

1 Analytical validation of HepatoPredict kit to assess hepatocellular carcinoma
2 prognosis prior to a liver transplantation.

3

4 Maria Gonçalves-Reis, MSc¹, Daniela Proença, BSc¹, Laura P. Frazão, PhD¹, João L. Neto, PhD¹, Sílvia
5 Silva, MD², Hugo Pinto-Marques, MD PhD^{2,3}, José B. Pereira-Leal, PhD¹, Joana Cardoso, PhD^{1*}

6

7 1- Ophiomics – Precision Medicine, Lisbon, Portugal

8 2- Hepato-Biliary-Pancreatic and Transplantation Centre, Curry Cabral Hospital, Centro Hospitalar
9 Universitário de Lisboa Central, Lisbon, Portugal;

10 3 – Chronic Diseases Research Center (CEDOC), NOVA Medical School, Universidade NOVA de Lisboa
11 (NMS/UNL), Lisbon, Portugal

12

13 * - Corresponding author: Joana Cardoso, PhD: jvaz@ophiomics.com; Ophiomics – Precision Medicine,
14 Pólo Tecnológico de Lisboa, Rua António Champalimaud, Lote 1 (Sala 14); 1600-514 Lisbon, Portugal.

15

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

16 Abstract

17 **Background:** The best curative treatment for hepatocellular carcinoma (HCC) is liver transplant (LT), but
18 the limited number of organs available for LT dictates strict eligibility criteria. Despite this patient
19 selection stringency, current criteria often fail in pinpointing patients at risk of HCC relapse and in
20 identifying good prognosis patients that could benefit from a LT. HepatoPredict kit was developed and
21 clinically validated to forecast the benefit of LT in patients diagnosed with HCC. By combining clinical
22 variables and a gene expression signature in an ensemble of machine learning algorithms,
23 HepatoPredict stratifies HCC patients according to their risk of relapse after LT.

24 **Methods:** Aiming at the characterization of the analytical performance of HepatoPredict kit in terms of
25 sensitivity, specificity and robustness, several variables were tested which included reproducibility
26 between operators and between RNA extractions and RT-qPCR runs, interference of input RNA levels
27 or varying reagent levels. The described methodologies, included in the HepatoPredict kit, were tested
28 according to analytical validation criteria of multi-target genomic assays described in guidelines such as
29 ISO201395-2019, MIQE, CLSI-MM16, CLSI-MM17, and CLSI-EP17-A. Furthermore, a new retrained
30 version of the HepatoPredict algorithms is also presented and tested.

31 **Results:** The results of the analytical performance demonstrated that the HepatoPredict kit performed
32 within the required levels of robustness ($p > 0.05$), analytical specificity (inclusivity $\geq 95\%$), and
33 sensitivity (LoB, LoD, linear range, and amplification efficiency between 90 – 110 %). The introduced
34 operator, equipment, input RNA and reagents into the assay had no significant impact on HepatoPredict
35 classifier results. As demonstrated in a previous clinical validation, a new retrained version of the
36 HepatoPredict algorithm still outperformed current clinical criteria, in the accurate identification of HCC
37 patients that more likely will benefit from a LT.

38 **Conclusions:** Despite the variations in the molecular and clinical variables, the prognostic information
39 obtained with HepatoPredict kit and does not change and can accurately identify HCC patients more
40 likely to benefit from a LT. HepatoPredict performance robustness also validates its easy integration
41 into standard diagnostic laboratories.

42

43 **Keywords:** HepatoPredict, Liver cancer, Hepatocellular carcinoma, Liver transplantation, Analytical

44 validation, Multi-target genomic assay, Prognostic test.

45

46 **Background:**

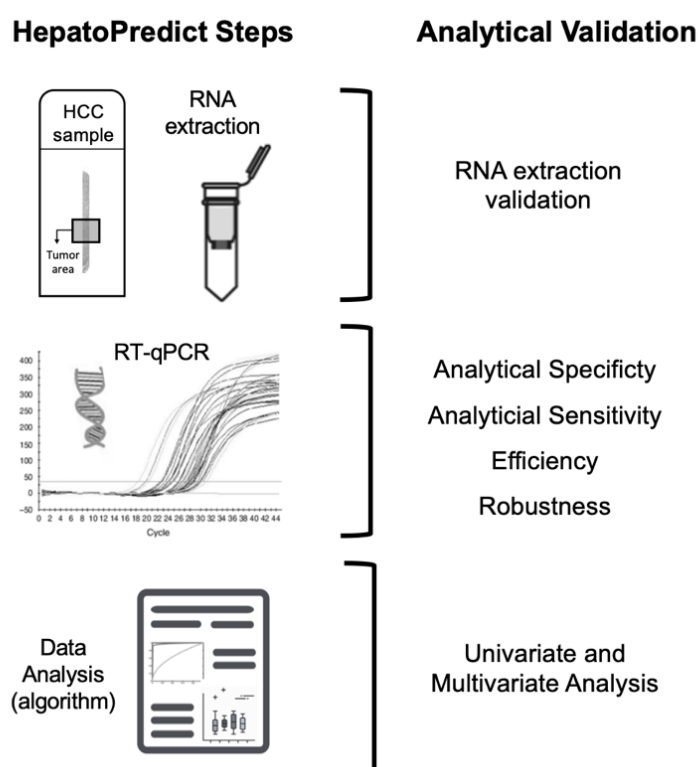
47 Primary liver cancer is the 6th most diagnosed cancer and the third leading cause of cancer death
48 worldwide [1]. Hepatocellular carcinoma (HCC) comprises 75-85 % of primary liver cancer cases and it
49 is associated with chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), aflatoxin-
50 contaminated foods, heavy alcohol intake, excess body weight, type 2 diabetes, and smoking [1]. About
51 30 % of HCC cases are considered for treatment with curative intent [2] which involves liver
52 transplantation (LT) or surgical resection [2,3]. In contrast with surgical resection, LT treats the HCC as
53 well as the underlying cirrhosis, reducing the patient's risk of death within the first 2 years of
54 diagnosis [3]. However, due to the shortage of liver donors, several different criteria mainly based on
55 tumor burden and protein biomarkers such as alpha-fetoprotein (AFP) and des- γ carboxyprothrombin
56 (DCP), have been developed for the identification of HCC patients most likely to benefit from LT [4–13].
57 Nevertheless, the limitations of these criteria are currently under discussion, mainly because they
58 exclude patients with an underlying good prognosis who can benefit from a LT and include bad
59 prognosis patients that will not benefit from the surgery [14,15].

60

61 The HepatoPredict kit intends to predict which patients diagnosed with HCC have a good prognosis and
62 thus will benefit from a LT. This is achieved by combining three clinical variables (tumor number, size
63 of the largest nodule, and total tumor volume) and a gene expression signature (includes *DPT*, *CLU*,
64 *CAPNS1*, and *SPRY2 genes*) and a proprietary algorithm. In short, the HepatoPredict kit can extract RNA
65 from formalin-fixed paraffin-embedded (FFPE) HCC samples and to perform gene expression analysis
66 through real-time quantitative reverse transcription polymerase chain reaction (RT-qPCR) technology.
67 The RT-qPCR results are then combined with the clinical data using a machine learning algorithm that
68 returns three different values concerning the predictive value (Class I and II) or its absence (Class 0) [16]
69 (Figure 1).

70

71 In a previous study [16], using a retrospective clinical validation cohort of patients diagnosed with HCC,
72 we have demonstrated that HepatoPredict outperforms the Milan [4], University of California San
73 Francisco (UCSF) [6], Up-to-seven [11], alpha fetoprotein (AFP) [7], Metroticket 2.0 [5], Total Tumour
74 Volume (TTV) [13], and TTV AFP [12] criteria in the selection of patients suitable for LT¹⁶. Apart from
75 its clinical utility, the technical performance of a prognostic test used in the diagnostic setting must also
76 be verified through its capacity to generate specific, sensitive, and robust data under standard
77 laboratory conditions (analytical validation). The aim of this work is to present evidence of the analytical
78 validation of the HepatoPredict kit based on comprehensive technical studies (**Figure 1**). This validation
79 also included the re-training of the HepatoPredict algorithm using a retrospective cohort of 162
80 patients diagnosed with HCC and submitted to LT.
81
82



83 **Figure 1 – HepatoPredict kit Analytical Validation.** The HepatoPredict kit intends to predict tumor
84 recurrence after a liver transplant in patients with HCC. The HepatoPredict kit uses FFPE HCC samples
85 from which the RNA is extracted and subsequently used as a template in 1-step RT-qPCR reactions

86 targeting reference genes (*RPL13A*, *GAPDH*, *TBP*), genes of interest (*DPT*, *CLU*, *CAPNS1*, *SPRY2*) and a
87 genomic DNA control (*Chr3*). Gene expression levels of the genes of interest are normalized to the
88 geometric mean of the reference genes and combined with clinical variables through an algorithm. The
89 analytical validation of the HepatoPredict kit comprised several different assays performed at different
90 steps of the kit: validation of the RNA extraction methodology, analytical specificity, sensitivity,
91 efficiency, and robustness of RT-qPCR reactions and univariate and multivariate analysis of the
92 HepatoPredict algorithm.

93 **Methods:**

94 Samples: In this study, HCC samples preserved as FFPE tissue were used. FFPE HCC samples were
95 acquired from four different suppliers: Biobank IRBLleida (PT20/00021), integrated in the Spanish
96 National Biobanks Network and Xarxa de Bancs de Tumors de Catalunya (XBTC) sponsored by Pla
97 Director d'Oncologia de Catalunya; Biobank ISABIAL, integrated in the Spanish National Biobanks
98 Network and in the Valencian Biobanking Network; and biorepositories from Amsbio (US) and Biotech
99 (US). All samples were processed following standard operating procedures with the appropriate
100 approval of the Ethical and Scientific Committees. Moreover, clinical samples from a retrospective
101 clinical study approved by the ethics authorities and taking place in the Curry Cabral Hospital (Lisbon,
102 Portugal), were also used. All HCC FFPE samples were acquired either sectioned with 3-5 μm thickness
103 or as paraffin blocks that were then cut in 3-5 μm thick slices using a microtome (Leica SM2010R Sliding
104 Microtome, Leica Biosystems) and mounted on a glass slide.

105

106 Histopathologic analysis: Prior to RNA extraction, HCC FFPE samples were analyzed by a certified
107 pathologist using an hematoxylin and eosin (H&E) stained tumor section. FFPE HCC slides (3 μm thick)
108 were first deparaffinized and stained using Harris Hematoxylin solution (#3801561E, Leica Biosystems,
109 Richmond, USA). The slide was then counterstained with Eosin Y solution (#2801601, Leica Biosystems,
110 Richmond, USA). Finally, slides were dehydrated in increasing alcohol concentrations, cleared in xylene
111 (#28973, VWR, Alfragide, Portugal), and mounted using a xylene-based mounting medium (#107961,
112 Merck, Darmstadt, Germany). After H&E staining, slides were observed under an optical upright
113 microscope (Panther L, #1100104600142, Motic®).

114

115 RNA extraction: For RNA extraction, an HCC area mimicking a needle biopsy was delimited in two
116 sequential 5 μm slides. Samples were initially deparaffinized and the RNA was extracted using the
117 RNeasy FFPE Kit (#73504, Qiagen, Hilden, Germany), according to the manufacturer's instructions with
118 two exceptions: proteinase K cell lysis and final elution volume.

119

120 DNA extraction: DNA was extracted from HCC FFPE samples using the QIAamp DNA FFPE Tissue Kit
121 (#56404, Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions.

122

123 RNA extraction method validation: For the validation of the RNA extraction method, 87 FFPE HCC
124 samples (from the four different suppliers) were used. Each sample was tested in duplicate and by two
125 different operators. Moreover, mirror sections of each sample were used to reduce the variability
126 between operators. Immediately after extraction, RNA was stored at -20 °C until further usage or used
127 straight away in RT-qPCR reactions (conditions described below) targeting *RPL13A* (reference gene) and
128 *Chr3* (genomic DNA control) to analyze the integrity [17] and the purity of the RNA samples,
129 respectively.

130

131 RT-qPCR reactions: 1-step RT-qPCR reactions were performed as previously described [16]. The
132 QuantStudio Design & Analysis Software v1.5.1 software was used for data acquisition and analysis. For
133 gene expression normalization, the geometric mean of the cycle threshold (Cq) of the reference genes
134 (*RPL13A*, *GAPDH*, and *TBP*) was subtracted from the Cq values of the genes of interest (*DPT*, *CLU*,
135 *CAPNS1*, and *SPRY2*).

136

137 Primer Specificity: RT-qPCR products were sequenced via Sanger sequencing outsourced to Eurofins
138 (<https://eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/supremerun-tube/>). In total,
139 16 different solutions (forward and reverse for 8 targets) were sent to Eurofins. Regarding RT-qPCR
140 products, a 2-step RT-qPCR reaction was performed using the SuperScript™ VILO™ cDNA Synthesis Kit
141 (#11754050, Thermo Fisher Scientific, Bleiswijk, Netherlands) to synthesize cDNA and the Invitrogen™
142 Platinum™ SuperFi™ PCR Master Mix with the SuperFi™ GC Enhancer (#12358010, Thermo Fisher
143 Scientific, Bleiswijk, Netherlands) in qPCR. An RNA pool (composed of 8 different FFPE HCC samples)
144 was used as template. The size of each RT-qPCR product was assessed by electrophoresis in a 4 %

145 agarose gel (#G401004, Thermo Fisher Scientific, Bleiswijk, Netherlands), using a DNA ladder
146 (#10488096, Thermo Fisher Scientific, Vilnius, Lithuania) and nuclease-free water (#129114, Qiagen,
147 Hilden, Germany) as a negative control in an electrophoresis system (#G8300, Thermo Fisher Scientific,
148 Vilnius, Lithuania).

149
150 RT-qPCR inclusivity: Different FFPE HCC samples were used as templates in RT-qPCR reactions targeting
151 all the genes included in HepatoPredict kit. Each RT-qPCR reaction was performed in duplicate and by
152 two different operators. For each sample, each operator used the same batch of extracted RNA or DNA.
153 Both nucleic acid extraction and RT-qPCR reactions were performed as described above.

154
155 Limit of Detection (LoD) determination: For LoD determination, a pool composed of 8 FFPE HCC samples
156 was used. The samples composing the pool reflected high and low expression levels of each target and
157 were associated with a bad prognosis (recurrence, n = 4) and a good prognosis (no recurrence, n = 4).
158 Both pools, of DNA and RNA, were created using the same samples. DNA pool was directly used for
159 serial dilutions (at least 11 per target) and the RNA pool was diluted 1:4 to create the starting sample
160 for the serial dilutions. In total, 21 replicates were done for each dilution (triplicates in each of the 7
161 RT-qPCR reactions), per lot number of reagents, on three different days (2-3 RT-qPCR reactions per
162 day). For each target, all reactions were performed by the same operator with the same equipment.
163 RT-qPCR reactions using reagents from different lots were analyzed separately. Data was analyzed in
164 accordance with the Probit model, which implied the creation of a regression representing the
165 probability vs \log_2 dilution for each target assuring at least 3 dilutions with hit rates within 0.10 - 0.90
166 and at least one exceeding 0.95. Moreover, to minimize the influence of the model limit ranges of
167 probability, dilutions with a 100 % fail or success rate were included in each analysis. LoDs were
168 independently calculated for each lot and the maximum LoD (concentration) was taken as the reported
169 value for the measurement procedure. To determine the Cq value associated with the LoD, a linear
170 regression was applied between the Cq values and the \log_2 (dilution factor). All the \log_2 (dilution factor)

171 until the one immediately after the LoD were considered. The values outside the confidence interval
172 (CI) at 99 % were considered outliers and were removed. Linear, quadratic, and cubic polynomial
173 functions were fitted to the Cq values using log₂ dilution values. If none of the non-linear coefficients
174 was different from zero, the target was considered linear (*GAPDH*, *TBP* and *Chr3*). Otherwise, the
175 absolute difference between the model that best fits the data (smallest mean squared error) and the
176 linear model was calculated. When the difference was less than 1 Cq value, the target was considered
177 linear (*RPL13A*, *DPT*, *CAPNS1*, *CLU* and *SPRY2*).

178

179 Linearity: The linear range of each target included in the HepatoPredict kit was determined for RT-qPCR
180 reactions using FFPE HCC samples (previously used for LoD determination) and reference materials
181 (#636690, Takara, Saint Germain en Laye, France) to cover a broad range of nucleic acids
182 concentrations in linearity determination. For that, seven serial dilutions of reference RNA were used
183 with 3 replicates per dilution and at least eleven dilutions of nucleic acids pools, obtained from HCC
184 FFPE samples, were used with 7 replicates per dilution. Finally, the Cq values and the dilution factors
185 were plotted in a base 2 logarithmic graph and R² (> 0.90) was calculated for all targets.

186

187 Amplification efficiency: The reaction efficiency was calculated for each target included in the
188 HepatoPredict kit. It was determined from the slope of the log-linear portion of each target curve:
189 amplification efficiency = $(2^{-1/\text{slope}} - 1) \times 100$.

190

191 Robustness of RT-qPCR reactions: Plackett and Burman tables [18] were used to design the robustness
192 assay: alterations in the concentrations ($\pm 30\%$) of the master mix (#A15300, Thermo Fisher Scientific,
193 Bleiswijk, Germany), primers and probes were implemented as well as different final reaction volumes
194 ($\pm 5\%$) and annealing temperatures ($\pm 1\text{ }^{\circ}\text{C}$), as demonstrated in **Table 1**. Two independent assays per
195 each target were performed using the same sample pool (see LoD) in triplicate and all reactions were
196 performed using sample concentration near the LoD (RNA pool serial dilution 2⁻² and DNA pool without

197 further dilutions). Three standard conditions were incorporated in the assay for data analysis: standard
198 (STD) (no changes), STD1 (-1 °C annealing temperature), and STD2 (+1 °C annealing temperature).

199

200 Determination of Cq values below LoD and within the linear range for each target: Serial dilutions of
201 the reference RNA (#636690, Takara, Saint Germain en Laye, France) were used as templates for RT-
202 qPCR reactions targeting all genes included in the HepatoPredict kit. The Cq values above the LoD and
203 outside the linear range for each target were identified and the maximum acceptable Cq value for each
204 target was determined.

205

206 Precision studies: The conditions under which repeated measurements were made determine the type
207 of precision being analyzed – reproducibility (daily, lot-to-lot, operator, and inter-assay) and
208 repeatability. For the daily reproducibility, for the same sample, assays were performed by the same
209 operator, using the same sample and kit's lot on 4 different days. Regarding lot-to-lot reproducibility,
210 the same sample was analyzed by the same operator using kits from three different lots. Finally,
211 operator reproducibility was studied by using the same sample with HepatoPredict kits from the same
212 lot but performed by three different operators. For each condition, two HepatoPredict kits were used
213 (two independent assays). Repeatability was measured considering the triplicates of each
214 HepatoPredict kit run. In total, 3 different HCC FFPE samples were studied, thus, 48 HepatoPredict kits
215 were used (16 kits/sample). To further assess inter-assay reproducibility, 15 additional samples were
216 tested in duplicate by different operators, using different lots of the HepatoPredict kit, and on different
217 days (total n = 18).

218

219

220 Table 1 – Design of the robustness assay for the HepatoPredict kit.

Factor	Conditions										
	STD	STD1	STD2	1	2	3	4	5	6	7	8
PCR Equipment	A	A	A	A	A	A	A	A	A	A	A
[Master Mix]	NC	NC	NC	+ 30 %	+ 30 %	- 30 %	- 30 %	+ 30 %	+ 30 %	- 30 %	- 30 %
[Primers]	NC	NC	NC	NC	- 30 %	NC	+ 30 %	NC	- 30 %	NC	+ 30 %
[Probe]	NC	NC	NC	NC	- 30 %	+ 30 %	NC	- 30 %	NC	NC	+ 30 %
Final Reaction Volume	NC	NC	NC	- 5 %	- 5 %	+ 5 %	+ 5 %	+ 5 %	+ 5 %	- 5 %	- 5 %
Annealing Temperature	NC	- 1 °C	+ 1 °C	+ 1 °C	- 1 °C	+ 1 °C	- 1 °C	- 1 °C	+ 1 °C	- 1 °C	+ 1 °C

STD – standard. NC – no changes regarding the standard protocol.

221

222 HepatoPredict algorithm training: A dataset with 162 patients diagnosed with HCC and submitted to
223 liver transplant was used (**Supplementary File 1**). Concerning the dataset, different models were tested,
224 such as Naive Bayes, support-vector machine (SVM) with different kernel functions, and Extreme
225 Gradient Booster (XGBoost). Moreover, synthetic minority oversampling technique (SMOTE) was also
226 used for data imbalances. Python 3.8 was used with scikit-learn 1.0.2 (<https://scikit-learn.org/stable/>),
227 XGBoost 1.6.1 (<https://xgboost.readthedocs.io/en/stable/>), imbalance-learn ([https://imbalanced-](https://imbalanced-learn.org/stable/)
228 [learn.org/stable/](https://imbalanced-learn.org/stable/)), and Optuna 2.10.0 (<https://optuna.org/>). Each model was fed with 4 molecular (*DPT*,
229 *CLU*, *CAPNS1*, and *SPRY2* gene expression) and 3 clinical variables (tumor number, largest tumor size
230 and total tumor volume). As previously described [16], the algorithm was developed as a two-level
231 predictor.

232 HepatoPredict algorithm univariate analysis: The univariate analysis of the HepatoPredict algorithm
233 consisted in calculating the error (i.e., counting each time the algorithm would fail the correct prognosis
234 classification) when altering the Cq mean values of each gene (prior normalization) and varying the
235 normalized Cq values of *DPT*, *CLU*, *SPRY2* and *CAPNS1* and the clinical variables (tumor number,
236 diameter of the largest tumor, and total tumor volume). Thus, Cq mean values of the genes of interest
237 (*DPT*, *CLU*, *SPRY2* and *CAPNS1*) were replaced by 40 Cq (the maximum number of cycles allowed) and
238 their respective LoD and the Cq means of the reference genes (*RPL13A*, *GAPDH*, and *TBP*) were
239 removed and replaced by their respective LoD. Regarding the variations of the normalized Cq values
240 (for *DPT*, *CLU*, *SPRY2* and *CAPNS1* genes), variations of 0.1 Cq were performed. Alterations in clinical
241 variables included the variations in the tumor number (1-2 units), and in the diameter of the largest
242 tumor (cm) and the total tumor volume (cm³) by 2 %.

243 HepatoPredict algorithm multivariate analysis: The multivariate analysis of the HepatoPredict algorithm
244 consisted in calculating the error associated with the alteration of more than one variable at a time.
245 Thus, a range for each variable variation was defined (based on algorithm univariate analysis) and
246 random combinations of 2, 3, 4, 5, 6 and 7 altered variables were tested. All possible combinations of
247 variables were performed and for each combination the assay was repeated 10,000 times with random

248 variable alterations within the defined range. Finally, two types of errors were calculated: error type A
249 (between Class I and Class II) and error type B (between Class I or II and Class 0).

250 Statistical analysis: Statistical analysis was performed using the R language for Statistical Computing (v
251 4.1.1) and GraphPad Prism 7 (GraphPad Software, Inc. 2016). For RNA extraction validation, as the data
252 followed a normal distribution, the Paired Student t-test was applied. For robustness assay, Dunn's
253 multiple comparisons test was applied. Regarding precision, due to data size, a non-parametric test
254 (Friedman test) was used. A $p < 0.05$ was considered statistically significant.

255

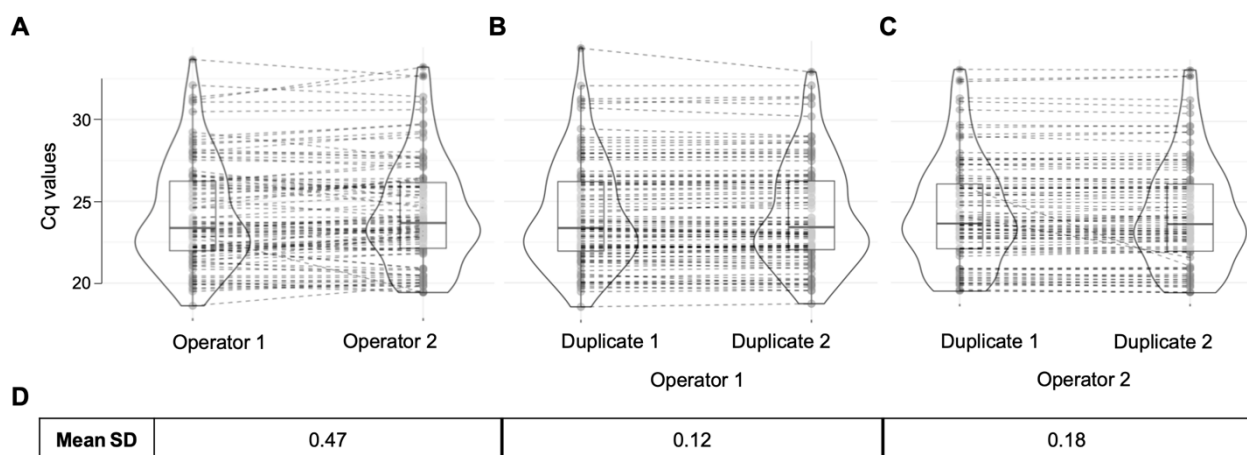
256 **Results:**

257 Validation of RNA extraction method:

258 For the RNA extraction from HCC FFPE tissues, the RNeasy FFPE kit was used. This RNA extracted
 259 method was validated by using 87 HCC FFPE samples handled by two different operators. After
 260 extraction, RNA was used as a template in RT-qPCR reactions targeting the *RPL13A* gene and a DNA-
 261 specific target (*Chr3*), assuring both the integrity and the purity of the extracted RNA. Each sample was
 262 analyzed in duplicate by each operator. Regarding *RPL13A* expression, no significant differences were
 263 observed between operators for each sample ($p = 0.27$, **Figure 2A**) and the mean standard deviation
 264 (SD) between samples tested by the two operators was 0.47 Cq (**Figure 2D**). Moreover, no statistically
 265 significant differences were observed between each operator's duplicates ($p = 0.99$ for operator 1 and
 266 $p = 0.13$ for operator 2, **Figure 2B and 2C** respectively). The mean SD between duplicates of each sample
 267 for operator 1 was 0.12 Cq while it was 0.18 Cq for operator 2 (**Figure 2D**). Regarding *Chr3*, residual
 268 genomic DNA (gDNA) contamination (Cq mean > 34) was identified in 8 samples (9.19 %) handled by
 269 operator 2 and in one sample (1.15 %) handled by operator 1 (data not shown).

270

271



272 **Figure 2 – RNA extraction validation using *RPL13A* gene expression.** The RNA extraction method was
 273 validated by two different operators ($p = 0.27$, Paired Student's t-test) (**A**). Moreover, each operator
 274 performed the assays in duplicate: operator 1 ($p = 0.99$, Paired Student's t-test) (**B**) and operator 2
 275 ($p = 0.13$, Paired Student's t-test) (**C**). The mean SD between operators and within operators is also

276 represented (D). For each violin plot (A, B, and C), dots represent Cq mean values (A) and Cq values (B
277 and C) of *RPL13A* gene. Dashed lines represent the correspondence of samples between groups.

278

279

280 Analytical specificity – Primer specificity and RT-qPCR inclusivity:

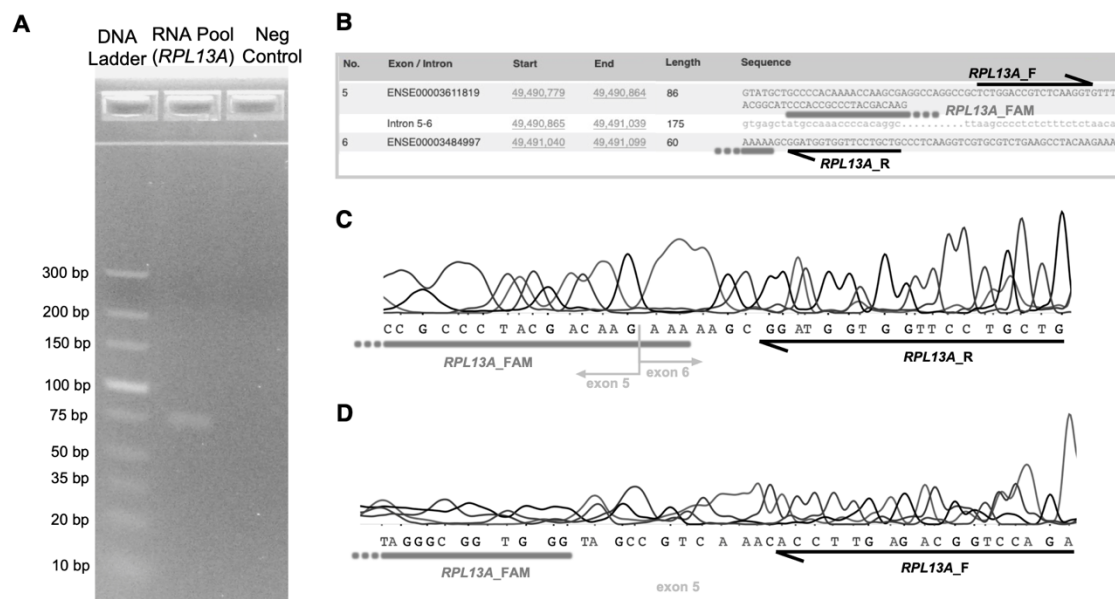
281 To demonstrate primers' uniqueness for each target, primers and RT-qPCR products were sequenced
282 and the specificity of each primer pair was confirmed by aligning the Sanger sequencing
283 electropherograms from each primer with the respective PCR amplicon, as suggested by ISO
284 20395:2019 and MIQE guidelines [19,20]. Before sequencing, the amplicon size was verified by
285 electrophoresis in an agarose gel (Figure 3A, *RPL13A* depicted as an example). For the *RPL13A* gene,
286 the electrophoresis band corresponded to 75 bp (Figure 3A), in accordance with the expected amplicon
287 size [16]. All the amplified amplicons (for the remaining HepatoPredict targets) corresponded to the
288 expected size [16] and no extra bands of unspecific PCR products were observed (data not shown)
289 confirming the specificity of the primer pairs for the desired target. The sequences of all primers and
290 probes were successfully aligned for all targets included in the HepatoPredict (data not shown) in the
291 DNA sequence displayed in the Sanger electropherogram of the respective PCR amplicon (Figure 3B-D,
292 example for *RPL13A*). Due to the very small size of the amplicons (between 71 and 108 bp [16]) and to
293 limit the baseline noise always present at the beginning and end of Sanger electropherograms, the
294 Sanger sequencing was performed for both forward and reverse strands (Figure 3C and 3D,
295 respectively). This allowed for the successful sequencing of forward and reverse primers and respective
296 probe positions in all HepatoPredict amplicons.

297

298 The inclusivity of the RT-qPCR reactions included in the HepatoPredict kit was demonstrated as
299 described in CLSI-MM17 guideline [21]. An inclusivity of 100 % was demonstrated for all targets, except
300 for *DPT* which had an inclusivity of 95 % (Supplementary File 2).

301

302



303 **Figure 3 – Primer Specificity (for *RPL13A* as an illustrative example).** After a RT-qPCR reaction targeting
 304 the *RPL13A* gene, the size of the amplicon was assessed by electrophoresis in an agarose gel (~75 bp),
 305 using a DNA ladder and a negative control (nuclease-free water) (A). ENSEMBL canonical transcript
 306 sequence (ENST00000391857.9, RefSeq NM_012423), represented from 5' to 3' (B). Alignment of the
 307 *RPL13A* forward (*RPL13A_F*) and reverse (*RPL13A_R*) primers and probe (*RPL13A_FAM*) with the Sanger
 308 sequencing electropherogram results for the *RPL13A* amplicon in forward (5' to 3') (C) and reverse (3'
 309 to 5') (D) directions. The presented image is cropped. The full size original image can be found in
 310 *Supplementary File 3*.

311

312 Limit of Detection (LoD), Limit of Blank (LoB), Linearity and Efficiency:

313 For LoD, LoB, linearity and efficiency determination of each RT-qPCR reaction included in HepatoPredict
 314 kit, an RNA pool of FFPE HCC samples was used in accordance with MM16-A guideline [22]. The LoD,
 315 for each of the 8 targets included in the HepatoPredict kit, was determined based on ISO 20395:2019,
 316 CLSI-MM17, MIQE, and CLSI-EP17-A guidelines [19–21,23]. The Probit model was used, and **Figure 4**
 317 exemplifies the application of the model to the *RPL13A* target. The Probit model was applied to two
 318 different data sets obtained using different reagent lots (**Figure 4A and 4B**). The LoD was defined as the

319 lowest concentration of target that could be detected in ≥ 95 % of the samples, as represented in **Figure**
320 **4A** and **4B**. To determine the Cq value corresponding to the LoD, a linear regression was performed
321 (**Figure 4C**). For all targets, Pearson's correlation coefficient (R^2) was higher than 0.90 (data not shown),
322 except for *Chr3* ($R^2 = 0.73$, data not shown). The highest LoD (nucleic acid concentration) between lots
323 was taken as the reported value for the measurement. The LoD of the 7 RNA targets included in the
324 HepatoPredict kit ranged from 34.75 Cq to 36.89 Cq (**Table 2**). Regarding *Chr3* the LoD was defined at
325 33.95 Cq (**Table 2**).

326

327 To apply the Probit model, it is necessary to assume that all blank or negative samples are reported as
328 negative. These assumptions are true for the HepatoPredict kit, in fact, if a valid Cq value was obtained
329 (< 40 cycles) for just one of the replicates of the NTC (no template control) the entire assay was
330 considered invalid. Thus, the LoB was assumed to be zero, i.e., Cq results for the NTC samples, for all
331 valid RT-qPCR reactions, did not cross the threshold within the 40 cycles and were considered
332 "undetermined" (> 40 Cq).

333

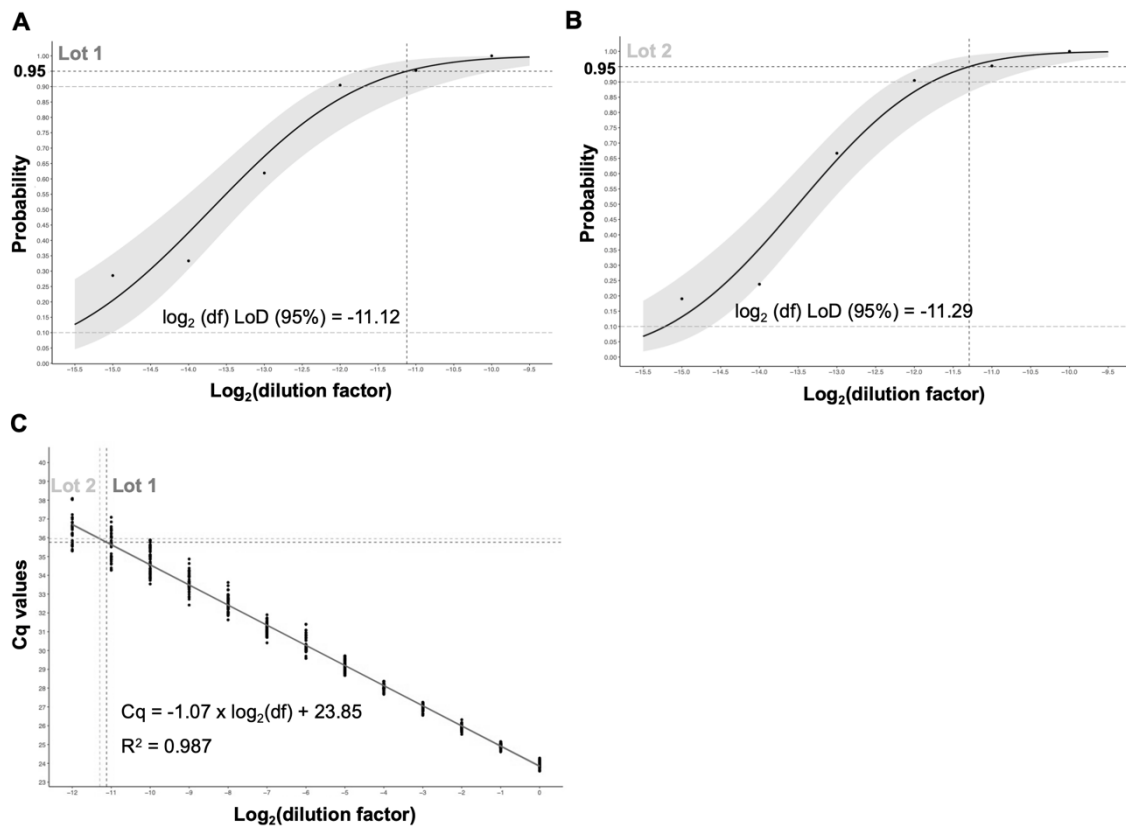
334 The linear range of each target was also determined in accordance with ISO 20395:2019 [19] with
335 $R^2 = 0.99$ for *RPL13A*, *GAPDH*, *DPT*, and *CAPNS1*, $R^2 = 0.98$ for *TBP* and *CLU*, $R^2 = 0.97$ for *SPRY2* and
336 $R^2 = 0.93$ for *Chr3*. All targets were linear at least within 24.63 and 35.11 Cq (**Table 2**). Moreover, the
337 amplification efficiencies, determined in accordance with ISO 20395:2019 and MIQE guidelines [19,20],
338 ranged from 91.02 to 110.26 Cq for all targets (**Table 2**).

339

340 Considering that the HepatoPredict kit analyzes 7 different genes and a DNA-specific target, it is
341 important to assure that all targets can be detected within their linear range and below their LoDs. As
342 represented in **Table 3**, the maximum RNA input to assure an ideal performance of the HepatoPredict
343 kit is 0.031 ng/ μ L, corresponding to Cq values of 28.26 for *RPL13A*, 28.34 for *GAPDH*, 33.95 for *TBP*,
344 34.44 for *DPT*, 29.42 for *CLU*, 31.61 for *CAPNS1*, and 33.96 for *SPRY2*.

345

346



347 Figure 4 – Estimation of the Limit of Detection (LoD) in Cq values (for *RPL13A* as an illustrative example).

348 The Probit approach was used to determine the LoD of *RPL13A* gene for two different lots of reagents.

349 The LoD was defined as the concentration (\log_2 dilution) at a probability of 95 %. The grey areas

350 represent the confidence interval at 99 % (A and B). The linear dynamic range was also estimated and

351 the Cq value associated with the LoD was determined (C). For the example of the *RPL13A* gene, the

352 highest LoD was obtained with Lot 1 and the correspondent Cq value was 35.75.

353

354 Table 2 - LoB, LoD, linear range and PCR efficiency of the 8 targets included in HepatoPredict kit.

355

Target	LoB	LoD	Linear Range		Efficiency
	Cq value	Cq value	Cq value	R ²	%
<i>RPL13A</i>	> 40	35.75	17.63 – 36.30	0.99	91.02
<i>GAPDH</i>	> 40	36.32	18.02 – 38.70	0.99	104.76
<i>TBP</i>	> 40	34.75	24.63 – 39.62	0.98	110.01

<i>DPT</i>	> 40	35.11	24.41 – 37.68	0.99	102.97
<i>CLU</i>	> 40	35.35	19.07 – 39.44	0.98	110.26
<i>CAPNS1</i>	> 40	36.89	21.22 – 38.44	0.99	101.79
<i>SPRY2</i>	> 40	34.93	23.27 – 38.02	0.97	97.70
<i>Chr3</i>	> 40	33.95	23.90 – 35.11	0.93	91.02

356

357 Table 3 – Acceptable maximum Cq values for each HepatoPredict kit’s target to be detected below their LoD and within the linear range.

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

Reference RNA (ng/μL)	Targets (Cq mean)						
	<i>RPL13A</i>	<i>GAPDH</i>	<i>TBP</i>	<i>DPT</i>	<i>CLU</i>	<i>CAPNS1</i>	<i>SPRY2</i>
8	20.19	20.40	25.67	25.95	21.38	23.43	25.15
4	21.14	21.36	26.59	26.94	22.44	24.45	26.30
2	22.28	22.37	27.66	27.85	23.42	25.46	27.31
1	23.33	23.27	28.82	28.89	24.44	26.45	28.26
0,5	24.06	24.28	29.76	30.02	25.40	27.68	29.26
0.25	25.09	25.33	30.67	31.12	26.58	28.70	30.50
0.125	26.12	26.43	31.93	32.04	27.40	29.46	31.34
0.063	27.06	27.47	32.76	32.47	28.60	30.75	32.58
0.031	28.26	28.34	33.95	34.44	29.42	31.61	33.96
0.016	29.12	29.38	35.47	35.64	30.53	32.86	34.33
0.008	30.31	30.34	35.31	36.68	31.79	33.89	37.39
0.004	31.20	31.54	36.40	N/A	32.67	35.02	36.80
0.002	32.36	32.41	N/A	38.06	33.68	37.12	N/A
0.00100	33.22	33.33	37.71	N/A	34.57	37.10	38.47
0.00050	34.41	34.18	N/A	N/A	35.20	38.19	N/A
0.00024	35.65	36.31	N/A	37.19	37.15	39.17	N/A
0.00012	36.85	37.37	N/A	N/A	N/A	38.75	N/A

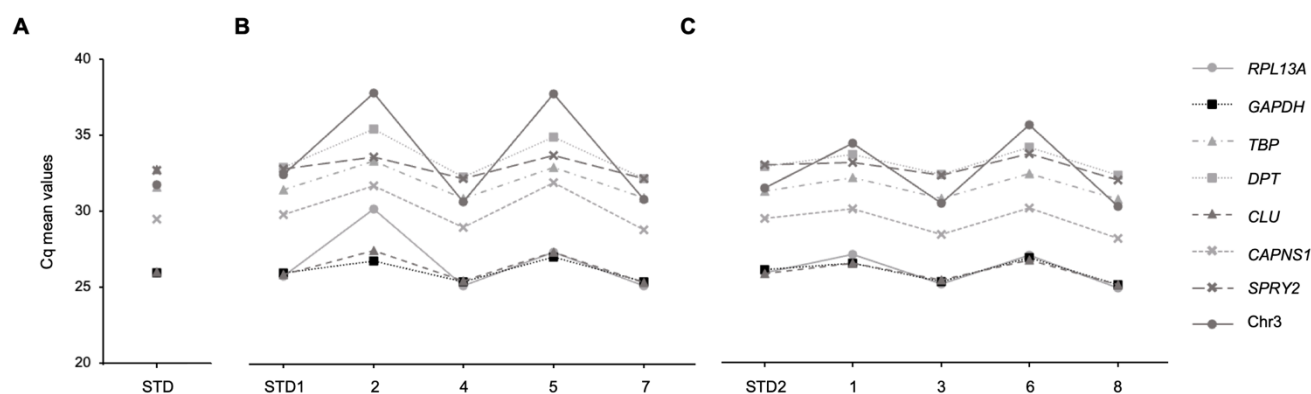
387 Cq mean values represented in bold are above the LoD of each target. Between double lines are represented the acceptable maximum Cq values for all targets assuring that all of them are
 388 detected below their LoD and within their linear range: [RNA] = 0.031 ng/μL. N/A – not applicable, Cq > 40.

389 Robustness of RT-qPCR reactions:

390 To study the robustness of the RT-qPCR reactions included in the HepatoPredict kit, alterations in the
391 concentrations and volumes of RT-qPCR reagents were performed as suggested in ISO 20395:2019 [19]
392 and represented in **Table 1**. The Cq mean values of two independent assays obtained for each target
393 under each condition are represented in **Figure 5A-C**. Conditions with the same annealing temperature
394 were compared with the respective standard condition – conditions 2, 4, 5, and 7 were compared with
395 STD1 (**Figure 5B**), while conditions 1, 3, 6, and 8 were compared with STD2 (**Figure 5C**) – and no
396 statistically significant differences were observed ($p > 0.05$). Furthermore, all conditions (from 1 to 8)
397 were compared with the STD condition representing no changes regarding the initial protocol (**Figure**
398 **5A**) and no statistically significant differences were observed ($p > 0.05$).

399

400



401 **Figure 5 – Robustness of the RT-qPCR reactions included in the HepatoPredict kit.** Representation of Cq
402 mean values for each condition (**Table 1**) of each target included in the HepatoPredict kit. The original
403 condition (STD) (**A**) was compared with all other conditions (1 to 8) and no statistically significant
404 differences were observed (Dunn's multiple comparisons test). Moreover, conditions with the same
405 annealing temperature were compared with the respective STD condition: (**B**) STD1 was compared with
406 condition 2, 4, 5 and 7 while (**C**) STD2 was compared with conditions 1, 3, 6, and 8. No statistically
407 significant differences were observed (Dunn's multiple comparisons test).

408

409 Precision of the HepatoPredict kit:

410 The precision of the HepatoPredict kit was determined as described in ISO 20395:2019, MIQE, and CLSI-
411 MM17 guidelines [19–21]. Precision data was transduced numerically using imprecision values such as
412 standard deviation (SD) and respective confidence interval (CI) at 95 % (**Table 4**). The HepatoPredict kit
413 reproducibility was verified by normalizing the gene expression level of the genes of interest (*DPT*, *CLU*,
414 *CAPNS2*, and *SPRY2*) to the geometric mean of the reference genes (*RPL13A*, *GAPDH*, and *TBP*) (**Table**
415 **4**), as described for the standard use of the kit. In general, the SD for daily, lot-to-lot, and operator
416 reproducibility were higher for the *DPT* gene (0.38 – 1.36) when compared with the other genes of
417 interest included in the HepatoPredict kit (0.03 – 0.44). The same was verified for the inter-assay
418 reproducibility (SD calculated between all the independent assays for the same sample) and total SD
419 (square root of the daily, lot-to-lot, and operator variances) (**Table 4**). Furthermore, while all targets of
420 sample A were associated with higher SD in lot-to-lot reproducibility, sample C presented higher SD
421 values in daily reproducibility. Nevertheless, none of these were observed in sample B, suggesting that
422 the observed variability between independent assays is not dependent on a single factor. In fact, when
423 all varying factors were considered (inter-assay reproducibility and total SD), SD values were similar
424 between both samples. Additionally, inter-assay reproducibility was also determined for 18 different
425 HCC FFPE samples (calculation of the mean SD and respective confidence interval), corroborating the
426 previous results (**Table 4**). The repeatability was verified for each target included in the HepatoPredict
427 kit – reference genes (**Supplementary File 4**) and genes of interest (**Table 4**) – and similar SD were
428 obtained for all targets ranging from 0.05 to 0.14 Cq.

429 Table 4 – Reproducibility of the HepatoPredict kit.

430

Sample	Target	Repeatability (mean SD)	Daily Reproducibility		Lot-to-Lot Reproducibility		Operator Reproducibility		Inter-Assay Reproducibility		Total SD
			SD	95 % CI	SD	95 % CI	SD	95 % CI	SD	95 % CI	
Sample A	<i>DPT</i>	0.07	0.52	0.10 – 0.93	1.05	0.21 – 1.89	0.81	0.16 – 1.46	0.77	0.55 – 0.99	1.42
	<i>CLU</i>	0.07	0.11	0.02 – 0.19	0.29	0.06 – 0.53	0.04	0.01 – 0.07	0.28	0.20 – 0.35	0.31
	<i>CAPNS1</i>	0.07	0.24	0.05 – 0.43	0.35	0.07 – 0.64	0.17	0.04 – 0.37	0.26	0.19 – 0.34	0.46
	<i>SPRY2</i>	0.13	0.08	0.02 – 0.14	0.08	0.02 – 0.14	0.11	0.02 – 0.19	0.18	0.13 – 0.23	0.16
Sample B	<i>DPT</i>	0.12	0.38	0.08 – 0.68	0.58	0.12 – 1.05	0.58	0.12 – 1.04	0.81	0.58 – 1.04	0.90
	<i>CLU</i>	0.08	0.34	0.07 – 0.60	0.30	0.06 – 0.53	0.19	0.04 – 0.35	0.25	0.18 – 0.32	0.49
	<i>CAPNS1</i>	0.11	0.08	0.02 – 0.14	0.04	0.01 – 0.07	0.15	0.03 – 0.26	0.13	0.09 – 0.16	0.17
	<i>SPRY2</i>	0.08	0.03	0.01 – 0.06	0.12	0.02 – 0.21	0.29	0.06 – 0.52	0.23	0.17 – 0.30	0.32
Sample C	<i>DPT</i>	0.05	1.36	0.27 – 2.45	0.41	0.08 – 0.74	0.84	0.17 – 1.51	0.99	0.71 – 1.26	1.65
	<i>CLU</i>	0.14	0.31	0.06 – 0.55	0.27	0.05 – 0.48	0.05	0.01 – 0.09	0.36	0.26 – 0.47	0.41
	<i>CAPNS1</i>	0.09	0.10	0.02 – 0.17	0.04	0.01 – 0.07	0.02	0.00 – 0.03	0.12	0.09 – 0.15	0.11
	<i>SPRY2</i>	0.08	0.44	0.09 – 0.79	0.16	0.03 – 0.28	0.38	0.08 – 0.69	0.47	0.34 – 0.60	0.60
All samples (n=18)	<i>DPT</i>								0.69	0.53 – 0.84	
	<i>CLU</i>								0.26	0.20 – 0.31	
	<i>CAPNS1</i>	N/A		N/A		N/A		N/A	0.24	0.18 – 0.29	N/A
	<i>SPRY2</i>								0.20	0.16 – 0.25	

431

SD, standard deviation; CI, confidence interval; N/A, not applied.

432 Univariate and multivariate analysis of the new version of the HepatoPredict algorithm:

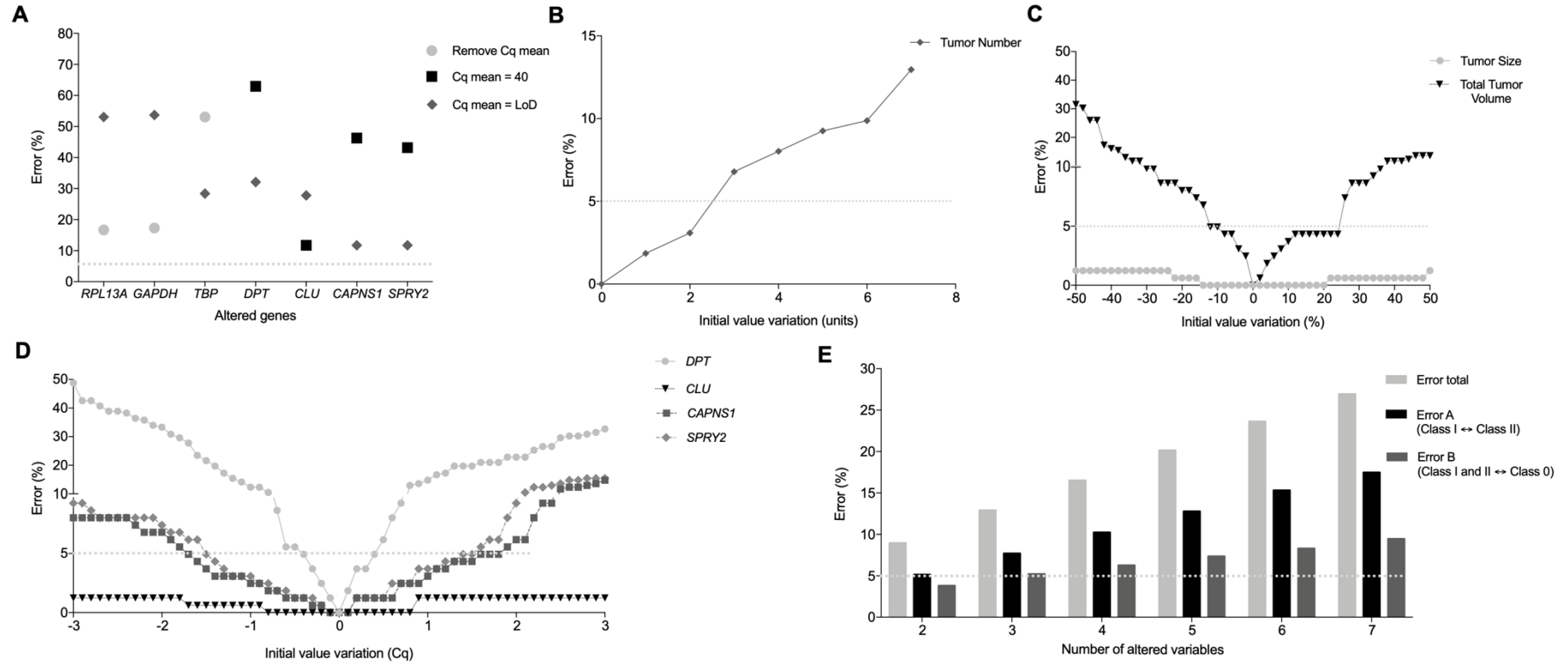
433 The HepatoPredict algorithm (V2.0) is a two-level predictor, but the first level (Class I), presents the
434 highest precision and uses the XGboost model (instead of SVM in V1.0) increasing the positive
435 predictive value (PPV) to 96.43 %. The second level (Class II) is a linear SVM model. The variables'
436 weights within each model are represented in **Supplementary File 5**, with *DPT* and *CLU* gene expression
437 levels and total tumor volume being the most important variables of the HepatoPredict algorithm.
438 Furthermore, the new HepatoPredict algorithm was also compared with other clinical criteria for the
439 identification of HCC patients suitable for liver transplantation (**Supplementary File 6**).

440

441 The analytical validation of the new HepatoPredict algorithm consisted in calculating the error (i.e., how
442 many times the correct classification is missed) associated with the alteration of one (univariate) or
443 more (multivariate) variables. **Figure 6A** represents the error associated with the alteration of the Cq
444 mean values before normalization. The errors were all superior to 10 % demonstrating that to maintain
445 a good HepatoPredict performance, no reference gene can be removed from the assay and the Cq
446 means of the genes cannot be replaced by 40 Cq or its respective LoD. Moreover, the variations allowed
447 for each variable, assuring a maximum error of 5 % (or a maximum variation of 2 Cq or 50 % from the
448 initial value), are represented in **Figure 6B-D**. In general, it was verified that the *DPT* normalized gene
449 expression level and the total tumor volume (in cm³) were the variables that least allowed for
450 alterations (for an error = 5 %, variation of ± 0.4 Cq and ± 12 %, respectively). On the other hand, *CLU*
451 normalized gene expression level and the tumor size (the diameter of the largest tumor in cm) were
452 the variables that tolerated greater variations (± 3 Cq and ± 50 % respectively, while maintaining an
453 error < 5 %) (**Figure 6C-D**). For these variables, the error at 5 % was not used as a threshold for the
454 multivariable analysis, but instead an acceptable absolute variation value was used (± 2 Cq for *CLU*
455 expression level and ± 50 % for tumor size). Furthermore, while maintaining an error < 5 %, the tumor
456 number was possible to vary in 2 units (**Figure 6B**) and the normalized gene expression levels of *CAPNS1*
457 and *SPRY2* varied ± 1.7 Cq and ± 1.5 Cq, respectively (**Figure 6D**).

458

459 The HepatoPredict algorithm multivariate analysis was based on the results presented in **Figure 6B-D**
460 and demonstrated that the error increases with the number of varying variables (**Figure 6E**). Moreover,
461 two types of errors were analyzed concerning the final HepatoPredict class. Error A corresponds to class
462 switching between HepatoPredict classes associated with a very high or high predicted benefit of LT,
463 respectively Class I and Class II [16]. Error B relates to a switch on HepatoPredict class with more impact
464 in the LT benefit, i.e., a switch from a class with no benefit to a class with LT benefit (Class 0 to Class I
465 or Class II) or from a class with LT benefit to a class with no benefit (Class I or Class II to Class 0). In
466 general, error A was ~1.6 times higher than error B (**Figure 6E**).



467
468

469 **Figure 6 – Robustness of the HepatoPredict algorithm.** Representation of errors associated with the univariate alteration of Cq mean values from each molecular
 470 variables before normalization (A), with the univariate alteration of the tumor number (B), or tumor size or total tumor volume (C). Errors associated with the
 471 univariate alteration of the molecular variables after normalization (D). Multivariate analyses of the HepatoPredict algorithm for distinct combinations of
 472 variables, ranging from 2 to 7 variables (E). The dashed line in light grey represents the errors at 5%.

473 **Discussion:**

474 The HepatoPredict kit uses an algorithm that combines molecular data (gene expression levels of *DPT*,
475 *CLU*, *CAPNS1*, and *SPRY2*) with clinical variables (tumor number, size of the largest nodule, and total
476 tumor volume) to classify the patients in two different classes associated with the benefit of a liver
477 transplant (Class I – very high confidence, and Class II – high confidence) or in Class 0 (no benefit of liver
478 transplant predicted). Some products already exist in the market focused on the prognostic prediction
479 of different tumors, such as breast [24] and prostate [25], but nothing specific for HCC is available.
480 While the successful clinical validation of HepatoPredict kit using a retrospective cohort was previously
481 published [16], the herein described new version of HepatoPredict algorithm presents several
482 improvements. To start, the new HepatoPredict algorithm is associated with a higher PPV in Class I,
483 which allows for the selection of very likely good prognosis candidates for LT, a potential advantage in
484 geographies where the time in a LT waiting lists is very long. In addition, the new HepatoPredict
485 algorithm presents a higher negative predictive value (NPV) when compared with different clinical
486 criteria (such as Milan [4], UCSF [6], Up to seven [11], AFP [7], Metroticket 2.0 [5], TTV [13], and TTV
487 AFP [12] criteria). The higher NPV of HepatoPredict translates into a higher probability of being correct
488 in terms of detecting a bad prognosis patient when the kit result is Class 0 (no benefit of a LT). This
489 reduces the misclassifications of patients that benefit from a LT and can avoid wasting a healthy organ
490 in a patient that very likely will face HCC recurrence. This correct prognosis assignment was also
491 corroborated by the results of multivariate analysis of the HepatoPredict algorithm, demonstrating a
492 higher rate of type A errors (switch between Class I and II with benefit prediction) than type B errors
493 (from good prognosis to bad prognosis and vice-versa). To further corroborate the clinical utility of the
494 HepatoPredict kit additional retrospective studies are being planned enrolling patients' cohorts from
495 different geographic localizations and HCC etiologies and a prospective study (NCT0449983) is currently
496 open and recruiting.
497

498 In the context of analytical validation of multi-target genomic assays (such as HepatoPredict kit), no
499 evaluation guidelines covering all the relevant aspects required for the diagnostic setting are available.
500 To fill this gap, different guidelines such as ISO201395-2019 [19], MIQE [20], CLSI-MM16 [22], CLSI-
501 MM17 [21], and CLSI-EP17-A [23] were followed where applicable, to demonstrate that the
502 HepatoPredict kit is a sensitive, specific, and robust test. Thus, the described analytical validation of the
503 HepatoPredict kit is in accordance with standard assay validation processes. This type of approach has
504 been previously used to validate similar prognostic [26–29] and diagnostic [30–32] tests and as a
505 reference for analytical validation of the test in different molecular pathology laboratories.

506

507 In diagnostic settings, FFPE is the most commonly used technique for long-term conservation of clinical
508 samples, since it preserves the proteins and vital structures within the tissue while it aids microscopic
509 diagnostic examination, experimental research, and diagnostic/drug development [33]. FFPE samples
510 were thus implemented for the HepatoPredict kit to simplify its adoption by molecular biology and
511 pathology laboratories. The RNA extraction method from FFPE HCC samples was demonstrated to be
512 repeatable (between duplicates) and reproducible (between operators). Regarding gDNA residual
513 contamination, Cq values above 34 in RT-qPCR reactions targeting *Chr3* were observed in 1.15 and
514 9.19 % of the samples (for operator 1 and 2, respectively). However, the LoD for *Chr3* was determined
515 at 33.95 Cq, meaning that above this Cq value, *Chr3* detection is likely invalid, suggesting that gDNA
516 contamination during RNA extraction from HCC FFPE samples is very residual and approaching 0 % with
517 HepatoPredict kit.

518

519 The RNA extracted from FFPE tissues is normally fragmented [34], thus, FFPE sections were digested
520 with heat application (56 °C) and proteinase K to decrease RNA fragmentation and chemical
521 modifications [35,36]. Moreover, the primers of the HepatoPredict kit were designed for the generation
522 of short amplicons to increase gene detection rate [17,37,38], and gene specific reverse transcription
523 and targeted cDNA amplification (1-step RT-qPCR) were performed to increase the accuracy and

524 sensitivity of the RT-qPCR reactions [39]. Nevertheless, the HepatoPredict kit includes a sample quality
525 control step comprising 1-step RT-qPCR reactions targeting *RPL13A* and *Chr3*. With this procedure, it is
526 possible to determine if the extracted sample contains enough RNA (shown by the Cq value from
527 *RPL13A*) and if gDNA contamination is present (reported by the Cq value from *Chr3*) [17] before
528 proceeding with the kit protocol. This assessment is important because it assures the reproducibility
529 and veracity of the experiments, avoids extra costs associated with the need of repeating the analysis
530 and the wasting of precious tumor samples [20,40]. Considering that HepatoPredict analyses the
531 expression level of 7 different genes, Cq values for each target, allowing all targets to be detected within
532 their linear ranges and below their LoDs, were determined. Thus, an acceptable Cq range for the sample
533 quality control was defined for the *RPL13A* gene: 18.32 to 28.26 Cq. Regarding *Chr3*, all Cq values above
534 its LoD (33.95 Cq) are acceptable since they represent no gDNA contamination.

535

536 Regarding RT-qPCR reactions, primer specificity for each target included in the HepatoPredict kit was
537 demonstrated. Although probes (TaqMan[®] technology) are also included in the RT-qPCR reactions, they
538 were not analyzed in the context of sequencing because they do not amplify PCR products and because
539 their fluorescence is only released in the context of highly specific annealing to the target sequences in
540 the PCR amplicons. Thus, if the Sanger sequencing proves that each primer pair-related PCR amplicon
541 is specific and no unspecific PCR products are detected, each probe can only anneal to the amplified
542 specific product. To further demonstrate that the RT-qPCR reactions included in the HepatoPredict kit
543 could distinguish between target and non-target sequences, an inclusivity of 100 % was demonstrated
544 for all targets, excluding the *DPT* gene which had an inclusivity of 95 %. This result was expected since
545 the downregulation of the *DPT* gene in HCC has been demonstrated and can be already associated with
546 HCC carcinogenesis and progression [41–43]. In contrast with diagnostic systems [31], the exclusivity
547 of the RT-qPCR reactions was not studied since the HepatoPredict kit analyzes the expression level of
548 genes that are not exclusively expressed on HCC cells. Nevertheless, to be analyzed by HepatoPredict
549 kit, each HCC sample needs to be collected by expert clinicians (surgeons or radiologists) and

550 subsequently analyzed at the microscopical level (H&E-stained tissue slides) by a certified pathologist,
551 assuring the specificity of each HCC biopsy submitted to the HepatoPredict kit test. Furthermore, the
552 sensitivity of the RT-qPCR reactions included in the HepatoPredict kit was determined by defining the
553 LoD for each target, as well as the respective linear range. The amplification efficiency was also
554 calculated for all targets being between 90 and 110 % as recommended by ISO 203095:2019.

555

556 The robustness of the qPCR reactions included in HepatoPredict kit was studied. It was demonstrated
557 that qPCR reactions were robust, not being affected by small changes either in annealing temperatures
558 and reagents' concentration and volumes.

559

560 Moreover, precision studies, assessing both the repeatability and reproducibility of the qPCR reactions
561 included in the HepatoPredict kit, demonstrated that the variability associated with normalized Cq
562 values for *DPT*, *CLU*, *CAPNS1* and *SPRY2* genes was not dependent on a single factor (day, lot, or
563 operator). Furthermore, the inter-assay SD within a sample is similar to the inter-assay SD between 18
564 samples, demonstrating the reproducibility of the assay independently of the sample used. The *DPT*
565 gene was associated with a higher SD in all assays, nevertheless no differences were observed in *DPT*
566 repeatability in comparison with the other targets. These results suggest that the lower reproducibility
567 of the *DPT* gene (i.e., higher SD), may be associated with *DPT* lower inclusivity (95 %) due to *DPT*
568 downregulation in HCC [41–43].

569

570 The robustness of the new HepatoPredict algorithm was also studied and the acceptable variation
571 range for each variable was determined. It was demonstrated that *DPT* gene expression level and total
572 tumor volume were the most sensitive variables. This was expected since in the XGBoost model (first
573 level), the *DPT* gene expression level and total tumor volume had an information gain of 3.75 and 0.29
574 (respectively), while the other variables had an information gain of zero. Moreover, the SVM model
575 (second level), also corroborated these results since the variables with higher SVM weights were *DPT*

576 gene expression level, total tumor volume, and *CLU* gene expression level. Total tumor volume is
577 related to the number of tumors and tumor diameter measurements; thus, errors in these variables
578 will influence its value. A recent study described a mean error of 0.81 cm when measuring the tumor
579 size using different magnetic resonance imaging pulse sequences [44]. This was reflected in an
580 HepatoPredict type B error of 4.94 %, demonstrating that the prognostic test handles common
581 measuring errors.

582

583

584 **Conclusions:**

585 Despite the introduction of perturbations mimicking real-life observed variations to the molecular and
586 clinical variables, the prognostic information achieved with the HepatoPredict kit does not change.
587 Prognosis variation is only expected under extreme and combined variations of multiple variables,
588 unlikely to occur in real life. In addition, the technical validation procedures presented in this study can
589 be used as a reference for the analytical validation of the HepatoPredict test in different molecular
590 diagnostic laboratories. The performance of the presented analytical testing also demonstrates that the
591 HepatoPredict kit can be easily integrated into routine molecular diagnostic procedures to accurately
592 identify HCC patients more likely to benefit from a liver transplant, contributing to the implementation
593 of a true precision medicine.

594

595 **List of Abbreviations:**

596 *AFP*: alpha-fetoprotein

597 *DNA*: Deoxyribonucleic acid

598 *CI*: Confidence interval

599 *CLSI*: Clinical laboratory standards institute

600 *Cq*: Cycle threshold

601 *DCP*: des- γ carboxyprothrombin

602 *FFPE*: formalin-fixed paraffin-embedded

603 *gDNA*: Genomic DNA

604 *HBV*: Hepatitis B virus

605 *HCC*: Hepatocellular carcinoma

606 *HCV*: Hepatitis C virus

607 *ISO*: International organization for standardization

608 *LoB*: Limit of blank

609 *LoD*: Limit of detection

610 *LT*: Liver transplantation
611 *MIQE*: Minimum information for publication of quantitative real-time PCR experiments
612 *NPV*: Negative predictive value
613 *NTC*: No template control
614 *N/A*: Not applied
615 *p*: p-value
616 *RNA*: Ribonucleic acid
617 *RT-qPCR*: real-time quantitative reverse transcription polymerase chain reaction
618 *SD*: Standard deviation
619 *STD*: Standard
620 *TTV*: Total tumor volume
621 *UCSF*: University of California San Francisco

622

623 **Declarations:**

624 Ethics approval and consent to participate: The retrospective clinical study was approved by the ethics
625 authorities (Comissão de Ética para a Saúde) from the Centro Hospitalar de Lisboa Central (Process
626 number 144/2014). Being a retrospective study focused on tumoral tissue, the informed consent was
627 waived by the same ethics authorities referred above. Samples were used in accordance with the
628 Declaration of Helsinki.

629 Consent for publication: Not applicable.

630 Availability of data and materials: The datasets used and/or analyzed during the current study are
631 available from the corresponding author on reasonable request.

632 Competing interests: The work described here is subject to patent WO 2021/064230 A1; JPL, JC, and
633 HPM declare an ownership interest in the company Ophiomics. MGR, DP, LPF, and JLN are employees
634 at Ophiomics. SS has no competing interests.

635 Funding: This work was partly funded by a grant from the European Innovation Council under the EIC
636 Accelerator scheme (Contract N°946364).

637 Authors' contributions: MGR and JC conceived and designed the study. MGR and DP performed the
638 experiments. MGR, LPF, JLN, and JC analyzed and interpreted the data. LPF drafted the manuscript. SS
639 and HPM coordinated clinical sample collection. MGR, DP, JLN, JBPL, JC edited and revised the
640 manuscript. All authors read and approved the final manuscript.

641 Acknowledgements: The authors wish to thank to the patients, to *Neuralshift* and to the pathology
642 team from the Curry Cabral Hospital, particularly Clara Rodrigues and António Figueiredo. Moreover,
643 the authors particularly acknowledge the Biobank IRBLleida (PT20/00021) integrated in the Spanish
644 National Biobanks Network and Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director
645 d'Oncologia Catalunya (XBTC), as well as the Biobank ISABIAL integrated in the Spanish National
646 Biobanks Network and in the Valencia Biobanking Network for their collaboration.

647

648 **References:**

649 1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics
650 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA
651 Cancer J Clin. Wiley; 2021;71:209–49.

652 2. Silva MF, Sherman M. Criteria for liver transplantation for HCC: What should the limits be? J Hepatol.
653 Elsevier B.V.; 2011;55:1137–47.

654 3. Golabi P, Fazel S, Otgonsuren M, Sayiner M, Locklear CT, Younossi ZM. Mortality assessment of
655 patients with hepatocellular carcinoma according to underlying disease and treatment modalities.
656 Medicine (United States). Lippincott Williams and Wilkins; 2017;96:e5904.

657 4. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for
658 the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med. M;
659 1996;334:693–702.

- 660 5. Mazzaferro V, Sposito C, Zhou J, Pinna AD, de Carlis L, Fan J, et al. Metroticket 2.0 Model for Analysis
661 of Competing Risks of Death After Liver Transplantation for Hepatocellular Carcinoma.
662 *Gastroenterology*. W.B. Saunders; 2018;154:128–39.
- 663 6. Yao FY, Ferrell L, Bass NM, Watson JJ, Bacchetti P, Venook A, et al. Liver transplantation for
664 hepatocellular carcinoma: Expansion of the tumor size limits does not adversely impact survival.
665 *Hepatology*. W.B. Saunders; 2001;33:1394–403.
- 666 7. Notarpaolo A, Layese R, Magistri P, Gambato M, Colledan M, Magini G, et al. Validation of the AFP
667 model as a predictor of HCC recurrence in patients with viral hepatitis-related cirrhosis who had
668 received a liver transplant for HCC. *J Hepatol*. Elsevier B.V.; 2017;66:552–9.
- 669 8. Halazun KJ, Najjar M, Abdelmessih RM, Samstein B, Griesemer AD, Guarrera J v., et al. Recurrence
670 after liver transplantation for hepatocellular carcinoma. *Ann Surg*. Lippincott Williams and Wilkins;
671 2017;265:557–64.
- 672 9. Sasaki K, Morioka D, Conci S, Margonis GA, Sawada Y, Ruzzenente A, et al. The Tumor Burden Score:
673 A New “metro-ticket” Prognostic Tool for Colorectal Liver Metastases Based on Tumor Size and Number
674 of Tumors. *Ann Surg*. Lippincott Williams and Wilkins; 2018;267:132–41.
- 675 10. Kaido T, Ogawa K, Mori A, Fujimoto Y, Ito T, Tomiyama K, et al. Usefulness of the Kyoto criteria as
676 expanded selection criteria for liver transplantation for hepatocellular carcinoma. *Surgery (United
677 States)*. 2013;154:1053–60.
- 678 11. Lei JY, Wang WT, Yan LN. Up-to-seven criteria for hepatocellular carcinoma liver transplantation: A
679 single center analysis. *World J Gastroenterol*. Baishideng Publishing Group Co; 2013;19:6077–83.
- 680 12. Toso C, Meeberg G, Hernandez-Alejandro R, Dufour JF, Marotta P, Majno P, et al. Total Tumor
681 Volume and Alpha-Fetoprotein for Selection of Transplant Candidates With Hepatocellular Carcinoma:
682 A Prospective Validation. *Hepatology*. John Wiley and Sons Inc; 2015;62:158–65.
- 683 13. Macaron C, Hanouneh IA, Lopez R, Aucejo F, Zein NN. Total tumor volume predicts recurrence of
684 hepatocellular carcinoma after liver transplantation in patients beyond Milan or UCSF criteria.
685 *Transplant Proc*. 2010. p. 4585–92.

- 686 14. Mehta N. Liver Transplantation Criteria for Hepatocellular Carcinoma, including Posttransplant
687 Management. *Clin Liver Dis* (Hoboken). 2021;17:332–6.
- 688 15. Santopaolo F, Lenci I, Milana M, Manzia TM, Baiocchi L. Liver transplantation for hepatocellular
689 carcinoma: Where do we stand? *World J Gastroenterol*. Baishideng Publishing Group Co;
690 2019;25:2591–602.
- 691 16. Pinto-Marques H, Cardoso J, Silva S, Neto JL, Gonçalves-Reis M, Proença D, et al. A gene expression
692 signature to select hepatocellular carcinoma patients for liver transplantation. *Ann Surg*. 2022;
- 693 17. Antonov J, Goldstein DR, Oberli A, Baltzer A, Pirotta M, Fleischmann A, et al. Reliable gene
694 expression measurements from degraded RNA by quantitative real-time PCR depend on short
695 amplicons and a proper normalization. *Laboratory Investigation*. 2005;85:1040–50.
- 696 18. Plackett RL, Burman JP. Biometrika Trust *The Design of Optimum Multifactorial Experiments*.
697 *Biometrika*. 1946;33:305–25.
- 698 19. International Organisation of Standardization. ISO 20395:2019. *Biotechnology – Requirements for*
699 *evaluating the performance of quantification methods for nucleic acid target sequences – qPCR and*
700 *dPCR*. ISO 20395:2019. International Organisation of Standardization; 2019.
- 701 20. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines:
702 Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*.
703 2009;55:611–22.
- 704 21. Clinical & Laboratory Standards Institute (CLSI). *Verification and validation of multiplex nucleic acid*
705 *assays*. CLSI document MM17-A. Clinical and Laboratory Standards Institute; 2008.
- 706 22. Clinical & Laboratory Standards Institute (CLSI). *Use of external RNA controls in gene expression*
707 *assays*. CLSI document MM16-A. Clinical and Laboratory Standards Institute; 2006.
- 708 23. Clinical & Laboratory Standards Institute (CLSI). *Protocols for determination of limits of detection*
709 *and limits of quantitation*. NCCLS document EP17-A. NCCLS; 2004.

- 710 24. Varnier R, Sajous C, de Talhouet S, Smentek C, Péron J, You B, et al. Using breast cancer gene
711 expression signatures in clinical practice: unsolved issues, ongoing trials and future perspectives.
712 *Cancers (Basel)*. MDPI; 2021;13:1–24.
- 713 25. Choudhury A, West CML. Translating prognostic prostate cancer gene signatures into the clinic.
714 *Transl Cancer Res*. AME Publishing Company; 2017. p. S405–8.
- 715 26. Kronenwett R, Bohmann K, Prinzler J, Sinn B v., Haufe F, Roth C, et al. Decentral gene expression
716 analysis: Analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test.
717 *BMC Cancer*. 2012;12.
- 718 27. Cronin M, Sangli C, Liu ML, Pho M, Dutta D, Nguyen A, et al. Analytical validation of the oncoType
719 DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-
720 negative, estrogen receptor-positive breast cancer. *Clin Chem*. 2007;53:1084–91.
- 721 28. Laible M, Schlombs K, Kaiser K, Veltrup E, Herlein S, Lakis S, et al. Technical validation of an RT-qPCR
722 in vitro diagnostic test system for the determination of breast cancer molecular subtypes by
723 quantification of ERBB2, ESR1, PGR and MKI67 mRNA levels from formalin-fixed paraffin-embedded
724 breast tumor specimens. *BMC Cancer*. BioMed Central Ltd.; 2016;16.
- 725 29. Gruselle O, Coche T, Louahed J. Development of a quantitative real-time RT-PCR assay for the
726 detection of MAGE-A3-positive tumors. *Journal of Molecular Diagnostics*. Elsevier B.V.; 2015;17:382–
727 91.
- 728 30. Potter NT, Hurban P, White MN, Whitlock KD, Lofton-Day CE, Tetzner R, et al. Validation of a real-
729 time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. *Clin*
730 *Chem*. American Association for Clinical Chemistry Inc.; 2014;60:1183–91.
- 731 31. Brown JT, Beldorth IJ, Laosinchai-Wolf W, Fahey ME, Jefferson KL, Ruskin AK, et al. Analytical
732 Validation of a Highly Sensitive, Multiplexed Chronic Myeloid Leukemia Monitoring System Targeting
733 BCR-ABL1 RNA. *Journal of Molecular Diagnostics*. Elsevier B.V.; 2019;21:718–33.

- 734 32. Gürtler C, Laible M, Schwabe W, Steinhäuser H, Li X, Liu S, et al. Transferring a Quantitative
735 Molecular Diagnostic Test to Multiple Real-Time Quantitative PCR Platforms. *Journal of Molecular*
736 *Diagnostics*. Elsevier B.V.; 2018;20:398–414.
- 737 33. Seiler C, Sharpe A, Barrett JC, Harrington EA, Jones E v., Marshall GB. Nucleic acid extraction from
738 formalin-fixed paraffin-embedded cancer cell line samples: a trade off between quantity and quality?
739 *BMC Clin Pathol*. BioMed Central; 2016;16:1–13.
- 740 34. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE
741 samples. *PLoS One*. 2007;2.
- 742 35. Chung J-Y, Braunschweig T, Hewitt SM. Optimization of Recovery of RNA From Formalin-fixed,
743 Paraffin-embedded Tissue. *Diagnostic Molecular* [Internet]. 2006; Available from:
744 www.cancer.gov/tarp
- 745 36. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA
746 from formalin-fixed samples and optimization of molecular biology applications for such samples.
747 *Nucleic Acids Res*. 1999;27:4436–43.
- 748 37. Specht K, Richter T, Mü U, Walch A, Werner M, Hö H. Technical Advance Quantitative Gene
749 Expression Analysis in Microdissected Archival Formalin-Fixed and Paraffin-Embedded Tumor Tissue.
750 *American Journal of Pathology*. 2001;158.
- 751 38. Sánchez-Navarro I, Gámez-Pozo A, González-Barón M, Pinto-Marín Á, Hardisson D, López R, et al.
752 Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen
753 and formalin-fixed, paraffin-embedded breast cancer tissues. *Biotechniques*. Eaton Publishing
754 Company; 2010;48:389–97.
- 755 39. Zeka F, Vanderheyden K, de Smet E, Cuvelier CA, Mestdagh P, Vandesomepele J. Straightforward and
756 sensitive RT-qPCR based gene expression analysis of FFPE samples. *Sci Rep*. Nature Publishing Group;
757 2016;6.
- 758 40. Die J v., Román B. RNA quality assessment: a view from plant qPCR studies. *J Exp Bot*. Oxford
759 University Press; 2012;63:6069–77.

- 760 41. Fu Y, Feng M-X, Yu J, Ma M-Z, Liu X-J, Li J, et al. DNA methylation-mediated silencing of matricellular
761 protein dermatopontin promotes hepatocellular carcinoma metastasis by $\alpha 3\beta 1$ integrin-Rho GTPase
762 signaling. *Oncotarget* [Internet]. 2014;5:6701–15. Available from:
763 www.impactjournals.com/oncotarget/
- 764 42. Li X, Feng P, Ou J, Luo Z, Dai P, Wei D, et al. Dermatopontin is expressed in human liver and is
765 downregulated in hepatocellular carcinoma. *Biochemistry (Moscow)*. Maik Nauka-Interperiodica
766 Publishing; 2009;74:979–85.
- 767 43. Liu S, Qiu J, He G, Geng C, He W, Liu C, et al. Dermatopontin inhibits WNT signaling pathway via
768 CXXC finger protein 4 in hepatocellular carcinoma. *J Cancer*. Ivyspring International Publisher;
769 2020;11:6288–98.
- 770 44. Armbruster M, Guba M, Andrassy J, Rentsch M, Schwarze V, Rübenthaler J, et al. Measuring hcc
771 tumor size in mri—the sequence matters! *Diagnostics*. MDPI; 2021;11.
772
773
774