A Computational Approach to Predict the Role of Genetic Alterations in Methyltransferase Histones Genes With Implications in Liver Cancer

Cancer Informatics
Volume 22: 1–10

DOI: 10.1177/11769351231161480

Tania Isabella Aravena[1*], Elizabeth Valdés[1*], Nicolás Ayala[2] and Vívian D'Afonseca[3]

[1]Facultad de Ciencias Agrarias y Forestales, Universidad Católica del Maule, Talca, Chile.
[2]Departamento de Genética, Microbiología y Estadística, Universidad de Barcelona, España.
[3]Departamento de Ciencias Preclínicas, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile.

ABSTRACT: Histone methyltransferases (HMTs) comprise a subclass of epigenetic regulators. Dysregulation of these enzymes results in aberrant epigenetic regulation, commonly observed in various tumor types, including hepatocellular adenocarcinoma (HCC). Probably, these epigenetic changes could lead to tumorigenesis processes. To predict how histone methyltransferase genes and their genetic alterations (somatic mutations, somatic copy number alterations, and gene expression changes) are involved in hepatocellular adenocarcinoma processes, we performed an integrated computational analysis of genetic alterations in 50 HMT genes present in hepatocellular adenocarcinoma. Biological data were obtained through the public repository with 360 samples from patients with hepatocellular carcinoma. Through these biological data, we identified 10 HMT genes (*SETDB1, ASH1L, SMYD2, SMYD3, EHMT2, SETD3, PRDM14, PRDM16, KMT2C*, and *NSD3*) with a significant genetic alteration rate (14%) within 360 samples. Of these 10 HMT genes, *KMT2C* and *ASH1L* have the highest mutation rate in HCC samples, 5.6% and 2.8%, respectively. Regarding somatic copy number alteration, *ASH1L* and *SETDB1* are amplified in several samples, while *SETD3, PRDM14*, and *NSD3* showed a high rate of large deletion. Finally, *SETDB1, SETD3, PRDM14*, and *NSD3* could play an important role in the progression of hepatocellular adenocarcinoma since alterations in these genes lead to a decrease in patient survival, unlike patients who present these genes without genetic alterations. Our computational analysis provides new insights that help to understand how HMTs are associated with hepatocellular carcinoma, as well as provide a basis for future experimental investigations using HMTs as genetic targets against hepatocellular carcinoma.

KEYWORDS: Histone methyltransferases, hepatocellular carcinoma, somatic copy number alteration, patient survival and prognosis

RECEIVED: December 13, 2022. **ACCEPTED:** February 16, 2023.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared that they have no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Vívian D'Afonseca, Universidad Católica del Maule, Av. San Miguel 3605, Talca, 3460000, Chile. Email: vdafonseca@ucm.cl

Introduction

Liver cancer is the third leading cause of cancer deaths in the world, about 830 000 people worldwide died of this type of tumor in 2020, according to the World Health Organization (WHO).[1,2] Liver cancer has an incidence that represents the fifth most frequent cancer in men (7.5% of the total) and the ninth in women (3.4%), with an unfavorable prognosis (mortality/incidence: 0.95) and increasing incidence.[1-3] Hepatocarcinoma (HCC) represents 70 to 85% of all primary liver cancers diagnosed.[3,4] The initial process of HCC development has been mainly associated with liver cirrhosis, which often is related to chronic hepatitis virus B (HBV) infection, in a percentage range from 50 to 80% of HCC cases. The family history of primary liver cancer has also been associated with HCC, and a synergistic effect with HBV infection.[4-6]

In addition to risk factors, it has long been known that cancer cells, such as liver cancer cells, undergo genetic and epigenetic changes. Genomic analyses have revealed the widespread occurrence of mutations in epigenetic regulators and a large number of epigenome alterations in tumor cells.[7-9] However, it is known that genetic and epigenetic mechanisms influence each other and work cooperatively to allow the acquisition of the pathological features of cancer.[8-10] Recent studies have shown that dysregulation of a group of proteins called histone methyltransferases (HMTs) leads to aberrant histone methylation patterns and contributes to the pathogenesis of many human cancers,[11] including hepatocellular carcinoma.[11,12] Methylation of the lysine residue in histones, which is controlled by histone methyltransferases and demethylases, is an important player in epigenetic regulation. More than 50 HMTs have been identified in humans.[12,13] Structurally, HMTs are a diverse group of proteins that can be broadly classified into 2 functional enzyme families, the SET domain-containing methyltransferases (variegation suppressor, zeste enhancer, Trithorax) and the DOT1-like lysine methyltransferases.[14,15] Emerging evidence indicates that genetic alterations of several HMTs that have oncogenic or tumor suppressor functions play an important role in initiation and progression of cancer.[8,13,16] These aforementioned alterations can affect gene transcription (increasing the expression of oncogenes and/or suppressing tumor suppressor genes), DNA repair (increasing genomic instability), cell replication (allowing evasion of various checkpoints during the cell cycle), and contribute to the development of the carcinogenesis process and resistance to conventional therapies.[17-19]

*These authors share the same authorship in this work.

During the development of HCC, mutations and deregulations of histone-modifying enzymes have been described. For example, overexpression of EZH2, a methyltransferase of histone 3 lysine 27 (H3K27) that functions as a catalytic subunit of the repressive Polycomb complex 2, has been associated with poor prognosis and an aggressive phenotype in patients with HCC due to the regulation of genes related to resistance to chemotherapy.[20,21] A recent study demonstrated an interregulation between EZH2 and cell cycle-related kinase (CCRK) critical in the hepatocarcinogenic process and tumor progression.[22] Similarly, both JMJD1A, a histone 3 lysine 9 demethylase (H3K9) and KDM5B and LSD1 (histone 3 lysine 4 demethylases) are overexpressed in HCC patients and are associated with poor prognosis and increased invasiveness.[23] Another important HMT is KMT2C (lysine methyltransferase 2C) or also called MLL3 (mixed lineage leukemia 3), which is associated with the H3K4 methylation and is mutated in 8% of tumors in general.[24] Whole genome sequencing (WGS) studies have identified genetic alterations in genes and pathways involved in HCC development such as the KMT2B gene that is a cognate of the KMT2C gene.[25]

In addition, the upregulation of histone methyltransferase SETDB1 (forked SET domain 1), an epigenetic regulator responsible for methylating the amino acid residues (lysine 9) on histone H3 (H3K9), is associated with HCC progression, aggressiveness, and poor prognosis.[26] SETDB1 inactivation prevents cancer cells migration by eliminating of lung metastasis in mice model.[27] Finally, another HMT group, the NSD family (NSD1-3), which plays an important role in the expansion of different tumors,[28] when they are overexpressed, as in the case of the NSD1 gene, in tissues and cell lines, they are associated with a poor prognosis for HCC. The deletion of this gene has been shown to inhibit the proliferation, migration, and invasion of cancer cells.[29]

This new knowledge has positioned the enzymes involved in epigenetic pathways as new therapeutic targets for HCC. For this, it is important to know the behavior of methyltransferase alterations associated with various types of cancer in order to generate a specific genetic atlas that provides new target genes for cancer control. The computational analysis of the behavior of these HMTs genes in terms of their alterations, mutations, among other genetic and genomic aspects, opens the way to start new *in vitro* studies. Therefore, we propose to determine through a computational approach a specific genomic landscape for HMT in HCC, its relationship with patient prognosis, and to propose HMT genes as a new target for HCC management.

Methodology

Biological data of HMT genes associated to HCC from public database

Biological and clinical data were obtained through the public repository CBioPortal (www.cbioprtal.org).[30] For the analysis, it was used a total of 366 samples from patients with hepatocellular carcinoma. The data was harbored in the "Liver Hepatocellular Carcinoma (TCGA, Firehose Legacy)" study.[31] The analysis was started with the 50 HMT genes (*ASH1L*, *EHMT1*, *EHMT2*, *EZH1*, *EZH2*, *KMT2A*, *KMT2B*, *KMT2C*, *KMT2D*, *KMT2E*, *KMT5A*, *KMT5B*, *KMT5C*, *MECOM*, *NSD1*, *NSD2*, *NSD3*, *PRDM1*, *PRDM10*, *PRDM11*, *PRDM12*, *PRDM13*, *PRDM14*, *PRDM15*, *PRDM16*, *PRDM2*, *PRDM4*, *PRDM5*, *PRDM6*, *PRDM7*, *PRDM8*, *PRDM9*, *SETD1A*, *SETD1B*, *SETD2*, *SETD3*, *SETD4*, *SETD5*, *SETD6*, *SETD7*, *SETDB1*, *SETDB2*, *SETMAR*, *SMYD1*, *SMYD2*, *SMYD3*, *SMYD4*, *SMYD5*, *SUV39H1*, *SUV39H2*) described for human.

Categories of biological data associated to HCC used in the computational data mining

Only the samples that presented the data of interest (360 samples) were considered in our study. Follow is the categories of biological data used in this study.

a) *somatic mutations*: the mutations in the histone methyltransferase genes were identified using the COSMIC database (Catalog Of Somatic Mutation In Cancer), the searches were performed automatically by the repository used.

b) Somatic copy number alterations (SCNA): identified by the GISTIC 2.0 program. The program finds 5 categories of somatic copy number alterations (SCNA) and attributes a value to each one (deep deletion: −2; shallow deletion: −1; diploid: 0 (contains no alteration); gain: 1; and amplification: 2).

c) mRNA expression: Identified by RNASeq V2 Illumina sequencing methodology. The HMT gene expression data used were normalized with diploid samples that do not contain SCNA-like alteration (data provided by cBioPortal). Data is provided as relative gene expression and values are represented as Z-scores.

d) In addition, the clinical data of the patients were accessed, such as weight (kg), vital status (alive or dead), follow-up time of the cancer since diagnosis (months), genetic alterations mentioned above, and histological neoplasm grade (G1, G2, G3, and G4).

Genetic alteration map of HMT genes in HCC samples

We selected HMT genes with an overall genetic alteration rate greater than 14% in 360 hepatocellular carcinoma samples, proceeding with these genes in all subsequent steps of the study. The genes were the *SETDB1* (SET Domain Bifurcated Histone Lysine Methyltransferase 1), *ASH1L* (Histone-Lysine N-Methyltransferase ASH1L), *SMYD2* (SET And MYND Domain-Containing Protein 2), *SMYD3* (SET And MYND

Domain-Containing Protein 3), *EHMT2* (Euchromatic Histone Lysine Methyltransferase 2), *SETD3* (SET Domain Containing 3, Actin Histidine Methyltransferase), *PRDM14* (PR Domain Zinc Finger Protein 14), *PRDM16* (PR Domain Zinc Finger Protein 14), *KMT2C* (Lysine (K)-specific methyltransferase 2C), and *NSD3* (Nuclear Receptor Binding SET Domain Protein 3) genes. The genetic map indicates the genetic alterations (%) and describes the 10 HMT genes. This map was obtained from the OncoPrint program associated with the CBioPortal database.

Characterization of somatic mutation in HMT genes from HCC samples

Each HMT gene who presented a rate greater than 14% of genetic alterations in HCC samples reported in CBioPortal repository was evaluated regarding their mutation content. The types of mutations present in each gene were evaluated based on their descriptions. In addition, available annotations regarding biological function or effect of each mutation were evaluated for each gene studied. The figures of the mutated genes were generated in the MutationMapper program available within the repository used.

Correlation between mRNA expression and SCNA-like alterations for HMT genes in hepatocellular carcinoma samples

We evaluated the correlation between the relative values of gene expression and the presence of different SCNA-like alterations of the 10 HMT genes. For this analysis, *Z*-score data of relative expression of mRNA normalized with normal samples (diploid) and biological data of SCNA-like alterations (deep deletion, deletion, diploid, gain, and amplification) were used. For the correlation and validation analysis, the Spearman and Kruskal-Wallis non-parametric tests were used. Statistical significance was evaluated with *P*-values less than 0.05.

Stratification of patient data that presented genetic alteration in HMT genes

We evaluated the correlation between some clinical attributes of the patients such as altered genomic fraction, weight (kg) and histological neoplasm grade (G1-G4) with the presence and/or absence of the genetic alterations described in this study (somatic mutations, SCNA, and differential mRNA expression). The data of each studied gene *SETDB1, ASH1L, SMYD2, SMYD3, EHMT2, SETD3, PRDM14, PRDM16, KMT2C*, and *NSD3* which presented genetic alteration were associated with the clinical data that contained the mentioned clinical information. A pool of 67 unaltered samples was used in this analysis. To validate the analyses, non-parametric statistics such as *t* test (Mann-Whitney) and chi-square were used. Statistical significance was evaluated with *P*-values less than 0.05 and Kaplan-Meyer test.

Computational analysis of patient survival

The analysis of survival probability in individuals with hepatocellular carcinoma represented in the 360 samples was evaluated through the genetic alterations in the *SETDB1, ASH1L, SMYD2, SMYD3, EHMT2, SETD3, PRDM14, PRDM16, KMT2C*, and *NSD3* genes. We used as variables a group where the 10 gene samples are genetically altered with a group where these gene samples are not altered. For both groups, clinical data such as vital status (alive or dead) and lifetime (given in months) were considered. Individuals who died were censored from the statistics. This analysis provides a probability of survival of an individual over time. Statistical significance was evaluated with *P*-values less than 0.05. The error risks of the work are minimal, once the results have been produced computationally. The error risks that the work presents is, for instance, the propagation of an error from the database such as some wrong sample or wrong values. However, a statistic was applied to validate the analysis performed and reduce these risks.

Results

The database studied presents 366 samples from patients with hepatocellular carcinoma. However, we used 360 samples, which presented the genetic features selected for this study such as mutations, SCNA, and differential gene expression for HMT genes. The biological data is composed of 248 (67.8%) samples from men and 118 (32.2%) samples from women. Additionally, the rate of survival after 5 years was about 47% (overall survival) and the rate of disease free over time reached around 11% after 10 years, demonstrating a tendency for a decreased patient survival rate in this data set. The ethnicity linked to samples showed 322 patients no Hispanic (90.7%), 18 Hispanic or Latin (4.9%), and 16 without ethnicity classification (4.4%). Concerning the age of patients, the diagnosis time showed a higher number for 45 to 80 years old; however, the range between 65 and 70 years old presented a higher number of cases (67). Furthermore, the race of the patients comprised White (179—48.9%), Asian (158—43.2%), Black or African (17—4.6%), and American Indian (2—0.5%). A percentage of 2.7% of samples had no race declared. All samples were primary tumor. The samples used showed a notary prevalence in no Hispanic white and Asian people with age range around 65 to 70.[6] Other clinical data are summarized in Table 1.

Overview of genetic alteration in HMT genes from HCC

Among the 50 human HMT genes described in the literature, 10 of them were selected for the present study, based on their alteration rate found in HCC samples (more than 14%) (Figure 1). These genes were *SETDB1, ASH1L, SMYD2, SMYD3, EHMT2, SETD3, PRDM14, PRDM16, KMT2C*, and *NSD3*, which were altered in 293 (81%) of HCC studied samples. Additionally, these set of genes presented some types

Table 1. Summary of clinical data from patients with hepatocellular carcinoma (data from CBioPortal www.cbioportal.org, accessed in November, 2022)[a].

FEATURES	DESCRIPTION OF FEATURES	NUMBER OF SAMPLES
Tumor type	Hepatocellular carcinoma	366
Biological data	Mutations	362
	mRNA expression *z*-scores relative to diploid samples (RNA Seq V2 RSEM)	360
	Putative copy-number alterations from GISTIC	359
Gender	Male	248
	Female	118
Age	Average	57.5 years old
	Range	25-90
Race	White	179
	Asian	158
	Black or African	17
	American Indian	2
	NA	10
Ethnicity	Not Hispanic or Latin	332
	Hispanic or Latin	18
	NA	16
Tumor type	Primary	366
	No	220
	Yes	13
	NA	133
Metastasis stage	M0	264
	MX	98
	M1	4
Tumor stage code	T1	179
	T2	91
	T3	45
	T3a	28
	T4	13
	T3b	7
	TX	1
	NA	2

[a]All data summarized in this table are publicly available in the CBioPortal database and were not generated in this study.

of genetic alteration such as somatic mutation, SNCA-like alterations, and changes of mRNA expression. The rate of alteration varied from 14% (*PRMD16, KMT2C*, and *NSD3*) to 43% (SETDB1). The *SETDB1* (43%) and *ASH1L* (27%) were the most altered genes in HCC cohort. In contrast, *SETD3* (19%) and *NSD3* (19%) in terms of gene expression showed mRNA underexpression in several samples (heat map), Figure 1.

KMT2C *the most mutated HMT gene in hepatocellular cancer*

All 10 HMT genes showed somatic mutations in different proportions. The most common type of mutation was missense (Figure 2). The *ASH1L* and *KMT2C* genes presented a rate of 2.8% and 5.6% of somatic mutations in HCC samples. A total of 10 mutations in *ASH1L* gene were identified, consisting of 8 missense mutations (D2461G, N2876T, N401I, L734I, S1368G, P1198T, R1516P, I2797S), 1 frameshift mutation of the deletion type (I2935Sfs*9), and 1 mutation that compromises the splice process (X1696_splice). In *KMT2C* gene, a total of 20 mutations were identified, consisting of 15 nonsense mutations (S777C, R380C, A4683P, L3004V, S2004C, S1148P, C310F, P4119L, S842Y, T2008I, D2843N, H3323P, V9M, D3803G, Q3183G), 3 frameshift deletions (K4212Nfs*6, K2815Nfs*8, S965Vfs*66), 1 nonstop mutation (*4912Lext*36), and 1 nonsense mutation (E1623*). Five somatic mutations in *KMT2C* (K4212Nfs*6, K2815Nfs*8, S965Vfs*66, E1623*, and *4912Lext*36) present a biological effect as loss-of-function and a probable oncogenic effect.

The distribution of mutations occurred along all gene lengths for all 10 HMT genes. However, for *SETDB1* gene, all mutations described (S962*, N1234K, and I836Sfs*2) occurs inside the SET domain and for *EHMT2* gene, one missense mutation (E1016G) occurs inside the pre-SET domain.

Analysis of SCNA-like revealed that SETDB1, ASH1L, *and* PRDM14 *are amplified in hepatocellular cancer*

In particular, the 10 HMT genes showed different rates of SCNA-like alteration. However, genes with the highest amplification and deletion rates can be observed. *ASH1L* (48), *SETDB1* (43), and *PRDM14* (42) had high-level amplification in more than 10% of the samples, being considered the most amplified genes for this HCC cohort. In contrast, 3 HMT-encoding genes, *NSD3* (189), *PRDM16* (145), and *SETD3* (129), were deleted in more than 30% of hepatocellular carcinoma samples, demonstrating an expressive genetic loss for these elements.

Furthermore, regarding the level of mRNA expression, for example, SETDB1 (30%), SMYD2 (15%), and EHMT2 (14%) showed a distinct pattern of mRNA overexpression, with

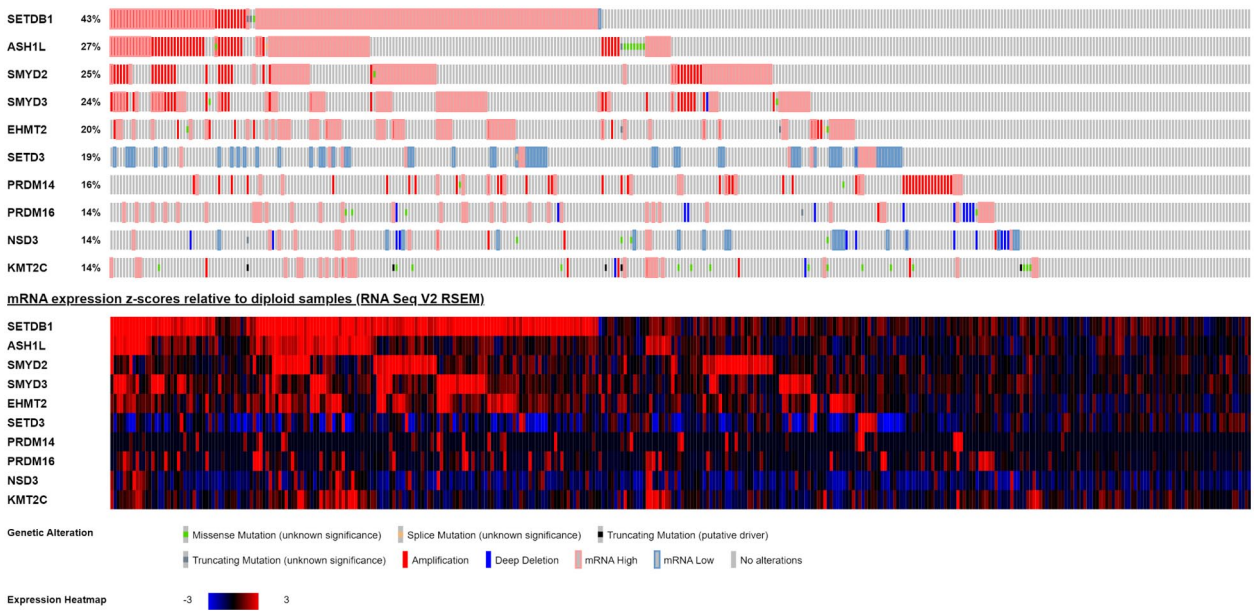

Figure 1. Overview of genetic alteration in most altered HMT genes: SETDB1, ASH1L, SMYD2, SMYD3, EHMT2, SETD3, PRDM14, PRDM16, KMT2C, and NSD3 from HCC samples.

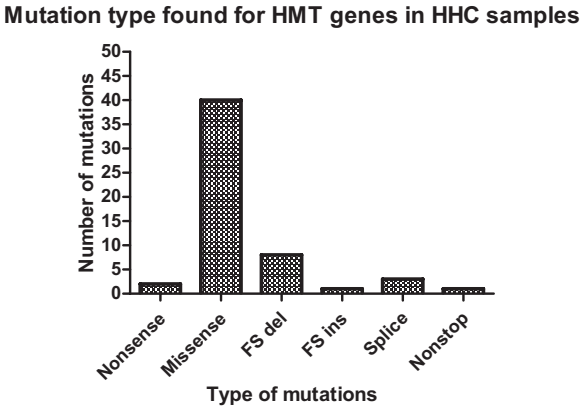

Figure 2. Summary of type of mutation found in the 10 HMT genes. FS del, frameshift deletion; FS ins, frameshift insertion.

SETDB1 being the most overexpressed HMT gene in the study. In contrast, the *SETD3* and *NSD3* genes showed the highest rates of mRNA underexpression in the HCC cohort, 13.7% and 4.5%, respectively. When comparing mRNA expression with the type of SCNA alteration, it is possible to observe a trend of increased expression of certain genes when there is gain or amplification and decreased expression when there is gene deletion. This type of gene response can be seen in Figure 3, which shows the analysis of correlation between mRNA expression and the type of SCNA alteration in the *SETD3, NSD3*, and *KMT2C* genes. The correlation between the genetic event of genomic gain or loss and level of mRNA expression is described for several genes present in malignant cells.[8,13]

Altered HMT genes decrease the overall patient survival

Concerning the analysis of overall patient survival, *SETDB1, SETD3, PRDM14*, and *NSD3* altered genes were linked with decreased survival probabilities in patients with HCC over time (months). Here are represented the only statistically significant cases within the 10 HMT genes studied. This data can be seen in Figure 4. On the *x*-axis it represents the time given in months, while on the *y*-axis it represents the probability of the patient with HCC surviving, either it present a certain HMT altered or not. Both groups of samples were considered, samples where the HMT gene were altered and samples that this same gene does not present any alteration. The crosses in the figure indicate the patients who died from the disease. The blue line indicates patients who do not have altered HMT gene, while the red line indicates patients who have altered HMT gene. Therefore, the graph shows that over time there are probabilities that these patients will survive or not, regarding absence or presence of alterations in HMT genes. Since as time passes, the probability of these people dying decreases or increases. Our results show a higher probability of survival of the patient belonging to a group whose HMTs genes are unaltered compared to the group which HMTs genes are altered for the 4 indicated cases. For example, for *SETD3* altered gene (Figure 4a), since the patient diagnosis time until 40 months after, it can be observed a decrease in survival probability, from 100% to 50%. In comparison with unaltered group, this decrease is weaker in this time (40 months), reaching around 70% of survival probability. After 40 months, the percentage of probability is the same in both groups. After 100 months, the probability to survival with *SETD3* altered is around 30%. However, after 80 months can be observed a slightly better scenario for patients with *SETD3* altered in contrast with unaltered group. Similar behavior is seen for *SETDB1* gene (Figure 4b). For *SETDB1*, in the first 60 to 70 months from the time of diagnosis, there is a difference in the survival profile between the unaltered group and the altered group. The

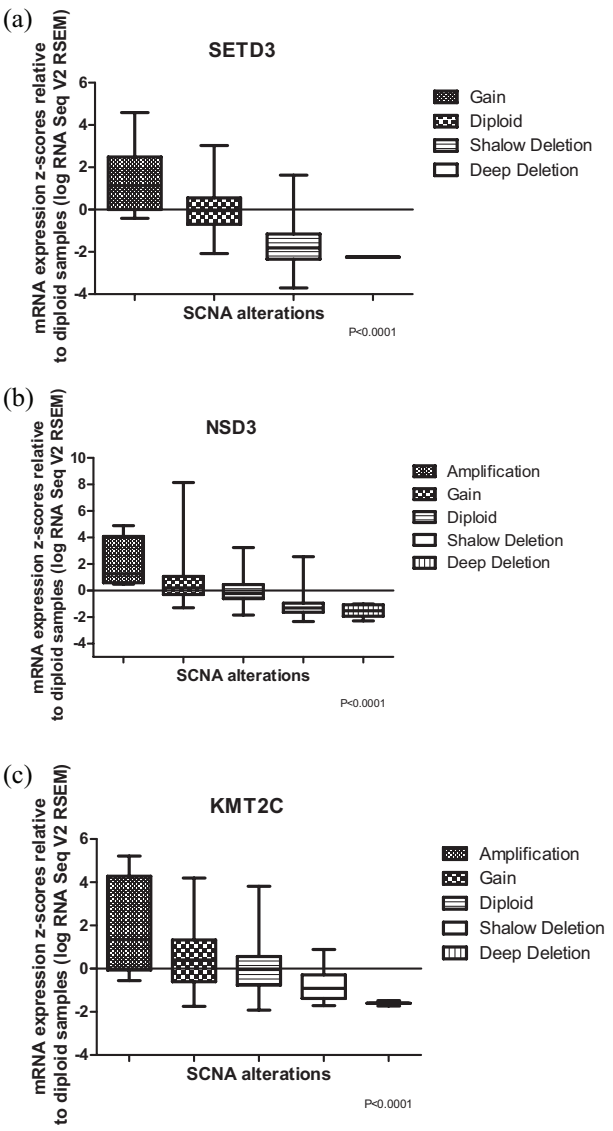

Figure 3. Analysis of correlation between SCNA-like alteration and mRNA expression level in HMT genes. (a) SETD3, (b) NSD3, and (c) KMT2C. *Y*-axis: mRNA expression *Z*-score values relative to diploid samples (log RNA Seq V2 RSEM). *X*-axis: SCNA-like alteration (amplification, gain, diploid (no alteration), shallow deletion, deep deletion). Data obtained from CBioPortal (www.cbioportal.org, accessed in November 2022). Statistical analysis: Kruskal-Wallis test, *P* value: .005.

survival probability is less for the altered group, from 100% to 50% in contrast to the group unaltered from 100% to 60%. However, after this period, the survival percentage remains the same in both groups. For *PRDM14* (Figure 4c) is observed a decrease of survival probability to 40% around 60 months and after 80 months the probability decrease until reach zero in the altered group. In the unaltered group, the probability also is lower after 100 months, around 15%. However, in the unaltered group, the patient can live more time in comparison the group with *PRDM14* altered. Finally, when the *NSD3* gene is altered before 20 months from patient diagnosis, the survival probability decrease from 100% to 60% in comparison to the unaltered group (Figure 4d). At 60 months, for the altered group, the probability decreases below 40%, and this pattern of decline continues over time, settling at 20% at 90-month follow-up.

Altered HMT genes are probably involved in HCC progression

The clinical data of the patients were stratified to perform various statistical analyzes to find the relationship between the genetic alterations in the HMT genes and some attributes of the patients such as weight and cancer progression described in the neoplasm histologic grade (G1-G4) (Figure 5). We found that HMT genes such as *SETDB1* and *EHMT2* are more altered in patients with lower body weight whereas in patients who do not show genetic alteration in this group of HMT genes the weight is higher. For the *SETDB1* gene (Figure 5a), the patients with this altered gene presented their body weight in a range of 60 kg to 80 kg. In the unaltered group, this range was from 60 kg to almost 100 kg. Similar results were observed for the *EHMT2* gene (Figure 5b), where the altered group and the unaltered group presented a similar pattern for weight range values.

Another result found is related to the neoplasm histologic grade. For the *SETDB1* and SMYD3 genes, it is observed that the advanced grade G3 mainly, and G4 for *SETDB1* is found in a higher proportion in the altered group than in the unaltered group (Figure 5c and d), suggesting a tendency to cancer spread and progression in these altered group in comparison to unaltered group.

Discussion

In the last decades, a biological revolution has taken place, in which an enormous amount of biological/biomedical information has been generated and made available in public databases. Many databases and computational tools are created daily to harbor and deliver new biological information about all forms of life. For humans, many databases share biological/biomedical information in raw or processed format, be it genomics, proteomics, metabolomics, and drug design, for example. However, rather than making the data available, it is important to process and extract useful information from these repositories, which could answer various questions about human health.

One way to get to this point is through *in silico* analysis of genetic and genomic data from diseases, such as cancer. Another important point is to choose well the useful data set to answer biological questions that affect human health through the computational approach. In our work, we selected a group of genes that could be involved in epigenetic regulation in both healthy cells and malignant cells, acting in different ways. Because epigenetic changes are reversible, they provide a unique opportunity for pharmacological intervention through inhibitors designed as a new class of anticancer drugs.[32,33] Recent evidence shows that the aberrant activity of HMTs, due to amplification, deletion, or mutation of their corresponding genes, contributes to the initiation and progression of

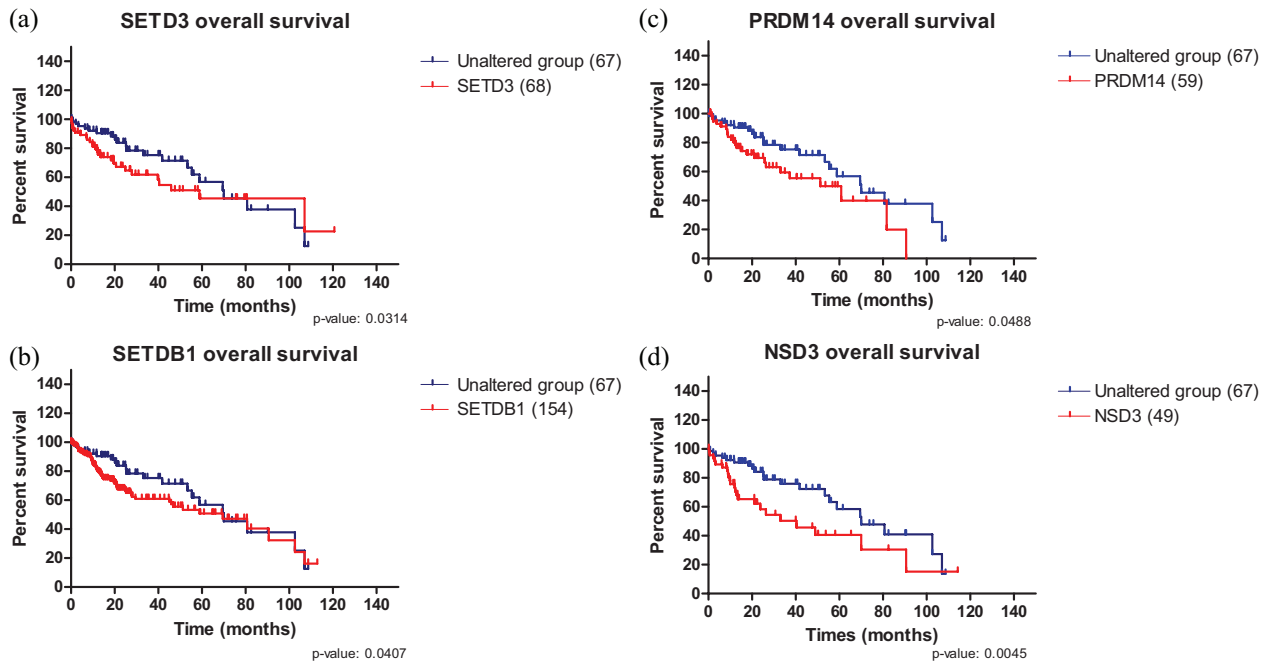

Figure 4. Analysis of overall patient survival linked to altered HMT genes: (a) SETD3, (b) SETDB1, (c) PRDM14, and (d) NSD3. *Y*-axis: percentage of survival probability (%). *X*-axis: time in months. Data obtained from CBioPortal (www.cbioportal.org, accessed in November 2022). Statistical analysis: Kaplan-Meyer test, *P* value: .005.

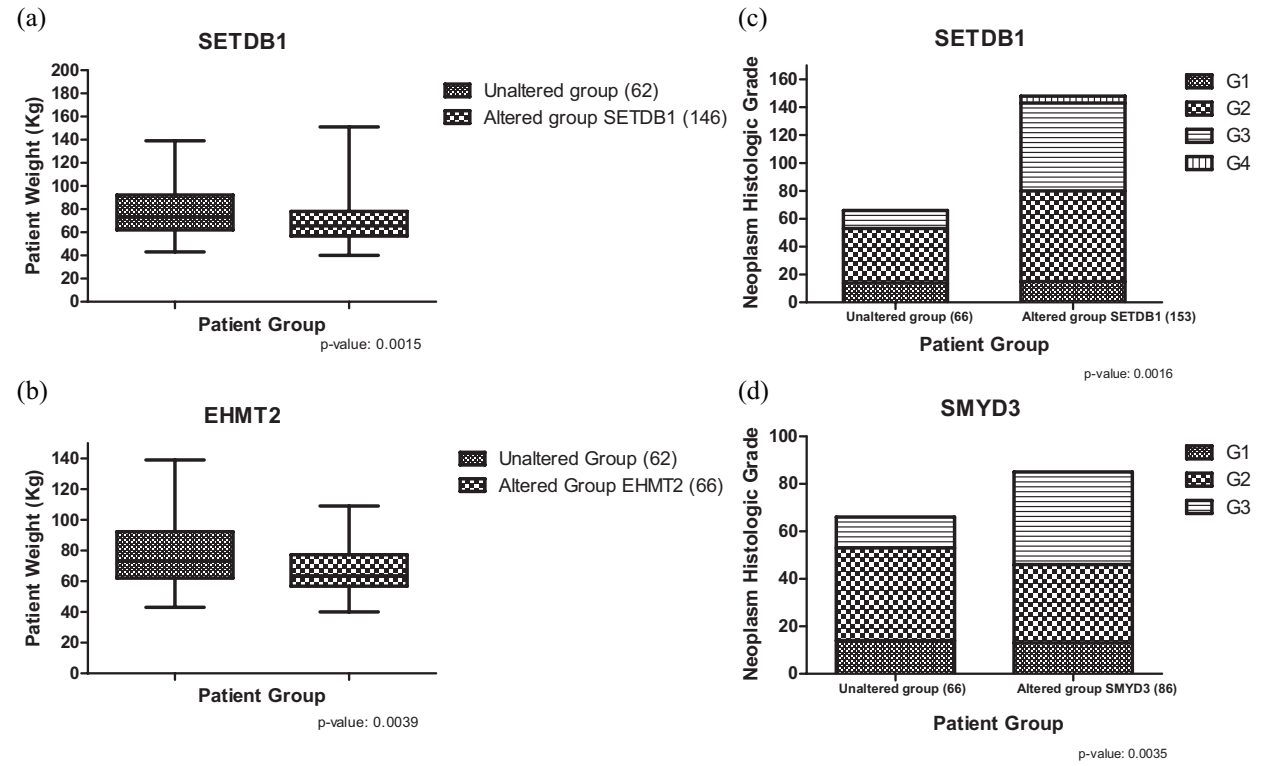

Figure 5. Statistical analysis of relationship between patient clinical attributes and alteration in HMT genes: (a) SETDB1 and (b) EHMT2. For a and b graphics, *Y*-axis: weight; *X*-axis: patient group (altered and unaltered). Statistical analysis: Mann Whitney test, *P* value: .005. (c) SETDB1 and (d) SMYD3. For c and d graphics, *Y*-axis: number of cases in each neoplasm histologic grade; *X*-axis: patient group (altered and unaltered). Statistical analysis: Chi-square test, *P* value: .005. Data obtained from CBioPortal (www.cbioportal.org, accessed in November 2022).

cancer.[8,13,33] Consequently, a promising strategy could target populations of patients who are carriers of these alterations. For years, it has been known that different types of cancer share common molecular mechanisms whose dysfunction allows uncontrolled cell proliferation through the deregulation or mutation of genes that positively or negatively influence the

regulation of cell proliferation, migration and differentiation.[34] Although the genetic term generally leads to understanding cancer as a hereditary disease, this only occurs in a small percentage. In most tumors, the alterations described are only somatic and, therefore, cannot be transmitted to offspring[35,36] so it is important to study mutations linked to cancers. Here we evidence 55 mutations in the 10 HMTs genes studied. Although none of them had a biological effect annotation, it is known that indels (insertions and deletions) mutations, as well as premature termination of protein synthesis, have serious implications in the functioning of the affected protein.[37,38] According to the My Cancer Genome database (www.mycancergenome.org), the *KMT2C* gene encoding a histone methyltransferase H3K4 is mutated at a rate higher than 5% in many solid tumors. In the present study, the same mutation pattern was found,[39] which this gene presented a mutation rate of 5.6% with the presence of many nonsense mutations that could lead to its inactivation. It is known that the disruption of *KMT2C* could be related to the cancer process through transcriptional deregulation in several pathways.[40] This hypothesis is described for colorectal cancer,[40] for example, and the data analyzed have shown a similar pattern of genetic alteration, which could have implications in HCC cancer.

In addition, for some time the mutagenic mechanisms have not explain all cancer of cases. Thus, from the etiopathogenic point of view, epigenetic changes have been implicated in the development of different cancers.[8,13] Another highly relevant alteration, are somatic alterations in the number of gene copies (SCNA), which have been widely described in cancer cells. The amplification of gene regions that encompass total genes are seen as driver alterations, they lead to the initiation of a process of cellular malignancy. We found that *ASH1L, SETDB1*, and *PRDM14* are at a high level of amplification, which could reveal that they are located in genomic regions that are amplified in around 10% of the studied samples, often leading to increased of their expressions, as we demonstrated in that study. In contrast, there are the *NSD3, PRDM16*, and *SETD3* genes, which present homozygous deletion in 30% of HCC samples, reducing their expressions in HCC studied samples. Both amplification (gain of function) and deletion (chromosomal instability and loss of function) could lead to cancer development processes, which is largely being determined experimentally.[41,42]

With respect to alterations in the expression of the HMT gene, *SETDB1*, is upregulated in our study and several other human cancers, such as ovarian cancers, endometrial cancers, lung adenocarcinoma, breast cancers, and HCC. Upregulation of *SETDB1* leads to several alterations in HCC tissues.[27,43] Another HMT gene found upregulated is the *ASH1L*, which encodes a member of the trithorax group of transcriptional activators, also is overexpression in liver cancer.[17,44] Furthermore, *SMYD2* and *SMYD3* genes presented high expression of in the current investigation. The *SMYD2* gene have been reported with mRNA overexpressed in pediatric acute lymphoblastic leukemia, gastric and liver cancer. Some research articles linked *SMYD2* gene with inhibitory functions in tumor suppressor proteins such as p53, Rb, and PTEN.[45-47] Additionally, *SMYD3* has been linked to several human cancers. High levels of this enzyme are expressed in colorectal, liver, and breast cancers.[48,49] Three other HMTs genes presented overexpression, *EHMT2* gene, has been related with fundamental functions in embryogenesis in genetic mouse models. Deletion of *EHMT2* gene in mice resulted in embryonic lethality.[50,51] The *EHMT2* overexpression has been reported in several types of cancers such as lung cancer, multiple myeloma, ovarian carcinoma, and liver cancer.[52,53] In addition, such overexpression of the *EHMT2* gene is associated with decreased of patient survival.[54,55] Already the *PRDM14* gene plays an important role in resetting and maintaining pluripotency in embryonic cells. *PRDM14* expression has not been detected in healthy adult tissues; however, genomic amplification, methylation and misexpression of *PRDM14* have been detected in several cases of human tumors. In addition, *PRDM14* gene has been associated with initiation of several cancers.[56,57] Finally, *KMT2C* are often deleted in myeloid leukemias[58] in our findings, this HMT gene present an elevation in its expression.

However, in our results, *SETD3* and *NSD3* showed an under expression in several samples. Generally, *NSD3* gene are amplified in some cancer such as cancer colorectal.[13] Already SETD3 gene, a SET domain-containing 3 (SETD3), member of the protein lysine methyltransferase family and with function to catalyze the addition of methyl group to lysine residues[59] are over expressed in lymphoma, kidney tumor, and invasive breast cancer.[60,61,62] Another research showed that *SETD3* gene level is correlated with cell proliferation of liver cancer cells in a xenograft mouse model.[63]

Regarding the relation between alterations in HMT genes and survival patient, our study revealed that *SETDB1, PRDM14, SETD3*, and *NSD3*, could affect the survival patient leading to poor prognosis. *SETDB1* is overexpression in several cancers such as breast cancer, non-small cell lung cancer, prostate cancer, colorectal cancer, acute myeloid leukemia, glioma, melanoma, pancreatic ductal adenocarcinoma, liver cancer, nasopharyngeal carcinoma, gastric carcinoma, and endometrial cancer. In colorrectal cancer, this genetic alteration is related with poor prognosis and decrease of survival of patients.[26] Similarly, *NSD3* is described as a regulator of the apoptotic process of lymphocytes. The high expression of *NSD3* could play an important role in the progression of breast cancer and colorectal cancer, leading to the worst prognosis.[13,64] Here, we report these genes as likely targets to better study HCC prognosis, since they are involved in decreased patient survival. In addition, these results could open up new treatment pathways and establish new protocols for evaluating the prognosis of patients. In addition, the *SETD3* gene in patients with breast cancer (triple negative) is associated with a poor prognosis. If the patient harbors a mutation in the p53 genes in addition to altered expression of *SETD3*, their prognosis is worse even in

patients with ER-positive tumors.[62] Finally, for the *PRDM14* gene, an abnormal expression associated with metastasis and invasion is observed in patients with colorectal cancer. In addition, *PRDM14* overexpression is related to stage III colorectal cancer, which enhanced the invasive, drug-resistant, and *in vitro* cell dividing properties of the colon cancer cells.[65]

An important finding presented here is about the relationship between the patient's weight and genetic alterations in the *SETDB1* and *EHMT2* genes. Both genes are more altered in patients with lower weight compared to the control group (without alteration in both genes). In the control group, the patients showed a higher body mass index (BMI). It is known that the body mass index could be closely related to the prognosis and mortality of various diseases such as cancer.[66] For example, overweight men have a better prognosis than normal-weight men with HCC; however, normal-weight women have a better prognosis than overweight women.[66] No study have related the weight of HCC patients with alterations in the *SETDB1* and *EHMT2* gene. It is likely that these findings could indicate a worse prognosis for patients with alterations in *SETDB1* and *EHMT2* and who presented a decrease in their BMI.

Our findings indicated that *SETDB1* and *SMYD3* are more altered in patients with histological grades 3 and 4 than in the control group (without alteration in any grade G1-G4 for these genes). This finding likely indicates that *SETDB1* and *SMYD3*, when disrupted, might be involved in the process of HCC growth and propagation. *SETDB1* was related to the histological grade (initial grade) in patients with breast cancer.[67] *SMYD3* is a histone methyltransferase previously linked to cancer cell invasion and migration. In breast cancer, it promotes the epithelial-mesenchymal transition in breast cancer.[68] These alterations in advanced histological neoplasm degrees such as G3 and G4, as demonstrated here, could represent their roles as invasion and migration factors in HCC cells.

It is interesting to note that the *SETDB1* gene probably has valuable importance in HCC. *SETDB1* was one of the most altered genes presenting a high level of amplification and it is overexpressed in several samples. In addition, it was associated with decreased patient survival, it is altered in patients with reduced body mass index and its genetic alterations seen here are related to advanced histological neoplasm grades (G3-G4). All these features lead to the understanding of *SETDB1* as an HCC cancer driver gene and *in vitro* experiments might be needed to better study the behavior of this gene in HCC.

Conclusions

For the scientific community, it is important that different types of biological data are available in public repositories. Computational analyzes allow us to understand the details of biological processes faster and more accurately. For several studies, the first steps are with a computational approach. This type of analysis can improve accuracy and drive experiments for further *in vitro* studies. Currently, instead of generating a significant amount of biological data, it is important to extract useful biological information from available genomic data, for example. Therefore, biological data mining can provide the crucial foundation for experimental research. In this approach, through the extraction of biological data from public repositories, we were able to identify that certain alterations in HMT genes could likely be involved with hepatocellular carcinoma.

Our findings strongly evidenced that genetic alterations such as somatic mutation, SCNA, and gene expression changes of HMT genes, may play an important role in the generation and development of hepatocellular carcinoma, laying a solid foundation for future studies. Furthermore, our work provides a genetic content for future studies, which could use HMT genes such as *SETDB1, PRMD14, NSD3*, and *KMT2C* in *in vitro* assays as interesting targets for HCC study.

Author contributions

VD: Conception of the research, organizing of database, analysis of data, writing the manuscript. EV, NA, and TIA: analysis of data, writing the manuscript.

REFERENCES

1. World Health Organisation. 2022. Cancer General Data. Accessed December 5, 2022. https://www.who.int/news-room/fact-sheets/detail/cancer/
2. Rumgay H, Arnold M, Ferlay J, Lesi O, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. *J Hepatol*, 2022;77:1598-1606.
3. Ahmed F, Perz JF, Kwong S, Jamison PM, Friedman C, Bell BP. National trends and disparities in the incidence of hepatocellular carcinoma, 1998-2003. *Prev Chronic Dis*. 2008;5:A74.
4. Erstad DJ, Tanabe KK. Prognostic and therapeutic implications of microvascular invasion in hepatocellular carcinoma. *Ann Surg Oncol*. 2019;26:1474-1493.
5. Loomba R, Liu J, Yang HI, et al. Synergistic effects of family history of hepatocellular carcinoma and hepatitis B virus infection on risk for incident hepatocellular carcinoma. *Clin Gastroenterol Hepatol*. 2013;11:1636-45.e1-3.
6. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepatocellular carcinoma. *Hepatology*. 2021;73: 4-13.
7. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3:415-428.
8. Reyes DA, Saure-Sarría VM, Salazar-Viedma M, D'Afonseca V. Histone methyltransferases useful in gastric cancer research. *Cancer Inform*. 2021;20: 1-11.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646-674.
10. Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153:38-55.
11. Liu L, Kimball S, Liu H, Holowatyj A, Yang ZQ. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget*. 2015;6:2466-2482.
12. Tian X, Zhang S, Liu H-M, et al. Histone lysine-specific methyltransferases and demethylases in carcinogenesis: new targets for cancer therapy and prevention. *Curr Cancer Drug Targets*. 2013;13:558-579.
13. D'Afonseca V, Gónzalez G, Salazar M, Arencibia AD. Computational analyses on genetic alterations in the NSD genes family and the implications for colorectal cancer development ecancer. *Ecancermedicalscience*. 2020;14:1001.
14. Roche J, Bertrand P. Inside HDACs with more selective HDAC inhibitors. *Eur J Med Chem*. 2016;121:451-483.
15. Wahid B, Ali A, Rafique S, Idrees M. New insights into the epigenetics of hepatocellular carcinoma. *Biomed Res Int*. 2017;2017:1609575.
16. Wolinska E, Skrzypczak M. Epigenetic changes affecting the development of hepatocellular carcinoma. *Cancers*. 2021;13(16):4237.
17. Jones M, Chase J, Brinkmeier M, Xu J, et al. Ash1l controls quiescence and self-renewal potential in hematopoietic stem cells. *J Clin Invest*. 2015;125:2007.
18. Fernández-Barrena MG, Arechederra M, Colyn L, Berasain C, Avila MA. Epigenetics in hepatocellular carcinoma development and therapy: the tip of the iceberg. *JHEP Rep*. 2020;2:100167.

19. Khemlina G, Ikeda S, Kurzrock R. The biology of Hepatocellular carcinoma: implications for genomic and immune therapies. *Mol Cancer*. 2017;16(1):149.
20. Xiao G, Jin LL, Liu CQ, et al. EZH2 negatively regulates PD-L1 expression in hepatocellular carcinoma. *J Immunother Cancer*. 2019;7:300.
21. Revia S, Seretny A, Wendler L, et al. Histone H3K27 demethylase KDM6A is an epigenetic gatekeeper of mTORC1 signalling in cancer. *Gut*. 2022;71: 1613-1628.
22. Sun H, Yang W, Tian Y, et al. An inflammatory-CCRK circuitry drives mTORC1-dependent metabolic and immunosuppressive reprogramming in obesity-associated hepatocellular carcinoma. *Nat Commun*. 2018;9:5214.
23. Wu ZH, Yang DL, Wang L, Liu J. Epigenetic and immune-cell infiltration changes in the tumor microenvironment in hepatocellular carcinoma. *Front Immunol*. 2021;12:793343.
24. Rabello D, Ferreira VD, Berzoti-Coelho MG, et al. MLL2/KMT2D and MLL3/KMT2C expression correlates with disease progression and response to imatinib mesylate in chronic myeloid leukemia. *Cancer Cell Int*. 2018;18: 26.
25. Kan Z, Zheng H, Liu X, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res*. 2013;23:1422-1433.
26. Lazaro-Camp VJ, Salari K, Meng X, Yang S. SETDB1 in cancer: overexpression and its therapeutic implications. *Am J Cancer Res*. 2021;11:1803-1827.
27. Wong CM, Wei L, Law CT, et al. Up-regulation of histone methyltransferase SETDB1 by multiple mechanisms in hepatocellular carcinoma promotes cancer metastasis. *Hepatology*. 2016;63:474-487.
28. Yi L, Yi L, Liu Q, Li C. Downregulation of NSD3 (WHSC1L1) inhibits cell proliferation and migration via ERK1/2 deactivation and decreasing CAPG expression in colorectal cancer cells. *Onco Targets Ther*. 2019;12:3933-3943.
29. Zhang S, Zhang F, Chen Q, Wan C, Xiong J, Xu J. CRISPR/Cas9-mediated knockout of NSD1 suppresses the hepatocellular carcinoma development via the NSD1/H3/Wnt10b signaling pathway. *J Exp Clin Cancer Res*. 2019;38:467.
30. Cerami E, Gao J, Dogrusoz U, et al. CBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401-404.
31. Broad Institute TCGA Genome Data Analysis Center. *Analysis Overview for Liver Hepatocellular Carcinoma (Primary Solid Tumor Cohort*. Broad Institute of MIT and Harvard; 2016.
32. Moldón YR, Lara AE, Leyva BG, Cabodevilla RC. Tratamiento del cáncer basado en la epigenética. *Correo Científico Médico* 2021;25:1-25.
33. Yang C, Zhang J, Ma Y, Wu C, Cui W, Wang L. Histone methyltransferase and drug resistance in cancers. *J Exp Clin Cancer Res*. 2020;39:173.
34. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. 2018;173(2):321-337.e10.
35. Hurtado C, Encina G, Wielandt A, et al. Prevalencia y características de mutaciones somáticas del gen KRAS en pacientes chilenos con cáncer colorrectal. *Revista médica de Chile* 2014;142:1407-1414.
36. Frank SA, Nowak MA. Problems of somatic mutation and cancer. *Bioessays*. 2004;26:291-299.
37. Zhang M, Yang D, Gold B. Origins of nonsense mutations in human tumor suppressor genes. *Mutat Res*. 2021;823:111761.
38. D'Afonseca V, Arencibia AD, Echeverría-Vega A, et al. Identification of altered genes in gallbladder cancer as potential driver mutations for diagnostic and prognostic purposes: a computational approach. *Cancer Inform*. 2020;19:1-14.
39. The AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discovery*. 2017;7:818-831.
40. Larsson C, Cordeddu L, Siggens L, et al. Restoration of KMT2C/MLL3 in human colorectal cancer cells reinforces genome-wide H3K4me1 profiles and influences cell growth and gene expression. *Clin Epigenet*. 2020;12:74.
41. Chen Y, McGee J, Chen X, et al. Identification of druggable cancer driver genes amplified across TCGA datasets. *PLoS ONE*. 2014;9:e98293.
42. Cheng J, Demeulemeester J, Wedge DC, et al. Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors. *Nat Commun*. 2017;8:1221.
43. Fei Q, Shang K, Zhang J, et al. Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. *Nat Commun*. 2015; 6:8651.
44. Li G, Ye Z, Shi C, et al. The histone methyltransferase Ash1l is required for epidermal homeostasis in mice. *Sci Rep*. 2017;7:45401.
45. Huang J, Perez-Burgos L, Placek BJ, et al. Repression of p53 activity by Smyd2-mediated methylation. *Nature*. 2006;444:629-632.
46 Saddic LA, West LE, Aslanian A, et al. Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem*. 2010;285:37733-37740.
47. Nakakido M, Deng Z, Suzuki T, Dohmae N, Nakamura Y, Hamamoto R. Dysregulation of AKT pathway by SMYD2-mediated lysine methylation on PTEN. *Neoplasia*. 2015;17:367-373.
48. Hamamoto R, Silva FP, Tsuge M, et al. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci*. 2006;97:113-118.
49. Hamamoto R, Furukawa Y, Morita M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol*. 2004;6:731-740.
50. Casciello F, Windloch K, Gannon F, Lee JS. Functional role of G9a histone methyltransferase in cancer. *Front Immunol*. 2015;6:487.
51. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87-108.
52. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:l1.
53. Lu C, Ward PS, Kapoor GS, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. 2012;483:474-478.
54. Hua K-T, Wang M-Y, Chen M-W, et al. The H3K9 methyltransferase G9a is a marker of aggressive ovarian cancer that promotes peritoneal metastasis. *Mol Cancer*. 2014;13:189.
55. Wozniak RJ, Klimecki WT, Lau SS, Feinstein Y, Futscher BW. 5-Aza-2'-deoxycytidine-mediated reductions in G9A histone methyltransferase and histone H3 K9 di-methylation levels are linked to tumor suppressor gene reactivation. *Oncogene*. 2007;26:77-90.
56. Nishikawa N, Toyota M, Suzuki H, et al. Gene amplification and overexpression of PRDM14 in breast cancers. *Cancer Res*. 2007;67:9649-9657.
57. Carofino BL, Ayanga B, Justice MJ. A mouse model for inducible overexpression of Prdm14 results in rapid-onset and highly penetrant T-cell acute lymphoblastic leukemia (T-ALL). *Dis Model Mech*. 2013;6:1494-1506.
58. Hu X, Biswas A, De S. KMT2C-deficient tumors have elevated APOBEC mutagenesis and genomic instability in multiple cancers. *NAR Cancer*. 2022; 4:zcac023.
59. Cohn O, Feldman M, Weil L, Kublanovsky M, Levy D. Chromatin associated SETD3 negatively regulates VEGF expression. *Sci Rep*. 2016;6:37115.
60. Chen Z, Yan CT, Dou Y, Viboolsittiseri SS, Wang JH. The role of a newly identified SET domain-containing protein, SETD3, in oncogenesis. *Haematologica*. 2013;98:739-743.
61. Pires-Luís AS, Vieira-Coimbra M, Vieira FQ, et al. Expression of histone methyltransferases as novel biomarkers for renal cell tumor diagnosis and prognostication. *Epigenetics*. 2015;10:1033-1043.
62. Hassan N, Rutsch N, Győrffy B, Espinoza-Sánchez NA, Götte M. SETD3 acts as a prognostic marker in breast cancer patients and modulates the viability and invasion of breast cancer cells. *Sci Rep*. 2020;10:2262.
63. Cheng X, Hao Y, Shu W, et al. Cell cycle-dependent degradation of the methyltransferase SETD3 attenuates cell proliferation and liver tumorigenesis. *J Biol Chem*. 2017;292:9022-9033.
64. Kim HS, Min KW, Kim DH, Son BK, Kwon MJ, Hong SM. High WHSC1L1 expression reduces survival rates in operated breast cancer patients with decreased CD8+ T cells: machine learning approach. *J Pers Med*. 2021;11:636.
65. Igarashi H., Taniguchi H., Nosho K., et al. PRDM14 promotes malignant phenotype and correlates with poor prognosis in colorectal cancer. *Clin Transl Oncol*. 2020;22:1126-1137.
66. Cha B, Yu JH, Jin YJ, Suh YJ, Lee JW. Survival outcomes according to body mass index in hepatocellular carcinoma patient: analysis of nationwide cancer registry database. *Sci Rep*. 2020;10:8347.
67. Zhou Z, Wu B, Tang X, Yang W, Zou Q, Wang H. High SET domain bifurcated 1 (SETDB1) expression predicts poor prognosis in breast carcinoma. *Med Sci Monit*. 2020;26:e922982.
68. Fenizia C, Bottino C, Corbetta S, et al. SMYD3 promotes the epithelial-mesenchymal transition in breast cancer. *Nucleic Acids Res*. 2019;47:1278-1293.