Molecular Classification and Therapeutic Targets in Extrahepatic Cholangiocarcinoma

Authors:

Robert Montal, Daniela Sia, Carla Montironi, Wei Q. Leow, Roger Esteban-Fabró, Roser Pinyol, Miguel Torres-Martin, Laia Bassaganyas, Agrin Moeini, Judit Peix, Laia Cabellos, Miho Maeda, Carlos Villacorta-Martin, Parissa Tabrizian, Leonardo Rodriguez-Carunchio, Giancarlo Castellano, Christine Sempoux, Beatriz Minguez, Timothy M. Pawlik, Ismail Labgaa, Lewis R. Roberts, Manel Sole, Maria I. Fiel, Swan Thung, Josep Fuster, Sasan Roayaie, Augusto Villanueva, Myron Schwartz, Josep M. Llovet.

Supplementary Material Contents:

Supplementary Methods

Pathological characterization

Further pathological characterization of tumors in terms of histological type, cell differentiation, perineural invasion, vascular invasion, precursor lesions, growth pattern, tumor purity, presence of fibrotic tissue (0 = none; 1 = mild; 2 = moderate; 3 = marked) and presence of lymphocyte infiltration (0 = absence; 1 = minimal; 2 = mild; 3 = moderate; 4 = severe; samples with staining between 0 and 2 were considered with low immune infiltration whereas 3 and 4 scores were classified as high immune infiltration) was conducted by 2 independent liver pathologists (W.Q.L. and C.M.) (Table 1, Table S1).

RNA and DNA isolation

Tumoral tissue sections were macro-dissected to avoid contamination of non-cancerous tissue. Total RNA was isolated from three freshly cut 5µm-thick FFPE sections using QIAcube (Qiagen). RNA quantity was assessed using Quant-iT Ribogreen RNA assay kit (Invitrogen). RNA quality was checked by real-time quantitative reverse transcription PCR (qRT-PCR) of RPL13A (cut-off Ct<28 cycles). Genomic DNA was isolated from seven freshly cut 5µm-thick FFPE sections using QIAcube (Qiagen). DNA quantity was assessed using a Quant-It PicoGreen dsDNA Assay kit (Invitrogen). To determine DNA quality, we used qRT-PCR of RNase P (Applied Biosystems).

Whole-genome expression

Unsupervised clustering

Principal Component Analysis (PCA) was initially conducted in the whole eCCA cohort using the prcomp function from the R package stats v3.6.2 in order to obtain the distribution of samples depending on two principal components (Fig. S19). pCCA and dCCA were not differentially distributed in the plot, indicating that it was reasonable to group together these two anatomical locations of eCCA for the transcriptome-based molecular clustering.

Non-negative matrix factorization (NMF) from NMFConsensus module in GenePattern[1] was employed to identify stable gene expression clusters. NMF parameters: k = 1 to k = 5 clusters; number of clusterings to build consensus matrix = 20; number of iterations = 2000; error function = Euclidean. In order to remove noise, the top 1696 most variable genes, identified with the Preprocess Dataset module in Genepattern[1], were used as input[2]. The preferred clustering result was determined using the observed cophenetic correlation, which measures the stability based on distances between clusters (Fig. 2A). Previous studies showed that transcriptome-based clustering was consistent independently of the exclusion of low purity samples[2]. No batch effect was observed between center of origin and molecular classes (Fig. 3A).

One of the molecular classes obtained with unsupervised clustering presented overexpression of classic hepatocyte markers such as albumin, transferrin and *CYP3A4* (Fig. 2F). A similar finding was observed during the molecular classification of CCA conducted in the TCGA project[3]. However, they considered this expression profile as a

result of contamination by even a small amount of non-tumoral liver. To exclude that a potential hepatic contamination could be the determinant of this biological traits, we assessed hematoxylin and eosin slides to quantify the percentage of non-tumoral liver in our macro-dissected samples (Fig. S6A). Of the 182 samples with available transcriptome, 62 (34%) had >1% [range 1-40%] non-tumoral liver inside the macro-dissected area, without a significant association with any molecular class despite a trend was observed in the Metabolic. Next, we repeated the unsupervised clustering in the eCCA cohort including only those samples without any non-tumoral liver in the slide (in and out the macro-dissected area) to grant exclusion of any potential hepatic contamination (n=93) (Fig. S6B). The four molecular classes obtained paralleled the ones obtained with the whole eCCA cohort. Specifically, 68% of the hallmarks defining each molecular class persisted significant even with a lower statistical power. Furthermore, to minimize the impact of non-tumoral liver expression from the transcriptome, we filtered out 386 liver-specific genes derived from the GTEx normal tissue expression database as was done in the TCGA study[3] (Fig. S6C). Unsupervised clustering in four molecular classes obtained an almost perfect overlap with the previously proposed classes (99% of the samples fell into the same class when these genes were subtracted). Finally, to determine a specific subset of genes particularly defining non-tumoral liver expression in our dataset we applied NMF to perform virtual microdissection of gene expression data as previously described[4]. A liver-related expression factor comprising 149 genes (Table S14) was unveiled by computing overlaps of selected genes with curated gene sets from MSigDB collections[5] (Table S15). Again, unsupervised clustering excluding these 149

genes obtained an almost perfect overlap with the previously proposed classes (97% of the samples fell into the same class) (Fig. S6D).

Gene Set Enrichment Analysis

Hallmark gene sets[5] from MSigDB collections, representing 50 well-defined biological states or processes collections, were evaluated using single-sample Gene Set Enrichment Analysis (ssGSEA) Projection from GenePattern[1]. Each enrichment score represents the degree of which the genes in a particular gene set are coordinately up- or down-regulated within a sample (Fig. 2B).

Virtual microdissection of tumor-microenvironment using gene expression data was conducted using the Estimation of STromal and Immune cells in MAlignant Tumors (ESTIMATE) package[6]. This method based on ssGSEA algorithm allows the calculation of the stromal and immune compartment in tumoral tissue (Fig. 2D-E).

The same ssGSEA tool and the Pan-cancer Immune Metagenes described in The Cancer Immunome Atlas[7] were used to estimate the infiltration in the tumors of 28 immune subpopulations including TILs as well as cell types related to innate immunity (Fig. 2C).

The Tumor Immune Dysfunction and Exclusion (TIDE) transcriptome-based algorithm[8] was applied to quantify dysfunction and exclusion of infiltrating cytotoxic T lymphocytes (Fig. S13).

Upstream transcriptional regulators

Genes differentially expressed between molecular classes (FDR<0.01) were identified with the Comparative Marker Selection module from GenePattern[1]. Ingenuity Pathway analysis software (Qiagen) was used for the inference of putative upstream regulators explaining the observed gene expression changes among the identified molecular classes (Table S8).

Molecular class prediction

Prediction in the eCCA cohort of previously reported mRNA-based molecular classes of CCA[9–11], hepatocellular carcinoma[12–14] and pancreatic adenocarcinoma[4,15] was performed using the Nearest Template Prediction method, as implemented in the specific module of GenePattern[1] (Fig. 3A). In addition, the similarity between transcriptome profiles of two independent data sets was analyzed with the SubMap module from GenePattern[1] (Fig. 3F-G).

Gene expression signature design

The Class Neighbors tool from GenePattern[1] was used to determine based on a signal-to-noise distance function which genes were most closely correlated with a specific molecular class template and how significant the correlation was compared with random permutation versions of the phenotype (intersection of observed data with 1% significance level)[16]. The selection of up to 25, 50, 75 and 100 genes per class for the construction of the gene expression signature seemed likely to be large enough to be robust against noise and small enough to be applied in a clinical setting (Fig. S14). The accuracy of

gene-expression signatures was tested on the discovery data set using the Nearest Template Prediction method.

Targeted DNA-sequencing

Mutation calling and interpretation

All sequenced genes had homogenous mean coverage allowing an unbiased interpretation of structural genomic aberrations (Fig. S20). Cancer-specific variant calling was performed with SNPPET algorithm (Agilent) and the following criteria: Variant score threshold = 0.3, minimum quality for base = 30, variant call quality threshold = 100, minimum allele frequency = 0.01, minimum number of reads supporting variant allele = 10.

Interpretation of variants was conducted by Cancer Genome Interpreter[17], a software that relies on existing knowledge collected from several resources (DoCM, ClinVar, OncoKB and IARC) and on a computational method that estimates the oncogenic effect of variants of uncertainty significance (OncodriveMUT). Candidate mutations were considered to be the ones already known to be oncogenic as well as the predicted drivers in Tier 1. Variants with an allele frequency lower than 5% were ultimately excluded in order to enrich the potential biological and clinical impact of results (Table S3). Accurate interpretation of PMS2 and KMT2C mutations was not feasible due to the interferences of pseudogenes (highly homologous sequences)[18]. The packages maftools and PathwayMapper were used for visualizing the data.

Copy number analysis

In order to detect copy number variations (CNV) from the targeted DNA-sequencing panel, we used the multifactor normalization tool ONCOCNV[19] version 6.9, designed specifically for CNV analysis of amplicon panels. A total of 15 matched non-neoplastic bile ducts were used for normalization, and 150 tumors were evaluated. Focal amplifications were called at segments with ≥6 copies and homozygous deletions at segments with 0 copies[20] (Table S4).

Immunohistochemistry

HER2 evaluation was done according to the recommendations for its testing in breast cancer[21], being positive when IHC was 3+ (circumferential membrane staining that is complete, intense, and within > 10% of tumor cells). PD-1 positivity was defined by an unequivocal cytoplasmic staining of lymphocytes in >5% over the total number of intra-tumoral lymphocytes[22]. PD-L1 positive samples were defined by unequivocal membranous staining of tumor cells or stromal cells in >1% over the total number of cells[23]. For α-SMA evaluation, the intensity of the staining in vascular smooth muscle cells was used as the reference staining value. α-SMA staining in tumor fibroblasts was qualitatively classified into 4 groups (0 = absence; 1 = much lower intensity; 2 = slightly lower intensity; 3 = equal intensity)[24]. Positive α-SMA was defined by ≥2 intensity in >50% of fibroblasts. Tumor testing for DNA MMR deficiency with immunohistochemistry for MMR proteins was conducted as recommended for screening of Lynch syndrome in hereditary nonpolyposis colorectal cancer[25]. Hep Par 1 and CK19 staining was qualitatively classified into 3 groups (0 = absence; 1 = low intensity; 2 = high intensity).

Positive Hep Par 1 and CK19 was defined by any positive intensity in >10% of tumoral cells. Ki67 percentage evaluation was determined in tumoral cells as previously described in breast cancer[26].

In situ hybridization

FISH testing was performed on serially cut 3µm paraffin-embedded tissue sections to validate ERBB2 amplifications identified by DNA-targeted exome sequencing (Fig. S4). One slide was stained with H&E and reviewed by a pathologist to identify areas of tumor. The *HER2* genetic testing was performed using the XL ERBB2 (HER2/NEU) amp probe (D-6010-100-OG, Metasystems, Altlußheim, Germany) according to the manufacturer's instructions. Samples were analyzed independently by two evaluators using the platform slide scanning system Metafer version 3.5 (MetaSystems) combined with image analysis (MetaSystems). At least 20 neoplastic cells were examined for each sample. Amplification of *HER2* was defined when the average copy number ratio, *HER2*/CEN17 was ≥ 2.0 or when the average of *HER2* signals per nuclei was ≥ 6.

ICGC and TCGA RNAseq analysis

Regarding BTC-ICGC, in order to remove biases in downstream analysis caused by differences in sequencing depth, we computed sample normalization factors using edgeR[27] and scaled the library sizes accordingly. Specifically, we used the trimmed mean of M values (TMM method) for estimating the library size before scaling and then normalized by transcript length too, resulting in the RPKM matrix. One sample (BD20) was not successfully normalized. Data from HCC-TCGA and PDAC-TCGA were already normalized when downloaded at https://www.cbioportal.org. Prediction in the external

cohort of the eCCA classifier was performed using the Nearest Template Prediction method, as implemented in the specific module of GenePattern[1] (Fig. 4A, Fig. S16).

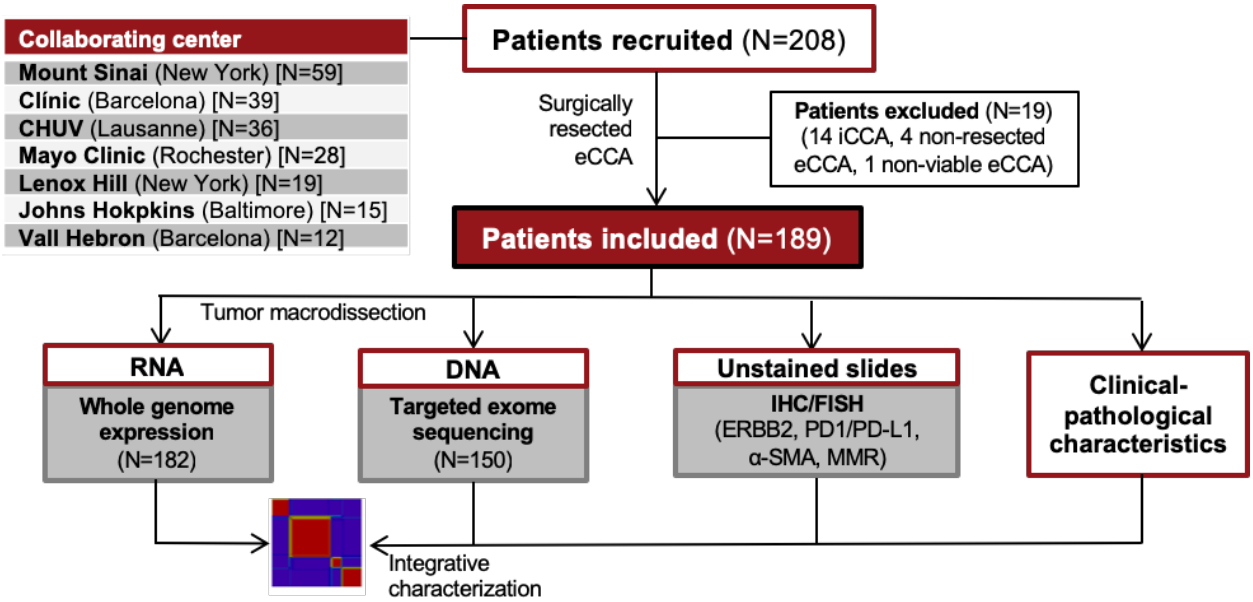
TIGER-LC metabolomics analysis

To better delineate metabolites defining the eCCA Metabolic class, we used gene expression data deposited in the NCBI GEO under accession code GSE76297 to infer hepatobiliary tumors (HCC and iCCA) from the TIGER-LC Consortium[28] recapitulating the proposed eCCA classes. Prediction in the external cohort of the eCCA classifier was performed using the Nearest Template Prediction method, as implemented in the specific module of GenePattern[1]. Metabolome data was obtained in 140 samples from Metabolon's Discover HD4 Platform (718 metabolites). Metabolite Set Enrichment Analysis from MetaboAnalyst 4.0[29] was used to identify biologically meaningful patterns defined by the eCCA Metabolic class (Fig. S17).

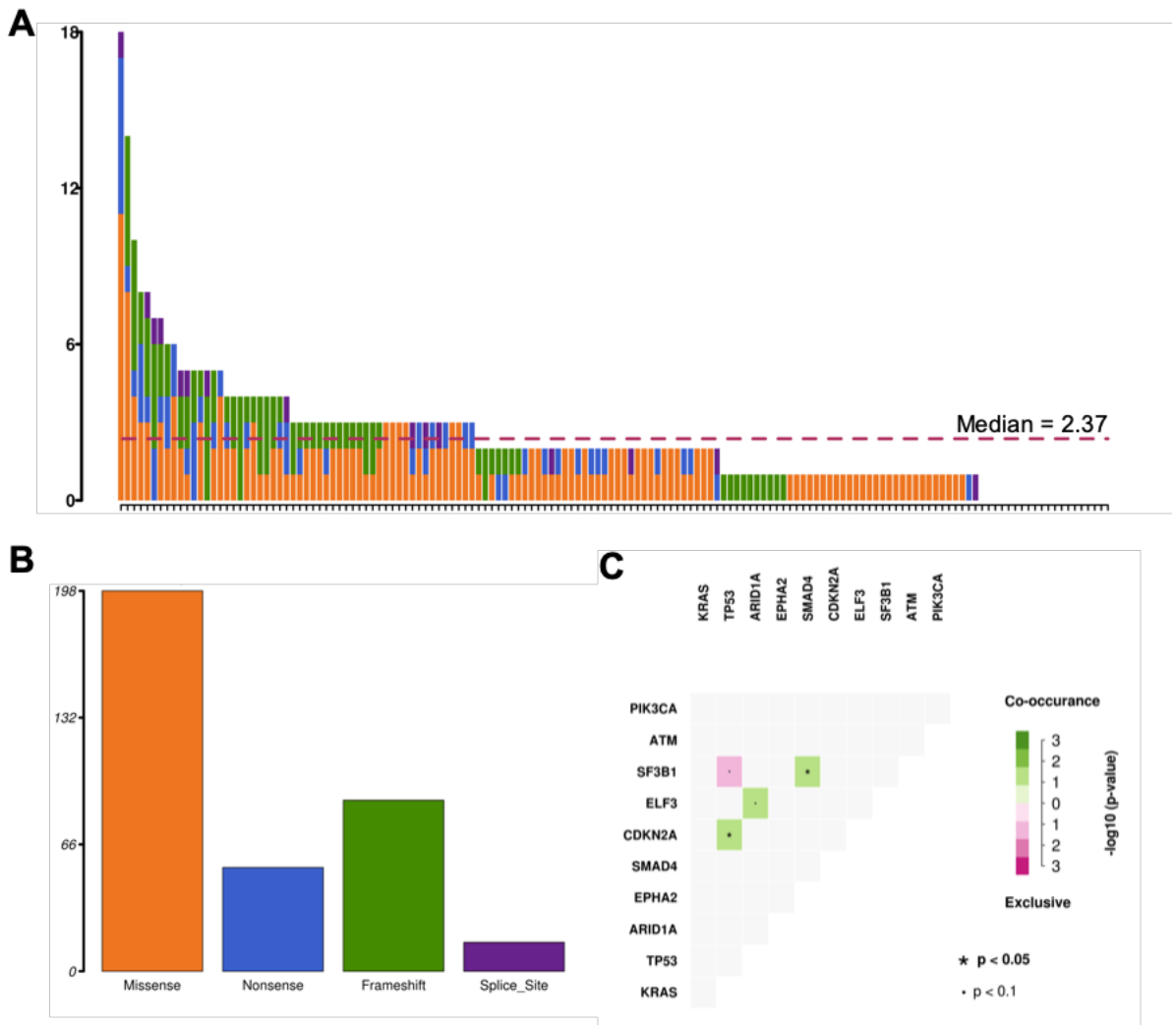
Ongoing clinical trials

Data of ongoing clinical trials was obtained in March 2019 from the ClinicalTrials.gov database. Keyword searches for "cholangiocarcinoma" and "biliary tract cancer" were used to identify active clinical trials (recruiting, not yet recruiting, active, not recruiting, enrolling by invitation) assessing targeted therapies for advanced eCCA. Basket trials assessing solid tumors other than hepato-biliary-pancreatic tumors were excluded (Table S13).

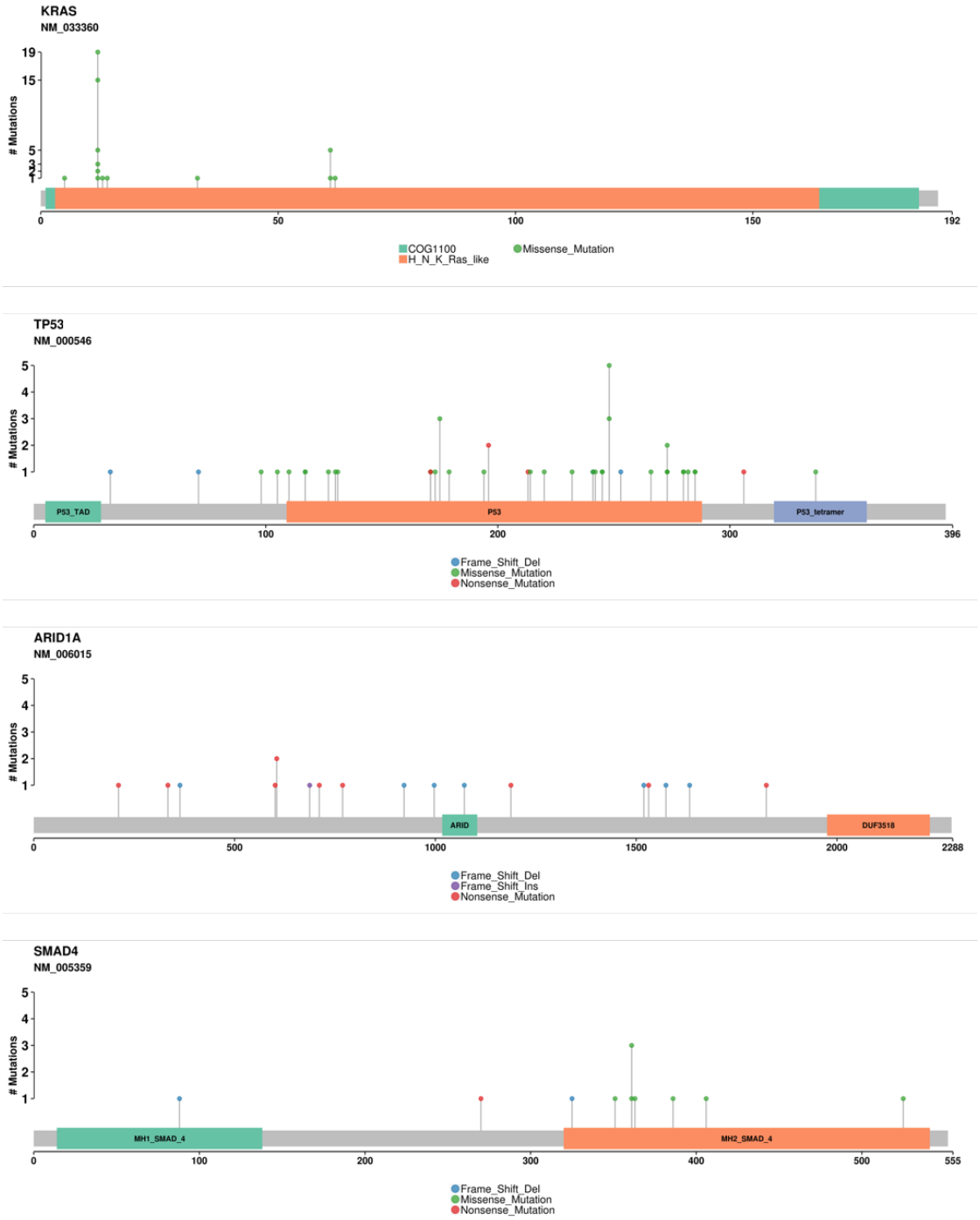
Supplementary Figures



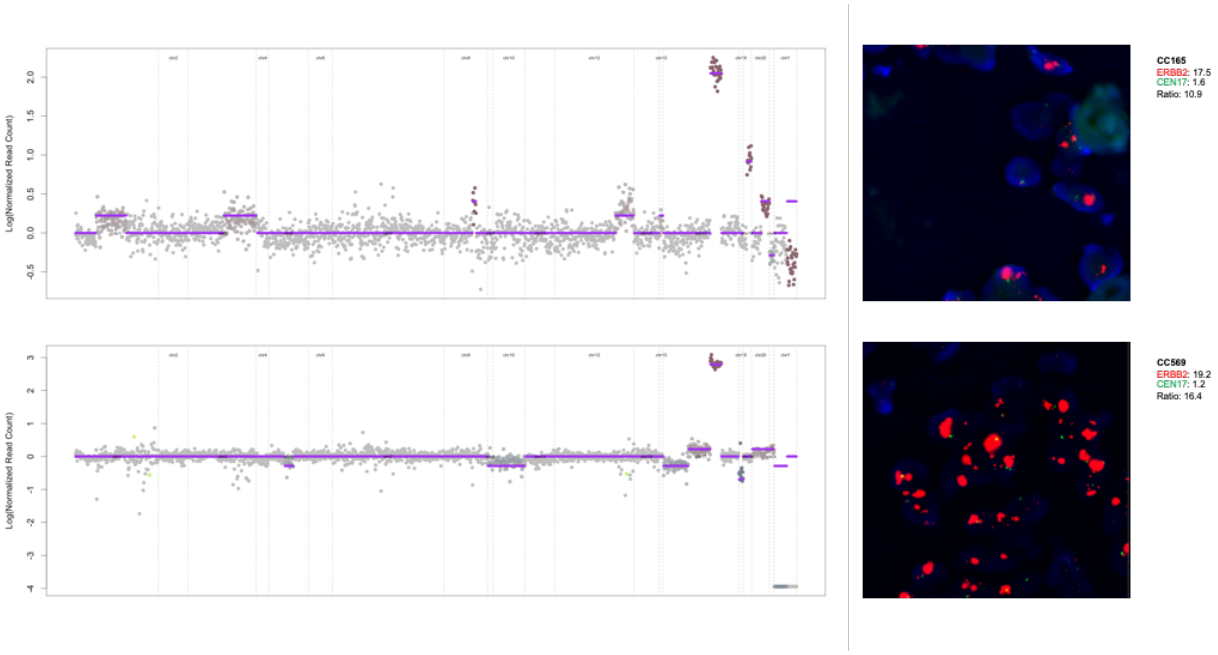
Supplementary Fig. 1. Flow chart of the eCCA study. Samples from surgically resected eCCA were collected from 7 international centers and analyzed using whole-genome expression, targeted DNA-sequencing and IHC/FISH.



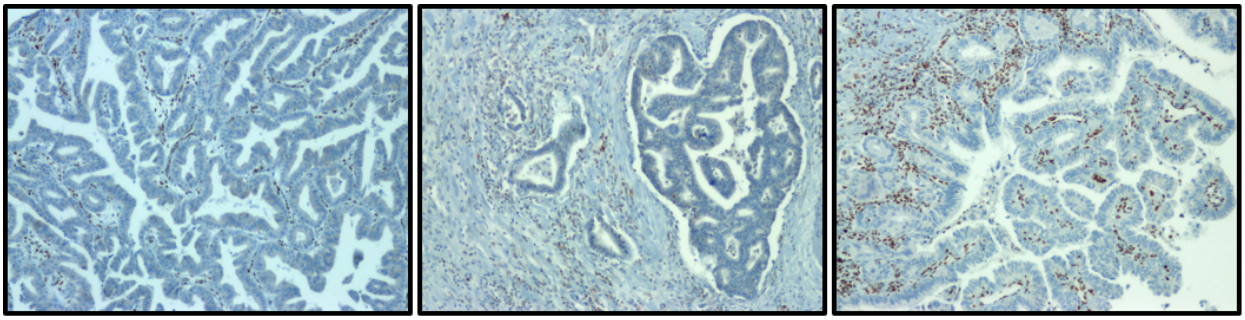
Supplementary Fig. 2. Tumor mutational burden in eCCA. (A), Total number of mutations per eCCA sample ranked by their mutational burden (maximum = 18, minimum = 0, median = 2.37). Color of bars represents the type of structural genomic alteration: orange: missense; purple: nonsense; green: frameshift indel; blue: splice site. **(B),** Prevalence of type of structural genomic alteration in the eCCA cohort. **c,**Top 10 most frequently mutated genes and their co-occurrence or mutually exclusivity. P values were calculated using a two-sided Fisher's exact test.



Supplementary Fig. 3. Distribution of mutations in *KRAS*, *TP53*, *ARID1A* and *SMAD4*. Mutation diagrams for the four most mutated genes in eCCA. Known or predicted driver mutations are visualized within the functional domains of the respective protein using maftools.

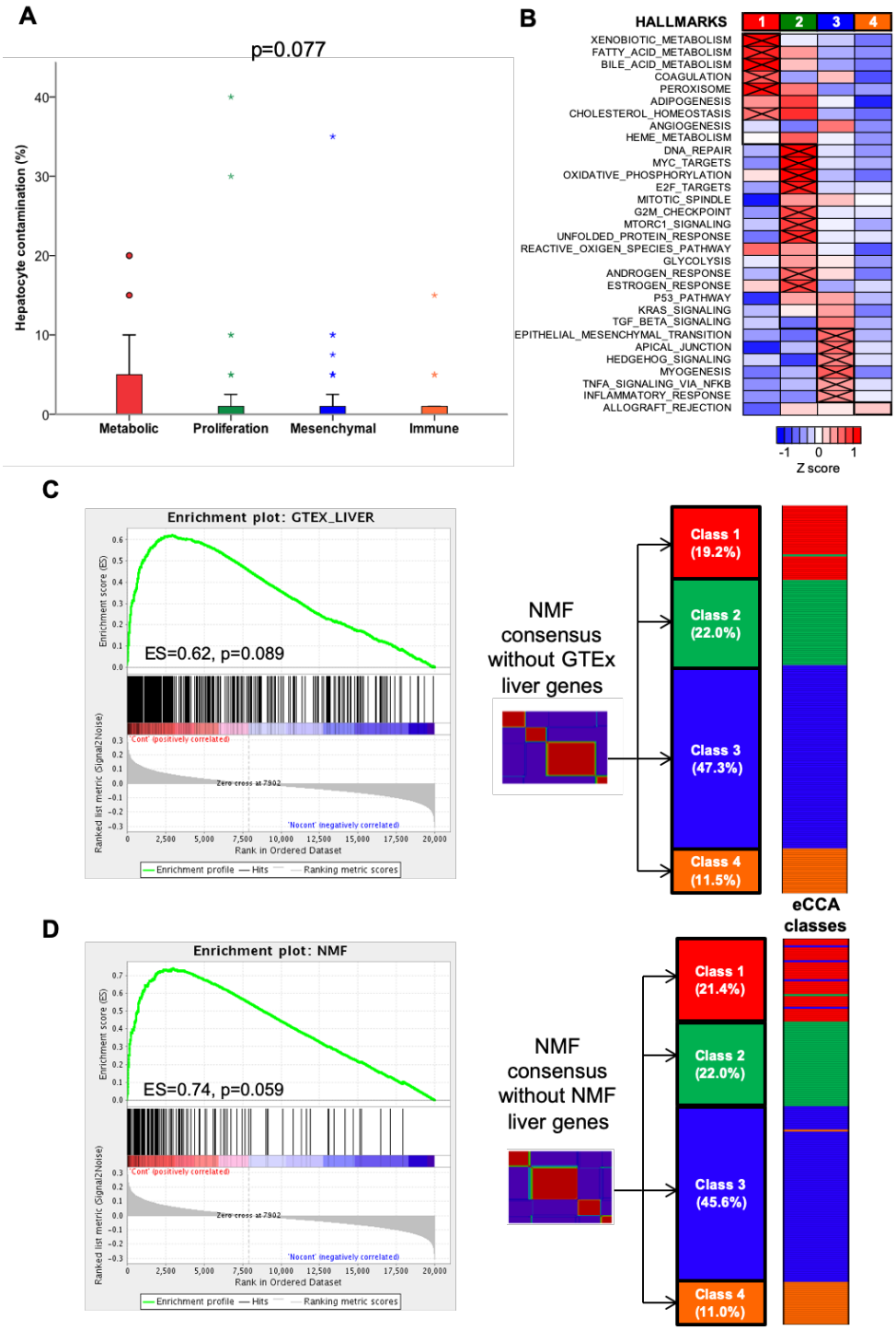


Supplementary Fig. 4. *ERBB2* amplifications in eCCA by FISH. Focal amplifications of *ERBB2* (segments with ≥6 copies) identified in two eCCA samples by targeted DNA-sequencing and subsequent validation by FISH.



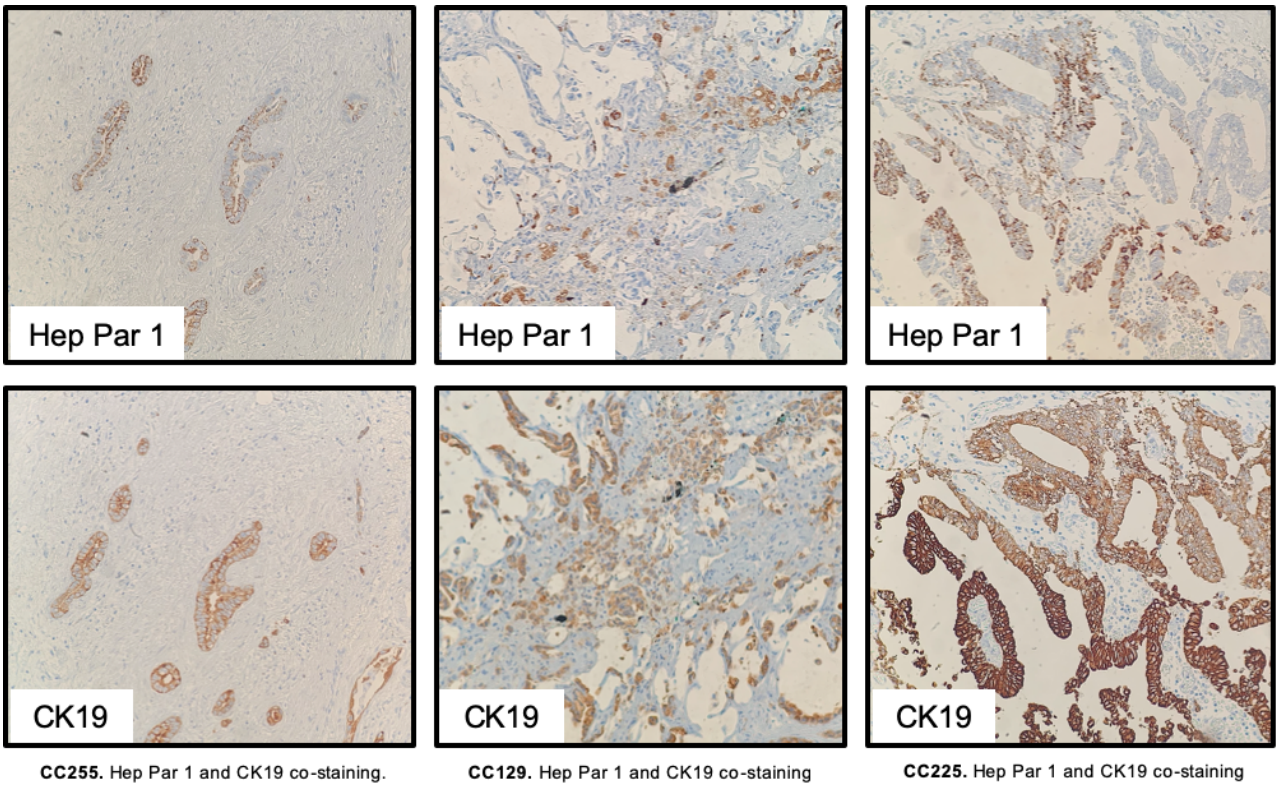
CC411. MSH2 mutation and loss of expression. **CC499.** MSH2 mutation and loss of expression. **CC593.** MSH6 mutation and loss of expression.

Supplementary Fig. 5. MMR repair deficiency in eCCA. Loss of expression of MMR proteins (*MSH2* and *MSH6*) by IHC in three of the four eCCA samples with available tissue and with the presence of mutations in the same gene detected by targeted DNA-sequencing.

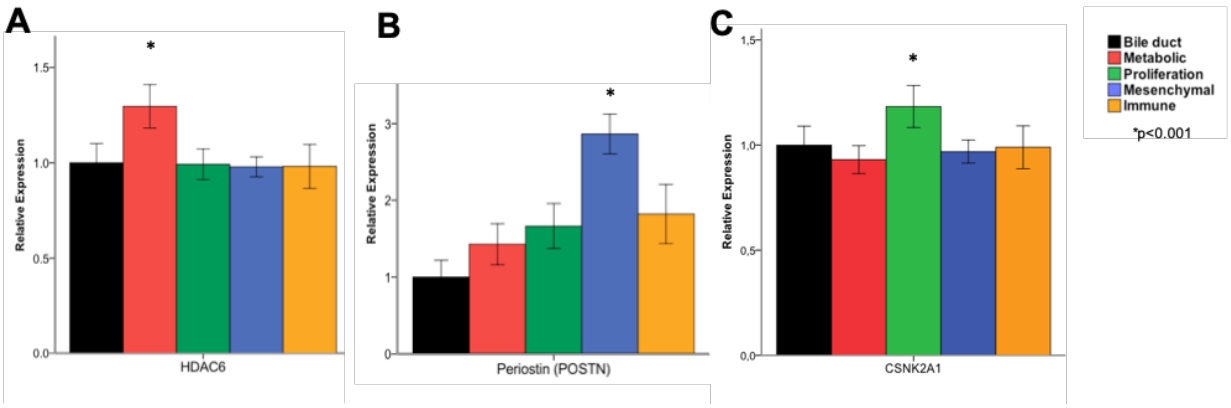


Supplementary Fig. 6. Molecular eCCA classes and potential hepatic contamination. (A), Percentage of non-tumoral liver contamination in macro-dissected samples assessed from hematoxylin and eosin slides. P value was calculated using

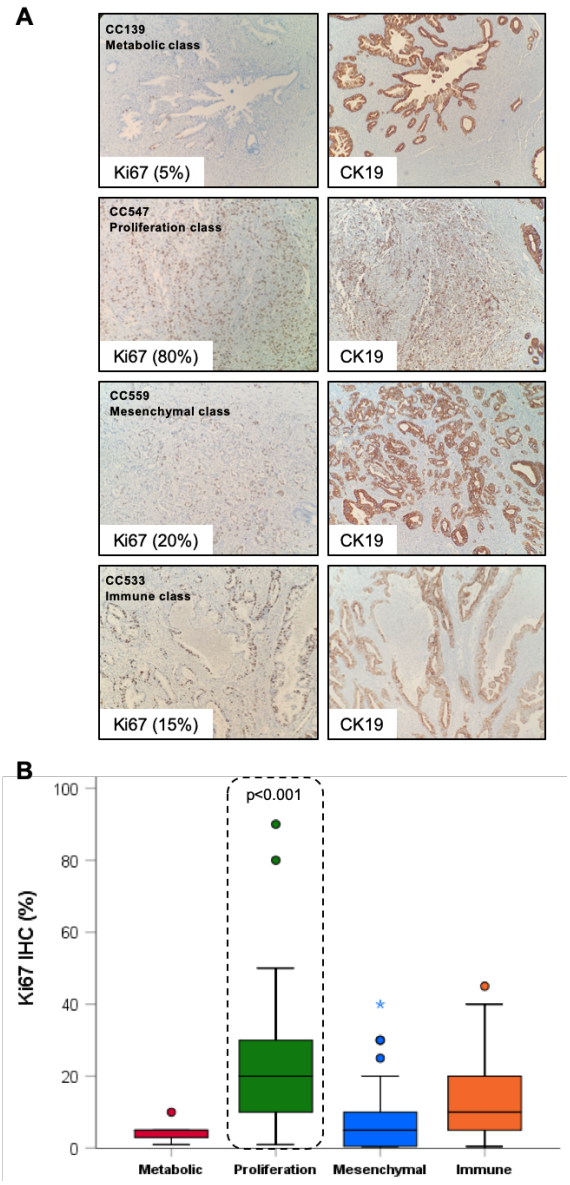
Kruskal-Wallis Test. **(B),** Heatmap of hallmark gene sets from MSigDB collections in four molecular classes obtained from unsupervised clustering of eCCA samples without potential hepatic contamination (n=93). The four molecular classes resemble the proposed Metabolic, Proliferation, Mesenchymal and Immune classes obtained with the whole eCCA cohort (Fig. 1b). Single-sample Gene Set Enrichment Analysis (ssGSEA) was used to obtain the enrichment score, representing the degree of which the genes in a particular gene set are coordinately up- or down-regulated. Samples from the same molecular class were represented with a normalized enrichment score. P values between a specific molecular class and the rest were calculated using T-Test, being crossed cells lower than 0.05. **(C),** Enrichment plot of 386 liver-specific genes derived from the GTEx normal tissue expression database[3] in eCCA samples with potential hepatic contamination and transcriptome-based unsupervised classification of eCCA filtering out these genes. The four molecular classes identified have an almost perfect overlap with the previously proposed clustering using the whole transcriptome: One Proliferation class tumor classified now as a Metabolic class. **(D),** Enrichment plot of 149 liver genes identified by NMF in eCCA samples with potential hepatic contamination and transcriptome-based unsupervised classification of eCCA filtering out these genes. The four molecular classes identified have an almost perfect overlap with the previously proposed clustering using the whole transcriptome: Four Mesenchymal class tumors classified now as a Metabolic class; One Proliferation class tumor classified now as a Metabolic class; and one Immune class tumor classified now as Mesenchymal class. GSEA was used to obtain the enrichment plot. Unsupervised classification of eCCA was done by non-negative matrix factorization consensus.



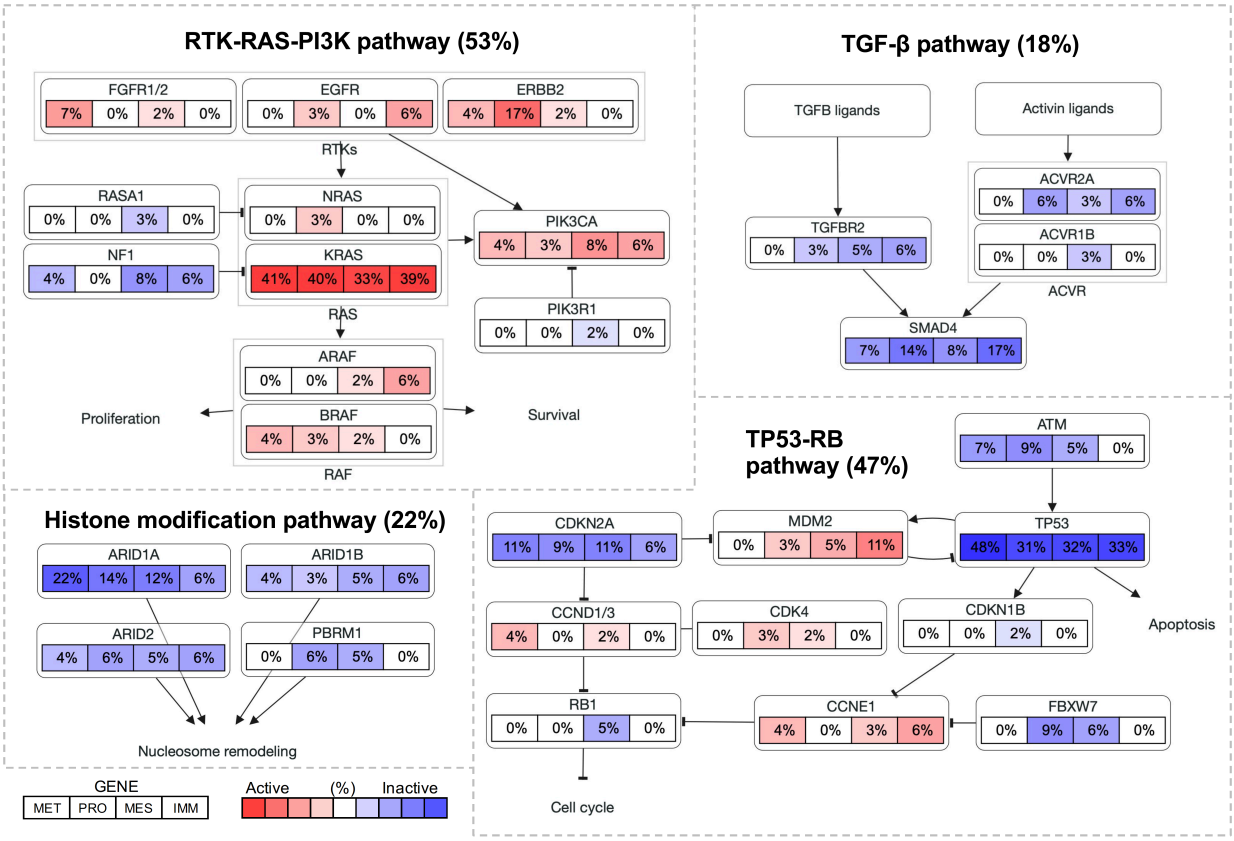
Supplementary Fig. 7. Hep Par 1 and CK19 co-staining in Metabolic eCCA class. Hep Par 1 (hepatocyte marker) and CK19 (cholangiocyte maker) IHC was conducted on a subset of the eCCA cohort (n=53) including Metabolic (n=23) and non-Metabolic tumors (Proliferation=6, Mesenchymal=19 and Immune=5). All tumors (100%) had a positive staining for CK19. On the other hand, positive staining for Hep Par 1 was observed in 14 tumors, most of them from the Metabolic class (Metabolic=43% vs Rest=13%, p=0.026). P value was calculated using a two-sided Fisher's exact test.



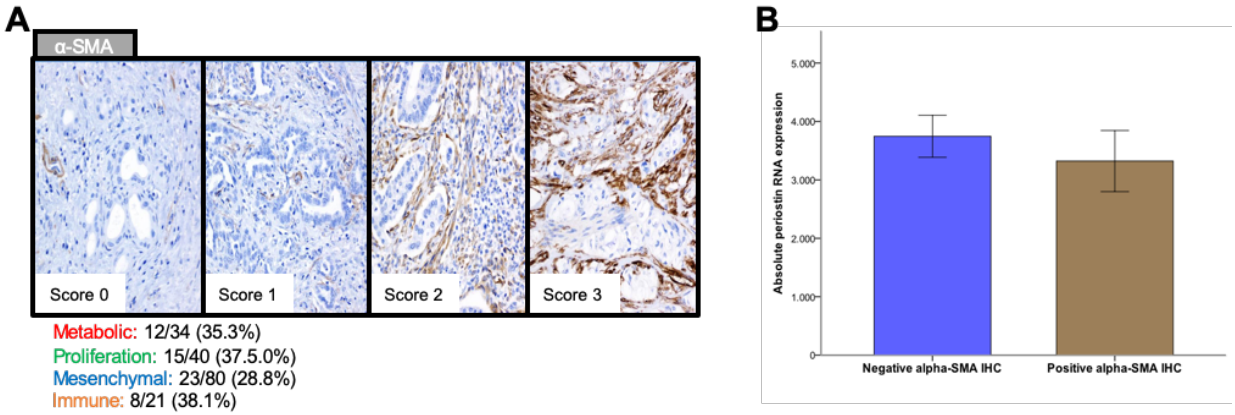
Supplementary Fig. 8. Overexpressed genes defining eCCA molecular classes. Relative RNA expression of **(A),** Histone deacetylase 6 (*HDAC6*); **(B),** Periostin (*POSTN*); and **(C),** Casein kinase 2 (*CSNK2A1*) in the four molecular eCCA classes in comparison to normal bile duct. P values were calculated using a two-sided T-test. Error bars represent 95% confidence intervals.



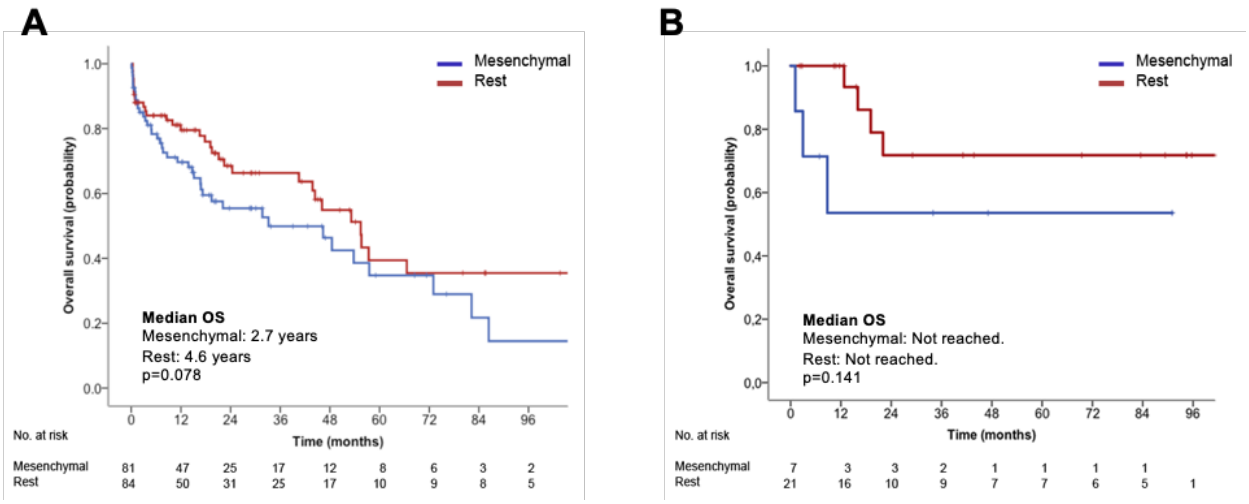
Supplementary Fig. 9. Ki67 staining in eCCA tumors according to their molecular class. Ki67 IHC was conducted in the eCCA cohort (Metabolic=8, Proliferation=33, Mesenchymal=67 and Immune=21). **(A)** Representative samples of each molecular class together with paired CK19 staining. **(B)** Box plots representing Ki67 index for eCCA molecular classes showing the highest percentage of staining in the Proliferation class ($p<0.001$). P values between the Proliferation class and the rest were calculated using T-Test.



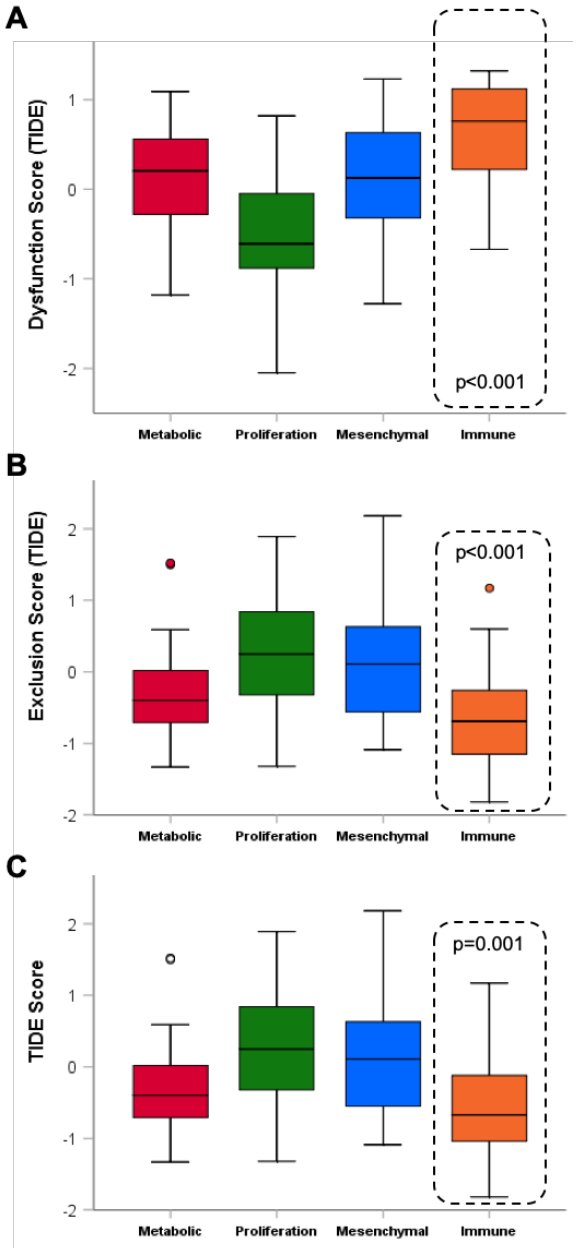
Supplementary Fig. 10. Landscape of structural genomic alterations in eCCA molecular classes. Pathway diagrams showing the percentage of samples from each molecular class with structural genomic alterations in genes from RTK-RAS-PI3K, TP53-RB, histone modification and TGFβ pathways. Red and blue mean alterations leading to activation or inactivation of the gene, respectively.



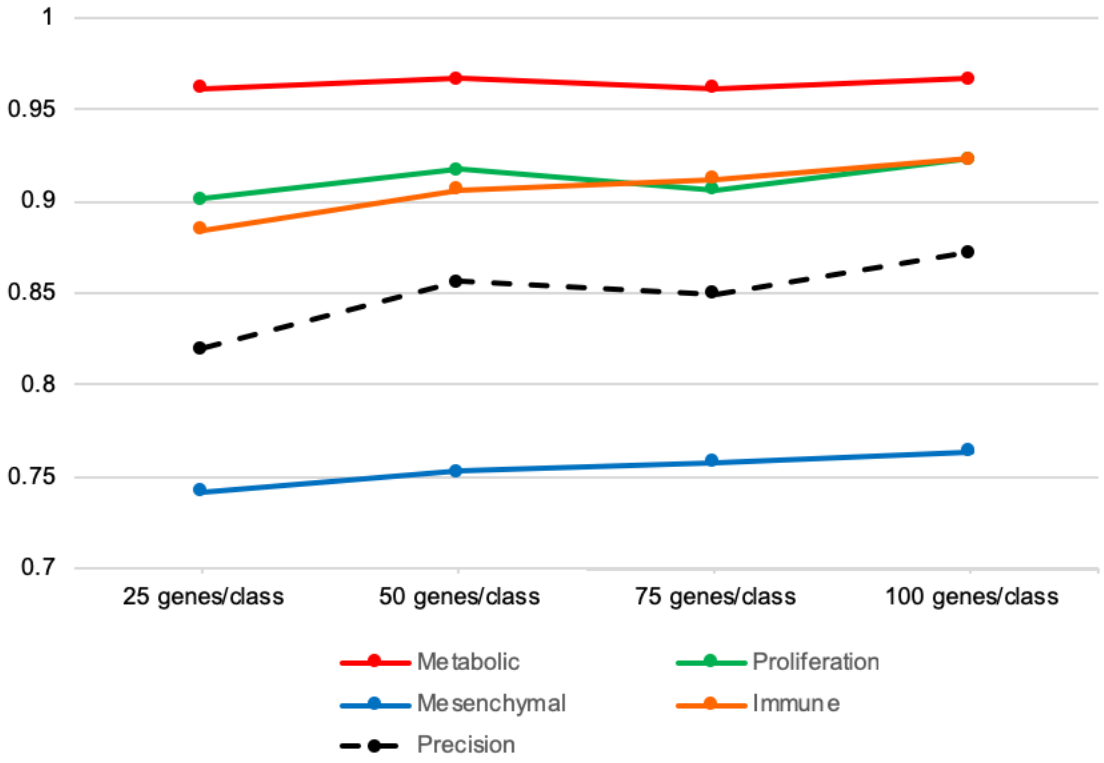
Supplementary Fig. 11. α-SMA and periostin in eCCA molecular classes. (A), Protein expression of α-SMA assessed by IHC. The total number of eCCA samples with a positive staining (≥2 intensity in >50% of fibroblasts) per molecular class is presented. **(B),** Absolute RNA expression of periostin depending on α-SMA positivity. Error bars represent 95% confidence intervals.



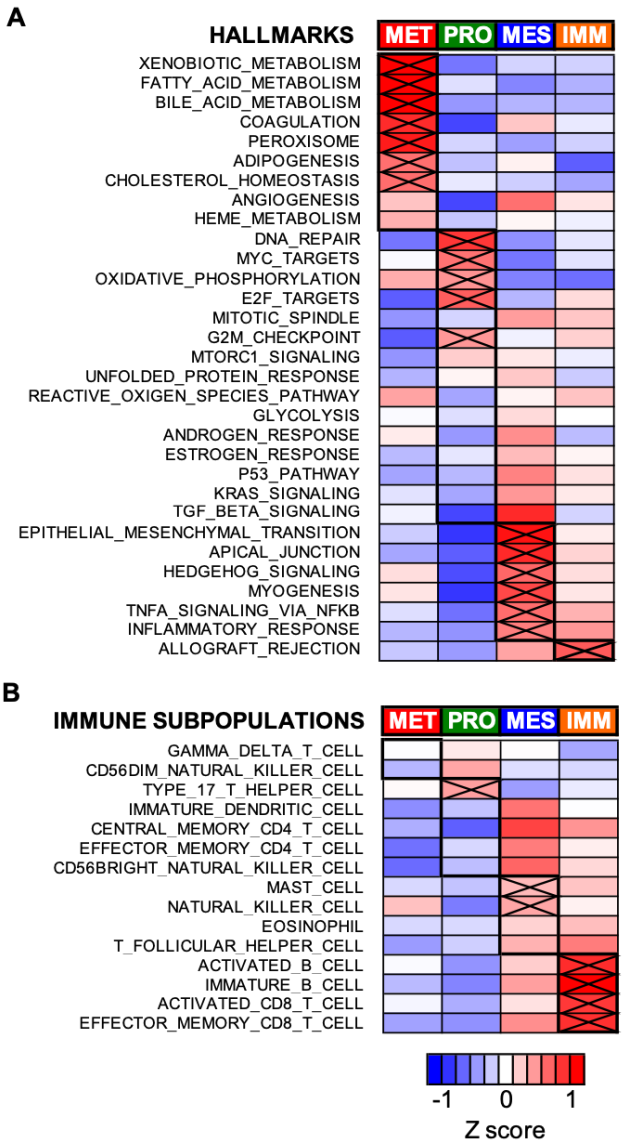
Supplementary Fig. 12. Clinical outcome of mesenchymal eCCA patients. Kaplan-Meier curves comparing OS in the mesenchymal eCCA class vs rest in the: **(A),** internal; and **(B),** external (ICGC) eCCA cohorts. Log-rank test was used to analyze survival data.



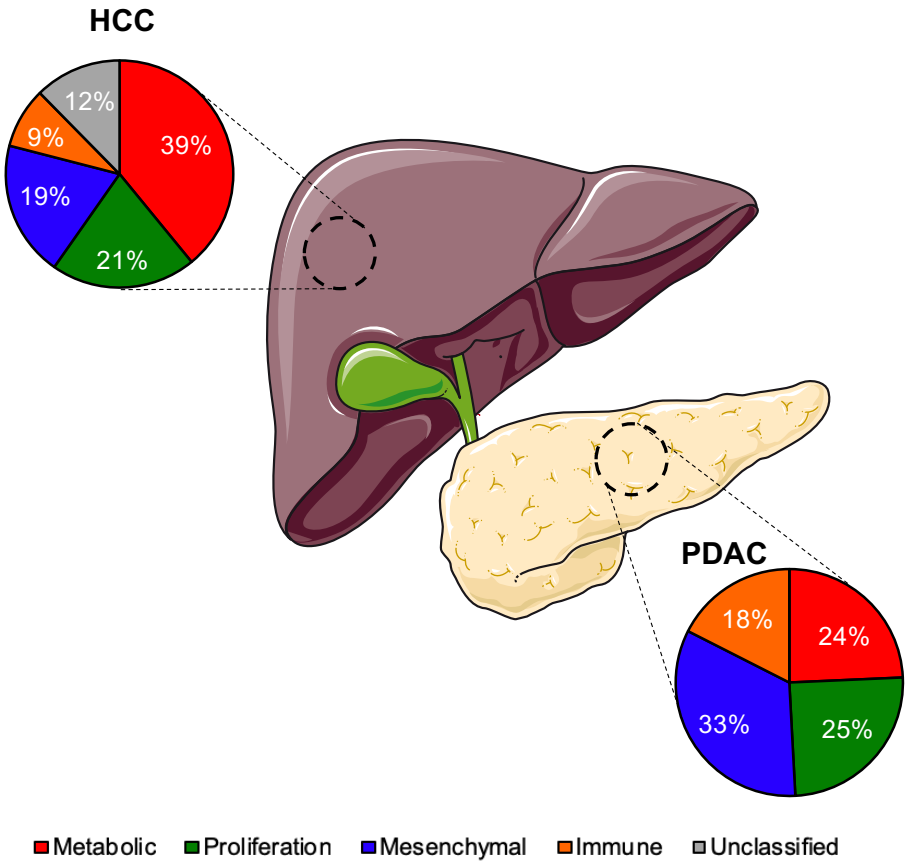
Supplementary Fig. 13. T cell functionality in eCCA molecular classes. Box plots representing the estimation of: **(A)** T cell dysfunction; and **(B)** T cell exclusion in each eCCA molecular class using gene expression data (TIDE software). TIDE score **(C)** merges the weight of the two previous T cell categories in order to predict response to immune checkpoint inhibitors (low scores indicating high probability of clinical benefit). P values between the eCCA Immune class and the rest were calculated using T-Test.



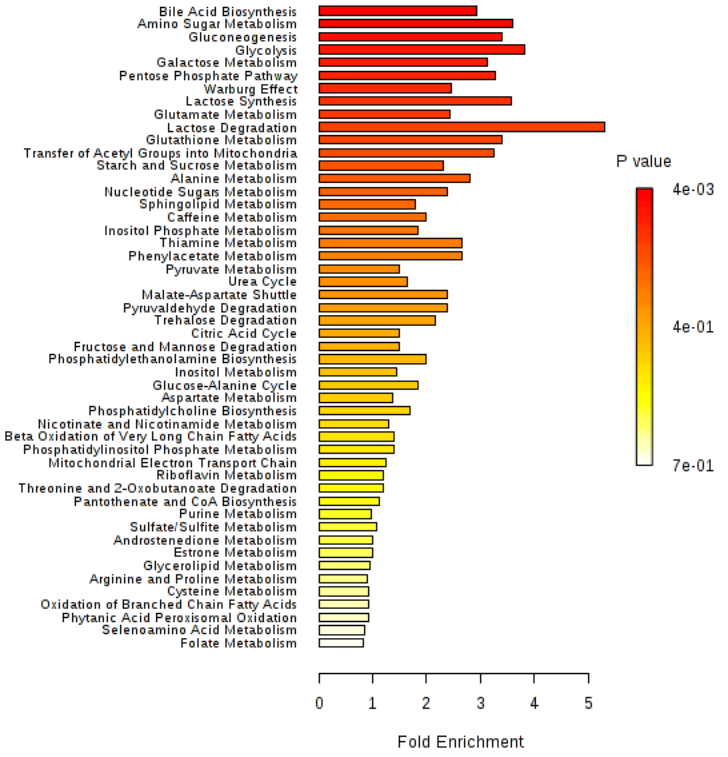
Supplementary Fig. 14. Accuracy of gene-expression eCCA classifiers. Line graph representing the accuracy per class (true positives + true negatives / eCCA samples) of different proposed gene-expression classifiers based on the maximum number of genes used for defining each class (25, 50, 75 and 100). Precision refers to positive predictive value (number of true positives / number of positive calls).



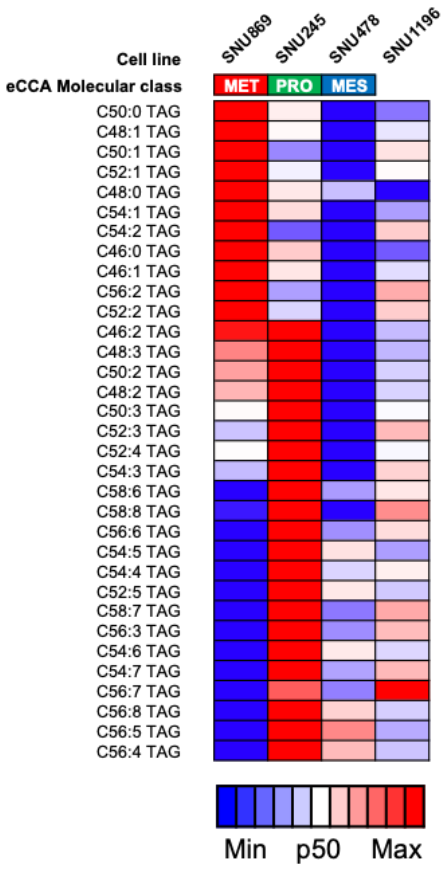
Supplementary Fig. 15. External validation of biological features defining eCCA molecular classes. (A), Heatmap of hallmark gene sets from MSigDB collections in the four molecular classes of eCCA inferred in the external ICGC cohort of CCA. **(B),** Heatmap of immune subpopulations inferred by gene expression of immune metagenes described in The Cancer Immunome Atlas in the four molecular classes of eCCA inferred in the external ICGC cohort of CCA. P values between a specific molecular class and the rest were calculated using T-Test, being crossed cells lower than 0.05.



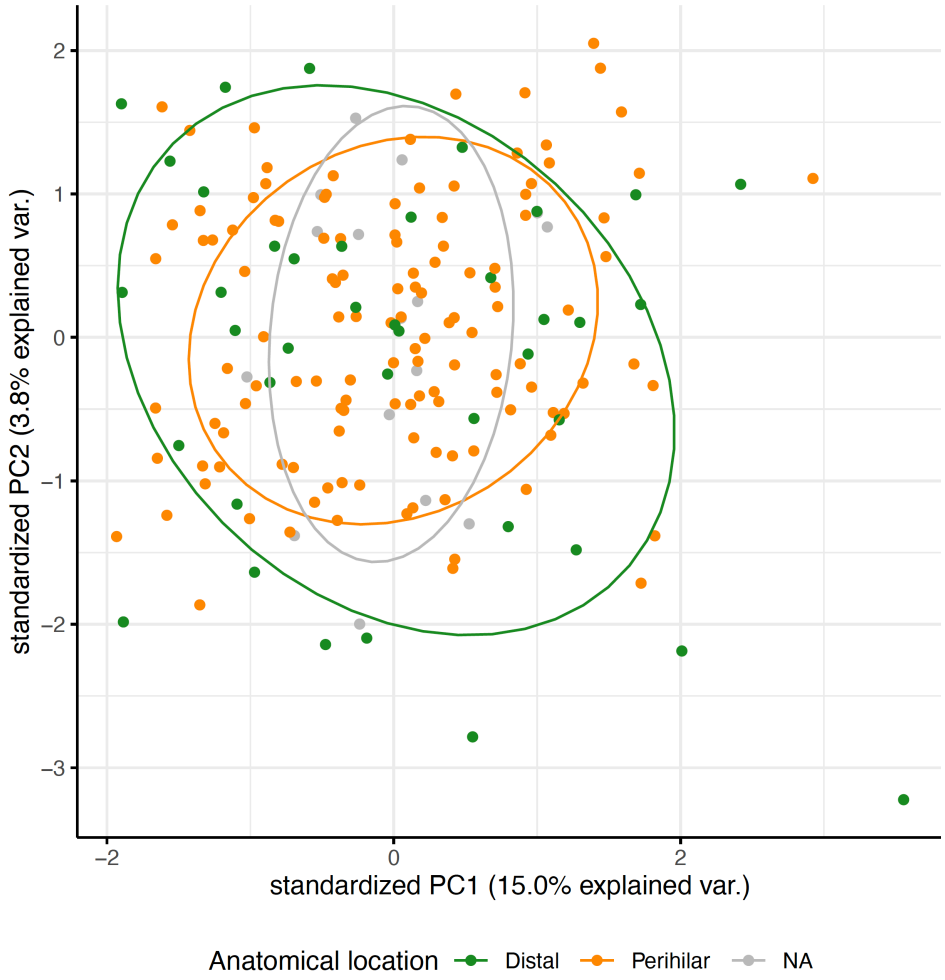
Supplementary Fig. 16. Distribution of eCCA molecular classes in HCC and PDAC. Prediction of eCCA molecular classes in samples from HCC-HEPTROMIC (n=228)[30], HCC-TCGA (n=362)[31] and PDAC-TCGA (n=177)[32] projects applying the 174-gene classifier. HCC: Hepatocellular carcinoma; PDAC: Pancreatic ductal adenocarcinoma.



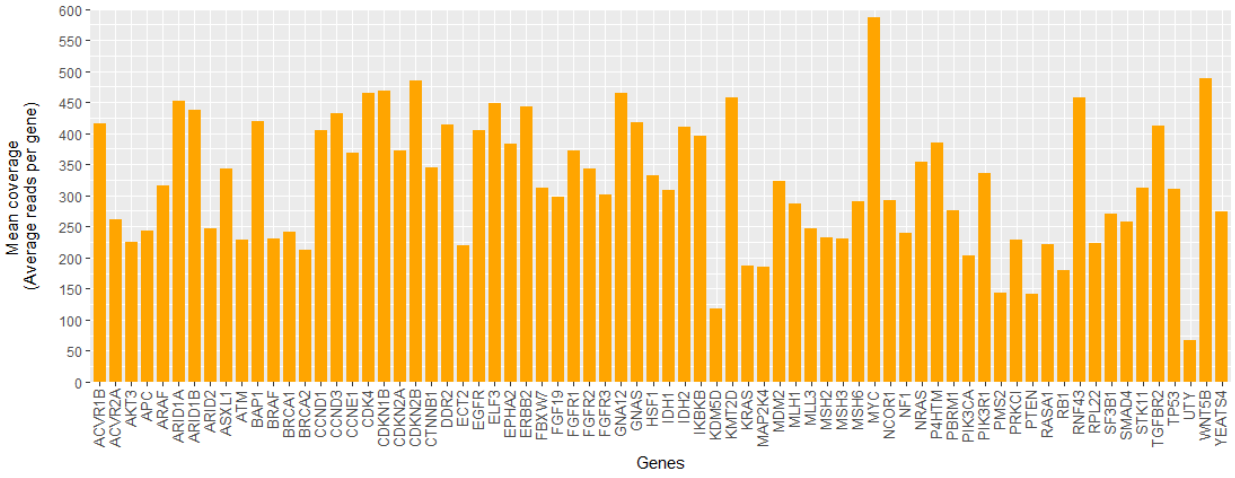
Supplementary Fig. 17. Metabolite enrichment analysis of TIGER-LC Metabolic-like tumors. Of the 140 liver tumors (HCC and iCCA) with paired transcriptome-metabolome in the TIGER-LC Consortium, 50 (36%) had a transcriptomic profile resembling the eCCA Metabolic class. Up to 51 metabolites were significantly more abundant in these tumors in comparison to the rest ($p<0.05$, two-sided T-test). Using Metabolite Set Enrichment Analysis from MetaboAnalyst 4.0[29], a summary plot for Over Representation Analysis (ORA) was obtained. ORA was implemented using the hypergeometric test to evaluate whether a particular metabolite set is represented more than expected by chance within the given compound list. One-tailed p values are provided after adjusting for multiple testing.



Supplementary Fig. 18. Differential abundance of triacylglycerol (TAG) species in eCCA cell lines. Heatmap showing relative levels of triacylglycerol (TAG) species in 4 eCCA cell lines from the Cancer Cell Line Encyclopedia (CCLE). Using the eCCA gene-expression classifier, SNU869, SNU245 and SNU478 resembled Metabolic, Proliferation and Mesenchymal eCCA classes, respectively. SNU869 Metabolic-like cell line presented high levels of monounsaturated fatty acids (MUFA) and low levels of polyunsaturated TAG, a pattern predicted to be sensitive to the loss of CTNNB1[33].



Supplementary Fig. 19. Distribution of pCCA and dCCA samples after a Principal Component Analysis (PCA) with transcriptome data. The data was inputted to the prcomp function from the R package stats v3.6.2 to perform PCA after scaling variables to have unit variance and shifting them to be zero centered. The resulting distribution of the samples according to PC1 and PC2 was visualized using the R package ggbiplot v0.55.



Supplementary Fig. 20. Mean coverage per gene. Average read depth of the 72 genes analyzed by exome-sequencing in 150 eCCA tumoral samples. Genes are sorted alphabetically.

Supplementary Tables:

Supplementary Table 1. Baseline characteristics of eCCA patients according to center of origin.

	eCCA cohort n=189	Mount Sinai n=50	Mayo n=28	Johns Hopkins n=15	Lenox Hill n=9	America n=102	Clinic n=39	CHUV n=36	Vall Hebron n=12	Europe n=87	America vs Europe p value
Gender, n (%) [1]											
Male	117 (67)	35 (70)	20 (71)	NA	5 (56)	60 (69)	28 (72)	22 (61)	7 (58)	57 (66)	0.747
Female	57 (33)	15 (30)	8 (29)	NA	4 (44)	27 (31)	11 (28)	14 (39)	5 (42)	30 (34)	
Age, median (range) [1]	65 (32-86)	64 (35-86)	63 (33-83)	NA	60 (41-81)	63 (32-86)	64 (40-84)	69 (46-86)	69 (49-74)	67 (40-86)	0.034
Race, n (%) [2]											
Caucasian	141 (89)	34 (68)	21 (95)	NA	NA	55 (76)	38 (97)	36 (100)	12 (100)	86 (99)	<0.001
Hispanic	8 (5)	8 (16)	0 (0)	NA	NA	8 (11)	0 (0)	0 (0)	0 (0)	0 (0)	
Asian	8 (5)	3 (6)	1 (5)	NA	NA	8 (11)	0 (0)	0 (0)	0 (0)	0 (0)	
African	2 (1)	1 (2)	0 (0)	NA	NA	1 (1)	1 (3)	0 (0)	0 (0)	1 (1)	
Risk factor, n (%) [3]											
Primary sclerosing cholangitis	6 (4)	4 (8)	NA	NA	NA	4 (8)	2 (5)	0 (0)	0 (0)	2 (2)	0.289
Chronic liver disease (HCV, HBV, alcohol, NASH)	8 (6)	3 (6)	NA	NA	NA	3 (6)	1 (3)	2 (6)	2 (17)	5 (6)	
Clinical presentation, n (%) [4]											
Local signs (jaundice/cholangitis)	115 (80)	43 (86)	20 (87)	NA	NA	63 (86)	22 (92)	22 (63)	8 (73)	52 (74)	NA[15]
Systemic signs (weight loss/anorexia)	43 (30)	0 (0)	14 (61)	NA	NA	14 (19)	5 (21)	20 (57)	4 (36)	29 (41)	
Asymptomatic	9 (6)	3 (6)	0 (0)	NA	NA	3 (4)	2 (8)	2 (6)	2 (18)	6 (9)	
Type of surgery, n (%) [5]											
Hepatic and bile duct resection	126 (78)	44 (88)	19 (70)	NA	NA	63 (82)	34 (87)	20 (56)	9 (90)	63 (74)	NA[15]
Pancreaticoduodenectomy	51 (31)	6 (12)	21 (78)	NA	NA	27 (35)	4 (10)	20 (56)	0 (0)	24 (28)	
Liver transplantation	5 (3)	0 (0)	2 (7)	NA	NA	2 (3)	2 (5)	0 (0)	1 (10)	3 (4)	
Regional lymphadenectomy	113 (70)	43 (86)	22 (81)	NA	NA	65 (84)	30 (77)	13 (36)	5 (50)	48 (56)	
Anatomical subtype, n (%) [6]											
Perihilar	130 (76)	38 (78)	23 (85)	NA	6 (67)	67 (79)	36 (92)	16 (44)	11 (92)	63 (72)	0.377
Distal	42 (24)	11 (22)	4 (15)	NA	3 (33)	18 (21)	3 (8)	20 (56)	1 (8)	24 (28)	
Tumor diameter (mm), median (range) [7]	25 (5-120)	22 (5-70)	25 (14-120)	NA	31 (12-52)	25 (5-120)	28 (15-73)	25 (9-55)	20 (10-60)	25 (9-73)	0.602
TNM stage AJCC 7th ed. (perihilar), n (%) [8]											
I	5 (4)	1 (3)	2 (9)	NA	0 (0)	3 (4)	1 (3.1)	0 (0)	1 (10)	2 (3)	0.262
II	50 (40)	16 (42)	9 (39)	NA	3 (60)	28 (42)	13 (41)	5 (31)	4 (40)	22 (38)	
IIIA	19 (15)	7 (18)	4 (17)	NA	0 (0)	11 (17)	2 (6)	2 (12)	4 (40)	8 (14)	
IIIB	36 (29)	12 (32)	8 (35)	NA	1 (20)	21 (32)	9 (28)	6 (37)	2 (12)	15 (26)	
IVA	5 (4)	0 (0)	0 (0)	NA	1 (20)	1 (1)	3 (9)	1 (6)	2 (12)	4 (7)	
IVB	9 (7)	2 (5)	0 (0)	NA	0 (0)	2 (3)	4 (12)	2 (12)	1 (10)	7 (12)	
TNM stage AJCC 7th ed. (distal), n (%) [8]											
IA	1 (2)	0 (0)	0 (0)	NA	0 (0)	0 (0)	0 (0)	1 (5)	0 (0)	1 (4)	0.408
IB	7 (17)	1 (9)	1 (25)	NA	1 (33)	3 (17)	0 (0)	3 (15)	1 (100)	4 (17)	
IIA	14 (33)	3 (27)	2 (50)	NA	0 (0)	5 (28)	2 (67)	7 (35)	0 (0)	9 (37)	
IIB	17 (40)	7 (64)	1 (25)	NA	2 (67)	10 (56)	1 (33)	6 (30)	0 (0)	7 (29)	
III	2 (5)	0 (0)	0 (0)	NA	0 (0)	0 (0)	0 (0)	2 (10)	0 (0)	2 (8)	
IV	1 (2)	0 (0)	0 (0)	NA	0 (0)	0 (0)	0 (0)	1 (5)	0 (0)	1 (4)	
Resection margins, n (%) [9]											
R0	107 (64)	27 (56)	21 (84)	NA	7 (87)	55 (68)	23 (59)	25 (69)	4 (36)	52 (60)	0.337
R1	60 (36)	421 (44)	4 (14)	NA	1 (12)	26 (32)	16 (41)	11 (31)	7 (64)	34 (39)	
Bilirubin (mg/dl), median (range) [10]	3.3 (0.3-73.4)	3.0 (0.3-40.0)	2.4 (0.6-15.6)	NA	NA	3 (0.3-40.0)	4.8 (0.3-38.7)	3.7 (0.4-73.4)	1.6 (0.5-21.1)	3.7 (0.3-73.4)	0.408
ALT (UI/l), median (range) [11]	99 (17-715)	98 (17-388)	106 (31-344)	NA	NA	101 (17-388)	114 (21-547)	75 (18-682)	97 (34-715)	90 (18-715)	0.453
Albumin (mg/dl), median (range) [12]	37 (22-49)	38 (22-46)	40 (29-45)	NA	NA	38 (22-46)	36 (26-49)	40 (30-44)	33 (26-44)	35 (26-49)	0.290
CA19.9 (UI/ml), median (range) [13]	142 (0-159600)	115 (1-26259)	105 (0-1880)	NA	NA	106 (0-26259)	210 (20-159600)	89 (2-10882)	409 (13-22876)	177 (2-159600)	0.160
CEA (ng/ml), median (range) [14]	2.4 (0-135)	2.2 (0.5-77.5)	1.9 (0-135)	NA	NA	2.1 (0-135)	3.4 (0.1-112.9)	1.8 (0.5-6.9)	3.2 (1.0-7.8)	3.1 (0.1-112.9)	0.992
Precursor lesions, n (%)											
BilIN	52 (28)	26 (52)	2 (7)	4 (27)	0 (0)	32 (31)	10 (26)	10 (28)	0 (0)	20 (25)	0.364
IPNB	18 (9)	4 (8)	1 (4)	1 (7)	2 (22)	8 (8)	4 (10)	5 (14)	1 (8)	10 (11)	
No	119 (63)	20 (40)	23 (89)	10 (67)	9 (78)	62 (61)	25 (64)	21 (58)	11 (92)	57 (65)	
Histological type, n (%)											
Adenocarcinoma	149 (79)	42 (84)	25 (89)	10 (67)	4 (44)	81 (79)	33 (85)	25 (69)	10 (83)	68 (78)	0.681
Mucinous	22 (12)	3 (6)	2 (7)	4 (27)	2 (22)	11 (11)	3 (8)	6 (22)	2 (17)	11 (13)	
Papillary	12 (6)	4 (8)	0 (0)	0 (0)	2 (22)	6 (6)	3 (8)	3 (8)	0 (0)	6 (7)	
Clear cell	2 (1)	0 (0)	0 (0)	1 (7)	0 (0)	1 (1)	0 (0)	1 (3)	0 (0)	1 (1)	
Adenosquamous	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)	0 (0)	1 (1)	
Signet ring cell	2 (1)	1 (2)	1 (4)	0 (0)	0 (0)	2 (2)	0 (0)	0 (0)	0 (0)	0 (0)	
Small cell carcinoma	1 (1)	0 (0)	0 (0)	0 (0)	1 (11)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	
Cell differentiation, n (%)											
G1	27 (14)	10 (20)	2 (7)	4 (27)	2 (22)	18 (18)	6 (15)	3 (8)	0 (0)	9 (10)	0.034
G2	136 (72)	34 (68)	18 (64)	9 (60)	4 (44)	65 (64)	30 (77)	30 (83)	11 (92)	71 (82)	
G3	23 (12)	6 (12)	6 (21)	1 (7)	3 (33)	16 (16)	7 (8)	3 (8)	1 (8)	7 (8)	
G4	3 (2)	0 (0)	2 (7)	1 (7)	0 (0)	3 (3)	0 (0)	0 (0)	0 (0)	0 (0)	
Growth pattern, n (%)											
Periductal	168 (89)	47 (94)	24 (86)	11 (73)	8 (89)	90 (88)	35 (90)	33 (92)	10 (83)	78 (90)	0.308
Exophytic	11 (6)	3 (6)	4 (14)	4 (27)	0 (0)	8 (8)	0 (0)	2 (6)	1 (8)	3 (3)	
Intraductal	10 (5)	0 (0)	0 (0)	0 (0)	1 (11)	4 (4)	4 (10)	1 (3)	1 (8)	6 (7)	
Perineural invasion, n (%)											
Yes	114 (60)	38 (76)	15 (54)	7 (47)	4 (44)	64 (63)	25 (64)	19 (53)	6 (50)	50 (57)	0.551
No	75 (40)	12 (24)	13 (46)	8 (53)	5 (56)	38 (37)	14 (36)	17 (42)	6 (50)	37 (42)	
Vascular invasion, n (%)											
Yes	29 (15)	13 (26)	3 (11)	2 (13)	3 (33)	21 (21)	4 (10)	3 (8)	1 (8)	8 (9)	0.042
No	160 (85)	37 (74)	25 (89)	13 (87)	6 (67)	81 (79)	35 (90)	33 (92)	11 (92)	79 (91)	

Clinical annotation data for all patients and samples included in this study. Associations between categorical variables were analyzed by Fisher's exact test. T-test was used for the comparison of categorical and continuous variables. HCV: Hepatitis C virus; HBV: Hepatitis B virus; NASH: Non-alcoholic steatohepatitis; BilIN: Biliary intraepithelial neoplasia; IPNB: intraductal papillary neoplasm of the bile duct. Non available in [1](15),

[2](30), [3](52), [4](46), [5](27), [6](17), [7](31), [8](21), [9](22), [10](33), [11](34), [12](62), [13](78) and [14](95) patients. [15]More than one possible.

Supplementary Table 2. Genes included in the targeted DNA-sequencing panel.

Gene	Alteration type	BTC (%)	iCCA (%)	eCCA (%)
ACVR1B	Mut	2%	1%	2%
ACVR2A	Mut	3-4%	3%	6%
AKT3	Amp	3%		
APC	Mut	5-7%	5%	6-11%
ARAF	Mut	5%	5%	
ARID1A	Mut	11-17%	15-19%	7-15%
ARID1B	Mut	4%	2%	7%
ARID2	Mut	5-6%	4%	5-8%
ASXL1	Mut	2-3%	2%	3-4%
ATM	Mut	4%		6%
BAP1	Mut	8%	11-12%	3-4%
BRAF	Mut	3%	3%	2%
BRCA1	Mut	1%		1%
BRCA2	Mut	4-5%	3%	4%
CCND1	Amp	3%	3%	5%
CCND3	Amp	3%	3%	
CCNE1	Amp	2%	0%	5%
CDK4	Amp	5%		
CDKN1B	Mut	1%	1%	1%
CDKN2A	Del	5%	4%	5-28%
CDKN2B	Del			15%
CTNNB1	Mut	1%	1%	2%
DDR2	Amp	4%		
ECT2	Amp	3%		
EGFR	Amp	5%	4%	1%
ELF3	Mut	5-6%	4%	10%
EPHA2	Mut	5%	10%	0%
ERBB2	Mut/Amp	5%	5%	4-9%
FBXW7	Mut	3-4%		4-5%
FGF19	Amp	3%		5%
FGFR1	Amp	4%	7%	0%
FGFR2	Amp/Fus/Mut	5%	8%	0%
FGFR3	Amp	4%	5%	3%
GNA12	Amp	3%		
GNAS	Mut	6%		5%
HSF1	Amp	4%		
IDH1	Mut	3%	4-6%	0-2%
IDH2	Mut	1%	2%	1%
IKBKB	Amp	3%		
KDM5D	Del	2%	2%	4%
KMT2D	Mut	7%	7%	5%
KRAS	Mut	17-18%	20-25%	10-43%
MAP2K4	Mut	2%	2%	2%
MDM2	Amp	5%	5%	5%
MLH1*	Mut			
MLL3	Mut	3%	3%	4%
MSH2*	Mut			
MSH3	Mut	2%	1%	4%
MSH6*	Mut			
MYC	Amp	5%	6%	4-6%
NCOR1	Mut	2%	2%	2%
NF1	Mut	5%	5%	5-6%
NRAS	Mut	2-3%	3-5%	1-5%
P4HTM	Mut	1%	1%	1%
PBRM1	Mut	5-7%	8%	0-4%
PIK3CA	Mut	7%	8%	4%
PIK3R1	Mut	3%	2%	3-4%
PMS2*	Mut			
PRKCI	Amp	4%		
PTEN	Mut	3%	1%	5-7%
RASA1	Mut	4%	4%	4%
RB1	Mut	3%	2%	4%
RNF43	Mut	5%	4%	5%
RPL22	Mut	3%		
SF3B1	Mut	5%	4%	6%
SMAD4	Mut	9-13%	10-11%	10-16%
STK11	Mut	3-5%	4%	7%
TERT	Mut (promoter)	3%		
TGFBR2	Mut	2-3%	2%	3%
TP53	Mut/Del	26-32%	23-30%	26-45%
UTY	Del	2%	2%	3%
WNT5B	Amp	4%		
YEATS4	Amp	5%		

Structural genomic alterations recurrently observed in BTC and specifically in eCCA based on bibliography[3,34–37]. *Due to clinical relevance, MMR genes were included despite being rarely mutated in BTC.

Supplementary Table 3. Driver mutations identified in eCCA.

Sample	Class	Gene	Allele frequency	Role	Protein	Consequence	Driver
CC129	Metabolic	*KRAS*	0.18	Act	p.G12D	Missense	known
CC129	Metabolic	*SMAD4*	0.27	LoF	p.C363F	Missense	predicted
CC129	Metabolic	*TP53*	0.19	LoF	p.R337C	Missense	known
CC131	Proliferation	*IDH1*	0.20	Act	p.R132C	Missense	known
CC133	Mesenchymal	*CDKN2A*	0.26	LoF	p.D74Y	Missense	known
CC133	Mesenchymal	*EPHA2*	0.19	LoF	p.V152Dfs*12	Frameshift	predicted
CC133	Mesenchymal	*TP53*	0.14	LoF	p.V73Rfs*76	Frameshift	known
CC135	Metabolic	*ERBB2*	0.50	Act	p.I654V	Missense	known
CC135	Metabolic	*KRAS*	0.08	Act	p.G12V	Missense	known
CC135	Metabolic	*TP53*	0.10	LoF	p.R248W	Missense	known
CC139	Metabolic	*KRAS*	0.08	Act	p.G12R	Missense	known
CC139	Metabolic	*SMAD4*	0.11	LoF	p.R361H	Missense	known
CC139	Metabolic	*TP53*	0.12	LoF	p.R273S	Missense	known
CC141	Mesenchymal	*CDKN2A*	0.12	LoF	p.W110*	Nonsense	known
CC141	Mesenchymal	*KRAS*	0.20	Act	p.G12V	Missense	known
CC141	Mesenchymal	*TP53*	0.05	LoF	p.C242Afs*5	Frameshift	predicted
CC145	Mesenchymal	*CDKN2A*	0.12	LoF	p.P48L	Missense	known
CC145	Mesenchymal	*KRAS*	0.25	Act	p.G12V	Missense	known
CC145	Mesenchymal	*TP53*	0.13	LoF	p.R248W	Missense	known
CC147	Immune	*EPHA2*	0.44	LoF	p.G391R	Missense	known
CC147	Immune	*KRAS*	0.14	Act	p.Q61H	Missense	known
CC147	Immune	*SMAD4*	0.07	LoF	p.V492*fs*1	Frameshift	predicted
CC149	Metabolic	*ATM*	0.55	LoF	p.Y2791Gfs*14	Frameshift	predicted
CC149	Metabolic	*KRAS*	0.22	Act	p.E62V	Missense	predicted
CC149	Metabolic	*KRAS*	0.22	Act	p.Q61R	Missense	known
CC151	Proliferation	*ERBB2*	0.49	Act	p.I654V	Missense	known
CC151	Proliferation	*TP53*	0.09	LoF	p.L35Cfs*9	Frameshift	predicted
CC155	NA	*TGFBR2*	0.15	LoF	.	SpliceDonorInsertion	predicted
CC159	Proliferation	*BAP1*	0.62	LoF	p.Y129*	Nonsense	predicted
CC159	Proliferation	*NRAS*	0.22	Act	p.Q61R	Missense	known
CC159	Proliferation	*PBRM1*	0.58	LoF	.	SpliceDonorSNV	predicted
CC163	Metabolic	*FGFR1*	0.07	Act	p.R501H	Missense	predicted
CC163	Metabolic	*TP53*	0.14	LoF	p.R282W	Missense	known
CC163	Metabolic	*TP53*	0.11	LoF	p.A276Lfs*29	Frameshift	predicted
CC165	Proliferation	*FBXW7*	0.11	LoF	p.R465C	Missense	known
CC165	Proliferation	*TP53*	0.17	LoF	p.R196*	Nonsense	known
CC169	Mesenchymal	*ACVR2A*	0.11	LoF	p.K437Rfs*5	Frameshift	predicted
CC169	Mesenchymal	*MLH1*	0.46	LoF	.	SpliceAcceptorDeletion	known
CC169	Mesenchymal	*MSH3*	0.08	LoF	p.K383Rfs*32	Frameshift	predicted
CC169	Mesenchymal	*RNF43*	0.48	LoF	p.P192L	Missense	predicted
CC169	Mesenchymal	*SF3B1*	0.06	Act	p.R625C	Missense	known
CC173	Proliferation	*ARID1A*	0.07	LoF	p.Q211*	Nonsense	predicted
CC173	Proliferation	*ARID1B*	0.14	LoF	p.D666Rfs*21	Frameshift	predicted
CC173	Proliferation	*BRAF*	0.13	Act	p.V600E	Missense	known
CC177	Proliferation	*ERBB2*	0.07	Act	p.S310F	Missense	known
CC177	Proliferation	*IKBKB*	0.05	Ambiguous	p.R220W	Missense	predicted
CC177	Proliferation	*TP53*	0.07	LoF	p.E171*	Nonsense	predicted
CC187	Immune	*BAP1*	0.10	LoF	p.D11Y	Missense	predicted
CC187	Immune	*KRAS*	0.08	Act	p.G12D	Missense	known
CC189	Immune	*TP53*	0.10	LoF	p.V122Dfs*26	Frameshift	known
CC191	Mesenchymal	*KRAS*	0.14	Act	p.G12D	Missense	known
CC191	Mesenchymal	*PIK3CA*	0.06	Act	p.E707K	Missense	predicted
CC201	Proliferation	*ERBB2*	0.49	Act	p.I654V	Missense	known
CC203	Proliferation	*BRCA1*	0.42	LoF	p.S157*fs*1	Frameshift	known
CC203	Proliferation	*CDKN2A*	0.06	LoF	p.R80*	Nonsense	known
CC203	Proliferation	*KRAS*	0.15	Act	p.G12A	Missense	known
CC203	Proliferation	*SF3B1*	0.07	Act	p.W658C	Missense	predicted
CC205	Mesenchymal	*BAP1*	0.62	LoF	p.P629Tfs*14	Frameshift	predicted
CC205	Mesenchymal	*EPHA2*	0.49	LoF	p.G391R	Missense	known
CC205	Mesenchymal	*FBXW7*	0.20	LoF	p.H540Ifs*16	Frameshift	predicted
CC205	Mesenchymal	*SF3B1*	0.18	Act	p.D894Y	Missense	predicted
CC211	Proliferation	*ATM*	0.05	LoF	p.F897Sfs*2	Frameshift	predicted
CC211	Proliferation	*SMAD4*	0.05	LoF	p.A190Qfs*10	Frameshift	predicted
CC213	Proliferation	*ARID2*	0.13	LoF	p.Q288*	Nonsense	predicted

CC213	Proliferation	*KRAS*	0.11	Act	p.G12D	Missense	known
CC215	Mesenchymal	*ACVR1B*	0.26	LoF	p.P470Q	Missense	predicted
CC215	Mesenchymal	*ACVR1B*	0.10	LoF	p.H358N	Missense	predicted
CC215	Mesenchymal	*APC*	0.37	LoF	p.S1857*	Nonsense	predicted
CC215	Mesenchymal	*ARID1B*	0.23	LoF	p.G1159W	Missense	predicted
CC215	Mesenchymal	*ARID2*	0.36	LoF	p.E194*	Nonsense	predicted
CC215	Mesenchymal	*CCND1*	0.20	Act	p.T286K	Missense	predicted
CC215	Mesenchymal	*FBXW7*	0.08	LoF	p.C386F	Missense	predicted
CC215	Mesenchymal	*FGFR1*	0.16	Act	p.W99L	Missense	predicted
CC215	Mesenchymal	*IDH2*	0.26	Act	p.G325C	Missense	predicted
CC215	Mesenchymal	*MAP2K4*	0.07	LoF	p.E202*	Nonsense	predicted
CC215	Mesenchymal	*NF1*	0.28	LoF	p.S1329*	Nonsense	predicted
CC215	Mesenchymal	*NF1*	0.23	LoF	p.E715*	Nonsense	predicted
CC215	Mesenchymal	*PBRM1*	0.34	LoF	p.G1422C	Missense	predicted
CC215	Mesenchymal	*PIK3CA*	0.10	Act	p.T1052K	Missense	predicted
CC215	Mesenchymal	*RASA1*	0.44	LoF	.	SpliceDonorSNV	predicted
CC215	Mesenchymal	*SF3B1*	0.32	Act	p.G915C	Missense	predicted
CC215	Mesenchymal	*SMAD4*	0.11	LoF	p.G270*	Nonsense	predicted
CC215	Mesenchymal	*TP53*	0.34	LoF	p.V173L	Missense	known
CC217	Mesenchymal	*ARAF*	0.14	Act	p.G322S	Missense	known
CC219	Immune	*KRAS*	0.36	Act	p.G12V	Missense	known
CC221	Mesenchymal	*RNF43*	0.49	LoF	p.G336D	Missense	predicted
CC221	Mesenchymal	*SMAD4*	0.10	LoF	.	SpliceAcceptorSNV	predicted
CC223	Mesenchymal	*ARID2*	0.08	LoF	p.Q917*	Nonsense	predicted
CC223	Mesenchymal	*ELF3*	0.08	LoF	p.F303L	Missense	predicted
CC225	Metabolic	*ARID1A*	0.17	LoF	p.S11Afs*90	Frameshift	predicted
CC225	Metabolic	*BRCA2*	0.48	LoF	p.S1982Rfs*22	Frameshift	known
CC225	Metabolic	*ELF3*	0.18	LoF	p.K304Qfs*167	Frameshift	predicted
CC225	Metabolic	*RNF43*	0.33	LoF	p.L7Sfs*12	Frameshift	predicted
CC227	Mesenchymal	*KRAS*	0.06	Act	p.G12D	Missense	known
CC229	Metabolic	*ARID1A*	0.18	LoF	p.Y1055Cfs*49	Frameshift	predicted
CC229	Metabolic	*KRAS*	0.19	Act	p.G12V	Missense	known
CC229	Metabolic	*TP53*	0.34	LoF	p.R175H	Missense	known
CC233	Proliferation	*KRAS*	0.07	Act	p.G12D	Missense	known
CC233	Proliferation	*TP53*	0.08	LoF	p.H214R	Missense	known
CC237	Immune	*TP53*	0.15	LoF	p.V122Dfs*26	Frameshift	known
CC241	Proliferation	*BAP1*	0.16	LoF	.	SpliceDonorSNV	predicted
CC241	Proliferation	*ELF3*	0.16	LoF	p.Q65Sfs*92	Frameshift	predicted
CC241	Proliferation	*RNF43*	0.47	LoF	p.R221Q	Missense	predicted
CC241	Proliferation	*STK11*	0.22	LoF	p.Q159*	Nonsense	predicted
CC241	Proliferation	*TGFBR2*	0.13	LoF	p.P154Afs*3	Frameshift	predicted
CC243	NA	*EPHA2*	0.44	LoF	p.G391R	Missense	known
CC245	Mesenchymal	*ARID1A*	0.08	LoF	p.S769*	Nonsense	predicted
CC245	Mesenchymal	*ARID1A*	0.06	LoF	p.V2041Gfs*58	Frameshift	predicted
CC245	Mesenchymal	*MLL2*	0.08	LoF	p.P4820Qfs*38	Frameshift	predicted
CC245	Mesenchymal	*RASA1*	0.12	LoF	p.R398*	Nonsense	predicted
CC245	Mesenchymal	*TP53*	0.09	LoF	p.R306*	Nonsense	known
CC247	Proliferation	*EPHA2*	0.50	LoF	p.G391R	Missense	known
CC247	Proliferation	*IDH2*	0.52	Act	p.I98T	Missense	predicted
CC249	NA	*ARID1A*	0.08	LoF	p.Q605*	Nonsense	predicted
CC249	NA	*KRAS*	0.09	Act	p.Q61H	Missense	known
CC255	Metabolic	*ATM*	0.45	LoF	p.L2307F	Missense	predicted
CC255	Metabolic	*TP53*	0.06	LoF	p.R273H	Missense	known
CC257	Metabolic	*BRCA2*	0.55	LoF	p.E1688*	Nonsense	predicted
CC257	Metabolic	*CDKN2A*	0.13	LoF	p.L78Hfs*41	Frameshift	predicted
CC257	Metabolic	*KRAS*	0.16	Act	p.G12V	Missense	known
CC257	Metabolic	*TP53*	0.15	LoF	p.L194R	Missense	predicted
CC259	Metabolic	*KRAS*	0.12	Act	p.G12D	Missense	known
CC261	Immune	*ARID1A*	0.06	LoF	p.P109Efs*3	Frameshift	predicted
CC261	Immune	*KRAS*	0.05	Act	p.G12D	Missense	known
CC261	Immune	*SMAD4*	0.06	LoF	p.R361S	Missense	known
CC265	Metabolic	*ARID1A*	0.06	LoF	p.S1791Qfs*15	Frameshift	predicted
CC265	Metabolic	*IDH2*	0.06	Act	p.R172S	Missense	known
CC267	Mesenchymal	*KRAS*	0.07	Act	p.G12D	Missense	known
CC271	Metabolic	*BRCA2*	0.35	LoF	p.V1610Gfs*4	Frameshift	known
CC275	Metabolic	*CDKN2A*	0.14	LoF	p.Y44*fs*1	Nonsense	predicted

CC275	Metabolic	*TP53*	0.12	LoF	p.R248W	Missense	known
CC299	Metabolic	*KRAS*	0.06	Act	p.G12R	Missense	known
CC301	Mesenchymal	*ATM*	0.06	LoF	p.K1625Tfs*12	Frameshift	predicted
CC301	Mesenchymal	*TGFBR2*	0.07	LoF	p.R520*	Nonsense	predicted
CC315	NA	*NCOR1*	0.06	LoF	p.E179K	Missense	predicted
CC315	NA	*SMAD4*	0.21	LoF	p.E374Sfs*3	Frameshift	predicted
CC315	NA	*TP53*	0.23	LoF	p.L130V	Missense	predicted
CC317	Mesenchymal	*ACVR1B*	0.08	LoF	p.R485*	Nonsense	predicted
CC317	Mesenchymal	*ARID1A*	0.11	LoF	.	SpliceDonorSNV	predicted
CC323	Mesenchymal	*CTNNB1*	0.21	Act	p.S45P	Missense	known
CC323	Mesenchymal	*EPHA2*	0.39	LoF	p.T511M	Missense	predicted
CC323	Mesenchymal	*FBXW7*	0.23	LoF	p.Y519C	Missense	predicted
CC323	Mesenchymal	*RB1*	0.19	LoF	p.Q257*	Nonsense	predicted
CC323	Mesenchymal	*RB1*	0.07	LoF	p.Q344*	Nonsense	predicted
CC323	Mesenchymal	*TP53*	0.19	LoF	p.R248Q	Missense	known
CC333	Metabolic	*ARID2*	0.20	LoF	p.Q1124*	Nonsense	predicted
CC333	Metabolic	*FGFR2*	0.35	Act	p.R165Q	Missense	predicted
CC333	Metabolic	*TP53*	0.15	LoF	p.G117V	Missense	predicted
CC333	Metabolic	*TP53*	0.15	LoF	p.G117W	Missense	predicted
CC333	Metabolic	*TP53*	0.14	LoF	p.P98S	Missense	predicted
CC335	Proliferation	*KRAS*	0.08	Act	p.G12D	Missense	known
CC337	Proliferation	*APC*	0.42	LoF	p.T1556Nfs*3	Frameshift	predicted
CC337	Proliferation	*KRAS*	0.34	Act	p.G12V	Missense	known
CC337	Proliferation	*SMAD4*	0.43	LoF	p.G386V	Missense	predicted
CC339	Mesenchymal	*KRAS*	0.21	Act	p.G12V	Missense	known
CC339	Mesenchymal	*RNF43*	0.09	LoF	p.Q733*	Nonsense	predicted
CC345	Proliferation	*KRAS*	0.12	Act	p.G12D	Missense	known
CC349	Mesenchymal	*SMAD4*	0.24	LoF	p.I525V	Missense	predicted
CC351	Proliferation	*ELF3*	0.10	LoF	.	SpliceAcceptorDeletion	predicted
CC351	Proliferation	*FBXW7*	0.26	LoF	p.R13*	Nonsense	predicted
CC351	Proliferation	*TP53*	0.22	LoF	p.Y220N	Missense	predicted
CC355	Proliferation	*CDKN2A*	0.25	LoF	p.R58*	Nonsense	known
CC355	Proliferation	*MSH6*	0.41	LoF	p.D1255Y	Missense	predicted
CC357	Mesenchymal	*ASXL1*	0.18	LoF	p.G645Vfs*58	Frameshift	predicted
CC357	Mesenchymal	*PBRM1*	0.07	LoF	p.R58*	Nonsense	predicted
CC359	Metabolic	*KRAS*	0.22	Act	p.G12V	Missense	known
CC359	Metabolic	*TP53*	0.12	LoF	p.R175H	Missense	known
CC361	Metabolic	*EPHA2*	0.21	LoF	p.W801*	Nonsense	predicted
CC361	Metabolic	*NCOR1*	0.32	LoF	p.R1229Q	Missense	predicted
CC361	Metabolic	*PIK3CA*	0.11	Act	p.E707K	Missense	predicted
CC363	Metabolic	*RNF43*	0.42	LoF	p.S419Tfs*25	Frameshift	predicted
CC365	Metabolic	*IDH2*	0.29	Act	p.R172K	Missense	known
CC367	Mesenchymal	*KRAS*	0.10	Act	p.G12A	Missense	known
CC367	Mesenchymal	*SF3B1*	0.06	Act	p.R625C	Missense	known
CC369	Proliferation	*ARID1A*	0.18	LoF	p.Q605*	Nonsense	predicted
CC369	Proliferation	*PBRM1*	0.15	LoF	p.R876C	Missense	predicted
CC369	Proliferation	*TP53*	0.13	LoF	.	SpliceAcceptorSNV	predicted
CC369	Proliferation	*TP53*	0.10	LoF	p.R196*	Nonsense	known
CC371	Mesenchymal	*TP53*	0.17	LoF	p.I254Hfs*10	Frameshift	predicted
CC383	Proliferation	*ATM*	0.28	LoF	p.P2842L	Missense	predicted
CC383	Proliferation	*KRAS*	0.38	Act	p.G12C	Missense	known
CC383	Proliferation	*STK11*	0.48	LoF	p.K44Sfs*7	Frameshift	predicted
CC383	Proliferation	*TP53*	0.49	LoF	p.E285V	Missense	known
CC385	Mesenchymal	*KRAS*	0.17	Act	p.Q61H	Missense	known
CC387	Mesenchymal	*ARID1A*	0.07	LoF	p.S711*	Nonsense	predicted
CC389	Mesenchymal	*KRAS*	0.16	Act	p.G12A	Missense	known
CC391	Mesenchymal	*KRAS*	0.08	Act	p.G12D	Missense	known
CC393	Mesenchymal	*APC*	0.46	LoF	p.R2505Q	Missense	predicted
CC411	Proliferation	*ARID1A*	0.27	LoF	p.M923Hfs*13	Frameshift	predicted
CC411	Proliferation	*ASXL1*	0.29	LoF	p.G645Vfs*58	Frameshift	predicted
CC411	Proliferation	*BAP1*	0.32	LoF	p.E200K	Missense	predicted
CC411	Proliferation	*BAP1*	0.30	LoF	p.Q280*	Nonsense	predicted
CC411	Proliferation	*EGFR*	0.23	Act	p.R451C	Missense	known
CC411	Proliferation	*EPHA2*	0.47	LoF	p.G391R	Missense	known
CC411	Proliferation	*ERBB2*	0.34	Act	p.V842I	Missense	known
CC411	Proliferation	*GNAS*	0.38	Act	p.R201H	Missense	known

CC411	Proliferation	*KRAS*	0.29	Act	p.D33E	Missense	known
CC411	Proliferation	*MSH2*	0.43	LoF	p.L414Hfs*24	Frameshift	predicted
CC411	Proliferation	*MSH2*	0.16	LoF	p.G174Mfs*33	Frameshift	predicted
CC411	Proliferation	*MSH3*	0.66	LoF	p.K383Rfs*32	Frameshift	predicted
CC411	Proliferation	*SF3B1*	0.47	Act	p.G640S	Missense	predicted
CC411	Proliferation	*SMAD4*	0.61	LoF	p.R361H	Missense	known
CC435	Metabolic	*ARID1B*	0.27	LoF	p.H1387Tfs*61	Frameshift	predicted
CC435	Metabolic	*CDKN2A*	0.23	LoF	p.E69Afs*50	Frameshift	predicted
CC435	Metabolic	*ELF3*	0.11	LoF	p.R339W	Missense	predicted
CC435	Metabolic	*EPHA2*	0.18	LoF	p.Q352Sfs*29	Frameshift	predicted
CC441	Proliferation	*ACVR2A*	0.18	LoF	p.K437Rfs*5	Frameshift	predicted
CC441	Proliferation	*KRAS*	0.25	Act	p.Q61H	Missense	known
CC445	Proliferation	*KRAS*	0.30	Act	p.G12C	Missense	known
CC445	Proliferation	*TP53*	0.17	LoF	p.I232N	Missense	predicted
CC451	Mesenchymal	*CDKN2A*	0.10	LoF	p.G111Pfs*8	Frameshift	predicted
CC451	Mesenchymal	*TP53*	0.15	LoF	p.E285K	Missense	known
CC453	Mesenchymal	*KRAS*	0.14	Act	p.G12D	Missense	known
CC453	Mesenchymal	*NF1*	0.26	LoF	.	SpliceAcceptorInsertion	predicted
CC459	Mesenchymal	*PIK3CA*	0.06	Act	p.E707K	Missense	predicted
CC461	Mesenchymal	*EPHA2*	0.08	LoF	p.A316Gfs*65	Frameshift	predicted
CC465	Mesenchymal	*ACVR2A*	0.41	LoF	p.K437Rfs*5	Frameshift	predicted
CC465	Mesenchymal	*ARID1A*	0.19	LoF	p.D1850Tfs*33	Frameshift	predicted
CC465	Mesenchymal	*ARID1A*	0.18	LoF	p.S366Afs*25	Frameshift	predicted
CC465	Mesenchymal	*ATM*	0.20	LoF	.	SpliceDonorSNV	predicted
CC465	Mesenchymal	*CDKN2A*	0.40	LoF	p.R80*	Nonsense	known
CC465	Mesenchymal	*TGFBR2*	0.25	LoF	p.K153Afs*3	Frameshift	predicted
CC465	Mesenchymal	*TP53*	0.34	LoF	p.R213*	Nonsense	known
CC469	Metabolic	*KRAS*	0.22	Act	p.G12V	Missense	known
CC469	Metabolic	*TP53*	0.22	LoF	p.R248W	Missense	known
CC473	Mesenchymal	*KRAS*	0.17	Act	p.G12D	Missense	known
CC475	Mesenchymal	*CDKN1B*	0.11	LoF	p.A115Gfs*9	Frameshift	predicted
CC475	Mesenchymal	*CDKN1B*	0.10	LoF	p.S112Tfs*11	Frameshift	predicted
CC475	Mesenchymal	*NF1*	0.13	LoF	p.E649Dfs*39	Frameshift	predicted
CC475	Mesenchymal	*TP53*	0.07	LoF	p.G245D	Missense	known
CC477	Mesenchymal	*KRAS*	0.33	Act	p.G12A	Missense	known
CC477	Mesenchymal	*NF1*	0.55	LoF	.	SpliceDonorSNV	predicted
CC477	Mesenchymal	*SMAD4*	0.29	LoF	p.R361H	Missense	known
CC485	Metabolic	*BRAF*	0.08	Act	p.D594G	Missense	known
CC485	Metabolic	*TP53*	0.12	LoF	p.R248Q	Missense	known
CC489	Proliferation	*SF3B1*	0.13	Act	p.K700E	Missense	known
CC489	Proliferation	*SMAD4*	0.06	LoF	p.A406T	Missense	predicted
CC491	Mesenchymal	*ASXL1*	0.12	LoF	p.E518*	Nonsense	predicted
CC491	Mesenchymal	*BAP1*	0.22	LoF	p.R237C	Missense	predicted
CC491	Mesenchymal	*KRAS*	0.29	Act	p.G12A	Missense	known
CC495	Metabolic	*ARID1A*	0.20	LoF	p.P1115Rfs*7	Frameshift	predicted
CC495	Metabolic	*NF1*	0.14	LoF	p.W561C	Missense	predicted
CC495	Metabolic	*TP53*	0.10	LoF	p.E171G	Missense	predicted
CC499	Immune	*ACVR2A*	0.10	LoF	p.K437Rfs*5	Frameshift	predicted
CC499	Immune	*ARAF*	0.10	Act	p.R411W	Missense	predicted
CC499	Immune	*ARID1B*	0.06	LoF	p.R1519H	Missense	predicted
CC499	Immune	*EPHA2*	0.09	LoF	p.P460Rfs*33	Frameshift	predicted
CC499	Immune	*KRAS*	0.08	Act	p.V14I	Missense	known
CC499	Immune	*MSH2*	0.20	LoF	p.V606Sfs*29	Frameshift	predicted
CC499	Immune	*MSH6*	0.07	LoF	p.F1088Sfs*2	Frameshift	known
CC499	Immune	*RNF43*	0.10	LoF	p.G659Vfs*41	Frameshift	predicted
CC499	Immune	*RNF43*	0.07	LoF	p.R113*	Nonsense	predicted
CC499	Immune	*TP53*	0.06	LoF	p.R273C	Missense	known
CC501	Mesenchymal	*ARID1A*	0.38	LoF	p.S334*	Nonsense	predicted
CC501	Mesenchymal	*ARID1A*	0.28	LoF	p.E1531*	Nonsense	predicted
CC501	Mesenchymal	*ARID1B*	0.32	LoF	p.R1102*	Nonsense	predicted
CC501	Mesenchymal	*ARID2*	0.56	LoF	p.E98K	Missense	predicted
CC501	Mesenchymal	*CDKN2A*	0.45	LoF	p.D84N	Missense	known
CC501	Mesenchymal	*ELF3*	0.37	LoF	p.Y352Sfs*96	Frameshift	predicted
CC501	Mesenchymal	*SMAD4*	0.34	LoF	p.I326Lfs*10	Frameshift	predicted
CC501	Mesenchymal	*TP53*	0.60	LoF	p.R110H	Missense	known
CC503	Mesenchymal	*DDR2*	0.07	Act	p.R165W	Missense	predicted

CC503	Mesenchymal	*ERBB2*	0.46	Act	p.I654V	Missense	known
CC503	Mesenchymal	*TP53*	0.06	LoF	p.C242Y	Missense	known
CC507	Mesenchymal	*BRAF*	0.13	Act	p.R726H	Missense	predicted
CC507	Mesenchymal	*CDKN2A*	0.31	LoF	p.R80*	Nonsense	known
CC509	Proliferation	*KRAS*	0.40	Act	p.G12D	Missense	known
CC509	Proliferation	*PIK3CA*	0.70	Act	p.E545G	Missense	known
CC515	Metabolic	*ARID1A*	0.09	LoF	p.F1859Lfs*40	Frameshift	predicted
CC525	Mesenchymal	*ELF3*	0.69	LoF	p.M49Nfs*43	Frameshift	predicted
CC525	Mesenchymal	*KRAS*	0.20	Act	p.G12S	Missense	known
CC525	Mesenchymal	*RB1*	0.13	LoF	p.R661W	Missense	known
CC525	Mesenchymal	*TP53*	0.21	LoF	p.Q144Afs*4	Frameshift	predicted
CC527	Mesenchymal	*KRAS*	0.06	Act	p.G12V	Missense	known
CC529	Immune	*EPHA2*	0.14	LoF	p.E403*	Nonsense	predicted
CC529	Immune	*KRAS*	0.08	Act	p.G12V	Missense	known
CC529	Immune	*TGFBR2*	0.10	LoF	p.P154Afs*3	Frameshift	predicted
CC529	Immune	*TP53*	0.11	LoF	p.G105R	Missense	predicted
CC531	Mesenchymal	*IDH1*	0.08	Act	p.R132H	Missense	known
CC533	Immune	*IDH1*	0.47	Act	p.R314G	Missense	predicted
CC533	Immune	*KRAS*	0.16	Act	p.G12V	Missense	known
CC533	Immune	*STK11*	0.17	LoF	p.D194V	Missense	known
CC535	Proliferation	*TP53*	0.07	LoF	p.R273H	Missense	known
CC539	Proliferation	*ARID1A*	0.05	LoF	p.H688Afs*129	Frameshift	predicted
CC541	Proliferation	*ACVR2A*	0.48	LoF	p.K437Rfs*5	Frameshift	predicted
CC541	Proliferation	*APC*	0.29	LoF	p.Q1529*	Nonsense	predicted
CC541	Proliferation	*ARID2*	0.06	LoF	p.R143C	Missense	predicted
CC541	Proliferation	*ELF3*	0.17	LoF	.	SpliceDonorDeletion	predicted
CC541	Proliferation	*FBXW7*	0.21	LoF	p.S668Vfs*39	Frameshift	predicted
CC541	Proliferation	*KRAS*	0.45	Act	p.G12D	Missense	known
CC541	Proliferation	*MSH2*	0.39	LoF	p.G548C	Missense	predicted
CC541	Proliferation	*MSH6*	0.06	LoF	p.F1088Lfs*5	Frameshift	known
CC543	Mesenchymal	*ARID1A*	0.13	LoF	p.K997Nfs*42	Frameshift	predicted
CC545	Metabolic	*ARID1A*	0.15	LoF	p.Q1188*	Nonsense	predicted
CC545	Metabolic	*ELF3*	0.39	LoF	p.S330R	Missense	predicted
CC545	Metabolic	*KRAS*	0.14	Act	p.K5N	Missense	known
CC547	Proliferation	*KRAS*	0.10	Act	p.G12V	Missense	known
CC547	Proliferation	*TP53*	0.18	LoF	p.R280T	Missense	known
CC549	Immune	*NF1*	0.05	LoF	p.A2321Cfs*5	Frameshift	predicted
CC549	Immune	*SF3B1*	0.07	Act	p.K700E	Missense	known
CC549	Immune	*SMAD4*	0.08	LoF	p.G89Dfs*5	Frameshift	predicted
CC553	Immune	*TP53*	0.25	LoF	p.G266R	Missense	known
CC555	Immune	*TP53*	0.10	LoF	p.S241F	Missense	known
CC557	Mesenchymal	*TP53*	0.26	LoF	p.C277Lfs*67	Frameshift	predicted
CC561	Mesenchymal	*EPHA2*	0.43	LoF	p.G391R	Missense	known
CC561	Mesenchymal	*KRAS*	0.17	Act	p.G12R	Missense	known
CC561	Mesenchymal	*TGFBR2*	0.08	LoF	p.Y495*	Nonsense	predicted
CC561	Mesenchymal	*TGFBR2*	0.08	LoF	p.K493Ifs*2	Frameshift	predicted
CC561	Mesenchymal	*TGFBR2*	0.07	LoF	p.Y495S	Missense	predicted
CC565	Proliferation	*KRAS*	0.48	Act	p.G12D	Missense	known
CC565	Proliferation	*SMAD4*	0.73	LoF	p.D351H	Missense	known
CC567	Immune	*CDKN2A*	0.24	LoF	p.Y129*	Nonsense	predicted
CC567	Immune	*TP53*	0.16	LoF	p.A138P	Missense	known
CC569	Proliferation	*CDKN2A*	0.40	LoF	p.E69*	Nonsense	known
CC569	Proliferation	*EPHA2*	0.52	LoF	p.G391R	Missense	known
CC569	Proliferation	*TP53*	0.29	LoF	p.R280K	Missense	known
CC571	Proliferation	*ARID1A*	0.06	LoF	p.Q1519Pfs*13	Frameshift	predicted
CC571	Proliferation	*ATM*	0.55	LoF	p.Y2677Lfs*5	Frameshift	predicted
CC571	Proliferation	*NCOR1*	0.52	LoF	p.R1628H	Missense	predicted
CC573	Mesenchymal	*ARID1B*	0.15	LoF	p.R1990*	Nonsense	predicted
CC573	Mesenchymal	*TP53*	0.12	LoF	p.R248W	Missense	known
CC577	Mesenchymal	*PIK3CA*	0.06	Act	p.E707K	Missense	predicted
CC579	Mesenchymal	*KRAS*	0.29	Act	p.Q61H	Missense	known
CC579	Mesenchymal	*PBRM1*	0.32	LoF	p.Q779Pfs*13	Frameshift	predicted
CC579	Mesenchymal	*TP53*	0.25	LoF	p.H179P	Missense	predicted
CC581	Mesenchymal	*CTNNB1*	0.17	Act	p.R565C	Missense	predicted
CC583	Mesenchymal	*KRAS*	0.06	Act	p.G12D	Missense	known
CC583	Mesenchymal	*TP53*	0.11	LoF	p.G245S	Missense	known

CC585	Mesenchymal	*ARID1A*	0.41	LoF	p.Q601*	Nonsense	predicted
CC585	Mesenchymal	*ATM*	0.27	LoF	.	SpliceDonorSNV	predicted
CC585	Mesenchymal	*ATM*	0.19	LoF	p.P2699L	Missense	predicted
CC585	Mesenchymal	*ELF3*	0.27	LoF	p.S266Kfs*35	Frameshift	predicted
CC585	Mesenchymal	*EPHA2*	0.56	LoF	p.E387Sfs*6	Frameshift	predicted
CC585	Mesenchymal	*PIK3CA*	0.10	Act	p.E707K	Missense	predicted
CC585	Mesenchymal	*TP53*	0.45	LoF	p.F134L	Missense	predicted
CC591	Mesenchymal	*KRAS*	0.13	Act	p.G12V	Missense	known
CC591	Mesenchymal	*SF3B1*	0.06	Act	p.K700E	Missense	known
CC593	Mesenchymal	*ARID1A*	0.20	LoF	p.K1072Nfs*21	Frameshift	predicted
CC593	Mesenchymal	*KRAS*	0.17	Act	p.G13D	Missense	known
CC593	Mesenchymal	*MSH6*	0.12	LoF	p.E221*	Nonsense	predicted
CC593	Mesenchymal	*NF1*	0.13	LoF	p.R1968*	Nonsense	predicted
CC593	Mesenchymal	*RNF43*	0.19	LoF	p.G659Vfs*41	Frameshift	predicted
CC593	Mesenchymal	*TP53*	0.14	LoF	p.R248Q	Missense	known
CC595	Mesenchymal	*KRAS*	0.09	Act	p.G12D	Missense	known
CC595	Mesenchymal	*TP53*	0.10	LoF	p.R175H	Missense	known
CC599	Mesenchymal	*FBXW7*	0.12	LoF	p.D607Y	Missense	predicted
CC601	Mesenchymal	*APC*	0.11	LoF	p.T1556Nfs*3	Frameshift	predicted
CC601	Mesenchymal	*ELF3*	0.12	LoF	p.F303Lfs*167	Frameshift	predicted
CC601	Mesenchymal	*PIK3R1*	0.19	LoF	p.T500Dfs*14	Frameshift	predicted
CC601	Mesenchymal	*RB1*	0.33	LoF	p.A74Efs*4	Frameshift	predicted
CC601	Mesenchymal	*TP53*	0.18	LoF	.	SpliceDonorSNV	known
CC603	Immune	*APC*	0.20	LoF	p.Y956*	Nonsense	predicted
CC603	Immune	*APC*	0.14	LoF	p.T1556Nfs*3	Frameshift	predicted
CC603	Immune	*ARID2*	0.13	LoF	p.Q329Pfs*2	Frameshift	predicted
CC603	Immune	*PIK3CA*	0.21	Act	p.E81K	Missense	known
CC603	Immune	*PIK3CA*	0.11	Act	p.N345S	Missense	predicted

Candidate mutations were the ones already known to be oncogenic as well as the predicted drivers in Tier 1 according to Cancer Genome Interpreter.

Supplementary Table 4. Copy number alterations identified in eCCA.

Sample	Class	Gene	Copy number
CC275	Metabolic	*CCND3*	10
CC239	Immune	*CCNE1*	13.5
CC495	Metabolic	*CCNE1*	6
CC525	Mesenchymal	*CCNE1*	18
CC557	Mesenchymal	*CCNE1*	6
CC383	Proliferation	*CD4*	13.5
CC581	Mesenchymal	*CD4*	7
CC567	Immune	*EGFR*	7.5
CC165	Proliferation	*ERBB2*	15.5
CC569	Proliferation	*ERBB2*	32.5
CC155	NA	*MDM2*	6.5
CC171	Mesenchymal	*MDM2*	20.5
CC189	Immune	*MDM2*	16
CC191	Mesenchymal	*MDM2*	11
CC239	Immune	*MDM2*	9
CC563	Proliferation	*MDM2*	12.5
CC581	Mesenchymal	*MDM2*	12
CC383	Proliferation	*MYC*	9.5
CC137	Immune	*YEATS4*	8.5
CC155	NA	*YEATS4*	6.5
CC171	Mesenchymal	*YEATS4*	20.5
CC189	Immune	*YEATS4*	16
CC191	Mesenchymal	*YEATS4*	11
CC237	Immune	*YEATS4*	6
CC239	Immune	*YEATS4*	7
CC563	Proliferation	*YEATS4*	12.5
CC581	Mesenchymal	*YEATS4*	21

Copy number alterations were inferred from targeted DNA-sequencing panel using the multifactor normalization tool ONCOCNV. Focal amplifications were called at segments with ≥6 copies.

Supplementary Table 5. Prognosis of structural genetic alterations.

Structural genetic alteration	Frequency in eCCA (%)	HR	95% CI		p value
			Lower	Upper	
KRAS	36.7	1.298	0.747	2.253	0.355
TP53	34.7	1.723	1.010	2.938	0.046
ARID1A	14	1.503	0.727	3.107	0.271
SMAD4	10.7	1.262	0.567	2.811	0.568
EPHA2	10	0.721	0.275	1.889	0.506
CDKN2A	9.3	0.492	0.153	1.585	0.235
ELF3	7.3	1.815	0.761	4.331	0.179
SF3B1	6	0.357	0.086	1.479	0.156
YEATS4	6	0.209	0.029	1.520	0.122
ATM	5.3	1.335	0.321	5.554	0.691
ERBB2	5.3	0.494	0.151	1.614	0.243
PIK3CA	5.3	2.336	0.994	5.488	0.052
RNF43	5.3	0.934	0.291	2.998	0.908

Cox regression for overall survival (OS) of mutations / amplifications that were present in >5% of eCCA tumors. OS was defined as the time between surgical resection and death of any cause or lost follow-up.

Supplementary Table 6. Targeted therapies and their clinical evidence for the treatment of eCCA.

OncoKB	Drug	Target	Neoplasm (ORR in drugs tested in BTC)	Trial	Biomarker (% in eCCA)
Level 1 (FDA-recognized biomarker predictive of response to an FDA- approved drug in this indication)	Pembrolizumab (Le et al.)	PD-1	CCA (1/1)	Basket	MMR deficiency (2.0%)*
Level 2B (Standard care biomarker predictive of response to an FDA- approved drug in another indication but not standard care for this indication)	Rucaparib/Niraparib/Olaparib/Talazoparib	PARP	Ovarian/Breast	Phase 3	BRCA1/2 mut (2.7%)
	Afatinib/Erlotinib/Gefitinib/Osimertinib/Dacomitinib	EGFR	NSCLC	Phase 3	EGFR mut (0.7%)
	Trastuzumab/Lapatinib/Pertuzumab/Neratinib	HER2	Breast/Gastric	Phase 3	HER2 overexperession (3.9%)
	Abemaciclib/Palbociclib	CDK4/6	Liposarcoma	Phase 2	CDK4 amp (1.3%)
	Enasidenib	IDH2	Leukemia	Phase 1/2	IDH2 mut (2.7%)
Level 3A (Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication, but neither biomarker nor drug is standard care)	Neratinib (Hyman et al.)	HER2/EGFR	BTC (2/9)	Basket	HER2 mut (4.0%)
	EGFR inhibitor (Verlingue et al.)	EGFR	BTC (1/1)	Basket	EGFR amp (0.7%)
	Trastuzumab (Verlingue et al.)	HER2	GBC (1/1)	Basket	HER2 amp (1.3%)
	Ivosidenib (Lowery et al.)	IDH1	CCA (4/73)	Phase 1	IDH1 mut (2.0%)
	Vemurafenib (Hyman et al.)	BRAF	CCA (1/8)	Basket	BRAF mutation (2.0%)
Level 3B (Compelling clinical evidence supports the biomarker as being predictive of response to a drug in another indication, but neither biomarker nor drug is	Binimetinib	MEK	Melanoma	Phase 3	NRAS mut (0.7%)
	Alpelisib, Buparlisib, Copanlisib, GDC-0077, Serabelisib, Taselisib	PIK3CA	Breast	Phase 1	PIK3CA mut (5.3%)
	RG7112, DS-3032b	MDM2	Liposarcoma	Phase 1/2	MDM2 amp (4.7%)

Drug/biomarker pairs were categorized according to the OncoKB curated precision oncology knowledge base. The prevalence of each biomarker is based on the present study. *Evaluated in samples with mutations in MMR genes. NSCLC: Non-small cell lung carcinoma.

Supplementary Table 7. Risk factors and eCCA molecular classes.

	eCCA molecular class			
Risk factor, n	Metabolic	Proliferation	Mesenchymal	Immune
Alcohol	2	0	1	0
Hepatitis B	0	1	1	1
Hepatitis C	0	0	1	0
NASH	0	0	1	0
PSC	1	1	4	0

Distribution of known eCCA risk factors in each molecular class.

Supplementary Table 8. Upstream regulators of the transcriptome-based eCCA molecular classes.

	Upstream Regulator	Molecule Type	Activation z-score	p value of overlap
Metabolic	HNF4A	transcription regulator	6.125	1.24E-38
	HNF1A	transcription regulator	5.712	6.15E-37
	CEBPA	transcription regulator	3.814	5.11E-14
	PXR ligand-PXR-Retinoic acid-RXRα	complex	3.812	7.06E-13
	PPARG	ligand-dependent nuclear receptor	3.749	1.5E-11
	PPARGC1A	transcription regulator	3.709	6.56E-11
	PKD1	ion channel	3.362	0.00000175
	FXR ligand-FXR-Retinoic acid-RXRα	complex	3.062	1.35E-13
	Ncoa-Nr1i3-Rxra	complex	3	1.36E-09
	FOXA2	transcription regulator	2.996	1.53E-13
Proliferation	MYCN	transcription regulator	6.14	7.67E-12
	MYC	transcription regulator	5.087	0.000224
	EIF4E	translation regulator	4.519	0.0219
	GAST	other	3.467	0.0091
	NFE2L2	transcription regulator	3.394	0.0135
	NLRC5	transcription regulator	2.595	0.00359
	TFEB	transcription regulator	2.435	0.00689
	HSF2	transcription regulator	2.335	0.00565
	p70 S6k	group	2.219	0.0109
	ERBB2	kinase	2.184	0.0179
Mesenchymal	TGFB1	growth factor	3.309	0.000303
	TNF	cytokine	3.148	0.00259
	RICTOR	other	3.073	0.000224
	CTNNB1	transcription regulator	2.356	0.0000707
	CD3	complex	2.356	0.000512
	CD44	other	2.333	0.0315
	ABCC8	transporter	2.216	0.000607
	LRP6	transmembrane receptor	1.98	0.00551
	miR-23a-3p (and other miRNAs w/see	mature microrna	1.964	0.00977
	ESR1	ligand-dependent nuclear receptor	1.883	0.00212
Immune	CST5	other	3.606	0.0402
	RICTOR	other	3.582	0.029
	IFNG	cytokine	3.069	0.00199
	miR-124-3p (and other miRNAs w/see	mature microrna	2.783	0.00461
	TGM2	enzyme	2.724	0.0453
	BTNL2	transmembrane receptor	2.449	0.0162
	IL7R	transmembrane receptor	2.433	0.0089
	IKZF1	transcription regulator	2.345	0.0121
	SOX1	transcription regulator	2.236	0.0483
	CD5	transmembrane receptor	2.219	0.0143

Top 10 activated upstream regulators in the four eCCA molecular classes. Genes differentially expressed between molecular classes (FDR<0.01) were identified with the Comparative Marker Selection module from GenePattern. Ingenuity Pathway analysis software was used for the inference of putative upstream regulators explaining the observed gene expression changes among the identified molecular classes. The overlap

p-value measures whether there is a statistically significant overlap between the dataset genes and the genes that are regulated by a transcriptional regulator. It is calculated using Fisher's Exact Test. Activation z-score infer the activation states of predicted transcriptional regulators.

Supplementary Table 9. Compounds potentially effective for each eCCA molecular class.

	Compound	Description	Tau
Metabolic	GW-843682X	PLK inhibitor	-99.42
	triciribine	AKT inhibitor	-99.25
	kinetin-riboside	apoptosis inducer	-99.22
	vinorelbine	tubulin inhibitor, apoptosis stimulant, microtubule inhibitor, mitosis inhibitor, mitotic inhibitor, tubulin polymerisation inhibitor, vinca alkaloid	-99.15
	MLN-4924	nedd activating enzyme inhibitor	-99.06
	MST-312	telomerase inhibitor	-98.97
	ABT-751	tubulin inhibitor, dihydropteroate synthase inhibitor, microtubule inhibitor, PABA antagonist, tubulin polymerisation inhibitor	-98.44
	didanosine	nucleoside reverse transcriptase inhibitor, reverse transcriptase inhibitor	-98.42
	emetine	protein synthesis inhibitor	-98.03
	MLN-2238	proteasome inhibitor	-97.92
	nocodazole	tubulin inhibitor, Tubulin Polymerization Inhibitors	-97.9
	prostratin	NFkB pathway activator, PKC activator	-97.69
	flubendazole	acetylcholinesterase inhibitor, microtubule inhibitor, tubulin inhibitor	-97.01
	ingenol	PKC activator	-96.9
	helveticoside	ATPase inhibitor	-96.85
	HU-211	glutamate receptor antagonist, apoptosis stimulant, NFkB pathway inhibitor, reducing agent	-96.75
	NSC-632839	ubiquitin hydrolase inhibitor, ubiquitin isopeptidase inhibitor	-96.24
	SA-63133	casein kinase inhibitor, tubulin inhibitor	-95.97
	ICI-204448	opioid receptor agonist	-95.46
	vincristine	tubulin inhibitor, microtubule inhibitor, microtubule polymerization inhibitor, tubulin polymerisation inhibitor, vinca alkaloid	-94.04
	VX-222	HCV inhibitor, RNA-directed RNA polymerase inhibitor	-93.99
	LDN-193189	ALK inhibitor, serine/threonine protein kinase inhibitor	-93.71
	lobeline	acetylcholine receptor antagonist, dopamine receptor modulator, opioid receptor antagonist, vesicular monoamine transporter ligand	-93.56
	rifampicin	DNA directed RNA polymerase inhibitor, enzyme inducer	-93.54
	LY-2183240	FAAH inhibitor, FAAH reuptake inhibitor	-93.4
	ruxolitinib	JAK inhibitor, tyrosine kinase inhibitor	-93.03
	quinoclamine	algicide	-92.68
	phorbol-12-myristate-13-acetate	PKC activator, CD antagonist	-92.54
	SA-792574	microtubule inhibitor, tubulin inhibitor	-92.32
	ciclacillin	cell wall synthesis inhibitor	-90.52
	altanserin	serotonin receptor antagonist, collagen stimulant	-90.31
	15-delta-prostaglandin-j2	PPAR receptor agonist, FXR antagonist	-90.25
Proliferation	calyculin	protein phosphatase inhibitor	-99.27
	W-12	calmodulin antagonist	-98.07
	narciclasine	cofilin signaling pathway activator, LIM kinase activator, ROCK activator	-97.81
	dexbrompheniramine	histamine receptor antagonist	-97.42
	entinostat	HDAC inhibitor, cell cycle inhibitor	-96.56
	BI-2536	PLK inhibitor, apoptosis stimulant, cell cycle inhibitor, protein kinase inhibitor	-96.47
	dihydroergocristine	adrenergic receptor antagonist, prolactin inhibitor, adrenergic receptor partial agonist, dopamine receptor agonist, dopamine receptor partial agonist, dopamine re	-96.3
	JWE-035	Aurora kinase inhibitor	-95.53
	givinostat	HDAC inhibitor, interleukin receptor antagonist, interleukin synthesis inhibitor, tumor necrosis factor receptor antagonist, tumor necrosis factor release inhibitor	-95.33
	emetine	protein synthesis inhibitor	-94.86
	belinostat	HDAC inhibitor, cell cycle inhibitor	-94.65
	NCH-51	HDAC inhibitor	-94.63
	cephaeline	protein synthesis inhibitor	-94.51
	everolimus	mTOR inhibitor, angiogenesis inhibitor, cell cycle inhibitor, immunosuppressant, protein kinase inhibitor, rotamase inhibitor	-94.43
	formoterol	adrenergic receptor agonist	-94.31
	isoeugenol	nitric oxide production inhibitor	-93.57
	AZD-7762	CHK inhibitor	-92.92
	tubaic-acid	mitochondrial complex I inhibitor, NADH-ubiquinone oxidoreductase (Complex I) inhibitor	-92.91
	trichostatin-a	HDAC inhibitor, CDK expression enhancer, ID1 expression inhibitor	-92.9
	erythrosine	food coloring agent	-92.84
	III606050	cytochrome P450 inhibitor	-92.84
	puromycin	adenosine receptor agonist, protein synthesis inhibitor	-92.72
	AR-A014418	glycogen synthase kinase inhibitor	-92.28
	JNJ-7706621	CDK inhibitor, Aurora kinase inhibitor	-92.12
	mofezolac	cyclooxygenase inhibitor, cytochrome P450 inhibitor, platelet aggregation inhibitor	-92.05
	oxfendazole	anthelmintic agent	-91.97
	clobetasol	glucocorticoid receptor agonist	-91.93
	canrenoic-acid	aldosterone antagonist	-91.93
	dofetilide	polarization inhibitor, potassium channel antagonist, potassium channel blocker	-91.9
	cabergoline	dopamine receptor agonist, prolactin inhibitor, prolactin secretion inhibitor	-91.85
	ZSTK-474	PI3K inhibitor	-91.85
	L-655240	platelet aggregation inhibitor, prostanoid receptor antagonist, thromboxane receptor antagonist	-91.44
	XMD-892	MAP kinase inhibitor, BMK inhibitor, leucine rich repeat kinase inhibitor	-90.67
	moclobemide	monoamine oxidase inhibitor	-90.66
	linifanib	PDGFR tyrosine kinase receptor inhibitor, VEGFR inhibitor, angiogenesis inhibitor, colony stimulating factor receptor antagonist, colony stimulating factor recepto	-90.61
	remacemide	glutamate receptor antagonist, glutamate receptor agonist	-90.55
	ZM-306416	Src and Abl inhibitor, vascular endothelial growth factor receptor 1 (VEGFR1) inhibitor	-90.44
	etodolac	cyclooxygenase inhibitor, TRPV agonist	-90.4
	THM-I-94	HDAC inhibitor, apoptosis stimulant, cell cycle inhibitor	-90.34
	BRD-K53780220	casein kinase inhibitor, FLT3 inhibitor	-90.25
	parachlorophenol	disinfectant	-90.17

Mesenchymal	SJ-172550	MDM inhibitor	-96.86
	AC-55649	RAR agonist, retinoid receptor agonist	-95.9
	estrone	estrogen receptor agonist, estrogenic hormone	-95.45
	methyl-2,5-dihydroxycinnamate	EGFR inhibitor, tyrosine kinase inhibitor	-94.96
	tyrphostin-AG-1295	FLT3 inhibitor, PDGFR tyrosine kinase receptor inhibitor	-94.04
	BAS-09104376	HIV integrase inhibitor	-93.91
	fluocinonide	corticosteroid agonist, corticosteroid hormone receptor agonist	-93.67
	baeomycesic-acid	lipoxygenase inhibitor	-92.7
	oxymetholone	androgen receptor agonist, synthetic hormone with anabolic and androgenic properties	-92.7
	farnesol	amine oxidase B inhibitor, FXR agonist	-92.65
	diethyltoluamide	DEET, activator of fly antenna ionotropic receptor IR40a	-92.04
	lansoprazole	ATPase inhibitor	-91.07
	linezolid	50S ribosomal subunit inhibitor, protein synthesis inhibitor, monoamine oxidase inhibitor	-90.97
	dexamethasone	glucocorticoid receptor agonist, corticosteroid agonist, immunosuppressant	-90.67
	PT-630	dipeptidyl peptidase inhibitor, fibroblast activation protein inhibitor	-90.29
	esomeprazole	ATPase inhibitor, ABC transporter expression enhancer	-90.22
Immune	panobinostat	HDAC inhibitor, apoptosis stimulant, cell cycle inhibitor	98.87
	SJ-172550	MDM inhibitor	-98.66
	rucaparib	PARP inhibitor	-97.18
	doxylamine	histamine receptor antagonist	-97.06
	exemestane	aromatase inhibitor	-96.46
	eugenol	androgen receptor (AR) inhibitor, free radical scavenger, monoamine oxidase inhibitor, quorum sensing signaling modulator	-96.23
	CA-074-Me	cathepsin inhibitor, antiamyloidogenic agent	-96.21
	mupirocin	isoleucyl-tRNA synthetase inhibitor	-96.14
	deferiprone	chelating agent, cytochrome P450 inhibitor, iron absorption inhibitor, reducing agent	-96.01
	NAS-181	serotonin receptor antagonist	-96
	ST-91	adrenergic receptor agonist	-95.41
	verapamil	calcium channel blocker, L-type calcium channel blocker, dopamine receptor antagonist	-95.31
	memantine	glutamate receptor antagonist, glutamate release inhibitor	-95.08
	topiramate	carbonic anhydrase inhibitor, glutamate receptor antagonist, kainate receptor antagonist, GABA receptor agonist, Sodium Channel Blockers, voltage-gated sodiu	-95.03
	ibudilast	phosphodiesterase inhibitor, leukotriene receptor antagonist, toll-like receptor antagonist, macrophage migration inhibiting factor inhibitor, macrophage migration i	-94.86
	tienilic-acid	sodium/potassium/chloride transporter inhibitor, uric acid diuretic	-94.62
	lysylphenylalanyl-tyrosine	heparin activation inhibitor	-94.47
	latrepirdine	glutamate receptor antagonist, histamine receptor antagonist, serotonin receptor antagonist	-94.43
	tyrphostin-AG-82	EGFR inhibitor, epidermal growth factor receptor (EGFR) inhibitor, tyrosine kinase inhibitor	-94.41
	epigallocatechin	AP inhibitor, aromatase inhibitor, bacterial efflux pump inhibitor, beta amyloid aggregation inhibitor, beta amyloid protein neurotoxicity inhibitor, beta secretase inh	-94.38
	amisulpride	dopamine receptor antagonist	-94.31
	desoxypeganine	acetylcholinesterase inhibitor, monoamine oxidase inhibitor	-93.98
	cholic-acid	ferrochelatase inhibitor, unidentified pharmacological activity	-93
	eugenitol	androgen receptor (AR) inhibitor, free radical scavenger, monoamine oxidase inhibitor, quorum sensing signaling modulator	-92.99
	3-amino-benzamide	PARP inhibitor	-92.94
	fluoropyruvate	PDH inhibitor	-92.85
	BRD-K21009077	dual specificity protein phosphatase inhibitors	-92.71
	L-692585	growth hormone releasing peptide ligand agonist, growth hormone secretagogue	-92.53
	temozolomide	DNA alkylating drug, DNA damage inducer, DNA inhibitor, topoisomerase inhibitor	-92.51
	methylnorlichexanthone	Aurora kinase inhibitor, Pim kinase inhibitor, VEGFR inhibitor	-92.29
	resveratrol	apolipoprotein expression enhancer, beta-secretase inhibitor, cyclooxygenase inhibitor, cytochrome P450 inhibitor, lipid peroxidase inhibitor, MAP kinase inhibitor	-92.28
	AC-55649	RAR agonist, retinoid receptor agonist	-92.26
	rifaximin	50S ribosomal subunit inhibitor, DNA directed DNA polymerase inhibitor, PXR agonist, RNA synthesis inhibitor	-92.24
	probenecid	MRP inhibitor, TRPV agonist, uricosuric blocker	-92.19
	oxymetholone	androgen receptor agonist, synthetic hormone with anabolic and androgenic properties	-92.16
	rilmenidine	adrenergic receptor agonist, imidazoline receptor agonist	-92.16
	brinzolamide	carbonic anhydrase inhibitor	-92.03
	piceid	glucosidase inhibitor, ICAM1 expression inhibitor, VCAM expression inhibitor, xanthine oxidase inhibitor	-91.74
	tomelukast	leukotriene receptor antagonist	-91.72
	YM-298198	glutamate receptor antagonist	-91.52
	BRD-K57954781	apoptosis stimulant, DNA inhibitor	-91.48
	lonidamine	glucokinase inhibitor, protein synthesis inhibitor	-91.25
	AM-404	FAAH transport inhibitor, anandamide transport inhibitor, nucelar factor of activated T-cells inhibitor, TRPV agonist	-91.13
	valsartan	angiotensin receptor antagonist	-90.77
	ritonavir	HIV protease inhibitor, cytochrome P450 inhibitor	-90.7
	zacopride	serotonin receptor antagonist, serotonin receptor agonist	-90.62
	sitagliptin	dipeptidyl peptidase inhibitor, HMGCR inhibitor, insulin secretagogue, tumor necrosis factor expression inhibitor	-90.57
	nicotine	acetylcholine receptor agonist	-90.28
	valproic-acid	HDAC inhibitor, ABAT inhibitor, GABA receptor agonist, GABAergic transmission enhancer, voltage-gated sodium channel blocker	-90.07
	methantheline	acetylcholine receptor antagonist	-90.06

Top up-regulated genes in each eCCA molecular class were used to identify perturbations (treatments with 2837 small molecules in 9 cancer cell lines)[38] that elicit opposed expression signatures (tau < -90). A tau of -90 indicates that only 10% of reference perturbations were more dissimilar to the query.

Supplementary Table 10. Prognostic factors in terms of overall survival in eCCA.

	Univariate		Multivariate	
	HR (95% CI)	p value	HR (95% CI)	p value
Gender (male vs female)	1.74 (0.98-3.07)	0.058		
Age (>65 years)	1.21 (0.73-2.01)	0.564		
Anatomical subtype (perihilar vs distal)	1.16 (0.59-2.25)	0.669		
Tumor diameter (>25mm)	1.09 (0.63-1.88)	0.759		
Pathological lymph nodes (present vs absent)	1.66 (0.95-2.90)	0.072		
Resection margins (R1 vs R0)	1.84 (1.05-3.22)	0.034		0.154
Bilirubin (>3.3mg/dl)	1.18 (0.69-2.00)	0.551		
ALT (>99UI/l)	1.38 (0.80-2.36)	0.244		
Albumin (>37mg/dl)	0.77 (0.43-1.37)	0.369		
CA19.9 (>142UI/l)	1.48 (0.79-2.76)	0.220		
CEA (>2.4ng/ml)	1.52 (0.75-3.07)	0.244		
Cell differentiation (G3/G4 vs G1/G2)	2.84 (1.36-5.92)	0.005	3.04 (1.38-6.69)	0.006
Perineural invasion (present vs absent)	1.39 (0.82-2.37)	0.226		
Vascular invasion (present vs absent)	1.09 (0.54-2.23)	0.805		
Molecular class (mesenchymal vs rest)	1.85 (1.07-3.17)	0.027	1.95 (1.12-3.40)	0.018

Variables with $p<0.05$ in the univariate analysis were subsequently introduced in the stepwise multivariate model using Cox regression. Overall survival was defined as the time between surgical resection and death of any cause or lost follow-up. Patients with less than one month of follow-up (n=24) were excluded from the analysis of prognostic factors in order to minimize the effect of surgical complications as a determinant of clinical outcome.

Supplementary Table 11. eCCA classifier containing 174 genes.

Gene	Metabolic	Proliferation	Mesenchymal	Immune
ORM1	RPL28	POSTN	IGL@
FGA	RPL37A	THBS2	IGHA1
ALB	STAG3L2	SYTL4	PIK3IP1
HP	WAC	NRP2	CORO1A
APOA2	CLIC1	INHBA	IGJ
AGT	RPS3	NEURL	ATP2A3
ITIH3	POM121C	COL1A1	ARHGDIB
ITIH2	PIK3C2A	COL12A1	IGHM
SERPINA6	STAG3L3	OLFML2B	PTGDS
ITIH4	YWHAB	SPARC	HIST1H2AE
FGG	RPL11	COL10A1	SMAP2
PCK1	EEF2	CDH11	PAPSS1
VTN	PTPN2	TAGLN	ITGAX
SLC25A47	RPL36A	BHLHE40	CD4
FGB	SERP1	ITGA11	IL23A
AZGP1	EFTUD2	VCAN	ARHGAP9
SERPINA1	CNOT7	MYL9	CCDC69
SERPINC1	STK25	CALD1	HIST1H2BD
SLC13A5	UXT	STON1	CCR7
ITIH1	RPL41	GREM1	CYR61
KNG1	RPS29	LGALS1	NR4A1
CYP27A1	TMBIM6	VIM	HIST1H4E
MGST1	MARCH6	TGFBI	TSC22D3
CYP2B6	HDGF	COMP	TCF7
PBLD	PAPOLA	HTRA3	
APOC2	GNB2L1	FAM127A	
APOC3	SYNCRIP	COL11A1	
AHSG	ATP5L	CCDC80	
C9	MRPL42	EHD2	
ALDH4A1	FAU	OSBPL5	
CFB	KHSRP	COL5A1	
APOB	SPINT1	PALLD	
ADH1B	RPL4	WIPF1	
TF	RBM17	IVNS1ABP	
CYP3A4	RPL14	SULF1	
APOA1	SMARCE1	TIMP2	
CYP2E1	MORF4L2	SH3PXD2B	
VNN1	MRFAP1	FSTL1	
SEPP1	MRPS24	C5AR1	
SERPINA3	UBE2N	COL6A1	
APCS	UBA52	PTK7	
EPHX1	OST4	PLOD2	
CBS	TAPBP	DNAJB5	
RBP4	H2AFY	CDH13	
UGT2B4	RPS15A	CRISPLD2	
CP	SLC25A6	PLK3	
CYP1A2	UPF3A	COL1A2	
CPS1	RPL5	ROR2	
TTC39C	SNW1	SERPINE1	
CYP3A5	TUBB2C	LSAMP	

The Class Neighbors tool from GenePattern was used to determine based on a signal-to-noise distance function which genes were most closely correlated with a specific molecular class template and how significant the correlation was compared with random permutation versions of the phenotype (intersection of observed data with 1% significance level). The 174-gene classifier -composed by a maximum of 50 genes defining each class- was able to assign eCCA samples to one of the four molecular classes with a global precision of 86% in our discovery eCCA cohort.

Supplementary Table 12. External validation of eCCA molecular classes in the ICGC cohort.

Sample	Anatomical location	Inferred eCCA molecular class	OS status	OS days	ERBB2 mutation
BD109	iCCA	Metabolic	0	380	Absent
BD111	iCCA	Metabolic	1	681	Absent
BD125	iCCA	Metabolic	1	1017	Absent
BD137	iCCA	Metabolic	0	1150	Absent
BD15	iCCA	Metabolic	1	181	Absent
BD151	iCCA	Metabolic	0	419	Absent
BD159	iCCA	Metabolic	0	503	Absent
BD165	iCCA	Metabolic	0	358	Present
BD167	iCCA	Metabolic	0	134	Absent
BD168	iCCA	Metabolic	0	220	Absent
BD19	iCCA	Metabolic	0	2382	Absent
BD210	iCCA	Metabolic	0	59	Absent
BD212	iCCA	Metabolic	1	694	Absent
BD218	iCCA	Metabolic	0	601	Absent
BD231	iCCA	Metabolic	1	648	Absent
BD237	iCCA	Metabolic	0	3610	Absent
BD24	iCCA	Metabolic	0	2284	Absent
BD242	iCCA	Metabolic	0	2127	Absent
BD244	iCCA	Metabolic	1	976	Absent
BD27	iCCA	Metabolic	1	1849	Absent
BD318	iCCA	Metabolic	NA	NA	NA
BD36	iCCA	Metabolic	0	2234	Absent
BD40	iCCA	Metabolic	1	1791	Absent
BD42	iCCA	Metabolic	0	3081	Absent
BD78	iCCA	Metabolic	1	156	Absent
BD81	iCCA	Metabolic	1	1905	Absent
BD95	iCCA	Metabolic	0	2134	Absent
BD105	iCCA	Proliferation	0	1981	Absent
BD114	iCCA	Proliferation	1	181	Present
BD117	iCCA	Proliferation	1	894	Absent
BD124	iCCA	Proliferation	0	1266	Absent
BD132	iCCA	Proliferation	1	445	Absent
BD134	iCCA	Proliferation	1	90	Absent
BD141	iCCA	Proliferation	0	1756	Absent
BD197	iCCA	Proliferation	1	537	Absent
BD199	iCCA	Proliferation	1	603	Absent
BD214	iCCA	Proliferation	1	125	Absent
BD23	iCCA	Proliferation	1	319	Absent
BD28	iCCA	Proliferation	0	2204	Absent
BD29	iCCA	Proliferation	1	53	Absent
BD308	iCCA	Proliferation	NA	NA	NA
BD334	iCCA	Proliferation	NA	NA	NA
BD46	iCCA	Proliferation	1	1094	Absent
BD47	iCCA	Proliferation	0	2914	Absent
BD57	iCCA	Proliferation	0	2756	Absent
BD74	iCCA	Proliferation	1	1627	Absent
BD8	iCCA	Proliferation	0	2934	Absent
BD82	iCCA	Proliferation	1	428	Present
BD84	iCCA	Proliferation	0	1835	Absent
BD92	iCCA	Proliferation	0	2172	Absent
BD104	iCCA	Mesenchymal	1	168	Absent
BD118	iCCA	Mesenchymal	1	851	Absent
BD129	iCCA	Mesenchymal	1	386	Absent
BD135	iCCA	Mesenchymal	0	1911	Absent
BD14	iCCA	Mesenchymal	0	1624	Absent
BD143	iCCA	Mesenchymal	0	162	Absent
BD147	iCCA	Mesenchymal	0	83	Absent
BD148	iCCA	Mesenchymal	0	264	Absent
BD149	iCCA	Mesenchymal	0	403	Absent

BD152	iCCA	Mesenchymal	0	767	Absent
BD153	iCCA	Mesenchymal	0	624	Absent
BD154	iCCA	Mesenchymal	0	135	Absent
BD157	iCCA	Mesenchymal	1	472	Absent
BD169	iCCA	Mesenchymal	0	146	Absent
BD18	iCCA	Mesenchymal	1	299	Absent
BD200	iCCA	Mesenchymal	0	2259	Absent
BD226	iCCA	Mesenchymal	0	916	Absent
BD247	iCCA	Mesenchymal	0	72	Absent
BD3	iCCA	Mesenchymal	0	166	Absent
BD30	iCCA	Mesenchymal	0	2608	Absent
BD31	iCCA	Mesenchymal	1	882	Absent
BD310	iCCA	Mesenchymal	NA	NA	NA
BD312	iCCA	Mesenchymal	NA	NA	NA
BD313	iCCA	Mesenchymal	NA	NA	NA
BD45	iCCA	Mesenchymal	0	1977	Absent
BD5	iCCA	Mesenchymal	0	3111	Absent
BD54	iCCA	Mesenchymal	1	646	Absent
BD56	iCCA	Mesenchymal	1	1380	Absent
BD80	iCCA	Mesenchymal	1	181	Absent
BD97	iCCA	Mesenchymal	NA	NA	NA
BD138	iCCA	Immune	1	4	Absent
BD140	iCCA	Immune	1	264	Absent
BD146	iCCA	Immune	0	670	Absent
BD150	iCCA	Immune	0	582	Absent
BD195	iCCA	Immune	NA	NA	NA
BD219	iCCA	Immune	0	1349	Absent
BD224	iCCA	Immune	0	818	Absent
BD239	iCCA	Immune	1	1128	Absent
BD6	iCCA	Immune	1	784	Absent
BD87	iCCA	Immune	0	762	Present
BD10	iCCA	Unclassified	1	1153	Absent
BD101	iCCA	Unclassified	1	1829	Absent
BD115	iCCA	Unclassified	0	1604	Absent
BD12	iCCA	Unclassified	1	532	Absent
BD121	iCCA	Unclassified	0	1335	Absent
BD142	iCCA	Unclassified	0	1068	Absent
BD196	iCCA	Unclassified	NA	NA	NA
BD201	iCCA	Unclassified	1	1424	Absent
BD21	iCCA	Unclassified	1	452	Absent
BD211	iCCA	Unclassified	1	341	Absent
BD213	iCCA	Unclassified	1	208	Absent
BD220	iCCA	Unclassified	0	458	Absent
BD221	iCCA	Unclassified	0	1013	Absent
BD222	iCCA	Unclassified	0	1016	Absent
BD223	iCCA	Unclassified	1	722	Absent
BD227	iCCA	Unclassified	1	277	Absent
BD229	iCCA	Unclassified	1	144	Absent
BD230	iCCA	Unclassified	1	486	Absent
BD232	iCCA	Unclassified	1	1428	Absent
BD233	iCCA	Unclassified	1	196	Absent
BD234	iCCA	Unclassified	1	218	Absent
BD243	iCCA	Unclassified	1	334	Absent
BD25	iCCA	Unclassified	1	402	Absent
BD32	iCCA	Unclassified	1	749	Absent
BD33	iCCA	Unclassified	1	574	Absent
BD38	iCCA	Unclassified	1	712	Absent
BD41	iCCA	Unclassified	1	654	Absent
BD72	iCCA	Unclassified	1	474	Present
BD75	iCCA	Unclassified	0	1825	Absent
BD79	iCCA	Unclassified	0	1722	Absent
BD86	iCCA	Unclassified	1	174	Absent
BD88	iCCA	Unclassified	1	175	Absent

BD11	pCCA	Metabolic	1	583	Absent
BD112	pCCA	Proliferation	0	1254	Absent
BD49	pCCA	Proliferation	0	2879	Absent
BD53	pCCA	Proliferation	0	1334	Absent
BD155	pCCA	Mesenchymal	0	210	Absent
BD163	pCCA	Mesenchymal	1	267	Absent
BD52	pCCA	Mesenchymal	0	2776	Absent
BD9	pCCA	Mesenchymal	0	1037	Absent
BD91	pCCA	Mesenchymal	1	90	Absent
BD162	pCCA	Unclassified	0	314	Absent
BD166	pCCA	Unclassified	0	328	Absent
BD241	pCCA	Unclassified	0	2118	Absent
BD34	pCCA	Unclassified	0	355	Absent
BD306	dCCA	Metabolic	NA	NA	NA
BD13	dCCA	Proliferation	0	2882	Present
BD158	dCCA	Proliferation	0	477	Absent
BD16	dCCA	Proliferation	1	489	Absent
BD207	dCCA	Proliferation	0	68	Present
BD305	dCCA	Proliferation	NA	NA	NA
BD314	dCCA	Proliferation	NA	NA	NA
BD315	dCCA	Proliferation	NA	NA	NA
BD316	dCCA	Proliferation	NA	NA	NA
BD322	dCCA	Proliferation	NA	NA	NA
BD333	dCCA	Proliferation	NA	NA	NA
BD35	dCCA	Proliferation	1	672	Absent
BD4	dCCA	Proliferation	0	2547	Absent
BD48	dCCA	Proliferation	0	2920	Absent
BD55	dCCA	Proliferation	0	2725	Absent
BD7	dCCA	Proliferation	0	3126	Absent
BD128	dCCA	Mesenchymal	0	1436	Absent
BD160	dCCA	Mesenchymal	1	34	Absent
BD319	dCCA	Mesenchymal	NA	NA	NA
BD122	dCCA	Immune	0	887	Absent
BD161	dCCA	Immune	0	388	Absent
BD317	dCCA	Immune	NA	NA	NA
BD205	dCCA	Unclassified	0	83	Absent
BD321	dCCA	Unclassified	NA	NA	NA
BD336	dCCA	Unclassified	NA	NA	NA
BD37	dCCA	Unclassified	1	389	Absent
BD170	GBC	Unclassified	1	804	Absent
BD171	GBC	Mesenchymal	1	297	Absent
BD172	GBC	Mesenchymal	1	792	Absent
BD173	GBC	Unclassified	1	384	Absent
BD175	GBC	Unclassified	0	3365	Absent
BD176	GBC	Proliferation	1	435	Absent
BD179	GBC	Unclassified	1	92	Absent
BD180	GBC	Proliferation	1	823	Present
BD182	GBC	Unclassified	0	204	Absent
BD184	GBC	Proliferation	0	2029	Absent
BD186	GBC	Mesenchymal	0	1580	Absent
BD187	GBC	Unclassified	0	1432	Absent
BD189	GBC	Unclassified	1	704	Absent
BD190	GBC	Unclassified	0	157	Absent
BD191	GBC	Unclassified	1	328	Absent
BD192	GBC	Unclassified	0	700	Absent
BD194	GBC	Mesenchymal	0	79	Absent
BD202	GBC	Unclassified	0	112	Absent
BD206	GBC	Mesenchymal	0	99	Absent
BD335	GBC	Proliferation	NA	NA	NA

Fastq files of RNAseq from 182 samples of biliary tract cancer (iCCA=122, pCCA=14, dCCA=26, GBC=20) were downloaded from the European Genome-phenome Archive. One sample (BD20) was not successfully normalized. Prediction in the external cohort of the eCCA classifier was performed using the Nearest Template Prediction method, as implemented in the specific module of GenePattern. We correlated the proposed molecular classes of eCCA (Metabolic, Proliferation, Mesenchymal and Immune) with clinical variables and non-silent somatic mutations analyzed by whole-exome sequencing and available at the International Cancer Genome Consortium (ICGC) Data portal. OS: Overall survival.

Supplementary Table 13. Ongoing clinical trials assessing targeted therapies in eCCA.

NCT Number	Experimental drug	Target	Trial	Line	Biomarker (% in eCCA)
NCT02989857	AG-120	IDH1	Phase 3	Second	IDH1 mut (2.0%)
NCT03656536	Pemigatinib	FGFR1/2/3	Phase 3	First	FGFR2 rearrangement (0%)*
NCT03875235	Durvalumab	PD-L1	Phase 3	First	
NCT03478488	KN035	PD-L1	Phase 3	First	
NCT03093870	Varlitinib	pan-HER	Phase 2/3	Second	*Proliferation class* (22.5%)
NCT03873532	Surufatinib	Multi-TKI	Phase 2/3	Second	
NCT02162914	Regorafenib	Multi-TKI	Phase 2	Second	
NCT02232633	BBI503	Stemness kinase inhibitor	Phase 2	Second	
NCT02520141	Ramucirumab	VEGFR2	Phase 2	Second	
NCT02631590	Copanlisib	PI3K	Phase 2	First	
NCT02982720	Pembrolizumab + Sylatron	PD-1	Phase 2	Second	*Immune class* (11.5%)
NCT03111732	Pembrolizumab	PD-1	Phase 2	First	*Immune class* (11.5%)
NCT03201458	Atezolizumab + Cobimetinib	PD-L1/MEK	Phase 2	Second	
NCT03250273	Entinostat	HDAC/PD-1	Phase 2	Second	
NCT03377179	Yeliva	SK2 (Lipid metabolism)	Phase 2	Second	*Metabolic class* (18.7%)
NCT03473574	Durvalumab + Tremelimumab	PD-L1/CTLA4	Phase 2	First	
NCT03486678	SHR-1210	PD-1	Phase 2	First	*Immune class* (11.5%)
NCT03613168	Trastuzumab	ERBB2	Phase 2	First	ERBB2 overexpression/amp (3.9%)
NCT03833661	M7824	PD-L1/TGFβ	Phase 2	Second	*Mesenchymal class* (47.3%)
NCT03878095	Olaparib + Ceralasertib	PARP/ATR	Phase 2	Second	IDH1/2 mut (4.7%)
NCT03796429	Toripalimab	PD-1	Phase 2	First	*Immune class* (11.5%)
NCT03144856	Apatinib	Multi-TKI	Phase 2	Second	
NCT02579616	Lenvatinib	Multi-TKI	Phase 2	Second	
NCT03231176	Varlitinib	pan-HER	Phase 2	Second	Proliferation class (22.5%)
NCT03639935	Rucaparib + Nivolumab	PARP/PD-1	Phase 2	Second	
NCT03110328	Pembrolizumab	PD-1	Phase 2	Second	*Immune class* (11.5%)
NCT02829918	Nivolumab	PD-1	Phase 2	Second	*Immune class* (11.5%)
NCT03110484	Erlotinib + Pemetrexed	EGFR	Phase 2	Second	
NCT02711553	Ramucirumab / Meristinb	VEGFR2/TKI	Phase 2	First	
NCT03260712	Pembrolizumab	PD-1	Phase 2	Second	*Immune class* (11.5%)
NCT02703714	Pembrolizumab + GM-CSF	PD-1	Phase 2	Second	*Immune class* (11.5%)
NCT03046862	Durvalumab + Tremelimumab	PD-L1/CTLA4	Phase 2	First	
NCT03101566	Nivolumab + Ipilimumab	PD-1/CTLA4	Phase 2	First	
NCT02115542	Regorafenib	Multi-TKI	Phase 2	Second	
NCT02151084	Selumetinib	MEK	Phase 2	First	
NCT02265341	Ponatinib	Multi-TKI	Phase 2	Second	FGFR alterations (2%)
NCT03427242	Apatinib	Multi-TKI	Phase 2	Second	
NCT03092895	SHR-1210 + Apatinib	PD-1/Multi-TKI	Phase 2	Second	
NCT02128282	CX-4945	CK2 (Cell cycle/DNA repair)	Phase 1/2	First	*Proliferation class* (22.5%)
NCT03785873	Nivolumab + Nal-Irinotecan	PD-1	Phase 1/2	Second	
NCT02992340	Varlitinib	pan-HER	Phase 1/2	Second	*Proliferation class* (22.5%)
NCT02773459	MEK162	MEK	Phase 1/2	Second	
NCT02386397	Regorafenib	Multi-TKI	Phase 1/2	First	
NCT01828034	MEK162	MEK	Phase 1/2	First	
NCT03257761	Guadecitabine + Durvalumab	DNMT/PD-L1	Phase 1	Second	
NCT03267940	PEGPH20 + Atezplizumab	Hyaluronidase/PD-L1	Phase 1	First	*Mesenchymal class* (47.3%)

Data of ongoing clinical trials was obtained in March 2019 from the ClinicalTrials.gov database. Keyword searches for “cholangiocarcinoma” and “biliary tract cancer” were used to identify active clinical trials (recruiting, not yet recruiting, active, not recruiting, enrolling by invitation) assessing targeted therapies for advanced eCCA. Basket trials assessing solid tumors other than hepato-biliary-pancreatic tumors were excluded. Biomarkers in italic are suggested based on the present study. *FGFR2 rearrangements exclusively detected in iCCA according to literature.

Supplementary Table 14. Expression factor comprising 149 genes identified by NMF in eCCA.

Gene
HOMER2
VIL1
GABRB3
CMBL
FAM171A1
HPX
TTC39C
HMGN3
PRDX2
SPINK1
SSR2
KRT18
GGH
TFR2
AHSG
GYG2
SAA1
FMO3
HNF4G
GATM
ARG1
CYP2D7P1
PLGLB2
CYP2C8
G6PC
CTAGE5
ATF5
HPD
CYP2C18
RNF5
SLC47A1
TMEM176A
DDTL
LOC100291980
PXMP2
SEMA6A
PTPRF
PRAP1
UGT2B10
TTR
SERPINC1
APOH
ALDH4A1
AKR1C3
C9
SLC19A3
OCLN
AFF4
C19orf77
ST6GAL1
APCS
ITIH4
CYP1A2
MAGI1
LOC100291873
UQCRFS1
GSTM1
ITIH3
PBLD
UGT2B4
SLC35C1
SEPP1
HPS3
CYP2A6
GOT1
AMY1A
MTHFD1
FGG
CFB
AQP9
ADH1A
AZGP1
A2M
SLC22A1
SERPINA6
DHCR24
CP
GC
UGT1A10
ACADVL
APOC2
CLU
MT1F
CDH2
AKR1C2
ITIH1
RPLP0
CYP4A11
CYP2A6
ALAS1
SEBOX
CYP3A7
C1RL
MT1M
VTN
CD96
MT1JP
SC4MOL
CYP3A4
ADH1B
CYP2B6
CFH
PCK1
ASS1
C3
FABP5
CYP3A5
UGT2B7
MT1P3
EPHX1
APOB
CYP2B6
PLG
SORD
FGFRL1
AGT
RBP4
AKR1C1
HAMP
SLC13A5
APOC3
CBS
CPS1
CXADR
KNG1
DYNC1I2
APOA1
FABP1
CYP2E1
CRP
ORM1
CYP4A11
FGB
TF
AQP3
SERPINA3
SCD
APOC1
AKR1C1
ALB
SERPINA1
ITIH2
HP
MT1G
APOA2
ADH1A
DCAF6
FGA
ORM1

The top 1696 most variable genes in our dataset identified with the Preprocess Dataset module in Genepattern underwent Non-negative matrix factorization (NMF) in order to perform virtual microdissection of gene expression data. A factor composed by 149 genes was identified and further characterized as explained in Supplementary Table 12.

Supplementary Table 15. Identification of a liver-related expression factor in eCCA.

Gene Set Name	Genes in Gene Set	Description	Genes in Overlap	p-value	FDR q-value
HSIAO_LIVER_SPECIFIC_GENES	248	Liver selective genes	66	5.71E-117	1.29E-112
MODULE_23	562	Liver genes - metabolism and xenobiotics.	65	2.55E-89	2.88E-85
MODULE_55	830	Genes in the cancer module 55.	66	1.05E-79	7.94E-76
MODULE_88	833	Heart, liver, kidney and pancreas metabolic and xenobiotic response genes.	64	4.77E-76	2.70E-72
MODULE_24	452	Fetal liver genes - metabolism and xenobiotics.	55	8.49E-76	3.84E-72
GNF2_HPX	135	Neighborhood of HPX	38	1.12E-66	4.23E-63
GNF2_HPN	134	Neighborhood of HPN	36	2.92E-62	9.43E-59
GNF2_LCAT	124	Neighborhood of LCAT	33	7.52E-57	2.12E-53
CAR_HPX	73	Neighborhood of HPX	29	5.78E-56	1.45E-52
GNF2_TST	104	Neighborhood of TST	30	7.29E-53	1.65E-49

149 genes identified with Non-negative matrix factorization (NMF) were interrogated using curated gene sets from MSigDB collections. According to significant overlaps of selected genes, a liver-related expression factor was proposed.

Supplementary References:

[1] Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006;38:500–1.

[2] Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47–52.

[3] Farshidfar F, Zheng S, Gingras M-C, Newton Y, Shih J, Robertson AG, et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH - Mutant Molecular Profiles. Cell Rep 2017;18:2780–94.

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

[5] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst 2015;1:417–25.

[6] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.

[7] Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype

Relationships and Predictors of Response to Checkpoint Blockade. CellReports 2017;18:248–62.

[8] Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;24:1550–8.

[9] Sia D, Hoshida Y, Villanueva A, Roayaie S, Ferrer J, Tabak B, et al. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. Gastroenterology 2013;144:829–40.

[10] Andersen JB, Spee B, Blechacz BR, Avital I, Komuta M, Barbour A, et al. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology 2012;142:1021–31.

[11] Sulpice L, Rayar M, Desille M, Turlin B, Fautrel A, Boucher E, et al. Molecular profiling of stroma identifies osteopontin as an independent predictor of poor prognosis in intrahepatic cholangiocarcinoma. Hepatology 2013;58:1992–2000.

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

[13] Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–92.

[14] Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro De Moura

M, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology 2017;153:812–26.

[15] Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500–3.

[16] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science 1999;286:531–7.

[17] Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med 2018;10:1–8.

[18] Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, et al. Clinical analysis of PMS2 : mutation detection and avoidance of pseudogenes. Hum Mutat 2010;31:588–93.

[19] Boeva V, Popova T, Lienard M, Toffoli S, Kamal M, Le Tourneau C, et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. Bioinformatics 2014;30:3443–50.

[20] Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013;31:1023–31.

[21] Wolff a. C, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. J Clin Oncol 2013;31:3997–4013.

[22] D'Incecco A, Andreozzi M, Ludovini V, Rossi E, Capodanno A, Landi L, et al. PD-1 and PD-L1 expression in molecularly selected non-small-cell lung cancer patients. Br J Cancer 2015;112:95–102.

[23] Ueno M, Chung HC, Nagrial A, Marabelle A, Kelley RK, Xu L, et al. Pembrolizumab for advanced biliary adenocarcinoma: Results from the multicohort, phase 2 KEYNOTE-158 study. Ann Oncol 2018;29:(suppl_8): viii205-viii270.

[24] Chuaysri C, Thuwajit P, Paupairoj A, Chau-In S, Suthiphongchai T, Thuwajit C. Alpha-smooth muscle actin-positive fibroblasts promote biliary cell proliferation and correlate with poor survival in cholangiocarcinoma. Oncol Rep 2009;21:957–69.

[25] Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, et al. Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). N Engl J Med 2005;352:1851–60.

[26] Soliman NA, Yussif SM. Ki-67 as a prognostic marker according to breast cancer molecular subtype. Cancer Biol Med 2016;13:496–504.

[27] Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009;26:139–40.

[28] Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, et al. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. Cancer Cell 2017;32:57-70.e3.

[29] Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 2018;46:W486–94.

[30] Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology 2015;61:1945–56.

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

[32] Raphael B, Hruban R, Aguirre A, Moffitt R, Yeh J, Stewart C, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell 2017;32:185–203.

[33] Li H, Ning S, Ghandi M, Kryukov G V., Gopal S, Deik A, et al. The landscape of cancer cell line metabolism. Nat Med 2019;25:850–60.

[34] Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A, Kato M, et al. Genomic spectra of biliary tract cancer. Nat Genet 2015;47:1003–10.

[35] 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;7:1–20.

[36] Sia D, Losic B, Moeini A, Cabellos L, Hao K, Revill K, et al. Massive parallel sequencing uncovers actionable FGFR2–PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. Nat Commun 2015;6:6087.

[37] Lee H, Wang K, Johnson A, Jones DM, Ali SM, Elvin J a, et al. Comprehensive genomic profiling of extrahepatic cholangiocarcinoma reveals a long tail of therapeutic targets. J Clin Pathol 2016;69:403–8.

[38] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 2017;171:1437-1452.e17.