UNIVERSITY OF VERONA

Ph.D. School of Health and Life Sciences

Ph.D. Program in Inflammation, Immunity and Cancer

Cycle: XXXII

2018-2019

Doctoral Thesis

Integrating Clinical Data and Molecular Profiling of Hepato-Pancreato-Biliary Cancers: a Surgical-pathological Approach

Tutors: **Prof. Aldo Scarpa**

Prof. Alfredo Guglielmi

PhD Candidate: **Dr. Fabio Bagante**

The World-Famous Author, Snoopy.

Abstract

Background and Aims: The Cancer Genome Atlas (TGCA project has recently published a flagship paper reporting that Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer including hepato-pancreatic and biliary (HPB) malignancies. The aim of the current project was to investigate the molecular landscape of HPB cancers to apply in the clinical practice the molecular classifications resulting from the TGCA analyses.

Patients and Methods:

Machine learning models (artificial neural network, ANN) were trained to predict the molecular subtypes and Cell-of-Origin (iCluster) of HPB cancers. A survival analysis was performed using Cox's survival models and machine learning models (Random Survival Forest, RSF) to investigate impact of the molecular subtypes and iClusters classifications on prognosis of HPB patients. Whole exome sequencing (WES) data of TGCA patients with cholangiocarcinoma (CHOL), liver hepatocellular carcinoma (LIHC), and pancreatic adenocarcinoma (PAAD) were used to develop the ANNs. Two control groups including patients with gastrointestinal cancers and other type of cancers were used to train the ANNs. WES data of patients who underwent surgery at the Ohio State University (OSU) for HPB cancers and of patients participating to the International Cancer Gene Consortium (ICGC) were used to validate the ANNs.

Results:

The ANNs predicting the iClusters (i.e. from iCluster1 to iCluster28) demonstrated an accuracy of 99% in training set versus 74% in the test set. The ANNs predicting the molecular subtypes demonstrated an accuracy of 99% in training set versus 81% in the test set. The survival data of 362 (34 TGCA, 17 OSU, and 311 ICGC) CHOL patients were investigated using the RSF algorithm. The model identified the most important variables as AJCC stage, TP53 pathways status, molecular subtypes, lymph node status, and iCluster. In the multivariable Cox model, AJCC stage, TP53 pathways status, molecular subtypes, and iCluster were associated with patients' survival. Compared with METH-3 patients, patients in IDH and METH-2 subgroups

had almost 2.5- and 5-fold risk of death (IDH, HR 2.47, p=0.037; METH-2, HR 4.85, p<0.001). The c-index of the final model integrating clinical and molecular data resulted 0.72.

A total of 598 (341 TGCA, 30 OSU, and 227 ICGC) LIHC patients were investigated using the RSF algorithm. The model identified the most important variables as AJCC stage, molecular subtypes, AJCC T stages, TP53 pathway status, and TGF-beta pathway status. In the multivariable Cox model, AJCC stage, TP53 pathways status, and molecular subtypes were associated with patients' survival. Compared with patients with other molecular subtypes, patients in i-Cluster2 had almost 2.2-fold increased risk of death (i-Cluster2, HR 2.18, p<0.001). The c-index of the final model was 0.63.

The survival data of 1,022 (155 TGCA, 66 OSU, and 999 ICGC) PAAD patients were investigated using the RSF algorithm. The model identified the most important variables as age, AJCC stage, molecular subtypes, i-Cluster, TP53 pathway, MYC pathway, and Cell-cycle pathway status. In the multivariable Cox model, AJCC stage, TP53 pathways status, and molecular subtypes were associated with patients' survival. Compared with patients with KRAS_wt molecular subtypes, patients with a KRAS_mut PAAD subtype had almost 1.4-fold increased risk of death (KRAS_mut, HR 1.38, p=0.031). The c-index of the final model integrating clinical and molecular data was 0.61.

Conclusion:

TGCA project have reported a complex and interconnected landscape describing the molecular biology of HPB cancers. In this preliminary work, the WES of patients with HPB cancers was used to predict the molecular classifications proposed in the TGCA papers. Moreover, the molecular classifications of HPB malignancies when integrated with the clinical staging system demonstrated to improve our ability to predict the prognosis of HPB patients.

Introduction

The Genome Cancer Atlas (TGCA) project

Since 2005 the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) have supported the Cancer Genome Atlas (TGCA) aiming to obtain a thorough understanding of the genomic alterations that underlie all major cancers. Recently, the TGCA project has published a flagship paper reporting that Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer including hepato-pancreatic and biliary (HPB) malignancies.(1) Following "The Cancer Genome Atlas Pan-Cancer analysis project" papers published in 2014 that included 12 types of different cancers (Pan-Cancer-12), in 2018 a novel TGCA Pan-Cancer analysis reclassified 33 human cancers types based on molecular similarity and reported that the Cell-of-Origin patterns did not completely determine cancers classification but might inform future clinical trial design and interpretation.(1-3) Moreover, several associated papers (Companion Papers) which focused on groups of cancers including gynecologic and breast, gastrointestinal, squamous, and renal cancers, revealed new insights into classifications and characterizations of these subgroups of malignancies.(4-7) In the new Pan-Cancer analysis, the authors performed multi-omics (chromosome-arm-level aneuploidy, DNA hypermethylation, reverse-phase protein arrays, mRNA, and miRNA) molecular clustering, resulting in 28 different integrative clusters (i-Clusters). While i-Clusters were emphasized the dominant role of Cell-of-Origin patterns, clustering was primarily organized by histology, tissue type, or anatomic origin. Molecular similarities among histologically or anatomically related cancer types provide a basis for focused pan-cancer analyses, such as pan-gastrointestinal, pan-gynecological, pan-kidney, and pan-squamous cancers, and those related by stemness features, which in turn may inform strategies for future therapeutic development.(1)

Driver Genes and Oncogenic Signaling Pathways

In order to personalize the surgical and oncological treatment of patients with cancer, a precise classification of tumors and identification of molecular cancer drivers is a fundamental step towards precision oncology. Recently, Bailey and colleagues have reported a comprehensive characterization of cancer drive genes and mutations. The authors reported a PanCancer and PanSoftware analysis spanning 9,423 tumor exomes (comprising all 33 TGCA cancer types) and using 26 computational tools to catalogue driver genes and mutations identifying 299 driver genes with implications regarding their anatomical sites and cancer/cell types. The incidence of single driver genes among different cancer types demonstrated that about half of driver genes (142 genes) were associated with a single cancer, 87 genes had driver roles in two or more cancer types, while about 10% of genes (29 genes) uniquely identified using PanCancer approaches on all samples combined. In particular, *TP53* was identified as cancer driver gene in 27 cancer types, followed by *PIK3CA*, *KRAS*, *PTEN*, and *ARID1A* which were associated with 15 or more cancer types.(8) Moreover, Francisco Sanchez-Vega et colleagues investigated the Oncogenic Signaling Pathways in The Cancer Genome Atlas.(9) Common hallmarks of cancer included mutations in genes members of signaling pathways that control cell-cycle progression, apoptosis, and cell growth. The authors used mutations, copy-number variation, mRNA expression, gene fusions, and DNA methylation in 9,125 tumors included in TCGA to analyze the mechanisms and patterns of somatic alterations in ten canonical pathways including cell cycle, Hippo, Myc, Notch, Nrf2, PI-3-Kinase/Akt, RTK-RAS, TGFβ signaling, p53, and β-catenin/Wnt. The authors reported that among the 33 TGCA cancer types 89% of tumors had at least one driver gene mutated in one of the 10 pathways, and 57% of cancers had at least one mutation potentially targetable by currently available drugs while 30% percent of tumors had multiple targetable alterations amenable of combination therapies.(9)

Molecular Analysis of Gastrointestinal Adenocarcinomas

TGAC data were analyzed to provide a TumorMap with organ systems, integrating single tumors in comprehensive categories. In particular, Yang Liu et colleagues performed the comparative molecular analysis of gastrointestinal (GI) adenocarcinomas. The authors analyzed 921 adenocarcinomas of the esophagus (ESCA), stomach (STAD), colon (COAD), and rectum (READ) to examine shared and distinguishing molecular characteristics of GI tract adenocarcinomas. A total of five molecular subtypes were identified including subgroups of tumors characterized by a high Epstein-Barr virus (EBV) burden, by microsatellite instability (MSI), by hypermutated tumors with single-nucleotide variants (HM-SNV), by chromosomal instability (CIN), and by genome stable (GS). Tumors in the EBV subgroup was founded only in the stomach (n = 30) displaying the most extensive hypermethylation profile of any tumor type in TCGA. Hypermutated tumors (n = 157) were defined by mutation density >10 per megabase (Mb) and classified in two distinct subgroups distinct regardless of cancer type. In particular, cancers enriched for insertions/deletions were included in the MIS subgroups representing microsatellite instability cases with epigenetic silencing of MLH1 in the context of CpG island methylator phenotype. Conversely, hypermutated malignancies characterized by elevated single-nucleotide variants (SNV) associated with mutations in polymerase epsilon(POLE) were included in the POLE/HM-SNV subtype. A total of 734 patients had low mutation density and were classified in the two subtypes CIN and GS. Chromosomal instability (CIN) tumors (n=625, 48% gastro-esophageal cancers; 52% colon-rectum cancers) exhibited marked aneuploidy, that was largely determined by chromosome- and arm-level losses. Finally, the authors identified a group of 109 (47% gastro-esophageal cancers; 53% colon-rectum cancers) patients with a tumor in the colon and rectum lacking hypermutation and aneuploidy termed genome stable (GS) and enriched in DNA hypermethylation and mutations in *KRAS*, *SOX9*, and *PCBP1*.(5)

Molecular Analysis of Hepato-Pancreatic and Biliary Cancers

Even though the integrative analysis of GI cancers provided a classification of cancer overlapping the single tumors and molecular subtypes interconnected between the adenocarcinomas of the esophagus (ESCA), stomach (STAD), colon (COAD), and rectum (READ), this approach was not possible for the hepato-pancreatic and biliary (HPB) cancers. The analyses of the single HPB tumors including cholangiocarcinoma (CHOL), liver hepatocellular carcinoma (LIHC), and pancreatic ductal adenocarcinoma (PAAD) improved our understanding of the biology of these malignancies but did not identify potentially target therapies in common between these tumors as well as a comprehensive classification. The papers investigating single tumor type rather than regrouping these malignancies further subdivided the three HPB tumors (CHOL, LIHC, and PAAD) in 9 molecular subgroups and 15 iClusters.

Molecular Analysis of Cholangiocarcinoma

Farshidfar et colleagues performed an integrative genomic analysis of 38 cholangiocarcinoma (CHOL) identifying distinct IDH-Mutant molecular profiles.(10) The authors described the integrated analysis of gene-level mutations, fusions genes, copy number alterations, mRNA expression, miRNA, lncRNA, protein expression, and DNA methylation of a set of predominantly intrahepatic (32/38, 84%) CHOL cases and propose a molecular classification scheme. The authors identified four distinct molecular subgroups IDH, ECC, METH2, and METH3 subgroups. Isocitrate dehydrogenase (IDH) molecular subtype was characterized by the presence of *IDH* hotspot mutations ($p<0.001$). Three of four distal or perihilar cholangiocarcinoma were in the extrahepatic cholangiocarcinoma (ECC) molecular subtype which was characterized by a wild-type status for *FGFR2*, *ARID1A*, *BAP1*, and *PBRM1*, low methylation, and low incidence of copy number alterations. The methylation cluster 2 (METH2) was enriched for tumors with *CCND1* amplification and with the most highly hypermethylated profile. Finally, the methylation cluster 3 (METH3) was

characterized by high incidence of tumors with BAP1 mutations (8/12; 67%; p=0.01) and all cases of tumors with *FGFR2* fusion genes (5/5; p = 0.004). The study revealed insights into the molecular pathogenesis and heterogeneity of CHOL and provided a classification with potential therapeutic significance.(10)

Molecular Analysis of Hepatocellular Carcinoma

David A. Wheeler et colleagues analyzed 363 liver hepatocellular carcinoma (LIHC) patients with whole-exome sequencing and DNA copy number analyses, as well as a subgroup of 196 patients with available DNA methylation, RNA, miRNA, and proteomic expression data.(11) The authors performed an integrative molecular analysis with an unsupervised clustering of five data platforms (DNA copy number, DNA methylation, mRNA, miRNA, and proteomic expression) and in a joint multivariate regression approach identified three molecular subtypes (iClust1, iClust2, and iClust3).(12) A total of 65 patients were included in the iClust1, characterized by higher tumor grade and presence of macrovascular invasion, low frequency of *CDKN2A* silencing (32%) as compared to iClust2 (69%) and iClust3 (63%). iClust1 was characterized by a low frequency of *CTNNB1* mutation (12%) compared with iClust2 (38%) and iClust3 (43%), and low frequency of *TERT* promoter mutation as well as low *TERT* expression. miRNA expression demonstrated specific characteristics in iClust1 HCC as a high expression of miR-181a (a lipid metabolism regulator) and epigenetic silencing of miR-122. iClust1 was associated with overexpression of proliferation marker genes such as *MYBL2*, *PLK1*, and *MKI67*. A total of 55 and 66 patients were included in iClust2 and iClust3, respectively, and HCC in these subgroups exhibited a high frequency of *CDKN2A* silencing by DNA hypermethylation, a high frequency of *TERT* promoter mutations and *CTNNB1* mutations, and enrichment for *HNF1A* mutation. While iClust2 was associated with low-grade tumors ($p < 0.001$) and less microvascular invasion ($p = 0.01$), iClust3 was characterized by a higher degree of chromosomal instability with distinct 17p loss, high frequency of *TP53* mutation, and hypomethylation of multiple CpG sites.

Molecular Analysis of Pancreatic Ductal Adenocarcinoma

Benjamin J. Raphael et colleagues investigated the genomic characteristic of 150 patients with pancreatic ductal adenocarcinoma (PAAD).(13) Several authors have reported that PAADs are characterized by a low neoplastic cellularity (5-20%) with an important desmoplastic reaction and that this characteristics might confound the molecular analyses of this cancers.(Iacobuzio-Donahue et al., 2002)(Wood and Hruban, 2012) For these reasons, the authors of the TGCA paper performed a whole-exome sequencing (WES; mean coverage 405x) identified recurrent mutations in *KRAS*, *TP53*, *CDKN2A*, *SMAD4*, *RNF43*, *ARID1A*, *TGFβR2*, *GNAS*, *RREB1*, and *PBRM1* but to improve the detection of *KRAS* mutations, codon 12, 13, and 61 hotspots of *KRAS* were sequenced using a microfluidic PCR-based approach with very deep coverage (mean coverage about 30,000x). In addition, significantly mutated genes within the 150 TCGA patients, were sequenced to higher coverage (mean coverage about 644x). This deep sequencing approach identified that 93% of PAADs had a *KRAS* mutation.(13)

Aims

The aim of the current project was to investigate the molecular landscape of HPB cancers and to validate the current molecular classification resulting from the TGCA analyses. A machine learning algorithm – artificial neural network – was used to investigate the reproducibility of the molecular subtypes and Cell-of-Origin (iClusters) classifications for HPB tumors. Moreover, a survival analysis was performed using both the “classical” survival models (i.e. Kaplan Maier survival curves and Cox survival model) and machine learning based approach (Random Survival Forest) were used to analyze the prognostic impact of the molecular subtypes and Cell-of-Origin (iClusters) classifications for HPB cancers.

Methods

The Genome Cancer Atlas data

The Genome Cancer Atlas (TGCA) data were retrieved trough the Multi-Center Mutation Calling in Multiple Cancers (MC3) project including over 10,000 tumor-normal exome pairs across 33 different cancer types.(14) Whole exome sequencing data from patients with a diagnosis of cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM) were analyzed to identify somatic genes mutations. CHOL, LIHC, and PAAD patients were in the hepato-biliary-pancreatic (HPB) cohort, COAD, ESCA, READ, and STAD patients were in the gastrointestinal (GI) cohort while THCA and UVM patients were in the control cohort. In particular, 36 CHOL patients, 353 LIHC patients, 159 PAAD patients (HPB cohort, n= 548), 401 COAD patients, 153 ESCA patients, 148 READ patients, 398 STAD patients (GI cohort, n= 1,100), 467 THCA patients, and 80 patients UVM (control cohort, n= 547) were included in the analyses. Genes mutated with a frequency greater than 3% in a single tumor type were included in the list of genes used as input of the artificial neural network (ANN). All 299 genes identified by Bailey et al. as cancers driver genes were also included in the list of genes for the ANN models.(8) Oncogenic signaling pathways were identified using the analyses of Sanchez-Vega et al. including (1) cell cycle, (2) Hippo signaling, (3) Myc signaling, (4) Notch signaling, (5) oxidative stress response/Nrf2, (6) PI-3-Kinase signaling, (7) receptor-tyrosine kinase (RTK)/RAS/MAP-Kinase signaling, (8) TGFb signaling, (9) p53 and (10) b-catenin/Wnt signaling pathways.(9) All 334 genes involved in the ten pathways were also included in the list of genes for the ANN models. The mutational status of 5,079 genes was selected as input variables for the ANN models.

The Ohio State University data

The first control cohort (the Ohio State University, OSU cohort) included CHOL, LIHC, and PAAD patients who participated at the Oncology Research Information Exchange Network (ORIEN) Avatar Data Collection and Integration Process at the Ohio State University. Whole exome sequencing data from patients in the OSU cohort were included in the analysis. The ORIEN Avatar Data Collection and Integration Process include nucleic acid extraction and a quality control of germline DNA and tumor DNA while library preparation includes a whole exome sequencing (WES) pipeline of the germline and tumor DNA to detect genetic mutation. For the tumor whole exome sequencing (WES T) frozen tissue was analyzed using the Qiagen QIASymphony for DNA and formalin-fixed paraffin-embedded (FFPE) tissue was analyzed using the Qiagen All prep FFPE DNA/RNA kit. Samples were converted to a sequencing-ready library for WES using the Roche Nimblegen SeqCap EZ Exome v3.0 (64Mb, Three Library Hyb) kit. Illumina HiSeq4000 was used to sequence the libraries with a target coverage/reads ratio of 100x for the germline whole exome sequencing (WES G) and 300x for the WES T. Sentieon (v. 201704.03) was the pipeline method for variant calling using the hg19/GRCh37 as reference genome.

Molecular Data Quality Control Parameters for the OSU cohort

Picard was used to generate quality control metrics. Desired depth was 300X for the WES T and samples were consider "failed" if the actual coverage after duplicate removal was less than 100X. A sample with >80% of bases with 20X or more coverage passed the threshold, otherwise, it was considered a "failed" sample. The number of quality variants called was required >20,000 with >20X read depth (to identify when a less complex library was constructed) and a sample "fails" if less than 20,000 variants were detected.

Desired depth was 100X for the WES G and samples were consider "failed" if the actual coverage after duplicate removal is less than 50X. Predicted gender by X-

chromosome SNP heterozygosity and a sample was noted as “flag” if the predicted gender did not match the clinically reported gender. For WES data, single nucleotide polymorphism (SNP) concordance between tumor and germline samples from the same patients was “fail” if the SNP concordance was < 80% (mis-matched sample pair). A total of 96 samples were run per sequencing batch. Quality control was also performed at the batch level to catch any defects/mistakes not obvious at the individual sample level.

The International Cancer Genome Consortium data

The International Cancer Genome Consortium (ICGC) data were used as the second control cohort (ICGC cohort). ICGC cohort include WES data from patients in the Pancreatic Cancer – AU (PACA-AU), Pancreatic Cancer – CA (PACA-CA, Liver Cancer – CN (LICA-CN), Liver Cancer – FR (LICA-FR), Biliary Tract Cancer – JP (BTCA-JP), Biliary Tract Cancer – SG projects (BTCA-SG), and Thyroid Cancer - SA (THCA-SA) projects. In particular, 461 PACA-AU patients, 317 PACA-CA patients, 404 LICA-CN patients, 369 LICA- FR patients, 239 BTCA-JP patients, 71 BTCA-SG patients, 243 THCA-SA patients, were included in the analyses.

Artificial Neural Network

The WES of TCGA patients was used as input of three artificial neural networks to predict anatomical site, i-Cluster, and cancers molecular subtypes. In a preliminary analysis the hyperparameters of the artificial neural network were tuned to identify the most accurate model. In particular, different values of batch size (100, 250, 350), epochs (100, 150, 250), dropout (0%, 10%, 20%, and 50%), kernel initializer (“uniform”, “normal”), optimizer (“Adam”, “Nadam”), activation (“relu”, “elu”), as well as the percentage of patients in the training and test sets (95%/5%, 90%/10%, and 80%/20%) were evaluated. The final hyperparameters resulted a batch size of 250, 250 epochs, a dropout of 20%, a uniform kernel initializer, the Adam algorithm as optimizer, the relu function as activation, and 95% of patients in the training set versus 5% of patients in test sets. The Ohio Super Computer (OSC) center provides

the computational power and storage to run the ANN analyses. OSC's Owens clusters were used to train and test the ANN.(15)

Survival Analysis

The TCGA, OSU, and ICGC data were merged and the survival analysis was performed using both the “classical” survival analysis (Kaplan Meier curves and Cox survival models) and machine learning algorithm (Random survival forest). Continuous variables were summarized as medians with interquartile ranges (IQR) while categorical variables were reported as whole numbers and percentages. The outcome for survival analyses was overall survival (OS), defined as the time interval between the date of surgery and the date of death. Time was censored at the date of last follow-up for living patients. OS estimates were calculated using the Kaplan-Meier method. Cox proportional hazards models were used to evaluate associations between tumor stage and OS. The coefficients from the Cox models were subsequently reported as hazard ratio (HR) with corresponding 95% confidence intervals (CIs). In order to assess the performance of the survival models, the concordance index (C-index) was used.(16)

Statistical Analysis

Continuous variables were reported as median values with interquartile range (IQR), while discrete variables were reported as totals and frequencies. Univariate comparisons were performed using the chi-squared test or Fisher’s exact test as appropriate. A p-value of <0.05 (two-tailed) was considered statistically significant. All analyses were performed using R CRAN software, v. 3.5.1 and Python 3.6 with Anaconda 5.2. The R CRAN packages were used for the analyses included maftool, TCGAmutations. Python packages used for the analyses included keras, sklearn, and tensorflow.

Results

- *Artificial Neural Network for Anatomical Types*

The artificial neural network (ANN) to predict the anatomical types (i.e. CHOL, COAD, ESCA, LIHC, PAAD, READ, STAD, THCA, and UVM) was trained on 10,000 bootstrapped samples and validated on 106 samples. The ANN had 77,937,079 total parameters with 77,937,079 trainable parameters. Overall the ANN had a loss value of 0.47 with an accuracy of 99% in the training set compared with loss value of 2.92 with an accuracy of 67% in the test set.

Anatomical Types and Cholangiocarcinoma

Among cholangiocarcinoma (CHOL) patients, 34 patients were in the training set and the model correctly classify all 34 (34/34; 100.0%) patients (Table 1a). Conversely, two patients were included in the test set and no patients (0/2; 0%) were correctly classified (Table 1b). Similarly, the ANN was used to predict the anatomical origin of 17 CHOL patients in the OSU cohort and 311 CHOL patients in the ICGC cohort with no patients correctly classified (0/17; 0% and 0/311; 0%) (Table 2a and 2b).

Table 1a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - cholangiocarcinoma (CHOL), 34 (100.0%), training set, n= 34

CHOL, n=33	Predicted	Correct
CHOL	34 (100.0%)	34 (100.0%)
COAD		
ESCA		
LIHC		
PAAD		
READ		
STAD		
THCA		
UVM		
Total	34 (100.0%)	34 (100.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 1b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - cholangiocarcinoma (CHOL), test set, n=2

CHOL, n=1	Predicted	Correct
CHOL		
COAD		
ESCA	1 (50.0%)	
LIHC	1 (50.0%)	
PAAD		
READ		
STAD		
THCA		
UVM		
Total	2 (100.0%)	0 (0.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 2a. The Ohio State University - artificial neural network predicting anatomical types - cholangiocarcinoma (CHOL), n=17

CHOL, n=17	Predicted	Correct
CHOL		
COAD	4 (23.5%)	
ESCA	1 (5.9%)	
LIHC	4 (23.5%)	
PAAD	1 (5.9%)	
READ		
STAD		
THCA	7 (41.2%)	
UVM		
Total	17 (100.0%)	0 (0.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 2b. The International Cancer Genome Consortium - artificial neural network predicting anatomical types - cholangiocarcinoma (CHOL), n=311

CHOL, n=311	Predicted	Correct
CHOL	0 (0.0%)	0 (0.0%)
COAD	38 (12.2%)	
ESCA	17 (5.5%)	
LIHC	94 (30.2%)	
PAAD	26 (8.4%)	
READ	34 (10.9%)	
STAD	68 (21.9%)	
THCA	32 (10.3%)	
UVM	1 (0.3%)	
Total	311 (100.0%)	0 (0.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Anatomical Types and Hepatocellular Carcinoma

Among hepatocellular carcinoma (LIHC) patients, 319 were in the training set and the model correctly classify 318 (318/319; 99.7%) patients (Table 3a). A total of 23 patients were in test set and the ANN was able to correctly classify 17 (17/23; 73.9%) patients (Table 3b). Among the 30 OSU LIHC patients, the ANN correctly classified 22 (22/39; 73.3%) patients (Table 4a). In the ICGC cohort, 651 patients had LIHC and the ANN model was able to correctly classify 275 (275/651; 42.2%) patients (Table 4b).

Table 3a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - liver hepatocellular carcinoma (LIHC), training set, n=319

LIHC, n=319	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC	318 (99.7%)	318 (99.7%)
PAAD		
READ		
STAD	1 (0.3%)	
THCA		
UVM		
Total	319 (100.0%)	318 (99.7%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 3b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - liver hepatocellular carcinoma (LIHC), test set, n=23

LIHC, n=23	Predicted	Correct
CHOL		
COAD		
ESCA	2 (8.7%)	
LIHC	17 (73.9%)	17 (73.9%)
PAAD		
READ		
STAD	2 (8.7%)	
THCA	2 (8.7%)	
UVM		
Total	23 (100.0%)	17 (73.9%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 4a. The Ohio State University - artificial neural network predicting anatomical types - liver hepatocellular carcinoma (LIHC), n=30

LIHC, n=30	Predicted	Correct
CHOL		
COAD	2 (6.7%)	
ESCA		
LIHC	22 (73.3%)	22 (73.3%)
PAAD		
READ	1 (3.3%)	
STAD	1 (3.3%)	
THCA	4 (13.3%)	
UVM		
Total	30 (100.0%)	22 (73.3%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 4b. The International Cancer Genome Consortium - artificial neural network predicting anatomical types - liver hepatocellular carcinoma (LIHC), n=651

LIHC, n=651	Predicted	Correct
CHOL		
COAD	98 (15.1%)	
ESCA	20 (3.1%)	
LIHC	275 (42.2%)	275 (42.2%)
PAAD	24 (3.7%)	
READ	16 (2.5%)	
STAD	153 (23.5%)	
THCA	60 (9.2%)	
UVM	5 (0.8%)	
Total	651 (100.0%)	275 (42.2%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Anatomical Types and Pancreatic Ductal Adenocarcinoma

Among pancreatic ductal adenocarcinoma (PAAD) patients, 145 were in the training set and the model correctly classify 143 (143/145; 98.6%) patients (Table 5a). A total of 10 patients were included in the test set and the ANN was able to correctly classify 7 (7/10; 70.0%) patients (Table 5b). Among the 67 OSU PAAD patients, the ANN correctly classified 25 (25/67; 37.3%) patients (Table 6a). In the ICGC cohort, 1,033 patients had PAAD and the ANN model was able to correctly classify 261 (261/1,033; 25.3%) patients (Table 6b).

Table 5a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - pancreatic adenocarcinoma (PAAD), training set, n=145

PAAD, n=145	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC		
PAAD	143 (98.6%)	143 (98.6%)
READ	2 (1.4%)	
STAD		
THCA		
UVM		
Total	145 (100.0%)	143 (98.6%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 5b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - pancreatic adenocarcinoma (PAAD), test set, n=10

PAAD, n=10	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC	1 (10.0%)	
PAAD	7 (70.0%)	7 (70.0%)
READ		
STAD	1 (10.0%)	
THCA	1 (10.0%)	
UVM		
Total	10 (100.0%)	7 (70.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 6a. The Ohio State University - artificial neural network predicting anatomical types - pancreatic adenocarcinoma (PAAD), n=67

PAAD, n=67	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC	3 (4.5%)	
PAAD	25 (37.3%)	25 (37.3%)
READ		
STAD	3 (4.5%)	
THCA	36 (53.7%)	
UVM		
Total	67 (100.0%)	25 (37.3%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 6b. The International Cancer Genome Consortium - artificial neural network predicting anatomical types - pancreatic adenocarcinoma (PAAD), n=1,033

PAAD, n=1,033	Predicted	Correct
CHOL		
COAD	137 (13.3%)	
ESCA	153 (14.8%)	
LIHC	62 (6.0%)	
PAAD	261 (25.3%)	261 (25.3%)
READ	24 (2.3%)	
STAD	374 (36.2%)	
THCA	24 (2.1%)	
UVM		
Total	1,033 (100.0%)	261 (25.3%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Anatomical Types and Gastrointestinal Cancers

Anatomical Types and Esophagus Cancers

Among esophagus cancers (ESCA) patients, 145 were in the training set and the model correctly classify 145 (145/145; 100.0%) patients (Table 7a). A total of 8 patients were included in the test set and the ANN was able to correctly classify 2 (2/8; 25.0%) patients (Table 7b).

Anatomical Types and Gastric Adenocarcinoma

Among gastric adenocarcinoma (STAD) patients, 382 were in the training set and the model correctly classify 382 (382/382; 100.0%) patients (Table 8a). A total of 13 patients were included in the test set and the ANN was able to correctly classify 5 (5/13; 38.5%) patents (Table 8b).

Anatomical Types and Colon Adenocarcinoma

Among colon adenocarcinoma (COAD) patients, 344 were in the training set and the model correctly classify 343 (343/344; 99.7%) patients (Table 9a). A total of 20 patients were included in the test set and the ANN was able to correctly classify 17 (17/20; 85.0%) patients (Table 9b).

Anatomical Types and Rectal Adenocarcinoma

Among rectal adenocarcinoma (READ) patients, 126 were in the training set and the model correctly classify 125 (125/126; 99.2%) patients (Table 10a). A total of 5 patients were included in the test set and no patients were correctly classified by the ANN (0/5; 0%)(Table 10b).

Table 7a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - esophageal carcinoma (ESCA), training set, n=145

ESCA, n=145	Predicted	Correct
CHOL		
COAD		
ESCA	145 (100.0%)	145 (100.0%)
LIHC		
PAAD		
READ		
STAD		
THCA		
UVM		
Total	145 (100.0%)	145 (100.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 7b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - esophageal carcinoma (ESCA), test set, n=8

ESCA, n=8	Predicted	Correct
CHOL		
COAD		
ESCA	2 (25.0%)	2 (25.0%)
LIHC	3 (37.5%)	
PAAD		
READ		
STAD	3 (37.5%)	
THCA		
UVM		
Total	8 (100.0%)	2 (25.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 8a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - stomach adenocarcinoma (STAD), training set, n=382

STAD, n=382	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC		
PAAD		
READ		
STAD	382 (100.0%)	382 (100.0%)
THCA		
UVM		
Total	382 (100.0%)	382 (100.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 8b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - stomach adenocarcinoma (STAD), test set, n=13

STAD, n=13	Predicted	Correct
CHOL		
COAD	2 (15.4%)	
ESCA	3 (23.1%)	
LIHC	3 (23.1%)	
PAAD		
READ		
STAD	5 (38.5%)	5 (38.5%)
THCA		
UVM		
Total	13 (100.0%)	5 (38.5%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 9a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - colon adenocarcinoma (COAD), training set, n=344

COAD, n=344	Predicted	Correct
CHOL		
COAD	343 (99.7%)	343 (99.7%)
ESCA		
LIHC		
PAAD		
READ	1 (0.3%)	
STAD		
THCA		
UVM		
Total	344 (100.0%)	343 (99.7%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 9b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - colon adenocarcinoma (COAD), test set, n=20

COAD, n=20	Predicted	Correct
CHOL		
COAD	17 (85.0%)	17 (85.0%)
ESCA		
LIHC		
PAAD	1 (5.0%)	
READ	1 (5.0%)	
STAD	1 (5.0%)	
THCA		
UVM		
Total	20 (100.0%)	17 (85.0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 10a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - rectum adenocarcinoma (READ), training set, n=126

READ, n=126	Predicted	Correct
CHOL		
COAD		
ESCA	1 (0.8%)	
LIHC		
PAAD		
READ	125 (99.2%)	125 (99.2%)
STAD		
THCA		
UVM		
Total	126 (100.0%)	125 (99.2%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 10b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - rectum adenocarcinoma (READ), test set, n=5

READ, n=7	Predicted	Correct
CHOL		
COAD	5 (100.0%)	0 (0%)
ESCA		
LIHC		
PAAD		
READ		
STAD		
THCA		
UVM		
Total	5 (100.0%)	0 (0%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Anatomical Types and Control Cancers

Anatomical Types and Thyroid Carcinoma

Among thyroid carcinoma (THCA) patients, 439 were in the training set and the model correctly classify 439 (439/1439; 100%) patients (Table 11a). A total of 22 patients were included in the test set and no patients were correctly classified by the ANN (21/22; 95.5%)(Table 11b). In the ICGC cohort, 239 patients had THCA and the ANN model was able to correctly classify 216 (216/239; 90.4%) patients (Table 12).

Anatomical Types and Uveal Melanoma

Among uveal melanoma (UVM) patients, 77 were in the training set and the model correctly classify 125 (76/77; 98.7%) patients (Table 13a). A total of 3 patients were included in the test set and two patients (2/3; 66.7%) were correctly classified by the ANN (Table 13b).

Table 11a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - thyroid carcinoma (THCA), training set, n=439

THCA, n=439	Predicted	Correct
CHOL		
COAD	1 (0.2%)	
ESCA		
LIHC	1 (0.2%)	
PAAD		
READ		
STAD		
THCA	437 (99.5%)	437 (99.5%)
UVM		
Total	439 (100.0%)	437 (99.5%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 11b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - thyroid carcinoma (THCA), test set, n=22

THCA, n=22	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC		
PAAD	1 (5.5%)	
READ		
STAD		
THCA	21 (95.5%)	21 (95.5%)
UVM		
Total	22 (100.0%)	21 (95.5%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 12. The International Cancer Genome Consortium - artificial neural network predicting anatomical types - thyroid carcinoma (THCA), n=239

THCA, n=239	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC	13 (5.4%)	
PAAD	7 (2.9%)	
READ	1 (0.4%)	
STAD		
THCA	216 (90.4%)	216 (90.4%)
UVM	2 (0.8%)	
Total	239 (100.0%)	216 (90.4%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 13a. The Cancer Genome Atlas - artificial neural network predicting anatomical types - uveal melanoma (UVM), training set, n=77

UVM, n=77	Predicted	Correct
CHOL		
COAD		
ESCA	1 (1.3%)	
LIHC		
PAAD		
READ		
STAD		
THCA		
UVM	76 (98.7%)	76 (98.7%)
Total	77 (100.0%)	76 (98.7%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

Table 13b. The Cancer Genome Atlas - artificial neural network predicting anatomical types - uveal melanoma (UVM), test set, n=3

UVM, n=3	Predicted	Correct
CHOL		
COAD		
ESCA		
LIHC		
PAAD		
READ		
STAD	1 (33.3%)	
THCA		
UVM	2 (66.7%)	2 (66.7%)
Total	3 (100.0%)	2 (66.7%)

cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), pancreatic adenocarcinoma (PAAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uveal melanoma (UVM)

- *Artificial Neural Network for iClusters*

The artificial neural network (ANN) to predict the iCluster (i.e. from iCluster 1 to iCluster 28) was trained on 10,000 bootstrapped samples and validated on 106 samples. The ANN had 77,579,658 total parameters with 77,579,658 trainable parameters. Overall the ANN had a loss value of 0.41 with an accuracy of 99% in the training set compared with a loss value of 2.82 with and an accuracy of 74% in the test set.

iClusters among Cholangiocarcinoma

A total of 10 iClusters were identified for cholangiocarcinoma (CHOL) patients. Among CHOL patients, 34 patients were in the training set and the model correctly classify all 33 (33/34; 97.1%) patients (Table 14a). The training set includes patients in 10 iClusters and the ability to predict the correct iCluster ranged between 100% (iClusters 3, 8, 10, 14, 19, 25, 26, and 28) to 85.7% (iCluster 20)(Table 14a). Conversely, two patients were included in the test and one (1/2; 50%) patient was correctly classified (Table 14b). The test set includes patients in 2 iClusters and the ability to predict the correct iCluster ranged was 100% for iClusters 10 while one (1/2; 50%) patients was predicted to be in an "unexpected" iCluster (iCluster 4). Among the 17 OSU CHOL patients, the ANN predicted two patients in "expected" iClusters (2/17; 11.8%; iClusters 10 and 20) while 15 (15/17; 88.2%) patients were in "unexpected iClusters (iClusters 4 and 18)(Table 15a). In the ICGC cohort, 311 patients had CHOL and the ANN model was able to classify in "expected" iCluster 44 (44/311; 14.1%) patients while 267 (267/311; 85.8%) patients were in "unexpected iClusters (iClusters 4, 18, and 27)(Table 15b).

Table 14a. The Cancer Genome Atlas - artificial neural network predicting iClusters - cholangiocarcinoma (CHOL), 34 (100.0%), training set, n= 34

CHOL, n=34	Actual	Predicted	Correct
1			
2			
3	1 (2.9%)	1	1 (100.0%)
4			
5			
7			
8	2 (5.9%)	2	2 (100.0%)
9			
10	1 (2.9%)	1	1 (100.0%)
12			
13			
14	10 (29.4%)	10	10 (100.0%)
15			
17			
18			
19	2 (5.9%)	3	2 (100.0%)
20	7 (20.6%)	6	6 (85.7%%)
21			
22			
25	4 (11.8%))	4	4 (100%)
26	2 (5.9%)	2	2 (100%)
27			
28	5 (14.7%)	5	5 (100%)
Total	34 (100.0%)		33 (97.1%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 14b. The Cancer Genome Atlas - artificial neural network predicting iClusters - cholangiocarcinoma (CHOL), test set, n=2

CHOL, n=2	Actual	Predicted	Correct
1			
2			
3			
4		1	
5			
7	1 (50%)		
8			
9			
10	1 (50%)	1	1 (100%)
12			
13			
14			
15			
17			
18			
19			
20			
21			
22			
25			
26			
27			
28			
Total	2 (100.0%)		1 (50%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 15a. The Ohio State University - artificial neural network predicting iClusters - cholangiocarcinoma (CHOL), n=17

CHOL, n=17	Predicted	Correct
1		
2		
3		
4	13 (76.5%)	
5		
7		
8		
9		
10	1 (5.9%)	1
12		
13		
14		
15		
17		
18	2 (11.8%)	
19		
20	1 (5.9%)	1
21		
22		
25		
26		
27		
28		
Total	17 (100%)	2 (11.8%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 15b. The International Cancer Genome Consortium - artificial neural network predicting iClusters - cholangiocarcinoma (CHOL), n=311

CHOL, n=311	Predicted	Correct
1		
2		
3		
4	188 (60.5%)	
5		
7		
8		
9		
10	13 (4.2%)	13 (4.2%)
12		
13		
14	7 (2.3%)	7 (2.3%)
15		
17		
18	72 (23.2%)	
19	1 (0.3%)	1 (0.3%)
20	13 (4.2%)	13 (4.2%)
21		
22		
25	6 (1.9%)	6 (1.9%)
26		
27	7 (2.3%)	
28	4 (1.3%)	4 (1.3%)
Total	311 (100%)	44 (14.1%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

iClusters among Hepatocellular Carcinoma

A total of 11 iClusters were identified for hepatocellular carcinoma (LIHC) patients. Among LIHC patients, 319 patients were in the training set and the model correctly classify all 33 (318/319; 99.7%) patients (Table 16a). The training set includes patients in 11 iClusters and the ability to predict the correct iCluster ranged between 100% (iClusters 3, 7, 10, 13, 14, 17, 20, 22, 25, and 26) to 50% (iCluster 1)(Table 16a). A total of 23 patients were included in the test set and 21 (21/23; 91.3%) patients were correctly classified (Table 16b). The test set included patients in 2 iClusters and the ability to predict the correct iCluster was 100% for iCluster 26 while no patients were predicted to be in an "unexpected" iClusters. Among the 30 OSU LIHC patients, the ANN predicted 30 patients in "expected" iClusters (30/30; 100%; iClusters 3 and 26) and no patient was in "unexpected iCluster (Table 17a). In the ICGC cohort, 651 patients had LIHC and the ANN model was able to classify in "expected" iClusters 420 (420/651; 64.5%) patients while 231 (231/651; 35.5%) patients were in "unexpected" iClusters (iClusters 4 and 18)(Table 17b).

Table 16a. The Cancer Genome Atlas - artificial neural network predicting iClusters - liver hepatocellular carcinoma (LIHC), training set, n=319

LIHC, n=319	Actual	Predicted	Correct
1	2 (0.6%)	1	1 (50%)
2			
3	4 (1.3%)	4	4 (100%)
4			
5			
7	2 (0.6%)	2	2 (100%)
8			
9			
10	1 (0.3%)	1	1 (100%)
12			
13	1 (0.3%)	1	1 (100%)
14	4 (1.3%)	4	4 (100%)
15			
17	1 (0.3%)	1	1 (100%)
18			
19			
20	11 (3.4%)	12	11 (100%)
21			
22	2 (0.6%)	2	2 (100%)
25	5 (1.6%)	5	5 (100%)
26	286 (89.7%)	286	286 (100%)
27			
28			
Total	319 (100.0%)		318 (99.7%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 16b. The Cancer Genome Atlas - artificial neural network predicting iClusters - liver hepatocellular carcinoma (LIHC), test set, n=23

LIHC, n=23	Actual	Predicted	Correct
1			
2			
3			
4			
5			
7			
8			
9			
10			
12			
13			
14			
15			
17			
18			
19			
20	2 (8.7%)		
21			
22			
25			
26	21 (91.3%)	23	21 (100%)
27			
28			
Total	23 (100.0%)		21 (91.3%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 17a. The Ohio State University - artificial neural network predicting iClusters - liver hepatocellular carcinoma (LIHC), n=30

LIHC, n=30	Predicted	Correct
1		
2		
3	1 (3.3%)	1 (3.3%)
4		
5		
7		
8		
9		
10		
12		
13		
14		
15		
17		
18		
19		
20		
21		
22		
25		
26	29 (96.7%)	29 (96.7%)
27		
28		
Total	30 (100%)	30 (100%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 17b. The International Cancer Genome Consortium - artificial neural network predicting iClusters - liver hepatocellular carcinoma (LIHC), n=651

LIHC, n=651	Predicted	Correct
1		
2		
3		
4	52 (8%)	
5		
7		
8		
9		
10		
12		
13		
14		
15		
17		
18	179 (27.5%)	
19		
20	15 (2.3%)	15 (2.3%)
21		
22		
25		
26	405 (62.2%)	405 (62.2%)
27		
28		
Total	651 (100%)	420 (64.5%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

iClusters among Pancreatic Ductal Adenocarcinoma

A total of 11 iClusters were identified for pancreatic ductal adenocarcinoma (PAAD) patients. Among PAAD patients, 145 patients were in the training set and the model correctly classify 144 (144/145; 99.3%) patients (Table 18a). The training set includes patients in 11 iClusters and the ability to predict the correct iCluster ranged between 100% (iClusters 2, 3, 4, 5, 7, 9, 10, 13, 18, and 22) to 99.2% (iCluster 20)(Table 18a). A total of 10 patients were included in the test set and 9 (9/10; 90%) patients were correctly classified (Table 18b). The test set includes patients in 2 iClusters and the ability to predict the correct iCluster was 100% for iClusters 20 while no patients were predicted to be in an "unexpected" iClusters. Among the 67 OSU PAAD patients, the ANN predicted 66 patients in "expected" iClusters (66/67; 98.5%; iCluster 20) and one patient was in "unexpected" iCluster (iCluster 26)(Table 19a). In the ICGC cohort, 1,033 patients had PAAD and the ANN model was able to classify in "expected" iClusters 1,016 (1,016/1,033; 98.4%) patients while 17 patients were in "unexpected" iClusters (iCluster 26)(Table 19b).

Table 18a. The Cancer Genome Atlas - artificial neural network predicting iClusters - pancreatic adenocarcinoma (PAAD), training set, n=145

PAAD, n=145	Actual	Predicted	Correct
1			
2	1 (0.7%)	1	1 (100%)
3	1 (0.7%)	1	1 (100%)
4	1 (0.7%)	1	1 (100%)
5	3 (2.1%)	3	3 (100%)
7	3 (2.1%)	3	3 (100%)
8			
9	1 (0.7%)	1	1 (100%)
10	2 (1.4%)	2	2 (100%)
12			
13	1 (0.7%)	1	1 (100%)
14			
15			
17			
18	1 (0.7%)	1	1 (100%)
19			
20	128 (88.3%)	127	127 (99.2%)
21			
22	3 (2.1%)	3	3 (100%)
25			
26		1	
27			
28			
Total	145 (100.0%)		144 (99.3%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 18b. The Cancer Genome Atlas - artificial neural network predicting iClusters - pancreatic adenocarcinoma (PAAD), test set, n=10

PAAD, n=10	Actual	Predicted	Correct
1			
2			
3			
4		1	
5			
7			
8			
9			
10			
12			
13	1 (10%)		
14			
15			
17			
18			
19			
20	9 (90%)	9	9 (100%)
21			
22			
25			
26			
27			
28			
Total	10 (100.0%)		9 (90%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 19a. The Ohio State University - artificial neural network predicting iClusters - pancreatic adenocarcinoma (PAAD), n=67

PAAD, n=67	Predicted	Correct
1		
2		
3		
4		
5		
7		
8		
9		
10		
12		
13		
14		
15		
17		
18		
19		
20	66 (98.5%)	66 (98.5%)
21		
22		
25		
26	1 (1.5%)	
27		
28		
Total	67 (100%)	66 (98.5%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 19b. The International Cancer Genome Consortium - artificial neural network predicting iClusters - pancreatic adenocarcinoma (PAAD), n=1,033

PAAD, n=1,033	Predicted	Correct
1		
2	1 (0.1%)	1 (0.1%)
3		
4	111 (10.7%)	111 (10.7%)
5		
7		
8		
9		
10		
12		
13		
14		
15		
17		
18	587 (56.8%)	587 (56.8%)
19		
20	317 (30.7%)	317 (30.7%)
21		
22		
25		
26	17 (1.6%)	17 (1.6%)
27		
28		
Total	1,033 (100%)	1,016 (98.4%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

iClusters among Gastrointestinal Cancers

iClusters among Esophagus Cancers

A total of 10 iClusters were identified for esophagus cancers (ESCA) patients. Among ESCA patients, 145 patients were in the training set and the model correctly classify 145 (145/145; 100%) patients (Table 20a). The training set includes patients in 10 iClusters and the ability to predict the correct iCluster was 100% (iClusters 2, 4, 7, 10, 13, 17, 18, 20, 25, and 27)(Table 20a). A total of 8 patients were included in the test set and 2 (2/8; 25.0%) patients were correctly classified (Table 20b). The test set includes patients in 5 iClusters and the ability to predict the correct iCluster was 100% for iCluster 18 while 2 patients were predicted to be in an "unexpected" iClusters (iCluster 26)(Table 20b).

iClusters among Gastric Adenocarcinoma

A total of 15 iClusters were identified for gastric adenocarcinoma (STAD) patients. Among STAD patients, 382 patients were in the training set and the model correctly classify 380 (380/382; 99.5%) patients (Table 21a). The training set includes patients in 16 iClusters and the ability to predict the correct iCluster ranged from 100% (iClusters 1, 2, 3, 4, 7, 10, 13, 14, 17, 18, 20, 21, 22, 25, and 27) to 80% (iCluster 10)(Table 21a). A total of 13 patients were included in the test set and 7 (7/13; 53.8%) patients were correctly classified (Table 21b). The test set includes patients in 6 iClusters and the ability to predict the correct iCluster ranged from 100% (iCluster 4) to 50% (iCluster 18) while no patients were predicted to be in an "unexpected" iClusters.

iClusters among Colon Adenocarcinoma

A total of 5 iClusters were identified for colon adenocarcinoma (COAD) patients. Among COAD patients, 344 patients were in the training set and the model correctly classify 341 (341/344; 99.1%) patients (Table 22a). The training set includes patients in 5 iClusters and the ability to predict the correct iCluster ranged from 100% (iClusters 2, 4, and 14) to 97.5% (iClusters 18 and 20)(Table 22a). A total of 20 patients were included in the test set and 12 (12/20; 60%) patients were correctly

classified (Table 22b). The test set includes patients in 4 iClusters and the ability to predict the correct iCluster was 85.7% (iClusters 4) while no patients was predicted to be in an “unexpected” iClusters.

iClusters among Rectal Adenocarcinoma

A total of 4 iClusters were identified for rectal adenocarcinoma (READ) patients. Among READ patients, 126 patients were in the training set and the model correctly classify 126 (126/126; 100%) patients (Table 23a). The training set includes patients in 4 iClusters and the ability to predict the correct iCluster was 100% (iClusters 2, 4, 18, and 20)(Table 23a). A total of 5 patients were included in the test set and 3 (3/5; 60%) patients were correctly classified (Table 23b). The test set includes patients in 2 iClusters and the ability to predict the correct iCluster was 100% (iCluster 4) while no patients was predicted to be in an “unexpected” iClusters.

Table 20a. The Cancer Genome Atlas - artificial neural network predicting iClusters - esophageal carcinoma (ESCA), training set, n=145

ESCA, n=145	Actual	Predicted	Correct
1			
2	2 (1.4%)	2	2 (100%)
3			
4	5 (3.4%)	5	5 (100%)
5			
7	3 (2.1%)	3	3 (100%)
8			
9			
10	27 (18.6%)	27	27 (100%)
12			
13	1 (0.7%)	1	1 (100%)
14			
15			
17	2 (1.4%)	2	2 (100%)
18	40 (27.6%)	40	40 (100%)
19			
20	17 (11.7%)	17	17 (100%)
21			
22			
25	40 (27.6%)	40	40 (100%)
26			
27	8 (5.5%)	8	8 (100%)
28			
Total	145 (100.0%)		145 (100%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 20b. The Cancer Genome Atlas - artificial neural network predicting iClusters - esophageal carcinoma (ESCA), test set, n=8

ESCA, n=8	Actual	Predicted	Correct
1			
2			
3			
4		1	
5			
7	1 (12.5%)		
8			
9			
10	2 (25%)	1	
12			
13			
14			
15			
17			
18	2 (25%)	4	2 (100%)
19			
20	1 (12.5%)		
21			
22			
25			
26		2	
27	2 (25%)		
28			
Total	8 (100.0%)		2 (25%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 21a. The Cancer Genome Atlas - artificial neural network predicting iClusters - stomach adenocarcinoma (STAD), training set, n=382

STAD, n=382	Actual	Predicted	Correct
1	31 (8.1%)	32	31 (100%)
2	14 (3.7%)	14	14 (100%)
3	1 (0.3%)	1	1 (100%)
4	23 (6%)	23	23 (100%)
5			
7	11 (2.9%)	11	11 (100%)
8			
9			
10	10 (2.6%)	8	8 (80%)
12			
13	1 (0.3%)	1	1 (100%)
14	1 (0.3%)	1	1 (100%)
15			
17	1 (0.3%)	1	1 (100%)
18	110 (28.8%)	110	110 (100%)
19			
20	165 (43.2%)	166	165 (100%)
21	1 (0.3%)	1	1 (100%)
22	4 (1%)	4	4 (100%)
25	7 (1.8%)	7	7 (100%)
26			
27	2 (0.5%)	2	2 (100%)
28			
Total	382 (100.0%)		380 (99.5%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 21b. The Cancer Genome Atlas - artificial neural network predicting iClusters - stomach adenocarcinoma (STAD), test set, n=13

STAD, n=13	Actual	Predicted	Correct
1	1 (7.7%)		
2	2 (15.4%)		
3			
4	1 (7.7%)	4	1 (100%)
5			
7			
8			
9			
10			
12			
13			
14			
15			
17			
18	2 (15.4%)	1	1 (50%)
19			
20	6 (46.2%)	8	5 (83.3%)
21			
22			
25	1 (7.7%)		
26			
27			
28			
Total	13 (100.0%)		7 (53.8%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 22a. The Cancer Genome Atlas - artificial neural network predicting iClusters - colon adenocarcinoma (COAD), training set, n=344

COAD, n=344	Actual	Predicted	Correct
1			
2	2 (0.6%)	2	2 (100%)
3			
4	238 (69.2%)	239	238 (100%)
5			
7			
8			
9			
10			
12		1	
13			
14	1 (0.3%)	1	1 (100%)
15			
17			
18	80 (23.3%)	79	78 (97.5%)
19			
20	23 (6.7%)	22	22 (95.7%)
21			
22			
25			
26			
27			
28			
Total	344 (100.0%)		341 (99.1%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 22b. The Cancer Genome Atlas - artificial neural network predicting iClusters - colon adenocarcinoma (COAD), test set, n=20

COAD, n=20	Actual	Predicted	Correct
1			
2			
3			
4	14 (70%)	16	12 (85.7%)
5			
7	1 (5%)		
8			
9			
10			
12			
13			
14			
15			
17			
18	1 (5%)	4	
19			
20	4 (20%)		
21			
22			
25			
26			
27			
28			
Total	20 (100.0%)		12 (60%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 23a. The Cancer Genome Atlas - artificial neural network predicting iClusters - rectum adenocarcinoma (READ), training set, n=126

READ, n=126	Actual	Predicted	Correct
1			
2	2 (1.6%)	2	2 (100%)
3			
4	112 (88.9%)	112	112 (100%)
5			
7			
8			
9			
10			
12			
13			
14			
15			
17			
18	4 (3.2%)	4	4 (100%)
19			
20	8 (2.3%)	8	8 (100%)
21			
22			
25			
26			
27			
28			
Total	126 (100.0%)		126 (100%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 23b. The Cancer Genome Atlas - artificial neural network predicting iClusters - rectum adenocarcinoma (READ), test set, n=5

READ, n=5	Actual	Predicted	Correct
1			
2			
3			
4	3 (60%)	5	3 (100%)
5			
7			
8			
9			
10			
12			
13			
14			
15			
17			
18			
19			
20	2 (10%)		
21			
22			
25			
26			
27			
28			
Total	5 (100.0%)		3 (60%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

iClusters among Control Cancers

iClusters among Thyroid Carcinoma

A total of 4 iClusters were identified for thyroid carcinoma (THCA) patients. Among THCA patients, 439 patients were in the training set and the model correctly classify 438 (438/439; 99.8%) patients (Table 24a). The training set includes patients in 4 iClusters and the ability to predict the correct iCluster ranged from 100% (iClusters 12, 14, and 22) to 67% (iCluster 20)(Table 24a). A total of 22 patients were included in the test set and 22 (22/22; 100%) patients were correctly classified (Table 24b). The test set includes patients in one iCluster and the ability to predict the correct iCluster was 100% (iCluster 12) while no patients were predicted to be in an "unexpected" iClusters. In the ICGC cohort, 239 patients had THCA and the ANN model was able to classify in "expected" iCluster 234 (234/239; 97.9%) patients while one patient was in "unexpected" iClusters (iCluster 15)(Table 25).

iClusters among Uveal Melanoma

A total of 3 iClusters were identified for uveal melanoma (UVM) patients. Among UVM patients, 77 patients were in the training set and the model correctly classify 76 (76/77; 98.7%) patients (Table 26a). The training set includes patients in 3 iClusters and the ability to predict the correct iCluster ranged from 100% (iClusters 3 and 15) to 93% (iClusters 5) (Table 26a). A total of 3 patients were included in the test set and 1 (1/3; 33.3%) patients was correctly classified (Table 26b). The test set includes patients in 3 iClusters and the ability to predict the correct iCluster was 33.3% (iCluster 15) while one patient was predicted to be in an "unexpected" iCluster (iCluster 12).

Table 24a. The Cancer Genome Atlas - artificial neural network predicting iClusters - thyroid carcinoma (THCA), training set, n=439

THCA, n=439	Actual	Predicted	Correct
1			
2			
3			
4			
5			
7			
8			
9			
10			
12	433 (98.6%)	433	433 (100%)
13			
14	2 (0.5%)	2	2 (100%)
15			
17			
18			
19			
20	3 (0.7%)	2	2 (66.7%)
21			
22	1 (0.2%)	1	1 (100%)
25		1	
26			
27			
28			
Total	439 (100.0%)		438 (99.8%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 24b. The Cancer Genome Atlas - artificial neural network predicting iClusters - thyroid carcinoma (THCA), test set, n=22

THCA, n=22	Actual	Predicted	Correct
1			
2			
3			
4			
5			
7			
8			
9			
10			
12	22 (100%)	22	22 (100%)
13			
14			
15			
17			
18			
19			
20			
21			
22			
25			
26			
27			
28			
Total	22 (100.0%)		22 (100%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 25. The International Cancer Genome Consortium - artificial neural network predicting iClusters - thyroid carcinoma (THCA), n=239

THCA, n=239	Predicted	Correct
1	1 (0.4%)	
2		
3		
4	3 (1.3%)	
5		
7		
8		
9		
10		
12	234 (97.9%)	234 (97.9%)
13		
14		
15	1 (0.4%)	
17		
18		
19		
20		
21		
22		
25		
26		
27		
28		
Total	239 (100.0%)	234 (97.9%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 26a. The Cancer Genome Atlas - artificial neural network predicting iClusters - uveal melanoma (UVM), training set, n=77

UVM, n=77	Actual	Predicted	Correct
1			
2			
3	5 (6.5%)	5	5 (100%)
4			
5	15 (19.5%)	14	14 (93.3%)
7			
8			
9			
10			
12			
13			
14			
15	57 (74%)	58	57 (100%)
17			
18			
19			
20			
21			
22			
25			
26			
27			
28			
Total	77 (100.0%)		76 (98.7%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

Table 26b. The Cancer Genome Atlas - artificial neural network predicting iClusters - uveal melanoma (UVM), test set, n=3

STAD, n=3	Actual	Predicted	Correct
1			
2			
3			
4			
5		1	
7			
8			
9			
10			
12		1	
13			
14			
15	3 (100%)	1	1 (33.3%)
17			
18			
19			
20			
21			
22			
25			
26			
27			
28			
Total	3 (100.0%)		1 (33.3%)

In gray the iClusters identified in Hoadley KA et al. Cell. 2018;173(2):291-304 e6.

- *Artificial Neural Network for Molecular Subtypes*

The artificial neural network (ANN) was trained to predict the HPB molecular subtypes (i.e. ECC, IDH, METH-2, METH-3, iClust1, iClust2, iClust3, KRAS_mut, and KRAS _wt) while the integrative GI molecular subtypes (i.e. CIN, MSI, GS, ESCC, POLE, and EBV) were used as control groups in the ANN. The ANN was trained on 10,000 bootstrapped samples and validated on 67 samples. The ANN had 77,578,345 total parameters with 77,578,345 trainable parameters. Overall the ANN had a loss value of 0.54 with an accuracy of 99% in the training set compared with loss value of 1.98 with an accuracy of 81% in the training set.

Molecular Subtypes of Cholangiocarcinoma

A total of 4 molecular subtypes were identified for cholangiocarcinoma (CHOL) patients including ECC, IDH, METH 2, and METH-3. Among CHOL patients, 33 patients were in the training set and the model correctly classify all 32 (32/33; 97.0%) patients (Table 27a). The training set includes patients in 4 molecular subtypes and the ability to predict the correct cluster ranged between 100% (ECC, IDH, and METH-2) to 90.9% (METH-3) (Table 27a). Conversely, one patient was included in the test set and was wrongly classified (0/1; 0%)(Table 27b). The test set included patients in one molecular subtype and no patients was predicted to be in an "unexpected" molecular subtype. Among the 17 OSU CHOL patients, the ANN predicted 8 patients in "expected" molecular subtypes (8/17; 47.1%; METH-3) while no patients were in "unexpected" molecular subtypes (Table 28a). In the ICGC cohort, 311 patients had CHOL and the ANN model was able to classify in "expected" molecular subtypes 93 (93/311; 29.9%) patients while 224 (224/311; 72.0%) patients were in "unexpected" molecular subtypes (CIN, MSI, GS, POLE, and ESCC)(Table 28b).

Table 27a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - cholangiocarcinoma (CHOL), training set, n=33

CHOL, n=33	Actual	Predicted	Correct
ECC	6 (18.2%)	6	6 (100.0%)
IDH	9 (27.3%)	9	9 (100.0%)
METH-2	7 (21.2%)	7	7 (100.0%)
METH-3	11(33.3%)	10	10 (90.9%)
iCluster-2		1	
Total	33 (100.0%)		32 (97.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 27b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - cholangiocarcinoma (CHOL), test set, n=1

CHOL, n=1	Actual	Predicted	Correct
ECC			
IDH			
METH-2		1	
METH-3	1(100.0%)		
Total	1(100.0%)		0 (0.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 28a. The Ohio State University - artificial neural network predicting molecular subtypes - cholangiocarcinoma (CHOL), n=17

CHOL, n=17	Predicted	Correct
METH-3	8 (47.1%)	8 (47.1%)
ESCC	1 (5.9%)	
MSI	2 (11.8%)	
CIN	6 (35.3%)	
Total	17 (100.0%)	8 (47.1%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 28b. The International Cancer Genome Consortium - artificial neural network predicting molecular subtypes - cholangiocarcinoma (CHOL), n=311

CHOL, n=311	Predicted	Correct
CIN	116 (37.3%)	
MSI	45 (14.5%)	
GS	2 (0.6%)	
POLE	51 (16.4%)	
ESCC	4 (1.3%)	
ECC	7 (2.3%)	7 (2.3%)
IDH	8 (2.6%)	8 (2.6%)
METH-2	6 (1.9%)	6 (1.9%)
METH-3	72 (23.2%)	72 (23.2%)
Total	311 (100.0%)	93 (29.9%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Molecular Subtypes of Hepatocellular Carcinoma

A total of 3 molecular subtypes (iClust1, iClust2, and iClust3) were identified for liver hepatocellular carcinoma (LIHC) patients. Among LIHC patients, 161 patients were in the training set and the model correctly classify 158 (158/161; 98.1%) patients (Table 29a). The training set includes patients in 3 molecular subtypes and the ability to predict the correct molecular subtype ranged between 100% (iClust1 and iClust3) to 93.9% (iClust2)(Table 29a). A total of 7 patients were included in the test and 3 (3/7; 42.9%) patients were correctly classified (Table 29b). The test set includes patients in 3 molecular subtypes and the ability to predict the correct molecular subtype was 100% for iClust1 while one patient was predicted to be in an "unexpected" molecular subtype (CIN). Among the 30 OSU LIHC patients, the ANN predicted 26 patients in "expected" molecular subtypes (26/30; 86.7%; iClust1, iClust2, and iClust3) and 4 patients were in "unexpected" molecular subtypes (Table 30a). In the ICGC cohort, 651 patients had LIHC and the ANN model was able to classify in "expected" molecular subtypes 380 (380/651; 58.4%) patients while 271 (271/651; 41.6%) patients were in "unexpected" molecular subtypes (CIN, MSI, GS, POLE, ESCC, EBV, KRAS_mut, and KRAS_wt)(Table 30b).

Table 29a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - liver hepatocellular carcinoma (LIHC), training set, n=161

LIHC, n=161	Actual	Predicted	Correct
iCluster-1	58 (36.0%)	58	58 (100.0%)
iCluster-2	49 (30.4%)	46	46 (93.9%)
iCluster-3	54 (33.5%)	56	54 (100.0%)
MSI		1	
Total	161 (100.0%)		158 (98.1%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 29b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - liver hepatocellular carcinoma (LIHC), test set, n=7

LIHC, n=7	Actual	Predicted	Correct
iClust1	1 (14.3%)	2	1 (100.0%)
iClust2	2 (28.6%)	1	0 (0.0%)
iClust3	4 (57.1%)	3	2 (50.0%)
CIN		1	
Total	7 (100.0%)		3 (42.9%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 30a. The Ohio State University - artificial neural network predicting molecular subtypes - liver hepatocellular carcinoma (LIHC), n=30

LIHC, n=30	Predicted	Correct
iClust1	10 (33.3%)	10 (33.3%)
iClust2	9 (30.0%)	9 (30.0%)
iClust3	7 (23.3)	7 (23.3)
MSI	1 (3.3%)	
CIN	2 (6.7%)	
KRAS _wt	1 (3.3%)	
Total	30 (100.0%)	26 (86.7%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 30b. The International Cancer Genome Consortium - artificial neural network predicting molecular subtypes - liver hepatocellular carcinoma (LIHC), n=651

LIHC, n=651	Predicted	Correct
CIN	23 (3.5%)	
MSI	84 (12.9%)	
KRAS_mut	2 (0.4%)	
GS	1 (0.2%)	
POLE	152 (23.3%)	
ESCC	2 (0.4%)	
EBV	2 (0.4%)	
KRAS _wt	5 (0.8%)	
iClust1	173 (26.6%)	173 (26.6%)
iClust2	93 (14.3%)	93 (14.3%)
iClust3	114 (17.5%)	114 (17.5%)
Total	651 (100.0%)	380 (58.4%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Molecular Subtypes of Pancreatic Ductal Adenocarcinoma

A total of 2 molecular subtypes (KRAS_wt and KRAS_mut) were identified for pancreatic ductal adenocarcinoma (PAAD) patients. Among PAAD patients, 148 patients were in the training set and the model correctly classify 146 (146/148; 98.6%) patients (Table 31a). The training set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged between 99% (KRAS_mut) to 97.7% (KRAS_wt)(Table 31a). A total of 7 patients were included in the test set and 6 (6/7; 85.7%) patients were correctly classified (Table 31b). The test set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged from 100% for the KRAS_wt subgroup to 75% for the KRAS_mut subgroups while one patient was predicted to be in an "unexpected" molecular subtype (CIN). Among the 67 OSU PAAD patients, the ANN predicted 67 patients in "expected" molecular subtypes (67/67; 100%; KRAS_mut and KRAS_wt) and no patient was in a "unexpected" molecular subtype (Table 32a). In the ICGC cohort, 1,033 patients had PAAD and the ANN model was able to classify in "expected" molecular subtypes 302 (302/1,033; 29.2%) patients while 731 (731/1,033; 70.8%) patients were in "unexpected" molecular subtypes (CIN, MSI, POLE, and iClust1)(Table 32b).

Table 31a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - pancreatic adenocarcinoma (PAAD), training set, n=148

PAAD, n=148	Actual	Predicted	Correct
KRAS_mut	104 (70.3%)	103	103 (99.0%)
KRAS _wt	44 (29.7%)	43	43 (97.7%)
iClust2		1	
MSI		1	
Total	148 (100.0%)		146 (98.6%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 31b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - pancreatic adenocarcinoma (PAAD), test set, n=7

PAAD, n=7	Actual	Predicted	Correct
KRAS_mut	4 (57.1%)	3	3 (75.0%)
KRAS _wt	3 (42.9%)	3	3 (100.0%)
CIN		1	
Total	7 (100.0%)		6 (85.7%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 32a. The Ohio State University - artificial neural network predicting molecular subtypes - pancreatic adenocarcinoma (PAAD), n=67

PAAD, n=67	Predicted	Correct
KRAS_mut	17 (25.4%)	17 (25.4%)
KRAS _wt	50 (74.6%)	50 (74.6%)
Total	67 (100.0%)	67 (100.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 32b. The International Cancer Genome Consortium - artificial neural network predicting molecular subtypes - pancreatic adenocarcinoma (PAAD), n=1,033

PAAD, n=1,033	Predicted	Correct
KRAS_mut	259 (25.1%)	259 (25.1%)
KRAS _wt	43 (4.2%)	43 (4.2%)
CIN	42 (4.1%)	
MSI	25 (2.4%)	
iClust1	1 (0.1%)	
POLE	663 (64.2%)	
Total	1,033 (100.0%)	302 (29.2%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Molecular Subtypes of Gastrointestinal Cancers

Molecular Subtypes of Esophagus Cancers

A total of 5 molecular subtypes (CIN, MSI, GS, POLE, and ESCC) were identified for esophagus cancers (ESCA) patients. Among ESCA patients, 140 patients were in the training set and the model correctly classify 140 (140/140; 100%) patients (Table 33a). The training set included patients in 5 molecular subtypes and the ability to predict the correct molecular subtypes was 100% (CIN, MSI, GS, POLE, and ESCC)(Table 33a). A total of 11 patients were included in the test set and 8 (8/11; 72.7%) patients were correctly classified (Table 33b). The test set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtype ranged from 100% for the CIN subgroup to 62.5% for the ESCC subgroup while no patients was predicted to be in an "unexpected" molecular subtype.

Molecular Subtypes of Gastric Adenocarcinoma

A total of 5 molecular subtypes (CIN, MSI, GS, POLE, and EBV) were identified for gastric adenocarcinoma (STAD) patients. Among STAD patients, 352 patients were in the training set and the model correctly classify 349 (349/352; 99.1%) patients (Table 34a). The training set included patients in 5 molecular subtypes and the ability to predict the correct molecular subtypes ranged between 100% (MSI, GS, and POLE) to 92.9% (EBV)(Table 34a). A total of 23 patients were included in the test set and 20 (20/23; 87.0%) patients were correctly classified (Table 34b). The test set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged from 100% for the CIN and MSI subgroups to 0% for the GS and EBV subgroups while no patients was predicted to be in an "unexpected" molecular subtype.

Molecular Subtypes of Colon Adenocarcinoma

A total of 4 molecular subtypes (CIN, MSI, GS, and POLE) were identified for COAD patients. Among colon adenocarcinoma (COAD) patients, 327 patients were

in the training set and the model correctly classify 327 (327/327; 100%) patients (Table 35a). The training set includes patients in 4 molecular subtypes and the ability to predict the correct molecular subtypes was 100% (CIN, MSI, GS, and POLE)(Table 35a). A total of 11 patients were included in the test set and 11 (11/11; 100%) patients were correctly classified (Table 35b). The test set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes was 100% for the CIN and MSI subgroup while no patients was predicted to be in an "unexpected" molecular subtype.

Molecular Subtypes of Rectal Adenocarcinoma

A total of 4 molecular subtypes (CIN, MSI, GS, and POLE) were identified for READ patients. Among rectal adenocarcinoma (READ) patients, 110 patients were in the training set and the model correctly classify 110 (110/110; 100%) patients (Table 36a). The training set includes patients in 4 molecular subtypes and the ability to predict the correct molecular subtypes was 100% (CIN, MSI, GS, and POLE)(Table 36a). A total of 7 patients were included in the test set and 6 (6/7; 85.7%) patients were correctly classified (Table 36b). The test set included patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged from 100% for the CIN subgroup to 0% for the GS subgroup while no patients were predicted to be in an "unexpected" molecular subtype.

Table 32a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - esophageal carcinoma (ESCA), training set, n=140

ESCA, n=140	Actual	Predicted	Correct
CIN	64 (45.7%)	64	64 (100.0%)
MSI	2 (1.4%)	2	2 (100.0%)
GS	1 (0.7%)	1	1 (100.0%)
ESCC	71 (50.7%)	71	71 (100.0%)
POLE	2 (1.4%)	2	2 (100.0%)
Total	140 (100.0%)		140 (100.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 32b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - esophageal carcinoma (ESCA), test set, n=11

ESCA, n=11	Actual	Predicted	Correct
CIN	3 (27.3%)	6	3 (100.0%)
ESCC	8 (72.7%)	5	5 (62.5%)
Total	11 (100.0%)		8 (72.7%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 33a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - stomach adenocarcinoma (STAD), training set, n=352

STAD, n=352	Actual	Predicted	Correct
CIN	204 (58.0%)	203	203 (99.5%)
MSI	70 (19.9%)	72	70 (100.0%)
GS	43 (12.2%)	43	43 (100.0%)
POLE	7 (2.0%)	7	7 (100.0%)
EBV	28 (8.0%)	26	26 (92.9%)
iClust3		1	
Total	352 (100.0%)		349 (99.1%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 33b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - stomach adenocarcinoma (STAD), test set, n=23

STAD, n=23	Actual	Predicted	Correct
CIN	18 (78.3%)	21	18 (100.0%)
MSI	2 (8.7%)	2	2 (100.0%)
GS	1 (4.3%)		0 (0.0%)
EBV	2 (8.7%)		0 (0.0%)
Total	23 (100.0%)		20 (87.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 34a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - colon adenocarcinoma (COAD), training set, n=327

COAD, n=327	Actual	Predicted	Correct
CIN	218 (66.7%)	218	218 (100.0%)
MSI	56 (17.1%)	56	56 (100.0%)
GS	48 (14.7%)	48	48 (100.0%)
POLE	5 (1.5%)	5	5 (100.0%)
Total	327 (100.0%)		327 (100.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 34b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - colon adenocarcinoma (COAD), test set, n=11

COAD, n=11	Actual	Predicted	Correct
CIN	8 (72.7%)	8	8 (100.0%)
MSI	3 (27.3%)	3	3 (100.0%)
Total	11 (100.0%)		11 (100.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 35a. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - rectum adenocarcinoma (READ), training set, n=110

READ, n=110	Actual	Predicted	Correct
CIN	95 (86.4%)	95	95 (100.0%)
MSI	3 (2.7%)	3	3 (100.0%)
GS	8 (7.3%)	8	8 (100.0%)
POLE	4 (3.6%)	4	4 (100.0%)
Total	110 (100.0%)		110 (100.0%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Table 35b. The Cancer Genome Atlas - artificial neural network predicting molecular subtypes - rectum adenocarcinoma (READ), test set, n=7

READ, n=7	Actual	Predicted	Correct
CIN	6 (85.7%)	7	6 (100.0%)
GS	1 (14.3%)		0 (0.0%)
Total	7 (100.0%)		6 (85.7%)

high Epstein-Barr virus (EBV) burden, microsatellite instability (MSI), hypermutated tumors with single-nucleotide variants (HM-SNV/POLE), chromosomal instability (CIN), genome stable (GS), extrahepatic cholangiocarcinoma (ECC), Isocitrate dehydrogenase (IDH), methylation cluster 2 (METH2), methylation cluster 3 (METH3), HCC integrative cluster 1 (iClust1), HCC integrative cluster 2 (iClust2), HCC integrative cluster 3 (iClust3), PAAD KRAS mutated cluster (KRAS_mut), PAAD KRAS wild type cluster (KRAS_wt)

Survival Analysis

Survival Analysis of Cholangiocarcinoma Patients

The survival outcome of 362 (34 TGCA, 17 OSU, and 311 ICGC) patients who underwent surgery for cholangiocarcinoma (CHOL) was investigated using the Random Survival Forest algorithm. The model identified the most important variables as AJCC stage, TP53 pathways status, molecular subtypes, lymph node status, and iCluster (Figure 1). In the multivariable Cox model, AJCC stage, TP53 pathways status, molecular subtypes, and iCluster were confirmed as independent predictor of survival (Table 36). In particular, patients with mutations in the TP53 pathways had a 48% increased risk of death compared with patients without mutations in the TP53 pathways (HR 1.48, 95% CI, 1.02-2.15, p=0.037). Compared with the METH-3 subgroup, patients in IDH and METH-2 subgroups had almost 2.5- and 5-fold risk of death (IDH, HR 2.47, 95% CI, 1.06-5.78, p=0.037; METH-2, HR 4.85, 95% CI, 2.20-10.7, p<0.001). The c-index of the final model integrating clinical and molecular data resulted 0.72.

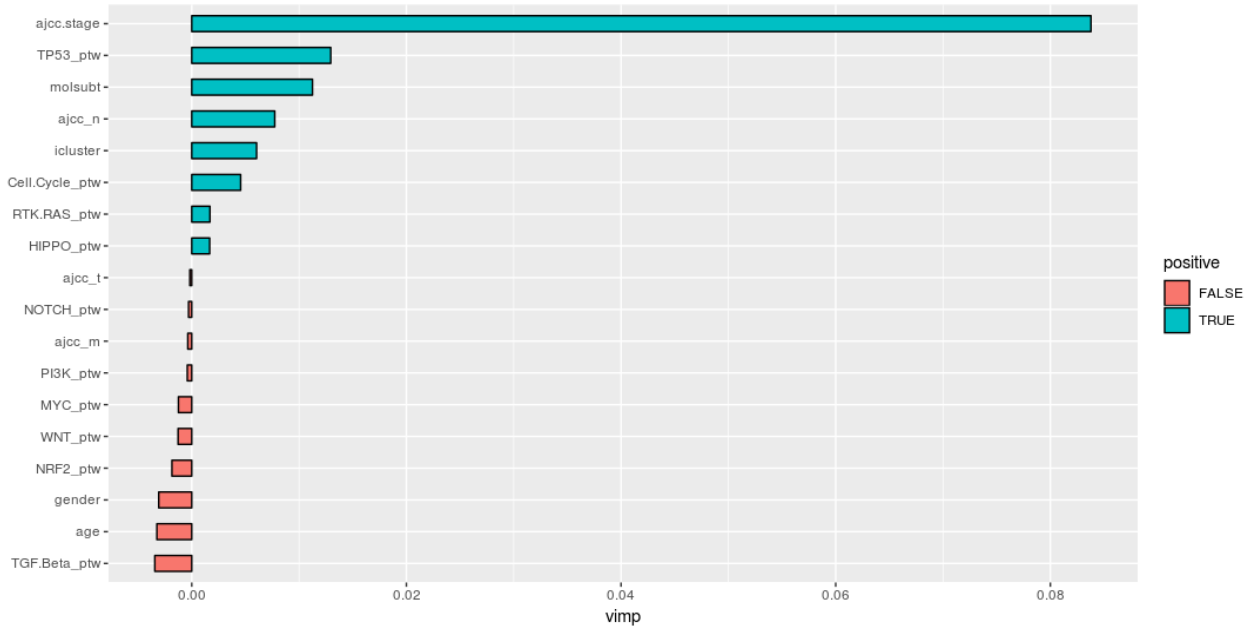

Figure 1. Variable importance rank resulting from the Random Survival Forest model for cholangiocarcinoma patients.

Table 36. Multivariable survival analysis: Cox model for cholangiocarcinoma patients

Variable	Multivariable Analysis		
	HR	**95% CI**	**p-value**
AJCC			
Stage I	Ref.	-	-
Stage II	1.89	0.99-3.62	0.053
Stage III	4.31	2.24-8.31	<0.001
Stage IV	6.92	3.76-12.7	<0.001
Not Staged	4.50	1.26-16.1	0.021
iCluster			
Other CHOL iClusters	Ref.	-	-
Non-CHOL iClusters	0.56	0.36-0.88	0.013
iCluster 14	0.66	0.25-1.76	0.41
iCluster 20	0.55	0.25-1.21	0.14
iCluster 26	4.51	0.92-22.0	0.063
TP53_ptw			
No	Ref.	-	-
Yes	1.48	1.02-2.15	0.037
Molecular Subtypes			
METH-3	Ref.	-	-
IDH	2.47	1.06-5.78	0.037
METH-2	4.85	2.20-10.67	<0.001
ECC	1.99	0.76-5.26	0.16
Non-CHOL Subtypes	1.96	1.21-3.16	0.006

Survival Analysis of Hepatocellular Carcinoma Patients

The survival outcome of 598 (341 TGCA, 30 OSU, and 227 ICGC) patients who underwent surgery for liver hepatocellular carcinoma (LIHC) was investigated using the Random Survival Forest algorithm. The model identified the most important variables as AJCC stage, molecular subtypes, AJCC T stages, TP53 pathway status, and TGF-beta pathway status (Figure 2). In the multivariable Cox model, AJCC stage, TP53 pathways status, and molecular subtypes were confirmed as independent predictor of survival (Table 37). In particular, patients with mutations in the TP53 pathways had a 52% increased risk of death compared with patients without mutations in the TP53 pathways (HR 1.52, 95% CI, 1.15-2.01, p=0.003). Compared with patients with other molecular subtypes, patients in iClust2 had almost 2.2-fold increased risk of death (iClust2, HR 2.18, 95% CI, 1.35-3.52, p<0.001). The c-index of the final model integrating clinical and molecular data resulted 0.63.

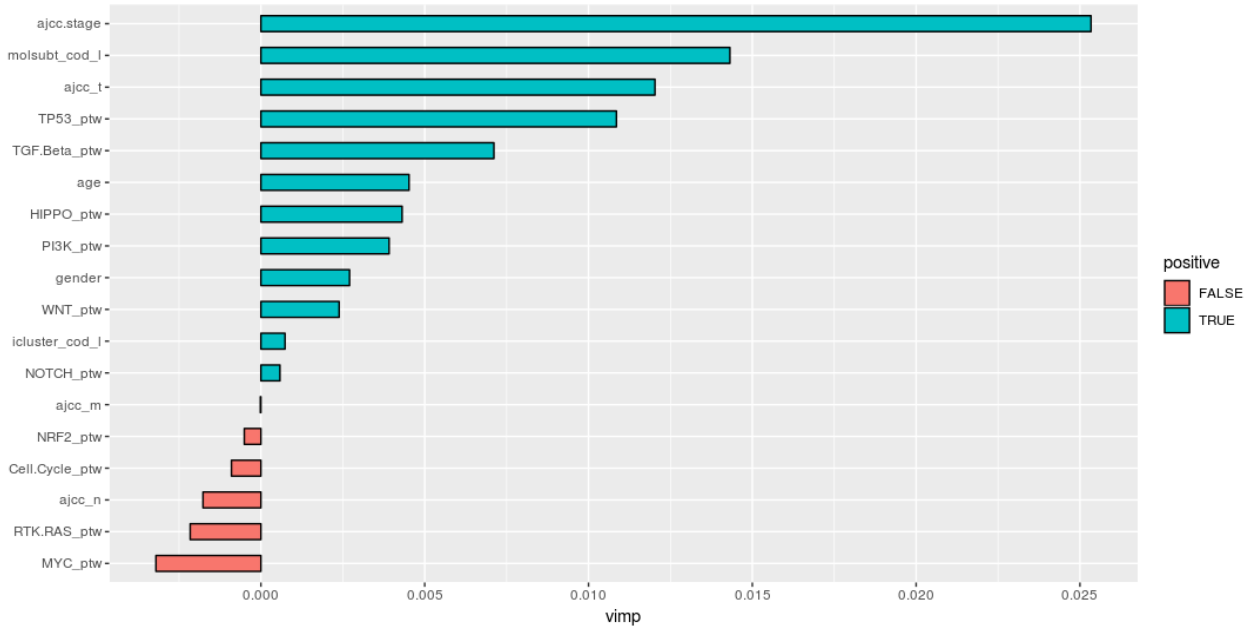

Figure 2. Variable importance rank resulting from the Random Survival Forest model for hepatocellular carcinoma patients.

Table 37. Multivariate survival analysis for LIHC

Variable	Multivariable Analysis		
	HR	**95% CI**	**p-value**
AJCC			
Stage I	Ref.	-	-
Stage II	1.04	0.72-1.51	0.82
Stage III	1.72	1.25-2.38	<0.001
Stage IV	2.71	1.61-4.59	<0.001
Not Staged	2.73	0.94-7.91	0.064
TP53_ptw			
No	Ref.	-	-
Yes	1.52	1.15-2.01	0.003
Molecular Subtypes			
Other Molecular Subtypes	Ref.	-	-
iClust1	1.78	1.14-2.79	0.011
iClust2	2.18	1.35-3.52	<0.001
iClust3	1.98	1.22-3.21	<0.001

Survival Analysis of Pancreatic Ductal Adenocarcinoma Patients

The survival outcome of 1,022 (155 TGCA, 66 OSU, and 999 ICGC) patients who underwent surgery for pancreatic ductal adenocarcinoma (PAAD) was investigated using the Random Survival Forest algorithm. The model identified the most important variables as age, AJCC stage, molecular subtypes, i-Cluster, TP53 pathway status, MYC pathway status, and Cell-cycle pathway status (Figure 3). In the multivariable Cox model, AJCC stage, TP53 pathways status, and molecular subtypes were confirmed as independent predictor of survival (Table 38). In particular, patients with mutations in the TP53 pathways had a 62% increased risk of death compared with patients without mutations in the TP53 pathways (HR 1.62, 95% CI, 1.32-1.99, $p<0.001$). Compared with patients with KRAS_wt molecular subtypes, patients with a KRAS_mut PAAD subtype had almost 1.4-fold increased risk of death (KRAS_mut, HR 1.38, 95% CI, 1.03-1.85, $p=0.031$). The c-index of the final model integrating clinical and molecular data resulted 0.61.

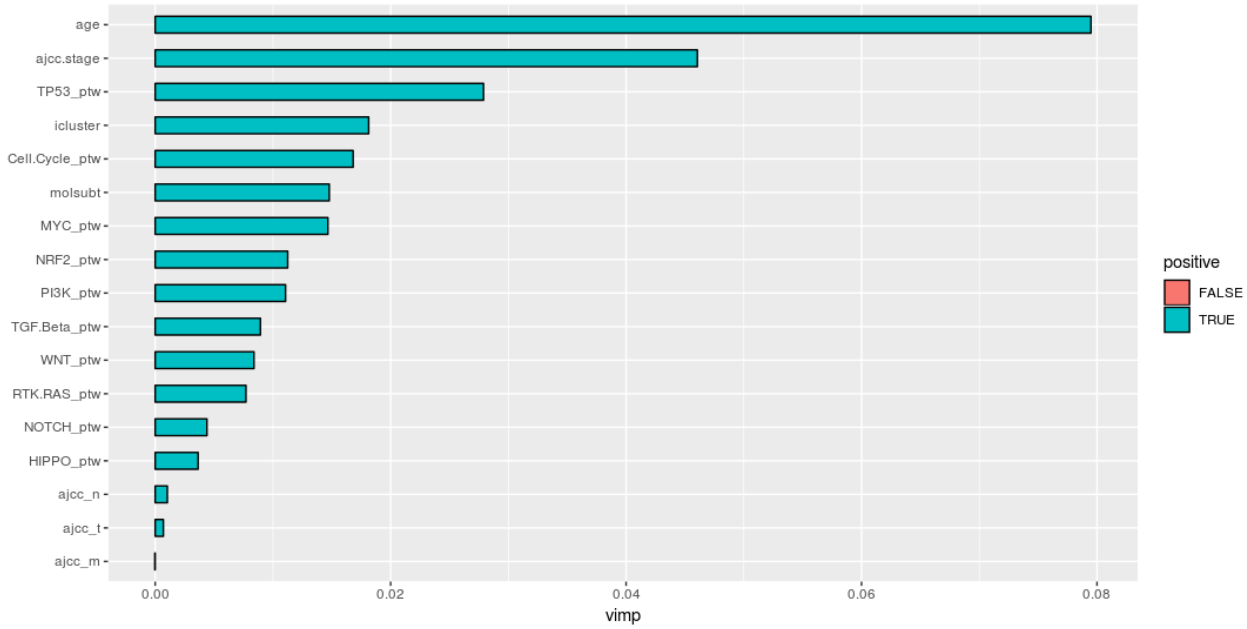

Figure 3. Variable importance rank resulting from the Random Survival Forest model for pancreatic ductal adenocarcinoma patients.

Table 38. Multivariate survival analysis for PAAD

Variable	Multivariable Analysis		
	HR	**95% CI**	**p-value**
Age	1.02	1.01-1.03	<0.001
AJCC			
Stage I	Ref.	-	-
Stage II	1.53	1.23-1.91	<0.001
Stage III	2.19	1.44-3.33	<0.001
Stage IV	1.92	1.25-2.96	0.002
Not Staged	1.36	1.10-1.68	0.004
TP53_ptw			
No	Ref.	-	-
Yes	1.62	1.32-1.99	<0.001
Molecular Subtypes			
KRAS_wt	Ref.	-	-
KRAS_mut	1.38	1.03-1.85	0.031
Other Molecular Subtypes	1.42	1.06-1.89	0.019

Discussion

Three artificial neural networks (ANNs) were developed to predict the anatomical types (i.e. CHOL, COAD, ESCA, LIHC, PAAD, READ, STAD, THCA, and UVM), the iClusters (i.e. from iCluster 1 to iCluster 28) representing the Cell-of-Origin, and the tumor molecular subtypes (i.e. ECC, IDH, METH-2, METH-3, iClust1, iClust2, iClust3, KRAS_mut, and KRAS _wt, CIN, MSI, GS, ESCC, POLE, and EBV). The networks were training using 10,000 bootstrapped samples and validated on 106 samples (67 samples for the molecular subtypes) from the TGCA dataset. While the input of the ANN was the whole exome sequencing of TGCA patients (plus the origin of tissue for the iClusters and the tumor molecular subtypes ANNs) with almost 5,100 variables, the final models had about 78,000,000 parameters. As expected, the accuracy of the three ANNs was 99% in training set but in the test set the accuracy decreased to 67% and 74% for the anatomical type and iClusters ANNs, respectively. Conversely, the ANN for the molecular subtypes demonstrated an accuracy of 81% in the training set.

Cholangiocarcinoma

A total of 38 patients with cholangiocarcinoma (CHOL) were included in the original TGCA paper and whole exome sequencing data which were used to train/test the ANN was available for 34 patients. A total of 10 iClusters and 4 molecular subtypes (ECC, IDH, METH 2, and METH-3) were identified for CHOL patients. In the training set, the ANNs for anatomical types, iClusters, and for the tumor molecular subtypes demonstrated an accuracy of 100% (34/34 patients), 97.1% (33/34 patients), and 97.0% (32/33 patients), respectively. In the test set, the ANNs demonstrated a very low ability to predict the correct anatomical origin (0%, 0/2 patients), molecular subtype (0%, 0/1 patients), and the correct iCluster (50%, 1/2 patients). The iCluster classification for cholangiocarcinoma included 10 clusters for 34 patients, the training set included patients in all the 10 clusters and the ability to predict the correct cluster ranged between 100% (iClusters 3, 8, 10, 14, 19, 25, 26, and 28) to 86% (iCluster

20). Conversely, two patients were included in the test but only one (50%, 1/2 patients) patient was correctly classified. The training set includes patients in 4 molecular subtypes and the ability to predict the correct cluster ranged between 100% (ECC, IDH, and METH-2) to 91% (METH-3). Conversely, one patient was included in the test and was wrongly classified (0%, 0/1 patient). The test set includes patients in one cluster and no patients were predicted to be in an “unexpected” molecular subtype. Interesting, the correct anatomical origin was predicted in 0% of patients among the OSU and ICGC cohorts (OSU, 0/17; ICGC 0/311). Among the OSU and ICGC patients, less than the 15% of patients (OSU, 2/17; 11.8%; ICGC 44/311; 14.1%) was in “expected” iClusters, while half of OSU (OSU cohort, 47%, 8/17) and one third of ICGC (ICGC cohort, 30%, 93/311) cholangiocarcinoma patients were in “expected” molecular subtypes. In the survival analysis, the predicted iClusters and molecular subtypes were used to investigate the impact of the biological characteristics of CHOL on the prognosis of patients undergoing surgery. A total of 362 (34 TGCA, 17 OSU, and 311 ICGC) patients were included in the final multivariable Cox model and the a mutation in genes of the TP53 pathways, the iCluster classification, as well as the CHOL molecular subtype demonstrated to improve the ability of the AJCC staging system in predicting the prognosis of CHOL patients. Of note, while in several publications the c-index of AJCC staging system for CHOL has been reported to range between 0.55 and 0.65, the Cox model integrating clinical and molecular characteristics of CHOL resulted in a c-index of 0.72.(17-32)

Hepatocellular Carcinoma

A total of 363 patients with liver hepatocellular carcinoma (LIHC) were included in the original TGCA paper and whole exome sequencing data which were used to train/test the ANN was available for 334 patients. A total of 11 iClusters and 3 molecular subtypes (iClust1, iClust2, and iClust3) were identified for LIHC patients. In the training set, the ANNs for anatomical types, iClusters, and for the tumor molecular subtypes demonstrated an accuracy of 100% (318/319 patients), 100% (318/319 patients), and 98% (158/161 patients), respectively. In the test set, the ANNs demonstrated a moderate to good ability to predict the correct anatomical origin (74%, 17/23 patients) and iCluster (93%, 21/23 patients). Conversely, only 3 (43%) patients among the 7 patients in the test set were assigned to the correct molecular subtype. The iCluster classification for hepatocellular carcinoma included 11 iClusters for 363 patients, the training set included patients in all the 11 clusters and the ability to predict the correct cluster ranged between 100% (iClusters 3, 7, 10, 13, 14, 17, 20, 22, 25, and 26) to 50% (iCluster 1). The test set includes patients in 2 iClusters and the ability to predict the correct iClusters was 100% for iClusters 26 while no patients were predicted to be in an “unexpected” iCluster. The training set includes patients in 3 molecular subtypes and the ability to predict the correct cluster ranged between 100% (iClust1 and iClust3) to 94% (iClust2). Interesting, the correct anatomical origin was predicted in 73% and 42% of patients among the OSU and ICGC cohorts (OSU, 22/39 patients: ICGC 275/651 patients), respectively. All patients in the OSU cohort (100%, OSU 30/30 patients) were assigned to “expected” iClusters compared with two third (65%; ICGC 420/651 patients) of ICGC patients. Among the OSU and ICGC patients, 87% of OUS patients and 58% of ICGC patients were in “expected” molecular subtypes. A total of 598 (341 TGCA, 30 OSU, and 227 ICGC) patients were included in the final multivariable Cox model and the mutation status of genes in the TP53 pathways, the LIHC molecular subtype demonstrated to improve the ability of the AJCC staging system in predicting the prognosis of LIHC patients. Of note, while in several publications AJCC staging system demonstrated to

be sub-optima to stage LIHC patients, the Cox model integrating clinical and molecular characteristics of LIHC resulted in a c-index of 0. 63.(32-35)

Pancreatic Ductal Adenocarcinoma

A total of 150 patients with pancreatic ductal adenocarcinoma (PAAD) were included in the original TGCA paper and whole exome sequencing data which were used to train/test the ANN was available for 155 patients. A total of 11 iClusters and 2 molecular subtypes (KRAS_wt and KRAS_mut) were identified for PAAD patients. In the training set, the ANNs for anatomical types, iClusters, and for the tumor molecular subtypes demonstrated an accuracy of 99% (143/145 patients), 99% (144/145 patients), and 98% (146/148 patients), respectively. In the test set, the ANNs demonstrated a good to optimal ability to predict the correct anatomical origin (70%, 7/10 patients), iClusters (90%, 9/10 patients), and molecular subtypes (86%, 6/7 patients). The iCluster classification for pancreatic ductal adenocarcinoma included 11 iClusters for 155 patients, the training set included patients in all the 11 clusters and the ability to predict the correct cluster ranged between 100% (iClusters 2, 3, 4, 5, 7, 9, 10, 13, 18, and 22) to 99.2% (iCluster 20). The test set includes patients in 2 clusters and the ability to predict the correct cluster was 100% for iClusters 20 while no patients were predicted to be in an “unexpected” iClusters. The training set includes patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged between 99% (KRAS_mut) to 98% (KRAS_wt). The test set includes patients in 2 molecular subtypes and the ability to predict the correct molecular subtypes ranged from 100% for the KRAS_wt subgroup to 75% for the KRAS_mut subgroups while one patient was predicted to be in an “unexpected” molecular subtype (CIN). Interesting, the correct anatomical origin was predicted only in 37% and 25% of patients among the OSU and ICGC cohorts (OSU, 25/67 patients; ICGC 261/1,033 patients), respectively. Almost all patients in the OSU and ICGC cohorts (99%, OSU 66/67 patients; 98%; ICGC 1,016/1,033 patients) were

assigned to “expected” iClusters. While among the OSU PAAD patients, the ANN predicted all patients (100%, 67/67 patients) in “expected” molecular subtype, in the ICGC cohort, the ANN predicted only 29% (302/1,033 patients) of patients in “expected” molecular subtypes. A total of 1,022 (155 TGCA, 66 OSU, and 999 ICGC) patients were included in the final multivariable Cox model and the mutation status of genes in the TP53 pathways as well as the PAAD molecular subtype demonstrated to improve the ability of the AJCC staging system in predicting the prognosis of PAAD patients. Of note, while several publications reported that several factors (i.e. lymph node, tumor grade, tumor size, Ca 19-9 level) are associated with the prognosis of patients with PAAD undergoing surgery, the Cox model integrating clinical and molecular characteristics of PAAD resulted in a c-index of 0. 61.(36-41)

Limitations:

Several limitations should be considered when interpreting the results. While WES TGCA data for HPB malignancies provide a consistent and reliable starting point to develop the artificial neuron network (ANN), when in the preliminary analysis the ANN hyperparameters were tuned to identify the most accurate model, up to 95% of patients were needed in the training set with only 5% of patients in test set. This approach resulted in a low number of patients in the test set for several disease (i.e. only two CHOL patients in the testing set) which might influence the results. For this reason, further analysis is needed to check the concordance between the results of ANN and the molecular characteristics of caners (i.e. driver mutations and altered pathways).

Conclusion:

TGCA project have reported a complex and interconnected landscape describing the molecular biology of HPB cancers. While these results data might guide physicians towards a personalized approach in the management of patients with HPB cancers, the TGCA analyses were based on multi-omics (i.e. chromosome-arm-level aneuploidy, DNA hypermethylation, reverse-phase protein arrays, mRNA, and miRNA) data and might not be easily translated in the clinical practice. Conversely,

the reduction in the price of WES as well as the diffusion of genes panels investigating the mutational status of hundreds of genes are revolutionizing the decision-making process for the treatment of cancers. In this preliminary work, the WES of patients with HPB cancers was used to predict the molecular classifications proposed in the TGCA papers. Moreover, the molecular classifications of HPB malignancies when integrated with the clinical staging system demonstrated to improve our ability to predict the prognosis of HPB patients.

References

1. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018;173(2):291-304 e6.
2. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113-20.
3. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014;158(4):929-44.
4. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell. 2018;33(4):690-705 e9.
5. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell. 2018;33(4):721-35 e8.

6. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, et al. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. Cell Rep. 2018;23(1):194-212 e6.
7. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep. 2018;23(1):313-26 e5.
8. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;173(2):371-85 e18.
9. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173(2):321-37 e10.
10. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. Cell Rep. 2017;18(11):2780-94.
11. Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas Research N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169(7):1327-41 e23.
12. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572-3.
13. Cancer Genome Atlas Research Network. Electronic address aadhe, Cancer Genome Atlas Research N. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2017;32(2):185-203 e13.
14. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst. 2018;6(3):271-81 e7.
15. Ohio Supercomputer Center. 1987. (https://www.osc.edu/supercomputing/hpc)
16. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-7.

17. Bagante F, Spolverato G, Merath K, Weiss M, Alexandrescu S, Marques HP, et al. Intrahepatic cholangiocarcinoma tumor burden: A classification and regression tree model to define prognostic groups after resection. Surgery. 2019;166(6):983-90.
18. Bagante F, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, Maithel SK, et al. Long-term outcomes of patients with intraductal growth sub-type of intrahepatic cholangiocarcinoma. HPB (Oxford). 2018;20(12):1189-97.
19. Ruzzenente A, Bagante F, Ardito F, Campagnaro T, Scoleri I, Conci S, et al. Comparison of the 7th and 8th editions of the American Joint Committee on Cancer Staging Systems for perihilar cholangiocarcinoma. Surgery. 2018;164(2):244-50.
20. Zhang XF, Chakedis J, Bagante F, Chen Q, Beal EW, Lv Y, et al. Trends in use of lymphadenectomy in surgery with curative intent for intrahepatic cholangiocarcinoma. Br J Surg. 2018;105(7):857-66.
21. Zhang XF, Bagante F, Chen Q, Beal EW, Lv Y, Weiss M, et al. Perioperative and long-term outcome of intrahepatic cholangiocarcinoma involving the hepatic hilus after curative-intent resection: comparison with peripheral intrahepatic cholangiocarcinoma and hilar cholangiocarcinoma. Surgery. 2018;163(5):1114-20.
22. Bagante F, Merath K, Squires MH, Weiss M, Alexandrescu S, Marques HP, et al. The Limitations of Standard Clinicopathologic Features to Accurately Risk-Stratify Prognosis after Resection of Intrahepatic Cholangiocarcinoma. J Gastrointest Surg. 2018;22(3):477-85.
23. Bagante F, Spolverato G, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Surgical Management of Intrahepatic Cholangiocarcinoma in Patients with Cirrhosis: Impact of Lymphadenectomy on Peri-Operative Outcomes. World J Surg. 2018;42(8):2551-60.
24. Bagante F, Spolverato G, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Defining Long-Term Survivors Following Resection of Intrahepatic Cholangiocarcinoma. J Gastrointest Surg. 2017;21(11):1888-97.
25. Buettner S, Galjart B, van Vugt JLA, Bagante F, Alexandrescu S, Marques HP, et al. Performance of prognostic scores and staging systems in predicting long-term

survival outcomes after surgery for intrahepatic cholangiocarcinoma. J Surg Oncol. 2017;116(8):1085-95.
26. Kim Y, Moris DP, Zhang XF, Bagante F, Spolverato G, Schmidt C, et al. Evaluation of the 8th edition American Joint Commission on Cancer (AJCC) staging system for patients with intrahepatic cholangiocarcinoma: A surveillance, epidemiology, and end results (SEER) analysis. J Surg Oncol. 2017;116(6):643-50.
27. Bagante F, Spolverato G, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Impact of Morphological Status on Long-Term Outcome Among Patients Undergoing Liver Surgery for Intrahepatic Cholangiocarcinoma. Ann Surg Oncol. 2017;24(9):2491-501.
28. Bagante F, Spolverato G, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Assessment of the Lymph Node Status in Patients Undergoing Liver Resection for Intrahepatic Cholangiocarcinoma: the New Eighth Edition AJCC Staging System. J Gastrointest Surg. 2018;22(1):52-9.
29. Spolverato G, Bagante F, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Comparative performances of the 7th and the 8th editions of the American Joint Committee on Cancer staging systems for intrahepatic cholangiocarcinoma. J Surg Oncol. 2017;115(6):696-703.
30. Conci S, Ruzzenente A, Sandri M, Bertuzzo F, Campagnaro T, Bagante F, et al. What is the most accurate lymph node staging method for perihilar cholangiocarcinoma? Comparison of UICC/AJCC pN stage, number of metastatic lymph nodes, lymph node ratio, and log odds of metastatic lymph nodes. Eur J Surg Oncol. 2017;43(4):743-50.
31. Spolverato G, Bagante F, Ethun CG, Poultsides G, Tran T, Idrees K, et al. Defining the Chance of Statistical Cure Among Patients with Extrahepatic Biliary Tract Cancer. World J Surg. 2017;41(1):224-31.
32. Bagante F, Tran T, Spolverato G, Ruzzenente A, Buttner S, Ethun CG, et al. Perihilar Cholangiocarcinoma: Number of Nodes Examined and Optimal Lymph Node Prognostic Scheme. J Am Coll Surg. 2016;222(5):750-9 e2.

33. Tsilimigras DI, Mehta R, Moris D, Sahara K, Bagante F, Paredes AZ, et al. Utilizing Machine Learning for Pre- and Postoperative Assessment of Patients Undergoing Resection for BCLC-0, A and B Hepatocellular Carcinoma: Implications for Resection Beyond the BCLC Guidelines. Ann Surg Oncol. 2019.
34. Tsilimigras DI, Bagante F, Moris D, Merath K, Paredes AZ, Sahara K, et al. Defining the chance of cure after resection for hepatocellular carcinoma within and beyond the Barcelona Clinic Liver Cancer guidelines: A multi-institutional analysis of 1,010 patients. Surgery. 2019;166(6):967-74.
35. Tsilimigras DI, Bagante F, Sahara K, Moris D, Hyer JM, Wu L, et al. Prognosis After Resection of Barcelona Clinic Liver Cancer (BCLC) Stage 0, A, and B Hepatocellular Carcinoma: A Comprehensive Assessment of the Current BCLC Classification. Ann Surg Oncol. 2019;26(11):3693-700.
36. Chun YS, Pawlik TM, Vauthey JN. 8th Edition of the AJCC Cancer Staging Manual: Pancreas and Hepatobiliary Cancers. Ann Surg Oncol. 2018;25(4):845-7.
37. Pu N, Yin L, Habib JR, Gao S, Hu H, Zhu Y, et al. Optimized modification of the eighth edition of AJCC TNM staging system for resected pancreatic ductal adenocarcinoma. Future Oncol. 2019;15(30):3457-65.
38. Fan Z, Cheng H, Jin K, Gong Y, Huang Q, Xu J, et al. AJCC 7th edition staging classification is more applicable than AJCC 8th edition staging classification for invasive IPMN. World J Surg Oncol. 2019;17(1):137.
39. Park MY, Shin SH, Song KB, Hwang D, Lee JH, Lee YJ, et al. Validation of the eighth edition of the American Joint Committee on Cancer staging system and proposal of an improved staging system for pancreatic ductal adenocarcinoma. Ann Hepatobiliary Pancreat Surg. 2019;23(1):46-55.
40. van Roessel S, Kasumova GG, Verheij J, Najarian RM, Maggino L, de Pastena M, et al. International Validation of the Eighth Edition of the American Joint Committee on Cancer (AJCC) TNM Staging System in Patients With Resected Pancreatic Cancer. JAMA Surg. 2018;153(12):e183617.

41. Kwon W, He J, Higuchi R, Son D, Lee SY, Kim J, et al. Multinational validation of the American Joint Committee on Cancer 8th edition pancreatic cancer staging system in a pancreas head cancer cohort. J Hepatobiliary Pancreat Sci. 2018;25(9):418-27.