HEPATOLOGY

HEPATOLOGY, VOL. 69, NO. 5, 2019

Integrative Analysis Defines Distinct Prognostic Subgroups of Intrahepatic Cholangiocarcinoma

Benjamin Goeppert,[1,2,*] Reka Toth,[3*] Stephan Singer,[1,4] Thomas Albrecht,[1] Daniel B. Lipka,[3] Pavlo Lutsik,[3] David Brocks,[3] Marion Baehr,[3] Oliver Muecke,[3] Yassen Assenov,[3] Lei Gu,[3,5] Volker Endris,[1] Albrecht Stenzinger,[1] Arianeb Mehrabi,[2,6] Peter Schirmacher,[1,2,7] Christoph Plass,[3,7] Dieter Weichenhan,[3,#] and Stephanie Roessler[1,2,#]

Intrahepatic cholangiocarcinoma (iCCA) is the second most common primary liver cancer. It is defined by cholangiocytic differentiation and has poor prognosis. Recently, epigenetic processes have been shown to play an important role in cholangiocarcinogenesis. We performed an integrative analysis on 52 iCCAs using both genetic and epigenetic data with a specific focus on DNA methylation components. We found recurrent isocitrate dehydrogenase 1 (*IDH1*) and *IDH2* (28%) gene mutations, recurrent arm-length copy number alterations (CNAs), and focal alterations such as deletion of 3p21 or amplification of 12q15, which affect BRCA1 Associated Protein 1, polybromo 1, and mouse double minute 2 homolog. DNA methylome analysis revealed excessive hypermethylation of iCCA, affecting primarily the bivalent genomic regions marked with both active and repressive histone modifications. Integrative clustering of genetic and epigenetic data identified four iCCA subgroups with prognostic relevance further designated as IDH, high (H), medium (M), and low (L) alteration groups. The IDH group consisted of all samples with *IDH1* or *IDH2* mutations and showed, together with the H group, a highly disrupted genome, characterized by frequent deletions of chromosome arms 3p and 6q. Both groups showed excessive hypermethylation with distinct patterns. The M group showed intermediate characteristics regarding both genetic and epigenetic marks, whereas the L group exhibited few methylation changes and mutations and a lack of CNAs. Methylation-based latent component analysis of cell-type composition identified differences among these four groups. Prognosis of the H and M groups was significantly worse than that of the L group. *Conclusion:* Using an integrative genomic and epigenomic analysis approach, we identified four major iCCA subgroups with widespread genomic and epigenomic differences and prognostic implications. Furthermore, our data suggest differences in the cell-of-origin of the iCCA subtypes. (Hepatology 2019;69:2091-2106).

Cholangiocarcinoma (CCA) is a rare malignancy of the intrahepatic or extrahepatic bile ducts with very limited treatment options and poor prognosis.[(1)] CCA is classified based on the anatomical location as intrahepatic (iCCA), perihilar (pCCA), and distal CCA (dCCA). The incidence and

Abbreviations: CCA, cholangiocarcinoma; CNA, copy number alteration; CpG, cytosine-guanine dinucleotide; dCCA, distal cholangiocarcinoma; FDR, false discovery rate; FFPE, formalin-fixed paraffin-embedded; HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma; IDH, isocitrate dehydrogenase; LMC, latent methylation component; LUMP, leukocytes unmethylation for purity; Myc, myelocytomatosis; OS, overall survival; pCCA, perihilar cholangiocarcinoma; PDAC, pancreatic adenocarcinoma; TCGA-CHOL, The Cancer Genome Atlas Cholangiocarcinoma Consortium; TSS, transcription start site.

Received September 1, 2018; accepted January 3, 2019.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.30493/suppinfo.

Supported by the German Research Foundation (CRC SFB/TR 209 Liver Cancer to P.S. [project Z01] and S.R. [project B01]); the European Union's Horizon 2020 Research and Innovation Program (667273 to P.S. and S.R); and the Helmholtz Foundation (to C.P.).

[*]These authors contributed equally to this work.

[#]These authors share last authorship.

etiologic factors of CCA vary in different geographic locations. In Southeast Asia, CCA is frequently caused by liver fluke infections, whereas the etiology is less clear in Western countries. Chronic inflammation and injury of bile duct cells are known CCA promoting conditions. Based on histology, iCCAs are subdivided into two groups: a bile duct type that resembles extrahepatic CCA with columnar cells with mucin production, and a cholangiolar type that recapitulates a genuine small-duct iCCA morphological pattern with cell-rich tubuli formed by cuboidal cells without extracellular mucin.[2] The bile duct type has a higher frequency of *KRAS* mutations, whereas the cholangiolar type shows a higher frequency of *IDH* mutations.[2] In addition, it was shown that the mutational landscape is partly subtype-specific, particularly displaying discriminating differences between iCCA versus pCCA and dCCA with, for example, isocitrate dehydrogenase (*IDH*) mutations almost exclusively detected in iCCA.[3,4] Frequent genetic alterations of epigenetic key players indicate a high impact of epigenetic processes in cholangiocarcinogenesis.[5] Deletions and mutations of genes encoding the chromatin remodeling enzymes BAP1, ARID1A, and PBRM1[6,7] and gain-of-function mutations of *IDH1* and *IDH2*[8] are the most common alterations perturbing the epigenetic landscape of iCCA.

Most epigenetic and genetic analyses were performed on mixed iCCA, pCCA, and dCCA cohorts, which may result in failure to detect variation within the iCCA subtype. In addition, the phenotypic and molecular heterogeneity of CCA in general, and iCCA in particular, is suspected to be a result of diverse cellular origins.[9,10] Potential cells-of-origin for iCCA are cholangiocytes, peribiliary glands, and hepatic stem/progenitor cells.[9,11-13] The methylation pattern of a tumor not only reflects tumorigenesis but also the methylation profile of the tumor-initiating cell types.[14] Therefore, we hypothesized that an integrative approach with special attention to cell type–composition differences may result in the identification of tumor subgroups with distinct clinical behavior. Here, we present a comprehensive integrated analysis on the genetic and epigenetic data of 52 iCCA patients of European descent with a non-liver fluke–associated etiology. We identified four distinct iCCA subgroups with prognostic relevance: an IDH group, a low (L group), a medium (M group), and a high genetic and epigenetic alteration group (H group). These four iCCA groups differ in the degree and pattern of DNA methylation, in their genomic alterations and gene mutations. Thus, these four iCCA subgroups might be clinically relevant for patient prognosis and treatment.

View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep.30493

Potential conflict of interest: Dr. Plass advises BioMed. Dr. Endris consults for Thermo Fisher. Dr. Stenzinger advises AstraZeneca, Bristol-Myers Squibb, Thermo Fisher, Novartis, and Illumina. Dr. Schirmacher is on the speakers' bureau and received grants from Novartis.

ARTICLE INFORMATION:

From the [1]Institute of Pathology, University Clinic of Heidelberg, Heidelberg, Germany; [2]Liver Cancer Center Heidelberg, Heidelberg, Germany; [3]Division of Cancer Epigenomics, German Cancer Research Center, Heidelberg, Germany; [4]Institute of Pathology, Ernst-Moritz-Arndt University, Greifswald, Germany; [5]Boston Children's Hospital, Boston, MA; [6]Department of General Visceral and Transplantation Surgery, University Hospital Heidelberg, Heidelberg, Germany; [7]German Consortium for Translational Cancer Research, Heidelberg, Germany.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Stephanie Roessler, Ph.D.
Institute of Pathology, University Hospital Heidelberg
Im Neuenheimer Feld 224
69120 Heidelberg, Germany
E-mail: Stephanie.Roessler@med.uni-heidelberg.de
Tel: +49-6221-56-35109
or

Benjamin Goeppert, M.D.
Institute of Pathology, University Hospital Heidelberg
Im Neuenheimer Feld 224
69120 Heidelberg, Germany
E-mail: Benjamin.Goeppert@med.uni-heidelberg.de
Tel: +49-6221-56-35109

Materials and Methods

STUDY POPULATION AND HISTOMORPHOLOGICAL SUBCLASSIFICATION

The study consisted of 52 iCCA patients (Table 1 and Supporting Table S1) and 12 nonneoplastic samples of cholangiocytes originating from the cystic duct of nonneoplastic cholecystectomies. All tissue samples were provided by the Tissue Bank of the National Center for Tumor Diseases (NCT, Heidelberg, Germany) in accordance with the regulations of the NCT Tissue Bank. Informed consent in writing was obtained from each patient and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki as reflected by *a priori* approval of the ethics committee of the University of Heidelberg (S-206/2005, S-207/2015, and S-539/2012). Each iCCA tumor sample was histologically confirmed by at least two experienced pathologists (B.G., S.S., and P.S.). In addition, a histomorphological subtyping into bile duct type or cholangiolar type according to Liau et al.[2] was performed (Supporting Fig. S1).

GENOMIC DNA ISOLATION

Genomic DNA was isolated from fresh frozen tissue using the QIAamp DNA micro kit (Qiagen, Hilden, Germany) for whole exome sequencing according to the manufacturer's instructions (Supporting Table S1). From formalin-fixed paraffin-embedded (FFPE) samples, genomic DNA was extracted using the AllPrep DNA/RNA FFPE Kit (Qiagen), as recommended by the manufacturer with the following modifications: After addition of xylene, samples were incubated at 56°C for 2 minutes followed by two ethanol washes. The first proteinase K digestion was performed with 20 μL at 56°C for 30 minutes. The DNA was eluted twice with 30 μL of H_2O.

EXOME SEQUENCING

Whole exome sequencing libraries were prepared from DNA isolated from fresh frozen tissue and from microdissected surrounding normal tissue to distinguish somatic from germline mutations. Sequencing of the libraries was done at the German Cancer Research Center (DKFZ) Genomics and Proteomics Core Facility using the Agilent SureSelectXT Human all Exon V4 kit and a HiSeq2000 instrument (Illumina, San Diego, CA) using the 100-bp paired-end mode.

PANEL SEQUENCING

To analyze the samples for genetic variations, a custom gene panel for massive parallel sequencing was used. This panel consisted of 285 amplicons covering 165 exons within 40 genes frequently mutated in biliary tract cancers (Supporting Table S2). For detailed information see the Supporting Information.

DNA METHYLATION ANALYSIS

DNA methylation profiles were determined by the Genomics and Proteomics Core Facility (DKFZ Heidelberg) using the Infinium HumanMethylation450 BeadChip from FFPE tissue–derived genomic DNA, as described previously.[15]

DATA ANALYSIS

Available data and the process of data analysis are shown in Fig. 1A.

Methylation data were processed using the R platform for statistical computing (see Supporting Information). Copy number alterations were assessed based on the signal intensities measured in the methylation array. Differential methylation analyses (between tumor and normal or between the identified subgroups) were conducted using linear models. Adjustments were made for patient age, gender, and the source of tissue in tumor-normal comparison and for age in the group-wise comparison among subgroups. A difference was considered to be significant if the false discovery rate (FDR) corrected *P* value (q value) was less than 0.05. Tumor purity was estimated using the LUMP (leukocytes unmethylation for purity) method.[16] To assess cell type heterogeneity and to trace the cell-of-origin in tumor samples, we used MeDeCom[17] to decompose methylation data into latent methylation components (LMCs) and to assess their proportions in each sample. The integrative clustering was confined to tumor samples and performed using the copy number, DNA methylation, and LMC data with the iClusterPlus R package.[18]

To identify possible candidate genes located in the respective chromosomal arms or focal

TABLE 1. Clinical Characteristics of the iCCA Study Population (n = 52)

				P Value*
Age	Mean (years)	59.91		0.51
Tumor size	Mean (mm)	7.19		0.034
Number of patients		n	Percentage	
Gender	Male	25	48.1	Ref.
	Female	27	51.9	0.31
Histological type	Cholangiolar type	40	76.9	Ref.
	Bile duct type	12	23.1	0.837
T stage	pT1	16	30.8	Ref.
	pT2	28	53.8	0.114
	pT3 (n = 7)/pT4 (n = 1)	8	15.4	0.661
Histologic grade	G1	4	7.7	Ref.
	G2	36	69.2	0.55
	G3	12	23.1	0.399
N	N0	17	32.7	Ref.
	N1	11	21.2	0.146
	N.A.	24	46.2	N.A.
M	M0	52	100.0	N.A.
L	L0	31	59.6	Ref.
	L1	16	30.8	0.343
	N.A.	5	9.6	N.A.
V	V0	33	63.5	Ref.
	V1	14	26.9	0.809
	N.A.	5	9.6	N.A.
R	R0	28	53.8	Ref.
	R1	16	30.8	0.016
	R2	3	5.8	0.964
	N.A.	5	9.6	N.A.
UICC	UICC 1	4	7.7	Ref.
	UICC 2	10	19.2	0.785
	UICC 3	15	28.8	0.436
	N.A.	23	44.2	N.A.
Hepatobiliary disease	Hepatitis B virus	3	5.8	0.741
	Hepatitis C virus	3	5.8	0.757
	Hepatic steatosis	4	7.7	0.999
	NAFLD	4	7.7	0.999
	Pre-existing inflammatory biliary tract disease†	8	15.4	0.421
	Chronic pancreatitis	1	1.9	N.A
	M. Wilson	1	1.9	N.A
	Thorotrast	1	1.9	N.A
	None detected	27	51.9	Ref.

*Cox regression *P* value.
†Cholecystitis and/or choledocholithiasis.
Abbreviations: N.A., not available; NAFLD, nonalcoholic fatty liver disease; Ref., reference; and UICC, International Union Against Cancer.

regions with significant copy number alterations (CNAs), we used the gene expression and copy number data of The Cancer Genome Atlas Cholangiocarcinoma Consortium (TCGA-CHOL) (https://portal.gdc.cancer.gov/projects/TCGA-CHOL). Pearson correlation was performed in chromosomal arm level analysis and Spearman correlation in the focal alterations. FDR values less than 0.05 were considered

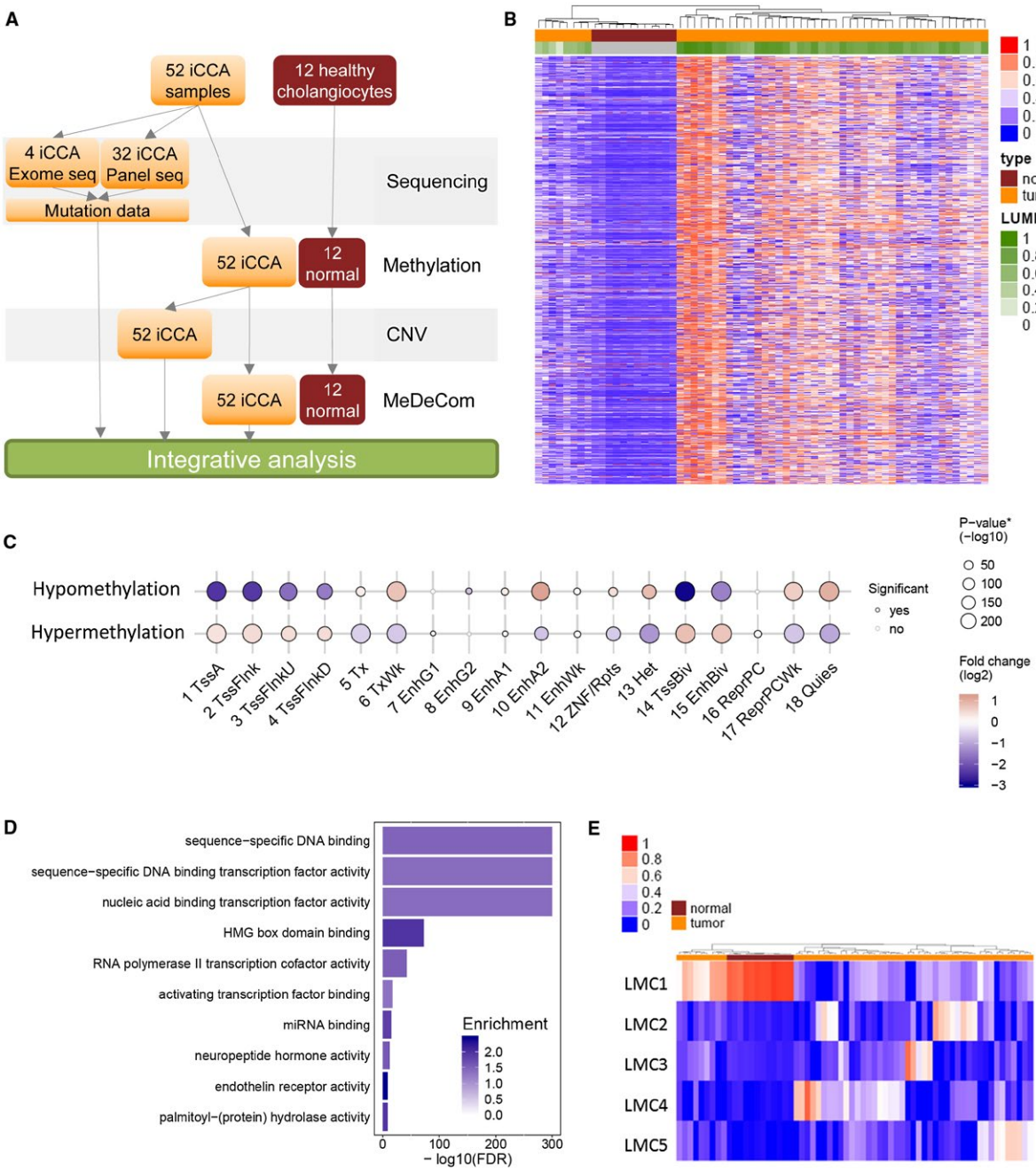

FIG. 1. DNA methylome landscape of iCCA. (A) Overview of the study. (B) Clustering of the methylation values of tumor and normal samples (columns), ranging from 0 to 1, of the 10,000 most variable CpG sites (rows). Tumor purity is represented by LUMP values ranging from 0 (low purity) to 1 (high purity). (C) Enrichment analysis of the hypomethylated and hypermethylated CpG sites compared with normal controls, using the 18-state ChromHMM model of Roadmap Epigenomics. The log2 fold change is represented by the color of the dots, whereas the size reflects the –log10(*P* value). The border of the dots shows whether the result is significant. (D) Enrichment analysis using the GO Molecular functions database for the hypermethylated CpG sites (|beta difference| > 0.2). FDR q values < 10^{-300} were set to 10^{-300}. (E) Clustering results of the MeDeCom analysis for tumor and normal samples (columns) with respect to the five LMCs (LMC1 to LMC5) (rows). Color represents the contribution of the given LMC to the respective sample. Abbreviations: 1 TssA, active TSS; 2 TssFlnk, flanking TSS; 3 TssFlnkU, flanking TSS upstream; 4 TssFlnkD, flanking TSS downstream; 5 Tx, strong transcription; 6 TxWk, weak transcription; 7 EnhG1, genic enhancer1; 8 EnhG2, genic enhancer 2; 9 EnhA1, active enhancer 1; 10 EnhA2, active enhancer 2; 11 EnhWk, weak enhancer; 12 ZNF/Rpts, ZNF genes and repeats; 13 Het, heterochromatin; 14 TssBiv, bivalent/poised TSS; 15 EnhBiv, bivalent enhancer; 16 ReprPC, repressed PolyComb; 17 ReprPCWk, weak repressed PolyComb; and 18 Quies, quiescent/low.

to be significant. Survival analysis was performed using Cox proportional hazards regression model.

Detailed information on the methods used is described in the Supporting Information.

Results

iCCA IS CHARACTERIZED BY FREQUENT TP53 AND IDH MUTATIONS AND LARGE CHROMOSOMAL ABERRATIONS

To genetically characterize our 52 iCCA samples, we performed whole exome or targeted panel sequencing of 40 genes commonly mutated in CCA in a subset of 36 iCCA tumor samples (Supporting Table S1), which revealed 24 mutations (Table 2 and Supporting Table S2). Recurrent missense mutations in *IDH1* were observed in 6 of 36 (17%) patients (Table 2). All *IDH1* mutations were located in a mutation hotspot and altered the same codon: p.R132C, p.R132G, or p.R132L.[19] *IDH2* encoding the mitochondrial isozyme of *IDH1* displayed three different missense mutations in the same mutation hotspot, leading to p.R172W, p.R172M, or p.R172S.[19] *IDH1* and *IDH2* mutations lead to the production of the oncometabolite 2-hydroxyglutarate, which was shown to inhibit histone and DNA demethylation.[20] *TP53* was the second most affected gene with two nonsense and two missense mutations in 4 patients. These mutations all impaired the DNA binding domain of the TP53 protein, as has been observed frequently in iCCA.[21] The mutations in *IDH1*, *IDH2,* and *TP53* were mutually exclusive with each other. In addition, two missense mutations were found in *TGFBR2* and one mutation was found in *ARID1A*, *SF3B1*, *ROBO1*, *ROBO2*, *FBXW7*, *BRAF*, *CDKN2A*, *FGFR2*, *KDM5A*, *KRAS*, *SMARCA4,* and *GNAS* each (Table 2).

We used HumanMethylation450 BeadChip data to identify recurrent DNA methylation and genomic

TABLE 2. Mutations in the Study Population (n = 36)

Gene	Amino Acid Change	Nucleotide Change	Type	Location
ARID1A	p.M890V	NM_006015:c.2668A>G	Missense	1p36.11
BRAF	p.V600E	NM_004333.4:c.1799T>A	Missense	7q34
CDKN2A	p.A97V	NM_058195:c.290C>T	Missense	9p21.3
FBXW7	p.R505H	NM_033632:c.1514G>A	Missense	4q31.3
FGFR2	p.C382R	NM_000141:c.1144T>C	Missense	10q26.13
GNAS	p.R201C	NM_000516:c.601C>T	Missense	20q13.32
IDH1	p.R132C	NM_001282386:c.394C>T	Missense	2q34
	p.R132G	NM_001282386:c.394C>G	Missense	
	p.R132L	NM_001282386:c.395G>T	Missense	
IDH2	p.R172W	NM_002168:c.514A>T	Missense	15q26.1
	p.R172S	NM_001290114:c.126G>T	Missense	
	p.R172M	NM_001290114:c.125G>T	Missense	
KDM5A	p.R604C	NM_001042603:c.1810C>T	Missense	12p13.33
KRAS	p.G12delinsGAG	NM_004985:c.35_36insAGCTGG	Insertion	12p12.1
ROBO1	p.V1454L	NM_002941:c.4360G>T	Missense	3p12.3
ROBO2	p.G866D	NM_002942:c.2597G>A	Missense	3p12.3
SF3B1	p.A702_S705delRTI	NM_012433:c.2104_2112del	Deletion	2q33.1
SMARCA4	p.E882D	NM_001128845:c.2646A>C	Missense	19p13.2
TGFBR2	p.R528C	NM_004333:c.1799T>A	Missense	3p24.1
	p.V412M	NM_001024847:c.1234G>A	Missense	
TP53	p.R248Q	NM_000546:c.743G>A	Missense	17p13.1
	p.E171*	NM_000546:c.511G>T	Nonsense	
	p.Y163*	NM_000546:c.489C>G	Nonsense	
	p.R290C	NM_000546:c.868C>T	Missense	

Note: All changes occurred once (2.8%) except for IDH1 p.R132C, which occurred 4 times (11.1%).

TABLE 3. Recurrent Arm-Length CNAs and Affected Genes (n = 52)

Deletion				Deleted Genes With Mutation in Study*
Arm	Genes (n)	Frequency	FDR q Value	
1p	2,121	0.18	0.0002	*ARID1A*
3p	1,062	0.38	0	*ROBO1, ROBO2, TGFBR2*
6q	839	0.48	0	—
8p	580	0.16	0.0002	—
9p	422	0.25	$9.64*10^{-13}$	*CDKN2A*
9q	1,113	0.25	$2.56*10^{-11}$	—
12q	1,447	0.12	0.0493	—
13q	654	0.21	$1.44*10^{-08}$	—
14q	1,341	0.23	$4.00*10^{-09}$	—
16q	702	0.12	0.0222	—
17p	683	0.12	0.0222	*TP53*
Amplification				
Arm	Genes (n)	Frequency	FDR q Value	
1q	1,955	0.21	$5.48*10^{-06}$	

*Mutated genes are listed in Table 2.

CNAs. We found recurrent deletions and amplifications of entire chromosome arms (Table 3) and of focal genomic regions (Table 4 and Supporting Fig. S2). Chromosome arms 1p, 3p, 6q, 8p, 9p, 9q, 12q, 13q, 14q, 16q, and 17p were affected by large deletions consisting of between 422 genes on 9p and up to 2,121 genes on 1p (Table 3). Some genes were found in both lists of deleted and mutated genes (namely, *ARID1A* on 1p; *ROBO1*, *ROBO2*, and *TGFBR2* on 3p; *CDKN2A* on 9p; and *TP53* on 17p). Chromosome arm 1q showed long-range amplifications covering up to 1,955 genes (Table 3).

We identified genes whose expression might be affected by focal deletions or amplifications in our study population by correlation analyses of publicly available CNA and gene expression data from TCGA-CHOL (n = 51) and considered information on known driver genes[22] (Table 4). For example, the focally deleted genes *BAP1* and *PBRM1* on 3p21.1 were also commonly affected by an arm-length deletion in our study population (38%)

TABLE 4. Recurrent Focal CNAs and Possible Candidate Genes in the Region

Deleted Regions				
Chromosome	Cytoband	Coordinates	q Value	Candidate Genes
1	1p13.2	100598444-116907532	0.0873	***SLC30A7, CTTNBP2NL, PHTF1, KCNC4, RHOC, SARS, PSMA5, NRAS, MOV10, TMEM167B, DDX20, GNAI3, CSDE1, CLCC1, BCAS2***
1	1p36.23	8014651-8874983	0.0045	
3	3p21.1	52318567- 52812032	0.0089	***BAP1***,*† ***SPCS1***, *PBRM1*†
4	4q35.2	171011578-191154276	0.0289	***CASP3, DCTD***, *FAT1*,† ***FRG1***,† ***ING2, FBXO8, UFSP2, LRP2BP, SNX25***
5	5q14.3	76116089-95991890	0.0029	*RASA1*,† ***TBCA, XRCC4***
7	7q21.11	79881211-85911765	0.0002	
11	11q22.3	104769088-105947697	0.0002	*CASP1*†
17	17q21.2	38808305-39465712	0.0289	
Amplified Regions				
Chromosome	Cytoband	Coordinates	q Value	Candidate Genes
1	1q21.3	150554876-152815842	0.0106	***ARNT, ENSA, PI4KB, VPS72, PIP5K1A, SETDB1, POGZ, GOLPH3L, CDC42SE1, ZNF687, PRUNE, GABPB2, LYSMD1***
2	2q31.1	176936398-177001961	0.0023	
4	4p15.32	16791107-17785836	1.57×10^{-31}	
12	12q15	66220166- 71083782	0.0023	***MDM2***,*† ***NUP107***,† ***CNOT2***

Note: Bold indicates a significant correlation between gene expression and CNA in TCGA-CHOL.
*Identified as possible driver gene in Ref. [50].
†Identified as driver gene in Ref. [22].

(Table 3) and by known inactivating mutations in CCA and renal cell carcinoma.[6,23] As an example of a candidate gene residing in the amplified 12q15 region, the *MDM2* oncogene showed a significant positive correlation (P = 0.016, r = 0.40) with its expression in the TCGA-CHOL cohort.

iCCA SHOWS EXTENSIVE HYPERMETHYLATION

Clustering of the 10,000 most variably methylated CpGs (cytosine-guanine dinucleotides) revealed distinct clusters for the normal samples and the iCCAs (Fig. 1B). One cluster of iCCA samples displayed DNA methylation levels similar to the nonneoplastic cholangiocyte samples, whereas the remaining iCCA cases displayed different patterns (Fig. 1B). The similarity in DNA methylation profiles between the group of nonneoplastic cholangiocyte samples and the normal-like iCCA group might in part be ascribed to a higher immune cell infiltration in both groups, as suggested by LUMP analysis (Fig. 1B). Primarily CpG hypermethylation was seen in the iCCA cohort with approximately 37,600 hypermethylated CpGs exhibiting increased methylation by more than 20%. In contrast, only 2,217 CpG sites were hypomethylated by more than 20%. We overlapped the hyperthylated and hypomethylated CpG sites with publicly available 18-state chromatin segmentation data of H1 ES cells and found that hypermethylated regions were enriched at transcription start sites (TSSs) and enhancers but depleted in heterochromatic regions and gene bodies. Hypomethylated regions were enriched in gene bodies, heterochromatic and quiescent regions, but depleted at TSSs (Fig. 1C).

To further characterize the hypermethylated regions, we used the GREAT tool and found genes involved in transcription factor and HMG (3-hydroxy-3-methylglutaryl) box domain binding to be enriched (Fig. 1D). We also performed MeDeCom analysis to dissect methylation patterns into LMCs to incorporate information on possible cell-type composition differences (Fig. 1E). Five major LMCs were identifiable, of which LMC1 appeared to be characteristic for normal bile duct samples (Supporting Fig. S3). LMC1 and LUMP values were strongly negatively correlated (r = -0.799, P < 0.01; Supporting Fig. S4).

FOUR iCCA SUBGROUPS ARE IDENTIFIED WITH INTEGRATIVE CLUSTERING

Using iCluster[18] on the tumor samples, we performed an integrative clustering combining CNA, methylation, and MeDeCom analysis data and identified four iCCA groups, named the IDH group, L (low alteration) group, M (medium alteration) group, and H (high alteration) group. These four groups are characterized by IDH mutation, the degree of acquired genetic and epigenetic alterations, and by LMC profiles (Fig. 2A, Supporting Fig. S3). The enrichment analyses of the LMC-specific hypermethylation show the involvement of developmental and differentiation processes for LMC3 and LMC4, whereas in LMC2, primarily metabolic processes are affected (Supporting Fig. S5). The L group partially overlaps with the normal-like samples (Fig. 1A) and is characterized by the highest proportion of LMC1 (48% versus 23%, 17% and 17%, respectively) and LMC5 (28% versus 22%, 2%, and 8%), low mutation number (4 mutations in 4 of 11 samples sequenced), low frequency of CNA (approximately 40% for 6q deletion and less for the other alterations), and generally low methylation levels. The high ratio of LMC1 may be partially explained by high infiltration of immune cells, as shown by the correlation of LMC1 with LUMP (Supporting Fig. S4).

Most of the samples in the IDH group are characterized by mutations in *IDH1* or *IDH2* (9 out of 10 sequenced), a unique pattern of relatively high methylation values, a high level of CNAs (up to 80% for 3p deletions), and high values of LMC2 (Fig. 2A and Supporting Fig. S6). The high level of methylation in the IDH group is consistent with the neo-enzymatic function of affected IDH proteins, resulting in the inhibition of TET-enzymes and ultimately in inhibition of active DNA demethylation.[20] The iCluster H group has high methylation levels, frequent CNAs (up to 70% for some chromosome arms), and high LMC4 values. Finally, the M group showed a mixed LMC composition, low frequency of deletions and amplifications, and a group-specific gain of the chromosome arm 8q harboring the *Myc* (myelocytomatosis) oncogene. Thus, iCluster analysis revealed four distinct iCCA subgroups with specific methylation patterns, LMCs, CNAs, and mutations.

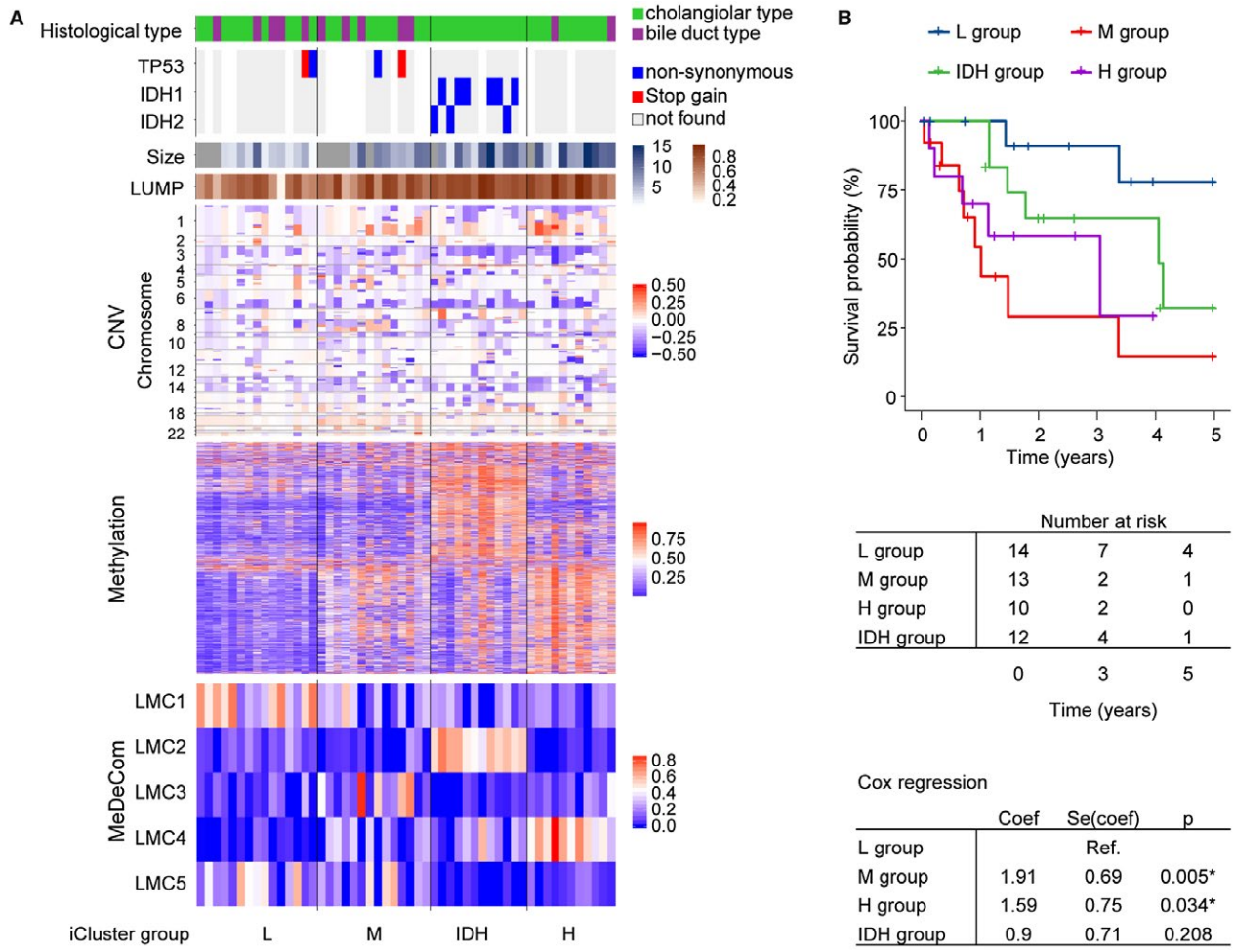

FIG. 2. Integrative cluster analysis identifying four subgroups in iCCA. (A) Integrative clustering using CNA, DNA methylation, and MeDeCom data splits the iCCA samples into four groups. In addition, the histological types of the iCCA samples, the mutation status of the three most commonly mutated genes, tumor size, and tumor purity according to LUMP are indicated. (B) Kaplan-Meier survival plots of the four groups. The tables indicate the number of patients at risk at given time points and the results of the Cox proportional hazard regression model using the L group as reference.

iCCA SUBGROUPS SHOW DIFFERENT MOLECULAR AND CLINICAL CHARACTERISTICS

The four iCCA subgroups did not show any significant associations with basic clinical characteristics, including age, gender, tumor staging, and underlying hepatobiliary disease (Supporting Table S3). In the IDH group, the only histological type was the cholangiolar type, whereas the other three groups included 18% to 36% samples with bile duct type (Fisher's exact test $P = 0.047$; comparison of IDH versus L, M, and H groups combined). In addition, we found that the molecular subgroups showed significantly different outcomes (Fig. 2B). Overall survival (OS) of the M group was the worst ($P = 0.005$, OS M group versus L group), followed by the H group ($P = 0.034$, OS H group versus L group). The L and the IDH groups showed the most favorable prognosis and did not significantly differ from each other (Fig. 2B).

Further analyses of known cancer driver genes showed that *RPL22*, *ROBO1*, *ROBO2*, *TGFBR1*, and *TGFBR2* were most frequently altered either by deletion or promoter hypermethylation in all groups (Fig. 3A).[22] The IDH and H groups frequently showed deletions on chromosome arms 3p and 6q, the former harboring the tumor-suppressor genes *BAP1*, *PBRM1*, *TGFBR2*, *ROBO1* and *ROBO2*,

the latter *ZNF292* and *EEF1A1*. Both the M and H groups shared amplification of chromosome 1q, whereas amplification of 8q, which harbors the *Myc* oncogene, predominated in the M group (Fig. 3B and Supporting Table S4). We narrowed down the amplification to a focal region in 1q21.3 containing a number of candidate genes (Table 4).

To better define candidate driver genes affected by genomic alteration in these regions, we correlated copy number gain with expression data obtained from TCGA-CHOL. Among the genes amplified on 1q, a strong correlation was observed for *PI4KB* (r = 0.71, FDR q = $3.66*10^{-5}$) and *PIP5K1A* (r = 0.69, FDR q = $5.91*10^{-5}$), which, together with *AKT3* (r = 0.28, FDR q = 0.19) (Supporting Fig. S7A), belong to the cancer-relevant phosphoinositide signaling pathway.[24] Another candidate was *YY1AP1* (r = 0.65, FDR q = $2.36*10^{-4}$), known as an oncogenic driver in EpCAM(+) AFP(+) hepatocellular carcinoma,[25] which shows features of hepatic stem/progenitor cells.[26] By high correlation (r = 0.82, FDR q=$1.09*10^{-6}$), *SETDB1*, encoding a histone lysine methyltransferase and known to be involved in breast cancer,[27] proved a candidate driver gene on the focally amplified 1q21.3 (Supporting Fig. S7B). *Myc* presented as a candidate on 8q, yet this role was not underscored by correlation with expression. However, potential candidates on 8q are *CHRAC1* (r = 0.61, FDR q = 0.010), *RAD21* (r = 0.51, FDR q = 0.023), and *TRAPPC9* (r = 0.50, FDR q = 0.030) (Supporting Fig. S7C), all known to be involved in breast cancer.[28,29] These results suggest the importance of CNA-driven gene-expression changes.

As the epigenomic landscape contributes to the properties and specification of tumor subgroups,[30,31] we searched for characteristics of the four iCCA groups regarding their methylation patterns. Globally, all four subgroups show a significant increase in their methylation levels, with significant differences between the L and the M versus the H and IDH mutant groups (Supporting Fig. S6). To decipher the group-specific changes, we performed differential methylation analysis comparing each group to all other tumor samples (Supporting Fig. S8). The IDH and H groups exhibited broad hypermethylation with highly group-specific patterns. This finding is supported by the number of unique and overlapping differentially methylated CpG sites using the L group as reference (Fig. 4A). In contrast, the M group showed only few CpGs with group-specific hypermethylation (Fig. 4A and Supporting Fig. S8).

Enrichment analysis of the specific hypermethylated sites in the IDH group showed increased appearance in bivalent TSSs and enhancers, but the enrichment is even more pronounced in the flanking regions of the TSSs (Fig. 4B). The H group exhibited an even stronger enrichment at the bivalent enhancers and TSSs, whereas the TSS flanking regions were depleted. Pathway analysis revealed that most frequently affected genes encode mostly homeobox proteins and transcription factors (Fig. 4C). However, the IDH group-specific alterations did not show enrichment of cancer-related pathways (Supporting Fig. S9) but were aggregated in CTCF and RAD21 binding sites and depleted in the binding sites of other transcription factors (Fig. 4D).

To compare the resemblance and the potential cell-of-origin of the four iCCA groups with hepatocellular carcinoma (HCC) and pancreatic adenocarcinoma (PDAC), we applied our MeDeCom model to TCGA DNA methylation data (see Supporting Information). We included 100 randomly selected HCCs and added all IDH-mutant HCCs that were not randomly selected (TCGA-LIHC, n = 106, including 9 IDH-mutant HCCs), 100 randomly selected PDACs and added all IDH-mutant PDACs (TCGA-PAAD, n = 101, including 1 IDH mutant), and all available CCAs (TCGA-CHOL; n = 45; Supporting Table S5). Clustering analysis of the tumor-specific LMC2-5 profiles showed that LMC2 co-occurred with *IDH1* and *IDH2* gain-of-function mutation of HCC and CCA but did not include any PDAC cases (Fig. 4E and Supporting Fig. S10). Furthermore, the TCGA-CHOL samples exhibited similar subgroups as our iCCA study population. A subgroup of PDAC cases formed a distinct cluster with primarily M group iCCA samples with high LMC3 ratio, whereas clusters with LMC4 and LMC5 included only HCC and CCA cases. H group iCCA shared similarities with LMC4-high HCC and CCA, but this cluster did not include any PDAC cases. Consistently, the bile duct–type iCCA clustered together with M group iCCA and PDAC, whereas the cholangiolar-type iCCA showed similar patterns to a subgroup of HCC high LMC4 or LMC5. In addition, iCCA cases with amplification of *YY1AP1*, a potential marker of HCC with hepatic stem/progenitor features, did not cluster together

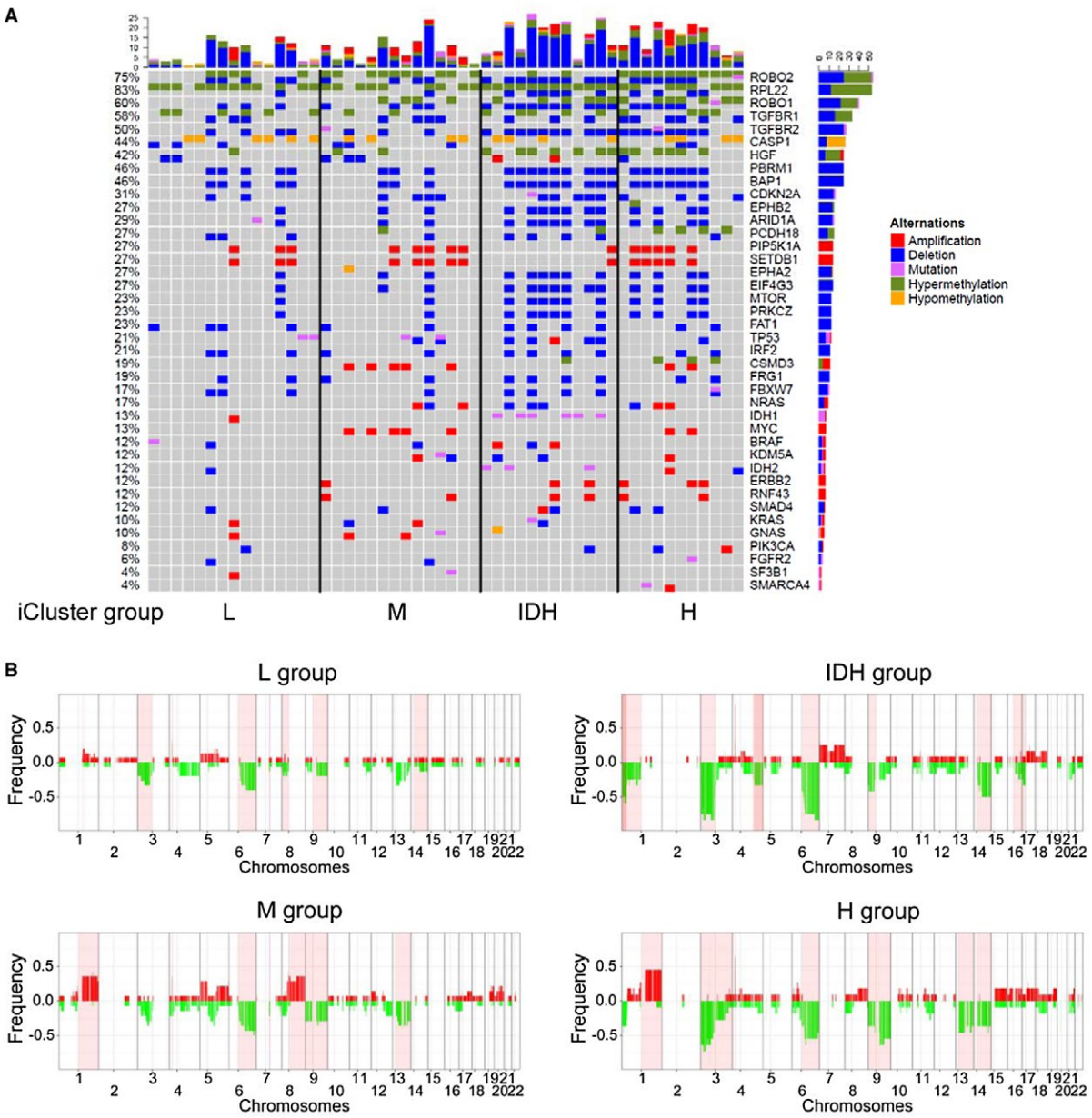

FIG. 3. Genetic profiles of the iCCA subgroups. (A) CNA, mutation, and DNA methylation differences of known tumor driver genes in the four groups. All driver genes located in a significantly amplified or deleted region, carrying a mutation in any of the samples or differentially methylated, are plotted. All samples were plotted, including those without sequencing data; thus, the number of mutations may be higher than indicated. (B) Frequency of CNAs in the different subgroups. Amplifications are indicated in red, deletions in green. Bar extent indicates the frequency of CNA. The light pink shading indicates significant CNAs based on GISTIC2.0 analysis.

with the IDH-mutant cases that are LMC2-high (Fig. 4E). *YY1AP1*-amplified iCCA showed high levels of LMC4 or LMC5 and clustered primarily with a subgroup of IDH wild-type HCC and CCA but not with PDAC. The remaining HCC and PDAC samples showed very low ratios of LMC2 to LMC5. Taken together, using three TCGA data sets consisting of HCC, CCA and PDAC cases, we found that IDH gain-of-function mutations resulted in distinct LMC2-high profiles, M group iCCA were enriched for bile duct–type iCCA and exhibited similarities with LMC3-high PDAC, whereas H group iCCA had higher LMC4 and the LMC4-high cluster was enriched for cholangiolar-type iCCA.

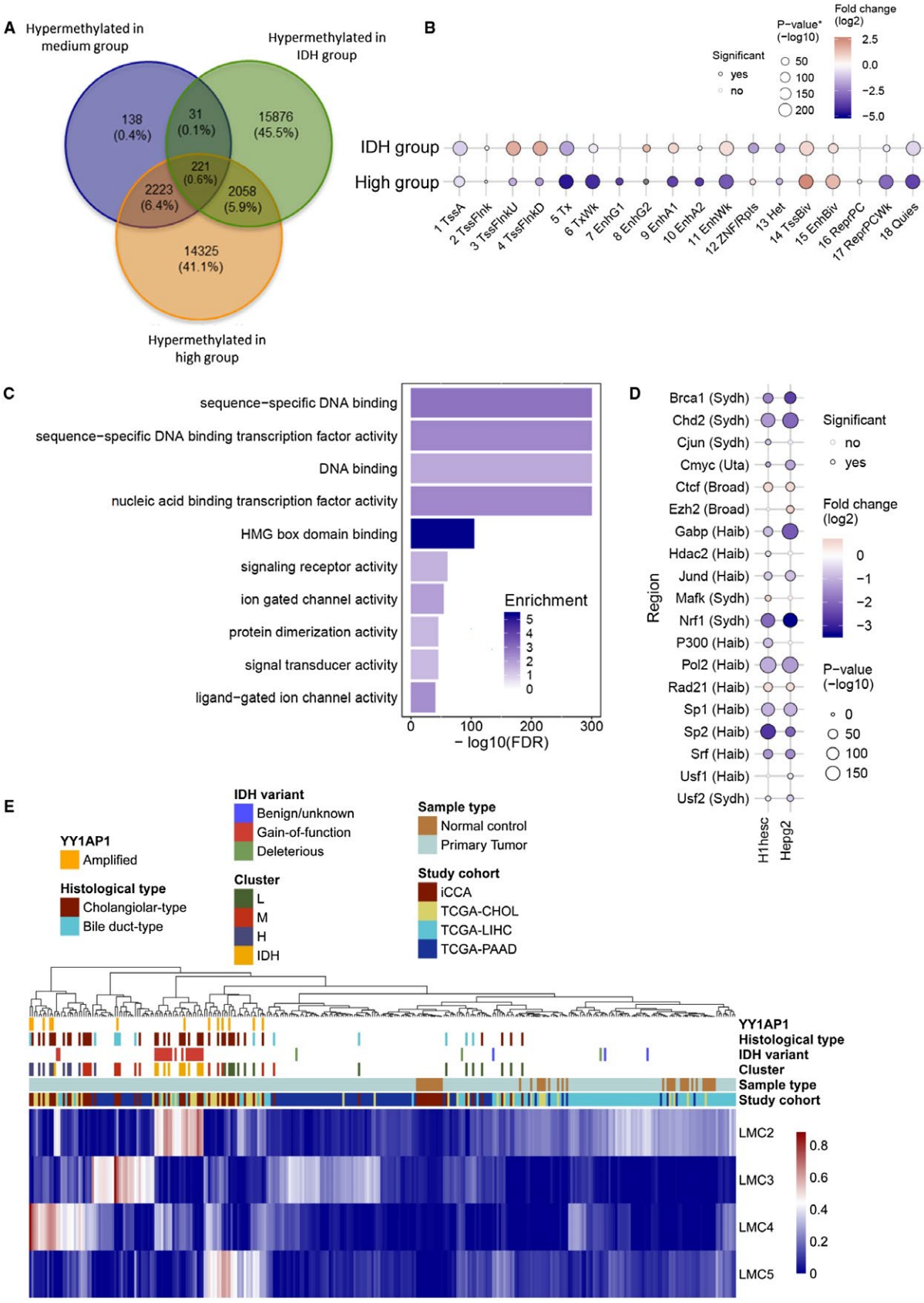
A
Hypermethylated in medium group
Hypermethylated in IDH group
Hypermethylated in high group
138 (0.4%)
31 (0.1%)
15876 (45.5%)
221 (0.6%)
2223 (6.4%)
2058 (5.9%)
14325 (41.1%)
B
P−value* (−log10)
Significant
yes
no
50
100
150
200
Fold change (log2)
2.5
0.0
−2.5
−5.0
IDH group
High group
1 TssA
2 TssFlnk
3 TssFlnkU
4 TssFlnkD
5 Tx
6 TxWk
7 EnhG1
8 EnhG2
9 EnhA1
10 EnhA2
11 EnhWk
12 ZNF/Rpts
13 Het
14 TssBiv
15 EnhBiv
16 ReprPC
17 ReprPCWk
18 Quies
C
sequence−specific DNA binding
sequence−specific DNA binding transcription factor activity
DNA binding
nucleic acid binding transcription factor activity
HMG box domain binding
signaling receptor activity
ion gated channel activity
protein dimerization activity
signal transducer activity
ligand−gated ion channel activity
Enrichment
5
4
3
2
1
0
0
100
200
300
− log10(FDR)
D
Brca1 (Sydh)
Chd2 (Sydh)
Cjun (Sydh)
Cmyc (Uta)
Ctcf (Broad)
Ezh2 (Broad)
Gabp (Haib)
Hdac2 (Haib)
Jund (Haib)
Mafk (Sydh)
Nrf1 (Sydh)
P300 (Haib)
Pol2 (Haib)
Rad21 (Haib)
Sp1 (Haib)
Sp2 (Haib)
Srf (Haib)
Usf1 (Haib)
Usf2 (Sydh)
Region
H1hesc
Hepg2
Significant
no
yes
Fold change (log2)
0
−1
−2
−3
P−value (−log10)
0
50
100
150
E
YY1AP1
Amplified
Histological type
Cholangiolar-type
Bile duct-type
IDH variant
Benign/unknown
Gain-of-function
Deleterious
Cluster
L
M
H
IDH
Sample type
Normal control
Primary Tumor
Study cohort
iCCA
TCGA-CHOL
TCGA-LIHC
TCGA-PAAD
YY1AP1
Histological type
IDH variant
Cluster
Sample type
Study cohort
LMC2
LMC3
LMC4
LMC5
0.8
0.6
0.4
0.2
0

FIG. 4. Methylation differences between the iCCA groups in relation to the L group. (A) Venn diagram shows the numbers of unique and overlapping hypermethylated sites for the IDH, H, and M groups. The reference group was the normal-like L group. (B) Enrichment analysis of the group-specific hypermethylated promoters, using the 18-state ChromHMM model of Roadmap Epigenomics for the IDH group (top) and the H group (bottom). The reference category was the combination of the other tumor groups. The log2 fold change is represented by the color of the dots, whereas the size reflects to the –log10(*P* value). The border of the dots shows whether the result is significant. (C) Pathway analysis of the group-specific hypermethylation in the H group. The analysis used the GREAT tool and the GO Molecular Function database. The blue shades reflect the strength of the enrichment as fold change, whereas the bars show the –log10(FDR) values. (D) Enrichment analysis of the IDH group–specific hypermethylation using transcription factor binding sites from ENCODE. The log2 fold change is represented by the color of the dots, whereas the size reflects the –log10(*P* value). The border of the dots shows whether the result is significant. (E) Heatmap applying LMCs 2-5 on DNA methylation data including CCA, HCC, and pancreatic adenocarcinoma (PDAC) cases from TCGA (TCGA-CHOL, TCGA-LIHC, TCGA-PAAD). The study cohort iCCA indicates the cohort used in this study, and the variable cluster reflects on the result of the integrative clustering, identifying the L, M, H, and IDH groups. The IDH "gain-of-function" category includes IDH mutations affecting IDH1 p.R132 and IDH2 p.R172. The IDH variant categories "deleterious" and "benign/unknown" depict other IDH1 and IDH2 variants by their predicted effect, as determined by PolyPhen. Abbreviations: 1 TssA, active TSS; 2 TssFlnk, flanking TSS; 3 TssFlnkU, flanking TSS upstream; 4 TssFlnkD, flanking TSS downstream; 5 Tx, strong transcription; 6 TxWk, weak transcription; 7 EnhG1, genic enhancer 1; 8 EnhG2, genic enhancer 2; 9 EnhA1, active enhancer 1; 10 EnhA2, active enhancer 2; 11 EnhWk, weak enhancer; 12 ZNF/Rpts, ZNF genes and repeats; 13 Het, heterochromatin; 14 TssBiv, bivalent/poised TSS; 15 EnhBiv, bivalent enhancer; 16 ReprPC, repressed PolyComb; 17 ReprPCWk, weak repressed PolyComb; and 18 Quies, quiescent/low.

Discussion

As a reflection of the growing urgency to better understand the mechanisms underlying iCCA, earlier *omic* studies analyzed the molecular landscape of CCA. These genome-wide studies focused primarily on the mutational and genomic CNA landscape of CCA.[6,7,21,32] Recent studies in cholangiocarcinoma used integrative analysis of genomic, transcriptomic, and epigenomic data in mixed cohorts of iCCA, pCCA, and dCCA.[33,34] Our present study included a clinicopathologically well-characterized European cohort of iCCA patients only (n = 52) with different underlying hepatobiliary disease but without liver fluke association (Supporting Table S3). Consistent with previous studies, pre-existing cholecystitis and/or choledocholithiasis was associated with bile duct–type iCCA (Supporting Table S6). Here, we integrated mutational data from 40 candidate genes, CNAs, whole genome DNA methylation changes, and publicly available TCGA-CHOL transcriptomes. We also harnessed the methylome data for tumor purity assessment and the identification of latent methylation components as proxies for different cell types, and thereby identified four distinct iCCA groups with prognostic significance. Importantly, the four iCCA groups resulting from the iCluster analysis are largely different in the proportions of LMCs, which recover cell type–specific hidden methylation patterns.[17]

Our sequencing analysis confirmed the recurrence of *IDH1* and *IDH2* gain-of-function mutations (28%) and *TP53* loss-of-function mutations (9%) in carcinogenesis of iCCA. Moreover, additional mutations in epigenetic genes other than *IDH1* and *IDH2* included *ARID1A* and *SMARCA*, both involved in chromatin remodeling, and *KDM5A*, encoding a histone demethylase, highlighting a strong epigenetic component in iCCA carcinogenesis. Recurrent long-range genomic alterations additionally underscored the importance of known iCCA candidate genes such as *BAP1* and *PBRM1* on deleted 3p21.1[6] or *MDM2* on amplified 12q15. The latter gene may initiate tumor onset and progression by negatively regulating tumor suppressor p53[35] or by a p53-independent mechanism.[36] In addition, *ARID1A*, *ROBO1*, *ROBO2*, *TGFBR2*, *CDKN2A*, and *TP53* were found to be affected by recurrent deletions or amplifications.

Methylome analysis showed excessive hypermethylation of iCCA. Hypermethylation disproportionally affected bivalent regulatory regions in this iCCA cohort and was enriched for differentiation and developmental processes. Hypermethylation of the regions showing bivalent characteristics in embryonic stem cells is a general phenomenon in many different cancers.[37,38] It usually affects transcription factors and developmental genes, especially those from the homeobox gene family. Our data show high concordance with these previous findings. Despite its widespread occurrence, the role of this specific pattern is still not completely understood; some studies showed that hypermethylation of bivalent regions continues the repression of

the genes involved,[37] but others found increased gene expression related to these regions.[38]

Integrative clustering is a powerful tool to identify patient subgroups. Our iCluster analysis revealed four different molecular iCCA groups, designated as the IDH, L, M, and H groups. The most conspicuous subgroup is the IDH group characterized by gain-of-function *IDH1* and *IDH2* hotspot mutations in 9 of 10 tumor samples sequenced, many CNAs, DNA hypermethylation, and high LMC2 values. The H group has a similarly disrupted genome as the IDH group with frequent deletions of chromosome arms 3p and 6q. The 3p deletions include tumor-suppressor genes like *BAP1* and *PBRM1*, found to be frequently inactivated in iCCA.[6] Amplification of 1q, which is seen in the M and H groups, is common in many different cancers and often associated with bad prognosis.[39] In HCC, this amplification occurs in more than 70% of cases[40] and has been described in iCCA as well.[41] It harbors the stem cell–related oncogene *YY1AP1* and genes of the phosphoinositide signaling pathway known to be involved in breast cancer.[24,25] A possible candidate gene of the 1q region, *PIP5K1A*, is a potential therapeutic target of kinase inhibitors, of which some are already in clinical trials.[22] *SETDB1*, which is amplified in almost 50% of the patients in the H and M groups, is discussed as a therapeutic target, showing sensitivity against mithramycin.[42]

Gain-of-function mutations of *IDH1* and *IDH2* are known to lead to DNA hypermethylation by the enrichment of the oncometabolite 2-hydroxyglutarate, which inhibits TET enzymes involved in active demethylation.[20] IDH mutations alone were shown to be sufficient to induce a hypermethylator phenotype,[43] and they tend to appear more frequently in recurrent tumors with gene-expression traits of epithelial-mesenchymal transition.[44] Targeted therapies for IDH-mutant tumors are already in clinical trials.[45,46] Our pathway analyses, consistent with the general hypermethylation of IDH-mutant tumors, showed no strong enrichment in any pathways (Supporting Fig. S9), but the transcription factor binding site-enrichment analysis revealed enrichment of hypermethylation at *CTCF* and *RAD21* binding sites. This is largely in line with the recent finding of Flavahan et al.,[47] showing that hypermethylation led to disruption of the insulator function of the CTCF protein in IDH-mutant gliomas and subsequent overexpression of the *PDGFRA* candidate oncogene. Our results suggest a similar mechanism in the IDH-mutant iCCAs. Both the H and M groups have a specific hypermethylation profile that is less pronounced in the M group and different from the pattern associated with the IDH group. This profile is enriched in embryonic stem cell–related bivalent regulatory regions that are indicators of pluripotency in stem cells and whose methylation manifests silencing during differentiation.[48] A strong difference between the M and H group is the enrichment of LMC4 in the H group. This enrichment, combined with the similar methylation profile, suggests related mechanisms of tumorigenesis, but probably a different cell-of-origin. Whether the L group and the IDH group with enrichment in high LMC1 and LMC2 values, respectively, may be traced to distinct cells-of-origin, or whether these profiles reflect properties independent thereof, remains to be clarified. Multiple cells-of-origin have been proposed for CCA, and the distinct cells-of-origin is supported by a recent pan-cancer integrative clustering that showed diverse cluster memberships for cholangiocarcinoma, where the clusters dominantly reflected on cells-of-origin-associated signals.[12,13,49] Our combined clustering based on the LMC patterns of PDAC, HCC, and CCA (Fig. 4E and Supporting Fig. S10) showed that bile duct–type iCCA resembled PDAC, whereas cholangiolar-type iCCA showed similar patterns to a subgroup of HCC. This supports the hypothesis that the cells-of-origin of cholangiolar type and bile duct–type iCCA differ, as previously proposed.[2] In addition, subsets of HCC and PDAC cases clustered together with the iCCA L and M groups, respectively, indicating potentially different cells-of-origin. Finally, IDH gain-of-function mutation is clearly associated with LMC2-high profiles in all cancer types. Therefore, IDH gain of function may overwrite parts of the cell-of-origin DNA methylation profiles dominating the resulting LMC profiles.

In this study, we used biliary epithelia microdissected from nonneoplastic cystic ducts as normal control tissues. Comparative methylation data analyses showed similar DNA methylation patterns and LMC profiles of L group iCCA and the normal control samples, indicating a large bile duct cholangiocyte as the cell-of-origin of the L group iCCA. However, due to the lack of direct matched-pair samples, direct comparisons cannot be conducted here. In addition, it would be very interesting to analyze matched-pair DNA methylation profiles of biliary precursor lesions

and invasive CCAs in future studies. Consistent with the low burden of mutations and epigenetic alterations compared with the normal cholangiocytes, the L group exhibited the most favorable outcome with a 3-year survival rate of 91%, compared with 65%, 50%, and 36% for the IDH, H, and M groups, respectively.

Further analyses are needed to reveal alterations in the background of the poor survival of the M group. A limitation of our study is that whole genome sequencing and gene-expression data that could reveal a hidden candidate gene affecting the prognosis in this group are missing. Our observation of group-specific clinical outcomes requires particular consideration: Both the L and IDH groups appear to have a better outcome than the other two groups. The favorable outcome of the IDH group is in agreement with two other studies.[19,34]

In summary, we identified four different iCCA groups that probably differ by their cell-of-origin, underlying tumorigenic mechanism, and clinical outcome. Thus, the four iCCA subgroups presented in this study demonstrate options for the stratification of patients according to our molecular profiling and may lead to subgroup-specific treatment modalities in the future.

Acknowledgment: We thank the Genomics and Proteomics Core Facility of the German Cancer Research Center for providing excellent exome sequencing and microarray services. We also thank Veronika Geißler (NCT Tissue Bank, Heidelberg) for the excellent technical assistance.

REFERENCES

1) Banales JM, Cardinale V, Carpino G, Marzioni M, Andersen JB, Invernizzi P, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). Nat Rev Gastroenterol Hepatol 2016;13:261-280.
2) Liau JY, Tsai JH, Yuan RH, Chang CN, Lee HJ, Jeng YM. Morphological subclassification of intrahepatic cholangiocarcinoma: etiological, clinicopathological, and molecular features. Mod Pathol 2014;27:1163-1173.
3) **Wardell CP**, **Fujita M**, Yamada T, Simbolo M, Fassan M, Karlic R, et al. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. J Hepatol 2018;68:959-969.
4) Lowery MA, Ptashkin RN, Jordan EJ, Berger MF, Zehir A, Capanu M, et al. Comprehensive molecular profiling of intra- and extrahepatic cholangiocarcinomas: potential targets for intervention. Clin Cancer Res 2018;24:4154-4161.
5) O'Rourke CJ, Munoz-Garrido P, Aguayo EL, Andersen JB. Epigenome dysregulation in cholangiocarcinoma. Biochim Biophys Acta 2018;1864:1423-1434.
6) **Jiao Y**, **Pawlik TM**, **Anders RA**, Selaru FM, Streppel MM, Lucas DJ, et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. Nat Genet 2013;45:1470-1473.
7) Chan-On W, Nairismagi ML, Ong CK, Lim WK, Dima S, Pairojkul C, et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. Nat Genet 2013;45:1474-1478.
8) Kipp BR, Voss JS, Kerr SE, Barr Fritcher EG, Graham RP, Zhang L, et al. Isocitrate dehydrogenase 1 and 2 mutations in cholangiocarcinoma. Hum Pathol 2012;43:1552-1558.
9) **Nakanuma Y**, **Sato Y**, **Harada K**, **Sasaki M**, **Xu J**, Ikeda H. Pathological classification of intrahepatic cholangiocarcinoma based on a new concept. World J Hepatol 2010;2:419-427.
10) Marquardt JU, Andersen JB, Thorgeirsson SS. Functional and genetic deconstruction of the cellular origin in liver cancer. Nat Rev Cancer 2015;15:653-667.
11) Bragazzi MC, Ridola L, Safarikia S, Matteo SD, Costantini D, Nevi L, et al. New insights into cholangiocarcinoma: multiple stems and related cell lineages of origin. Ann Gastroenterol 2018;31:42-55.
12) Carpino G, Cardinale V, Folseraas T, Overi D, Grzyb K, Costantini D, et al. Neoplastic transformation of peribiliary stem cell niche in cholangiocarcinoma arisen in primary sclerosing cholangitis. Hepatology. In press.
13) Komuta M, Spee B, Vander Borght S, De Vos R, Verslype C, Aerts R, et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. Hepatology 2008;47:1544-1556.
14) Baron U, Turbachova I, Hellwag A, Eckhardt F, Berlin K, Hoffmuller U, et al. DNA methylation analysis as a tool for cell typing. Epigenetics 2006;1:55-60.
15) Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics 2009;1:177-200.
16) Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun 2015;6:8971.
17) **Lutsik P**, **Slawski M**, Gasparoni G, Vedeneev N, Hein M, Walter J. MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. Genome Biol 2017;18:55.
18) Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics 2009;25:2906-2912.
19) **Wang P**, **Dong Q**, Zhang C, Kuan PF, Liu Y, Jeck WR, et al. Mutations in isocitrate dehydrogenase 1 and 2 occur frequently in intrahepatic cholangiocarcinomas and share hypermethylation targets with glioblastomas. Oncogene 2013;32:3091-3100.
20) **Xu W**, **Yang H**, **Liu Y**, Yang Y, Wang P, Kim SH, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 2011;19:17-30.
21) **Zou S**, **Li J**, Zhou H, Frech C, Jiang X, Chu JS, et al. Mutational landscape of intrahepatic cholangiocarcinoma. Nat Commun 2014;5:5696.
22) **Rubio-Perez C**, **Tamborero D**, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell 2015;27:382-396.
23) Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539-542.

24) Waugh MG. Amplification of chromosome 1q genes encoding the phosphoinositide signalling enzymes PI4KB, AKT3, PIP5K1A and PI3KC2B in breast cancer. J Cancer 2014;5:790-796.
25) Zhao X, Parpart S, Takai A, Roessler S, Budhu A, Yu Z, et al. Integrative genomics identifies YY1AP1 as an oncogenic driver in EpCAM(+) AFP(+) hepatocellular carcinoma. Oncogene 2015;34:5095-5104.
26) Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology 2009;136:1012-1024.
27) **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.
28) Mahmood SF, Gruel N, Chapeaublanc E, Lescure A, Jones T, Reyal F, et al. A siRNA screen identifies RAD21, EIF3H, CHRAC1 and TANC2 as driver genes within the 8q23, 8q24.3 and 17q23 amplicons in breast cancer with effects on cell growth, survival and transformation. Carcinogenesis 2014;35:670-682.
29) Lei B, Zhang XY, Zhou JP, Mu GN, Li YW, Zhang YX, et al. Transcriptome sequencing of HER2-positive breast cancer stem cells identifies potential prognostic marker. Tumour Biol 2016;37:14757-14764.
30) **Hovestadt V**, **Jones DT**, Picelli S, Wang W, Kool M, Northcott PA, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature 2014;510:537-541.
31) **Pajtler** KW, **Witt** H, **Sill** M, Jones DT, Hovestadt V, Kratochwil F, et al. Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. Cancer Cell 2015;27:728-743.
32) **Ong CK**, **Subimerb C**, Pairojkul C, Wongkham S, Cutcutache I, Yu W, et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. Nat Genet 2012;44:690-693.
33) Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, et al. Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. Cell Rep 2017;18:2780-2794.
34) **Jusakul A**, **Cutcutache I**, **Yong CH**, **Lim JQ**, Huang MN, Padmanabhan N, et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. Cancer Discov 2017;7:1116-1135.
35) Zhao Y, Yu H, Hu W. The regulation of MDM2 oncogene and its impact on human cancers. Acta Biochim Biophys Sin (Shanghai) 2014;46:180-189.
36) Manfredi JJ. The Mdm2-p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor. Genes Dev 2010;24:1580-1589.
37) **Easwaran H**, **Johnstone SE**, Van Neste L, Ohm J, Mosbruger T, Wang Q, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. Genome Res 2012;22:837-849.
38) Bernhart SH, Kretzmer H, Holdt LM, Juhling F, Ammerpohl O, Bergmann AK, et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. Sci Rep 2016;6:37393.
39) Muthuswami M, Ramesh V, Banerjee S, Viveka Thangaraj S, Periasamy J, Bhaskar Rao D, et al. Breast tumors with elevated expression of 1q candidate genes confer poor clinical outcome and sensitivity to Ras/PI3K inhibition. PLoS One 2013;8:e77553.
40) Kusano N, Shiraishi K, Kubo K, Oga A, Okita K, Sasaki K. Genetic aberrations detected by comparative genomic hybridization in hepatocellular carcinomas: their relationship to clinicopathological features. Hepatology 1999;29:1858-1862.
41) Dalmasso C, Carpentier W, Guettier C, Camilleri-Broet S, Borelli WV, Campos Dos Santos CR, et al. Patterns of chromosomal copy-number alterations in intrahepatic cholangiocarcinoma. BMC Cancer 2015;15:126.
42) Rodriguez-Paredes M, Martinez de Paz A, Simo-Riudalbas L, Sayols S, Moutinho C, Moran S, et al. Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. Oncogene 2014;33:2807-2813.
43) **Turcan S**, **Rohle D**, **Goenka A**, Walsh LA, Fang F, Yilmaz E, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 2012;483:479-483.
44) Peraldo-Neia C, Ostano P, Cavalloni G, Pignochino Y, Sangiolo D, De Cecco L, et al. Transcriptomic analysis and mutational status of IDH1 in paired primary-recurrent intrahepatic cholangiocarcinoma. BMC Genom. 2018;19:440.
45) Stein E, Yen K. Targeted differentiation therapy with mutant IDH inhibitors: early experiences and parallels with other differentiation agents. Annu Rev Cancer Biol 2017;1:379-401.
46) Saha SK, Gordan JD, Kleinstiver BP, Vu P, Najem MS, Yeo JC, et al. Isocitrate dehydrogenase mutations confer dasatinib hypersensitivity and SRC dependence in intrahepatic cholangiocarcinoma. Cancer Discov 2016;6:727-739.
47) **Flavahan WA**, **Drier Y**, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature 2016;529:110-114.
48) Coskun V, Tsoa R, Sun YE. Epigenetic regulation of stem cells differentiating along the neural lineage. Curr Opin Neurobiol 2012;22:762-767.
49) **Hoadley KA**, **Yau C**, **Hinoue T**, **Wolf DM**, **Lazar AJ**, **Drill E**, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 2018;173:291-304, e296.
50) Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., Kinzler KW. Cancer genome landscapes. Science 2013;339:1546-1558.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.30493/suppinfo.