Hepatocellular carcinoma in Mongolia delineates unique molecular traits and a mutational signature associated with environmental agents

Laura Torrens, Marc Puigvehí, Miguel Torres-Martín, Huan Wang, Miho Maeda, Philipp K Haber, Thais Leonel, Mireia García-López, Roger Esteban-Fabró, Wei Qiang Leow, et al.

Title: Hepatocellular carcinoma in Mongolia delineates unique molecular traits and a mutational signature associated with environmental agents

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

Affiliations:

[1]Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

[2]Translational research in Hepatic Oncology, Liver Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, University of Barcelona, Barcelona, Spain

[3]Hepatology Section, Gastroenterology Department, Parc de Salut Mar, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Catalonia, Spain

[4]Sema4, Stamford, Connecticut, USA

[5]Liver Unit, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), CIBEREHD, University of Barcelona, Barcelona, Spain

[6]Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore

[7]Hepato-Pancreatico-Biliary Surgery Department, National Cancer Center, Ulaanbaatar, Mongolia

[8]National Cancer Center, Ulaanbaatar, Mongolia

[9]Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, Université Paris 13, Functional Genomics of Solid Tumors laboratory, F-75006, Paris, France

[10]Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA

[11] Bristol Myers Squibb, Princeton, New Jersey, USA

[12]Department of Surgery, White Plains Hospital, White Plains, New York, USA

[13]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

[#] These authors contributed equally

Corresponding author:

* Josep M. Llovet, MD, PhD

Mount Sinai Liver Cancer Program, Division of Liver Diseases, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, Box 11-23, New York, NY 10029, USA

Tel.: +1 2126599503; fax: +1 212 849 2574

Email address: josep.llovet@mssm.edu

Running title: Molecular characterization of Mongolian HCC

Keywords: Liver Cancer; Mongolia; Molecular characterization; Mutational Signatures, Environmental Carcinogens

AUTHOR CONTRIBUTIONS

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

CONFLICT OF INTEREST

HW, ARV and PT are employed by Sema4. AV has received consulting fees from Guidepoint, Fujifilm, Boehringer Ingelheim, FirstWord, and MHLife Sciences; advisory board fees from Exact Sciences, Nucleix, Natera, Gilead and NGM Pharmaceuticals; and research support from Eisai. AU reports employment and stock ownership from Sema4. JN is employed by Bristol-Myers Squibb. JML is receiving research support from Bayer HealthCare Pharmaceuticals, Eisai Inc, Bristol-Myers Squibb, Boehringer-Ingelheim and Ipsen, and consulting fees from Eli Lilly, Bayer HealthCare Pharmaceuticals, Bristol-Myers Squibb, Eisai Inc, Celsion Corporation, Exelixis, Merck, Ipsen, Genentech, Roche, Glycotest, Nucleix, Sirtex, Mina Alpha Ltd and AstraZeneca. The remaining authors declare no competing interests.

FINANCIAL SUPPORT

This study was partially supported by Bristol-Myers Squibb. MP received a scholarship grant from Asociación Española para el Estudio del Hígado (AEEH). PKH is the recipient of a grant from the German Research Foundation (DFG, HA 8754/1-1). MGL is supported by the i-PFIS program (fellowship IFI18/00006) of the Instituto de Salud Carlos III (ISCIII), co-funded by the European Social Fund (ESF). AV is supported by the US Department of Defense grant (CA150272P3) and the Tisch Cancer Institute (Cancer Center grant P30 CA196521). SPP is supported by the ISCIII through the Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016 and 2017-2020 co-funded by the European Regional Development Fund (ERDF) (PI16/00111 and PI19/00036). JZR's team is supported by Inserm, Labex OncoImmunology Investissement d'Avenir and is "Equipe labellisée par la Ligue Nationale Contre le Cancer». XF is supported by the ISCIII through the Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016 and 2017-2020 co-funded by the ERDF (PI18/00079), by the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (grant 2017_SGR_1753) and by CERCA Programme/Generalitat de Catalunya. DS is supported by the Gilead Sciences Research Scholar Program in Liver Disease. JML is supported by grants from the Samuel Waxman Cancer Research Foundation, the Spanish National Health Institute (MICINN, PID2019-105378RB-I00), NIH (R01 DK128289-01), HUNTER (Ref. C9380/A26813) through a partnership between Cancer Research UK, Fondazione AIRC and Fundación Científica de la Asociacion Española Contra el Cáncer and by the Generalitat de Catalunya (AGAUR, SGR-1358).

ACKNOWLEDGMENENTS

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

STATEMENT OF TRANSLATIONAL RELEVANCE

Mongolia has the world's highest incidence of hepatocellular carcinoma (HCC), with ~100 cases per 100,000 inhabitants, far exceeding that of the rest of the world and surrounding countries. However, it is unclear whether this HCC burden is completely explained by the combination of risk factors (e.g. HBV, HCV, and HDV infections), or eventually other unknown agents might play a role. By performing a thorough molecular characterization of Mongolian HCC, we identified distinct genomic and transcriptomic footprints in Mongolian HCC, including higher number of mutations per tumor, a novel mutational signature, and two molecular clusters not observed in Western HCC. The presence of mutational fingerprints associated with exposure to the carcinogenic agent dimethyl sulfate – a derivate from coal combustion – suggests a role of environmental factors in this country. Thus, our findings could pave the way for the identification of environmental or genetic factors partially responsible for the increased HCC incidence in Mongolia.

ABSTRACT

Purpose: Mongolia has the world's highest incidence of hepatocellular carcinoma (HCC), with ~100 cases/10^5 inhabitants, although the reasons for this feature have not been thoroughly delineated.

Experimental Design: We performed a molecular characterization of Mongolian (n=192) compared to Western HCCs (n=187) by RNA-seq and WES to unveil distinct genomic and transcriptomic features associated with environmental factors in this population.

Results: Mongolian patients were younger, with higher female prevalence, and with predominantly HBV-HDV co-infection etiology. Mongolian HCCs presented significantly higher rates of protein-coding mutations (121 vs 70 mutations per tumor in Western), and in specific driver HCC genes (i.e. *APOB, TSC2*). Four mutational signatures characterized Mongolian samples, one of which was novel (SBS Mongolia) and present in 25% of Mongolian HCC cases. This signature showed a distinct substitution profile with a high proportion of T>G substitutions and was significantly associated with exposure to the environmental agent dimethyl sulfate (DMS, 71%), a 2A carcinogenic associated with coal combustion. Transcriptomic-based analysis delineated two molecular clusters, one with a highly inflamed profile, not present in Western HCC, and that were significantly associated with HBV-HDV etiology and female gender.

Conclusions: Mongolian HCC has unique molecular traits with a high mutational burden and a novel mutational signature associated with genotoxic environmental factors present in this country.

INTRODUCTION

Liver cancer is the second leading cause of cancer-related mortality, and its global burden has increased in recent years[1,2]. Hepatocellular carcinoma (HCC) accounts for 90% of liver cancer cases and arises almost unfailingly in the setting of chronic liver diseases. The worldwide incidence of HCC cases is heterogeneous, reflecting the distribution of known liver disease risk factors such as hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol consumption, and non-alcoholic steatohepatitis[2]. Mongolia, a landlocked East Asian country between Russia and China, shows the world's highest incidence of HCC, with a burden of 86 cases per 100,000 inhabitants[1]. This incidence far exceeds that of the surrounding countries (4x and >20x compared to China and Russia, respectively) or any other country worldwide[1], and has been attributed to a historical high prevalence of HBV (10.6%) and HCV (6.4%) viruses, as well as alcohol consumption[3–5]. Indeed, 90% of Mongolian HCC cases are positive for HBV, HCV, or both,[6] and co-infection with hepatitis delta virus (HDV), a defective virus that needs HBV for its replication cycle and has been associated with liver fibrosis and HCC development, occurs in 50-80% of HBV-infected individuals[7]. Despite the implementation of universal infant HBV vaccination in 1991[3] and the improved control of HCV blood-based transmission, the burden of HCC in Mongolia is increasing year after year, and its incidence is now ~10x compared to the 1960s[1]. Overall, the reasons for this extreme incidence have never been thoroughly understood.

During the last 10 years, large-scale next-generation sequencing studies have been key in deciphering the transcriptomic-based HCC subtypes and the molecular alterations occurring in HCC[8,9]. The presence of mutational signatures consisting of unique nucleotide substitution patterns has allowed to track the exposure of endogenous and exogenous factors in cancer[10]. For instance, HCC risk factors such as aristolochic acid, aflatoxin, alcohol, or tobacco smoking can be

related to specific signatures[8,11]. In this context, assessing the genetic alterations, mutational frequencies and transcriptomic profile of Mongolian HCC could provide relevant information that may help unveil the genetic factors and environmental exposures underlying the high incidence in this population. Previous studies have provided valuable data about the molecular landscape in Mongolian HCC[12]. However, further analyses comparing Mongolian tumors with a Western cohort are needed to understand the relevance of the molecular traits in Mongolian HCC. Additionally, viral characteristics in Mongolian HCC remain underexplored.

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

MATERIALS AND METHODS

Study design

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

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

This study complies with the Declaration of Helsinki and was performed according to ethics committee approval.

Whole exome sequencing

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

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

Mutational signature analysis

Somatic SNVs in exome region (defined by coding exons in Ensembl GRCh37 built) at allelic frequency cutoff of 0.05 and with gnomAD population frequency or ethnic-specific frequency ≤

0.5% were selected. Tumor samples with total SNV count ≥ 50 (after the aforementioned filtering) were used in downstream mutational signature analyses, resulting in a total of 254 samples (148 from the Mongolian cohort plus 106 from the Western cohort). The remaining samples (n=9) were considered negative for the signatures. All signature fitting and *de novo* signature extraction analyses were performed using exome region SNVs and trinucleotide frequencies normalized via `exome2genome` approach in the deconstructSigs R package[15].

RNA sequencing

RNA-seq data was processed by the RAPiD pipeline at the Mount Sinai Genomics Core Facility. Briefly, Fastq files were aligned using STAR (v 2.7.0f)[16] to hg19 with gencode annotation v19. QoRTs (v1.3.6)[17] was used for QC and obtaining raw counts. Batch correction was performed using RUVSeq[18]. Empirical method (RUVg) with 10.000 low expressed genes, and normalization for subsequent analysis was performed using VST method from DESeq2[19].

Unsupervised clustering analysis of the whole cohort (n=224) was performed using the Non-negative matrix factorization (NMFc) module from GenePattern and Euclidean distance and Ward's agglomerative procedure. Clustering of the Mongolian (n=118) and NCI Mongolian cohort[12] (n=70) was also performed using NMFc. Gene expression characterization was performed using Nearest Template Prediction (NTP), Gene Set Enrichment Analysis (GSEA), and single sample GSEA (ssGSEA) modules from GenePattern. To this end, Molecular Signature Database (MSigDB, www.broadinstitute.org/msigdb) and previously reported gene sets were used (**Supplementary Table 1**). Class comparison between molecular clusters was performed using subclass mapping analysis and NTP in the in-house and external cohorts. Finally, the stromal infiltration and relative tumor purity were assessed using the ESTIMATE R package[20].

Statistical analysis

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

Data Sharing

Total RNA-seq data and the original WES data from the Mongolia and Western cohorts have been deposited at the European Genome-Phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under the study accession code EGAS00001005364. HBV and HDV genotypes are deposited in Genebank (accession IDs MZ984226 - MZ984323 and OM981174 - OM981226, respectively). The remaining data are available in the Article, Supplementary Information, or available from the authors upon reasonable request".

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

RESULTS

Clinico-pathological characteristics of the cohorts

HCC patients from the Mongolian cohort were younger (61 vs 66 years old, p<0.001), with a higher rate of HBV/HDV co-infection (84% of HBV infected vs 7% in the Western cohort, p<0.001), and lower rate of non-infected cases (15% vs 40%, p<0.001) (**Table 1**, **Supplementary Table 2**). In line with previous data[21]. 54% of the HCC patients in Mongolia were male compared to 80% in the Western cohort (p<0.001). The rate of advanced hepatic fibrosis (F3-4) and cirrhosis (F4) was significantly lower in Mongolian compared to Western cases (38% vs 79% and 16% vs 60%, respectively), independently of etiology (**Supplementary Fig. 1a**). Tumor characteristics were similar in both cohorts, with most tumors within Barcelona Clinic Liver Cancer (BCLC) stages 0-A, and with alpha fetoprotein (AFP) <400 (IU/mL). However, tumors in the Mongolian cohort showed a lower differentiation grade (**Table 1**). No significant survival differences were observed between cohorts.

Viral characterization of Mongolian and Western individuals

To understand whether the particular clinicopathological features of the Mongolian HCCs were due to unique viral characteristics, we analyzed the phenotypes of HBV and HDV in infected patients. HBV genomes can be classified into 9 genotypes (A to I), according to differences in nucleotide sequences. HBV genotypes have a characteristic geographical and ethnic distribution, and HBV genotype D is known to be almost universal in Mongolia[22]. In the Western cohort, patients were infected mostly by genotypes C and D (27% and 43%, respectively, p<0.001) whereas in the Mongolian cohort, all HBV-infected individuals were genotype D, including 2 patients with recombinant forms of genotype C and D (**Supplementary Table 3)**. Interestingly,

genotype D has been previously associated with reduced HCC development as compared to genotype C[23]. The most frequent HBV sub-genotypes in Mongolian and Western cases were D1 (77%) and D3 (39%), respectively (**Supplementary Fig. 1b**). Among Western cohort samples, the most prevalent genotypes in European patients was genotype D3 (71%), whereas in patients from USA it was genotype C (60%) (**Supplementary Fig. 1c**). Regarding HDV infection, all Mongolian patients were genotype 1 (**Supplementary Table 3**). We then evaluated the presence of basal core promoter (BCP) and precore HBV mutations, which have been associated with liver disease progression and HCC development[24]. It has been previously suggested that these mutations are more frequent in HBV-infected patients with genotype D, with only 11% of patients with such genotype being wild-type[25]. We detected BCP A1762T, BCP G1764A and precore G1896A mutations in 14%, 21% and 29% of HBV-infected Mongolian patients, respectively (**Supplementary Fig. 1d, Supplementary Table 3**). Interestingly, the prevalence of these mutations was significantly higher in the Western cohort (60%, 66%, and 46%, respectively), ($p<0.001$, $p<0.001$, and $p=0.089$, respectively, **Supplementary Fig. 1d**). The percentages observed in our Western patients were similar to those previously reported[26]. In line with previously reported data[27], precore G1896A mutations were less frequent in HBV/HDV co-infected versus HBV mono-infected in Mongolia (20% vs 64%, $p=0.002$); however, we found no differences for both BCP A1762T and G1764A mutations (14% vs 14%, and 19% vs 29%, respectively, p=ns). Overall, these data suggest that the rate of BCP and precore HBV pro-oncogenic mutations in Mongolia is particularly low, despite the predominance of genotype D in this population.

The interaction between HBV and HDV viruses is complex and not fully understood. It has been proposed that HDV can suppress HBV[28,29], but HBV and HDV levels can fluctuate over time[30].

To further investigate this, we determined HBV-DNA and HDV-RNA levels in tumor-adjacent liver tissue in Mongolian and Western samples. Of note, high HBV-DNA levels in blood have been associated with more aggressive liver disease, including HCC development[31]. First, we analyzed the differences between HDV positive and negative samples. Intrahepatic HBV-DNA load was significantly higher in samples with HBV/HDV co-infection than those with HBV mono-infection (5.0 vs 3.8 log copies/µg total DNA, p=0.001, **Supplementary Fig. 2a**). Furthermore, high HBV-DNA levels were associated with advanced liver fibrosis, advanced tumor stage, and worse survival in Mongolian individuals (**Supplementary Table 4, Supplementary Fig. 2b**). Conversely, patients with high HDV-RNA levels showed significantly higher alanine aminotransferase levels, suggesting greater inflammation (**Supplementary Table 5)**. Finally, no significant differences in HBV-DNA load were found between Mongolian and Western samples (4.85 vs. 4.97 log copies/µg total DNA, respectively p=0.23, **Supplementary Fig. 2c, Supplementary Table 6**). Taken together, our analysis suggests that HBV and HDV viral characteristics in Mongolia are highly homogeneous and with low oncogenic potential.

Analysis of the genomic landscape in Mongolian HCC

To gain further insights into the molecular landscape of Mongolian HCC, we performed mutation and copy number variation (CNV) analysis. The pattern of broad gains and losses in our Western cohort was consistent with previous reports in HCC[8], with 1q and 8q gains and 8p losses being the most common alterations (**Supplementary Fig. 3a-b**, **Supplementary Table 7).** When we compared the broad chromosomal variation profiles between Mongolian and Western HCC patients, no difference was observed in terms of overall CNV burden (**Supplementary Fig. 3c-e**). Nonetheless, Mongolian HCCs showed a significantly higher occurrence of 1p gains, 9q gains, as well as fewer 9q losses, 1q losses, and 8p losses (**Supplementary Table 7**).

The average mutations per tumor was significantly higher in Mongolian patients compared to Western, with a median of 121 and 70 mutations/tumor, respectively (p<0.001) (**Fig. 1b**). Accordingly, the median tumor mutational burden (TMB) in the Mongolian and Western cohorts was 4.0 and 2.3 mutations/Mb, respectively (p<0.001). No significant differences were observed depending on the origin of Western samples (0.645, **Fig. 1c**). To rule out the possibility that the observed difference was due to random selection bias, we then compared the median protein-coding mutations with previously published cohorts in Western and Asian countries, applying the same filtering criteria[8,12–14]. The median number of mutations per tumor was 111 in the Mongolian NCI cohort[12], 76 in TCGA[14], 61 in the European[8] cohort, and 63 in the Korean cohort[32], corresponding to a TMB of 3.7, 2.5, 2.0, and 2.1, respectively (**Fig. 1b, Supplementary Table 8**) confirming that the number of mutations in both Mongolian HCC cohorts was significantly higher than in other countries (all <0.001). There was a positive association between the number of mutations and tumor grade in Mongolian cases, with a median of 100, 120, and 171 mutations/tumor in samples with good, moderate, or poor differentiation grades (p=0.004). We did not observe any significant difference in the median of mutations according to etiology in either cohort (**Supplementary Fig. 4a-c**).

Previous studies in HCC have suggested that highly mutated tumors (TMB ≥4 mutations/Mb) are enriched with mutations in DNA damage response (DDR) genes[33]. In our Mongolian cohort, 84 (56%) of samples showed ≥4 mutations/Mb, compared to 7 (6.3%) in the Western cohort (p<0.001). However, no association was observed with the presence of mutations in DDR genes (**Supplementary Table 9)**.

Mutational profile of Mongolian HCC

We then explored whether the higher frequency of mutations in the Mongolian cohort was due to enrichment in specific genes. Among the 100 genes with significant differences between cohorts, 89 were more mutated in Mongolian vs Western HCCs, suggesting that the higher number of mutations was broadly occurring across the whole genome and not concentrated in specific loci (**Supplementary Table 10**). Overall, we detected a significant mutational increase in multiple known HCC driver genes in Mongolia compared to the Western cohort, including *TP53* (46% vs 32%), *APOB* (15% vs 5%), *TSC2* (9% vs 1%), and *NFE2L2* (6% vs 1%), ($p<0.05$, **Fig. 2, Supplementary Fig. 5, Supplementary Tables 11-12**). Furthermore, genes belonging to the *KMT2* histone lysine methyltransferase family were significantly more mutated in Mongolian HCC (34% vs 18%, $p=0.005$, **Fig. 3a**). Mongolian HCC with mutations in the *KMT2* gene family displayed higher TMB than patients without these mutations in both Mongolian cohorts (median TMB 4.6 and 4.7 vs 3.9, $p=0.005$ and $p<0.0001$, **Fig. 3b**). The mutation rate of HCC drivers was also assessed in external HCC cohorts for comparison, confirming a similar frequency of mutations in the Mongolian NCI cohort for most HCC drivers including *APOB*, *TSC2*, *NFE2L2*, and *KMT2* family (all n.s. vs in-house Mongolian cohort) (**Fig. 2, Supplementary Table 12**).

Next, potential drivers in Mongolian HCC were further assessed by OncodriveCLUSTL and dN/dScv algorithms[34,35]. Among the genes significantly more mutated in the Mongolian cohort, 6 were enriched for damaging alterations, suggesting that they could exert a driver role in Mongolian HCC ($q<0.05$; **Fig. 3c-d**), including *TSC2* (9%). Similar *TSC2* mutation rates were confirmed in the Mongolian NCI cohort (7%) (**Fig. 3c**).

Finally, differences in the mutation profile depending on etiology were detected in the Mongolian cohort (**Supplementary Fig. 4d-e**). *CDKN2A* mutations were enriched in patients with HBV (7.4% vs 0%) and HDV (8.0% vs 0%) infection, while *ARID2* mutations were more common in HCV-infected patients (11.9% vs 1.2%) ($p<0.05$, **Supplementary Fig. 4d**). No significant differences between Western samples from Europe and USA were observed (**Supplementary Fig. 6, Supplementary Table 11)**.

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

Mutational signature analysis

We then analyzed the pattern of single base substitutions in the Mongolian and Western cohorts. Notably, Mongolian HCC was characterized by a higher proportion of T>G substitutions compared to Western tumors (**Fig. 4a, Supplementary Fig. 7a-b, Supplementary Table 13**). This was also observed in the Mongolian NCI cohort (**Supplementary Fig. 7c**).

De novo mutational signature extraction from the Mongolian and Western cohorts revealed four signatures (**Fig. 4b**), three of which were mapped to COSMICv3 signatures previously found in liver cancer: SBS22, a combination of SBS6-SBS40 and of SBS16-SBS26 (cosine similarity >0.90 in all cases)[36]. The fourth signature did not present strong similarities with any of the COSMIC signatures and was therefore considered novel (**Supplementary Tables 14-15**). We then performed signature fitting using the *de novo* signature 4 and HCC-specific COSMICv3 signatures by applying a bootstrap approach. Interestingly, the Mongolian cohort was enriched in the *de novo*

signature 4, henceforth renamed SBS Mongolia (SBSM) (25.2% [38/151 vs 4.5% [5/112] in Mongolian vs Western cohorts; p<0.0001) (**Fig. 5a-b**). Notably, Mongolian HCC presenting the SBSM signature showed a distinct substitution profile consisting of a high proportion of T>G substitutions (14% vs 8% in SBSM positive and negative samples, respectively [p<0.001], and 6% in Western HCC, **Fig. 5c**). Other COSMICv3 signatures previously reported in HCC[36] presented similar prevalence between cohorts (**Fig. 5a-b, Supplementary Fig. 7d-e, Supplementary Table 16**). High presence of the SBSM signature in Mongolian HCC was confirmed in the external NCI Mongolian cohort (31% of positive cases, 22/71), but not in another Asian (Korean) cohort (4.3% of positive cases, 10/231, **Supplementary Table 17**). No differences in the frequency of SBSM were detected in Western HCC depending on the sample origin (**Supplementary Table 16**). However, SBS22, associated with exposure to aristolochic acid, was more frequent in samples from the USA (6/43 vs 0/69 in Europe, p=0.02), likely due to the presence of Asian-American patients, which reportedly present higher exposure to this compound[37].

To further characterize SBSM-positive samples, we assessed the presence of mutational signatures linked to the effects of known or suspected environmental mutagens from the Compendium of Mutational Signatures of Environmental Agents[38]. Samples presenting SBSM were significantly enriched for the mutational signature associated with exposure to dimethyl sulfate (DMS) (71.1% [27/38] vs 26.5% [30/113], p<0.0001, **Fig. 5c**), as further confirmed in the Mongolian NCI cohort (**Fig. 5d**). In line with this, DMS was the only environmental-related signature significantly enriched in Mongolian HCC compared to Western (37.7% [57/151] vs 18.8% [21/112], **Supplementary Table 18**), a feature confirmed in external HCC cohorts (**Supplementary Table 19**), with no significant differences depending on the origin of Western samples (**Supplementary Table 18**). Patients presenting the DMS signature were older (64.3 vs 59.5 years) and

predominantly HCV-positive; **Supplementary Fig. 8**). No association between SBSM and TMB, etiology, fibrosis or other clinical and molecular variables were found (**Fig. 5c**).

Finally, we investigated the mutational profile in adjacent matched liver tissue of Mongolia and Western HCCs. Due to the small number of SNVs observed in the adjacent tissues, mutational signature fitting was performed on pooled variants from the Mongolian and Western cohorts using HCC-specific COSMICv3 signatures plus SBSM (**Supplementary Fig. 9**). SBSM was the only dominant signature in adjacent tissue from the Mongolian cohort (**Supplementary Fig. 9c-d**), suggesting that non-tumoral liver tissue in Mongolia presents the signature before HCC arises. SBS5, associated with age-related clock-like mutations[36], was the main signature in Western non-tumoral tissue. The DMS signature was the most dominant environmental signature in Mongolian non-tumoral tissues, while it was not detected in Western non-tumoral samples (**Supplementary Figure 9e-f**).

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

Identification of unique gene expression patterns in Mongolian HCC

We then investigated the transcriptome profiling of HCC samples using RNA-seq data to define the molecular patterns in Mongolian HCC tumors. Unsupervised clustering analysis of Mongolian and Western samples using non-negative matrix factorization (NMFc) identified two robust clusters (**Supplementary Fig. 10a-c**). Notably, 80% of Western tumors were included in one cluster, while the second cluster showed a strong enrichment of Mongolian HCC samples, suggesting that Mongolian HCC may present a distinct molecular profile. No clustering differences

were observed between Western samples from Europe and USA (**Supplementary Fig, 9c**). Furthermore, NMFc analysis of the Mongolian HCC samples alone revealed three main clusters - MGL1, MGL2 and MGL3- (**Supplementary Fig. 10b-c**), which overlapped with the classification of the whole cohort. Unsupervised clustering showed similar results, thus confirming robustness of our findings (**Supplementary Fig. 10d**).

To elucidate the transcriptomic differences between Mongolian and Western HCC, we performed ssGSEA and NTP analyses (**Supplementary Fig. 11a**). The Mongolian cohort presented an enrichment in Hoshida S1 and Proliferation classes (39% vs 20% and 36% vs 14%, $p<0.01$). In addition, Mongolian tumors showed enhanced inflammatory signaling (i.e., IFN and HCC Immune class), response to viral infection, and growth factor-related pathways (all $p<0.05$). Comparatively, the Western cohort was enriched in the Hoshida S3 class (26% vs 47%, $p<0.01$) and liver-related metabolic activation. Proliferative and inflammatory profiles similar to the in-house cohorts were confirmed in the Mongolian NCI cohort compared to non-Mongolian external HCC cohorts (**Supplementary Figure 12a**). Overall, Western HCC samples from Europe and USA showed a similar molecular profile (**Supplementary Fig. 11a**)

We then characterized each one of the MGL clusters. Patients belonging to MGL1 class (44% of the cohort) were more frequently HCV-infected, older and mostly males (female:male ratio of 1:2), (**Supplementary Table 20**) and with a molecular profile closer to Western HCCs than the rest of Mongolian HCC (**Fig. 6**). On the other hand, patients of the MGL2 (26%) and MGL3 clusters (30%) were significantly younger than MGL1 and enriched in HBV/HDV infection and triple infections (HBV/HDV/HCV). Interestingly, while patients of the MGL3 class showed a female:male ratio similar to MGL1 (1:2), the MGL2 class showed a female:male ratio of 2:1 and

higher AFP levels. None of the MGL clusters was associated with specific outcomes (**Fig. 6, Supplementary Fig. 13, Supplementary Table 20**).

In terms of molecular features (**Supplementary Tables 21-23**), MGL1 cluster was characterized by enrichment in *CTNNB1* class and activation of metabolic and liver-specific pathways (all $p<0.05$). MGL2 HCC cases displayed a proliferative HCC phenotype (i.e enrichment of Proliferation, G3, and Cluster A gene signatures), higher rates of RB1 mutations, and lower rates of *CTNNB1* mutations (**Fig. 6**). Finally, MGL3 tumors were characterized by lower rates of *TP53* mutations, fewer broad chromosomal alterations, and were particularly enriched in immune-related features (i.e immune class, interferon and inflammatory pathways and PD1 signaling).

To validate these results, we performed NMFc analysis using RNA-seq data from the Mongolian NCI cohort[12]. The clinico-pathological and molecular features of each MGL cluster aligned with what was described in the in-house Mongolian cohort (**Supplementary Fig. 12, Supplementary Table 24**). Subclass mapping and NTP analyses in the in-house Mongolian cohort indicated a good overlap of the MGL clusters with the previously published MO classification of Mongolian HCC[12] (**Supplementary Fig. 14**). Specifically, MGL1 aligned with the MO1 class, associated with good prognosis and metabolic processes (FDR < 0.018). MGL2 overlapped with the MO4 class (FDR = 0.012), characterized by poor prognosis and activation of proliferative pathways. Finally, MGL3 encompassed both MO2, a good prognosis class with inflammatory signaling, and MO3, associated with poor prognosis and metabolic pathway enrichment17 (both FDR = 0.012) (**Supplementary Fig. 14**). Overall, our results suggest that three distinct gene expression patterns characterize Mongolian HCC of whom two present unique clinicopathological features not observed in HCC samples from Western countries.

Immune characterization of the molecular classes of Mongolian HCC

We then further explored the inflammatory profile of Mongolian HCC using ESTIMATE[20] and ssGSEA in our in-house Mongolian cohort and the NCI cohort[12]. In both Mongolian cohorts, MGL3 showed significantly higher immune enrichment score, presence of both an innate and adaptive immune response, and signatures predicting response to immunotherapy (**Fig. 6, Supplementary Fig. 12c**). A certain degree of inflammation was also observed in MGL2, even if significantly lower than MGL3. Our previously reported HCC immune class was detected in most patients belonging to MGL3 (88% in the Mongolian cohort, 100% in the NCI Mongolian cohort[12], **Fig. 6, Supplementary Fig. 12c**). Compared to Western HCC, the immune class in Mongolian tumors was larger (42% versus 29%, p=0.05) with an inverted ratio of Exhausted/Active subtypes (65/35 vs 30/70, **Supplementary Fig. 11b**), indicating a more prominent immunosuppressive phenotype. Overall, we observed a higher presence of immune signaling in the Mongolia clusters MGL2 and MGL3. Considering that most patients belonging to these clusters were HBV/HDV infected and younger, this suggests that HDV infection could accelerate disease progression through inflammatory mechanisms.

DISCUSSION

Our study entails a comprehensive characterization of the molecular profile of HCC in Mongolia, the country with the highest global incidence. Mongolia has many particularities that might play a role in HCC burden, including specific risk factors, socioeconomic particularities, and genetic profile[12,39]. Despite a strikingly high prevalence of HBV (10.6%), HCV (6.4%), and HDV (70% of HBV-positive individuals) infections and alcohol consumption[3–5], it is unclear whether HCC

incidence is completely explained by the unique combination of risk factors, or eventually other unknown factors might be responsible. In addition, a direct comparison with a comprehensive Western cohort is necessary to understand the relevance of newly identified molecular traits. By analyzing WES and RNA-seq data from 379 new Mongolian and Western HCC samples, we identified unique genomic and transcriptomic footprints in Mongolian tumors that suggest a role of specific genetic and environmental factors in the country. The study provides novel information in three major areas: a) Clinical characteristics of Mongolian cases, b) Unique tumor mutational burden and mutational profile associated with environmental agents, and c) Unique transcriptomic-based molecular classes.

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

In our study, HBV characteristics such as genotype D1 and precore mutations in <30% of cases revealed traits associated with low oncogenic potential of the virus in Mongolia[23,25], whereas HBV load was similar to Western samples. Thus, other factors not associated with HBV infection might be responsible for the high HCC incidence rates in Mongolia[23,25]. In this sense, 84% of the HBV-infected Mongolian patients showed co-infection with HDV, which contrasts with the Western data (less than 7% of coinfection)[40]. This particular co-infection profile was associated with two molecular subclasses only identified in Mongolian patients (MGL2 and MGL3), thus pointing

towards a potential role of HDV in the oncogenic process, despite it is not currently considered a carcinogenic agent[7].

At the genomic standpoint, Mongolian HCCs had a high rate of protein-coding mutations, which almost doubled that in the Western in-house cohort (121 vs 70 mutations per tumor) and publicly available datasets[8,12–14]. This suggests the presence of intrinsic and/or extrinsic factors promoting mutagenesis in Mongolian HCC. Notably, while several HCC driver genes presented similar mutation rates in the Mongolian cohort compared to the Western tumors and previous studies[8,13,14] (e.g. *CTNNB1* and *ARID1A*), others such as *APOB*, and the *KMT2* gene family were significantly more mutated in Mongolian HCC (e.g. *APOB* in 15.2% vs 4.5% in Western). For instance, *KMT2* family mutations (34% vs 17% in Western) were associated with higher TMB in Mongolian HCC, suggesting that they could be partially responsible for the increased mutational burden. In line with this, loss of function in KMT2 methyltransferases has been proposed to induce DNA damage due to aberrant chromatin remodeling[41]. *TSC2* (9%) mutations were identified as potential drivers in Mongolian tumors. *TSC2*, a known cancer-related gene participating in the mTOR oncogenic pathway [8,42], has been proposed as an actionable alteration with level 2B evidence, as it could be a predictor of response to the FDA-approved drug everolimus[42].

To understand whether Mongolian HCCs present differential genomic and genotoxic footprints, we assessed the presence of distinct mutational signatures[10,38]. We identified a new mutational signature (SBS Mongolia) with no similarities with previously reported COSMIC signatures. SBS Mongolia was significantly enriched in Mongolian HCC (25% vs 4.5%), indicating unique substitution patterns characterized by increased T>G substitutions. Overall, this mutational landscape points towards specific exposure to environmental factors in Mongolian patients. In this regard, Mongolian HCC samples presenting the novel SBS Mongolia were significantly enriched

for the mutational signature associated with exposure to DMS (71.1% vs 26.5%). DMS has been classified as a probable carcinogenic hazard to humans by the International Agency for Research on Cancer[43] (category 2A carcinogen) and its production has been associated with coal combustion. Most of Mongolian population is currently exposed to coal combustion, as coal is used to fight against the intense cold weather both in urban and rural areas. Half of the 3-million population of Mongolia lives in Ulaanbaatar, an overpopulated capital with dismal environmental conditions[44], whereas the rest is still predominantly nomad and lives in traditional tents or *gers*, where coal is used both for cooking and heating. This fact has been recognized by international organizations as a major health threat in this country[44]. Considering this, we hypothesize that long-term exposure to DMS from coal combustion could be a risk factor for HCC development in Mongolia. Future functional and epidemiological studies will be required to confirm the association between the DMS-related signature and exposure to coal combustion in Mongolia. Specifically, DMS could act as a cofactor and exacerbate the risk of HCC in patients exposed to viral infection. A similar phenomenon is observed with the fungal compound aflatoxin B1, which acts synergistically with HBV to induce HCC in Southeast Asia and Sub-Saharan Africa[37].

The transcriptomic profile of Mongolian tumors was consistent with known HCC features[2]. Nonetheless, we observed two striking differences a) Mongolian HCC was characterized by an enhanced proliferative and immunological signaling, with a proportion of tumors belonging to proliferative/progenitor HCC classes (39%) doubling the one in Western HCC (20%) and previously published studies[45]; and b) Two out of the three identified molecular classes presented distinct molecular features compared to Western HCC, and thus are deemed unique for Mongolian tumors. Effectively, MGL2 (26%) and MGL3 classes (30%) were specific for Mongolian tumors but not for Western HCC and were enriched in HBV/HDV infection. Interestingly, MGL2 class,

was associated with both clinical and molecular features of aggressiveness and showed a female:male ratio of 2:1. The presence of this molecular class could be due to the increased HCC incidence in females in this population[21]. Finally, MGL3 presented an inflamed profile, potentially linked to the response to HBV/HDV infection[7]. Inflamed HCC tumors from the MGL3 class could be ideal candidates to receive immune checkpoint inhibitors and combination therapy approaches[46].

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

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021;71:209–49.
2. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nature Reviews Disease Primers [Internet]. 2021 [cited 2021 May 21];7:7. Available from: https://pubmed.ncbi.nlm.nih.gov/33479224/
3. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. The Lancet Gastroenterology and Hepatology. 2018;3:383–403.
4. Blach S, Zeuzem S, Manns M, Altraif I, Duberg AS, Muljono DH, et al. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. The Lancet Gastroenterology and Hepatology. 2017;2:161–76.
5. Demaio AR, Dugee O, de Courten M, Bygbjerg IC, Enkhtuya P, Meyrowitsch DW. Exploring knowledge, attitudes, and practices related to alcohol in Mongolia: a national population-based survey. BMC Public Health. 2013;13:178.
6. Dondog B, Lise M, Dondov O, Baldandorj B, Franceschi S. Hepatitis B and C virus infections in hepatocellular carcinoma and cirrhosis in Mongolia. Eur J Cancer Prev. 2011;20:33–9.
7. Puigvehí M, Moctezuma-Velázquez C, Villanueva A, Llovet JMJ, Moctezuma-Velazquez C, Villanueva A, et al. The oncogenic role of hepatitis delta virus in hepatocellular carcinoma. JHEP Rep [Internet]. 2019 [cited 2020 Jul 1];1:120–30. Available from: https://pubmed.ncbi.nlm.nih.gov/32039360/

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

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

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

11. Letouzé E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nature Communications [Internet]. 2017 [cited 2020 Jul 1];8:1315. Available from: https://pubmed.ncbi.nlm.nih.gov/29101368/

12. Candia J, Bayarsaikhan E, Tandon M, Budhu A, Forgues M, Tovuu LO, et al. The genomic landscape of Mongolian hepatocellular carcinoma. Nature Communications [Internet]. 2020 [cited 2020 Oct 29];11:4383. Available from: https://pubmed.ncbi.nlm.nih.gov/32873799/

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

14. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell [Internet]. 2017 [cited 2017 Jul 17];169:1327-1341.e23. Available from: /pmc/articles/PMC5680778/?report=abstract

15. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biology. 2016;17.

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

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

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

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

20. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013;4.

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

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

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

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

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

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

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

28. Giersch K, Homs M, Volz T, Helbig M, Allweiss L, Lohse AW, et al. Both interferon alpha and lambda can reduce all intrahepatic HDV infection markers in HBV/HDV infected humanized mice. Scientific Reports. 2017;7.

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

30. Schaper M, Rodriguez-Frias F, Jardi R, Tabernero D, Homs M, Ruiz G, et al. Quantitative longitudinal evaluations of hepatitis delta virus RNA and hepatitis B virus DNA shows a dynamic, complex replicative profile in chronic hepatitis B and D. Journal of Hepatology. 2010;52:658–64.

31. Chen CJ, Yang HI, Su J, Jen CL, You SL, Lu SN, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA Level. J Am Med Assoc. 2006;295:65–73.

32. Ahn S-M, Jang SJ, Shim JH, Kim D, Hong S-M, Sung CO, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology [Internet]. 2014;60:1972–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24798001

33. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nature Genetics. 2014;46:1267–73.

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

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

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

37. 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. Journal of Hepatology. 2012;56:908–43.

38. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell [Internet]. 2019 [cited 2020 Nov 3];177:821-836.e16. Available from: https://pubmed.ncbi.nlm.nih.gov/30982602/

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

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

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

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

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

44. WHO. Air pollution in Mongolia. Bull World Health Organ [Internet]. 2019;97:79:80. Available from: https://www.who.int/bulletin/volumes/97/2/19-020219/en/

45. Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Research. 2008;68:6779–88.

46. Montironi C, Castet F, Haber PK, Pinyol R, Torres-Martin M, Torrens L, et al. Inflamed and non-inflamed classes of HCC: a revised immunogenomic classification. Gut. 2022;gutjnl-2021-325918.

FIGURE LEGENDS

Figure 1. Flow chart of the study and mutational profile in Mongolian HCC. a A total of 192 HCC samples from Mongolia were used in this study. A Western cohort including 187 HCCs was used as internal control. **b** Mutations per tumor in the Mongolian (n=151) and Western (n=112) cohorts. Mongolian NCI (n=71), European (n=241), Korean (n=231) and TCGA (n=363) external cohorts are shown as reference. **c** Mutations per tumor in Mongolian HCC (n=151) compared to Western HCC from Europe (n=69) and USA (n=43). Y axis was cut at 300 mutations/tumor to facilitate data interpretation. P-value corresponds to Kruskal-Wallis test. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers).

Figure 2. Mutational landscape of Mongolian and Western HCC. Mutations present in HCC samples from the Mongolian (n=151) and Western (n=112) cohorts. The frequency of mutations in the Western and Mongolian NCI cohort are indicated for comparison (left). Genes with significant differences between the Mongolian and Western cohorts are highlighted in green (Fisher $p<0.05$). Top panel shows tumor mutational burden (TMB, mutations/Mb) per sample. Middle panel indicates the presence of mutations per sample (right) and overall percentage (left) in the most frequently mutated genes.

Figure 3. Potential driver and *KMT2* family mutations in Mongolian and Western cohorts. a Mutations in the *KMT2* gene family in the Mongolian (n=151) and Western cohorts (n=112). **b** TMB in samples with *KMT2* family mutations and wild type (wt). Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers). P-value corresponds to Kruskal-Wallis test. **c** Mutations in potential driver genes in the Mongolian and Western cohorts according to enrichment in damaging alterations. Percentage of mutations in external cohorts is indicated in the right panel. Top panel shows tumor mutational burden (TMB,

mutations/Mb) per sample. Middle panel indicates the presence of mutations per sample (left) and overall percentage (right) in the Mongolian and Western cohorts. Significant differences between the Mongolian and Western cohorts are indicated in green (Fisher $p < 0.05$). **d** Mutation distribution in the *TSC2* gene.

Figure 4. Single-base substitution profile in Mongolian HCC. a Differences in trinucleotide substitution frequency between the in-house Mongolian and Western cohorts. Bars indicate the median values for each substitution group. Significant substitutions differences in both in-house and NCI Mongolian cohorts are highlighted in brighter colors. * $p < 0.05$ (Kruskal-Wallis test). **b** *De novo* single-base substitution (SBS) signatures found in Mongolian and Western HCC samples. For each signature, the 96-substitution classification is displayed, including the substitution type and sequence context.

Figure 5. Mutational signatures in Mongolian and Western HCC. a,b Signature fitting results in Mongolian (**a**, n=151) and Western HCC (**b**, n=112) using HCC-specific COSMIC signatures and SBS Mongolia (SBSM). **c,d** Distribution of samples harboring SBSM in the Mongolian cohort (**c**) and in the NCI Mongolian cohort (**d**). Presence of dimethyl sulfate (DMS) mutational signature, main clinical variables and T>G substitution frequency are shown for each sample. FDR-adjusted p-values comparing SBSM positive and negative samples are indicated. Percentages in the lower panel refer to median T>G frequency.

Figure 6. Molecular classification of Mongolian HCC. Consensus-clustered classification and inflammatory profile of Mongolian HCC samples. In the heatmap, clinico-pathological characteristics, nearest template prediction and gene set enrichment in each sample are shown.

TABLES

Table 1. Baseline characteristics of the Mongolian and Western cohorts

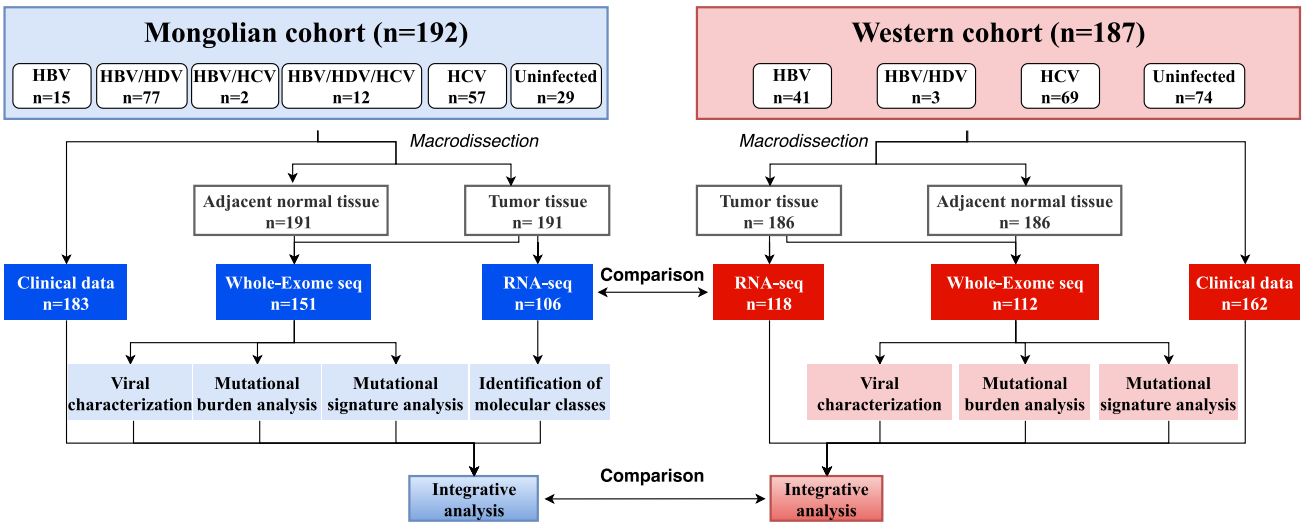
	Mongolian Cohort (n=192)	Western Cohort (n=187)	p value
Age (years)	61 (18-80)	66 (29-91)	<0.001
< 60 years (n, %)	82 (44.6)	32 (18.6)	<0.001
< 50 years (n, %)	20 (10.9)	7 (4.1)	0.017
Gender (male, %)	98 (53.6)	137 (79.7)	<0.001
Etiology			<0.001
HBV+ (n, %)	15 (7.8)	41 (21.9)	
HBV/HDV+ (n, %)	77 (40.1)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

HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis delta virus; AFP, alfa-fetoprotein; BCLC, Barcelona Clinic Liver Cancer The following variables have missing values for the Mongolian and Western cohorts, respectively: Age: 8 and 15 patients. Gender: 9 and 15 patients.

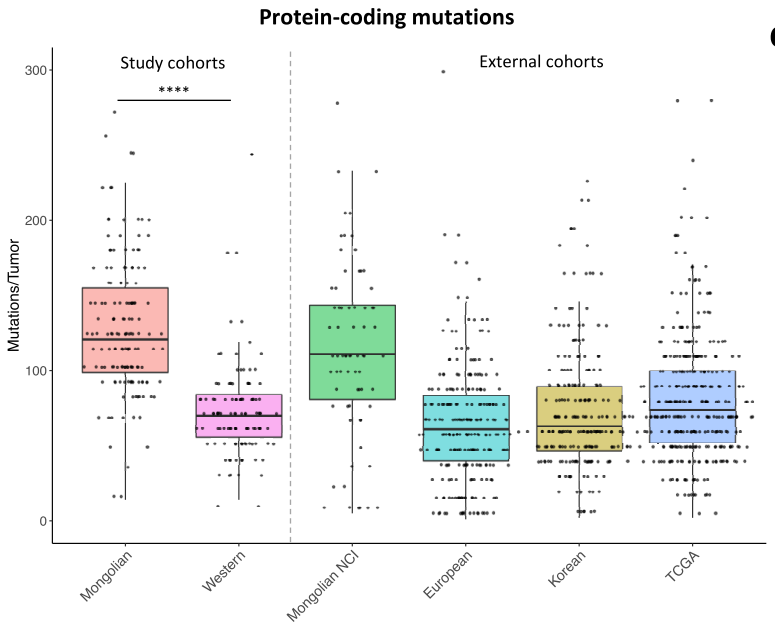
Bilirubin, albumin, and platelets: 19 and 17 patients. AFP: 39 and 28 patients. Tumor size: 16 and 25 patients. BCLC stage: 23 and 25 patients. Tumor number: 23 and 24 patients. Liver fibrosis in 24 and 52 patients. Microvascular invasion in 41 and 15 patients. Tumor grade in 88 and 44 patients.

Figure 1

a


Mongolian cohort (n=192)
HBV n=15
HBV/HDV n=77
HBV/HCV n=2
HBV/HDV/HCV n=12
HCV n=57
Uninfected n=29
Western cohort (n=187)
HBV n=41
HBV/HDV n=3
HCV n=69
Uninfected n=74
Macrodissection
Adjacent normal tissue n=191
Tumor tissue n= 191
Tumor tissue n= 186
Adjacent normal tissue n= 186
Clinical data n=183
Whole-Exome seq n=151
RNA-seq n=106
Comparison
RNA-seq n=118
Whole-Exome seq n=112
Clinical data n=162
Viral characterization
Mutational burden analysis
Mutational signature analysis
Identification of molecular classes
Viral characterization
Mutational burden analysis
Mutational signature analysis
Integrative analysis
Comparison
Integrative analysis

b


Protein-coding mutations
Study cohorts
External cohorts
****
Mutations/Tumor
300
200
100
0
Mongolian
Western
Mongolian NCI
European
Korean
TCGA

c

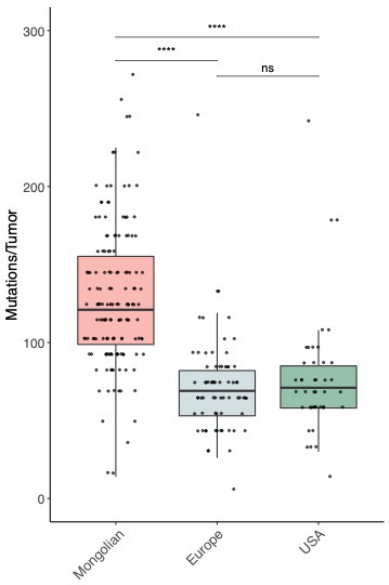

****
****
ns
Mutations/Tumor
300
200
100
0
Mongolian
Europe
USA

Figure 2

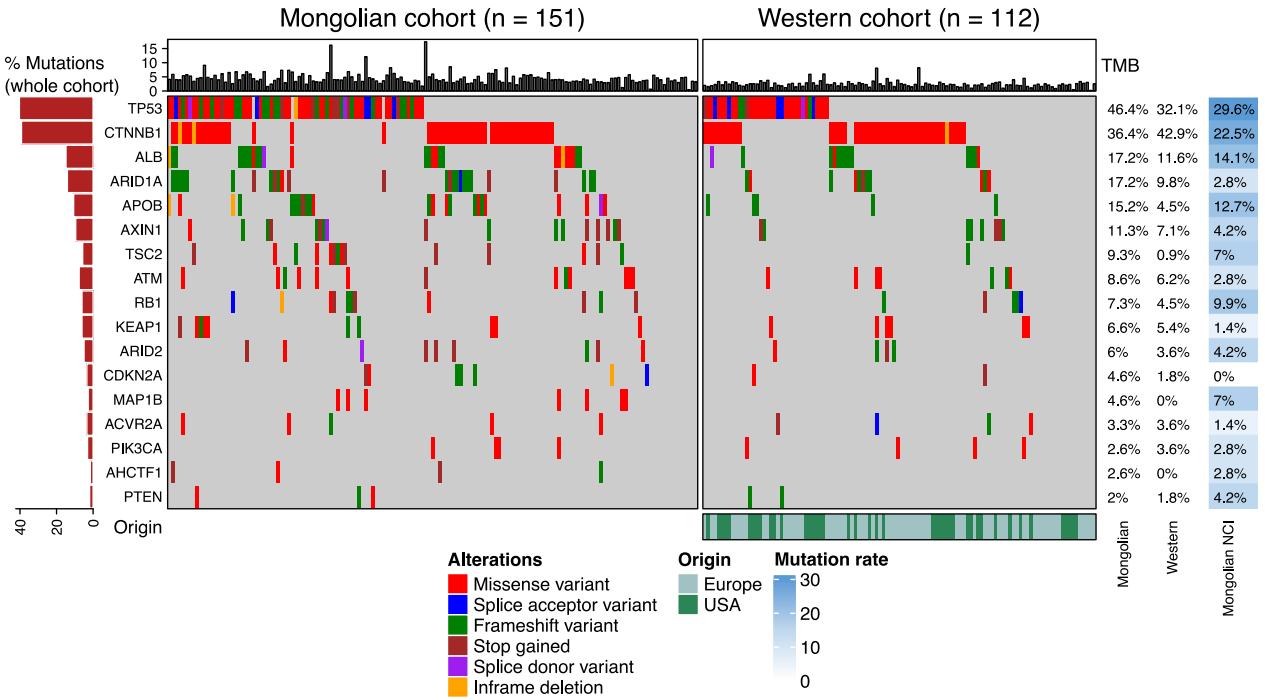

Mongolian cohort (n = 151)
Western cohort (n = 112)
% Mutations
(whole cohort)
TMB
TP53
CTNNB1
ALB
ARID1A
APOB
AXIN1
TSC2
ATM
RB1
KEAP1
ARID2
CDKN2A
MAP1B
ACVR2A
PIK3CA
AHCTF1
PTEN
Origin
Mongolian
Western
Mongolian NCI
Alterations
Missense variant
Splice acceptor variant
Frameshift variant
Stop gained
Splice donor variant
Inframe deletion
Origin
Europe
USA
Mutation rate

Figure 3

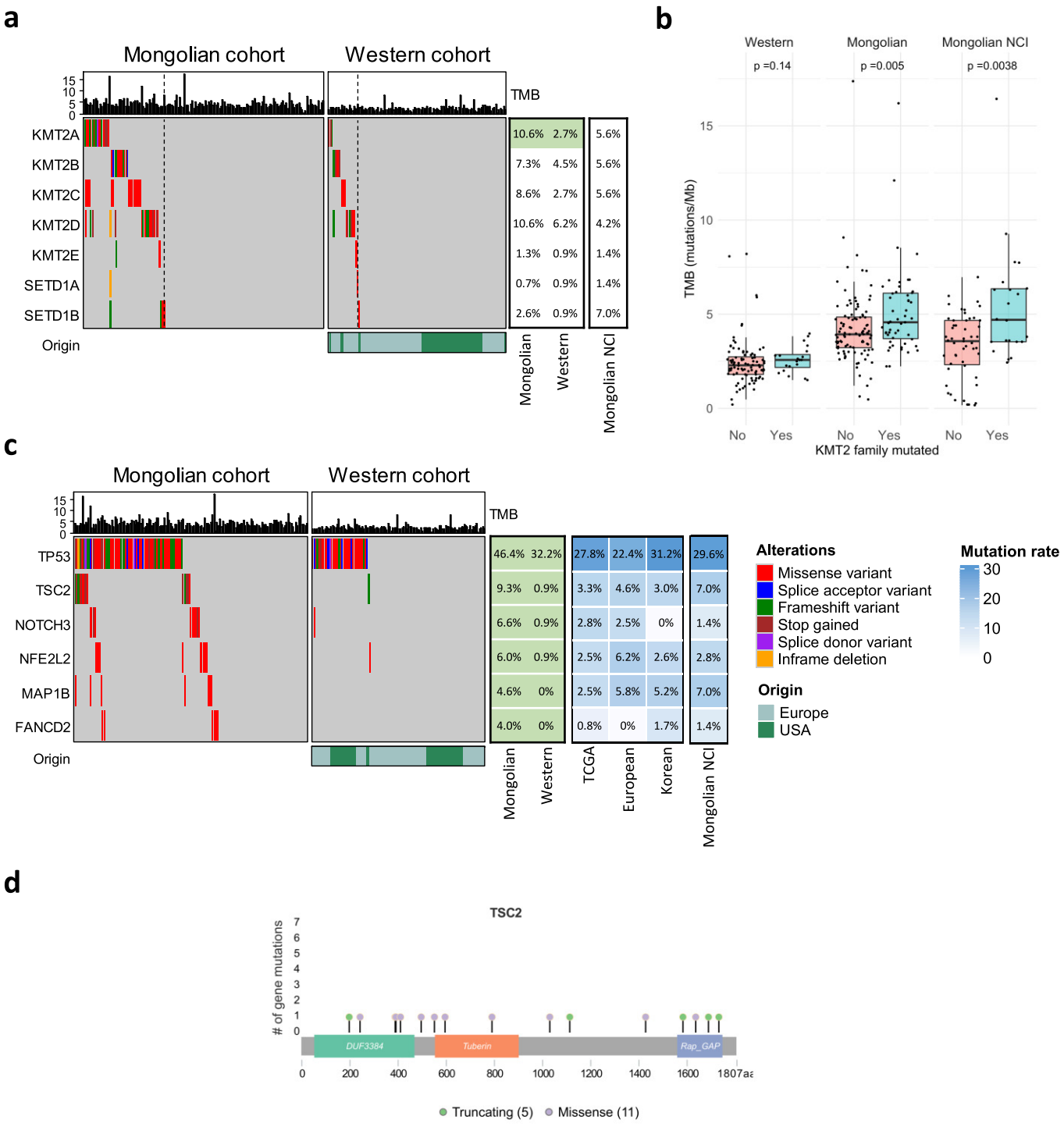
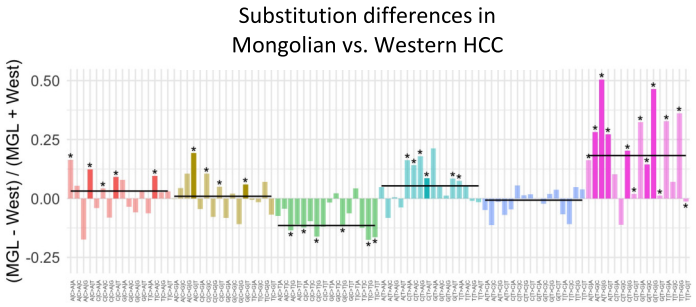

a
Mongolian cohort
Western cohort
TMB
KMT2A
KMT2B
KMT2C
KMT2D
KMT2E
SETD1A
SETD1B
Origin
Mongolian
Western
Mongolian NCI
b
Western
Mongolian
Mongolian NCI
p =0.14
p =0.005
p =0.0038
TMB (mutations/Mb)
No
Yes
KMT2 family mutated
c
Mongolian cohort
Western cohort
TMB
TP53
TSC2
NOTCH3
NFE2L2
MAP1B
FANCD2
Origin
Mongolian
Western
TCGA
European
Korean
Mongolian NCI
Alterations
Missense variant
Splice acceptor variant
Frameshift variant
Stop gained
Splice donor variant
Inframe deletion
Mutation rate
Origin
Europe
USA
d
TSC2
# of gene mutations
DUF3384
Tuberin
Rap_GAP
1807aa
Truncating (5)
Missense (11)

Figure 4

a


Substitution differences in
Mongolian vs. Western HCC
(MGL - West) / (MGL + West)
0.50
0.25
0.00
-0.25

b

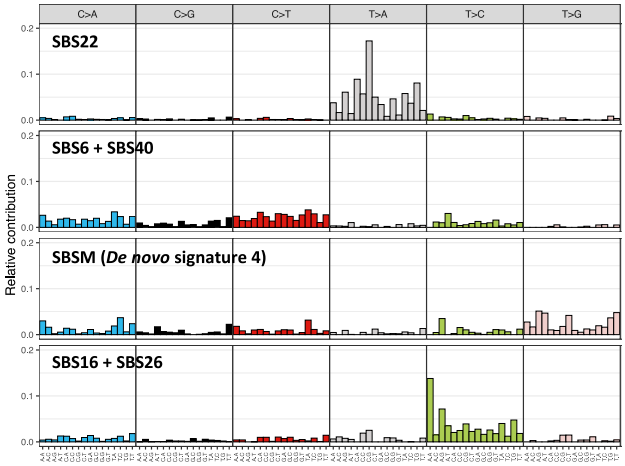

C>A
C>G
C>T
T>A
T>C
T>G
SBS22
SBS6 + SBS40
SBSM (De novo signature 4)
SBS16 + SBS26
Relative contribution

Figure 5

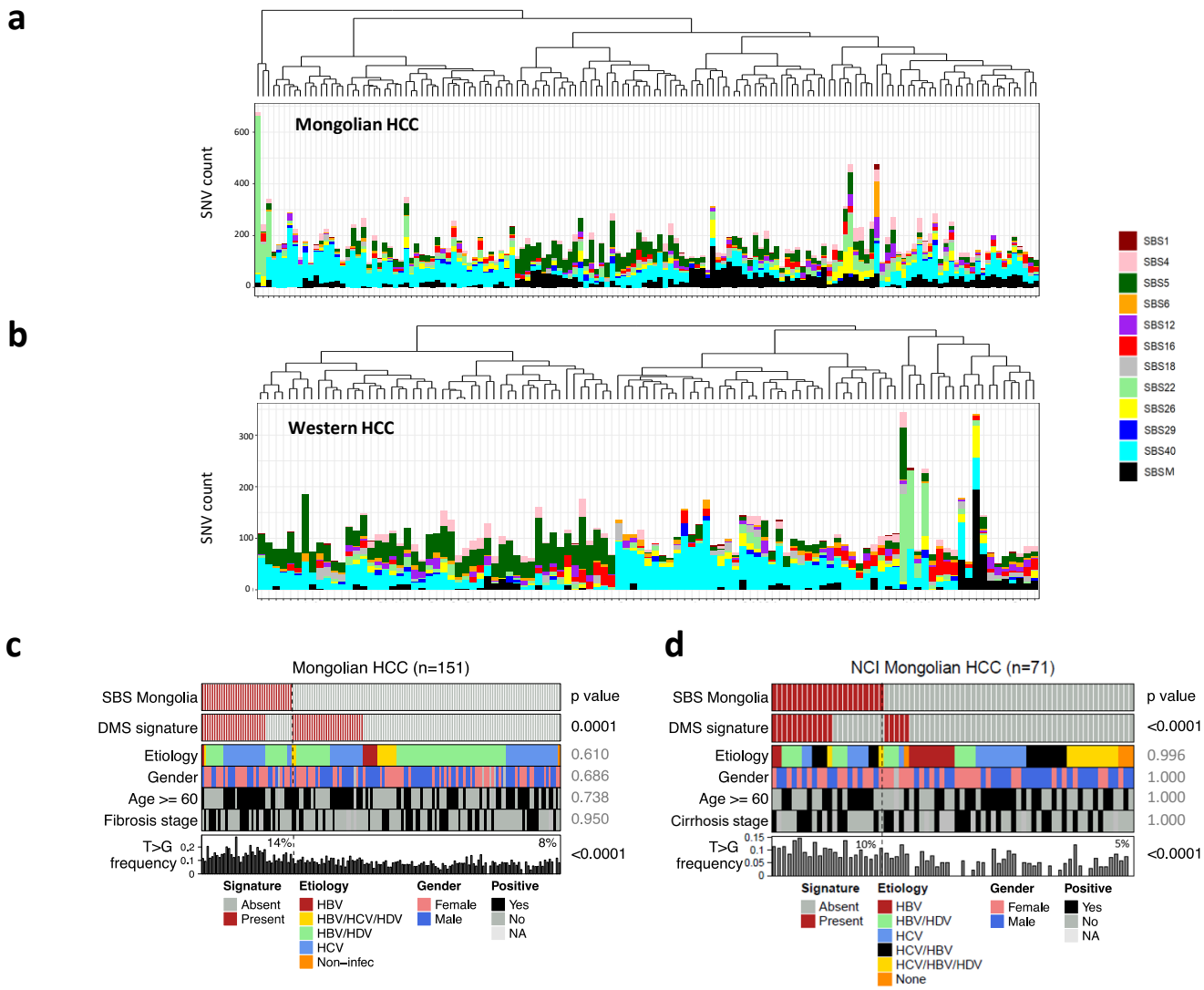
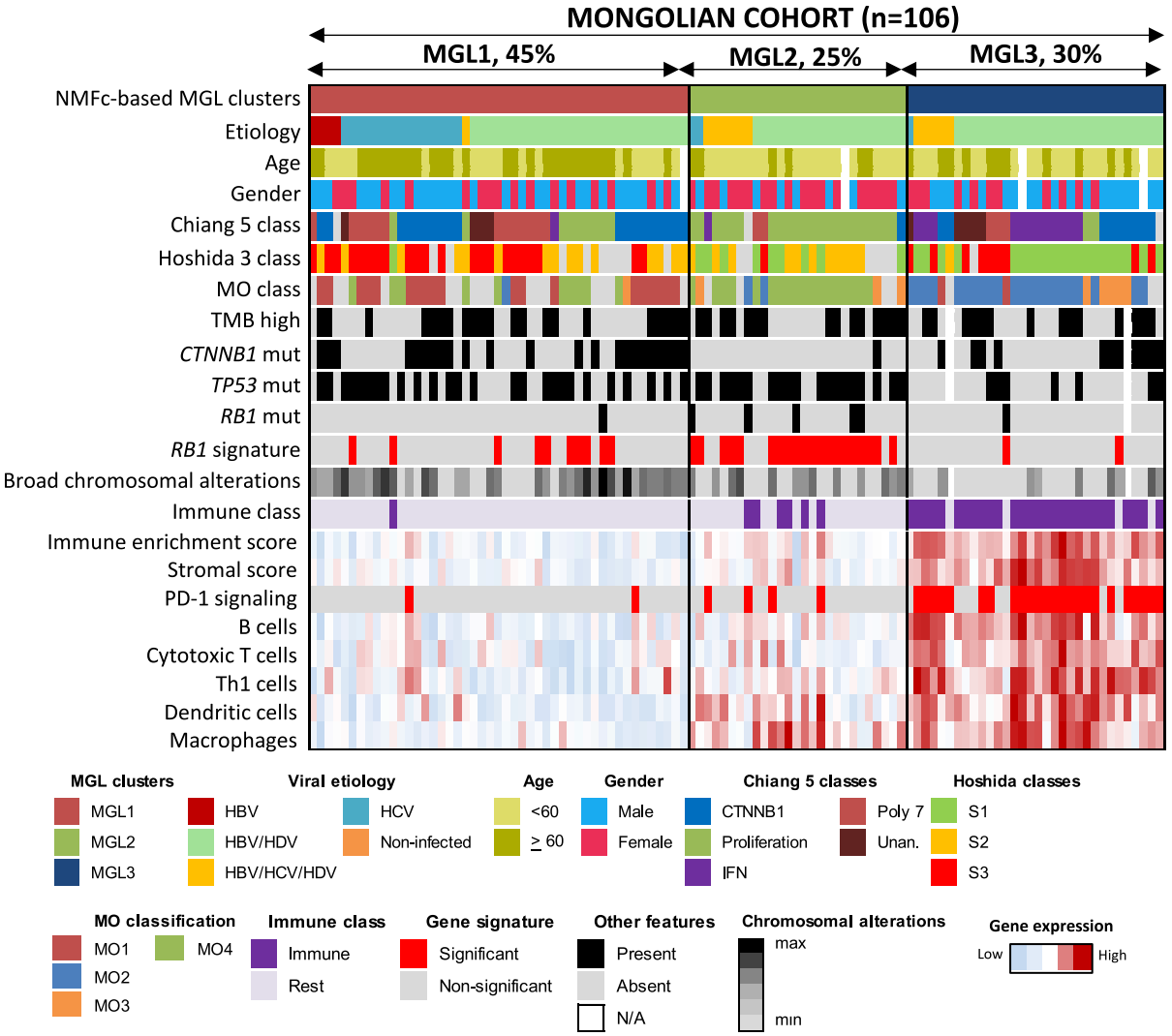

a
Mongolian HCC
SNV count
600
400
200
0
b
Western HCC
SNV count
300
200
100
0
SBS1
SBS4
SBS5
SBS6
SBS12
SBS16
SBS18
SBS22
SBS26
SBS29
SBS40
SBSM
c
Mongolian HCC (n=151)
SBS Mongolia
DMS signature
Etiology
Gender
Age >= 60
Fibrosis stage
T>G frequency
14%
8%
p value
0.0001
0.610
0.686
0.738
0.950
<0.0001
Signature
Absent
Present
Etiology
HBV
HBV/HCV/HDV
HBV/HDV
HCV
Non−infec
Gender
Female
Male
Positive
Yes
No
NA
d
NCI Mongolian HCC (n=71)
SBS Mongolia
DMS signature
Etiology
Gender
Age >= 60
Cirrhosis stage
T>G frequency
10%
5%
p value
<0.0001
0.996
1.000
1.000
1.000
<0.0001
Signature
Absent
Present
Etiology
HBV
HBV/HDV
HCV
HCV/HBV
HCV/HBV/HDV
None
Gender
Female
Male
Positive
Yes
No
NA

Figure 6


MONGOLIAN COHORT (n=106)
MGL1, 45%
MGL2, 25%
MGL3, 30%
NMFc-based MGL clusters
Etiology
Age
Gender
Chiang 5 class
Hoshida 3 class
MO class
TMB high
CTNNB1 mut
TP53 mut
RB1 mut
RB1 signature
Broad chromosomal alterations
Immune class
Immune enrichment score
Stromal score
PD-1 signaling
B cells
Cytotoxic T cells
Th1 cells
Dendritic cells
Macrophages
MGL clusters
MGL1
MGL2
MGL3
Viral etiology
HBV
HBV/HDV
HBV/HCV/HDV
HCV
Non-infected
Age
<60
≥ 60
Gender
Male
Female
Chiang 5 classes
CTNNB1
Proliferation
IFN
Poly 7
Unan.
Hoshida classes
S1
S2
S3
MO classification
MO1
MO2
MO3
MO4
Immune class
Immune
Rest
Gene signature
Significant
Non-significant
Other features
Present
Absent
N/A
Chromosomal alterations
max
min
Gene expression
Low
High