

Supplementary Materials and Methods

Bile collection and DNA extraction

Patients were fasted overnight and ERCPs were conducted by highly experienced endoscopists. During standard ERCP procedure, after cannulation of the bile duct, and in most cases before contrast injection (Omnipaque, iohexol), a bile sample of 2 to 6 ml from each patient was aspirated through the sphincterotome as we previously described [1].

A second group of patients (n=6) included healthy living liver donors from which gallbladder bile was collected at the time of surgery. These samples were collected at the Department of General, Transplant and Liver Surgery, Medical University of Warsaw, Warsaw, Poland, with the approval of the Institutional Ethics Committee (protocol# KB/49/2015). Informed consent was obtained from each patient and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

After collection, bile samples were maintained at 4°C, centrifuged for 10 min (4°C) at 3,500 g and stored in aliquots at -80°C in biobank facilities. All the process was performed in less than 2 hours. Prior to cfDNA isolation, bile was slowly thawed at 4°C and centrifuged for 10 min (4°C) at 13,000 g to ensure removal of impurities in the supernatant. Bile and, when available, plasma cfDNA were extracted with the Maxwell® RSC Automated cfDNA Plasma Kit (Promega, Madison, WI) using the Maxwell® Nucleic Acid Purification Instrument (Promega) according to the manufacturer's instructions. cfDNA was quantitated with the QuantiFluor® dsDNA Sample Kit (Promega) and cfDNA size distributions were analysed by Agilent 2100 Bioanalyzer (Agilent

Technologies, Santa Clara, CA). PCRs were performed to amplify different size fragments of the *TP53* gene in bile and plasma cfDNA. A 148bp fragment was amplified with primers located at exon 6 (forward: 5'-TGGGCCTGTGTTATCTCCTA-3'; reverse: 5'-GGCAAGTGGCTCCTGACCT-3') whereas a 957bp fragment was amplified with primers located at exon 5 (forward:5'-CCGCGCCATGGCCATCTACAAG-3') and exon 7 (reverse: 5'-GAGTCTTCCAGTGTGATGATGG-3') as we previously described [2]. For 30 patients paired bile and tumor tissue samples were available. Genomic DNA extracted from formalin-fixed and paraffin-embedded (FFPE) CCA and PDAC tissues, three pooled 5µm slices per sample, using the Maxwell® RSC DNA FFPE kit from Promega. All samples included in the study had an estimated tumor tissue content ≥ 40% based on histopathologic assessment.

Next Generation Sequencing DNA analyses

Coded bile and plasma cfDNA samples were blindly tested with the Oncomine™ Pan-Cancer Cell-Free Assay following the manufacturer's instructions (PanCancer, Thermo Fisher Scientific, Waltham, MA)[3]. This panel includes 52 genes enabling hotspot single nucleotide variation (SNV) and short indel as well as copy number variation (CNV) detection in key genes frequently mutated in multiple cancer types, including pancreatobiliary cancers. Four PanCancer libraries were manually prepared prior to be sequenced within an Ion 540 Chip in the Ion S5 system. Using an input of 50 ng of cfDNA, tagging individual DNA fragments with short random oligonucleotides called unique molecular identifiers and defining 2 as the minimum number of variants supporting functional families, this assay allowed variant detection as low as 0.02% given that on average ~8,000-10,000 functional families were identified

by sequenced base. A mutant allele frequency (MAF) value $\geq 0.15\%$ was defined as presence of mutation.

cfDNA extracted from bile samples and DNA extracted from FFPE CCA tissues was analyzed using the OncoPrint™ Comprehensive Assay panel v3 (OCA, Thermo Fisher), designed for FFPE tissue samples, following manufacturer's instructions. The OCA panel covers 161 genes enabling hotspot SNV and indel as well as CNV and fusion detection in cancer driver genes, also with extensive coverage of genes altered in pancreaticobiliary cancers. Eight OCA libraries were automatically prepared by the Ion Chef Instrument prior to be sequenced within an Ion 540 Chip in the Ion S5 System. This assay allowed variant allele detection as low as 5% given that on average, 2000x sequencing depth was obtained per base (only variants with $>500x$ were interpreted).

Statistical analyses

The sample size was determined according to previous studies addressing the same diagnostic issues [4]. Sensitivity and specificity were calculated using standard 2x2 contingency tables, essentially as described in previous similar studies [4].

References

- 1 Urman JM, Herranz JM, Uriarte I, *et al.* Pilot multi-omic analysis of human bile from benign and malignant biliary strictures: A machine-learning approach. *Cancers (Basel)* 2020;**12**:1–30. doi:10.3390/cancers12061644
- 2 Jiménez M, Urtasun R, Elizalde M, *et al.* Splicing events in the control of genome integrity: Role of SLU7 and truncated SRSF3 proteins. *Nucleic Acids Res* 2019;**47**:3450–66. doi:10.1093/nar/gkz014
- 3 Maclás M, Canáda-Higueras E, Alegre E, *et al.* Performance comparison of two next-generation sequencing panels to detect actionable mutations in cell-free DNA in cancer patients. *Clin Chem Lab Med* 2020;**58**:1341–8. doi:10.1515/cclm-2019-1267
- 4 Singhi AD, Nikiforova MN, Chennat J, *et al.* Integrating next-generation sequencing to endoscopic retrograde cholangiopancreatography (ERCP)-obtained biliary specimens improves the detection and management of patients with malignant bile duct strictures. *Gut* 2020;**69**:52–61. doi:10.1136/gutjnl-2018-317817