14q11.2	14q32.33
#623T	chr14	106236463	106237595	4	-0,6126	-1			14q32.33	14q32.33
#623T	chr14	106304722	107283203	120	0,211	0			14q32.33	14q32.33
#623T	chr15	20083862	28499559	337	0,1527	0			15q11.1	15q13.1
#623T	chr15	28501214	28518094	11	0,4644	1			15q13.1	15q13.1
#623T	chr15	28520085	63579683	2958	0,1318	0			15q13.1	15q22.2
#623T	chr15	63594607	63631080	6	0,4247	0			15q22.2	15q22.2
#623T	chr15	63632565	79088963	1403	0,162	0			15q22.2	15q25.1
#623T	chr15	79170577	79189366	12	-0,1472	0			15q25.1	15q25.1
#623T	chr15	79215397	102389374	1240	0,1916	0			15q25.1	15q26.3
#623T	chr16	81841	1889347	338	0,2741	0	AXIN1		16p13.3	16p13.3
#623T	chr16	1891897	1918135	10	-0,0623	0			16p13.3	16p13.3
#623T	chr16	1927389	55532265	3108	0,1946	0	TSC2,BRD7		16p13.3	16q12.2
#623T	chr16	55536730	55616917	16	-0,0443	0			16q12.2	16q12.2
#623T	chr16	55690728	90162029	2597	0,2124	0			16q12.2	16q24.3
#623T	chr17	11293	81052160	9104	0,1986	0	TP53		17p13.3	17q25.3
#623T	chr18	47960	77960703	2811	0,1097	0			18p11.32	18q23
#623T	chr19	182713	59093463	7255	0,2314	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#623T	chr20	68379	5528386	488	0,2229	0			20p13	20p12.3
#623T	chr20	5538670	5560715	13	-0,0369	0			20p12.3	20p12.3
#623T	chr20	5564947	62934540	3532	0,1887	0	DNAJC5		20p12.3	20q13.33
#623T	chr21	9483381	48084222	1822	0,1612	0	DYRK1A		21p11.2	21q22.3
#623T	chr22	16101435	51207918	3186	0,2238	0	ZNRF3,CLDN5		22q11.1	22q13.33
#623T	chrX	2700137	154774831	5817	0,1301	0	RPS6KA3		Xp22.33	Xq28
#623T	chrY	10011790	28584969	7	0,0468	0			Yp11.2	Yq11.23
#750T	chr1	721657	31418261	3652	0,1409	0	ARID1A,RPL22		1p36.33	1p35.2
#750T	chr1	31423042	31528222	16	0,4492	0			1p35.2	1p35.2
#750T	chr1	31532236	36039088	559	0,1703	0			1p35.2	1p34.3
#750T	chr1	36040575	36060460	6	-0,1775	-1			1p34.3	1p34.3
#750T	chr1	36068620	150483520	5910	0,2675	0	COL11A1		1p34.3	1q21.3
#750T	chr1	150484120	150532308	18	-0,0668	-1			1q21.3	1q21.3
#750T	chr1	150532973	249212472	7718	0,2503	0			1q21.3	1q44
#750T	chr2	40221	241870446	13193	0,2757	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#750T	chr2	241871823	241905587	9	-0,2731	-1		2q37.3	2q37.3	2q37.3
#750T	chr2	241922371	242946533	242	0,1197	0			2q37.3	2q37.3
#750T	chr3	239470	38524787	1924	0,2569	0			3p26.3	3p22.2
#750T	chr3	38537892	38548402	5	0,6099	1			3p22.2	3p22.2
#750T	chr3	38564685	196730870	8918	0,2532	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p22.2	3q29
#750T	chr3	196733513	196754704	12	-0,0767	-1			3q29	3q29
#750T	chr3	196770516	197769162	111	0,3065	0			3q29	3q29
#750T	chr4	53291	190947577	6954	0,3118	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#750T	chr5	140602	141380964	5590	0,3058	0	MEF2C,IL6ST,ADAMTS19,TERT		5p15.33	5q31.3
#750T	chr5	141382740	141385800	3	-0,2313	-1			5q31.3	5q31.3
#750T	chr5	141387414	156723730	956	0,2472	0			5q31.3	5q33.3
#750T	chr5	156729850	156738730	5	-0,1781	-1			5q33.3	5q33.3
#750T	chr5	156741411	180687559	1362	0,2118	0	ATP10B,DOCK2		5q33.3	5q35.3
#750T	chr6	203615	170892714	9178	0,2715	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q27
#750T	chr7	195642	65422479	2835	0,2337	0			7p22.3	7q11.21
#750T	chr7	65425861	65557032	19	-0,1837	-1			7q11.21	7q11.21
#750T	chr7	65557596	158937319	5662	0,2449	0	MET		7q11.21	7q36.3
#750T	chr8	190866	37641778	1721	0,2522	0			8p23.3	8p11.23
#750T	chr8	37672449	37700667	14	-0,1677	-1			8p11.23	8p11.23
#750T	chr8	37703248	38270536	110	0,2715	0			8p11.23	8p11.23
#750T	chr8	38271492	38275818	7	-0,107	-1			8p11.23	8p11.23
#750T	chr8	38277152	142186774	3602	0,3172	0	MYC		8p11.23	8q24.3
#750T	chr8	142190899	146280815	675	0,04	0	SLURP1		8q24.3	8q24.3
#750T	chr9	214910	141017498	7117	0,2148	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#750T	chr10	226000	88277586	3671	0,2598	0			10p15.3	10q23.2
#750T	chr10	88414530	88452310	19	-0,0512	-1			10q23.2	10q23.2
#750T	chr10	88459415	99639482	1006	0,2795	0	PTEN		10q23.2	10q24.2
#750T	chr10	99640077	99656706	5	-0,2509	-1			10q24.2	10q24.2
#750T	chr10	99661328	135439464	2522	0,2363	0			10q24.2	10q26.3
#750T	chr11	193118	63755839	4683	0,2094	0			11p15.5	11q13.1
#750T	chr11	63756162	63919814	10	-0,1101	-1			11q13.1	11q13.1
#750T	chr11	63953094	134856548	5522	0,2109	0	HMBS		11q13.1	11q25
#750T	chr12	208345	49920264	3614	0,2742	0	ARID2,CDKN1B		12p13.33	12q13.12
#750T	chr12	49934798	49950206	10	-0,2037	-1			12q13.12	12q13.12
#750T	chr12	49950430	54792431	963	0,1883	0			12q13.12	12q13.13
#750T	chr12	54793485	54796038	8	-0,1375	-1			12q13.13	12q13.13
#750T	chr12	54796800	133196073	5835	0,252	0	HNF1A		12q13.13	12q24.33
#750T	chr12	133196300	133197894	6	-0,3137	-1		12q24.33	12q24.33	12q24.33
#750T	chr12	133198515	133811687	159	0,1572	0			12q24.33	12q24.33
#750T	chr13	19600450	115091029	3095	0,3087	0	RB1		13q12.11	13q34
#750T	chr14	20201689	107283203	5961	0,2461	0			14q11.2	14q32.33
#750T	chr15	20170066	102389376	6299	0,2395	0			15q11.1	15q26.3
#750T	chr16	97096	50324469	3989	0,1477	0	AXIN1,TSC2		16p13.3	16q12.1
#750T	chr16	50326661	50338852	7	-0,193	-1			16q12.1	16q12.1
#750T	chr16	50339455	90161144	3409	0,1705	0	BRD7		16q12.1	16q24.3
#750T	chr17	6088	81052160	10417	0,1698	0	TP53		17p13.3	17q25.3
#750T	chr18	117194	77941018	2745	0,2998	0			18p11.32	18q23
#750T	chr19	107510	59083958	9626	0,1109	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#750T	chr20	68379	62906077	4528	0,188	0	DNAJC5		20p13	20q13.33
#750T	chr21	9825922	48084620	1859	0,2181	0	DYRK1A		21p11.2	21q22.3
#750T	chr22	17060842	24628888	915	0,1087	0	CLDN5		22q11.1	22q11.23
#750T	chr22	24698258	24761522	11	0,4377	0			22q11.23	22q11.23
#750T	chr22	24765236	45732250	2375	0,1449	0	ZNRF3		22q11.23	22q13.31
#750T	chr22	45736512	45804684	20	0,4072	0			22q13.31	22q13.31
#750T	chr22	45809418	51219140	576	0,0671	0			22q13.31	22q13.33
#750T	chrX	2700138	49098548	1972	0,2715	0	RPS6KA3		Xp22.33	Xp11.23
#750T	chrX	49099404	49105675	9	-0,2016	-1			Xp11.23	Xp11.23
#750T	chrX	49107417	51640936	112	0,2726	0			Xp11.23	Xp11.22
#750T	chrX	51641249	52677378	6	-0,0809	-1			Xp11.22	Xp11.22
#750T	chrX	52681349	154774846	3866	0,2574	0			Xp11.22	Xq28

#757T	chr1	69550	249211681	20304	0,6618	0	ARID1A,RPL22,COL11A1,JAK1,NRAS		1p36.33	1q44
#757T	chr2	41618	242814646	15037	0,6698	0	ACVR2A,NFE2L2,EPHA4,APOB,ALK		2p25.3	2q37.3
#757T	chr3	361505	197765488	11714	0,6711	0	CTNNB1,BAP1,KLF15,SETD2,FANCD2,TNFSF10,PIK3CA		3p26.3	3q29
#757T	chr4	53384	113242	7594	0,6696	0			4p16.3	4p16.3
#757T	chr5	92369	180794757	8649	0,6691	0	P10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT,APC,MAP1B		5p15.33	5q35.3
#757T	chr6	292550	117150079	7094	0,6552	0	CDKN1A,EEF1A1,FRK,VEGFA		6p25.3	6q22.1
#757T	chr6	117198550	117593637	23	0,2028	-1			6q22.1	6q22.1
#757T	chr6	117609810	117645537	10	0,5867	0			6q22.1	6q22.1
#757T	chr6	117647482	117725517	29	0,1954	-1			6q22.1	6q22.1
#757T	chr6	117730775	170893528	2613	0,6668	0			6q22.1	6q27
#757T	chr7	193502	158937438	9235	0,6615	0	MET,HGF,EGFR,KMT2C,BRAF,NUP205		7p22.3	7q36.3
#757T	chr8	116555	146279474	6612	0,6679	0	MYC,RSPO2		8p23.3	8q24.3
#757T	chr9	14874	141016186	7946	0,6606	0	CDKN2A,TSC1,NOTCH1,PTPN3		9p24.3	9q34.3
#757T	chr10	93526	135370294	8093	0,667	0	PTEN		10p15.3	10q26.3
#757T	chr10	135370616	135373626	6	1,0887	2			10q26.3	10q26.3
#757T	chr10	135378997	135440158	5	0,6815	0			10q26.3	10q26.3
#757T	chr11	193127	134257498	11028	0,656	0	HMBS,FGF19,IGF2,CDKN1C,ATM		11p15.5	11q25
#757T	chr12	176326	133810818	11210	0,6667	0	ARID2,HNF1A,GLI1,KMT2D,KRAS,PTPRB		12p13.33	12q24.33
#757T	chr13	19748151	115090537	3449	0,6716	0	RB1,MYCBP2		13q12.11	13q34
#757T	chr14	19378084	104433107	5584	0,6596	0	DICER1		14q11.2	14q32.33
#757T	chr14	104436918	104460898	6	1,1162	2			14q32.33	14q32.33
#757T	chr14	104462115	105995651	288	0,6257	0			14q32.33	14q32.33
#757T	chr15	20740212	102462803	6943	0,6637	0			15q11.2	15q26.3
#757T	chr16	97494	90142284	8326	0,6511	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#757T	chr17	193127	81052161	11692	0,6561	0	TP53,STAT3,ERBB2,NF1		17p13.3	17q25.3
#757T	chr18	158706	78005195	3022	0,6676	0			18p11.32	18q23
#757T	chr19	111138	59082559	11244	0,6383	0	KEAP1,CCNE1,JAK3,KMT2B,DNAJB1,PRKACA		19p13.3	19q13.43
#757T	chr20	68379	62904763	4813	1,01	1	GNAS	BCL2L1,ZNF217,20q13.33	20p13	20q13.33
#757T	chr21	45658349	48084246	2031	0,6508	0			21q22.3	21q22.3
#757T	chr22	19378084	51220669	4095	0,6493	0	ZNRF3,NF2		22q11.21	22q13.33
#757T	chrX	200918	155239777	6822	0,6001	0	RPS6KA3		Xp22.33	Xq28
#757T	chrY	150918	59252516	118	0,4603	0			Yp11.32	Yq12
#757T	chrY	59272417	59342783	12	0,7472	0			Yq12	Yq12
#835T	chr1	861348	249212472	19263	-1,0751	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#835T	chr2	40221	242842554	14692	-1,0726	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#835T	chr3	239470	197769162	11873	-1,0763	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#835T	chr4	53291	190947044	7735	-1,0695	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#835T	chr5	140602	180687559	8651	-1,0709	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#835T	chr6	203615	68598996	5143	-1,0792	0	CDKN1A,SPATS1,VEGFA		6p25.3	6q12
#835T	chr6	69348939	93967210	1151	-1,5062	-1	EEF1A1		6q12	6q16.1
#835T	chr6	93967909	167608890	3440	-1,0738	0			6q16.1	6q27
#835T	chr6	167704934	168070076	16	-1,5068	-1			6q27	6q27
#835T	chr6	168187893	168764159	71	-2,0972	-1			6q27	6q27
#835T	chr6	168910680	170866070	120	-1,5087	-1			6q27	6q27
#835T	chr6	170871100	170892714	11	-1,0903	0			6q27	6q27
#835T	chr7	193386	158937319	9196	-1,0631	0	MET		7p22.3	7q36.3
#835T	chr8	190866	146280815	6700	-1,0766	0	SLURP1,MYC		8p23.3	8q24.3
#835T	chr9	213465	141070232	7640	-1,0757	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#835T	chr10	226000	135438930	7841	-1,0726	0	PTEN		10p15.3	10q26.3
#835T	chr11	193118	134856548	10894	-1,0807	0	HMBS		11p15.5	11q25
#835T	chr12	208345	133811687	11410	-1,0865	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#835T	chr13	19600450	115091029	3456	-1,0658	0	RB1		13q12.11	13q34
#835T	chr14	20201689	107283203	6477	-1,0745	0			14q11.2	14q32.33
#835T	chr15	20170066	102389376	6766	-1,0867	0			15q11.1	15q26.3
#835T	chr16	96492	90161144	7793	-1,0816	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#835T	chr17	6088	81052160	10955	-1,0804	0	TP53		17p13.3	17q25.3
#835T	chr18	117194	77960704	3041	-1,0592	0			18p11.32	18q23
#835T	chr19	281462	59083958	10169	-1,0731	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#835T	chr20	68379	62906077	4809	-1,0721	0	DNAJC5		20p13	20q13.33
#835T	chr21	9825922	48084620	2008	-1,0665	0	DYRK1A		21p11.2	21q22.3
#835T	chr22	17060842	51219140	4075	-1,0738	0	ZNRF3,CLDN5		22q11.1	22q13.33
#835T	chrX	2700138	154774846	6399	-1,069	0	RPS6KA3		Xp22.33	Xq28
#873T	chr1	762325	249212472	18763	-0,2675	0	ARID1A,RPL22,COL11A1		1p36.33	1q44
#873T	chr2	40221	242842554	14284	-0,2698	0	ACVR2A,NFE2L2,EPHA4,APOB		2p25.3	2q37.3
#873T	chr3	239470	197769162	11615	-0,2735	0	CTNNB1,BAP1,PDE12,KLF15,SETD2		3p26.3	3q29
#873T	chr4	53291	190947044	7541	-0,2773	0	ALB,FGA,GABRA2,IRF2		4p16.3	4q35.2
#873T	chr5	140602	180687559	8421	-0,2731	0	ATP10B,MEF2C,IL6ST,DOCK2,ADAMTS19,TERT		5p15.33	5q35.3
#873T	chr6	203615	92231416	6118	-0,2706	0	CDKN1A,EEF1A1,SPATS1,VEGFA		6p25.3	6q15
#873T	chr6	93951498	105736700	269	-0,6049	-1			6q16.1	6q21
#873T	chr6	105770558	170892714	3314	-0,2715	0			6q21	6q27
#873T	chr7	195642	158937319	8937	-0,2685	0	MET		7p22.3	7q36.3
#873T	chr8	190866	146280815	6533	-0,2689	0	SLURP1,MYC		8p23.3	8q24.3
#873T	chr9	214910	141017498	7440	-0,2648	0	CDKN2A,TSC1,MTAP,PTPN3		9p24.3	9q34.3
#873T	chr10	226000	135438930	7635	-0,2674	0	PTEN		10p15.3	10q26.3
#873T	chr11	193118	134856548	10613	-0,2642	0	HMBS		11p15.5	11q25
#873T	chr12	208345	133811687	11140	-0,2715	0	ARID2,HNF1A,CDKN1B		12p13.33	12q24.33
#873T	chr13	19600450	115091029	3375	-0,2767	0	RB1		13q12.11	13q34
#873T	chr14	20201689	107283203	6306	-0,2738	0			14q11.2	14q32.33
#873T	chr15	20170066	102389376	6593	-0,2661	0			15q11.1	15q26.3
#873T	chr16	96492	90142284	7558	-0,2577	0	AXIN1,TSC2,BRD7		16p13.3	16q24.3
#873T	chr17	6088	81052160	10619	-0,2568	0	TP53		17p13.3	17q25.3
#873T	chr18	117194	77960704	2956	-0,2703	0			18p11.32	18q23
#873T	chr19	281462	59083958	9716	-0,2455	0	KEAP1,SEPW1,TMEM160		19p13.3	19q13.43
#873T	chr20	68379	29652196	1447	-0,2639	0			20p13	20q11.21
#873T	chr20	29845513	62906077	3240	0,008	0	DNAJC5		20q11.21	20q13.33
#873T	chr21	9825922	22912907	167	-0,254	0			21p11.2	21q21.1
#873T	chr21	26946324	48084620	1778	-0,607	-1	DYRK1A	RG,DSCAM,TMPRSS2,PTTG1	21q21.3	21q22.3
#873T	chr22	17060842	51219140	3935	-0,2511	0	ZNRF3,CLDN5		22q11.1	22q13.33
#873T	chrX	2700138	154774846	4699	-0,2802	0	RPS6KA3		Xp22.33	Xq28
#873T	chrY	2655318	23762108	150	-0,3081	0			Yp11.31	Yq11.223

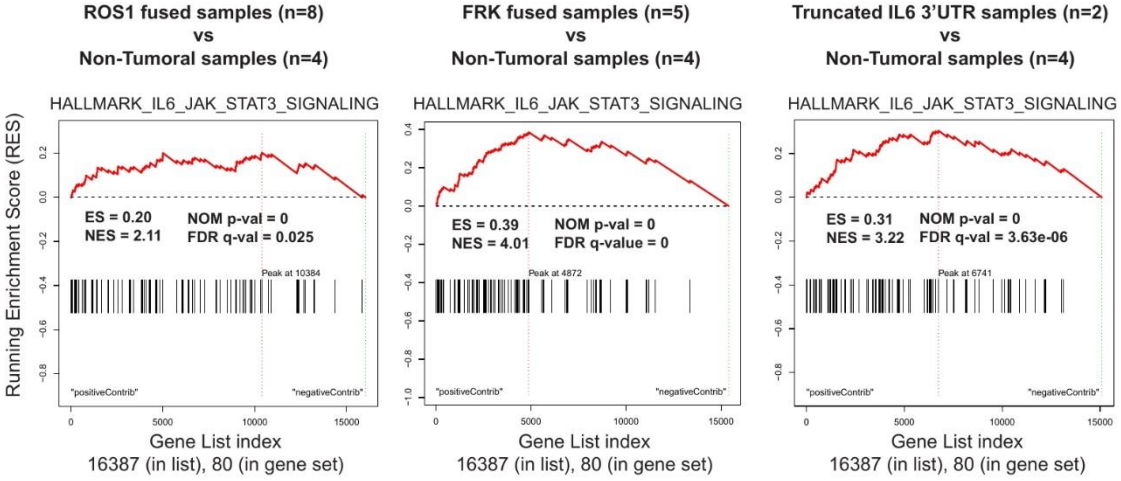
Abbreviations:
LL=Gain, Normal, Loss
LRR=LogR ration

Supplementary Table 6: *ROS1* and *FRK* fusions identified from RNAseq

Sample	Diagnosis	Fus_type	5prim_gene	chr1	pos1_hg38	3prim_gene	chr2	pos2_hg38	Supporting_reads	Oncofuse_InFrame	Oncofuse_DRIVER_PROB	Oncofuse_3_DOMAINS_RETAINED	Oncofuse_3_DOMAINS_BROKEN
#1439T	HCA	TRANS	PLEKHA5	chr12	19348520	FRK	chr6	115968739	11	Yes	0.99928	Tyrosine-protein kinase, active site[Active_site]	SH2 domain[Domain]
#835T	HCA	TRANS	MIA3	chr1	222660315	FRK	chr6	115968739	58	Yes	0.99928	Tyrosine-protein kinase, active site[Active_site]	SH2 domain[Domain]
#222T	HCA	TRANS	MIA2	chr14	39327023	FRK	chr6	115968739	14	Yes	0.981333	Tyrosine-protein kinase, active site[Active_site]	SH2 domain[Domain]
#873T	HCA	TRANS	LMO7	chr13	75853310	FRK	chr6	115968739	25	Yes	0.99928	Tyrosine-protein kinase, active site[Active_site]	SH2 domain[Domain]
#471T	HCA	TRANS	SEC16B	chr1	177936297	FRK	chr6	116060149	9	Yes	0.946629	Tyrosine-protein kinase, active site[Active_site]	Src homology-3 domain[Domain]
#1343T	HCA	TRANS	RBP4	chr10	93600392	ROS1	chr6	117321394	60	Yes	0.911829	Tyrosine-protein kinase, active site[Active_site]	
#1343T	HCA	TRANS	RBP4	chr10	93600392	ROS1	chr6	117324415	205	Yes	0.911829	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]
#375T	HCA	INV	PLG	chr6	160741418	ROS1	chr6	117324415	284	Yes	0.999594	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]
#2729T	HCA	INV	PLG	chr6	160718839	ROS1	chr6	117324415	21	Yes	0.996059	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]
#477T	HCA	TRANS	RBP4	chr10	93593822	ROS1	chr6	117321394	2	Yes	0.911829	Tyrosine-protein kinase, active site[Active_site]	
#744T	HCA	INV	PLG	chr6	160741418	ROS1	chr6	117324415	360	Yes	0.999594	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]
#533T	HCA	TRANS	APOB	chr2	21040937	ROS1	chr6	117320030	500	Yes	0.911829	Tyrosine-protein kinase, active site[Active_site]	
#533T	HCA	TRANS	APOB	chr2	21040937	ROS1	chr6	117326414	414	Yes	0.911829	Tyrosine-protein kinase, active site[Active_site]	Immunoglobulin-like fold[Domain];Fibronectin,
#4110T	HCA	INV	PLG	chr6	160741418	ROS1	chr6	117324415	147	Yes	0.999594	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]
#757T	HCA	INV	PLG	chr6	160741418	ROS1	chr6	117324415	21	Yes	0.999594	Tyrosine-protein kinase, active site[Active_site]	Fibronectin, type III[Domain]

Abbreviations:
HCA=Hepatocellular Adenoma
FRK=Fyn Related Src Family Tyrosine Kinase
ROS1=ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
PLEKHA5=Pleckstrin Homology Domain Containing A5
MIA2=Melanoma Inhibitory Activity 2
MIA3=Melanoma Inhibitory Activity 3
LMO7=LIM Domain 7
SEC16B=SEC16 Homolog B, Endoplasmic Reticulum Export Factor
RBP4=Retinol Binding Protein 4
PLG= Plasminogen
APOB=Apolipoprotein B
TRANS=Translocation
INV=Inversion

Supplementary Figure 3



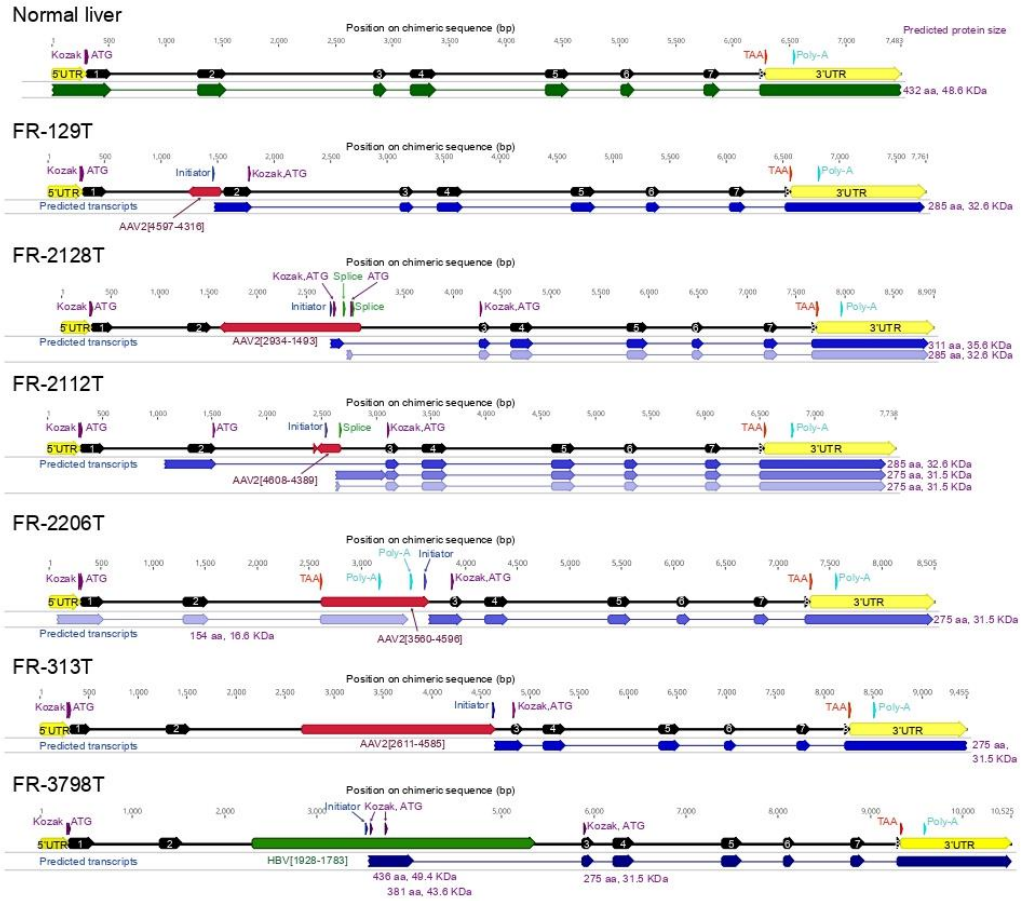
2. Annex 2: Supplementary data Article 4

Supplementary Information

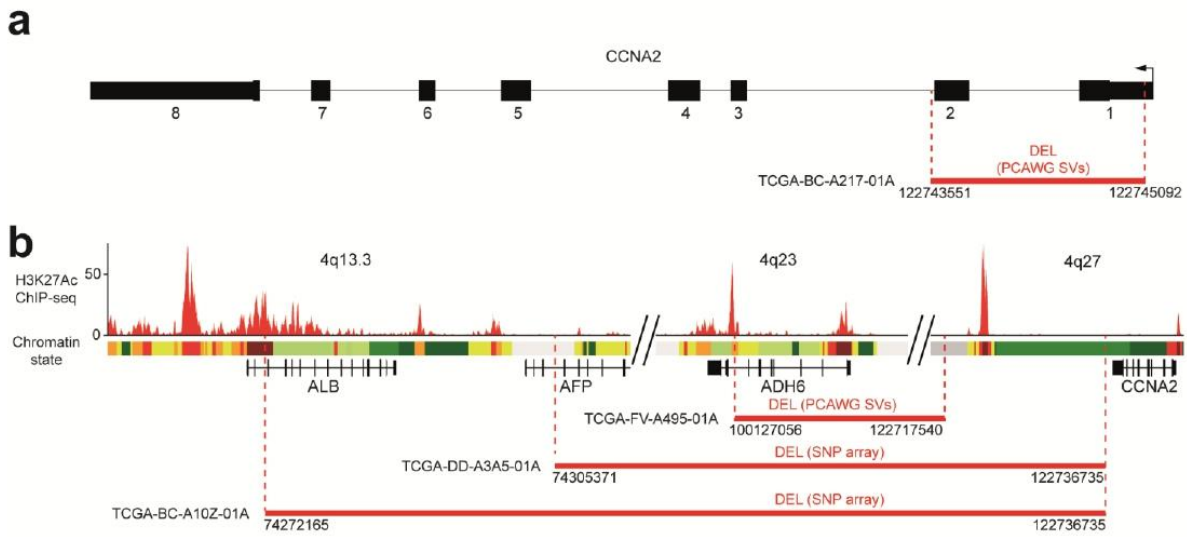
Cyclin A2 and E1 genomic alterations define a specific subclass of hepatocellular carcinomas

Bayard *et al.*

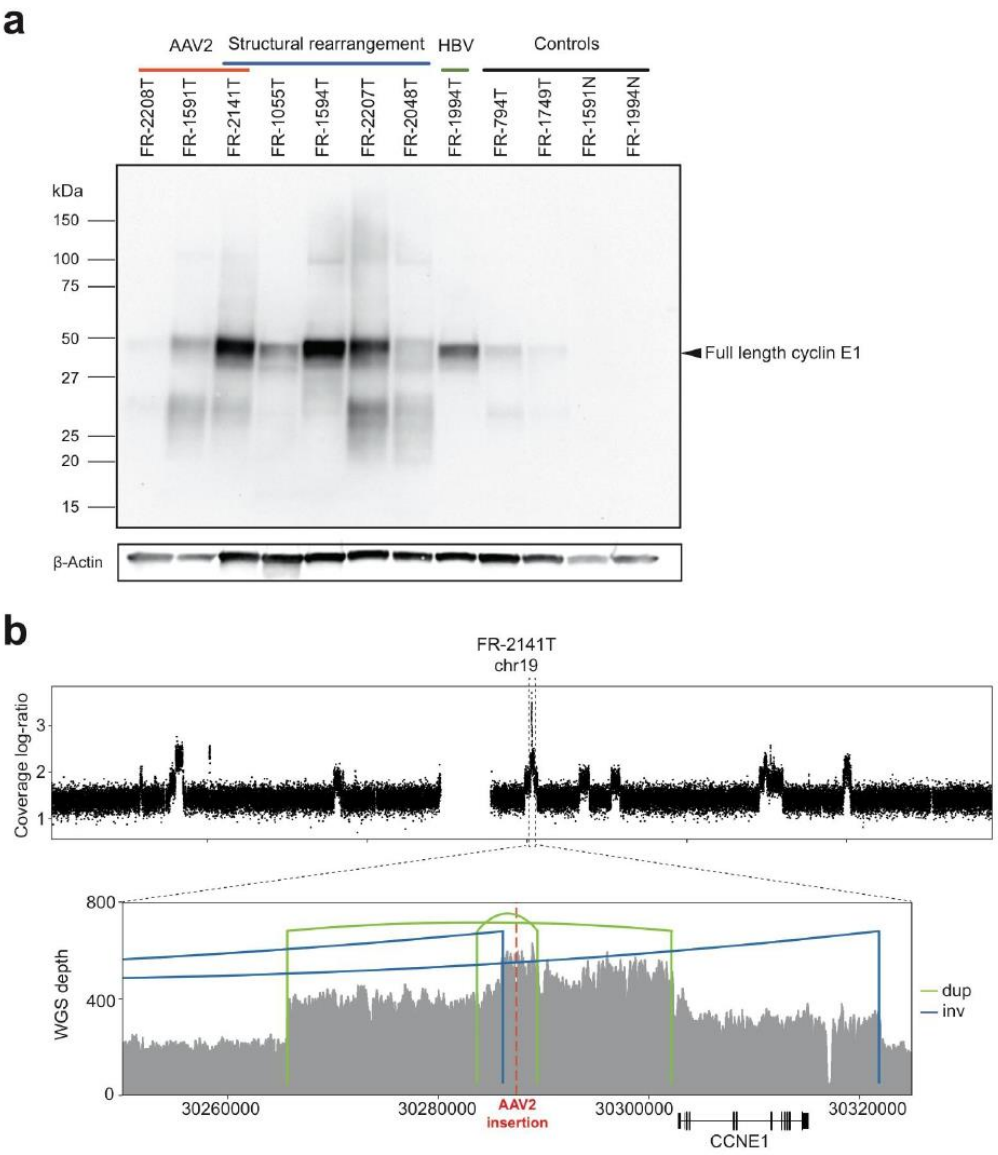
Supplementary Figures



Supplementary Figure 1. Consequences of viral insertions in *CCNA2* gene
Five AAV2 and 1 HBV insertions were identified in *CCNA2* in the LICA-FR series. Precise insertion boundaries were identified by WGS or viral capture, and RNA-seq reads were aligned on the chimeric sequence. Here, the different transcripts predicted by Cufflinks are represented for each case, ordered by transcript abundance. Predicted functional elements (Transcription initiator, splice and poly-A sites, Kozak sequence, initiator and terminator codons) are annotated on the chimeric DNA sequence, and the predicted protein sizes resulting from the translation of each transcript are annotated on the right. Only the most abundant abnormal transcripts were represented in **Fig. 1c**.



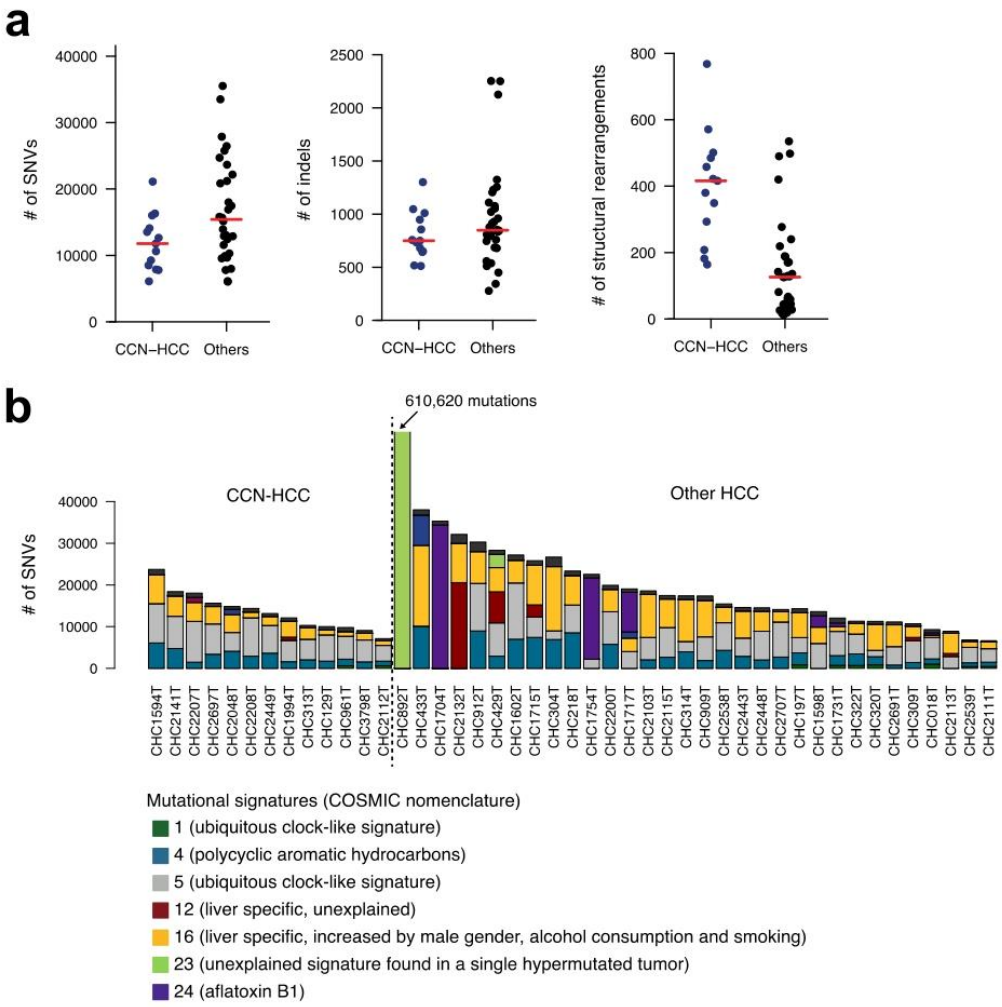
Supplementary Figure 2. Deletions associated with *CCNA2* deregulation
a Focal deletion of *CCNA2* exons 1 & 2 identified in TCGA tumor TCGA-BC-A217.
b Deletions identified in 3 TCGA tumors linking *CCNA2* downstream region with the highly expressed genes *ALB*, *AFP* and *ADH6*.



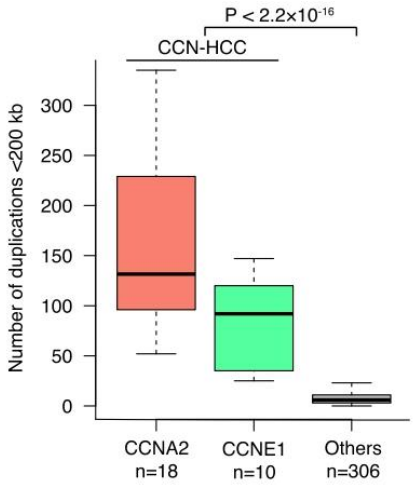
Supplementary Figure 3. Cyclin E1 overexpression induced by viral insertions and structural rearrangements

a Western blot analysis of cyclin E1. Tumors with viral insertions or structural rearrangements are compared with tumors without *CCNE1* alteration and non-tumoral liver controls.

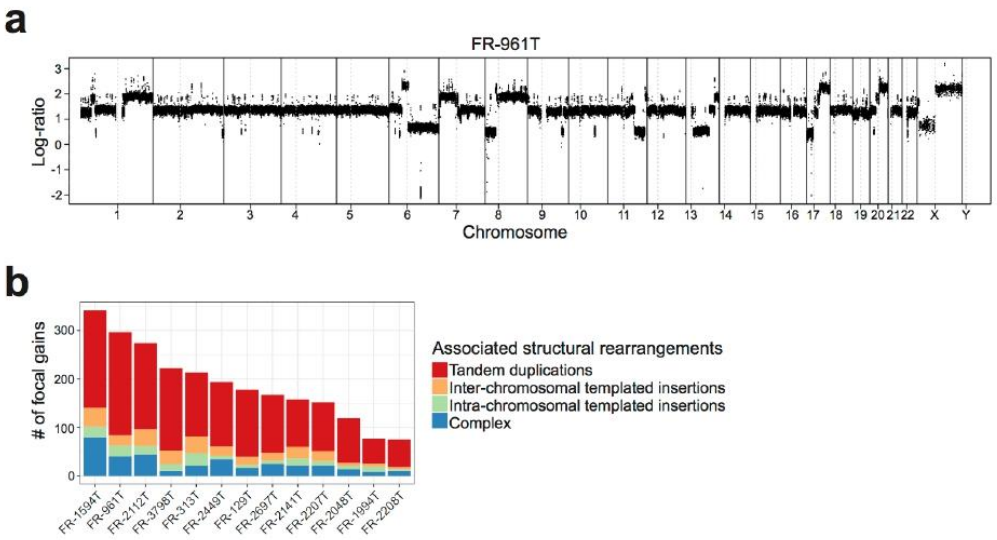
b Tumor FR-2141T displays both an AAV2 insertion in *CCNE1* regulatory region and a high-level amplification of the locus. The top panel displays the coverage log-ratio along chromosome 19 in this tumor. The bottom panel displays the coverage of WGS reads aligned to the chimeric sequence of *CCNE1* locus including AAV2 insertion together with structural rearrangement breakpoints. It shows that the most strongly amplified region includes *CCNE1* regulatory region, and in particular the locus of AAV2 insertion.



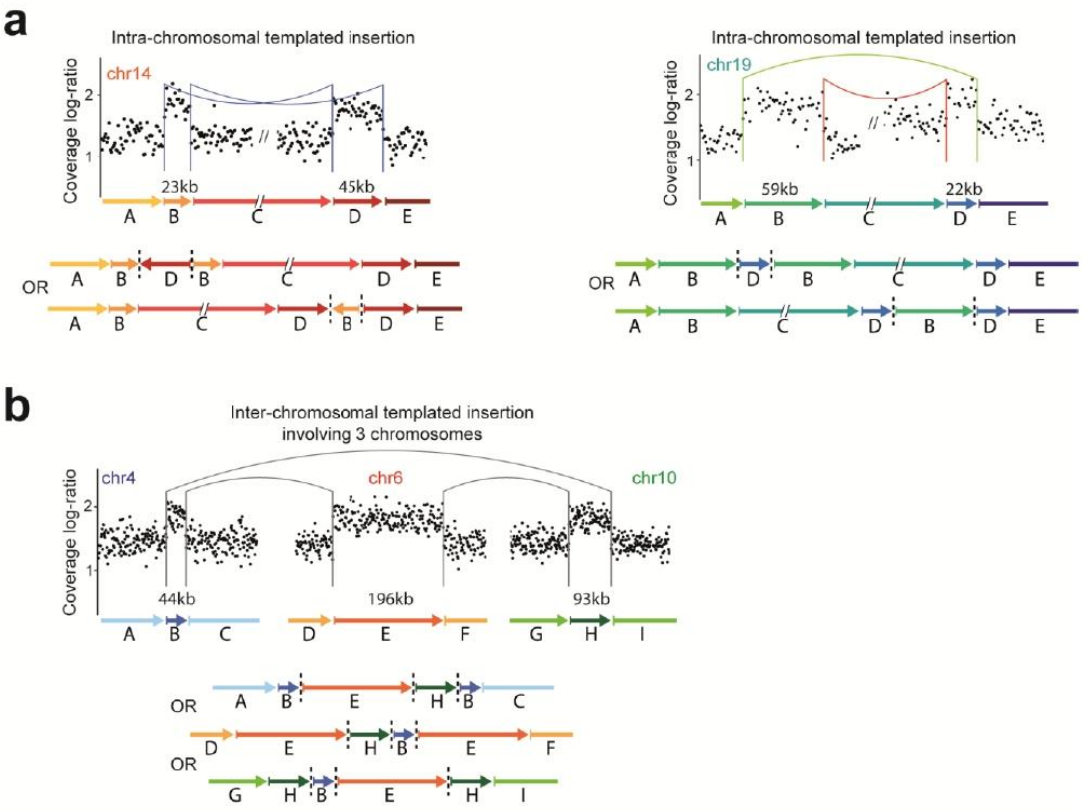
Supplementary Figure 4. Mutational signature analysis of CCN-HCC
a Comparison of the number of single-nucleotide variants (SNVs), indels and structural rearrangements in CCN-HCC vs others (LICA-FR data, 45 WGS). Red line represent median.
b Contribution of the different mutational signatures known to be operative in liver cancers to the mutational burden of CCN-HCC and other HCC.



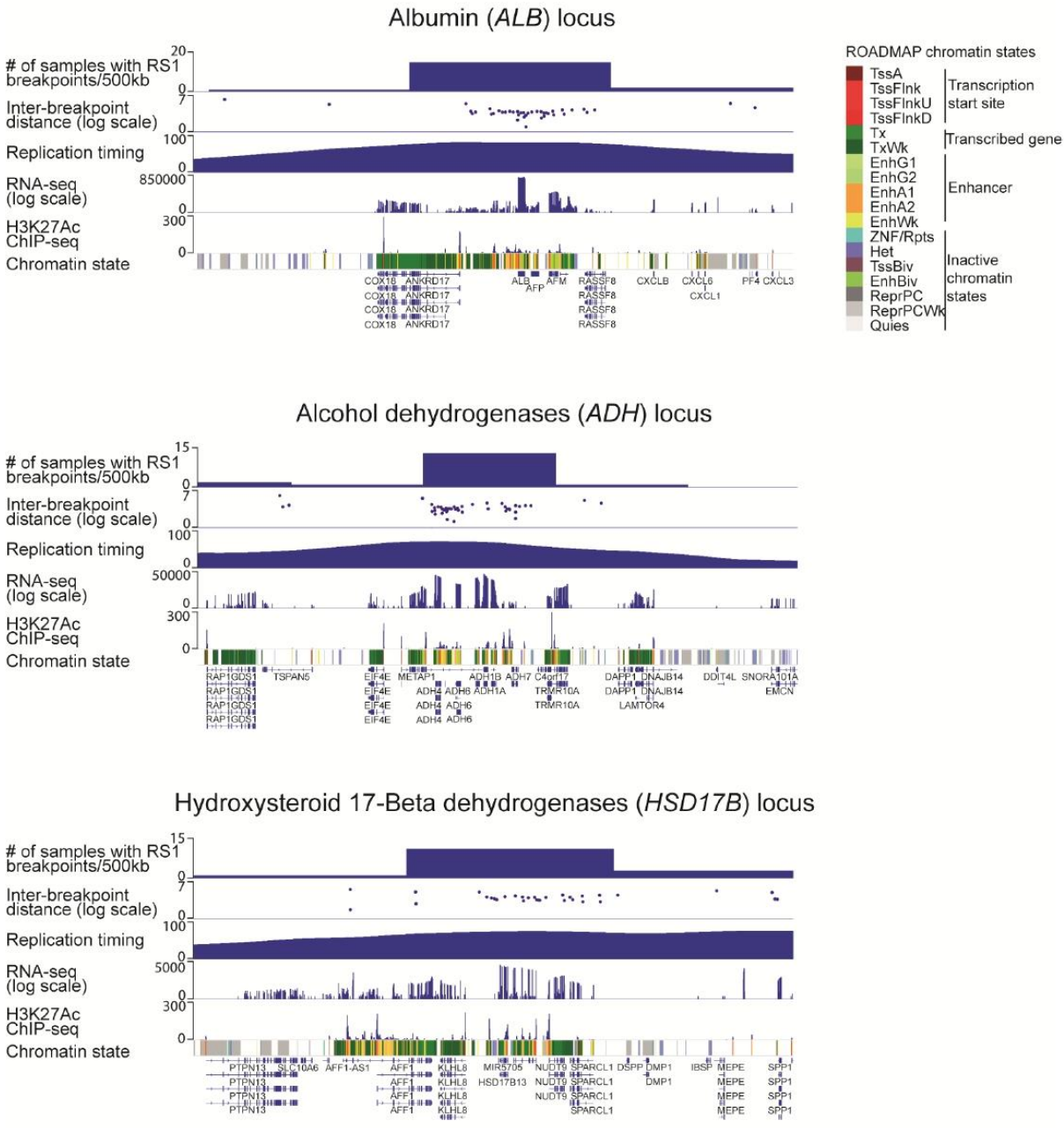
Supplementary Figure 5. SNP array analysis of focal duplications in CCN-HCC from the TCGA series.
Focal deletions (<200 kb) were quantified across 334 tumors from the TCGA series as a surrogate marker of the RS1 signature, and compared between CCNA2-activated, CCNE1-activated and other HCC. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range.



Supplementary Figure 6. Characteristic copy-number profile of CCN-HCC
a Typical copy-number profile of a CCN-HCC, showing hundreds of focal gains scattered throughout the genome.
b Proportion of focal gains in each CCN-HCC attributed to each rearrangement mechanism.

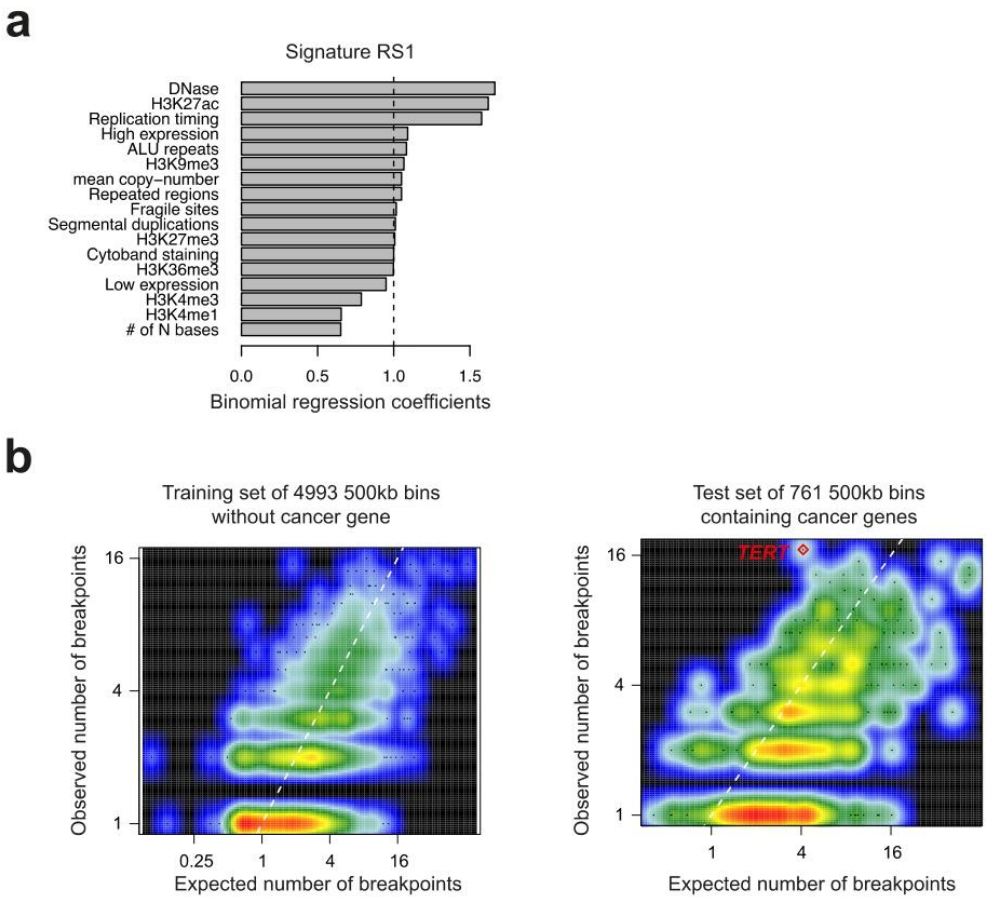


Supplementary Figure 7. Examples of intra-chromosomal templated insertions and inter-chromosomal templated insertion cycles involving several chromosomes
a Intra-chromosomal templated insertions appear as couples of inversions (left) or deletion and duplication (right) depending on the orientation of aberrant junctions.
b Example of inter-chromosomal templated insertion involving 3 different chromosomes.

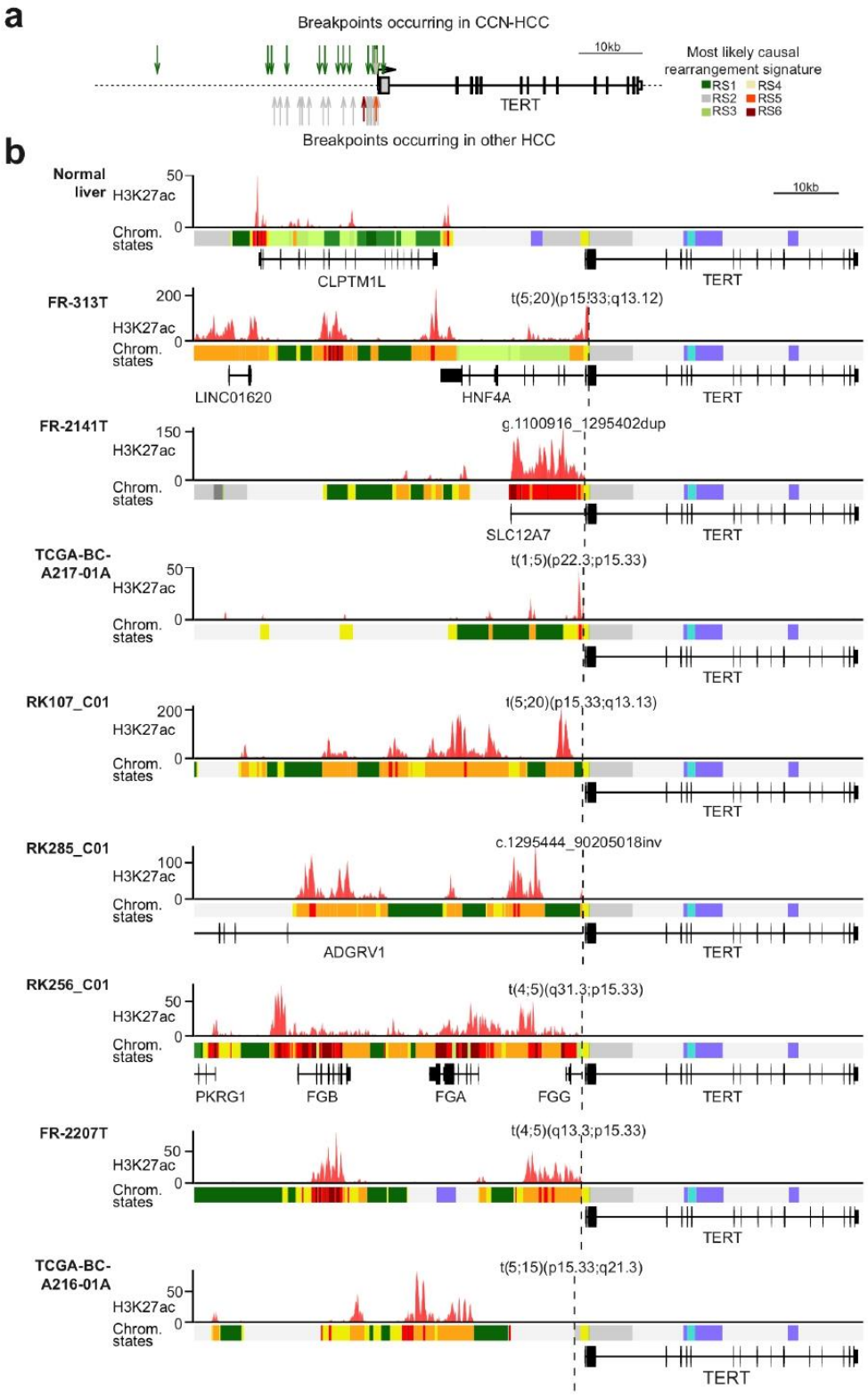


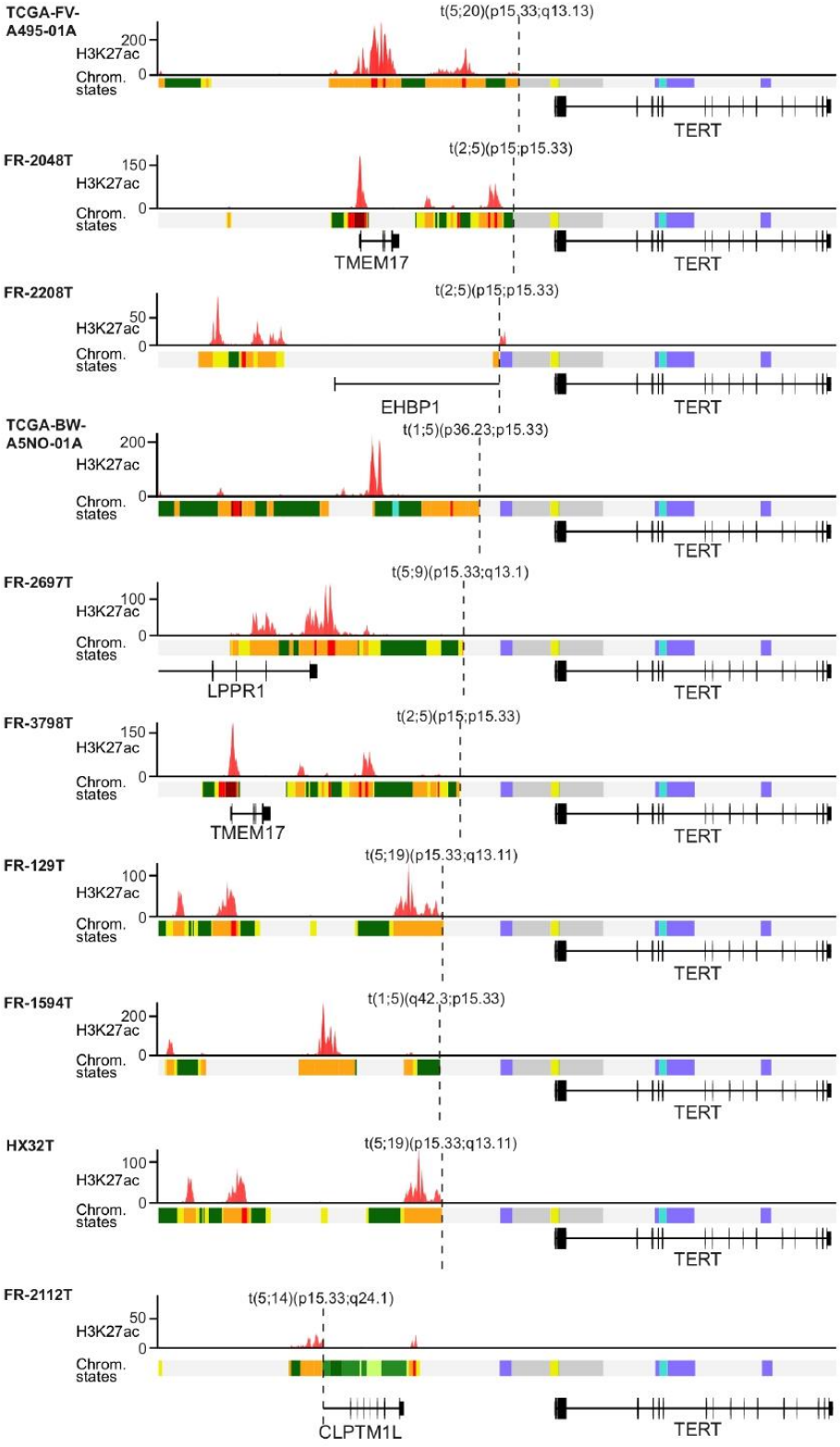
Supplementary Figure 8. RS1 breakpoint hotspots involving highly expressed liver enzymes

The density of RS1 breakpoints, replication timing, RNA-seq expression, H3K27Ac and ROADMAP chromatin states are displayed for 3 representative hotspots involving the very highly expressed liver enzymes albumin, alcohol dehydrogenases and hydroxysteroid 17-Beta dehydrogenases.



Supplementary Figure 9. Binomial regression modeling of rearrangement breakpoint density
a Regression coefficients of the 17 genomic features used to predict the density of signature RS1 breakpoints.
b Correlation between the number of observed RS1 breakpoints per 500 kb bin and the expected number predicted by the binomial regression model. Left: Within 4933 bins without any cancer gene used as training set. Right: Within 761 bins containing cancer genes (test set). The bin corresponding to *TERT* promoter region is highlighted in red.

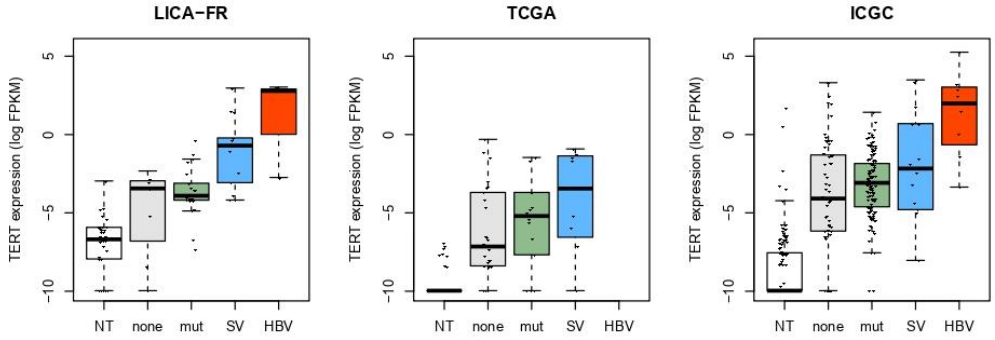

a
Breakpoints occurring in CCN-HCC
10kb
Most likely causal rearrangement signature
RS1
RS2
RS3
RS4
RS5
RS6
TERT
b
Breakpoints occurring in other HCC
Normal liver
H3K27ac
Chrom. states
CLPTM1L
TERT
FR-313T
t(5;20)(p15.33;q13.12)
LINC01620
HNF4A
FR-2141T
g.1100916_1295402dup
SLC12A7
TCGA-BC-A217-01A
t(1;5)(p22.3;p15.33)
RK107_C01
t(5;20)(p15.33;q13.13)
RK285_C01
c.1295444_90205018inv
ADGRV1
RK256_C01
t(4;5)(q31.3;p15.33)
PKRG1
FGB
FGA
FGG
FR-2207T
t(4;5)(q13.3;p15.33)
TCGA-BC-A216-01A
t(5;15)(p15.33;q21.3)


TCGA-FV-A495-01A
t(5;20)(p15.33;q13.13)
H3K27ac
Chrom. states
TERT
FR-2048T
t(2;5)(p15;p15.33)
TMEM17
TERT
FR-2208T
t(2;5)(p15;p15.33)
EHBP1
TERT
TCGA-BW-A5NO-01A
t(1;5)(p36.23;p15.33)
TERT
FR-2697T
t(5;9)(p15.33;q13.1)
LPPR1
TERT
FR-3798T
t(2;5)(p15;p15.33)
TMEM17
TERT
FR-129T
t(5;19)(p15.33;q13.11)
TERT
FR-1594T
t(1;5)(q42.3;p15.33)
TERT
HX32T
t(5;19)(p15.33;q13.11)
TERT
FR-2112T
t(5;14)(p15.33;q24.1)
CLPTM1L
TERT

Supplementary Figure 10. Rearrangements affecting *TERT* promoter region in 350 HCC genomes

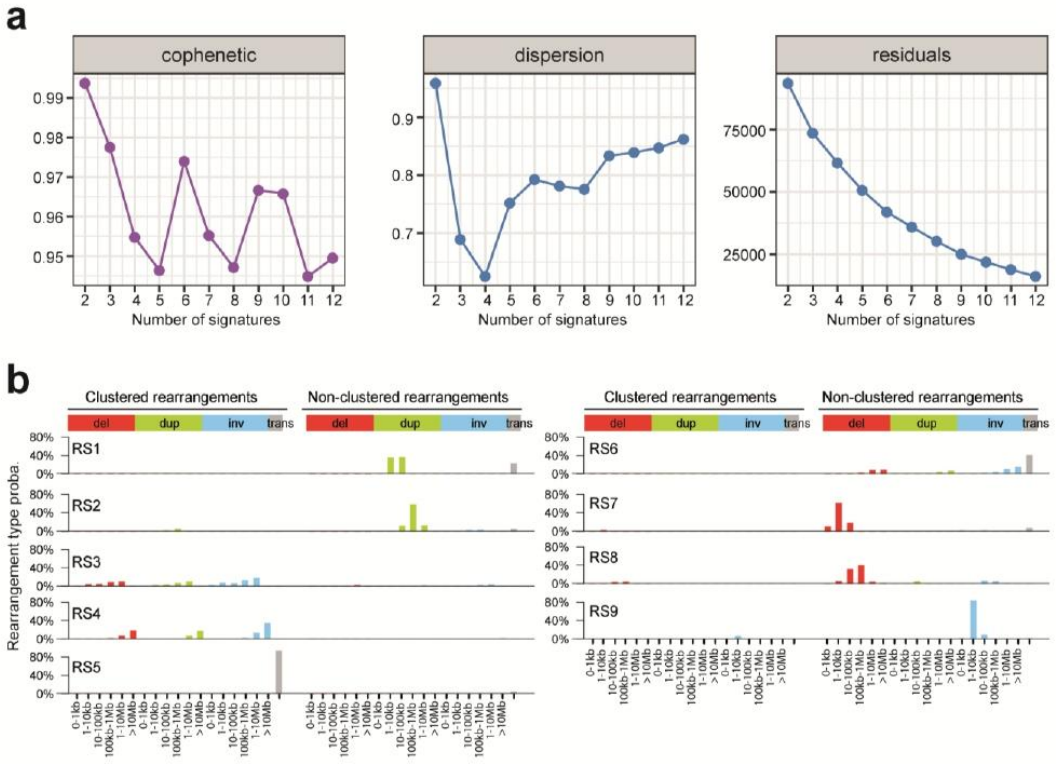
a Summary of rearrangement breakpoints in *TERT* regulatory region. Each arrow indicates a rearrangement breakpoint. The color indicates the rearrangement signature of the most likely causal process. Breakpoints occurring in CCN-HCC are represented above the scheme. Those occurring in other HCC are represented below.

b Functional consequences of structural rearrangements affecting *TERT* regulatory region in CCN-HCC. Chromatin states and H3K27 acetylation in normal adult liver (top) are compared with predicted chromatin states and H3K27 acetylation resulting from the 18 structural rearrangements of *TERT* regulatory regions identified in CCN-HCC. H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) signal were obtained from the ROADMAP consortium. The color code of chromatin states is the same as in **Supplementary Fig. 8**



Supplementary Figure 11. Modulation of *TERT* expression by different types of genomic alterations.

The expression of *TERT* is shown in the 3 data sets as a function of the type of alteration identified in *TERT* promoter. Only samples with both whole genome and RNA sequencing were considered. No HBV insertion in *TERT* was identified among the 48 TCGA samples with whole genome sequencing.
NT: non-tumor liver sample; none: HCC without *TERT* alteration identified; mut: HCC with *TERT* promoter mutation; SV: HCC with structural rearrangement affecting *TERT* promoter; HBV: HCC with HBV insertion in *TERT* promoter. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range.



Supplementary Figure 12. Rearrangement signatures identified in the pan-cancer ICGC series

a Non-negative matrix factorization (NMF) metrics used to determine the optimal number of signatures. With 9 signatures, we obtain good cophenetic coefficient and dispersion score.

b Frequency of the 38 structural rearrangement categories in the 9 signatures.

Description of Additional Supplementary Files

File Name: Supplementary Table 1
Description: Clinical annotations for the 160 samples of the LICA-FR series

File Name: Supplementary Table 2
Description: Main clinical characteristics of the LICA-FR, TCGA and ICGC-JP cohorts

File Name: Supplementary Table 3
Description: Viral insertions identified at CCNA2 and CCNE1 loci

File Name: Supplementary Table 4
Description: Structural rearrangements identified at CCNA2 and CCNE1 loci

File Name: Supplementary Table 5
Description: Significantly de-regulated pathways in CCN-HCC

File name: Supplementary Table 6
Description: Significantly enriched and depleted driver genes in CCN-HCC

File Name: Supplementary Table 7
Description: Contribution of rearrangement signatures to the genomes of 350 tumors from the LICA-FR, TCGA and ICGC series

File Name: Supplementary Table 8
Description: Rearrangements affecting TERT promoter region in 350 HCC genomes

File Name: Supplementary Table 9
Description: Number of RS1 events and association with CCNA2/E1 alterations across cancer types

Supplementary Table 1: Clinical annotations for the 160 samples of the LICA-FR series

Sample	CCNA2/E1 alteration	CCNA2/E1 alteration type	Gender	Age	Alcohol intake	Hepatitis B	Hepatitis C	Metabolic syndrome	Fibrosis stage	Edmonson grade	Largest nodule diameter (mm)	Vascular invasion	Last survival status	Last survival delay (months)
BCB307T	None	NA	M	59	no	yes	no	no	F4	I-II	30	no	death from liver cancer	33
BCM257T	CCNE1	AAV2	M	53	yes	no	no	no	F4	III-IV	30	no	death from liver cancer	29
BCM269T	None	NA	M	67	no	yes	no	no	F4	III-IV	26	no	death from liver cancer	23
CHC018T	None	NA	F	35	no	yes	no	no	F2-F3	III-IV	170	yes	death from liver cancer	35
CHC051T	None	NA	F	69	no	no	yes	no	F4	III-IV	60	no	death from liver cancer	54
CHC1010T	None	NA	F	53	yes	no	no	no	F0-F1	III-IV	100	no	alive without relapse	56
CHC1040T	None	NA	M	73	yes	no	no	no	F2-F3	III-IV	160	yes	alive with relapse	36
CHC1041T	None	NA	M	69	no	no	no	no	F0-F1	I-II	100	no	alive without relapse	36
CHC1052T	None	NA	M	75	yes	no	no	no	F2-F3	III-IV	130	yes	alive without relapse	1
CHC1053T	None	NA	M	74	yes	no	no	no	F4	III-IV	35	yes	death from liver cancer	20
CHC1055T	CCNE1	amplification	M	68	yes	no	no	no	F2-F3	III-IV	200	yes	death from liver cancer	6
CHC1061T	None	NA	F	79	no	no	no	yes	F0-F1	I-II	150	yes	alive with relapse	40
CHC1137T	None	NA	M	57	no	no	yes	no	F4	III-IV	60	yes	death from other etiology	8
CHC1148T	None	NA	M	69	yes	no	no	no	F0-F1	I-II	90	yes	death from liver cancer	15
CHC1180T	None	NA	M	65	yes	no	no	no	F2-F3	I-II	30	no	alive without relapse	89
CHC1183T	None	NA	M	60	no	no	no	no	F2-F3	I-II	180	no	death from liver cancer	49
CHC1185T	None	NA	M	53	yes	yes	no	no	F4	III-IV	30	no	alive without relapse	92
CHC1207T	None	NA	M	60	yes	no	no	no	F0-F1	I-II	210	yes	alive without relapse	66
CHC1208T	None	NA	M	47	no	yes	no	no	F4	III-IV	60	yes	death from liver cancer	11
CHC1210T	None	NA	F	44	no	yes	no	no	F2-F3	III-IV	70	no	alive without relapse	23
CHC1211T	None	NA	F	32	no	yes	no	no	F0-F1	III-IV	130	no	alive without relapse	82
CHC121T	None	NA	M	67	yes	no	no	yes	F0-F1	I-II	120	no	death from other etiology	110
CHC129T	CCNA2	AAV2	F	62	no	no	no	no	F0-F1	I-II	130	yes	death from liver cancer	5
CHC1530T	None	NA	M	64	yes	yes	no	no	F2-F3	III-IV	75	yes	death from liver cancer	16
CHC1534T	None	NA	M	67	yes	no	no	no	F2-F3	III-IV	55	no	alive without relapse	53
CHC1539T	None	NA	M	45	no	no	yes	no	F4	III-IV	32	yes	alive without relapse	59
CHC1545T	None	NA	M	77	yes	no	yes	no	F4	I-II	40	no	death from liver cancer	37
CHC1548T	None	NA	M	55	no	yes	no	no	F4	III-IV	25	no	alive without relapse	87
CHC1568T	None	NA	M	71	no	yes	no	no	F4	III-IV	30	yes	death from other etiology	45
CHC1594T	CCNE1	Rearrangement regulatory region	F	76	yes	no	no	no	F0-F1	I-II	100	yes	death from other etiology	25
CHC1598T	None	NA	F	76	no	yes	no	no	F0-F1	III-IV	90	yes	alive without relapse	61
CHC1602T	None	NA	M	71	no	no	no	no	F0-F1	I-II	75	yes	death	88
CHC1600T	None	NA	M	69	yes	no	no	no	F0-F1	I-II	80	yes	alive without relapse	66
CHC1704T	None	NA	M	43	no	yes	no	no	F2-F3	III-IV	140	yes	alive without relapse	34
CHC1603T	None	NA	M	78	yes	no	no	no	F4	III-IV	50	yes	death from liver cancer	14
CHC1604T	None	NA	M	57	no	no	no	no	F2-F3	III-IV	18	no	alive with relapse	67
CHC1616T	None	NA	F	77	no	no	yes	no	F4	III-IV	100	no	death from liver cancer	46
CHC1626T	None	NA	M	75	yes	no	no	yes	F0-F1	III-IV	100	yes	death from liver cancer	51
CHC1629T	None	NA	M	64	yes	no	no	no	F0-F1	I-II	70	yes	death	69
CHC1715T	None	NA	M	72	yes	no	no	no	F0-F1	I-II	60	no	alive without relapse	69
CHC1717T	None	NA	M	50	no	yes	no	no	F4	I-II	55	yes	alive with relapse	25
CHC1731T	None	NA	F	55	no	no	no	no	F0-F1	I-II	110	no	alive without relapse	21
CHC1725T	None	NA	F	83	no	no	yes	no	F2-F3	I-II	60	yes	alive without relapse	11
CHC1754T	None	NA	M	34	no	yes	no	no	F2-F3	III-IV	170	yes	death from liver cancer	5
CHC1732T	None	NA	M	49	yes	no	yes	no	F4	III-IV	60	yes	death from liver cancer	10
CHC1736T	None	NA	M	58	no	yes	no	no	F4	III-IV	45	yes	death from liver cancer	3
CHC1737T	None	NA	M	73	no	no	no	yes	F2-F3	III-IV	32	no	alive with relapse	61
CHC1739T	None	NA	M	55	yes	no	no	yes	F4	III-IV	50	no	death from other etiology	19
CHC1741T	None	NA	M	57	yes	no	no	no	F4	I-II	32	no	alive with relapse	60
CHC1742T	None	NA	M	67	no	no	no	no	F0-F1	III-IV	33	yes	death from liver cancer	17
CHC1744T	None	NA	M	50	no	yes	yes	no	F4	III-IV	70	yes	death from liver cancer	28
CHC1745T	None	NA	F	69	no	no	yes	yes	F4	III-IV	60	yes	death from other etiology	14
CHC1746T	None	NA	M	75	no	no	no	no	F2-F3	III-IV	40	yes	alive with relapse	65
CHC1747T	None	NA	M	54	yes	no	yes	no	F4	III-IV	40	yes	death from liver cancer	27
CHC1749T	None	NA	M	66	no	yes	no	no	F0-F1	III-IV	150	yes	alive without relapse	12
CHC1753T	None	NA	M	65	no	no	yes	no	F2-F3	III-IV	25	no	alive with relapse	112
CHC197T	None	NA	M	73	yes	no	no	no	F2-F3	III-IV	130	yes	alive with relapse	68
CHC1756T	None	NA	M	73	yes	yes	no	no	F4	III-IV	45	yes	alive with relapse	144
CHC1757T	None	NA	M	41	yes	no	no	yes	F4	I-II	12	no	alive without relapse	66
CHC1763T	None	NA	M	75	yes	no	no	yes	F0-F1	III-IV	60	no	death from other etiology	21
CHC1994T	CCNE1	HBV	M	57	no	yes	no	no	F0-F1	III-IV	190	yes	alive with relapse	52
CHC2048T	CCNE1	Rearrangement regulatory region	M	65	yes	no	no	no	F0-F1	III-IV	100	yes	death from liver cancer	21
CHC2043T	None	NA	F	21	no	no	no	no	F0-F1	III-IV	50	no	alive without relapse	67
CHC2103T	None	NA	M	57	yes	no	yes	no	F0-F1	III-IV	28	yes	death	49
CHC2111T	None	NA	F	56	no	no	no	no	F0-F1	I-II	60	no	alive without relapse	61
CHC2112T	CCNA2	AAV2	F	48	no	no	no	yes	F0-F1	III-IV	190	yes	death from other etiology	13
CHC2128T	CCNA2	AAV2	F	53	no	no	no	yes	F0-F1	I-II	200	yes	alive without relapse	45
CHC2113T	None	NA	M	61	yes	no	no	no	F0-F1	III-IV	90	no	death from other etiology	57
CHC2115T	None	NA	M	75	yes	no	no	yes	F0-F1	III-IV	100	yes	death from liver cancer	17
CHC2206T	CCNA2	AAV2	M	90	no	no	no	no	F0-F1	III-IV	40	no	alive without relapse	41
CHC2132T	None	NA	M	57	no	no	no	yes	F0-F1	III-IV	180	no	alive without relapse	51
CHC2135T	None	NA	F	57	yes	yes	no	yes	F0-F1	III-IV	25	no	death	36
CHC2141T	CCNE1	AAV2	M	74	yes	no	no	no	F2-F3	III-IV	65	no	death	16
CHC218T	None	NA	M	69	no	no	no	yes	F0-F1	III-IV	130	yes	death from liver cancer	2
CHC2200T	None	NA	M	69	no	no	no	yes	F0-F1	III-IV	110	yes	alive with relapse	40
CHC1591T	CCNE1	AAV2	M	60	no	no	no	yes	F0-F1	III-IV	120	yes	death from liver cancer	46
CHC2207T	CCNE1	Rearrangement regulatory region	M	49	no	no	no	no	F0-F1	III-IV	90	yes	alive with relapse	51
CHC2208T	CCNE1	AAV2	M	53	no	no	no	no	F0-F1	III-IV	20	no	alive without relapse	31
CHC2210T	None	NA	M	66	no	no	no	yes	F0-F1	III-IV	50	yes	death	43
CHC2415T	None	NA	M	68	no	no	no	yes	F0-F1	III-IV	180	yes	alive without relapse	36
CHC2443T	None	NA	M	74	yes	no	no	yes	F0-F1	III-IV	48	yes	death	24
CHC2448T	None	NA	M	82	no	no	no	yes	F0-F1	I-II	75	yes	death from other etiology	26
CHC2449T	CCNA2	fusion	M	81	no	no	no	yes	F0-F1	III-IV	90	no	alive with relapse	45
CHC2491T	None	NA	M	66	yes	no	no	no	F4	III-IV	30	yes	death from other etiology	86
CHC2538T	None	NA	F	76	no	no	no	yes	F2-F3	I-II	40	no	alive with relapse	37
CHC2539T	None	NA	F	41	no	no	no	no	F0-F1	III-IV	160	yes	alive with relapse	35
CHC253T	None	NA	M	67	no	no	no	yes	F4	III-IV	80	yes	death from other etiology	0
CHC2558T	None	NA	M	70	yes	no	no	yes	F0-F1	I-II	30	no	alive with relapse	34
CHC2560T	None	NA	M	74	no	no	no	yes	F0-F1	I-II	70	no	alive without relapse	28
CHC2686T	None	NA	F	52	no	no	no	no	F0-F1	III-IV	110	yes	alive without relapse	NA
CHC2687T	None	NA	M	76	no	no	no	yes	F0-F1	III-IV	120	yes	NA	NA
CHC2691T	None	NA	M	68	yes	no	no	yes	F0-F1	III-IV	60	no	NA	NA
CHC2695T	None	NA	M	94	no	no	no	no	F0-F1	III-IV	83	no	NA	NA
CHC2696T	None	NA	M	66	yes	no	no	no	F0-F1	III-IV	110	yes	NA	NA
CHC2697T	CCNA2	fusion	M	64	yes	no	no	no	F0-F1	III-IV	110	yes	death	11

Sample	CCNA2/E1 alteration	CCNA2/E1 alteration type	Gender	Age	Alcohol intake	Hepatitis B	Hepatitis C	Metabolic syndrome	Fibrosis stage	Edmonson grade	Largest nodule diameter (mm)	Vascular invasion	Last survival status	Last survival delay (months)
CHC2706T	None	NA	M	70	yes	no	no	no	F4	III-IV	50	yes	alive with relapse	35
CHC2707T	None	NA	M	79	no	no	no	no	F0-F1	I-II	30	yes	alive without relapse	33
CHC2844T	None	NA	M	55	no	no	no	no	F0-F1	I-II	120	yes	alive without relapse	21
CHC2899T	None	NA	M	52	no	no	yes	no	F4	III-IV	62	yes	alive with relapse	30
CHC3029T	None	NA	F	54	no	no	yes	no	F4	III-IV	70	yes	alive with relapse	26
CHC302T	None	NA	M	72	no	no	yes	no	F2-F3	I-II	45	no	alive without relapse	66
CHC304T	None	NA	M	77	yes	no	no	no	F0-F1	III-IV	180	yes	death from liver cancer	22
CHC306T	None	NA	M	68	no	no	yes	no	F4	I-II	20	no	alive without relapse	65
CHC309T	None	NA	F	69	no	no	yes	no	F2-F3	III-IV	20	yes	death from other etiology	2
CHC313T	CCNA2	AAV2	F	43	no	no	yes	no	F0-F1	III-IV	130	yes	death from liver cancer	11
CHC314T	None	NA	M	71	yes	no	yes	no	F2-F3	I-II	45	no	alive without relapse	60
CHC320T	None	NA	M	65	yes	no	yes	no	F4	III-IV	35	no	death from other etiology	1
CHC322T	None	NA	M	74	yes	no	no	no	F4	III-IV	40	no	death from liver cancer	21
CHC3238T	None	NA	F	58	no	no	no	no	F0-F1	III-IV	160	no	death from liver cancer	7
CHC327T	None	NA	M	63	no	no	yes	no	F4	I-II	25	no	death from other etiology	5
CHC3618T	None	NA	M	53	no	yes	no	no	F0-F1	III-IV	130	no	death from liver cancer	1
CHC3619T	None	NA	M	66	no	no	yes	no	F4	I-II	50	yes	death from liver cancer	15
CHC3620T	None	NA	M	35	no	yes	no	no	F4	III-IV	40	yes	alive	15
CHC3621T	None	NA	M	55	yes	no	no	no	F4	III-IV	60	yes	alive	11
CHC3624T	None	NA	M	64	no	no	no	no	F4	III-IV	60	no	alive	21
CHC3626T	None	NA	M	82	yes	no	no	no	F4	I-II	160	no	death from liver cancer	7
CHC3627T	None	NA	M	55	no	no	yes	no	F4	I-II	63	no	death from liver cancer	20
CHC3628T	None	NA	M	86	no	no	no	yes	F4	I-II	20	no	alive	24
CHC3631T	None	NA	M	85	yes	no	no	no	F0-F1	I-II	130	no	death from liver cancer	2
CHC3633T	None	NA	M	70	yes	no	no	yes	F4	III-IV	110	yes	death from liver cancer	1
CHC3638T	None	NA	M	84	no	no	no	yes	F0-F1	III-IV	65	no	alive	28
CHC3641T	CCNE1	AAV2	M	47	no	no	no	yes	F0-F1	I-II	140	no	alive	29
CHC3643T	None	NA	M	53	no	yes	no	no	F0-F1	III-IV	200	no	death from liver cancer	9
CHC3644T	None	NA	M	50	yes	no	no	no	F4	III-IV	100	yes	NA	NA
CHC3647T	None	NA	M	56	no	no	yes	no	F4	III-IV	50	yes	death from other etiology	11
CHC3650T	None	NA	F	67	no	no	yes	no	F4	III-IV	110	no	death from liver cancer	1
CHC3716T	None	NA	F	53	no	no	yes	no	F4	III-IV	20	no	death from liver cancer	4
CHC3788T	None	NA	M	58	no	yes	no	no	F4	III-IV	28	yes	alive with relapse	16
CHC3798T	CCNA2	HBV	F	51	no	yes	no	no	F2-F3	III-IV	120	yes	NA	NA
CHC3825T	None	NA	M	74	no	yes	no	no	F2-F3	III-IV	43	yes	NA	NA
CHC3864T	None	NA	M	69	yes	no	no	no	F4	I-II	32	no	alive without relapse	13
CHC3880T	CCNA2	fusion	F	32	no	no	no	no	F0-F1	NA	80	yes	death from liver cancer	34
CHC3894T	None	NA	M	56	yes	no	no	no	F0-F1	III-IV	160	yes	alive	69
CHC3914T	None	NA	M	68	no	yes	no	no	F4	NA	40	no	alive	10
CHC4041T	None	NA	F	64	no	no	yes	no	F2-F3	III-IV	120	yes	death from liver cancer	0
CHC4042T	None	NA	F	58	no	no	no	no	F0-F1	I-II	80	no	death from liver cancer	4
CHC4043T	None	NA	M	72	no	no	no	yes	F0-F1	III-IV	50	no	alive	11
CHC4046T	None	NA	M	58	yes	no	no	yes	F2-F3	III-IV	90	no	death from liver cancer	8
CHC4049T	None	NA	M	42	no	yes	no	no	F4	III-IV	25	yes	alive	22
CHC4051T	None	NA	M	72	yes	no	no	no	F0-F1	III-IV	190	no	death from liver cancer	8
CHC4055T	None	NA	M	62	no	no	yes	no	F4	I-II	35	yes	death from liver cancer	38
CHC429T	None	NA	F	64	no	no	no	no	F0-F1	III-IV	45	yes	alive with relapse	65
CHC432T	None	NA	M	70	yes	no	no	no	F2-F3	I-II	70	yes	death from liver cancer	34
CHC433T	None	NA	M	70	yes	no	no	yes	F0-F1	I-II	180	yes	death from liver cancer	16
CHC609T	None	NA	M	60	yes	yes	no	yes	F2-F3	III-IV	50	yes	alive with relapse	50
CHC614T	None	NA	M	61	no	no	no	yes	F0-F1	III-IV	30	yes	death from other etiology	12
CHC736T	None	NA	M	77	no	yes	no	no	F0-F1	III-IV	160	yes	alive without relapse	61
CHC793T	None	NA	M	61	no	no	no	no	F0-F1	III-IV	80	yes	death from liver cancer	37
CHC794T	None	NA	M	73	no	no	no	no	F0-F1	III-IV	160	yes	death from liver cancer	6
CHC796T	None	NA	M	76	yes	no	no	no	F2-F3	I-II	48	no	alive without relapse	31
CHC884T	None	NA	M	75	yes	no	no	no	F2-F3	III-IV	130	yes	alive without relapse	54
CHC889T	None	NA	M	71	no	no	no	yes	F2-F3	I-II	85	yes	death from other etiology	1
CHC891T	None	NA	F	73	no	no	no	no	F4	III-IV	45	yes	death from liver cancer	17
CHC892T	None	NA	F	72	no	no	no	no	F0-F1	I-II	55	no	alive with relapse	49
CHC898T	None	NA	M	71	no	no	no	no	F2-F3	III-IV	80	yes	death from liver cancer	47
CHC900T	None	NA	M	41	yes	no	yes	no	F4	I-II	17	no	alive without relapse	110
CHC909T	None	NA	M	70	no	no	no	no	F0-F1	III-IV	210	yes	alive without relapse	37
CHC912T	None	NA	M	78	yes	no	yes	no	F0-F1	III-IV	60	yes	death from liver cancer	26
CHC961T	CCNA2	fusion	M	57	yes	no	no	no	F0-F1	III-IV	190	yes	death from liver cancer	17
CHC985T	None	NA	M	55	yes	no	no	no	F4	I-II	25	no	death from other etiology	2

Supplementary Table 2: Main clinical characteristics of the LICA-FR, TCGA and ICGC-JP cohorts. For each clinical characteristic, % are given as proportion of samples with available data.

LICA-FR COHORT (n=160)

Age	Median [range]	66 [21-94]
Gender	M	126 (79%)
	F	34 (21%)
Etiology	Alcohol	63 (39%)
	Metabolic disease	37 (23%)
	HCV	30 (19%)
	HBV	30 (19%)
Largest nodule diameter (mm)	Median [range]	67 [12-210]
Fibrosis (METAVIR score)	F0-F1	75 (47%)
	F2-F3	32 (20%)
	F4	53 (33%)
Edmonson grade	I-II	48 (30%)
	III-IV	110 (70%)
	Not determined	2
Vascular invasion	yes	93 (58%)
	no	67 (42%)

TCGA COHORT (n=334)

Age	Median [range]	61 [16-90]
Gender	M	224 (67%)
	F	110 (33%)
Etiology	Alcohol	106 (32%)
	NAFLD*	19 (6%)
	HCV	52 (16%)
	HBV	86 (26%)
Fibrosis (Ishak score)	0 No Fibrosis	85 (36%)
	1,2 Portal Fibrosis	35 (15%)
	3,4 Fibrous Speta	28 (12%)
	5 Nodular Formation and Incomplete Cirrhosis	9 (4%)
	6 Established Cirrhosis	78 (33%)
	Not determined	99
Histological grade	G1	50 (15%)
	G2	160 (48%)
	G3	113 (34%)
	G4	7 (2%)
	Not determined	4
Vascular invasion	Macro	14 (5%)
	Micro	80 (29%)
	None	186 (66%)
	Not determined	54

*NAFLD: Non-Alcoholic Fatty Liver Disease

ICGC-JP COHORT (n=257)

Age	Median [range]	68 [31-86]
Gender	M	196 (76%)
	F	61 (24%)
Etiology	Alcohol (>60g per day)	33 (14%)
	HCV	147 (57%)
	HBV	76 (30%)
Tumor Size (mm)	Median [range]	30 [8-300]
Liver fibrosis (New Inuyama Classification)	0	11 (4%)
	1	28 (11%)
	1-2	12 (5%)
	2	53 (21%)
	3	61 (24%)
	4	92 (36%)
Edmonson grade	I	21 (8%)
	I-II	11 (4%)
	II	149 (58%)
	II-III	43 (17%)
	III	30 (12%)
	IV	2 (1%)
	Not determined	1
Vascular invasion	yes	86 (34%)
	no	169 (66%)
	Not determined	2

Supplementary Table 3: Viral insertions identified at CCNA2 and CCNE1 loci.

Series	Sample	Human target gene	Virus	Breakpoint location on human genome (hg19)	Breakpoint location on human genome (hg19)	Breakpoint location on viral genome	Breakpoint location on viral genome	Data used to map breakpoints	Published/new
LiC1162	CHC313T	CCNA2	AAV2	chr4:122,742,417	chr4:122,742,409[a]	2611	4585	ViralCapture	Nault et al., 2015
LiC1162	CHC2112T	CCNA2	AAV2	chr4:122,742,660	chr4:122,742,667	4608	4389	ViralCapture	Nault et al., 2015
LiC1162	CHC2128T	CCNA2	AAV2	chr4:122,743,475	chr4:122,743,452[b]	2934	1493	ViralCapture	Nault et al., 2015
LiC1162	CHC2206T	CCNA2	AAV2	chr4:122,742,478	chr4:122,742,456	3560	4596	ViralCapture	Nault et al., 2015
LiC1162	CHC129T	CCNA2	AAV2	chr4:122,743,835[b]	chr4:122,743,844	4316	4597	WGS	new
TCGA	TCGA-G3-AAUZ	CCNA2	AAV2	chr4:122,743,570	ND	4613	ND	WES	new
TCGA	TCGA-ZS-A9CF	CCNA2	AAV2	chr4:122,742,400	ND	3412	ND	RNA-seq	new
TCGA	TCGA-BC-A10Y	CCNA2	AAV2	chr4:122,743,756	chr4:122,743,742	4064	4586	WES	new
TCGA	TCGA-DD-AADF	CCNA2	HBV	chr4:122,743,189	ND	1807	ND	RNA-seq	new
ICGC-JP	RK107	CCNA2	HBV	chr4:122,742,221	chr4:122,742,581	1817	1639	WGS	Fujimoto et al., 2016
LiC1162	CHC3798T	CCNA2	HBV	chr4:122,742,792	chr4:122,742,755	1928	1783	WGS	new
LiC1162	CHC1591T	CCNE1	AAV2	chr19:30,303,035	chr19:30,303,053[b]	4340	4598	ViralCapture	Nault et al., 2015
LiC1162	CHC2141T	CCNE1	AAV2	chr19:30,287,313	chr19:30,287,316[b]	4599	4379	ViralCapture	Nault et al., 2015
LiC1162	CHC2208T	CCNE1	AAV2	chr19:30,304,532	chr19:30,304,542[b]	4604	4237	ViralCapture	Nault et al., 2015
LiC1162	CHC3641T	CCNE1	AAV2	chr19:30,291,515	ND	4239	ND	WES	new
LiC1162	BCM257T	CCNE1	AAV2	chr19:30,303,743[c]	chr19:30,303,740	4253	4590	WES	new
LiC1162	CHC1994T	CCNE1	HBV	chr19:30,291,247	chr19:30,291,287	1674	2849	WGS	new
TCGA	TCGA-BC-A10T	CCNE1	AAV2	chr19:30,302,891	ND	4626	ND	WES	new
TCGA	TCGA-DD-AACV	CCNE1	HBV	chr19:30,300,450	ND	885	ND	RNA-seq	new
TCGA	TCGA-ED-A7XO	CCNE1	HBV	chr19:30,303,439	ND	2453[d]2477[d]3024*	ND	RNA-seq	TCGA, 2017
TCGA	TCGA-DD-AADV	CCNE1	HBV	chr19:30,302,064	ND	1830	ND	RNA-seq	new
ICGC-JP	RK256	CCNE1	HBV	chr19:30,304,909	chr19:30,305,243	2040	1808	WGS	Fujimoto et al., 2016
ICGC-JP	HX032	CCNE1	AAV2	chr19:30,304,517	chr19:30,304,511	4386	4600	WGS	Fujimoto et al., 2016

ND: Not determined

*The HBV insertion in TCGA-ED-A7XO was identified by RNA-seq. Junctions were found between chr19:30,303,439 and 3 different positions on the virus, corresponding to 3 different splicing sites.

Supplementary Table 4: Structural rearrangements identified at CCNA2 and CCNE1 **loci.**

Series	Sample	Gene	Type of rearrangement	Breakpoint 1	Breakpoint 2	Data used to map breakpoints	Functional consequence	Detail
TCGA	TCGA-BC-A217	CCNA2	deletion	chr4:122743551	chr4:122745092	WGS	intragenic deletion	deletion of the exons 1 & 2
LiC1162	CHC2449T	CCNA2	inversion	chr4:122742941	chr4:106090217	WGS	gene fusion	TET2-CCNA2 fusion
LiC1162	CHC3880T	CCNA2	inversion*	chr4:122742245	chr4:106068135	RNAseq	gene fusion	TET2-CCNA2 fusion
LiC1162	CHC2697T	CCNA2	inter-chromosomal translocation	chr4:122743069	chr16:12307494	WGS	gene fusion	SNX29-CCNA2 fusion
LiC1162	CHC961T	CCNA2	inversion	chr4:122743077	chr4:106634881	WGS	gene fusion	GSTCD-CCNA2 fusion
TCGA	TCGA-GI-A6C0	CCNA2	inter-chromosomal translocation*	chr4:122742245	chr15:58703017	RNAseq	gene fusion	LIPC-CCNA2 fusion
TCGA	TCGA-DD-A3A5	CCNA2	tandem duplication*	chr4:122742245	chr4:68566767	RNAseq	gene fusion	UBA6-CCNA2 fusion
TCGA	TCGA-DD-A1EK	CCNA2	inversion*	chr4:122742245	chr4:123075267	RNAseq	gene fusion	KIAA1109-CCNA2 fusion
TCGA	TCGA-CC-5263	CCNA2	inversion*	chr4:122742245	chr4:123075267	RNAseq	gene fusion	KIAA1109-CCNA2 fusion
TCGA	TCGA-PD-A5DF	CCNA2	inversion	chr4:122743197	chr4:123091320	WGS	gene fusion	KIAA1109-CCNA2 fusion
TCGA	TCGA-BC-A3KG	CCNA2	inversion*	chr4:122742245	chr4:152403800	RNAseq	gene fusion	FAM160A1-CCNA2 fusion
TCGA	TCGA-BC-A216	CCNA2	inversion	chr4:122743078	chr4:156815968	WGS	gene fusion	TDO2-CCNA2 fusion
ICGC-JP	RK285	CCNA2	tandem duplication	chr4:122743507	chr4:122617787	WGS	gene fusion	ANXA5-CCNA2 fusion
TCGA	TCGA-BC-A10Z	CCNA2	deletion	chr4:122736735	chr4:74272165	SNP	rearrangement of the regulatory region	deletion linking ALB with the intergenic region 3' to CCNA2
TCGA	TCGA-XR-A8TG	CCNA2	deletion	chr4:122736735	chr4:74305371	SNP	rearrangement of the regulatory region	deletion linking AFP with the intergenic region 3' to CCNA2
TCGA	TCGA-FV-A495	CCNA2	deletion	chr4:122755879	chr4:100065889	WGS	rearrangement of the regulatory region	deletion linking ADH6 with the intergenic region 3' to CCNA2
LiC1162	CHC1055T	CCNE1	amplification	chr19:30303894	chr19:30314622	WES	high-level amplification	high-level amplification of the whole CCNE1 locus
LiC1162	CHC2207T	CCNE1	inter-chromosomal translocation	chr19:30268298	chr2:204506174	WGS	rearrangement bringing enhancers of the APH1B region and CCNE1 region	rearrangement bringing enhancers of the APH1B region and CCNE1 regulatory region
LiC1162	CHC2048T	CCNE1	inter-chromosomal translocation	chr19:30276189	chr1:8076609	WGS	rearrangement of the regulatory region	CCNE1 promoter translocation with ERRFI1
TCGA	TCGA-BW-A5NO	CCNE1	inter-chromosomal translocation	chr19:30276809	chr8:9176252	WGS	rearrangement bringing distal enhancers into CCNE1 regulatory region	rearrangement bringing distal enhancers into CCNE1 regulatory region
LiC1162	CHC1594T	CCNE1	inter-chromosomal translocation	chr19:30300307	chr18:71877647	WGS	rearrangement bringing enhancers of the CYB5A region and CCNE1 regulatory region	rearrangement bringing enhancers of the CYB5A region and CCNE1 regulatory region

*For rearrangements identified only at the transcript level only (RNA-seq data), breakpoint locations correspond to exon boundaries in each side of the junction, and the most likely types of rearrangement were selected from the location and orientation of gene partners.

Supplementary Table 5: Significantly de-regulated pathways in CCN-HCC.

UP-REGULATED PATHWAYS

Geneset	Geneset type	Geneset size	Normalized enrichment score LICA-FR	Normalized enrichment score TCGA	FDR q-value LICA-FR	FDR q-value TCGA	Description
HALLMARK_E2F_TARGETS	Hallmark	197	6,0755	7,5188	0	0	Genes encoding cell cycle related targets of E2F transcription factors.
HALLMARK_G2M_CHECKPOINT	Hallmark	195	5,4502	6,1559	0	0	Genes involved in the G2/M checkpoint, as in progression through the cell division cycle.
REACTOME_CELL_CYCLE	REACTOME	386	5,0887	5,7322	0	0	Genes involved in Cell Cycle
REACTOME_CELL_CYCLE_MITOTIC	REACTOME	302	4,7241	5,6609	0	0	Genes involved in Cell Cycle, Mitotic
REACTOME_GENERIC_TRANSCRIPTION_PATHWAY	REACTOME	333	4,4213	5,0535	0	0	Genes involved in Generic Transcription Pathway
REACTOME_CHROMOSOME_MAINTENANCE	REACTOME	111	3,965	4,5359	0	0	Genes involved in Chromosome Maintenance
REACTOME_DNA_REPLICATION	REACTOME	185	3,9481	4,5097	0	0	Genes involved in DNA Replication
REACTOME_G2_M_CHECKPOINTS	REACTOME	41	3,517	4,0963	0	0	Genes involved in G2/M Checkpoints
REACTOME_MITOTIC_M_M_G1_PHASES	REACTOME	165	3,4981	4,0636	0	0	Genes involved in Mitotic M-M/G1 phases
REACTOME_MITOTIC_PROMETAPHASE	REACTOME	83	3,4656	3,9825	0	0	Genes involved in Mitotic Prometaphase
REACTOME_DNA_STRAND_ELONGATION	REACTOME	30	3,4275	3,9651	0	0	Genes involved in DNA Strand Elongation
REACTOME_DEPOSITION_OF_NEW_CENPA_CONTAINING_NUCLEOSOMES	REACTOME	56	3,1044	3,7764	1,29E-05	0	Genes involved in Deposition of New CENPA-containing Nucleosomes at the Centromere
REACTOME_MEIOSIS	REACTOME	100	3,0982	3,7687	1,28E-05	0	Genes involved in Meiosis
REACTOME_G0_AND_EARLY_G1	REACTOME	23	3,018	3,6524	1,80E-05	0	Genes involved in G0 and Early G1
PID_ATR_PATHWAY	Canonical pathway	39	2,9928	3,5532	2,38E-05	0	ATR signaling pathway
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	REACTOME	30	2,974	3,5156	2,34E-05	0	Genes involved in Activation of the pre-replicative complex
REACTOME_UNWINDING_OF_DNA	REACTOME	11	2,9421	3,4882	4,03E-05	0	Genes involved in Unwinding of DNA
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	KEGG	124	2,935	3,4604	3,98E-05	0	Systemic lupus erythematosus
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRE	REACTOME	35	2,9311	3,4085	3,94E-05	0	Genes involved in Activation of ATR in response to replication stress
REACTOME_MEIOTIC_RECOMBINATION	REACTOME	72	2,8936	3,3583	5,34E-05	0	Genes involved in Meiotic Recombination
PID_FANCONI_PATHWAY	Canonical pathway	47	2,8911	3,353	5,31E-05	0	Fanconi anemia pathway
PID_PLK1_PATHWAY	Canonical pathway	45	2,867	3,3285	7,15E-05	0	PLK1 signaling events
REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	REACTOME	18	2,8521	3,3108	7,53E-05	0	Genes involved in G1/S-Specific Transcription
REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION	REACTOME	33	2,7989	3,2926	0,00011519	0	Genes involved in E2F mediated regulation of DNA replication
REACTOME_TELOMERE_MAINTENANCE	REACTOME	68	2,7784	3,246	0,00013589	0	Genes involved in Telomere Maintenance
KEGG_CELL_CYCLE	KEGG	120	2,6965	3,2354	0,00033299	0	Cell cycle
BIOCARTA_MCM_PATHWAY	BioCarta	18	2,6002	3,1199	0,00060631	2,12E-05	CDK Regulation of DNA Replication
HALLMARK_MITOTIC_SPINDLE	Hallmark	196	2,5385	3,0785	0,00099436	4,08E-05	Genes important for mitotic spindle assembly.
PID_FOXM1_PATHWAY	Canonical pathway	39	2,532	3,0502	0,00010158	5,98E-05	FOXM1 transcription factor network
REACTOME_FANCONI_ANEMIA_PATHWAY	REACTOME	21	2,4473	3,0251	0,0018333	5,89E-05	Genes involved in Fanconi Anemia pathway
REACTOME_MITOTIC_G1_G1_S_PHASES	REACTOME	129	2,4278	3,0211	0,0020988	6,85E-05	Genes involved in Mitotic G1-G1/S phases
PID_E2F_PATHWAY	Canonical pathway	71	2,4026	2,9845	0,0025517	9,57E-05	E2F transcription factor network
BIOCARTA_G2_PATHWAY	BioCarta	24	2,3998	2,9334	0,0025907	9,25E-05	Cell Cycle: G2/M Checkpoint
KEGG_DNA_REPLICATION	KEGG	36	2,3994	2,899	0,0026003	0,00011723	DNA Replication
PID_AURORA_B_PATHWAY	Canonical pathway	39	2,3981	2,8886	0,002616	0,00013464	Aurora B signaling
REACTOME_MITOTIC_G2_G2_M_PHASES	REACTOME	78	2,396	2,852	0,0026849	0,00017477	Genes involved in Mitotic G2-G2/M phases
SA_G2_AND_M_PHASES	Canonical pathway	8	2,3507	2,8325	0,0035699	0,00021432	Cdc25a activates the Cdc2/cyclin B complex to induce the G2/M transition.
REACTOME_MEIOTIC_SYNAPSIS	REACTOME	64	2,3082	2,6857	0,0046463	0,00060251	Genes involved in Meiotic Synapsis
REACTOME_CYCLIN_A_B1_ASSOCIATED_EVENTS_DURING_G2_M_TRANS	REACTOME	15	2,2912	2,6798	0,0051626	0,00063702	Genes involved in Cyclin A/B1 associated events during G2/M transition
REACTOME_LAGGING_STRAND_SYNTHESIS	REACTOME	19	2,2588	2,6222	0,0063416	0,00097661	Genes involved in Lagging Strand Synthesis
SA_REG_CASCADE_OF_CYCLIN_EXPR	Canonical pathway	11	2,2242	2,51	0,0077682	0,0019643	Expression of cyclins and regulates progression through the cell cycle by activating cyclin-dependent kinases.
REACTOME_EXTENSION_OF_TELOMERES	REACTOME	27	2,2002	2,4553	0,0089946	0,002975	Genes involved in Extension of Telomeres
REACTOME_S_PHASE	REACTOME	105	2,1638	2,424	0,011202	0,003654	Genes involved in S Phase
REACTOME_SYNTHESIS_OF_DNA	REACTOME	90	2,163	2,4108	0,011251	0,0039285	Genes involved in Synthesis of DNA
BIOCARTA_CELLCYCLE_PATHWAY	BioCarta	22	2,105	2,3928	0,015678	0,0043915	Cyclins and Cell Cycle Regulation
REACTOME_OLFACTORY_SIGNALING_PATHWAY	REACTOME	54	2,0453	2,3774	0,021773	0,0050119	Genes involved in Olfactory Signaling Pathway
BIOCARTA_ATRBRCA_PATHWAY	BioCarta	21	2,0238	2,3665	0,024483	0,0053759	Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility
REACTOME_KINESINS	REACTOME	23	2,0211	2,3083	0,024765	0,0075011	Genes involved in Kinesins
SA_G1_AND_S_PHASES	Canonical pathway	13	1,9834	2,3069	0,030465	0,0075558	Cdk2, JNK and p56lck bind cyclin D in G1, while cdk2/cyclin E promotes the G1/S transition.
REACTOME_CDC6_ASSOCIATION_WITH_THE_ORC_ORIGIN_COMPLEX	REACTOME	11	1,9705	2,1614	0,032586	0,018177	Genes involved in CDC6 association with the ORC origin complex
REACTOME_POL_SWITCHING	REACTOME	13	1,9699	2,1325	0,02263	0,021213	Genes involved in Polymerase Switching
NABA_COLLAGENS	Canonical pathway	44	1,9543	2,1079	0,035196	0,024326	Genes encoding collagen proteins
REACTOME_PROCESSIVE_SYNTHESIS_ON_THE_LAGGING_STRAND	REACTOME	15	1,9489	2,0283	0,035951	0,036694	Genes involved in Processive Synthesis on the Lagging Strand

DOWN-REGULATED PATHWAYS

Gene set	Gene set type	Gene set size	Normalized enrichment score LICA-FR	Normalized enrichment score TCGA	FDR q-value LICA-FR	FDR q-value TCGA	Description
HALLMARK_OXIDATIVE_PHOSPHORYLATION	Hallmark	192	-8,5766	-3,715	0	0	Genes encoding proteins involved in oxidative phosphorylation.
HALLMARK_MYC_TARGETS_V1	Hallmark	191	-6,9431	-3,6501	0	0	A subgroup of genes regulated by MYC - version 1 (v1).
REACTOME_METABOLISM_OF_PROTEINS	REACTOME	414	-6,5794	-3,6293	0	0	Genes involved in Metabolism of proteins
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	REACTOME	108	-6,5467	-3,5416	0	0	Genes involved in The citric acid (TCA) cycle and respiratory electron transport
REACTOME_TRANSLATION	REACTOME	141	-6,4403	-3,5255	0	0	Genes involved in Translation
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	REACTOME	104	-6,4298	-3,5173	0	0	Genes involved in 3'-UTR-mediated translational regulation
KEGG_OXIDATIVE_PHOSPHORYLATION	KEGG	109	-6,3007	-3,4372	0	0	Oxidative phosphorylation
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_	REACTOME	74	-6,1286	-3,3745	0	0	Genes involved in Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	REACTOME	98	-6,0795	-3,3411	0	0	Genes involved in Influenza Viral RNA Transcription and Replication
REACTOME_METABOLISM_OF_RNA	REACTOME	247	-5,9311	-3,2594	0	0	Genes involved in Metabolism of RNA
REACTOME_PEPTIDE_CHAIN_ELONGATION	REACTOME	83	-5,7991	-3,1913	0	0	Genes involved in Peptide chain elongation
REACTOME_INFLUENZA_LIFE_CYCLE	REACTOME	131	-5,7795	-3,0016	0	0,00018015	Genes involved in Influenza Life Cycle
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_J	REACTOME	104	-5,7731	-2,9741	0	0,00026093	Genes involved in Nonsense Mediated Decay Enhanced by the Exon Junction Complex
KEGG_RIBOSOME	KEGG	81	-5,7619	-2,6431	0	0,0028391	Ribosome
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING	REACTOME	105	-5,6704	-2,5242	0	0,0056459	Genes involved in SRP-dependent cotranslational protein targeting to membrane
REACTOME_METABOLISM_OF_MRNA	REACTOME	205	-5,5643	-2,4233	0	0,0093172	Genes involved in Metabolism of mRNA
HALLMARK_MYC_TARGETS_V2	Hallmark	56	-4,589	-2,4207	0	0,0094132	A subgroup of genes regulated by MYC - version 2 (v2).
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP	REACTOME	56	-4,4127	-2,3906	0	0,011144	Genes involved in Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S
REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUE	REACTOME	48	-4,1903	-2,3599	0	0,013414	Genes involved in Formation of the ternary complex, and subsequently, the 43S complex
REACTOME_CYTOSOLIC_TRNA_AMINOACYLATION	REACTOME	23	-3,8832	-2,349	0	0,014192	Genes involved in Cytosolic tRNA aminoacylation
REACTOME_TRNA_AMINOACYLATION	REACTOME	41	-3,7096	-2,3453	0	0,014476	Genes involved in tRNA Aminoacylation
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	REACTOME	49	-3,5366	-2,3204	0	0,016459	Genes involved in Mitochondrial Protein Import
HALLMARK_REACTIVE_OXIGEN_SPECIES_PATHWAY	Hallmark	46	-3,4888	-2,3168	0	0,016849	Genes up-regulated by reactive oxigen species (ROS).
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	KEGG	41	-3,3865	-2,2181	0	0,025562	Aminoacyl-tRNA biosynthesis
KEGG_LYSOSOME	KEGG	119	-3,0686	-2,218	3,56E-06	0,025538	Lysosome
REACTOME_TRANS_GOLGI_NETWORK_VESICLE_BUDDING	REACTOME	56	-3,0595	-2,1774	1,06E-05	0,031602	Genes involved in trans-Golgi Network Vesicle Budding
KEGG_VIBRIO_CHOLERAE_INFECTION	KEGG	51	-2,5394	-2,1678	0,00049064	0,032626	Vibrio cholerae infection
KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTI	KEGG	67	-2,4174	-2,1254	0,0011	0,039087	Epithelial cell signaling in Helicobacter pylori infection
REACTOME_RETROGRADE_NEUROTROPHIN_SIGNALLING	REACTOME	12	-2,263	-2,1124	0,0028905	0,041073	Genes involved in Retrograde neurotrophin signalling
REACTOME_INSULIN_RECEPTOR_RECYCLING	REACTOME	21	-1,9752	-2,1042	0,014178	0,042399	Genes involved in Insulin receptor recycling
REACTOME_NFKB_ACTIVATION_THROUGH_FADD_RIP1_PATHWAY_MEDI	REACTOME	11	-1,7305	-2,1028	0,044336	0,042454	Genes involved in NF-kB activation through FADD/RIP-1 pathway mediated by caspase-8 and -10

Supplementary Table 6: Significantly enriched and depleted driver genes in CCN-HCC.

LICA-FR SERIES

Gene	% alteration in CCN-HCC	% alteration in other HCC	Fisher's exact test p-value	q-value
PTEN	0,50	0,078571	0,000004	0,000110
RB1	0,70	0,178571	0,000009	0,000110
TERT promoter	0,21	0,688525	0,000160	0,001331
CTNNB1	0,00	0,407143	0,001321	0,008255
TP53	0,50	0,378571	0,004942	0,024708

TCGA SERIES

Gene	% alteration in CCN-HCC	% alteration in other HCC	Fisher's exact test p-value	q-value
PTEN	0,500000	0,084967	0,000000	0,000002
CDKN1A	0,392857	0,055556	0,000001	0,000004
RB1	0,571429	0,169935	0,000001	0,000004
FGA	0,357143	0,124183	0,000013	0,000066
CTNNB1	0,000000	0,316993	0,000549	0,002306
TERT promoter*	0,464286	0,715686	0,009426	0,034562

*TERT promoter screening available for 186 tumors.

Supplementary Table 7: Contribution of rearrangement signatures to the genomes of 350 tumors from the LICA-FR, TCGA and ICGC series

Sample	Series	CCNA2/E1 alteration	CCNA2/E1 alteration type	# of events attributed to signature RS1	# of events attributed to signature RS2	# of events attributed to signature RS3	# of events attributed to signature RS4	# of events attributed to signature RS5	# of events attributed to signature RS6
CHC2449T	LICA-FR	CCNA2	fusion	572	0	0	0	0	0
CHC1594T	LICA-FR	CCNE1	SV	524	41	39	163	0	0
TCGA-BC-A217	TCGA	CCNA2	del_ex12	459	0	113	0	0	0
CHC961T	LICA-FR	CCNA2	fusion	455	0	42	62	0	0
RK107	ICGC-JP	CCNA2	HBV	424	42	48	60	43	0
CHC129T	LICA-FR	CCNA2	AAV2	414	0	0	0	32	0
TCGA-PD-A5DF	TCGA	CCNA2	fusion	408	44	72	0	0	0
CHC3798T	LICA-FR	CCNA2	HBV	363	24	0	0	29	0
CHC313T	LICA-FR	CCNA2	AAV2	316	29	0	50	0	0
CHC2112T	LICA-FR	CCNA2	AAV2	293	81	54	119	0	0
CHC2697T	LICA-FR	CCNA2	fusion	276	0	0	0	36	0
RK285	ICGC-JP	CCNA2	fusion	275	23	0	0	54	0
CHC2207T	LICA-FR	CCNE1	SV	262	32	0	91	34	0
TCGA-BW-A5NO	TCGA	CCNE1	SV	215	69	0	122	0	0
HX32	ICGC-JP	CCNE1	AAV2	208	29	0	0	0	0
CHC2141T	LICA-FR	CCNE1	AAV2	199	57	0	92	0	0
TCGA-FV-A495	TCGA	CCNA2	SV	190	39	76	88	0	0
CHC2048T	LICA-FR	CCNE1	SV	188	31	0	0	0	0
TCGA-BC-A216	TCGA	CCNA2	fusion	171	96	38	80	60	0
CHC1994T	LICA-FR	CCNE1	HBV	123	39	15	18	13	0
CHC2208T	LICA-FR	CCNE1	AAV2	110	27	0	0	35	0
RK126	ICGC-JP	WT	NA	95	0	0	0	68	0
TCGA-DD-A4NA	TCGA	WT	NA	89	57	0	0	18	0
CHC2115T	LICA-FR	WT	NA	65	18	0	0	123	0
RK111	ICGC-JP	WT	NA	60	96	0	20	0	0
TCGA-DD-A1EB	TCGA	WT	NA	56	175	0	0	0	0
TCGA-ED-A459	TCGA	WT	NA	56	0	211	0	42	0
RK172	ICGC-JP	WT	NA	55	0	103	10	0	0
CHC218T	LICA-FR	WT	NA	46	29	0	0	488	0
CHC322T	LICA-FR	WT	NA	43	13	0	53	30	0
CHC912T	LICA-FR	WT	NA	39	0	0	122	281	0
RK023	ICGC-JP	WT	NA	34	18	0	0	105	0
RK079	ICGC-JP	WT	NA	30	0	29	15	0	6
CHC909T	LICA-FR	WT	NA	29	0	0	0	38	0
TCGA-G3-A25T	TCGA	WT	NA	28	0	0	0	69	14
CHC2200T	LICA-FR	WT	NA	27	145	34	21	23	0
CHC304T	LICA-FR	WT	NA	26	9	0	30	82	0
TCGA-FV-A496	TCGA	WT	NA	26	254	0	20	0	0
RK211	ICGC-JP	WT	NA	26	118	31	36	16	0
RK326	ICGC-JP	WT	NA	26	0	12	0	10	0
RK083	ICGC-JP	WT	NA	24	0	0	0	82	0
RK187	ICGC-JP	WT	NA	24	8	0	19	18	0
RK263	ICGC-JP	WT	NA	24	0	0	0	4	0
RK070	ICGC-JP	WT	NA	23	217	0	42	0	0
CHC314T	LICA-FR	WT	NA	22	20	0	104	11	0
TCGA-DD-A1EJ	TCGA	WT	NA	22	13	11	69	8	8
CHC2443T	LICA-FR	WT	NA	20	12	20	53	22	0
RK019	ICGC-JP	WT	NA	20	48	105	150	0	0
RK162	ICGC-JP	WT	NA	20	5	0	4	4	0
RK191	ICGC-JP	WT	NA	19	0	39	0	6	0
RK027	ICGC-JP	WT	NA	18	18	0	35	58	0
RK207	ICGC-JP	WT	NA	18	19	32	0	52	0
CHC1754T	LICA-FR	WT	NA	17	0	89	50	30	11
TCGA-G3-A25S	TCGA	WT	NA	17	0	128	65	20	15
RK229	ICGC-JP	WT	NA	17	2	0	4	4	0
TCGA-DD-A3A7	TCGA	WT	NA	16	35	43	47	29	0
TCGA-DD-A4NE	TCGA	WT	NA	16	0	28	93	17	0
HX27	ICGC-JP	WT	NA	16	0	0	0	11	0
RK115	ICGC-JP	WT	NA	16	0	0	2	10	6
CHC320T	LICA-FR	WT	NA	15	14	16	5	42	0
TCGA-BW-A5NP	TCGA	WT	NA	15	13	0	66	16	0
RK026	ICGC-JP	WT	NA	15	0	0	0	15	4
RK030	ICGC-JP	WT	NA	15	26	7	17	31	38
RK062	ICGC-JP	WT	NA	15	2	0	12	2	0
RK157	ICGC-JP	WT	NA	15	5	0	5	21	6
RK256	ICGC-JP	CCNE1	HBV	15	19	18	33	6	0
RK280	ICGC-JP	WT	NA	15	77	0	10	10	0
TCGA-FV-A23B	TCGA	WT	NA	14	0	37	31	0	0
RK001	ICGC-JP	WT	NA	14	5	0	9	18	3
RK046_2	ICGC-JP	WT	NA	14	0	6	8	44	0
RK128	ICGC-JP	WT	NA	14	0	0	89	0	0
CHC2103T	LICA-FR	WT	NA	13	0	0	15	20	0
RK228	ICGC-JP	WT	NA	13	20	0	9	22	5
RK261_2	ICGC-JP	WT	NA	13	11	15	26	54	47
CHC1704T	LICA-FR	WT	NA	12	0	0	30	101	0
TCGA-CC-5261	TCGA	WT	NA	12	0	0	42	19	0
TCGA-DD-A1EI	TCGA	WT	NA	12	0	156	43	0	11
TCGA-CC-A1HT	TCGA	WT	NA	12	11	0	22	44	0
RK257	ICGC-JP	WT	NA	12	4	0	8	3	0
CHC1598T	LICA-FR	WT	NA	11	7	18	33	10	0
HX28	ICGC-JP	WT	NA	11	0	30	119	0	0
RK075	ICGC-JP	WT	NA	11	11	0	39	61	0
RK224	ICGC-JP	WT	NA	11	0	52	75	0	0
RK266	ICGC-JP	WT	NA	11	40	0	0	39	0
CHC433T	LICA-FR	WT	NA	10	0	0	11	54	6

TCGA-CC-5260	TCGA	WT	NA	10	15	17	73	11	0
RK005	ICGC-JP	WT	NA	10	0	0	62	0	6
RK087	ICGC-JP	WT	NA	10	0	9	15	15	12
RK135	ICGC-JP	WT	NA	10	0	100	31	12	39
RK277	ICGC-JP	WT	NA	10	4	0	4	24	3
RK287	ICGC-JP	WT	NA	10	61	0	34	26	0
CHC2113T	LICA-FR	WT	NA	9	0	0	9	18	0
TCGA-FV-A3R2	TCGA	WT	NA	9	4	0	18	2	0
HX11	ICGC-JP	WT	NA	9	0	7	7	16	3
HX21	ICGC-JP	WT	NA	9	0	59	14	41	11
HX34	ICGC-JP	WT	NA	9	0	0	0	17	2
RK034	ICGC-JP	WT	NA	9	0	4	18	18	0
RK092	ICGC-JP	WT	NA	9	8	12	26	28	6
RK120	ICGC-JP	WT	NA	9	0	0	19	19	5
CHC309T	LICA-FR	WT	NA	8	5	0	13	12	0
TCGA-CC-5262	TCGA	WT	NA	8	0	8	5	17	0
TCGA-DD-A1EG	TCGA	WT	NA	8	7	31	71	0	7
TCGA-DD-A4NG	TCGA	WT	NA	8	0	18	27	12	0
RK028	ICGC-JP	WT	NA	8	1	0	0	7	2
RK059	ICGC-JP	WT	NA	8	7	0	6	11	0
RK074	ICGC-JP	WT	NA	8	13	6	25	15	0
RK082	ICGC-JP	WT	NA	8	4	0	6	13	4
RK089	ICGC-JP	WT	NA	8	0	0	27	5	17
RK104	ICGC-JP	WT	NA	8	24	0	51	0	5
RK170	ICGC-JP	WT	NA	8	24	9	16	31	0
RK201	ICGC-JP	WT	NA	8	8	0	22	23	4
RK265	ICGC-JP	WT	NA	8	0	0	0	17	0
CHC892T	LICA-FR	WT	NA	7	0	0	0	12	5
TCGA-DD-A1EL	TCGA	WT	NA	7	0	0	20	11	4
TCGA-FV-A3R3	TCGA	WT	NA	7	7	0	0	12	0
HX17	ICGC-JP	WT	NA	7	0	0	6	31	0
HX30	ICGC-JP	WT	NA	7	0	0	11	14	0
RK021	ICGC-JP	WT	NA	7	0	0	78	0	5
RK025	ICGC-JP	WT	NA	7	5	22	24	12	0
RK044	ICGC-JP	WT	NA	7	2	0	14	5	2
RK058	ICGC-JP	WT	NA	7	4	0	0	10	11
RK205	ICGC-JP	WT	NA	7	0	0	4	10	5
RK233_2	ICGC-JP	WT	NA	7	7	0	6	7	0
RK261_1	ICGC-JP	WT	NA	7	10	0	27	22	0
RK262	ICGC-JP	WT	NA	7	0	6	22	36	5
CHC2448T	LICA-FR	WT	NA	6	0	0	0	19	0
HX10	ICGC-JP	WT	NA	6	0	26	10	9	0
HX25	ICGC-JP	WT	NA	6	0	64	40	0	0
HX37	ICGC-JP	WT	NA	6	0	0	45	27	0
RK042	ICGC-JP	WT	NA	6	0	0	31	55	0
RK147	ICGC-JP	WT	NA	6	21	0	19	6	0
RK215	ICGC-JP	WT	NA	6	13	0	0	3	0
TCGA-FV-A2QQ	TCGA	WT	NA	5	10	38	22	8	0
TCGA-DD-A4ND	TCGA	WT	NA	5	3	35	0	6	0
TCGA-ED-A4XI	TCGA	WT	NA	5	0	0	0	70	0
TCGA-G3-A5SL	TCGA	WT	NA	5	5	3	9	11	3
HX16	ICGC-JP	WT	NA	5	4	0	19	4	5
RK007	ICGC-JP	WT	NA	5	0	0	7	18	2
RK050	ICGC-JP	WT	NA	5	13	0	22	8	0
RK068	ICGC-JP	WT	NA	5	0	3	28	0	5
RK088	ICGC-JP	WT	NA	5	8	18	10	19	15
RK140	ICGC-JP	WT	NA	5	0	0	10	17	4
RK141	ICGC-JP	WT	NA	5	0	0	20	0	5
RK185	ICGC-JP	WT	NA	5	0	0	21	7	0
RK199	ICGC-JP	WT	NA	5	0	0	10	0	8
RK210	ICGC-JP	WT	NA	5	0	0	2	30	0
RK217	ICGC-JP	WT	NA	5	0	0	35	4	0
RK223	ICGC-JP	WT	NA	5	0	0	3	30	0
RK234	ICGC-JP	WT	NA	5	9	0	15	17	5
RK237	ICGC-JP	WT	NA	5	19	21	22	0	0
RK264	ICGC-JP	WT	NA	5	0	0	0	53	8
CHC1717T	LICA-FR	WT	NA	4	0	0	10	6	0
CHC1731T	LICA-FR	WT	NA	4	0	0	17	21	3
CHC2132T	LICA-FR	WT	NA	4	0	19	20	26	0
CHC429T	LICA-FR	WT	NA	4	7	0	24	14	0
HX4	ICGC-JP	WT	NA	4	3	6	20	9	4
HX26	ICGC-JP	WT	NA	4	5	0	13	6	0
RK006_1	ICGC-JP	WT	NA	4	16	0	9	17	0
RK018	ICGC-JP	WT	NA	4	0	0	7	10	5
RK024	ICGC-JP	WT	NA	4	0	0	14	7	0
RK037	ICGC-JP	WT	NA	4	1	0	0	10	0
RK046_1	ICGC-JP	WT	NA	4	1	0	5	14	2
RK052	ICGC-JP	WT	NA	4	0	12	0	18	0
RK118	ICGC-JP	WT	NA	4	3	0	0	15	21
RK144	ICGC-JP	WT	NA	4	5	0	1	2	0
RK150	ICGC-JP	WT	NA	4	11	0	6	3	33
RK152	ICGC-JP	WT	NA	4	17	12	6	8	6
RK156	ICGC-JP	WT	NA	4	10	0	13	5	6
RK159	ICGC-JP	WT	NA	4	3	0	12	5	0
RK181	ICGC-JP	WT	NA	4	9	0	9	22	3
RK197	ICGC-JP	WT	NA	4	8	10	30	4	0
RK216	ICGC-JP	WT	NA	4	0	12	22	0	0
RK230	ICGC-JP	WT	NA	4	15	0	0	7	4
RK232	ICGC-JP	WT	NA	4	11	3	20	12	5

Supplementary Table 8: Rearrangements affecting TERT promoter region in 350 HCC genomes.

Sample	Cyclin.Status	Type	Most likely causal rearrangement signature	Breakpoint1 (TERT 5' flanking region)	Distance between breakpoint1 and TERT TSS	Breakpoint1 chromatin state	Breakpoint2	Breakpoint2 functional category	Breakpoint2 closest gene*	Breakpoint2 chromatin state
TCGA-BC-A217	CCNA2	inter-chromosomal translocation	RS1	chr5:1295282	-120	EnhWk	chr1:87326526	IGR	SEP15	EnhA2
CHC313T	CCNA2	inter-chromosomal translocation	RS1	chr5:1295392	-230	EnhWk	chr20:43039447	Intron	HNF4A	EnhA1
RK285	CCNA2	inversion	RS1	chr5:1295444	-282	EnhWk	chr5:90205018	Intron	GPR98	EnhA1
RK107	CCNA2	inter-chromosomal translocation	RS1	chr5:1295485	-323	EnhWk	chr20:49035293	IGR	[PTPN1]	TxWk
TCGA-BC-A216	CCNA2	inter-chromosomal translocation	RS1	chr5:1296703	-1541	ReprPCWk	chr15:53763959	IGR	[WDR72]	Quies
TCGA-FV-A495	CCNA2	inter-chromosomal translocation	RS1	chr5:1300634	-5472	ReprPCWk	chr20:48901456	IGR	[CEBPB]	EnhA1
CHC2697T	CCNA2	inter-chromosomal translocation	RS1	chr5:1309009	-13847	Quies	chr9:104109584	IGR	[BAAT]	EnhA1
CHC3798T	CCNA2	inter-chromosomal translocation	RS1	chr5:1309561	-14399	Quies	chr2:62698661	IGR	[TMEM17]	EnhWk
CHC129T	CCNA2	inter-chromosomal translocation	RS1	chr5:1311967	-16805	Quies	chr19:33772231	lincRNA	CTD-2540B15.12	EnhWk
CHC2112T	CCNA2	inter-chromosomal translocation	RS1	chr5:1330055	-34893	Tx	chr14:69147921	Intron	RAD51B	EnhA1
CHC2141T	CCNE1	tandem duplication	RS2	chr5:1295402	-240	EnhWk	chr5:1100916	Intron	SLC12A7	EnhA1
RK256	CCNE1	inter-chromosomal translocation	RS4	chr5:1295548	-386	EnhWk	chr4:155527854	Intron	FGG	EnhG2
CHC2207T	CCNE1	inter-chromosomal translocation	RS1	chr5:1295640	-478	EnhWk	chr4:74249043	IGR	[ALB]	EnhA1
CHC2048T	CCNE1	inter-chromosomal translocation	RS1	chr5:1301474	-6312	ReprPCWk	chr2:62710191	IGR	[TMEM17]	TxWk
CHC2208T	CCNE1	inter-chromosomal translocation	RS1	chr5:1303534	-8372	Quies	chr2:62926001	Intron	EHBP1	TssFlnkU
TCGA-BW-A5NO	CCNE1	inter-chromosomal translocation	RS1	chr5:1304409	-9247	Quies	chr1:8224257	IGR	[ERRFI1]	Quies
HX32	CCNE1	inter-chromosomal translocation	RS1	chr5:1312193	-17031	Quies	chr19:33770341	IGR	[CEBPA]	EnhA1
CHC1594T	CCNE1	inter-chromosomal translocation	RS1	chr5:1312557	-17395	Quies	chr1:234994146	IGR	[IRF2BP2]	TxWk
RK260	WT	inter-chromosomal translocation	RS4	chr5:1295147	15	EnhWk	chr8:126644024	IGR	[TRIB1]	Quies
RK077	WT	deletion	RS5	chr5:1295439	-277	EnhWk	chr5:1345091	5'UTR	CLPTM1L	TssFlnk
CHC1598T	WT	inter-chromosomal translocation	RS4	chr5:1295562	-400	EnhWk	chr3:58007495	Intron	FLNB	EnhA1
RK222	WT	inversion	RS4	chr5:1295601	-439	EnhWk	chr5:67466718	IGR	[RP11-404L6.2]	EnhA1
RK237	WT	inter-chromosomal translocation	RS4	chr5:1295752	-590	EnhWk	chr8:26665185	Intron	ADRA1A	EnhA2
RK261_1	WT	inter-chromosomal translocation	RS4	chr5:1295800	-638	ReprPCWk	chr19:31858861	5'UTR	NCLN	TssA
RK012	WT	inversion	RS4	chr5:1296174	-1012	ReprPCWk	chr5:156985	Exon	PLEKHG4B	Quies
CHC1602T	WT	inter-chromosomal translocation	RS4	chr5:1296214	-1052	ReprPCWk	chr2:63960694	IGR	[WDPCP]	EnhA2
RK278	WT	inter-chromosomal translocation	RS4	chr5:1296719	-1557	ReprPCWk	chr1:8223434	IGR	[ERRFI1]	Quies
RK122	WT	inversion	RS4	chr5:1296781	-1619	ReprPCWk	chr5:14730258	Intron	ANKH	EnhA2
RK044	WT	inter-chromosomal translocation	RS4	chr5:1296896	-1734	ReprPCWk	chr2:20798818	IGR	[HS1BP3]	EnhA1
TCGA-DD-A1EI	WT	inter-chromosomal translocation	RS4	chr5:1297114	-1952	ReprPCWk	chr14:37580176	Intron	SLC25A21	Quies
CHC1731T	WT	inversion	RS6	chr5:1297385	-2223	ReprPCWk	chr5:1299728	IGR	[TERT]	ReprPCWk
RK254	WT	inter-chromosomal translocation	RS4	chr5:1299077	-3915	ReprPCWk	chr12:57326014	Intron	SDR9C7	EnhWk
TCGA-G3-A25S	WT	inter-chromosomal translocation	RS4	chr5:1300602	-5440	ReprPCWk	chr1:249239943	IGR	[PGBD2]	ZNF/Rpts
RK065	WT	inter-chromosomal translocation	RS4	chr5:1302992	-7830	Het	chr10:23847990	IGR	[OTUD1]	TssFlnkU
RK283	WT	inter-chromosomal translocation	RS4	chr5:1303811	-8649	Quies	chr4:74268328	5'Flank	ALB	EnhA1
TCGA-DD-A4NE	WT	inter-chromosomal translocation	RS4	chr5:1306070	-10908	Quies	chr14:69253156	IGR	ZFP36L1	EnhA2
TCGA-DD-A3A8	WT	inter-chromosomal translocation	RS4	chr5:1307144	-11982	Quies	chr8:98176329	IGR	[TSPYL5]	EnhA2
HX26	WT	inter-chromosomal translocation	RS4	chr5:1307525	-12363	Quies	chr6:64283298	Intron	PTP4A1	TssA
RK227	WT	inter-chromosomal translocation	RS4	chr5:1309508	-14346	Quies	chr19:45247948	IGR	[BCL3]	EnhA1
RK054	WT	inter-chromosomal translocation	RS4	chr5:1311591	-16429	Quies	chr1:235105047	IGR	[TOMM20]	EnhA1
RK034	WT	deletion	RS4	chr5:1311668	-16506	Quies	chr5:86730606	IGR	[CCNH]	ReprPCWk

*For intergenic breakpoints, the name of the gene with the closest transcription start site is indicated between brackets.

Supplementary Table 9: Number of RS1 events and association with CCNA2/E1 alterations across cancer types.

Cancer type	Cancer_code	ICGC/TCGA projects regrouped	Number of tumors	Mean number of RS1 events per tumor	% of tumors with at least 5 RS1 events	Association with *CCNA2/E1* rearrangement (Fisher's exact test p-value)
Acute Myeloid Leukemia	AML	LAML-KR	8	0,75	0	1
Biliary Tract Cancer	BILCA	BTCA-SG	12	3,666666667	0	1
Chronic Lymphocytic Leukemia	CLL	CLLE-ES	91	0,175824176	0	1
Chronic Myeloid Disorders	CMD	CMDI-UK	21	0,380952381	0	1
Diffuse Large B-cell Lymphoma	DLBCL	DLBC-US	7	1,571428571	0	1
Glioblastoma Multiforme	GBM	GBM-US	40	1,6	0	1
Kidney Cancer	KIRC	KIRC-US/KIRP-US/RECA-EU/KICH-US	176	0,977272727	0	1
Lower Grade Glioma	LGG	LGG-US	17	0,529411765	0	1
Oral Cancer	OCA	ORCA-IN	12	4,333333333	0	1
Pediatric Brain Cancer	PEDBCA	PBCA-DE	204	0,828431373	0	1
Sarcoma	SARC	SARC-US	34	4,323529412	0	1
Thyroid Cancer	THYCA	THCA-US	30	0,2	0	1
Malignant Lymphoma	ML	MALY-DE	99	3,525252525	1	0,292929293
Bone Cancer	BOCA	BOCA-UK	60	4,616666667	2	1
Head and Neck Squamous Cell Carcinoma	HNSCC	HNSC-US	44	10,56818182	2	1
Pancreatic Cancer	PACA	PACA-AU/PACA-CA/PAEN-AU/PAEN-IT	307	5,296416938	2	0,890566495
Prostate Cancer	PROST	PRAD-US/PRAD-CA/PRAD-UK/EOPC-DE	278	11,83093525	2	1
Skin Cancer	SKIN	SKCM-US/MELA-AU	106	6,29245283	2	0,372675236
Colorectal Cancer	COLORECT	COAD-US/READ-US	60	12,78333333	3	1
Liver Cancer	LIHC	LIHC-US/LICA-FR/LIRI-JP/LINC-JP	345	7,808695652	3	2,54E-10
Lung Adenocarcinoma	LUAD	LUAD-US	36	4,416666667	3	1
Bladder Cancer	BLCA	BLCA-US	23	9,434782609	4	1
Lung Squamous Cell Carcinoma	LUSC	LUSC-US	48	22,39583333	4	1
Cervical Squamous Cell Carcinoma	CSCC	CESC-US	20	9,1	5	1
Uterine Cancer	UTCA	UCEC-US	47	13,55319149	6	1
Esophageal Cancer	ESOCA	ESAD-UK	87	22,06896552	11	0,372675236
Gastric Cancer	GASTR	STAD-US/GACA-CN	72	24,88888889	15	0,890566495
Breast Cancer	BRCA	BRCA-US/BRCA-EU/BRCA-UK	209	46,04784689	18	0,334057764
Ovarian Cancer	OVCA	OV-US/OV-AU	112	64,54464286	33	0,664892237

3. Annex 3: Article 5

Research Article
Hepatic and Biliary Cancer

JOURNAL OF HEPATOLOGY

BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA

Théo Z. Hirsch[1,2,†], Ana Negulescu[1,2,†], Barkha Gupta[1,2], Stefano Caruso[1,2], Bénédicte Noblet[1,2], Gabrielle Couchy[1,2], Quentin Bayard[1,2], Léa Meunier[1,2], Guillaume Morcrette[1,2,3], Jean-Yves Scoazec[4], Jean-Frédéric Blanc[5,6,7], Giuliana Amaddeo[8], Jean-Charles Nault[1,2,9], Paulette Bioulac-Sage[6,7], Marianne Ziol[10], Aurélie Beaufrère[11], Valérie Paradis[11,12], Julien Calderaro[1,2,13], Sandrine Imbeaud[1,2], Jessica Zucman-Rossi[1,2,14,*,‡]

[1]*Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, F-75006 Paris, France;* [2]*Functional Genomics of Solid Tumors laboratory, équipe labellisée Ligue Nationale contre le Cancer, Labex OncoImmunology, F-75006, Paris, France;* [3]*Service de Pathologie Pédiatrique, APHP, Hôpital Robert Debré, F-75019 Paris, France;* [4]*Service d'anatomie et de cytologie pathologiques, Gustave Roussy Cancer Center, F-94800 Villejuif, France;* [5]*Service Hépato-Gastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, CHU de Bordeaux, F-33000 Bordeaux, France;* [6]*Service de Pathologie, Hôpital Pellegrin, CHU de Bordeaux, F-33076 Bordeaux, France;* [7]*Université Bordeaux, Inserm, Research in Translational Oncology, BaRITOn, F-33076 Bordeaux, France;* [8]*Service d'Hépato-Gastro-Entérologie, Hôpital Henri Mondor, APHP, Université Paris Est Créteil, Inserm U955, Institut Mondor de Recherche Biomédicale, F-94010 Créteil, France;* [9]*Service d'Hépatologie, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, Université Sorbonne Paris Nord, F-93140 Bondy, France;* [10]*Service d'Anatomie Pathologique, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, Université Sorbonne Paris Nord, F-93140 Bondy, France;* [11]*Service de pathologie, Hôpital Beaujon, APHP, F-92110 Clichy, France;* [12]*Université de Paris, CNRS, Centre de Recherche sur l'Inflammation (CRI), Paris, F-75890, France;* [13]*Service d'Anatomopathologie, Hôpital Henri Mondor, APHP, Institut Mondor de Recherche Biomédicale, F-94010 Créteil, France;* [14]*Hôpital Européen Georges Pompidou, APHP, F-75015 Paris, France*

Background & Aims: *DNAJB1-PRKACA* fusion is a specific driver event in fibrolamellar carcinoma (FLC), a rare subtype of hepatocellular carcinoma (HCC) that occurs in adolescents and young adults. In older patients, molecular determinants of HCC with mixed histological features of HCC and FLC (mixed-FLC/HCC) remain to be discovered.

Methods: A series of 151 liver tumors including 126 HCC, 15 FLC, and 10 mixed-FLC/HCC were analyzed by RNAseq and whole-genome- or whole-exome sequencing. Western blots were performed to validate genomic discoveries. Results were validated using the TCGA database.

Results: Most of the mixed-FLC/HCC RNAseq clustered in a robust subgroup of 17 tumors, which all had mutations or translocations inactivating *BAP1*, the gene encoding BRCA1-associated protein-1. Like FLC, BAP1-HCC were significantly enriched in females, patients with a lack of chronic liver disease, and fibrotic tumors compared to non-BAP1 HCC. However, patients were older and had a poorer prognosis than those with FLC. BAP1 tumors were immune hot, showed progenitor features and did not show *DNAJB1-PRKACA* fusion, while almost none of these tumors had mutations in *CTNNB1*, *TP53* and *TERT* promoter. In contrast, 80% of the BAP1 tumors showed a chromosome gain of *PRKACA* at 19p13, combined with a loss of *PRKAR2A* (coding for the inhibitory regulatory subunit of PKA) at 3p21, leading to a high PRKACA/PRKAR2A ratio at the mRNA and protein levels.

Conclusion: We have characterized a subgroup of *BAP1*-driven HCC with fibrolamellar-like features and a dysregulation of the PKA pathway, which could be at the root of the clinical and histological similarities between BAP1 tumors and *DNAJB1-PRKACA* FLCs.

Lay summary: Herein, we have defined a homogeneous subgroup of hepatocellular carcinomas in which the *BAP1* gene is inactivated. This leads to the development of cancers with features similar to those of fibrolamellar carcinoma. These tumors more frequently develop in females without chronic liver disease or cirrhosis. The presence of PKA activation and T cell infiltrates suggest that these tumors could be treated with PKA inhibitors or immunomodulators.

© 2019 Published by Elsevier B.V. on behalf of European Association for the Study of the Liver.

Keywords: Liver cancer; Fibrolamellar carcinoma; BAP1; PRKACA; HCC; Immunotherapy; Genomics.
Received 6 September 2019; received in revised form 4 December 2019; accepted 5 December 2019; available online 18 December 2019

* Corresponding author. Address: Centre de recherche des Cordeliers, Functional genomics of solid tumors, 15 rue de l'école de médecine, 75006 Paris, France. Tel.: +33 6 01 07 78 75.
E-mail address: jessica.zucman-rossi@inserm.fr (J. Zucman-Rossi).
† Co-first authors, contributed equally.
‡ Senior, corresponding author.

https://doi.org/10.1016/j.jhep.2019.12.006

Introduction

Fibrolamellar carcinoma (FLC) is a rare subtype of hepatocellular carcinoma (HCC) mostly diagnosed in adolescents and young adults. It was originally defined by specific histological features of both the tumor cells and their stroma, with the presence of abundant fibrosis arranged in a lamellar fashion around deeply eosinophilic large neoplastic hepatocytes, frequent central scar and calcifications.[1,2] FLCs define a specific subgroup of HCC since they have peculiar clinical features compared to classical HCC, such as a young age at onset between 10 to 35 years old, a balanced sex ratio, an absence of underlying liver disease or risk factors and a better prognosis, with 80% survival at 5-years after resection.[3,4] Biologically, FLCs show a high number of mitochondria and progenitor features, suggesting that the tumor cells

JOURNAL
OF HEPATOLOGY

are blocked at a specific stage of differentiation with hepatocellular (HEPAR1), biliary (CK7) and CD68 co-expressed markers.[1,5] In 2014, Honeyman and colleagues discovered a specific 400 kb chromosome deletion at chromosome 19 in FLC leading to a recurrent chimeric *DNAJB1-PRKACA* gene fusion.[6] *DNAJB1* encodes HSP40, a member of the heat shock protein family, while *PRKACA* codes for the cAMP-dependent protein kinase (PKA) catalytic subunit alpha; the chimeric gene results in PRKACA catalytic domain overexpression and subsequent PKA activation. In addition, rare FLC without *DNAJB1-PRKACA* fusion were identified in patients with Carney disease due to *PRKAR1A* germline mutations that also led to PKA activation and a similar phenotype.[7] In contrast, PKA activation is only rarely identified in HCC (<1% *GNAS* mutations) or in cholangiocarcinoma (around 6% of *GNAS* mutations and rare fusions of *PRKACA*, *PRKACB* and *PRKAR1B*).[8,9]

Frequently, FLC diagnosis can be difficult and tumors with histological features of both HCC and FLC have been described as mixed-FLC/HCC.[10–12] In comparison with FLC, mixed-FLC/HCC patients were older, all above 35 years, and had a poor prognosis. Moreover, transcriptomic analyses showed a different profile of expression in the mixed-FLC/HCC and a lack of *DNAJB1-PRKACA* fusion.[13,14] Therefore, mixed-FLC/HCC should be better defined in the phenotypic/molecular diversity of liver tumors, in particular to identify similarities and differences with FLC of the young and other HCC subtypes.[12]

In this study, to identify specific molecular driver(s) of the mixed-FLC/HCC subgroup of tumors, we performed an integrated genomic analysis using RNA sequencing (RNAseq) and whole-genome- or whole-exome sequencing (WGS/WES) of 151 liver tumors classified as HCC, FLC and mixed-FLC/HCC by pathological review.

Material and methods

Patients and tumors

We assembled a series of 151 patients (LICA-FR cohort) enriched with 15 FLC, 10 mixed-FLC/HCC cases and including 126 HCC controls;[8,15] genomic data (WGS or WES and RNAseq) and data from pathological review were available for all cases (see below). The study was approved by the local Ethics Committee (CCPPRB Paris Saint-Louis IRB00003835), and informed consent was obtained in accordance with French legislation for all patients. Tumor and corresponding non-tumor samples were frozen at -80°C after tumor resection and all tumor samples were primary tumors except 8 samples collected at relapse; clinical features are summarized in Table 1. For validation, we used 345 tumors (339 HCC, 4 FLC, 2 mixed-FLC/HCC) from the TCGA (The Cancer Genome Atlas) cohort,[16] whose clinical features are also described in Table 1.

Pathological review

Multiple slides of the same tumor were reviewed by at least 2 liver pathologists for several histological criteria previously described[11,17,18] including: tumor differentiation (WHO grade and Edmonson grade), vascular invasion (microscopic), tumor architecture (macrotrabecular, microtrabecular, pseudoglandular), abundant fibrous stroma (>20%), steatosis, histological patterns (scirrhous, macrotrabecular massive), lymphocytic inflammation (>20%), tumor invasion (biliary tract, perineural) and clear cell presence. HCC containing at least 1 area with FLC-like features were annotated as mixed-FLC/HCC as previously described[11] (Fig. S1). For the TCGA cohort, histological features were reviewed by at least 2 liver pathologists from the virtual slides available online via cBioPortal (http://www.cbioportal.org/). Tertiary lymphoid structures were annotated on hematein-eosin-safran slides as in[19,20] for a subset of the LICA-FR cohort (55 tumors) and all tumors from the TCGA cohort.

Genomic sequencing

A total of 151 tumors and their corresponding non-tumor tissues were sequenced by WES (n = 115 including 43 new cases) or WGS (n = 36 including 3 new cases) as previously described.[8,21,22] For samples without WGS, targeted analysis of *TERT* promoter was performed with Sanger (n = 100) and/or miSeq (n = 38) sequencing as previously described,[8] while *BAP1* was re-sequenced with miSeq technology in 24 samples (the list of primer sequences is provided in Table S1). Copy number analysis was performed as previously described[22] and manually corrected with the help of the GAP tool.[23]

RNAseq and transcriptomic analyses

mRNAs were extracted from 151 tumor (including 64 new cases) and 3 non-tumor tissues and sequenced using Illumina TruSeq or Illumina TruSeq Stranded mRNA kit, libraries were sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 2,000 or 4,000 paired-end 75 bp or 100 bp reads as previously described.[24] Fastq files were aligned to the reference human genome GRCh38 using TopHat2.[25] Reads mapping to multiple locations were removed and we used HTSeq[26] to obtain the number of reads associated with each gene in the Gencode v27 database, restricting reads to protein-coding genes, pseudogenes, antisense and lincRNAs (n = 58,288). We used the Bioconductor DESeq2 package[27] to import raw HTSeq counts for each sample into R statistical software and applied variance stabilizing transformation to the raw count matrix to obtain an expression matrix without variance-mean dependence (vst matrix). FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalizing the count matrix for the library size and the coding length of each gene. We used the area under the ROC curve (AUC) to identify and remove 2,404 genes with a significant batch effect (AUC >0.95 between one sequencing project and others). Unsupervised hierarchical clustering was done using Cosine distance and Ward method on most 5,000 differentially expressed genes from the vst matrix as previously described.[24]

Tumors were classified in the G1-G6 classification as previously described for the LICA-FR cohort[28] and the TCGA cohort.[29] We used the Bioconductor *limma* package[30] to test for differential gene expression between 2 groups of RNAseq samples of interest. All genes expressed in at least 5 samples (FPKM>0) were considered for the differential expression analysis. We applied a q-value threshold of ≤0.05 to define differentially expressed genes. We used an in-house adaptation of the gene-set enrichment analysis (GSEA) method[31] to identify gene sets from the MSigDB (v. 6.0) database overrepresented among up- and downregulated genes. Single sample GSEA scores were calculated as gene-set variation analysis (GSVA) enrichment scores, using the GSVA package.[32]

The relative abundance of immune and stromal cell infiltration within tumors was assessed through microenvironment cell population (MCP) scores computed with the MCP-counter method[33] adapted to RNAseq data (exp matrix) after filtering

Table 1. Clinical, histological and molecular features of tumors stratified by molecular alterations.

	LICA-FR					TCGA[a]				
	Non-BAP1/ fusPRKACA tumors (n = 118)	BAP1 tumors (n = 18)		fusPRKACA tumors (n = 15)		Non-BAP1/ fusPRKACA tumors (n = 323)	BAP1 tumors (n = 16)		fusPRKACA tumors (n = 6)	
	n/N (%)	n/N(%)	p	n/N (%)	p	n/N (%)	n/N(%)	p	n/N (%)	p
Histological Diagnosis										
HCC	116/118 (98)	10/18 (56)	***###	0/15 (0)	‡‡‡	323/323 (100)	16/16 (100)	###	0/6 (0)	‡‡‡
FLC	0/118 (0)	2/18 (11)	*###	13/15 (87)	‡‡‡	0/323 (0)	0/16 (0)	##	4/6 (67)	‡‡‡
mixed-FLC/HCC	2/118 (2)	6/18 (33)	***	2/15 (13)		0/323 (0)	0/16 (0)		2/6 (33)	‡‡‡
Age at sampling (years)	64 (18-90)	51 (27-73)	**###	26 (17-36)	‡‡‡	61 (20-90)	61 (44-77)	###	22 (17-35)	‡‡‡
Female gender	23/118 (19)	12/18 (67)	***	9/15 (60)	‡‡	92/323 (28)	12/16 (75)	***	3/6 (50)	
Alcohol intake	47/116 (41)	5/16 (31)		0/12 (0)	‡‡	107/304 (35)	3/14 (21)		1/4 (25)	
Hepatitis B	39/118 (33)	1/17 (6)	*	0/12 (0)	‡	102/304 (34)	2/14 (14)		0/4 (0)	
Hepatitis C	15/116 (13)	0/15 (0)		0/12 (0)		47/304 (15)	1/14 (7)		0/4 (0)	
Without etiology	14/118 (12)	8/16 (50)	***##	12/12 (100)	‡‡‡	67/300 (22)	7/13 (54)	*	3/4 (75)	‡
Non-tumor liver histology***‡‡‡*‡										
F0-F1	52/118 (44)	16/17 (94)		15/15 (100)		92/192 (48)	9/12 (75)		5/5 (100)	
F2-F3	36/118 (31)	1/17 (6)		0/15 (0)		29/192 (15)	2/12 (17)		0/5 (0)	
F4	30/118 (25)	0/17 (0)		0/15 (0)		71/192 (37)	1/12 (8)		0/5 (0)	
Abundant fibrous stroma	32/115 (28)	13/18 (72)	***#	15/15 (100)	‡‡‡	36/323 (11)	6/16 (38)	**	4/6 (67)	‡‡
Tumor steatosis	19/100 (19)	10/15 (67)	***##	2/13 (15)		105/323 (33)	12/16 (75)	***#	1/6 (17)	
Biliary Tract Invasion	2/104 (2)	5/16 (31)	***	1/13 (8)		7/323 (2)	2/16 (13)		0/6 (0)	
Perineural invasion	0/104 (0)	2/15 (13)	*	2/13 (15)	‡	0/323 (0)	1/16 (6)	*	1/6 (17)	‡
Macrotrabecular massive	13/114 (11)	0/13 (0)		0/15 (0)		62/323 (19)	1/16 (6)		0/6 (0)	
Scirrhous pattern	9/117 (8)	3/18 (17)		0/13 (0)		5/319 (2)	0/16 (0)		0/6 (0)	
Lymphocytic inflammation	23/110 (21)	11/15 (73)	***##	1/12 (8)		70/323 (22)	8/16 (50)	*	2/6 (33)	
Boyault transcriptomic class			***##		‡‡			***		
G1	13/118 (11)	14/18 (78)		3/15 (20)		41/292 (14)	10/16 (63)		1/6 (17)	
G2	17/118 (14)	1/18 (6)		0/15 (0)		46/292 (16)	1/16 (6)		0/6 (0)	
G3	28/118 (24)	1/18 (6)		3/15 (20)		29/292 (10)	0/16 (0)		1/6 (17)	
G4	22/118 (19)	2/18 (11)		9/15 (60)		83/292 (28)	3/16 (19)		3/6 (50)	
G5	21/118 (18)	0/18 (0)		0/15 (0)		17/292 (6)	0/16 (0)		0/6 (0)	
G6	17/118 (14)	0/18 (0)		0/15 (0)		76/292 (26)	2/16 (13)		1/6 (17)	
BAP1 altered	0/118 (0)	18/18 (100)	***###	0/15 (0)		0/323 (0)	16/16 (100)	***###	0/6 (0)	
CTNNB1 mutated	41/118 (35)	1/18 (6)	*	1/15 (7)	‡	92/323 (28)	1/16 (6)		0/6 (0)	
TP53 mutated	47/118 (40)	2/18 (11)	*	0/15 (0)	‡‡	107/323 (33)	3/16 (19)		0/6 (0)	
TERT altered	63/116 (54)	0/17 (0)	***	2/14 (14)	‡‡	89/169 (53)	0/8 (0)	**	2/4 (50)	
DNAJB1-PRKACA fusion	0/118 (0)	0/18 (0)	###	15/15 (100)	‡‡‡	0/323 (0)	0/16 (0)	###	6/6 (100)	‡‡‡

FLC, fibrolamellar carcinoma; HCC, hepatocellular carcinoma.
Wilcoxon test, trend χ^2 test, χ^2 test and Fisher exact test were used respectively for continuous, ordinal, categorical and binary data.
*BAP1 *vs.* non-BAP1/fusPRKACA; #BAP1 *vs.* fusPRKACA; ‡fusPRKACA *vs.* non-BAP1/fusPRKACA.
###, ***, ‡‡‡ p <0.001; ##, **, ‡‡ p <0.01; #, *, ‡ p <0.05.
[a]For the TCGA series, only 1 or 2 slides were available for histological review.

out genes expressed in hepatocytes (the detailed list of genes used for each population is available in Fig. S2).

Immunoblotting

Proteins were extracted from frozen tissues samples using modified Laemmli lysis buffer (50 mM Tris pH = 6.8, 2% SDS, 5% glycerol, 2 mM DTT, 2.5 mM EDTA/EGTA) supplemented with 2x phosphatase and protease inhibitor cocktail (#78444, Thermo Fisher Scientific, Waltham, MA, USA), 2 mM Na_3VO_4 (#P0758S, BioLabs, Ipswich, MA, USA) and 10 mM NaF (#674141, Sigma Aldrich, St. Louis, MO, USA), homogenized by Qiagen TissueLyser II and boiled for 10 min. Protein concentration was measured with the BCA assay kit (Pierce Biotechnology, Rockford, IL, USA). Equal amounts of protein were deposed on polyacrylamide gel (Bio-Rad, Hercules, CA, USA), separated by electrophoresis, transferred on 0.2 μm nitrocellulose membrane (Bio-Rad) and incubated at 4°C with the following primary antibodies: anti-BAP1 (sc-28383 c4, Santa Cruz Biotechnologies, Dallas, TX, USA), anti-PKA-Cα (ab76238, Abcam, Cambridge, UK) and anti-PKA-R2 (ab32514, Abcam, Cambridge, UK) at 1:10,000 dilution. Anti-mouse (#7076, Cell Signaling) or anti-rabbit (#7074, Cell Signaling) IgG horseradish peroxidase (HRP) linked secondary antibodies were used at 1:2,000 dilution. Proteins were detected using SuperSignal West Pico Plus kit (#34580, Thermo Fisher Scientific) and the signal was captured by the ChemiDoc XRS system. Quantification was done by measuring the intensity of each band using Image Lab software (Bio-Rad) and normalized by the Ponceau whole lane charge evaluated with ImageJ software (National Institute of Health, Bethesda, MD, USA).

Immunohistochemistry

Immunohistochemistry was performed on formalin-fixed paraffin-embedded tumor tissues with the Leica Bond automated stainer. The percentage of EpCAM positive tumor cells was assessed in 52 tumors from the LICA-FR series stained with an EpCAM monoclonal antibody (Dako, Santa Clara, California, Clone Ber-Ep4, dilution 1/50). Quantitative scores of infiltration by T cells and B cells were assessed in 41 tumors with antibodies

JOURNAL
OF HEPATOLOGY

against CD3 (Dako, Clone F7.2.38, dilution 1/50) and CD20 (Dako, Clone L26, Dilution 1/500) respectively, with a separate estimation of positive cells within the tumor stroma and within the tumor parenchyma.

Statistical analysis

For statistical analysis of gene expression data, R version 3.3.2 (R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, http://www.R-project.org) and Bioconductor version 3.4 were used. Wilcoxon, Fisher or χ^2 statistical tests were applied with respect to the type of variable. Survival analysis was performed in patients treated by R0 liver resection (n = 112) as previously described[8] including 5 patients with FLC without good quality RNAseq data available. We assessed overall survival defined by the interval between surgery and death (whatever the cause), and survival curves were represented using the Kaplan-Meier method compared with the Log-Rank test. Univariate analysis was performed using the Cox model. A p value <0.05 was considered significant.

Data availability

The sequencing data of the LICA-FR cohort reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (RNAseq accession [EGAS00001002879], [EGAS00001001284] and [EGAS00001003837]; WES accessions [EGAS00001000217], [EGAS00001001002], [EGAS00001003063], [EGAS00001000679] and [EGAS00001003837]; WGS accessions [EGAS00001002408], [EGAS00001000706], [EGAS00001002888] and [EGAS00001003837]), through the ICGC (International Cancer Genome Consortium) data access committee. The sequencing data from the TCGA cohort were retrieved from http://www.cbioportal.org/.

Results

FLC and mixed-FLC/HCC cluster into 2 distinct transcriptomic subgroups of tumors

We performed RNA sequencing of 151 liver tumor samples which included 126 HCC, 15 FLC and 10 mixed-FLC/HCC, defined by pathological review (Fig. S1) and with clinical features described in Table 1. Out of 15 FLC, 13 clustered together with 2 mixed-FLC/HCC in a robust subgroup of 15 tumors (Fig. 1). Interestingly, all 15 patients in this cluster were young (age min = 17 years; max = 36 years, mean = 26 years) and all the tumors harbored the *DNAJB1-PRKACA* fusion (Fig. 1). Six out of the 8 remaining mixed-FLC/HCC clustered in a second group of tumors also including 2 FLC and 9 HCC (Fig. 1). This cluster of 17 tumors highly enriched

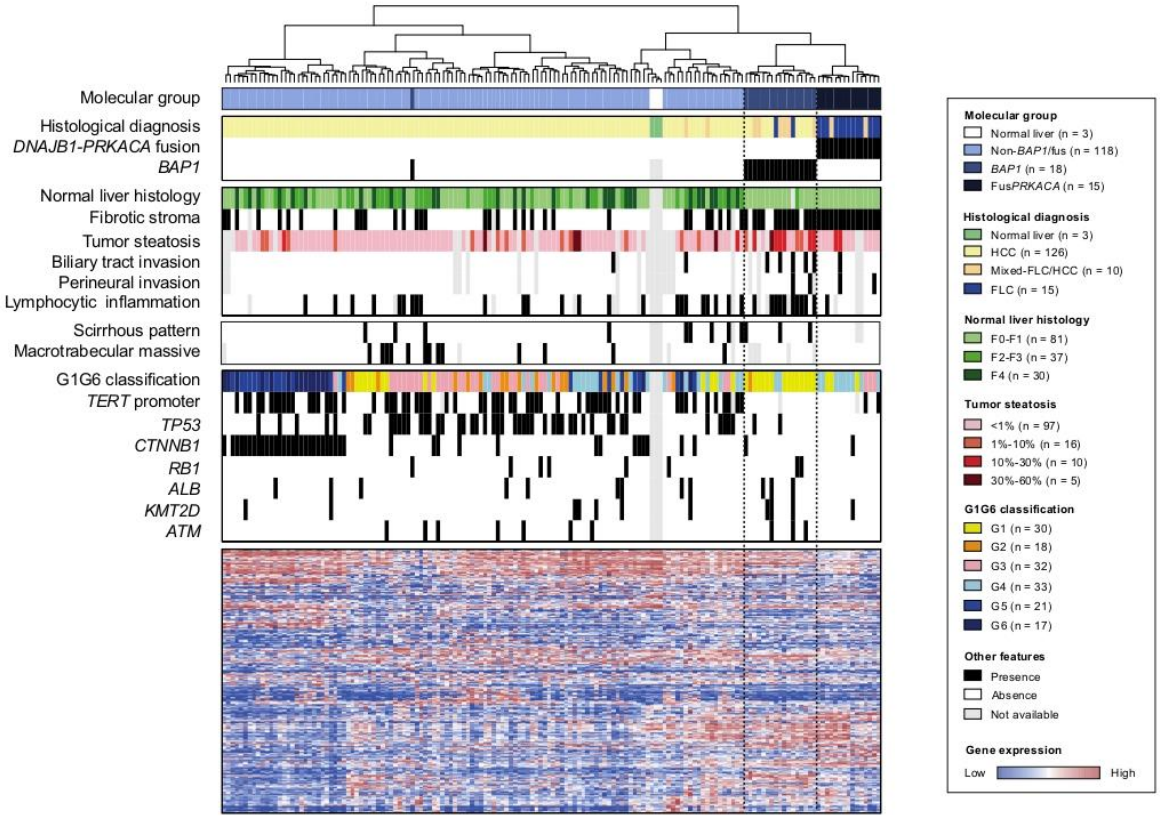


Fig. 1. FLC cluster together and are divided between *BAP1*-driven and *DNAJB1-PRKACA*-driven tumors. Hierarchical clustering based on the expression of the 5,000 genes with the higher variability in RNAseq analysis of 151 liver tumors and 3 non-tumor-liver samples. The top heatmaps show clinical, histological and molecular annotation of the samples, while the lower heatmap shows the expression pattern of the 5,000 genes. FLC, fibrolamellar carcinoma; RNAseq, RNA sequencing. (This figure appears in color on the web.)

Table 2. Detailed *BAP1* alterations in the 18 patients from the LICA-FR series.

Case #	Gender	Age	Histological Diagnosis	Etiology / background	Non-tumor liver	*BAP1* genomic alteration (GRCh37, chr3)	BAP1 protein change	*BAP1* copy number variation
#794T[a]	M	73	HCC	Hemochromatosis	F0-F1	52443600del	E31fs	Focal del
#1010T	F	53	HCC	Alcohol	F0-F1	52436869_52436870delinsAA	K637*	Focal del
#3894T	M	56	HCC	Alcohol	F0-F1	52443876C>A	E7*	Focal del
#3907T	M	39	HCC	Alcohol	F0-F1	52441270del	A167fs	Focal del
#2135T	F	57	HCC	Alcohol / HBV / NASH	F0-F1	52441301del + 52436304T>G	N157fs + *730C	
#2211T	F	37	HCC	Chronical intrahepatic obstructive cholestasis	F0-F1	52439814_52439830del	VLEANR295fs[b]	Focal del
#228T	M	48	HCC	Without etiology	F0-F1	52440294_52440295insTT	Q253fs	Focal del
#1182T	F	51	HCC	Without etiology	F0-F1	52442618_52442627del	PA42fs (splice)	large del
#800T	M	73	HCC	Without etiology	F2-F3	52441229del	F181fs	Chr3p del
#141T	F	67	HCC	Without etiology	F0-F1	52441252T>C	Y173C	Focal del
#412T	F	51	FLC	n.a.				Homozygous del
#411T	F	65	FLC	Alcohol	F0-F1			Homozygous del
#3919T	F	54	mixed-FLC/HCC	n.a.	F0-F1	Inversion (3;3)	Frameshift fusion	Focal del
#4211T	F	40	mixed-FLC/HCC	Without etiology	F0-F1	Translocation t(2;3)	Frameshift fusion	Chr3p del
#026T	F	38	mixed-FLC/HCC	Without etiology	F0-F1	52439136del	P370fs[b]	Focal del
#255T	F	39	mixed-FLC/HCC	Without etiology	F0-F1	52441473T>C + 52443758G>A	K127E + G13G (splice)	Focal del
#187T	M	27	mixed-FLC/HCC	Without etiology	F0-F1	52441991_52441992del	K120fs[b]	Focal del
#906T	F	50	mixed-FLC/HCC	NASH	F0-F1	52443580del	K38fs	Focal del

F0-F1: no fibrosis; F2-F3: mild fibrosis; F4: cirrhosis.
FLC, fibrolamellar carcinoma; HCC, hepatocellular carcinoma; NASH, non-alcoholic steatohepatitis; n.a., not available.
[a]Of note, patient #794T was found outside the cluster of 17 *BAP1* altered tumors (Fig. 1) but was later included in the group of BAP1 tumors for further analysis.
[b]3 mutations were not detected in the first analysis of whole exome sequencing data but were later identified in target sequencing of *BAP1* with a higher depth.

in mixed-FLC/HCC was located in proximity to the FLC cluster, but patients were older (age min = 27 years; max = 73 years, mean = 51 years) including the 2 FLC that developed at 51 years and 65 years. Interestingly all 17 of these tumors lacked the *DNAJB1-PRKACA* fusion but showed alterations in the *BAP1* gene. Only 1 HCC mutated for *BAP1* (#794T), without FLC features, was outside this FLC/HCC cluster. Finally, only 2 mixed-FLC/HCC, not mutated for *BAP1* nor harboring *DNAJB1-PRKACA* fusion, were distributed outside the 2 clusters (Fig. 1).

In agreement with the tumor suppressor function of BAP1 in various tumor types,[16,34–37] *BAP1* alterations were predicted to inactivate the gene either by somatic mutations (9 frameshift, 2 nonsense, 1 splice site, 2 missense and 1 nonstop), by gene fusion in 2 cases (*DAG1-BAP1* in #3919T and a complex event leading to both *BAP1-LINC01460* and *RAB3GAP1-BAP1* fusions in #4211T, see Fig. S3), or by homozygous deletion in 2 tumors (Table 2). In all tumors, *BAP1* alterations were biallelic with 2 mutations in 1 case (#2211T) or a deletion of the second allele of *BAP1* (Table 2).

Since no FLC carrying the *DNAJB1-PRKACA* fusion showed *BAP1* alterations, we define 3 distinct groups of tumors: altered for *BAP1* (BAP1 tumors), with *DNAJB1-PRKACA* fusion (fusPRKACA tumors) or without either alterations (non-BAP1/fusPRKACA tumors), as represented in Fig. 1 (Molecular group annotation).

BAP1 altered hepatocellular carcinomas show distinct clinical and histological features

In our cohort (LICA-FR), patients with BAP1 tumors (n = 18) were younger (median 51 years) than patients with non-BAP1/fusPRKACA tumors (n = 118, median 64 years, p = 0.006) but older than patients with fusPRKACA tumors (n = 15, median 26 years, p = 3.3e-6, Fig. 2A). Compared to non-BAP1/fusPRKACA tumors, BAP1 tumors were more frequent in females (67% *vs.* 19%, p = 9.7e-5), in non-fibrotic liver (F0-F1), and in patients without known etiologies of HCC (50% *vs.* 12%, p = 0.0008, Fig. 2C-E and Table 1). Patients with BAP1 tumors had similar overall survival as those with non-BAP1/fusPRKACA HCC, whereas patients with fusPRKACA HCC had a better prognosis (p = 0.02 in log-rank test, Fig. 2B). All of these clinical relations, except younger age, were also found in the TCGA cohort that included 16 BAP1 tumors and 6 fusPRKACA tumors out of 345 HCC (Fig. 2A-E and Table 1). Of note, a 67-year-old female patient in TCGA harbored both a *DNAJB1-PRKACA* fusion and a *BAP1* mutation in the tumor and was excluded from the statistical analysis.

BAP1 tumors were enriched in mixed-FLC/HCC (6 out of 18), characterized by an abundant fibrous stroma (72%), intratumor steatosis (67%), biliary tract invasion (31%), perineural invasion (13%), and a high lymphocytic infiltration (73%) (Fig. 1 & 2F-J, Table 1). Tertiary lymphoid structures were frequently identified in BAP1 tumors (70%) ranging from simple aggregate up to secondary follicles harboring a clear germinal center as previously described in HCC[20] (Fig. 2K). The MCP-counter method[33] applied to RNAseq data showed a high component of the microenvironment in *BAP1* tumors, enriched in T and B lymphocytes, endothelial cells and myeloid dendritic cells when compared with non-BAP1/fusPRKACA tumors (Fig. 2L, Fig. S2). In accordance with these results, CD3 immunohistochemistry revealed a higher infiltration of T cells in the stroma of BAP1 tumors compared to non-BAP1/fusPRKACA tumors (Fig. 2M,N). Infiltrated $CD20^+$ B cells were also observed within BAP1 tumors but in levels comparable to non-BAP1/fusPRKACA HCC (Fig. 2O and Fig. S2). In the TCGA cohort, we confirmed that BAP1 tumors were highly fibrous, steatotic, enriched in perineural invasion and with frequent lymphocytic infiltration organized in tertiary lymphoid structures (Fig. 2F-K and Table 1).

BAP1 tumors demonstrate a specific genomic profile with frequent *PRKACA* gain and *PRKAR2A* deletion

Analysis of the profile of gene mutations and chromosome alterations in BAP1 tumors showed a significant exclusion from alterations in the 3 most frequent drivers of HCC, namely *TERT* promoter, *CTNNB1* and *TP53*[22] (Fig. 1 and Table 1). This exclusion was confirmed in the TCGA cohort and also observed in *DNAJB1-PRKACA* fusion tumors (Fig. 1 and Table 1). In contrast, BAP1

JOURNAL
OF HEPATOLOGY

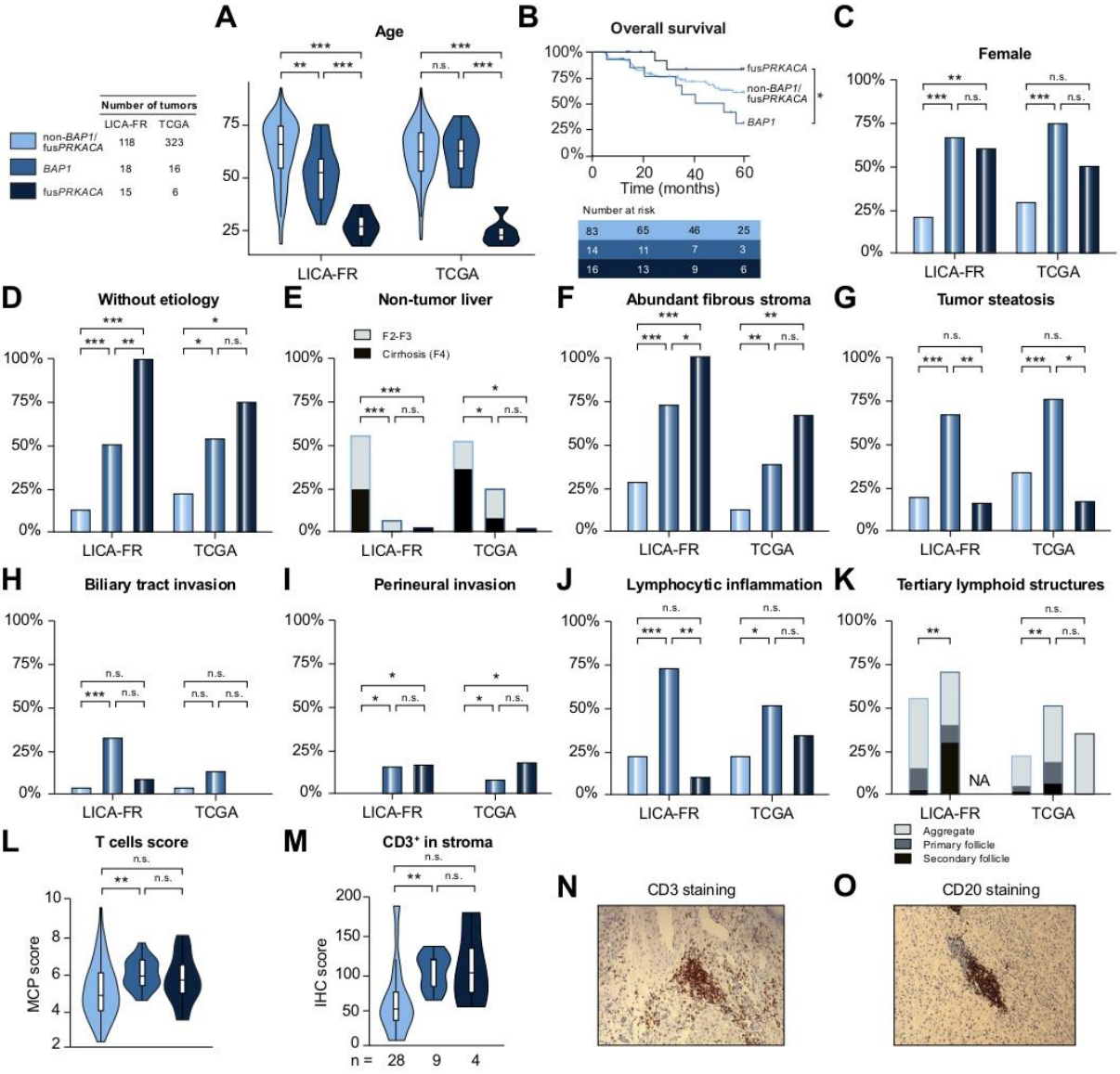


Fig. 2. Clinical and histological features of BAP1, fusPRKACA and non-BAP1/fusPRKACA tumors. (A) Distribution of age at surgery. (B) Kaplan-Meier plot showing the overall survival in the LICA-FR series. (C-E) Repartition of female gender (C), absence of liver disease etiology (D) and non-tumor liver histology (F2-F3: mild fibrosis; F4: cirrhosis) (E). (F-K) Histological review of abundant fibrous stroma (F), tumor steatosis (G), biliary tract invasion (H), perineural invasion (I), lymphocytic inflammation (J) and tertiary lymphoid structures (K). (L) T cell score calculated with the MCP-counter method applied to RNAseq data of the LICA-FR series. (M) Quantification of T cell abundance in tumor stroma with immunohistochemistry in a subset of tumors from the LICA-FR series. (N, O) Representative images of CD3 (N) and CD20 (O) staining in BAP1 tumors (#411T and #412T respectively), showing the aggregation of T and B cells within aggregate type tertiary lymphoid structures. Wilcoxon test, trend χ^2 test, Fisher exact test and Log-rank test were used respectively for continuous (A, L), ordinal (C, K), binary (B, D, F-J) and survival data (E): $^{***}p$ <0.001, $^{**}p$ <0. 01, $^{*}p$ <0.05. MCP, microenvironment cell population; RNAseq, RNA sequencing. (This figure appears in color on the web.)

tumors were slightly enriched in *RB1*, *KMT2D*, *ATM* and *ALB* mutations (n = 3 for each gene, Fig. 1) but these associations did not reach significance.

In order to find alternative driver events associated with *BAP1* alterations in HCC, we compared the frequency of copy number alterations between the 3 molecular groups of tumors. In 15 out of the 18 BAP1 tumors (83%) we identified a specific recurrent chromosome gain at the 19p13 locus centered on the *PRKACA* gene found in only 6% of the non-BAP1/fusPRKACA tumors (p = 3e-12) and in none of the fusPRKACA tumors (Fig. 3A-B). In addition, deletions at chromosome 3p21 identified in the BAP1 tumors encompassed both *BAP1* and *PRKAR2A* (coding for

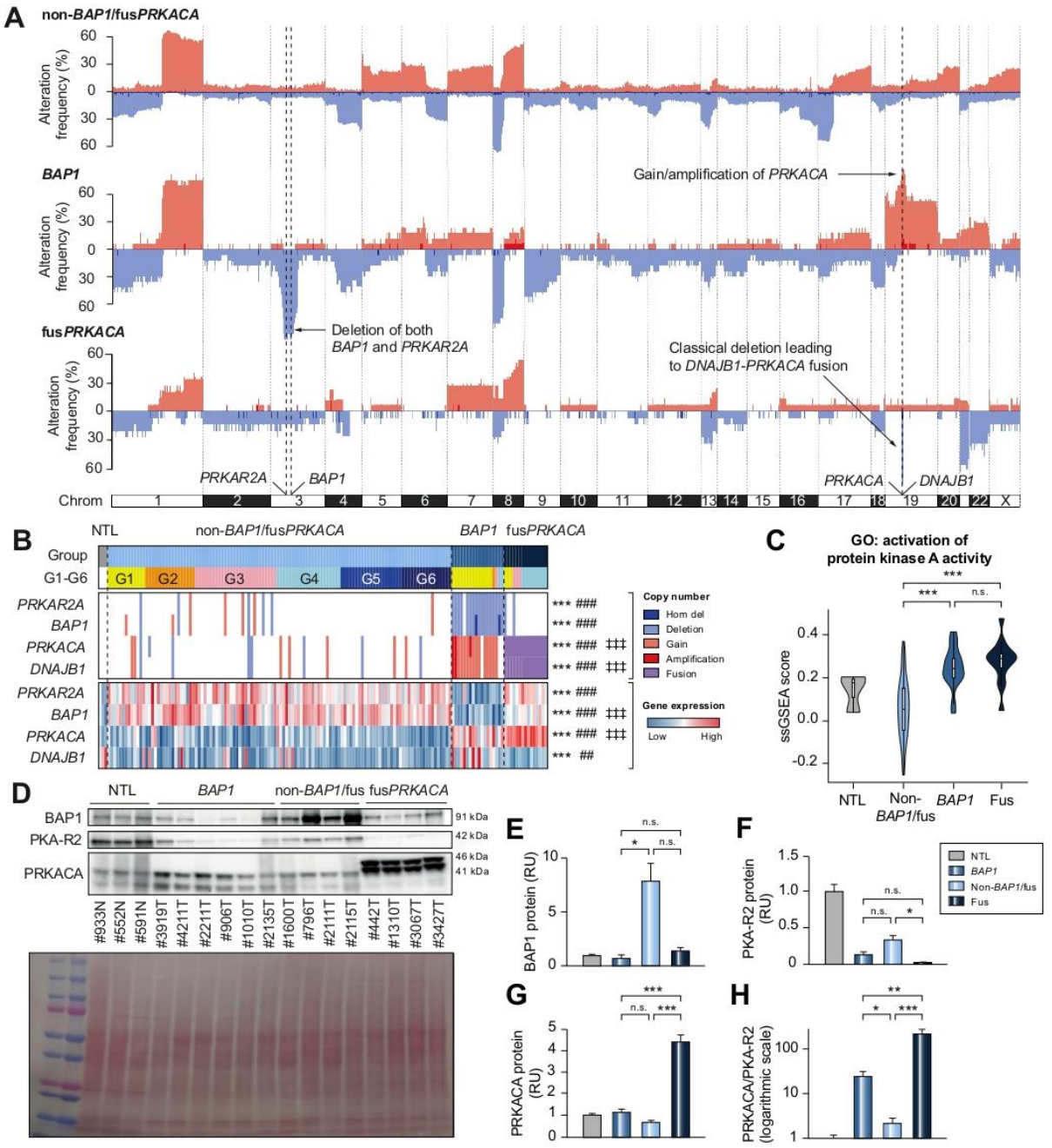


Fig. 3. Alternative copy number alterations activate the PKA pathway in BAP1 tumors. (A) Frequency of copy number alterations in the 3 molecular groups of HCC. (B) Heatmap representation of copy number alterations and relative gene expression for *PRKAR2A*, *BAP1*, *PRKACA* and *DNAJB1* assessed by RNAseq. Statistical comparisons with exact Fisher test (copy number alterations) and Wilcoxon test (gene expression): *BAP1 *vs.* non-BAP1/fusPRKACA, #BAP1 *vs.* fusPRKACA, ‡fusPRKACA *vs.* non-BAP1/fusPRKACA. (C) Comparison of single sample GSEA scores for the gene set GO: Activation of PKA activity between the 3 molecular groups of tumors. (D) Western-blot of BAP1, PKA-R2 and PRKACA in normal liver samples (NTL, n = 3) and in different liver tumors: BAP1 (n = 6), non-BAP1/fusPRKACA (n = 4), fusPRKACA (n = 4). In fusPRKACA tumors, the higher mass of the bands stained by the PRKACA antibody corresponds to the already described DNAJB1-PRKACA chimeric protein. Bottom image represents Ponceau coloration. (E-H) Quantification of protein expression of BAP1 (E), PKA-R2 (F), PRKACA (G) and of relative PKA activation measured by the ratio between PRKACA and PKA-R2 (H), compared with Wilcoxon test. ***, ###, ‡‡‡ $p < 0.001$; **, ## $p < 0.01$; *$p < 0.05$. GO, gene ontology; GSEA, gene-set enrichment analysis; HCC, hepatocellular carcinoma; NTL, non-tumor liver; RNAseq, RNA sequencing. (This figure appears in color on the web.)

JOURNAL
OF HEPATOLOGY

PKA-R2α, a negative regulatory subunit of the cAMP-dependent protein kinase PKA) genes distant by 3.5 Mb (94% *vs.* 4% in non-BAP1/fusPRKACA tumors, $p = 2e\text{-}16$, Fig. 3A-B and Fig. S4). In agreement with these findings, RNAseq data showed significantly lower *PRKAR2A* and higher *PRKACA* gene expression in BAP1 tumors compared to non-BAP1/fusPRKACA tumors and these results were confirmed at the protein level (Fig. 3B,D-F). A significantly increased ratio of PRKACA to PKA-R2 proteins, reflecting PKA activity, was monitored in both BAP1 and fusPRKACA groups of tumors compared to non-BAP1/fusPRKACA HCC (Fig. 3H). Accordingly, PKA activity assessed by single sample GSEA was increased in both BAP1 and fusPRKACA tumors (Fig. 3C). These results suggest that like FLC with *DNAJB1-PRKACA* fusion, the PKA pathway is activated in BAP1 tumors but via an alternative mechanism associated with *PRKACA* chromosome gains and *PRKAR2A* deletions (Fig. S4).

BAP1 and fusPRKACA tumors share common proliferation and differentiation programs, distinct from other hepatocellular tumors

At the transcriptomic level, 78% of the BAP1 tumors were classified in the G1 subgroup of HCC, known to be associated with progenitor features,[17,28] compared to only 11% of the non-BAP1/fusPRKACA tumors ($p = 9.5e\text{-}9$) and 20% of the fusPRKACA FLC and these results were confirmed in TCGA (63% *vs.* 14%, $p = 7.0e\text{-}5$, Table 1). Single sample GSEA revealed that BAP1 and fusPRKACA tumors were highly enriched for stem cell and neuronal features, vasculature development, extracellular matrix and epithelial-mesenchymal transition compared to non-BAP1/fusPRKACA HCC (Fig. 4A). FusPRKACA tumors appeared to be less proliferative than non-BAP1/fusPRKACA tumors, as illustrated by lower levels of cell cycle genes such as *CCNB1*, *PCNA* or *BIRC5*, while the tendency toward lower proliferation in BAP1 tumors did not reach significance (Fig. 4B).

Interestingly, expression of the 3 TGF-β ligands was increased in both BAP1 and fusPRKACA tumors (Fig. 4B), together with an enrichment in the signature of TGF-β signaling in fibroblasts (Plasari TGFb1 Targets 10hr Up, Fig. 4A). In contrast, a signature of TGF-β activation in hepatocytes defined by Coulouarn *et al.*[38] was not retrieved in BAP1 nor fusPRKACA tumors, suggesting that activation of the TGF-β pathway in fibroblasts but not in tumor hepatocellular cells could be at the root of the fibrotic phenotype in both groups of tumors.

In accordance with the enrichment in stem cell features detected by GSEA, BAP1 and fusPRKACA tumors shared high expression of hepatic stem cell genes such as *EPCAM*, *VIM* or *THY1* (CD90) (Fig. 4B). Immunohistochemistry performed on a subset of tumors confirmed the higher percentage of EpCAM-positive progenitor cells within both BAP1 tumors and fusPRKACA tumors while this staining was absent in most non-BAP1/fusPRKACA tumors (Fig. 4C, D). Furthermore, we also monitored high expression of markers of the common hepato-pancreatic progenitor[39,40] such as *PDX1* or *SOX17* (Fig. 4B), 2 markers which have previously been detected by immunohistochemistry in most FLC samples.[5] Among the differentiation markers, we observed a significant decrease of hepatocyte-specific genes *ALB*, *PROX1* or *HNF4A* and a significant increase of cholangiocyte-specific gene *KRT7* in fusPRKACA compared to BAP1 or non-BAP1/fusPRKACA tumors (Fig. S5). Also, several markers of pancreatic lineage were overexpressed in BAP1 and/or fusPRKACA tumors, but at various levels in the 2 groups of tumors. For instance, among the pancreatic markers, high expression of *PDX1*, *PAX6*, *PAK3*, and *MNX1* was identified in BAP1 and fusPRKACA tumors while *NEUROG3* (NGN3), *INSM1* (IA-1), *PTF1A*, *TRIM50*, *GP2* and *TGIF2* overexpression was restricted to BAP1 tumors and only fusPRKACA tumors overexpressed the pancreatic/neuroendocrine gene *PCSK1* (Fig. 4B and Fig. S5). Other neuroendocrine genes such as *CALCA*, *NTS*, *DNER* or *SSTR5* were overexpressed with different patterns in the 2 groups of tumors (Fig. S5). Moreover, we observed a high expression of genes coding for proteins involved in neural infiltration and guidance, such as neuron guide molecules (*NGF*, *NTF3*, *BDNF*) and their binding receptors expressed on neurons (TrkA-C (*NTRK1-3*) and p75NTR (*NFGR*), Fig. S5), possibly accounting for the enrichment in perineural invasion in both BAP1 and fusPRKACA tumors (Fig. 2I). In line with a possible neural infiltration, other neuron expressed receptors such as UNC5A-D were found significantly overexpressed in BAP1 or fusPRKACA tumors (Fig. S5).

Lastly, the higher expression of the hepato-pancreatic progenitor marker PDX1 and the lack of expression of pancreatic-committed progenitor markers such as IA-1 or PTF1A in fusPRKACA compared to BAP1 tumors, suggest a hepato-pancreatic progenitor-like program in fusPRKACA tumors and a more pancreatic-engaged program in BAP1 tumors (Fig. 4B and Fig. 5).

Discussion

In this study, we showed that *BAP1* alterations delineate a specific subgroup of HCC with common clinical, histological and molecular features (Fig. 5A). BAP1 is a deubiquitinating enzyme active in both the cytosol, where it controls the stability of different proteins, and in the nucleus, where it targets H2A histones as part of the polycomb group repressive deubiquitinase complex, which is involved in development and stem cell pluripotency.[41,42] BAP1 is described as a tumor suppressor since it is lost in different tumors including cholangiocarcinoma (around 25%) and HCC (around 5%).[16,34–37] However BAP1 was recently shown to also have an anti-apoptotic role in the liver,[42] which could explain our observation of an increased expression of *BAP1* in HCC compared to normal liver (Fig. 3B,D-E). Here, we found that the BAP1 tumors include most of the mixed-FLC/HCC tumors and show several clinical and histological similarities with the classical FLC together with a common PKA pathway activation.

Fibrolamellar features in liver tumors are already known to be related to PKA activation through 3 different mechanisms: (i) the *DNAJB1-PRKACA* fusion in FLC,[6] (ii) inactivating mutations in *PRKAR1A* in rare FLC developed in Carney complex patients[7] and (iii) activating mutations in *GNAS* (leading to production of cAMP and subsequent activation of PKA) identified in rare hepatocellular adenomas characterized by fibrolamellar-like patterns.[43] Here, we discovered a fourth mechanism of PKA activation in BAP1 tumors resulting from recurrent gains of *PRKACA* and deletions of *PRKAR2A*, encoding respectively a catalytic subunit and an inhibitory regulatory subunit of PKA, with consequences at the mRNA and protein levels (Fig. 3 and Fig. S4).

An important characteristic of the *DNAJB1-PRKACA* and BAP1 tumors is their age distribution starting with FLC in the second and third decade, then BAP1 HCC occurring before non-BAP1/fusPRKACA HCC (Fig. 5A). This gradient could be related to the cell of origin of the malignant transformation since FLC has been proposed to originate from the biliary tree stem cell,[5] a recently discovered hepato-pancreatic progenitor,[39,40] in agreement with

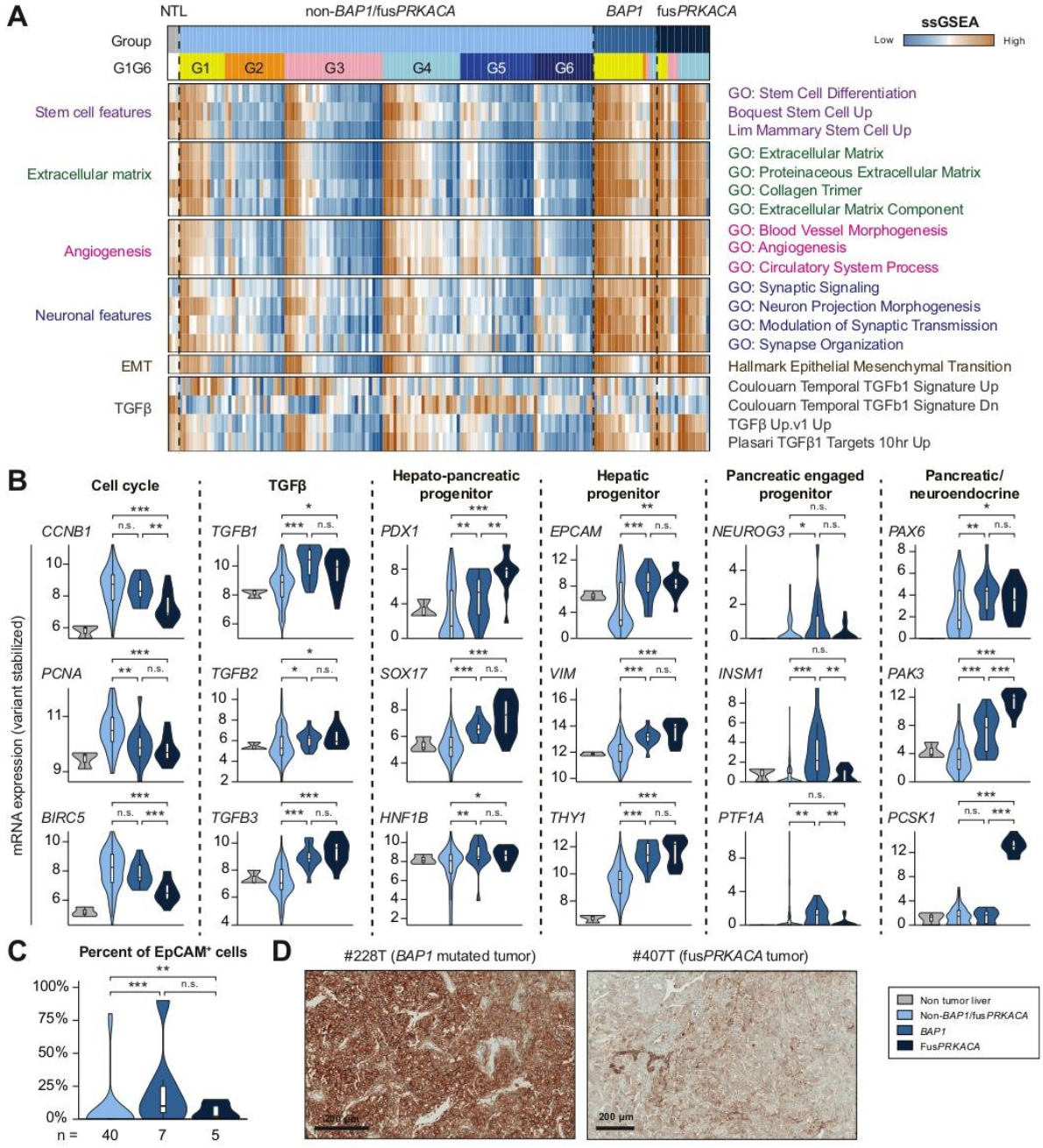


Fig. 4. Transcriptomic characterization of BAP1 and fusPRKACA tumors. (A) Heatmap of single sample GSEA for 3 normal liver samples and the different molecular groups of liver tumors ordered by G1-G6 classes. (B) RNAseq expression of representative cell cycle genes, TGFβ ligands as well as representative markers of the different progenitors found in the hepato-pancreatic lineage and of pancreatic / neuroendocrine markers. (C) Immunohistochemistry reveals a higher percentage of EpCAM positive tumor cells in both BAP1 and fusPRKACA tumors. (D) Representative images of EpCAM staining show the lighter staining in fusPRKACA tumors compared to BAP1 tumors. Wilcoxon test: ***p <0.001, **p <0.01, *p <0.05. GSEA, gene-set enrichment analysis. (This figure appears in color on the web.)

the hepato-pancreatic progenitor phenotype we identified in both fusPRKACA and BAP1 tumors (Fig. 5B). BAP1 loss may promote a dedifferentiation towards a progenitor phenotype since *BAP1* silencing in the Huh7 HCC cell line induces an increase in stemness marker expression, *e.g. EPCAM* or *PROM1*.[44] Similarly, CRISPR-based loss of *BAP1* in human liver organoids increases the expression of progenitor markers while reducing markers of liver function and ultimately epithelial homeostasis,

JOURNAL
OF HEPATOLOGY

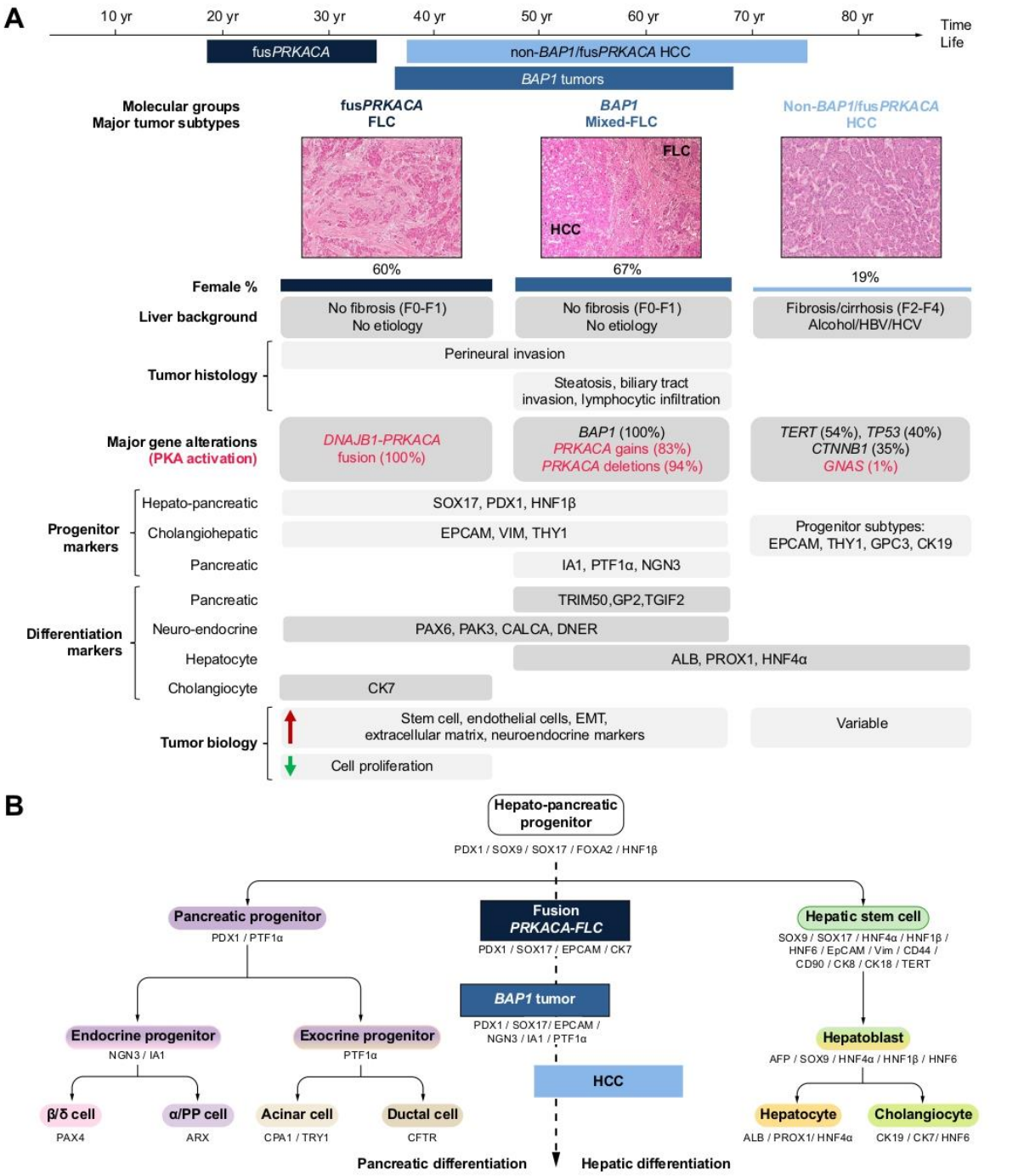


Fig. 5. Summarized characteristics of fusPRKACA and BAP1 tumors compared to non-BAP1/fusPRKACA HCC and relation to the common hepato-pancreatic lineage. (A) A recap of the most significant results in terms of histological, clinical and molecular features was represented. Age was represented from the 1st to 9th deciles. (B) Schematic differentiation program from the common hepato-pancreatic progenitor, adapted from Lanzoni *et al.*[40] FusPRKACA FLC are thought to be directly derived from hepato-pancreatic progenitors found in the young liver while BAP1 tumors have a more pancreatic-engaged program. FLC, fibrolamellar carcinoma; HCC, hepatocellular carcinoma. (This figure appears in color on the web.)

possibly via a direct regulation of transcription by BAP1 through its known role in the polycomb deubiquitinase complex.[42] In the TCGA cohort, the unique patient with HCC harboring both *DNAJB1-PRKACA* fusion and *BAP1* inactivation was old (67 years old), suggesting that the progenitor phenotype induced by BAP1 loss could be required for the oncogenic effect of PKA activation in an old liver. This is in line with our observation of a co-occurrence of BAP1 inactivation and PKA activation through copy number alterations. Remarkably, 2 independent teams recently reported the presence of *DNAJB1-PRKACA* fusions in rare pancreatic and biliary neoplasms, thus reinforcing the link between the oncogenic activation of PKA and the hepato-pancreatic progenitor lineage, while challenging the exclusivity between FLC and the *DNAJB1-PRKACA* fusion.[45,46]

Of note, HCC usually lacks lymph node involvement or perineural invasion, commonly found in other epithelial tumors such as pancreatic adenocarcinomas[47,48] or cholangiocarcinoma.[49,50] Conversely, there is a high rate of lymph node metastasis in FLC (up to 70%)[51] and here we identified an enrichment of perineural invasion in BAP1 and fusPRKACA tumors across the 2 series (Table 1). This enrichment could be a consequence of the overexpression of genes encoding neuron guide molecules (*NGF*, *NTF3*, *BDNF*) that allow the recruitment and growth of neurons expressing axonal guidance receptors (TrkA-C, p75NTR, UNC5A-D, see Fig. S5).

Strikingly, BAP1 tumors are never mutated in the *TERT* promoter (in both LICA-FR and TCGA cohort) while this alteration is found in 30–60% of HCC.[16,52] This could also be a consequence of the progenitor state of the tumor cells associated with BAP1 loss, since *in vitro* experiments suggest that *BAP1* overexpression can repress *TERT* transcription[53] and progenitors cells are supposed to maintain their telomeres throughout replication without requiring mutations. This would also explain the exclusion between *DNAJB1-PRKACA* fusion and *TERT* alteration, following the hypothesis that the cell of origin of FLC is a progenitor.[5]

While it has been suggested that endocrine/pancreatic features were specific of FLC and not found in mixed-FLC/HCC,[13] here we show that both fusPRKACA and BAP1 tumors express high levels of neuroendocrine and pancreatic markers but with different patterns (Fig. S5). These different profiles of expression could reflect different stages of differentiation along the hepato-pancreatic lineage in the 2 groups of tumors, as a consequence of their different cell of origin, as discussed above (Fig. 5B). Moreover, an association between neuroendocrine features and abundant fibrous stroma has already been described.[54]

Interestingly, *BAP1* mutations are recurrently identified in several tumor types including uveal or cutaneous melanoma, mesothelioma, renal cell carcinoma, mesothelioma and lung cancers.[55] In 2013, a pan-cancer study defined a typical morphology of *BAP1*-mutated cells as “rounded or polygonal cell shape with abundant amphophilic or eosinophilic cytoplasm”.[56] Interestingly, these features are also retrieved in FLC[1,4] as well as in BAP1 liver tumors (in both FLC and HCC components). This, along with the abundant fibrous stroma, hinders the distinction between BAP1 and fusPRKACA tumors based solely on histology in our cohort of tumors and in other reports.[13,57] These observations argue for the use of molecular characterization to distinguish BAP1 tumors not only from the fusPRKACA FLC but also from scirrhous HCC, another type of highly fibrous liver tumor usually not mutated in BAP1 (Fig. 1 and Table 1). However, it is worth noting that the detection of *BAP1* mutations is challenging due to the high level of stromal contamination: in our series, 3 out of the 15 *BAP1* mutations were not found at first in the analysis of whole exome data, but were later detected in target sequencing of *BAP1* with a higher depth.

From a clinical perspective, our observations of similarities between fusPRKACA and BAP1 tumors may be in favor of developing common therapeutic strategies. Indeed, the over-activation of PKA in both groups of tumors argues that potential future treatments targeting PKA, if successful in FLC, could be repositioned to also combat BAP1 tumors. From a therapeutic perspective, the 2 groups of tumors could benefit from anti-angiogenic drugs such as sorafenib[58] due to their common enrichment in markers of angiogenesis and endothelial cells (Fig. 4A and Fig. S2). Of note, a link between *BAP1* alterations and angiogenesis was also recently described in uveal melanoma.[59] Finally, as BAP1 HCC are highly infiltrated by lymphocytes (Fig. 2J), a feature also described in *BAP1*-mutated peritoneal mesothelioma,[60] they can be considered as good candidates for immunotherapy. Of note, despite this higher lymphocytic infiltration, we did not monitor a difference in the tumor mutational burden of BAP1 tumors compared to non-BAP1/fusPRKACA tumors, while both groups of tumors had more mutations than fusPRKACA tumors (Fig. S2B).

In conclusion, we characterized a new molecular subgroup of HCC, driven by *BAP1* loss of function and PKA activation in the context of hepato-pancreatic progenitor de-differentiation, associated with a specific pathological subtype of tumors, including abundant fibrous stroma, steatosis, lymphocyte infiltration, perineural and biliary tract invasion, which develop mainly in female patients without cirrhosis or chronic liver disease.

Abbreviations

AUC, area under the ROC curve; *BAP1*, *BRCA1*-associated protein 1; FLC, fibrolamellar carcinoma; GO, gene ontology; GSEA, gene-set enrichment analysis; HCC, hepatocellular carcinoma; MCP, microenvironment cell populations; NASH, non-alcoholic steatohepatitis; NTL, non-tumor liver; PKA, cAMP-dependent protein kinase; RNAseq, RNA-sequencing; WES, whole-exome sequencing; WGS, whole-genome sequencing.

Financial support

This work was supported by Institut National du Cancer (INCa) with the International Cancer Genome Consortium (ICGC LICA-FR project), INSERM with the « Cancer et Environnement » (plan Cancer), MUTHEC projects (INCa), and the HTE consortium, HetCoLi (plan Cancer). The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir), Coup d'Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM, FRM prix Rosen, Ligue contre le cancer comité de Paris (prix René et André Duquesne) and Fondation Mérieux.

Conflict of interest

The authors declare no conflicts of interest that pertain to this work.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Study concept and design: JZR, TZH, AN. Acquisition of data: JZR, TZH, AN, BG, SC, BN, GC, QB, LM, GM, JYC, JFB, GA, JCN, PBS, MZ,

JOURNAL
OF HEPATOLOGY

AB, VP, JC, SI. Analysis and data interpretation: TZH, AN, SC, QB, LM, JZR. Manuscript writing: JZR, TZH, AN and all. Obtained funding: JZR.

Acknowledgments

We warmly thank the team UMR-1138 for all the advice and support in particular Sandra Rebouissou and Eric Letouzé for helpful scientific discussions, Eric Trepo for the help in bioinformatics, Camille Péneau and Jie Yang for experimental help. A special thanks to all the clinician, surgeon, pathologist, hepatologist and oncologist who contributed to the tissue collection and clinical annotations within the GENTHEP network. T.Z.H. was supported by a fellowship from Cancéropole Ile de France and Fondation d'Entreprise Bristol-Myers Squibb pour la Recherche en Immuno-Oncologie, SC was funded by CARPEM and the Labex OncoImmunology, AN was funded by the Labex OncoImmunology, QB and LM were funded by a FRM and MRESI fellowships.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2019.12.006.

References

Author names in bold designate shared co-first authorship

[1] Craig JR, Peters RL, Edmondson HA, Omata M. Fibrolamellar carcinoma of the liver: a tumor of adolescents and young adults with distinctive clinico-pathologic features. Cancer 1980;46:372–379.
[2] Edmondson HA. Differential diagnosis of tumors and tumor-like lesions of liver in infancy and childhood. AMA J Dis Child 1956;91:168–186.
[3] Kassahun WT. Contemporary management of fibrolamellar hepatocellular carcinoma: diagnosis, treatment, outcome, prognostic factors, and recent developments. World J Surg Oncol 2016;14:151.
[4] Lin C-C, Yang H-M. Fibrolamellar carcinoma: a concise review. Arch Pathol Lab Med 2018;142:1141–1145.
[5] **Oikawa T**, **Wauthier E**, Dinh TA, Selitsky SR, Reyna-Neyra A, Carpino G, et al. Model of fibrolamellar hepatocellular carcinomas reveals striking enrichment in cancer stem cells. Nat Commun 2015;6:8070.
[6] **Honeyman JN**, **Simon EP**, **Robine N**, Chiaroni-Clarke R, Darcy DG, Lim IIP, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. Science 2014;343:1010–1014.
[7] Graham RP, Lackner C, Terracciano L, González-Cantú Y, Maleszewski JJ, Greipp PT, et al. Fibrolamellar carcinoma in the Carney complex: PRKAR1A loss instead of the classic DNAJB1-PRKACA fusion. Hepatology 2018;68:1441–1447.
[8] Nault J-C, Martin Y, Caruso S, Hirsch TZ, Bayard Q, Calderaro J, et al. Clinical impact of genomic diversity from early to advanced hepatocellular carcinoma. Hepatology 2020;71(1):164–182.
[9] **Nakamura H**, **Arai Y**, **Totoki Y**, **Shirota T**, **Elzawahry A**, Kato M, et al. Genomic spectra of biliary tract cancer. Nat Genet 2015;47:1003–1010.
[10] Malouf G, Falissard B, Azoulay D, Callea F, Ferrell LD, Goodman ZD, et al. Is histological diagnosis of primary liver carcinomas with fibrous stroma reproducible among experts? J Clin Pathol 2009;62:519–524.
[11] Malouf GG, Brugières L, Le Deley MC, Faivre S, Fabre M, Paradis V, et al. Pure and mixed fibrolamellar hepatocellular carcinomas differ in natural history and prognosis after complete surgical resection. Cancer 2012;118:4981–4990.
[12] Calderaro J, Ziol M, Paradis V, Zucman-Rossi J. Molecular and histological correlations in liver cancer. J Hepatol 2019;71:616–630.
[13] **Malouf GG**, **Job S**, Paradis V, Fabre M, Brugières L, Saintigny P, et al. Transcriptional profiling of pure fibrolamellar hepatocellular carcinoma reveals an endocrine signature. Hepatology 2014;59:2228–2237.
[14] Malouf GG, Tahara T, Paradis V, Fabre M, Guettier C, Yamazaki J, et al. Methylome sequencing for fibrolamellar hepatocellular carcinoma depicts distinctive features. Epigenetics 2015;10:872–881.
[15] **Nault JC**, **De Reyniès A**, Villanueva A, Calderaro J, Rebouissou S, Couchy G, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology 2013;145:176–187.
[16] 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.
[17] Calderaro J, Couchy G, Imbeaud S, Amaddeo G, Letouzé E, Blanc J-F, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. J Hepatol 2017;67:727–738.
[18] Ziol M, Poté N, Amaddeo G, Laurent A, Nault J-C, Oberti F, et al. Macrotrabecular-massive hepatocellular carcinoma: a distinctive histological subtype with clinical relevance. Hepatology 2018;68:103–112.
[19] Murakami J, Shimizu Y, Kashii Y, Kato T, Minemura M, Okada K, et al. Functional B-cell response in intrahepatic lymphoid follicles in chronic hepatitis C. Hepatology 1999;30:143–150.
[20] Calderaro J, Petitprez F, Becht E, Laurent A, Hirsch TZ, Rousseau B, et al. Intra-tumoral tertiary lymphoid structures are associated with a low risk of early recurrence of hepatocellular carcinoma. J Hepatol 2019;70:58–65.
[21] **Guichard C**, **Amaddeo G**, **Imbeaud S**, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694–698.
[22] **Schulze K**, **Imbeaud S**, **Letouzé E**, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505–511.
[23] Popova T, Manié E, Stoppa-Lyonnet D, Rigaill G, Barillot E, Stern MH. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome Biol 2009;10:R128.
[24] Bayard Q, Meunier L, Peneau C, Renault V, Shinde J, Nault J-C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. Nat Commun 2018;9:5235.
[25] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013;14:R36.
[26] Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–169.
[27] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
[28] **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.
[29] Calderaro J, Meunier L, Nguyen CT, Boubaya M, Caruso S, Luciani A, et al. ESM1 as a marker of macrotrabecular-massive hepatocellular carcinoma. Clin Cancer Res 2019;25(19):5859–5865.
[30] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[31] **Subramanian A**, **Tamayo P**, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–15550.
[32] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
[33] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218.
[34] Harbour JW, Onken MD, Roberson EDO, Duan S, Cao L, Worley LA, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. Science 2010;330:1410–1413.
[35] Bott M, Brevet M, Taylor BS, Shimizu S, Ito T, Wang L, et al. The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. Nat Genet 2011;43:668–672.
[36] Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, Pavía-Jiménez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. Nat Genet 2012;44:751–759.
[37] **Jiao Y**, **Pawlik TM**, **Anders RA**, Selaru FM, Streppel MM, Lucas DJ, et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. Nat Genet 2013;45:1470–1473.
[38] Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-β gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059–2067.

[39] **Cardinale V**, **Wang Y**, Carpino G, Cui C-B, Gatto M, Rossi M, et al. Multipotent stem/progenitor cells in human biliary tree give rise to hepatocytes, cholangiocytes, and pancreatic islets. Hepatology 2011;54:2159–2172.
[40] **Lanzoni G**, **Cardinale V**, Carpino G. The hepatic, biliary, and pancreatic network of stem/progenitor cell niches in humans: a new reference frame for disease and regeneration. Hepatology 2016;64:277–286.
[41] **Scheuermann JC**, **de Ayala Alonso AG**, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, et al. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature 2010;465:243–247.
[42] Artegiani B, van Voorthuijsen L, Lindeboom RGH, Seinstra D, Heo I, Tapia P, et al. Probing the tumor suppressor function of BAP1 in CRISPR-engineered human liver organoids. Cell Stem Cell 2019;24(6):927–943.e6.
[43] Nault JC, Fabre M, Couchy G, Pilati C, Jeannot E, Tran Van Nhieu J, et al. GNAS-activating mutations define a rare subgroup of inflammatory liver tumors characterized by STAT3 activation. J Hepatol 2012;56:184–191.
[44] Woo HG, Choi J-H, Yoon S, Jee BA, Cho EJ, Lee J-H, et al. Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. Nat Commun 2017;8:839.
[45] Vyas M, Hechtman JF, Zhang Y, Benayed R, Yavas A, Askan G, et al. DNAJB1-PRKACA fusions occur in oncocytic pancreatic and biliary neoplasms and are not specific for fibrolamellar hepatocellular carcinoma. Mod Pathol 2019. https://doi.org/10.1038/s41379-019-0398-2 [Epub ahead of print].
[46] Singhi AD, Wood LD, Parks E, Torbenson MS, Felsenstein M, Hruban RH, et al. Recurrent rearrangements in PRKACA and PRKACB in intraductal oncocytic papillary neoplasms of the pancreas and bile duct. Gastroenterology 2020;158(3):573–582.e2.
[47] Liu B, Lu K-Y. Neural invasion in pancreatic carcinoma. Hepatobiliary Pancreat Dis Int 2002;1:469–476.
[48] Winter JM, Cameron JL, Campbell KA, Arnold MA, Chang DC, Coleman J, et al. 1423 pancreaticoduodenectomies for pancreatic cancer: a single-institution experience. J Gastrointest Surg 2006;10:1199–1210. discussion 1210-1.
[49] Bhuiya MR, Nimura Y, Kamiya J, Kondo S, Fukata S, Hayakawa N, et al. Clinicopathologic studies on perineural invasion of bile duct carcinoma. Ann Surg 1992;215:344–349.
[50] Pichlmayr R, Weimann A, Klempnauer J, Oldhafer KJ, Maschek H, Tusch G, et al. Surgical treatment in proximal bile duct cancer. A single-center experience. Ann Surg 1996;224:628–638.
[51] Yamashita S, Vauthey J-N, Kaseb AO, Aloia TA, Conrad C, Hassan MM, et al. Prognosis of fibrolamellar carcinoma compared to non-cirrhotic conventional hepatocellular carcinoma. J Gastrointest Surg 2016;20:1725–1731.
[52] Nault J-C, Ningarhari M, Rebouissou S, Zucman-Rossi J. The role of telomeres and telomerase in cirrhosis and liver cancer. Nat Rev Gastroenterol Hepatol 2019;8:294–296.
[53] Linne H, Yasaei H, Marriott A, Harvey A, Mokbel K, Newbold R, et al. Functional role of SETD2, BAP1, PARP-3 and PBRM1 candidate genes on the regulation of hTERT gene expression. Oncotarget 2017;8:61890–61900.
[54] Laskaratos F-M, Rombouts K, Caplin M, Toumpanakis C, Thirlwell C, Mandair D. Neuroendocrine tumors and fibrosis: an unsolved mystery? Cancer 2017;123:4770–4790.
[55] Masoomian B, Shields JA, Shields CL. Overview of BAP1 cancer predisposition syndrome and the relationship to uveal melanoma. J Curr Ophthalmol 2018;30:102–109.
[56] Murali R, Wiesner T, Scolyer RA. Tumours associated with BAP1 mutations. Pathology 2013;45:116–126.
[57] **Griffith OL**, **Griffith M**, **Krysiak K**, Magrini V, Ramu A, Skidmore ZL, et al. A genomic case study of mixed fibrolamellar hepatocellular carcinoma. Ann Oncol 2016;27:1148–1154.
[58] Wilhelm SM, Adnane L, Newell P, Villanueva A, Llovet JM, Lynch M. Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. Mol Cancer Ther 2008;7:3129–3140.
[59] Brouwer NJ, Gezgin G, Wierenga APA, Bronkhorst IHG, Marinkovic M, Luyten GPM, et al. Tumour angiogenesis in uveal melanoma is related to genetic evolution. Cancers (Basel) 2019;11:1–12.
[60] **Shrestha R**, **Nabavi N**, Lin Y-Y, Mo F, Anderson S, Volik S, et al. BAP1 haploinsufficiency predicts a distinct immunogenic class of malignant peritoneal mesothelioma. Genome Med 2019;11:8.

BAP1 mutations define a homogenous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA

Théo Z Hirsch, Ana Negulescu, Barkha Gupta, Stefano Caruso, Bénédicte Noblet, Gabrielle Couchy, Quentin Bayard, Léa Meunier, Guillaume Morcrette, Jean-Yves Scoazec, Jean-Frédéric Blanc, Giuliana Amaddeo, Jean-Charles Nault, Paulette Bioulac-Sage, Marianne Ziol, Aurélie Beaufrère, Valérie Paradis, Julien Calderaro, Sandrine Imbeaud, Jessica Zucman-Rossi

Table of contents

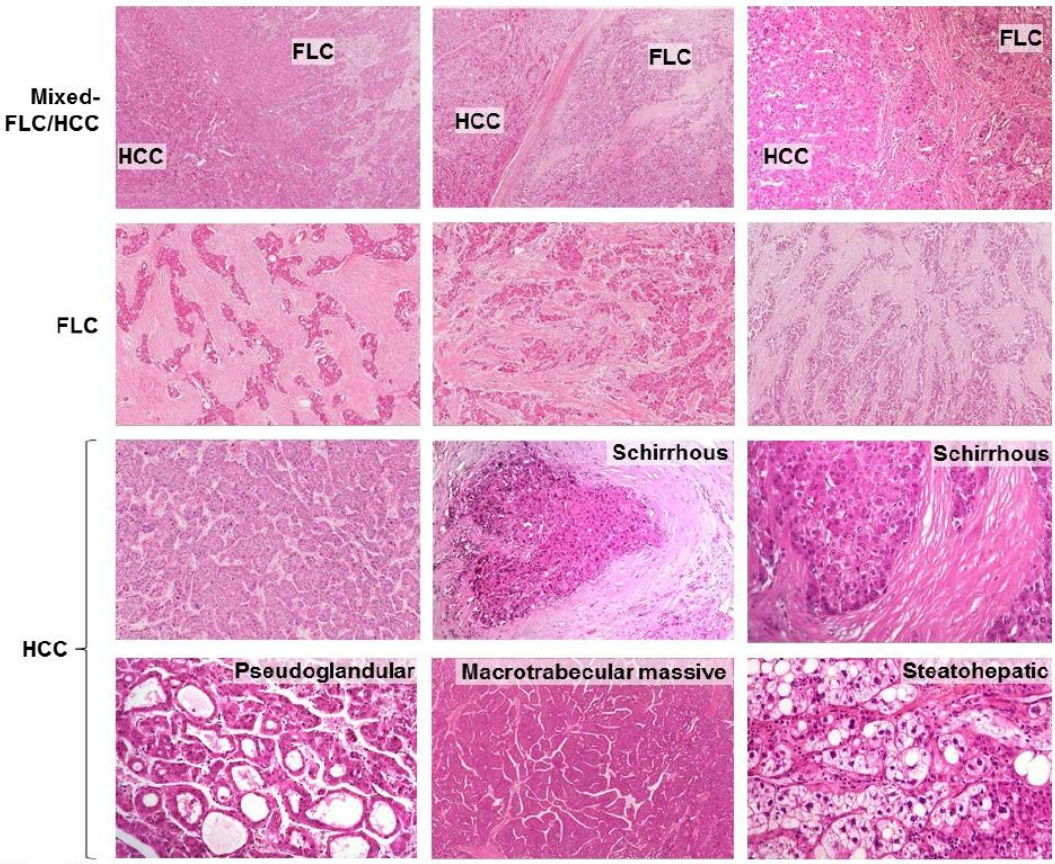


Fig. S1. Representative hematein-eosin-safran staining of different HCC subtypes

Mixed-FLC/HCC are defined as HCC containing at least one area with FLC features. Examples of distinct histological HCC subtypes (scirrhous, macrotrabecular massive, steatohepatic) as well as pseudoglandular pattern are shown.

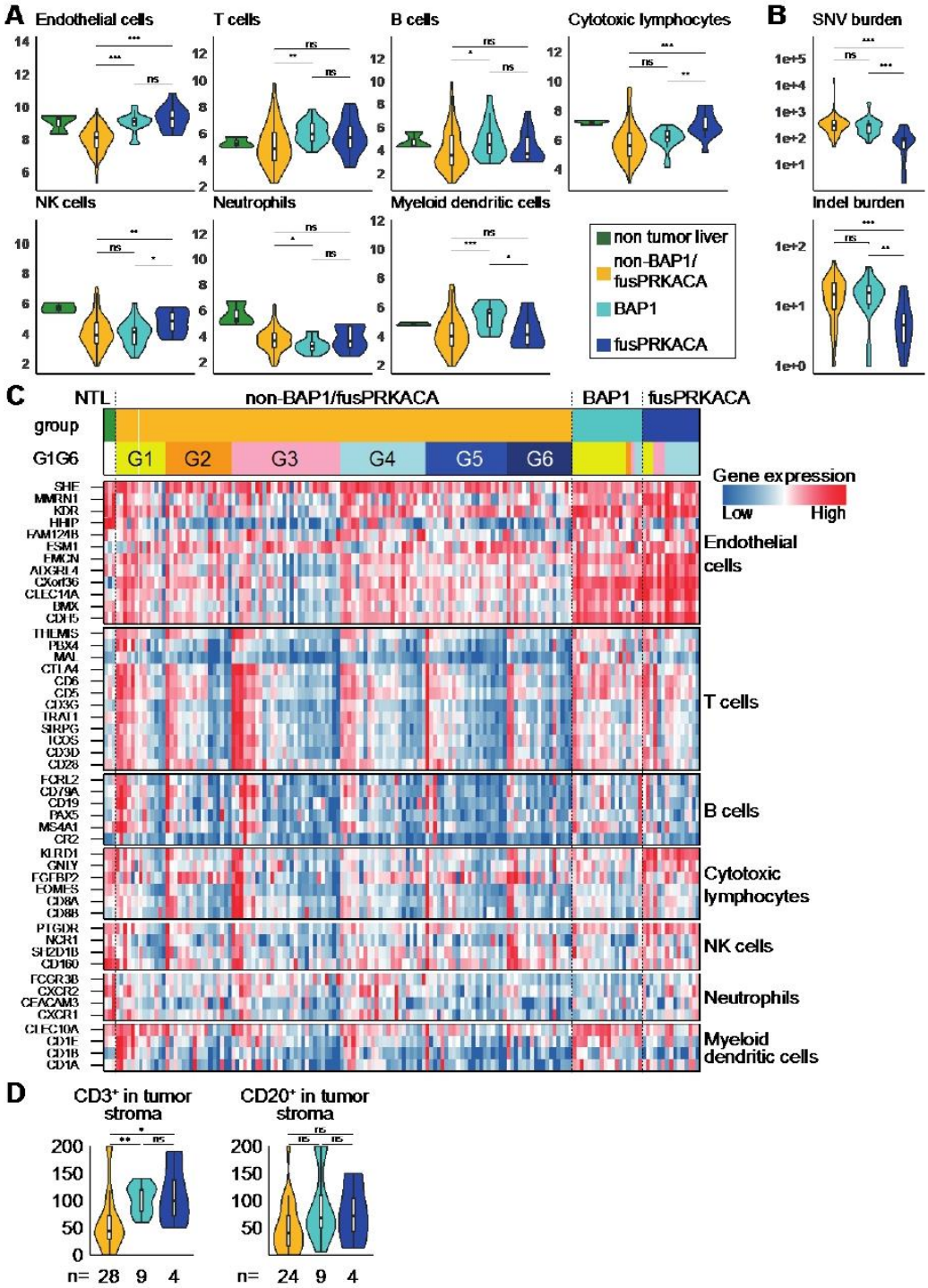


Fig. S2. RNAseq analysis of immune infiltrate

(A) Different immune scores calculated with the Microenvironment Cell Populations-counter method applied to RNAseq data. (B) Comparison of the tumor mutational burden (based on the number of mutations in whole-exome-sequencing with a variant allele frequency above 10%) split between single nucleotide variants (SNV, top) and

small insertions and deletions (indel, bottom). (C) Detailed expression of immune related genes used for determining Microenvironment Cell Populations scores in all samples of the LICA-FR cohort, ordered by molecular group, G1-G6 and *CD45* expression. (D) Quantification of $CD3^+$ T cells and $CD20^+$ B cells abundance in tumor stroma. Wilcoxon test: ***p-val < 0.001, **p-val < 0. 01, *p-val < 0.05.

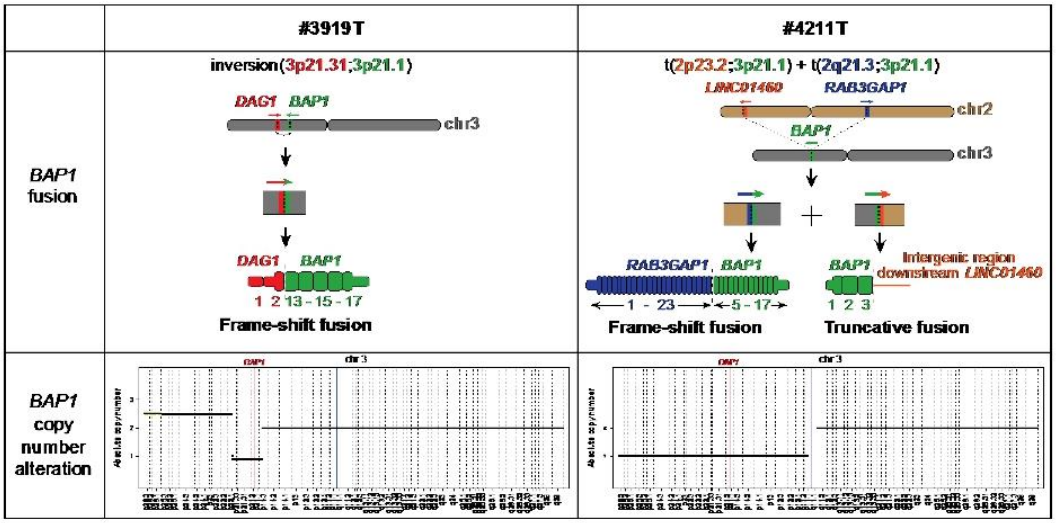


Fig. S3. Biallelic inactivation of *BAP1* through gene fusion and copy number alteration in 2 patients

In tumor #3919T, an unbalanced inversion between locus 3p21.31 and 3p21.1 gave rise to a frameshift fusion between *DAG1* and *BAP1* (top panel), while the second *BAP1* allele was lost in a focal deletion (lower panel). In tumor #4211T, a complex event implicating chromosome 2 and 3 led to two translocations, t(2p23.2;3p21.1) and t(2q21.3;3p21.1), with two subsequent gene fusions: a truncative fusion between *BAP1* and the intergenic region downstream to *LINC01460* and a frameshift fusion between *RBA3GAP1* and *BAP1*. The other *BAP1* allele was lost together with the whole 3p arm (lower panel).

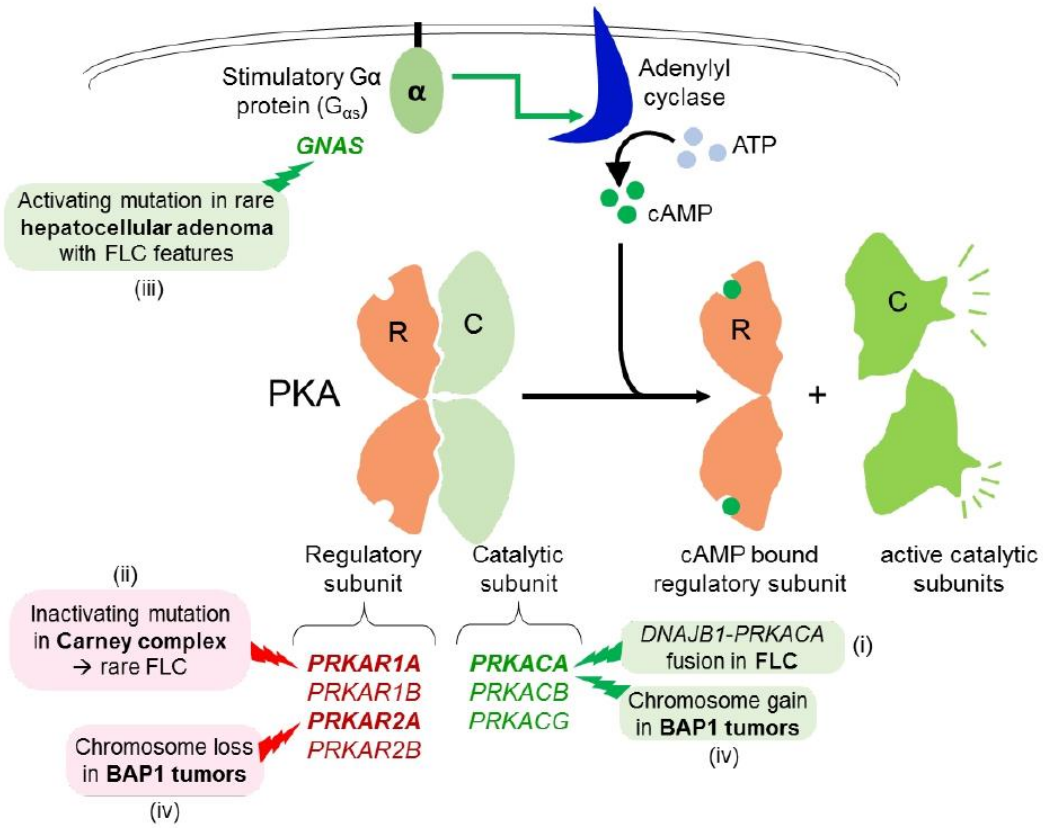


Fig. S4. Different mechanism of PKA activation in hepatocellular tumors
PKA (cAMP-dependent protein kinase) is composed of regulatory subunits (encoded by genes *PRKAR1A*/*B* and *PRKAR2A*/*B*) which normally repress the catalytic subunits (encoded by *PRKACA*/*B*/*C*). When cAMP bounds to the regulatory subunits, the catalytic subunits become active. Alpha subunits of stimulatory G-protein ($G_{αs}$, encoded by *GNAS*) activates the adenylyl cyclase, leading to production of cAMP and subsequent activation of PKA. Over-activation of PKA in liver tumors is associated with fibrolamellar features in 4 settings: (i) the *DNAJB1-PRKACA* fusion in FLC, (ii) inactivating mutations in *PRKAR1A* in rare FLC developed in Carney complex patients, (iii) activating mutations in *GNAS* in rare hepatocellular adenomas characterized by fibrolamellar-like patterns and (iv) recurrent gains of *PRKACA* and deletions of *PRKAR2A* in BAP1 tumors.

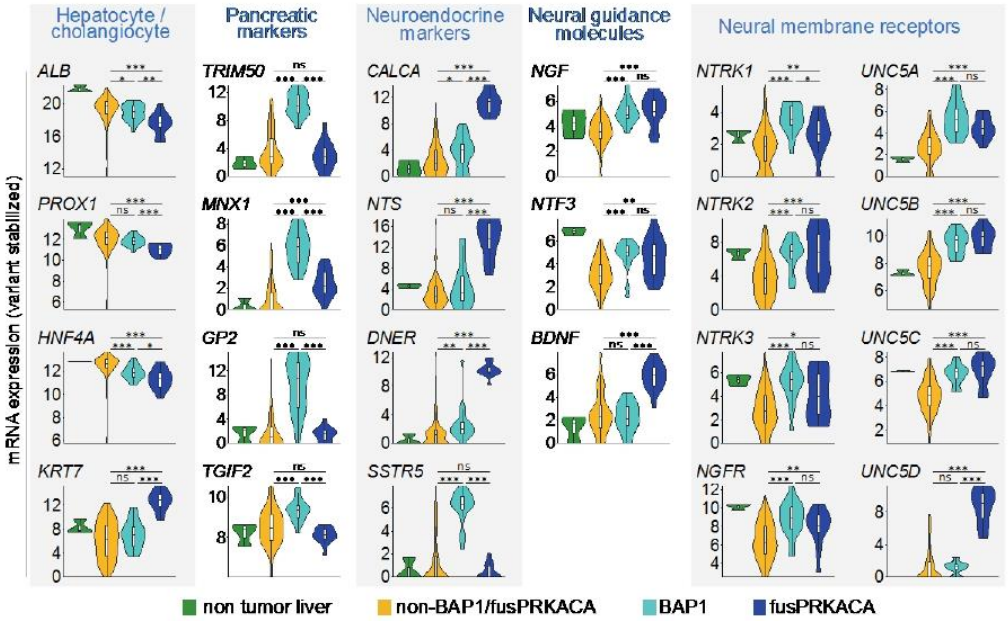


Fig. S5. mRNA expression of lineage and differentiation markers

mRNA expression of markers of hepatocytes, cholangiocytes (*KRT7*), pancreas, neuroendocrine markers, neural guidance molecules and receptors located on the membrane of the neurons assessed by RNAseq in normal liver (n=3), non-BAP1/fusPRKACA (n=118), BAP1 (n=18) and fusPRKACA (n=15) tumors. Wilcoxon test: ***p-val < 0.001, **p-val < 0. 01, *p-val < 0.05.

Table S1. Primers used for BAP1 sequencing in miSeq technology.

Amplicon name	Exon	Forward primer sequence w/o adapter	Reverse primer sequence w/o adapter
BAP1_Ex01_Amp01x1	exon01	CCATCCGGCCTCCCCAG	GCCGTCGCGCCTTGG
BAP1_Ex01_Amp01x2	exon01	AGGGAGAAAAGGCTCTTACCGA	GTTGTCTGTGTGTGGGACTGAG
BAP1_Ex01_Amp02x1	exon01	CCCGACCCTCCCCTTC	CTCAGTCCCACACACAGACAAC
BAP1_Ex01_Amp02x2	exon01	CTCAGTCCCACACACAGACAAC	CCCCTCTTCCCTTCGCCC
BAP1_Ex01_Amp03x1	exon01	GGGCGAAGGGGAGGGAC	CGCCCAGCACTTCCCG
BAP1_Ex01_Amp03x2	exon01	ACCTGCGCCCAGCACT	GCGCTCGAAGGCGAAC
BAP1_Ex02_Amp01x1	exon02	TGAGGAAAGGAAAGCAGTAGGGA	GCTGGGGAGGCCGGAT
BAP1_Ex02_Amp01x2	exon02	GGTCGTAGATCTCCTCCACTTG	CCCACCCAGGCCTCTTCA
BAP1_Ex02_Amp02x2	exon02	ATGAATAAGGGCTGGCTGGAG	CGATGAGGAAAGGAAAGCAGTA
BAP1_Ex03_Amp01x1	exon03	TAGGGTTCCTGGCACTGTCTTC	CTTTTCTCCCTGCCGGACC
BAP1_Ex03_Amp01x2	exon03	GCGGGCACTCAGAGAGCAA	GTCAAGGGGGTGCAAGTGG
BAP1_Ex03_Amp02x2	exon03	CAGCACTCTGGGTGTAAGGGG	AGCCTTTTCTCCCTGCCGG
BAP1_Ex04_Amp01x1	exon04	CCCTCCAAACAAAGCACAGAGT	TCTTCTCCCCTTTGGCTGATCT
BAP1_Ex04_Amp01x2	exon04	CTTGGTGGATGATACGTCCGTG	TACCCACTGGATATCTGAGGACA
BAP1_Ex04_Amp02x2	exon04	CCTGGTGGGCAAAGAACATGTT	CGCAGTGCAAAGGATTAATGGG
BAP1_Ex05_Amp01x1	exon05	GCTGTGAGCCAGGATGAAGG	CTTGCAGTGAGGGGTGCTG
BAP1_Ex05_Amp01x2	exon05	ACAATCATTTGAAAAAGAAACAGCCT	GTGCTCCTGAACTGCAGCAG
BAP1_Ex05_Amp02x2	exon05	GCTGAAACCCTTGGTGAAGTCC	TGCTCTGGATCCCCATTCTTGA
BAP1_Ex06_Amp01x1	exon06	AGGTGGCGTGGCTCGG	CCTCCAGGCTGCTCTCTGAAG
BAP1_Ex06_Amp01x2	exon06	CATGGTCCGCACTGCACTAAG	CAATGCCCCGGAGTTGGC
BAP1_Ex06_Amp02x2	exon06	CCTACTCCCACCCCACATCAG	TGCTCTCTGAAGCTTTGCCTTC
BAP1_Ex07_Amp01x1	exon07	CTAGGAGGTAGGCAGAGACACC	TTGGGCCCTGACTCTGTTTTTC
BAP1_Ex07_Amp01x2	exon07	GTGAAACCCCAGCCAGAGC	CCACTTTGTCAGCTATGTGCCT
BAP1_Ex07_Amp02x2	exon07	TGGTCAATGGGGTAGACCTTCA	TAATAGCCATGCCAGGTGTGTG
BAP1_Ex08_Amp01x1	exon08	CACCTGGATACTCTCTGTCCCT	CTGGCTCAACTGCTCTTCTCTG
BAP1_Ex08_Amp01x2	exon08	TGAGAACCCATGATCTAAGCCTGA	GACGAGGAGTGGACAGACAAGG
BAP1_Ex08_Amp02x2	exon08	CCAGAGAGTAGAACAGGGCAGG	CTGGCTCAACTGCTCTTCTCTG
BAP1_Ex09_Amp01x1	exon09	AATGCAGGGAGGGTTGGGC	CTGCTCTCACGGCTGCG
BAP1_Ex09_Amp01x2	exon09	GTGGTTAGCTGAAGCCCAGATC	GCAGGATCAAGTATGAGGCCAG
BAP1_Ex09_Amp02x2	exon09	CTGCTGCAGAGCCTCTAGTACT	CTGCCAGGATATCTGCCTCAAC
BAP1_Ex10_Amp01x1	exon10	CTCTGAGGTCCACAAGAGGTCC	TTTCTCCTCTGAGCCCTGGATT
BAP1_Ex10_Amp01x2	exon10	CTCTCCCTCTACCTTCTGACGG	TCCAAGTCAGCCAGCAACAAG
BAP1_Ex10_Amp02x2	exon10	CCTGTGTGGTTGCCCTCAGA	TTTAAGGTAGAAGCCCGGGTCT
BAP1_Ex11_Amp01x1	exon11	AGTAGAGAATCCTGCAAGGGTG	TGGAGATCCTACTGGAGGTCCT
BAP1_Ex11_Amp01x2	exon11	ATCAGGCAGAGGAACCTAGCAA	GCTAGTGGTGAAGCCTCCAGG
BAP1_Ex11_Amp02x2	exon11	GATTGTCTAGAAAGGCCGGCAG	CTTCTCTGGGAAGTGCTGGTTC
BAP1_Ex12_Amp01x1	exon12	ATGGGATCCGAAGCACCTAGAA	GCAGCACTTGTTTGTAACTGCC
BAP1_Ex12_Amp01x2	exon12	CAGCCCCATCATGCCAGC	ACCCCAGCAGTACTCAGATGAT
BAP1_Ex12_Amp02x2	exon12	CCTAAGGGCAGAGTTGGTGTTC	TCTCTGGCTGTGAGTGTCTAGG
BAP1_Ex13_Amp01x1	exon13	CAAGTGGCCAGTGAGCCAG	CAATGAGAGTACAGACACGGCC
BAP1_Ex13_Amp01x2	exon13	GGAACCCTTCCCACCCTCTG	CCCACATCTCCAAGGTGCTTTT

BAP1_Ex13_Amp02x1	exon13	GAAAGCACTGCCGATCTCAGAG	CCGTTCCCTTGCTTCACATCTT
BAP1_Ex13_Amp02x2	exon13	TGTAGCGTATGCAGTCAACACG	GCTGAGAAGCTCAAAGAGTCCC
BAP1_Ex13_Amp03x2	exon13	GGACAGAGGAATTGAGAGGTCCT	AGCATGGCTAGTTCAAGTTGCT
BAP1_Ex14_Amp01x1	exon14	GGCCTTACCCTCTGCCAGG	GTGTCCTGCACTCTGATGATTT
BAP1_Ex14_Amp01x2	exon14	CCACCCAACCCAGAAAGTCTTC	CAGCCCAGTGGAGAAGGAGG
BAP1_Ex14_Amp02x2	exon14	GTGAGTATTTCTCCCCACTCAAGG	CACTGGCCACTTGGTGCAC
BAP1_Ex15_Amp01x1	exon15	CGGGAGAGGCCAGATCAGG	TTTGTTGCTGGCCCGCC
BAP1_Ex15_Amp01x2	exon15	ACTCATCGTAGTTGTGGGTCCT	GTGGAGGCTGAGATTGCAAACT
BAP1_Ex15_Amp02x2	exon15	GCAACTGGAGAAATCACCCACC	CTGCTCATCCTTGCCTCTAGCT
BAP1_Ex16_Amp01x1	exon16	CAAGGTCTGCTCAAGCCTCAG	TTTCTCCAGTTGCCTGATCTGG
BAP1_Ex16_Amp01x2	exon16	CCTCAGGCCAGGAGCTGAG	AGGACCCACAACTACGATGAGT
BAP1_Ex16_Amp02x2	exon16	GGGCATTCCAGTTAAGACAGCA	TTTCTCCAGTTGCCTGATCTGG
BAP1_Ex17_Amp01x1	exon17	AGCCTCCATATCTCAGGCCAG	GACCTGATGCCAAGGGGTAATG
BAP1_Ex17_Amp01x2	exon17	ATTCCAAGGATCTCTGGCCCAA	AGCAGGCCCCTTTATCTGTACA
BAP1_Ex17_Amp02x1	exon17	CCATGGGGCTGTGTTGGG	CTCTGTCCAGCCCCTCAGTATT
BAP1_Ex17_Amp02x2	exon17	CTAGAGGGGTCAAGCTGTGCTC	AGCTCTTCCTGCTACTCCAGTT
BAP1_Ex17_Amp03x1	exon17	CCTATAAGCAACCCTGTCTCTGC	CCAGTGGGGTTGGGGTGA
BAP1_Ex17_Amp03x2	exon17	GCTCCAACGGGTTGTCTTGATC	GCTGTTCAAGACTGCTCTCCAT
BAP1_Ex17_Amp04x1	exon17	ACAAGGAAGACAGTTCACAGGC	GTCCTTGTATCATGCCACGGTC
BAP1_Ex17_Amp04x2	exon17	GGACAGAGACCTAGAGCCCAAT	TGTAAGGAAGCCAGGTCTTCTCT
BAP1_Ex17_Amp05x1	exon17	CTGGTGGCTGGGGCTAGAG	GAGGTACTGCAGCTTCCTCCA
BAP1_Ex17_Amp05x2	exon17	CTACTCCCAACCCAGCCCAG	TCTATTTTTCTGGGCTCCAACC
BAP1_Ex17_Amp06x1	exon17	AATATATATTCTGCTGCCCAGGC	CTGCAGCCCACTCTTGCC
BAP1_Ex17_Amp06x2	exon17	CAGGACAGCTCCTAGGAGAGAA	CTGGGAATGGGAGGAACCAGG
BAP1_Ex17_Amp07x1	exon17	GGACTCTCCAGCTGGGACTATT	TCAAGGGTCTCTACCTCTTCGC
BAP1_Ex17_Amp07x2	exon17	GATGCTGCCTCCTGAGCACTAT	GTCAGCATCGGCCGGC
BAP1_Ex17_Amp08x2	exon17	TCAGGGCCAGCAGTCCTC	CTGAGGGCCGTGTCCTTCAG

Adapter sequences were as follow:
Forward: CTACACGACGCTCTTCCGATCT
Reverse: CTCCACCGATAAATATTAGCCCGT