OPEN ACCESS

Original research

Molecular portraits of patients with intrahepatic cholangiocarcinoma who diverge as rapid progressors or long survivors on chemotherapy

Colm J O'Rourke,[1] Massimiliano Salati,[2,3] Colin Rae,[4] Guido Carpino ,[5] Holly Leslie,[4] Antonio Pea,[4] Maria G Prete,[4] Luca R Bonetti,[6] Francesco Amato,[4] Robert Montal,[7] Rosie Upstill-Goddard,[4] Colin Nixon,[8] Paula Sanchon-Sanchez,[4] Paolo Kunderfranco,[9] Daniela Sia ,[10] Eugenio Gaudio,[5] Diletta Overi ,[5] Stefano Cascinu ,[11] Dan Hogdall ,[1,12] Sian Pugh,[13] Enric Domingo ,[14] John N Primrose,[15] John Bridgewater,[16] Andrea Spallanzani,[2] Fabio Gelsomino,[2] Josep M Llovet,[17,18,19] Diego F Calvisi,[20,21] Luke Boulter ,[22,23] Francesco Caputo,[2] Ana Lleo ,[24,25] Nigel B Jamieson,[4,23] Gabriele Luppi,[2] Massimo Dominici,[2] Jesper B Andersen ,[1] Chiara Braconi [4,23,26]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2023-330748).

For numbered affiliations see end of article.

Correspondence to
Prof Chiara Braconi, School of Cancer Sciences, University of Glasgow, Glasgow G61 1QH, UK;
chiara.braconi@glasgow.ac.uk and Professor Jesper B Andersen, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Dk-2200, Denmark;
jesper.andersen@bric.ku.dk

CJO and MS contributed equally.

JBA and CB are joint senior authors.

Received 26 July 2023
Accepted 11 September 2023

To cite: O'Rourke CJ, Salati M, Rae C, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2023-330748

ABSTRACT

Objective Cytotoxic agents are the cornerstone of treatment for patients with advanced intrahepatic cholangiocarcinoma (iCCA), despite heterogeneous benefit. We hypothesised that the pretreatment molecular profiles of diagnostic biopsies can predict patient benefit from chemotherapy and define molecular bases of innate chemoresistance.

Design We identified a cohort of advanced iCCA patients with comparable baseline characteristics who diverged as extreme outliers on chemotherapy (survival <6 m in rapid progressors, RP; survival >23 m in long survivors, LS). Diagnostic biopsies were characterised by digital pathology, then subjected to whole-transcriptome profiling of bulk and geospatially macrodissected tissue regions. Spatial transcriptomics of tumour-infiltrating myeloid cells was performed using targeted digital spatial profiling (GeoMx). Transcriptome signatures were evaluated in multiple cohorts of resected cancers. Signatures were also characterised using in vitro cell lines, in vivo mouse models and single cell RNA-sequencing data.

Results Pretreatment transcriptome profiles differentiated patients who would become RPs or LSs on chemotherapy. Biologically, this signature originated from altered tumour-myeloid dynamics, implicating tumour-induced immune tolerogenicity with poor response to chemotherapy. The central role of the liver microenviroment was confrmed by the association of the RPLS transcriptome signature with clinical outcome in iCCA but not extrahepatic CCA, and in liver metastasis from colorectal cancer, but not in the matched primary bowel tumours.

Conclusions The RPLS signature could be a novel metric of chemotherapy outcome in iCCA. Further development and validation of this transcriptomic signature is warranted to develop precision chemotherapy strategies in these settings.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Cholangiocarcinoma (CCA) patient management continues to be dominated by an all-comer approach to chemotherapy in first-line despite heterogeneous benefit. Our inability to quantify the chemosensitivity of patients' disease remains a bottleneck to optimising their clinical management. Increasing knowledge of the molecular bases behind chemosensitivity can aid development of novel therapeutic strategies.

WHAT THIS STUDY ADDS

⇒ Pretreatment transcriptomic profiles of diagnostic biopsies differentiate intrahepatic CCA patients who become rapid progressors or long survivors on chemotherapy. The RPLS signature is associated with benefit from cytotoxic agents for patients with primary and liver-metastatic tumours, indicating a precision chemotherapy strategy may be feasible and identifying candidate therapeutic targets to boost chemosensitivity.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Pending further validation, the RPLS signature could provide a clinical-grade tool to inform on chemotherapy benefit before starting treatment, foregoing unnecessary toxicities from a regimen of limited therapeutic benefit on a patient-by-patient basis. In addition, it unveils the biology behind different long-term outcomes in patients receiving chemotherapy, providing the bases for development of novel therapeutics.

INTRODUCTION

Intrahepatic cholangiocarcinoma (iCCA) is a family of rare, heterogeneous tumours arising from the

intrahepatic biliary tree. Incidence and mortality rates of iCCA appear to be increasing.[1 2] Due to asymptomatic development, typical presentation without known risk factors and lack of early diagnostic biomarkers, more than 50% of patients are diagnosed with locally advanced and metastatic disease.[3] In this setting, systemic chemotherapy with gemcitabine and cisplatin remains the standard-of-care in first-line,[4] with reported additive benefit of durvalumab and pembrolizumab for a niche of patients.[5 6] Overall benefit from FOLFOX in second-line[7] and capecitabine in the adjuvant setting[8] reinforce chemotherapy as central to patients' management throughout their disease trajectory. However, benefit from cytotoxic agents is heterogeneous. In the ABC-02 trial, 11% of biliary tract cancer (BTC) patients were alive at 2 years following enrolment, greater than double the median overall survival (OS).[4 9] These long survivors (LSs) are contrasted with data indicating that chemotherapy fails to achieve disease control in 25%–28% of advanced BTC patients.[4 10] Understanding and predicting benefit of chemotherapy remains a critical unmet need for patients with iCCA, especially as novel targeted therapies become available in second-line for specific tumour genotypes[11 12] and might be superior in first-line for patients unlikely to benefit from chemotherapy.[13]

Precision approaches improve the outcome of patients with iCCA,[14] but are currently limited to targeted therapies. With an all-comer approach for first-line chemotherapy, it remains unclear which patients will not benefit from this regimen, with implications for optimal treatment and quality of life. Potentially actionable genomic alterations occur in up to 52% of iCCA,[15] but corresponding therapies are only available following progression on chemotherapy, a rate-limiting step as many patients deteriorate as a result of disease progression. Unlike targeted therapies with clear DNA-based indications in single genes, the molecular basis of innate sensitivity to chemotherapy remains unclear. DNA profiling of tumours is restricted to detecting genetic alterations in tumour cells, omitting the important impact of non-genetic tumour alterations and non-genetic microenvironment alterations on treatment outcome.[16] Transcriptome profiling captures a more holistic overview of the tumour biology (cell composition and behaviour) and is gaining clinical traction due to successes in matching patients to novel therapies and predicting their outcomes.[17]

We have integrated transcriptomic profiling of bulk and geospatially macrodissected diagnostic biopsies with digital pathology and digital spatial profiling in a cohort of clinically matched iCCA patients with extreme divergent outcomes on chemotherapy. This led to identification of the RPLS gene expression signature as a candidate metric of poor clinical outcome and innate chemoresistance. Modelling the RPLS signature in cell lines, single cell RNA-sequencing data, animal models and bulk transcriptomic data from iCCA implicated tumour-induced immune tolerogenicity as a defining hallmark of rapid progression on chemotherapy, as well as establishing a robust and feasible tool to validate for the development of precision chemotherapy in this rare cancer type.

MATERIALS AND METHODS

Comprehensive information on patients and methods are provided online in online supplemental information.

RESULTS

The RPLS cohort

Survival on chemotherapy is heterogeneous among patients with iCCA, including those with comparable clinical features at diagnosis.[9 18] To identify pretreatment molecular features associated with chemotherapeutic outcome that are currently overlooked in the clinic, we identified a cohort of patients with advanced iCCA who diverged as rapid progressors (RP; n=7) or LS (LS; n=6) on chemotherapy. All RP patients survived less than 6 months (half the median OS reported in the ABC-02 trial[4]), whereas all LS patients survived more than 23 months (double the median OS reported in the ABC-02 trial) (p=0.0003; figure 1A). Critically, these patients did not differ in baseline clinical features established during diagnostic workup (figure 1B, online supplemental table S1). No differences were found in haematological or systemic biochemical features, with the exception of higher platelets (p=0.03) and alkaline phosphatase (ALP) levels (p=0.01) in RP patients. All patients were treated with platinum-combination chemotherapy in first line, leading to greater radiological responses in LS patients (p=0.04; figure 1C). Of note, the response rate in the LS cohort was 66%, more than double compared with an unselected population, suggesting that the differences in the long-term control of the disease are not only related to slow growing tumours. Overall, LS patients received a greater number of lines of chemotherapy (p=0.002; online supplemental figure S1A,B). Collectively, the RPLS cohort epitomises extreme divergent outcomes on chemotherapy in an otherwise homogeneous patient population. As such, despite being of small size, we hypothesised that this cohort provides a prime setting, in which to apply molecular profiling to understand benefit from standard-of-care chemotherapy, as well as potentially overall prognosis.

Transcriptomic profiles of pretreatment biopsies differentiate RPs and LSs

We retrieved the pretreatment, diagnostic liver biopsies for patients in the RPLS cohort and performed digital histopathological evaluation (figure 2A). Pixel classification of the entire biopsy tissues showed no differences between RP and LS biopsies in tumorous (p=0.90), epithelial (p=0.39) or stromal (p=0.39) content. Cell detection and classification analysis also revealed no differences in the total number of epithelial (p=0.43) or tumour cells (p=0.39). However, RP biopsies had higher stromal (p=0.03) and lower immune cell (p=0.03) content, suggesting an association between microenvironment composition and chemotherapy outcome, consistently with previous studies.[19 20]

Next, we performed whole-transcriptome profiling of the bulk biopsies using Tempo-seq, a sequencing technology compatible with the limited and fragmented RNA retrievable from archival FFPE biopsies. In total, 504 genes were differentially expressed between RP and LS biopsies (fold-change≥2, $p<0.05$), including 310 genes higher expressed in LS ('LS-high') and 194 genes higher expressed in RP ('RP-high') tissues with distinct biological functions (figure 2B,C, online supplemental table S2). Expression of RP-high and LS-high genes anticorrelated in the biopsies (Spearman's $r=-0.92$, $p=9.4\times10^{-6}$), suggesting opposing biological functions that are associated with chemotherapy outcome (online supplemental figure S2A). Therefore, we derived a formula using the RPLS signature genes ($[\log_2(\Sigma\text{RP-high genes})-\log_2(\Sigma\text{LS-high genes})]_{\text{z-score}}$), with resulting RPLS scores being higher in RP compared with LS biopsies ($p=8.3\times10^{-6}$; figure 2D). As such, we hypothesised that the RPLS signature might represent a metric of the innate chemoresistance potential of iCCA. Inclusion of pretreatment systemic features that differed between patient subgroups (platelets, ALP; figure 2B) or an optimised systemic signature (defined by AIC backwards elimination using all haematological

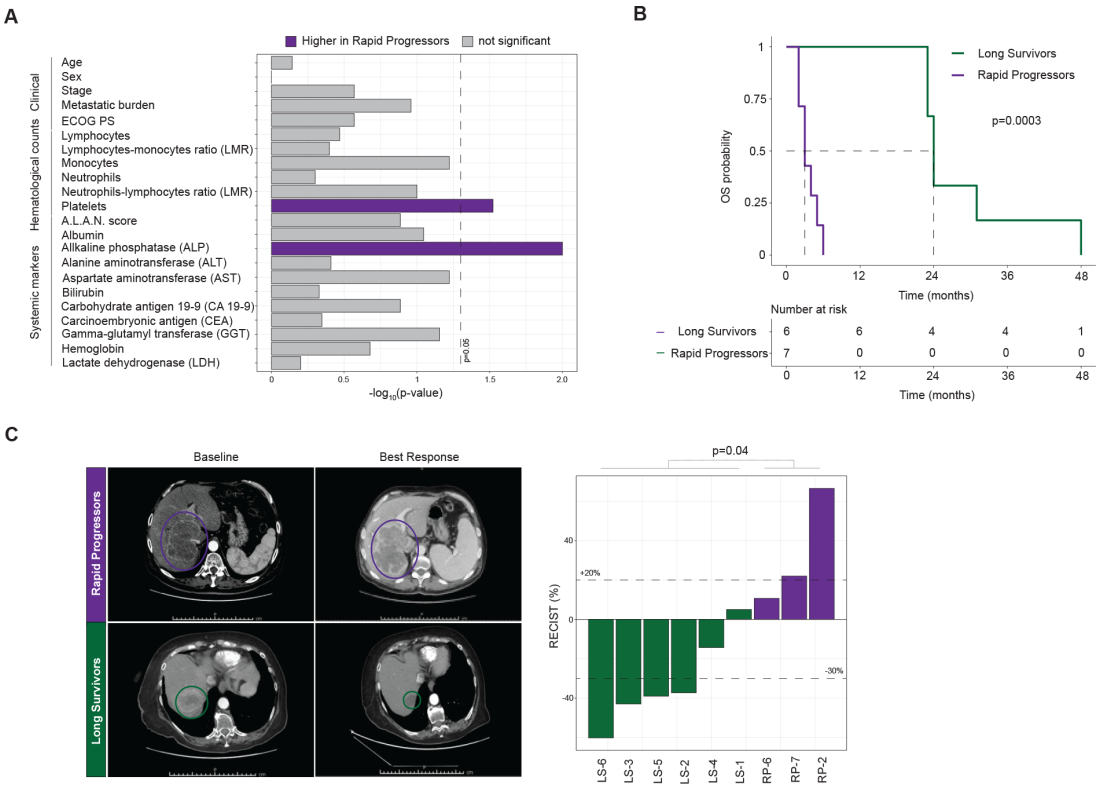


Figure 1 Clinical characteristics and chemotherapy response of patients with intrahepatic cholangiocarcinoma in the RPLS cohort. (A) Kaplan-Meier survival curves with log-rank statistics for overall survivalin the RPLS cohort. (B) Barplot of statistical differences in baseline characteristics between rapid progressor (RP) and long survivor (LS) patients. (C) Representative baseline and best response CT images for an RP and an LS patient. Barplot of best radiological response (Response Evaluation Criteria in Solid Tumours (RECIST)1; Welch t-test). Disease was not measurable for one RP patient, while three RP patients had clinical progression without radiological confirmation (RP-1/RP-3/RP-4/RP-5). ALAN, actual neutrophil count; lymphocyte-monocyetesratio; neutrophil-lymphocytesratio; albumin.

and systemic features; figure 2C) did not improve the predictive performance of the RPLS score in multivariable analysis (online supplemental figure S2B). This further supports the utility of the RPLS score during diagnostic workup of patients with advanced iCCA. Expression of genes involved in gemcitabine uptake/metabolism and previous transcriptomic predictors of chemosensitivity in other cancers failed to differentiate RP and LS biopsies (figure 2D). Collectively, these observations argue that the pretreatment molecular features associated with chemotherapy outcome are distinct in iCCA compared with other cancers, implicating disease and/or context (liver) specificity with treatment response.

Geospatially distinct biopsy regions harbour unique transcriptional programmes that differentiate RP and LS patients

Bulk biopsy transcriptomes capture global signalling but lose biological resolution of specific histopathological regions within the tissues. To address this limitation, we performed geospatial macrodissection followed by whole-transcriptome profiling (Tempo-seq) of tumour cores (TCs; 11 LS, 9 RP), tumour stroma (TSs; 5 LS, 6 RP), invasive fronts (IFs; 3 LS, 3 RP) and non-malignant regions (NRs; 4 LS, 4 RP) from the same RPLS biopsies (figure 2E, online supplemental figure S3). Intrasample transcriptomic heterogeneity and phylotranscriptomic analyses did not differentiate RP and LS biopsies, indicating that intrasample heterogeneity was not associated with outcomes (online supplemental figure S4A,B).

Intertumour heterogeneity is dictated by differential gene and pathway expression, modulated by cell-intrinsic transcriptional programmes. Therefore, we identified differentially expressed genes and pathways for each of the macrodissected regions, as well as predicting transcription factor (TF) activities. No significant differences were found for IFs, so these samples are excluded from further discussion.

In TCs, 639 genes (388 LS-high, 251 RP-high) were differentially expressed between LS and RP biopsies (figure 2F). TF activities of PRDM14, GATA2 and TP63 were higher in RP tissues, potentially regulating 17%, 2% and 2% of the RP-high genes, respectively. In LS biopsies, there was increased activity of SREBF1 and ZNF263, each potentially controlling expression of 1% of the LS-high genes. Within these tumour cell-enriched regions, Notch and Wnt pathways were elevated in RP patients, both developmental programmes associated with poor prognosis in iCCA with incremental potential for druggability.[21 22]

In TSs, LS and RP biopsies differed in expression of 704 genes (637 LS-high, 67 RP-high) (online supplemental figure S5A). CDX2 and KLF4 were more active in RP TSs, potentially regulating 9% and 1.5% of the RP-high genes, respectively. ZNF263 was more active in LS TSs and may control expression of approximately 4% of the LS-high genes. Notch signalling was higher in RP TSs, whereas metabolic processes and MYC targets were higher expressed in LS TS regions. Notably, previously reported signatures of cancer-associated fibroblasts (CAFs) in iCCA[23] did not differ between TSs in our cohort (online supplemental figure S5B).

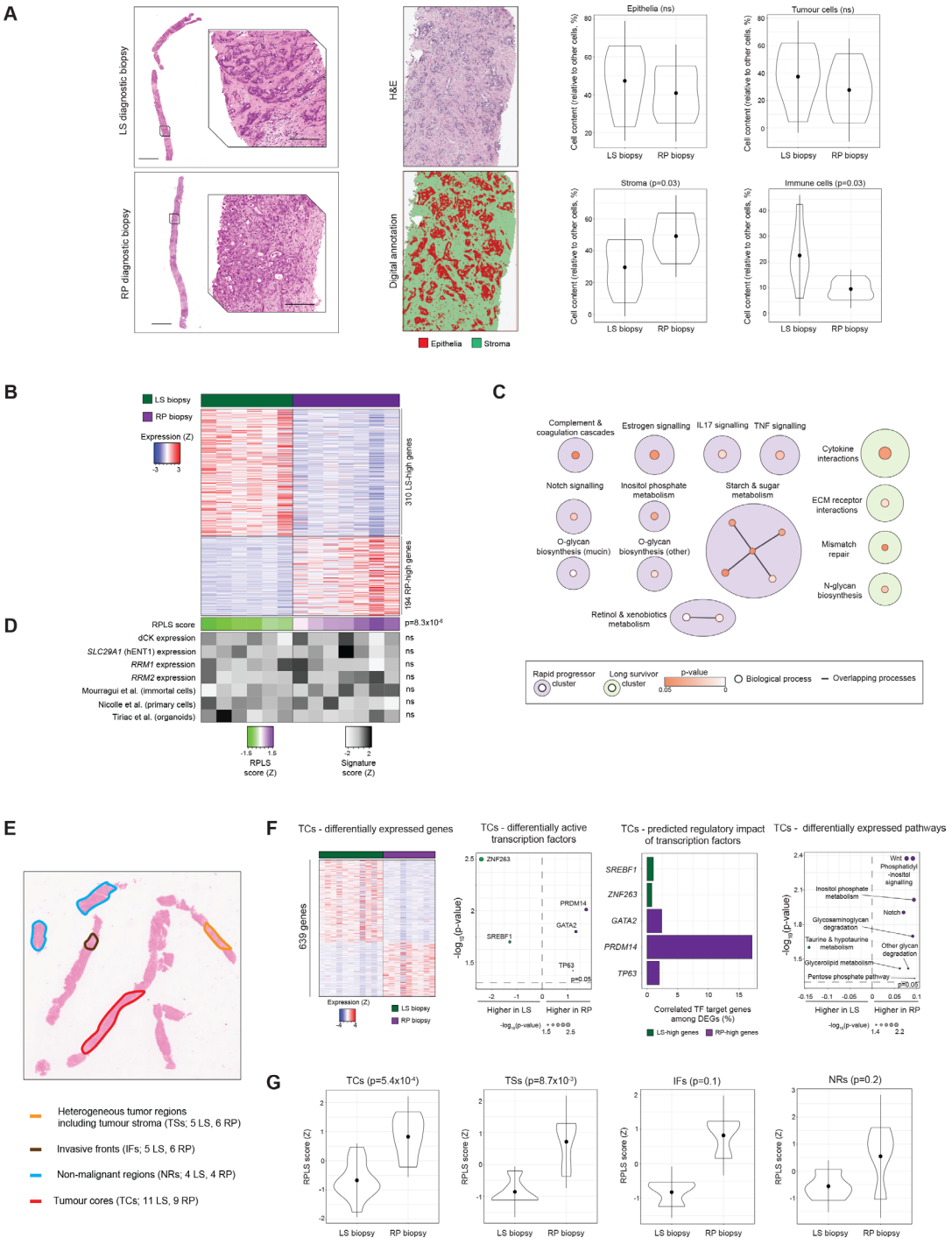


Figure 2 Histopathological and transcriptomic profiles of diagnostic biopsies from the RPLS cohort. (A) Representative H&E images (scale bars: 2 mm—top; 200 µm—bottom), characterisation of the epithelial component of diagnostic biopsies in each region of interest, and differential composition in tumourous, stromal and immune cells following cell segmentation (LS: n=6; RP: n=7; Welch t-test). (B) Heatmap of 504 differentially expressed genes (≥2 fold change, p<0.05; Wilcoxon) between LS and RP biopsies. (C) KEGG pathway over-representation analysis of LS-high and RP-high genes using EnrichmentMap. Overlapping pathways are connected by lines and annotated under a common theme using AutoAnnotate. KEGG: Kyoto Encyclopaedia of Genes and Genomes. (D) Heatmap and differential expression analysis of the RPLS score and previously published metrics of gemcitabine sensitivity in the RPLS cohort. P values were derived by Wilcoxon test. (E) Representative H&E stain of a diagnostic biopsy, indicating histological regions targeted by macrodissection. (F) Differentially expressed genes (≥2 fold change, p<0.05; Wilcoxon), differentially active transcription factors (p<0.05, Wilcoxon test; DoRothEA), differentially expressed pathways (p<0.05, Wilcoxon test; ssGSEA of KEGG and Hallmarks gene lists), and differentially active cytokines (p<0.05, Wilcoxon test; CytoSig) between RP and LS tumour cores (TCs; 11 LS, 9 RP). (G) Differential expression of the bulk tissue RPLS signature in TCs, tumour stroma (TSs; 5 LS, 6 RP), invasive fronts (IFs; 3 LS, 3 RP) and non-malignant regions (NRs; 4 LS, 4 RP) from RP and LS biopsies (Wilcoxon test). LS, long survivor; RP, rapid progressor.

In NRs, the expression of 269 genes (251 LS-high, 45 RP-high) differed between LS and RP biopsies (online supplemental figure S5C). Activities of CEBPB and E2F4 were elevated in RP NRs, potentially regulating 4% and 20% of the RP-high genes, respectively. PAX6 was more active in LS NRs and may control expression of 6% of LS-high genes. Diverse processes were more highly expressed in RP (fructose and mannose metabolism, G2M checkpoint, MTORC1 signalling) and LS (cytokine receptor interactions, haematopoietic cell lineage) NRs, the latter suggesting more widespread immune activity in the liver of LS patients.

Approximately one in four biopsies fail molecular profiling in BTCs due to low tumour cellularity.[24] Evaluation of the RPLS signature (derived from bulk biopsy) in macrodissected biopsy regions revealed RPLS scores to be elevated in both TCs ($p=5.4\times10^{-4}$) and TSs ($p=8.7\times10^{-3}$) from RP patients, but not in IFs (p=0.1) or NRs (p=0.2) (figure 2G). These data suggest that the RPLS signature is tumour-specific and may be sufficiently assessed in stroma-rich biopsies containing relatively few tumour cells, a critical limitation of DNA-based biopsy profiling in the clinic today.

The RPLS signature originates from tumour-intrinsic programmes associated with innate immune dysfunction

The RPLS signature can originate from differences in cell composition and/or behaviour. To investigate this, we pursued a digital cytometry approach (CIBERSORTx) to infer the cellular origin(s) of RPLS signature genes in our biopsies. Among assignable genes, the RPLS signature predominantly originated from tumour cells, followed by tumour-associated myeloid cells, B cells and CAFs (figure 3A). RP-high genes originating from tumour cells were over-represented in immune signalling pathways (IL-17, NFkB, TNF) and drug metabolism (figure 3B), highlighting two plausible mechanisms (impaired immunogenic cell death, enhanced metabolic inactivation) undermining chemotherapy efficacy.[16 25]

As these immune pathways typically require heterotypic signalling between tumour and immune cells, we hypothesised that defective tumour-innate immune dynamics are a defining characteristic of RP tumours. Consistent with this, immune infiltrates significantly differed between LS and RP TCs (figure 3C). Cytolytic scores were higher in LS TCs (p=0.04), indicating proficient anti-tumour cytotoxicity in these biopsies (figure 3D). Cytokine activity profiles also differed between LS and RP TCs (figure 3E). The interferon responsible for activation of antitumour immunity, interferon-γ (IFNG; p=0.02), and the proapoptotic cytokine, TNF-related apoptosis-inducing ligand (TRAIL; p=0.004), were both more active in LS TCs, indicating effective cytotoxicity and cell death. In contrast, granulocyte colony-stimulating factor (GSCF), a promyelopoiesis and anti-inflammatory cytokine associated with myeloid-derived suppressor cells,[26] had increased activity in RP TCs (p<0.05), supporting an immunosuppressive phenotype in these tumours.

To experimentally verify immune cell dysfunction, we performed digital spatial profiling (DSP, GeoMx) of tumour-infiltrating myeloid cells (CD68) using a targeted immunobiology panel (78 genes). Myeloid cells were chosen as they were the second most dominant contributor to the RPLS signature, pDCs were enriched in RP TCs and myeloid cells can promote chemosensitivity independent of adaptive immune cells.[27] In LS TCs, myeloid cells expressed high levels of cytotoxic effectors (*BCL2*, *GZMB*, *NKG7*, *TNF*) and cytokines (*CCL5*, *CXCR6*, *IL12B*, *IL15*, *TNF*) (figure 3F). RP tumour-infiltrating myeloid cells expressed high levels of molecules associated with immunosuppression (*CD58*, *CD80*, *CD163*), as well as monocyte activation and dendritic cell maturation (CD40). Consistent with failed immunogenic clearance, Ki-67 staining indicated tumour cells were proliferating faster in RP TCs (p=0.02; figure 3G). Altogether, these data implicate pretreatment antitumour immunity as a characteristic required for chemotherapy benefit.

Myeloid cells are highly diverse, so we analysed myeloid-specific RPLS scores (13 RPLS signature genes with predicted myeloid origin by digital cytometry) in scRNA-seq data generated from immune cells (CD45+) of 3 iCCA patients undergoing tumour resection.[28] Under adjuvant capecitabine, one patient exhibited short disease-free survival (9 months; S-DFS) and two others exhibited long disease-free survival (L-DFS; ≥24 months) (figure 3H). Among the eight identified cell subpopulations, CD14 monocytes were more abundant in the L-DFS cases (p=0.03). However, evaluation of myeloid-specific RPLS scores identified increased signature expression across diverse myeloid cell types in S-DFS ($CD140^+$ monocytes, $ID3^+$ macrophages, $MARCO^+$ macrophages, $TREM2^+$ macrophages, $CD1C^+$ cDC2, lymphoid-like cells). This implicates widespread behavioural changes of diverse myeloid subpopulations with diminished chemotherapy outcome.

RP-like and LS-like iCCA are dependent on unique gene networks for survival in vitro

In patient biopsies, the dominant origin of the RPLS signature is the tumour cells. To determine whether immortalised iCCA cell lines recapitulate aspects of these tumour-intrinsic programmes, we integrated the transcriptome profiles of 25 iCCA cell lines with transcriptome data from our biopsy TCs (online supplemental figure S7A), annotating 52% (13/25) of cell lines as RP-like and the remainder as LS-like (online supplemental figure S7B). RP-like cells had decreased in vitro gemcitabine sensitivity (p=0.02; Figure S7C), trended towards association with *KRAS* mutations (p=0.07) (online supplemental figure S7D), and differentially expressed pathways (KRAS and P53 pathways, glycolysis) compared with LS-like cells (online supplemental figure S7E). These observations suggest that immortalised iCCA cell lines can provide minimalistic avatars to study some tumour-intrinsic aspects of RP-like and LS-like patient phenotypes in vitro. As chemoresistance is associated with distinct biology, RP-like phenotypes should also be associated with fitness tradeoffs, specifically genes which become more or less important for tumour cell survival. Using genome-wide CRISPR inactivation data (DepMap), we identified differential gene dependencies (Wilcoxon p<0.05) between RP-likeand LS-like iCCA that fell into common biological networks (online supplemental table S3). RP-like iCCA was more dependent on 48 network-based genes for survival (Notch, p53 and TGF-β signalling; online supplemental figure S7F), whereas LS-like iCCA was more dependent on 62 network-based genes (Hedgehog, Ras signalling; online supplemental figure S7G). Approximately 44% (21/48) and 26% (16/62) of these genes are predicted to be potentially druggable for RP-like and LS-like phenotypes, respectively (online supplemental table S3). These subgroup-specific dependencies indicate that considerable drug development opportunities remain for patients with iCCA, including those with RP-like phenotypes on standard-of-care chemotherapy.

RP-like tumour cells engage immunosuppressive microenvironments via myeloid and T cell communication

Although in vitro models could partially recapitulate RPLS-associated oncogenic programmes, they lack interacting

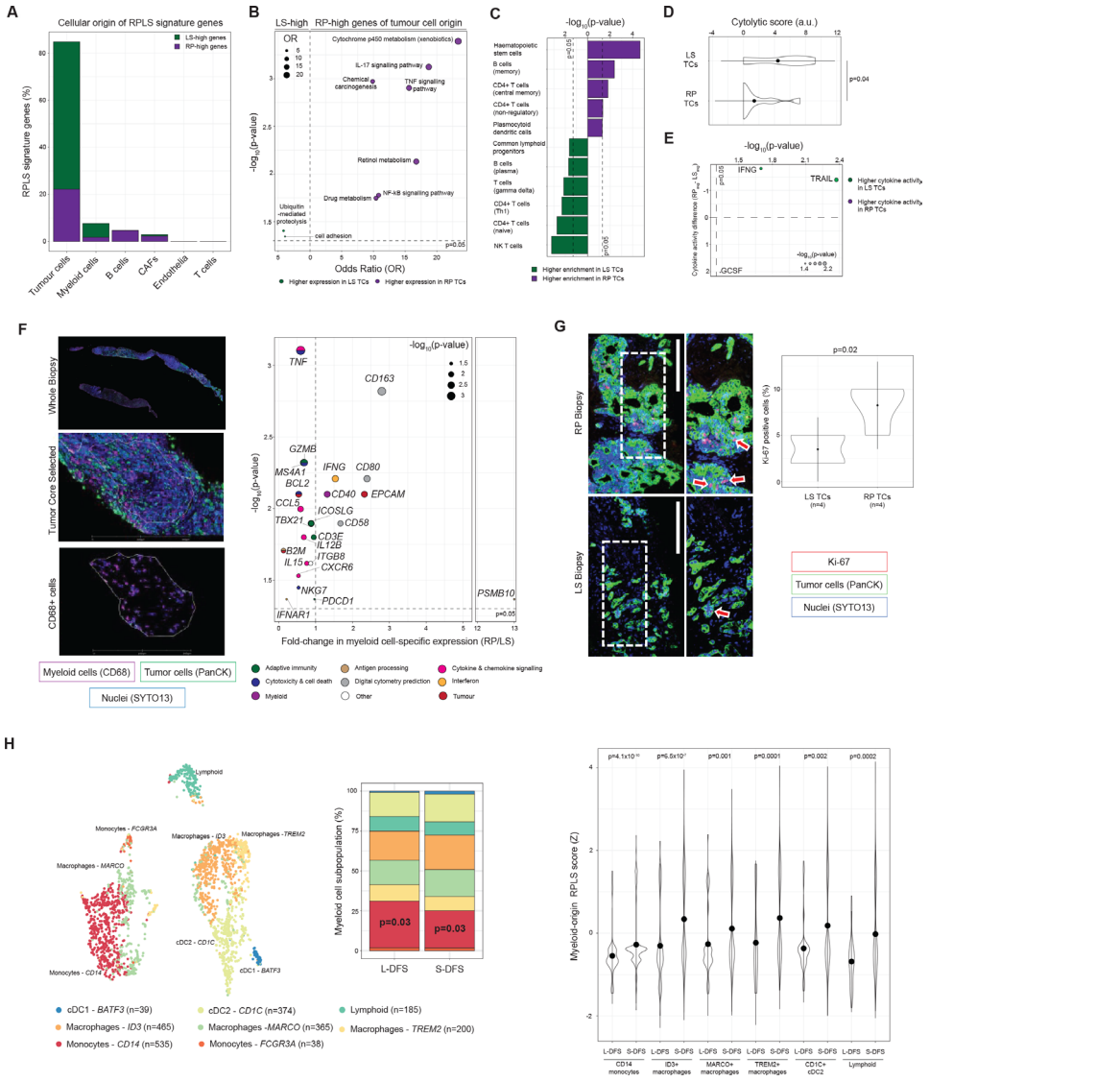


Figure 3 Cellular origins and tumour-immune dynamics associated with the RPLS signature. (A) Barplot of cell type-associations of RPLS signature genes. Genes were only assigned to cell types if their associations were supported by two independent single cell RNA-sequencing datasets. (B) KEGG pathway over-representation analysis of tumour-origin RPLS signature genes. (C) Differential enrichment (Wilcoxon test) of immune cell type signatures in rapid progressor (RP) and long survivor (LS) tumour cores (TCs) determined by cellular deconvolution (xCell). (D) Differential cytolytic scores (Wilcoxon test) between RP and LS TCs. (E) Differential cytokine activities (Wilcoxon test) between RP and LS TCs determined by CytoSig. (F) Representative multiplex immunofluorescence images of TCs undergoing RNA extraction from myeloid (CD68+) myeloid cells using the Digital Spatial Profiling (GeoMx) platform with regions of interest identifed. Volcano plot of differentially expressed genes (Immune Pathways Panel (NanoString) plus 5 custom targets derived from digital cytometry) in tumour-infiltrating myeloid cells from LS (n=6) and RP (n=6) TCs. P values were computed by Wilcoxon test. (G) Representative Ki-67 staining in an LS and RP TCs, including differential proliferation analysis (Welch t-test). (H) tSNE plot of myeloid subpopulations identified in immune-enriched single cell RNA-sequencing data from three resected iCCA. Frequency barplot (p values from χ^2 test) comparing the abundance of myeloid subpopulations in patients without (long disease-free survival, L-DFS) and with (short disease-free survival, S-DFS) recurrence under adjuvant treatment with capecitabine. Differential expression (Wilcoxon test) of the myeloid-origin RPLS signature in myeloid subpopulations. iCCA, intrahepatic cholangiocarcinoma.

microenvironmental cells that are pronouncedly reprogrammed between RP and LS patient biopsies. Using two scRNA-seq iCCA datasets, we annotated patient tumours as RP-like or LS-like (figure 4A; Methods). Tumour cells mirrored pathway expression as observed in the immortalised cell lines, including elevated cytosolic DNA sensing, glycolysis, gluconeogenesis, and P53 pathways in RP-like tumours (figure 4B, online supplemental table S4). RP-like and LS-like tumour cells were also characterised by unique cytokine and TF activity profiles (figure 4C, online supplemental table S5), highlighting stable cell behavioural states. Microenvironment cells (CAFs, myeloid, T cells) also exhibited unique cytokine and TF activity profiles that consistently differed between RP-like and LS-like tumours (figure 4C, online supplemental table S5). Metabolic flux modelling identified increased methionine and losartan utilisation by RP-like myeloid cells, metabolites that are required for anti-tumour immunity[29 30] but appear to be otherwise restricted by the myeloid compartment, potentially indicating metabolic competition as a contributory factor to immune dysfunction (online supplemental figure S8). To pinpoint tumour–microenvironment interactions that support

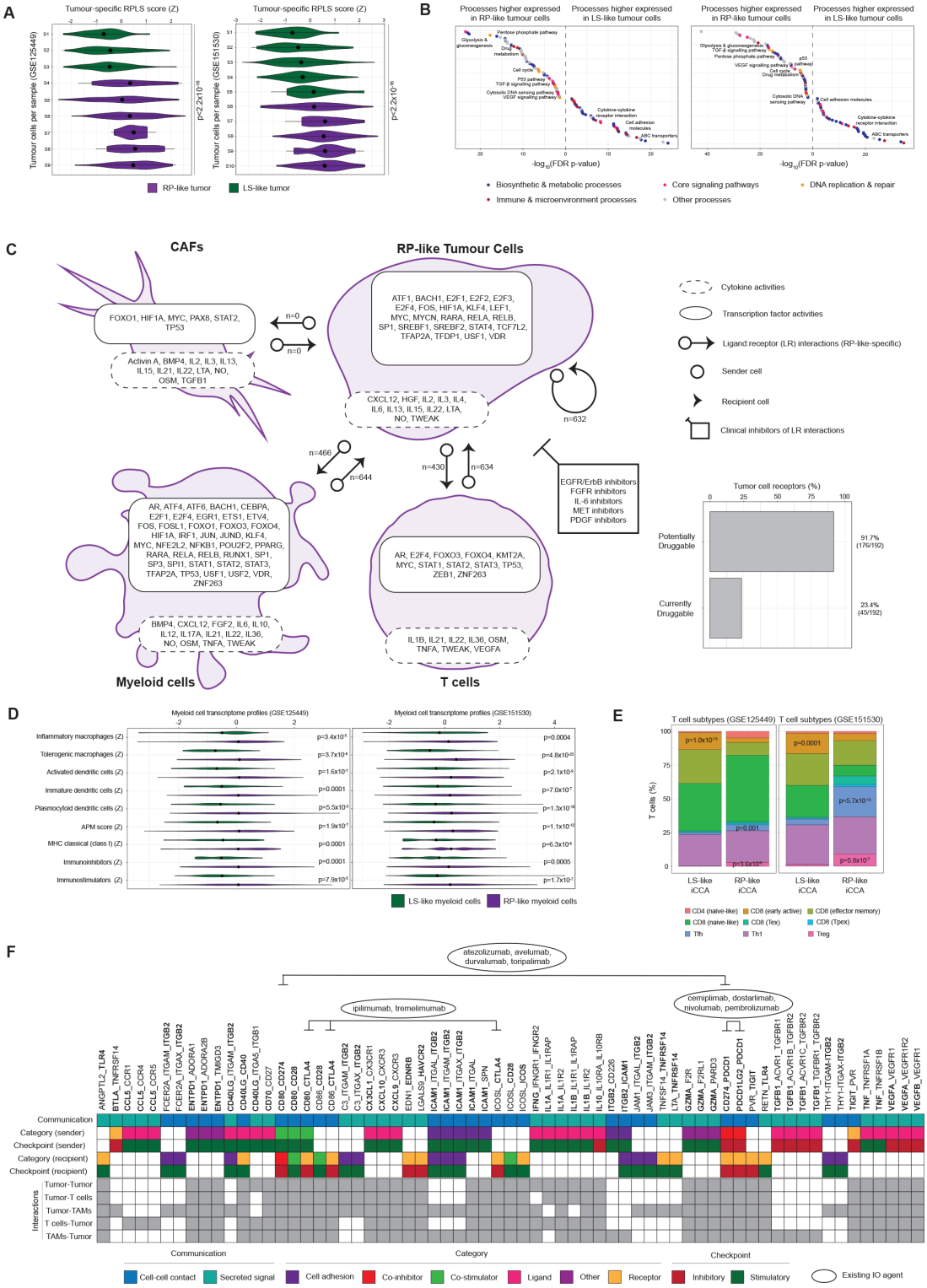


Figure 4 Modelling the RPLS signature in iCCA single cell RNA-sequencing data. (A) Annotation of tumours as long survivor (LS)-iike or rapid progressor (RP)-like based on tumour cell expression of the tumour-origin RPLS signature (ESCAPE tool, p values derived by Wilcoxon test). (B) Differentially expressed pathways and processes (ESCAPE with KEGG and Hallmarks gene lists) between LS-like and RP-like tumour cells. (C) Cell type-specific transcription factor activities (DoRothEA), cytokine activities (CytoSig) and ligand:receptor interactions (CellChat) unique to RP-like tumours in GSE125449 and GSE151530, including the potential and current druggability of tumour surface receptors. (D) Differential expression of myeloid cell type and functional signatures in LS-like and RP-like myeloid cells (ESCAPE, Wilcoxon test). (E) T cell subtype annotation using ProjectTILs (p values from Fisher's exact test). (F) RP-specific ligand-receptor interactions tumours in GSE125449 and GSE151530 involving immunomodulatory targets (highlighted in bold; defined by CRI iAtlas). iCCA, intrahepatic cholangiocarcinoma.

these altered behavioural states, we identified ligand-receptor (LR) interactons (CellChat) unique to RP-like and LS-like iCCA. Whereas no LR interactions were unique to and reproducible for LS-like iCCA, RP-like iCCAs were characterised by many unique LR interactions: tumour-to-tumour (n=632), tumour-to-myeloid (n=466), myeloid-to-tumour (n=644), tumour-to-T cell (n=430), T cell-to-tumour (n=634) (figure 4C; online supplemental table S6). No unique interactions were found for CAFs, but this may be due to their under-representation in non-enriched scRNA-seq datasets. Receptors expressed on the surface of tumour cells have historically provided impactful therapeutic targets. Focusing on the RP-like tumour 'surfaceome' with evidence of LR interactions, roughly 92% (176/192) are predicted to be potentially druggable, whereas 23% (45/192) are currently actionable with clinically approved compounds (online supplemental table S7). These agents include IL-6R monoclonal antibodies which have been shown to improve chemotherapy response in a subcutaneous transplant model of iCCA,[31] as well as several FGFR inhibitors already approved for second-line use in iCCA.[32]

Variation in microenvironment behavioural states might reflect differences in abundance of cell subpopulations. RP-like myeloid cells had higher signature expression for dendritic cells (figure 4D), including plasmocytoid dendritic cells that were increased in RP TCs (figure 3C), as well as inflammatory and immune tolerogenic subtypes of liver-associated macrophages. RP-like myeloid cells were defined by higher antigen expression scores (including classical MHC presentation), but also co-occuring expression of immunostimulator (including CD275 and IL6) and immunoinhibitor (including CSF1R, CTLA4, IDO1, LAG3, PDCD1) signatures (figure 4D). Subclassification of T cells into their functional ontogenies identified more regulatory T (Tregs) and T follicular helper cells in RP-like tumours, and a reduced amount of early active CD8 T cells (figure 4E). Collectively, these findings implicate tumour-induced immune tolerogenicity as a hallmark of RP phenotypes, defined by the activity of immunosuppressive cytokines (especially IL10[33]), antigen presentation in the presence of diverse immunoinhibitors and accumulation of regulatory T cells.

Combining chemotherapy with the PD-L1 inhibitor, durvalumab or the PD-1 inhibitor, pembrolizumab, improved OS of patients with BTC in the TOPAZ-1 trial[5] and KEYNOTE-966 trial,[6] respectively. Combination chemotherapy-immunotherapy represents the new standard-of-care, although mechanistic explanations and biomarkers for patient selection are lacking. By investigating LR interactions involving immunomodulatory targets (CRI-iAtlas), we identified interactions between CD274 (PD-L1), its receptor PDCD1 (PD-1) and its costimulator CD80 (B7-1) exclusively in RP-like tumours (figure 4F). These observations suggest that the immune escape mechanisms employed by RP-like tumours might render them susceptible to checkpoint inhibitors compared with LS-like tumours. Other RP-specific targets for potential immunomodulatory inhibitors include immunosuppressive cytokines (IL10), ligands (PDCD1LG2, TGFB1, VEGFA, VEGFB) and receptors (EDNRB).

The RPLS signature is prognostic and pathobiologically distinct in early-stage iCCA

Although the RPLS signature was identified in advanced iCCA biopsies, its origin in tumour-immune interactions suggests that the determinants of chemotherapy response might be established early during cholangiocarcinogenesis. Therefore, we investigated the RPLS signature in 637 fresh-frozen iCCA.[19 34–37] As with our biopsy cohort of advanced patients, LS-high and RP-high genes were anticorrelated in all resected cohorts, emphasising the reproducibility of the observed inter-network signalling (online supplemental figure S9). High RPLS scores (above median) were consistently associated with decreased 5-year survival in all cohorts (figure 5A). High RPLS scores were also associated with inferior survival in an FFPE cohort of 119 iCCA,[38] but not in 219 extrahepatic CCA[39] (online supplemental figure S10). As such, the RPLS signature appears to be prognostic exclusively in iCCA, further suggesting the importance of the liver microenvironment in the RPLS signature. Data on adjuvant or subsequent palliative therapy are not available and therefore association with chemoresistance cannot be deducted.

RPLS scores were lower in tumours harbouring *IDH1 or IDH2* mutations (all 3 cohorts with data available), as well as *FGFR2* fusions (2/3 cohorts) (figure 5B). Conversely, tumours with *KRAS* or *TP53* mutations had higher RPLS scores in all cohorts, indicative of a higher baseline level of innate chemoresistance. Evaluation of 17 mouse models of iCCA revealed RPLS scores to become elevated in 24% (4/17) of model tumours relative to their controls, among which three involved insults in *Kras* and/or *Tp53* (figure 5C). RPLS scores were also increased in intraductal papillary neoplasm of the bile duct relative to ductular proliferation and normal tissues in cholangiocyte-specific $Kras^{G12D}$-expressing mice (online supplemental figure S11A), suggesting that RPLS-associated chemoresistance is established early during cholangiocarcinogenesis. Combining RPLS scores with associated genomic alterations in multivariable models revealed the RPLS signature to consistently provide genotype-indepedent prognostic information across cohorts (figure 5D).

Clinically, RPLS scores were consistently higher in iCCA with liver fluke infection, advanced grade, perineural involvement and portal tract spreading (figure 5E; online supplemental table S8). RPLS scores were higher in tumours with advanced stage and lymph node invasion (3/4 cohorts with data available) and positively correlated with serum albumin, CA19-9, CEA and GGT (figure 5E). Higher RPLS scores were associated with large duct-type iCCA in a small reseceted cohort (p=0.007; GSE107943; online supplemental figure S11B), consistent with our mutation observations (*KRAS* and *TP53*) (figure 5B) and previous associations between morphology and chemotherapy outcome.[40] In the Dong cohort where transcriptome and clinicopathological data are present for all patients, the RPLS signature is an independent prognostic variable after correcting for its clinicopathological correlates (online supplemental figure S11C), highlighting the potential utility of this metric in the resected setting where surgical and post-surgical specimen evaluation is possible (unlike the advanced setting).

Transcriptomically, the RPLS signature was associated with expression of metabolic pathways (glycolysis and gluconeogenesis, pentose phosphate pathway, phosphatidylinositol signalling, citrate signalling) (figure 5F). The RPLS signature was also reproducibly associated with key TF activities, immune cell infiltrates and cytokine activities. Therefore, RPLS-associated oncogenic programmes exhibit robust pathobiological associations and reflect a significant source of intertumour heterogeneity.

The RPLS signature captures a liver-specific oncogenic programme and predicts chemotherapy outcome in liver-metastatic colorectal cancer

Our observations that the RPLS signature is prognostic in iCCA but not eCCA indicate that the liver microenvironment plays an important role. As we found the RPLS signature is not prognostic

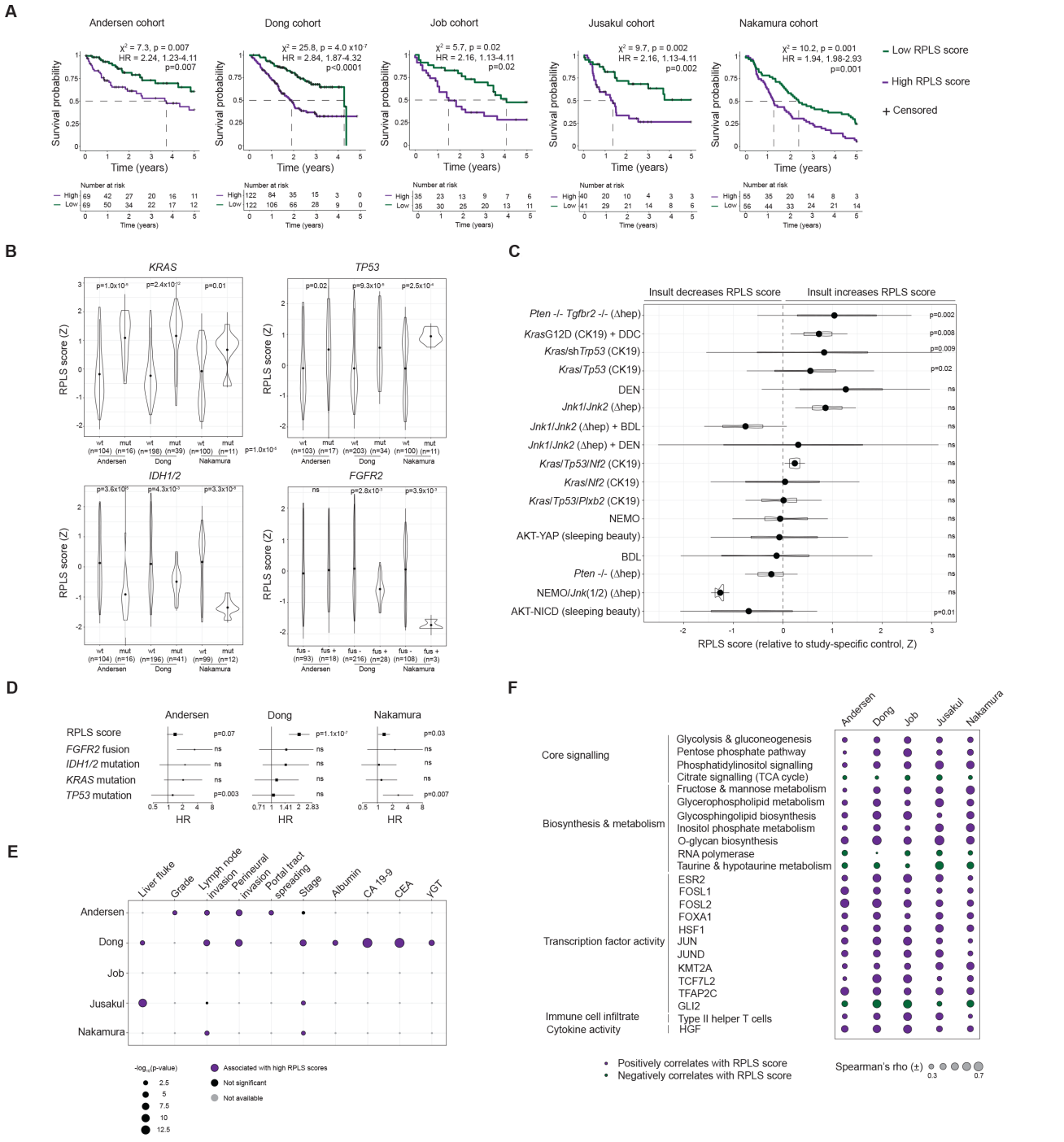


Figure 5 Prognostic, clinicogenomic and transcriptomic associations of the RPLS signature in 653 resected iCCA. (A) Kaplan-Meier survival curves with log-rank statistics of resected iCCA stratified into high (>median) or low (<median) RPLS score groups across five resected cohorts. (B) Differential expression (Wilcoxon test) of RPLS scores between iCCA stratified by genotype across three cohorts where RNA-profiling and DNA-profiling data are available. (C) Differential expression (Welch t-test) of RPLS scores between mouse models of iCCA and their study-specific controls. Cell type-specific induction of genetic insults are indicated as CK19 (biliary/progenitor cell) or hep (hepatocyte). Δ: deletion; (D) Forest plot of Cox proportional hazards statistics derived from multivariable analysis of RPLS scores and tumour genotypes. (E) Associations of RPLS scores with clinicopathological variables. γ-GT: γ-glutamyltransferase; CA 19–9: carbohydrate antigen 19–9; CEA: carcinoembryonic antigen. (E) Forest plot depicting ORs and p values from multivariable analysis of RPLS scores, tumour stages and genotypes across three resected cohorts. (F) Correlation plot of RPLS score with core signalling pathways (KEGG), biosynthesis and metabolic processes (KEGG), transcription factor activity (DoRothEA), immune cell infiltrate (xCell) and cytokine activity (CytoSig) across five resected cohorts. Spearman's r is only indicated for significantly correlated features (FDR p<0.05). BDL, bile duct ligation; DDC, 3,5-diethoxycarbonyl-1,4-dihydrocollidine; DEN, diethylnitrosamine; FDR, false discovery rate; iCCA, intrahepatic cholangiocarcinoma; mut, mutant; ns, not significant; wt, wild-type.

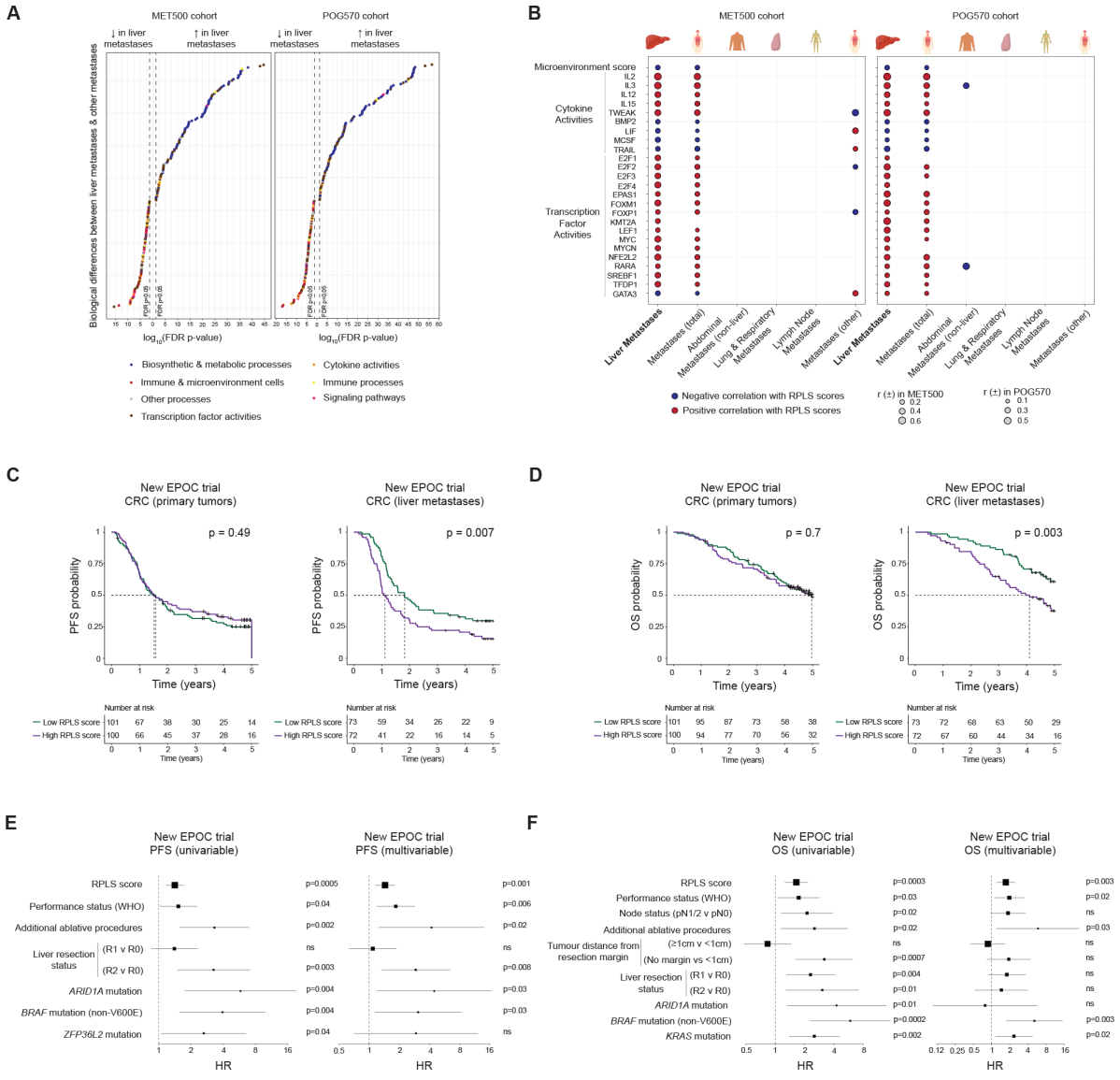


Figure 6 Pathobiological associations and predictive potential of the RPLS signature in liver metastases. (A) Differential expression of pathways and processes (ssGSEA with KEGG and Hallmarks gene lists) between liver metastases and other metastases in MET500 (n=490) and POG570 (n=438) cohorts (p values derived from Wilcoxon test). (B) Biological processes uniquely associated with the RPLS signature in liver metastases. (C, D) Kaplan-Meier survival curves with log-rank statistics of primary colorectal cancer tumours (n=204) and liver metastases (n=145) stratified by RPLS score (above and below median) for (C) progression-free survival and (D) overall survival in the New EPOC trial. (E–F) Univariable and multivariable Cox proportional hazards analysis of RPLS scores and other significant clinicogenomic variables for (E) progression-free and (F) overall survival. ns, not significant.

in HCC (online supplemental figure S12) which is treated using targeted therapy-based regimens, we hypothesised that the RPLS signature is predictive for primary or metastatic liver tumours treated with chemotherapy. Accordingly, we investigated the RPLS signature in two basket cohorts of metastatic cancers: MET500 (n=484) and POG-570 (n=438).[41 42] Consistent with the biological functions of the liver, liver metastases had higher expression of metabolic processes compared with other metastases in both cohorts (figure 6A, online supplemental table S9). These included glycolysis and gluconeogenesis, processes associated with increasing RPLS scores across iCCA models (advanced and resected patient tissues, immortalised cell lines, scRNA-seq). However, liver metastases were also remarkably depleted in diverse immune cells and immune processes compared with other metastatic sites, and associated with unique cytokine and TF activities, including a negative association with TRAIL activity as observed in iCCA biopsies (figure 6B). Exclusively in the liver, RPLS scores negatively correlated with microenvironment signalling, highlighting the extensive immunosuppressive capacity of the liver in coordination with specific oncogenic programmes.

Finally, we demonstrated the predictive utility of the RPLS signature in cancers with liver metastases undergoing chemotherapy. To evaluate this, we applied the RPLS signature to primary tumours and resected liver metastases from colorectal cancer patients receiving preoperative and postoperative chemotherapy with or without cetuximab in the phase III New EPOC trial.[43] In the total cohort composed of both treatment arms, RPLS scores were not associated with progression-free survival (PFS; p=0.49) or OS (OS; p=0.7) when evaluated in primary tumours (figure 6C,D). However, high RPLS scores were associated with inferior PFS (p=0.007) and OS (p=0.003) when measured in liver metastases. Liver metastasis RPLS scores were

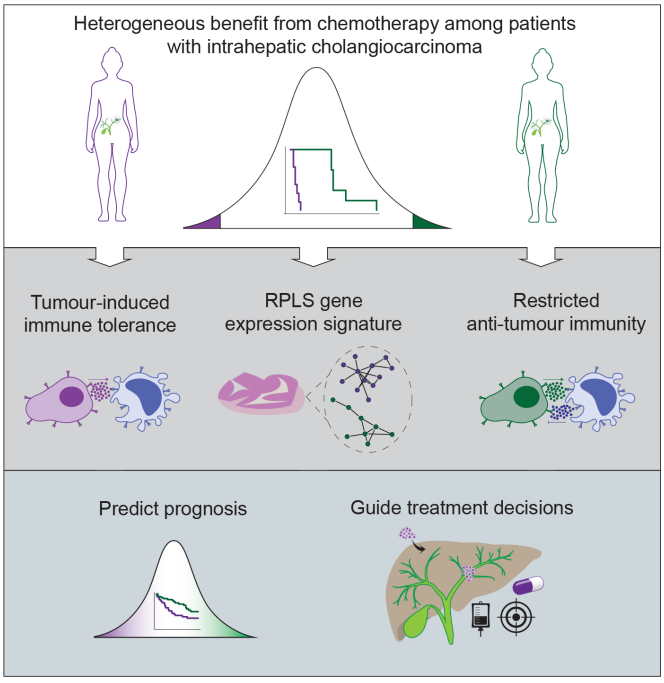


Figure 7 Graphical schematic of RPLS-associated chemosensitivity in iCCA. Heterogeneous benefit from chemotherapy is associated with tumour-induced tolerogenicity and restricted anti-tumour immunity. Pending further validation, the RPLS signature could clinically empower accurate prognostic prediction and guide treatment selection. iCCA, intrahepatic cholangiocarcinoma.

predictive of PFS (HR 1.4, 95% CI 1.2 to 1.7; online supplemental table S10) and OS (HR 1.6, 95% CI 1.3 to 2.2; online supplemental table S11) in univariable analyses in the New EPOC trial (figure 6E). These RPLS scores remained independent predictors of PFS (HR 1.4, 95% CI 1.2 to 1.8) and OS (HR 1.7, 95% CI 1.2 to 2.4) after adjusting for clinicopathologic and genomic predictors in multivariable analyses (figure 6F). Based on the robust predictive performance of the RPLS signature in primary (iCCA) and metastatic (colorectal) liver tumours treated with chemotherapy, continued clinical evaluation of this metric is warranted.

DISCUSSION

Heterogeneous benefit from chemotherapy challenges the paradigm of this universal standard-of-care for patients with iCCA. In this study, we aimed to identify the biology associated with different clinical outcomes in iCCA undergoing chemotherapy (figure 7). Based on our data, we hypothesise that multilayered mechanisms contribute to RPLS-associated chemoresistance, involving tumour-intrinsic processes, tumour-myeloid interactions and tumour-T cell interactions. Decreased in vitro sensitivity of RP-like cell lines likely originates from metabolic reprogramming such as increased glycolysis which has been shown to confer chemoresistance to cytotoxic agents in diverse cancers.[44 45] In immunosuppressive microenvironments, macrophages can further limit the antitumour effects of chemotherapy, by metabolically inactivating gemcitabine prior to drug uptake by cancer cells[27] and actively decreasing the duration of mitotic arrest of tumour cells following induction of DNA damage.[46] Copresentation of tumour antigens alongside immunosuppressive molecules by myeloid cells promotes regulatory T cell expansion and tumour tolerogenicity, culminating in a steady state of immunological inertia. In this context, the unique immunoregulatory capabilities of the liver microenvironment appear to be critical, replete with atypical dendritic cells that normally function to suppress systemic immune responses arising from continuous exposure to antigens and gut microbial byproducts. Such systemic regulation of immunity by the liver may explain why only RPLS scores from liver tumours (primary and metastatic) are associated with chemotherapy outcome in our studies. A similar phenomenon has been reported for immunotherapy, in which liver metastases uniquely blunt checkpoint inhibitor response through hepatic macrophage-mediated elimination of CD8 T cells.[47] A critical question emerging is why only certain tumours in the liver can trigger high RPLS scores leading to diminished chemotherapy response. One commonality across in vitro and ex vivo analyses was high glycolytic pathway expression in models with high RPLS scores. Increased glucose consumption and lactic acid production are associated with immunosuppressive microenvironments,[48] and myeloid cells are the highest cellular consumers of glucose,[49] potentially establishing a metabolically initiated and competitive tumour niche.

Advancing the RPLS signature into a clinical grade test will require further optimisation in large retrospective cohorts. This will include statistically optimising the signature into a smaller gene panel with a weighted formula and establishing reference value ranges for interpretation of individual patient risk (radiological response, PFS, OS) if treated with chemotherapy. Pending continued validation, clinical implementation of the optimised RPLS signature could prioritise patients for neoadjuvant chemotherapy (high-risk resectable or borderline resectable) and support earlier tumour molecular profiling in predicted RP patients to identify alternative first-line treatment strategies. For predicted RP patients currently lacking alternative treatment strategies, further therapeutic evaluation of RP-associated biology (Notch, TGF-β, IL-6, immune checkpoints) is warranted. A complementary transcriptome-driven approach will also be important to predict benefit from chemotherapy-immunotherapy combinations. Pursuing such a precision chemotherapy approach will be critical for optimising patient management and decision-making as the treatment landscape continues to evolve in the present and future.

Author affiliations
[1]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Department of Health and Medical Sciences, Copenhagen, Denmark
[2]Division of Oncology, Department of Oncology and Hematology, University Hospital Modena, Modena, Italy
[3]Clinical and Experimental Medicine, University of Modena and Reggio Emilia, Modena, Italy
[4]School of Cancer Sciences, University of Glasgow, Glasgow, UK
[5]Department of Anatomical, Histological, Forensic Medicine and Orthopaedic Sciences, Sapienza University of Rome, Roma, Italy
[6]Division of Pathology, University of Modena and Reggio Emilia, Modena, Italy
[7]Cancer Biomarkers Research Group, Department of Medical Oncology, Hospital Universitari Arnau de Vilanova, Lleida, Spain
[8]Cancer Research UK Beatson Cancer Research Institute, Glasgow, UK
[9]Bioinformatics Unit, IRCCS Humanitas Research Hospital, Rozzano, Italy
[10]Liver Cancer Translational Research Laboratory, BCLC Group, Liver Unit and Pathology Department, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[11]Medical Oncology, IRCCS Humanitas Research Hospital, Milan, Italy
[12]Department of Oncology, Herlev Hospital, Herlev, Denmark
[13]Addenbrooke's Hospital, Cambridge, UK
[14]Department of Oncology, University of Oxford, Oxford, UK
[15]Surgery, University of Southampton, Southampton, UK
[16]Department of Oncology, University College London, London, UK
[17]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clínic, University of Barcelona, Barcelona, Spain
[18]Icahn School of Medicine at Mount Sinai, New York, New York, USA
[19]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

[20]Institute of Pathology, University of Regensburg Faculty of Medicine, Regensburg, Germany
[21]Medical, Surgical, and Clinical Sciences, University of Sassari, Sassari, Italy
[22]MRC HGU, The University of Edinburgh MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK
[23]CRUK Scotland Cancer Centre, Glasgow-Edinburgh, UK
[24]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[25]Internal Medicine and Hepatology Unit, Department of Gastroenterology, IRCCS Humanitas Research Hospital, Milan, Italy
[26]Beatson West of Scotland Cancer Centre, Glasgow, UK

Twitter Antonio Pea @antonio pea, Ana Lleo @analleo, Nigel B Jamieson @nigeljamieson and Jesper B Andersen @jaboeje

Acknowledgements This work was performed under the frame of the European Network for the Study of Cholangiocarcinoma (ENSCCA) and the COST Action Collaboration (COST Action CA18122 European Cholangiocarcinoma Network Euro-Cholangio-Net). Part of the data used in the preparation of this manuscript were obtained and analyzed from datasets generated by the S:CORT consortium and from the New EPOC trial. S:CORT was funded by a UK Medical Research Council (MRC) Stratified Medicine Consortium programme grant (MR/M016587/1) with support also from Cancer Research-UK. New EPOC was also supported by Cancer Research-UK. The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The authors wish to recognise the contribution of the appropriate specimen donors and research groups underlying TCGA data.This work was supported by the CRUK Scotland Centre [grant number 100006]. For the purpose of open access, the author(s) has applied a CreaVve Commons AQribuVon (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Contributors Conceived the idea: CJO'R, MS, JBA and CB; designed and implemented the study: CJO'R, MS, NBJ, CB and JBA; performed experiments and data analyses: CJO'R, MS, CR, GC, HL, AP, MGP, LRB, FA, RU-G, CN, PS-S, EG, DO, SC, AS, FG, DFC, LB, NBJ, GL, MD, CB and JBA; interpreted omics data: CJO'R, RM, PK, DH, DS, CB and JBA; patient recruitment: MS, SC, SP, ED, JNP, JB, AS, FG, JML, AL, GL, MD and CB; clinical data collection and pathological assessment: GC and MS; wrote the manuscript: CJO'R, MS, CB and JBA; all authors read, edited and approved the manuscript. Guarantor of the study: CB and JBA.

Funding JBA is a member of the scientific advisory board at SEALD, Norway and reports scientific consultancies for QED Therapeutics and Flagship Pioneering. JBA has received research funding from Incyte not related to this study. CB received honoraria as speaker (Astrazeneca, Incyte) and consultant (Incyte, Servier, Boehringer Ingelheim, Astrazeneca), received research funds (Avacta, Medannex, Servier) and her spouse is an employee of Astrazeneca. JML is receiving research support from Eisai, Bayer HealthCare Pharmaceuticals, Ipsen and consulting fees from Eisai, Merck, Bristol-Myers Squibb, Eli Lilly, Roche, Genentech, Ipsen, Glycotest, AstraZeneca, Bayer HealthCare Pharmaceuticals, Omega Therapeutics, Mina Alpha, Boston Scientific, Exelixis, Bluejay and Captor Therapeutics. RM has received consulting and lecture fees from Servier and Roche and travel and education funding from MSD, Eli Lilly, Bayer, Roche, Astrazeneca. AL reports receiving consulting fees from Intercept Pharma, Alfa Sigma, Takeda, and Albireo Pharma, and speakers' fees from Gilead, Abbvie, MSD, Intercept Pharma, AlfaSigma, GSK and Incyte.

Competing interests JBA is a member of the scientific advisory board at SEALD, Norway and reports scientific consultancies for QED Therapeutics and Flagship Pioneering. JBA has received research funding from Incyte. CB received honoraria as speaker (Astrazeneca, Incyte) and consultant (Incyte, Servier, Boehringer Ingelheim, Astrazeneca), received research funds (Avacta, Medannex, Servier) and her spouse is an employee of Astrazeneca. JML is receiving research support from Eisai, Bayer HealthCare Pharmaceuticals, Ipsen, and consulting fees from Eisai, Merck, Bristol-Myers Squibb, Eli Lilly, Roche, Genentech, Ipsen, Glycotest, AstraZeneca, Bayer HealthCare Pharmaceuticals, Omega Therapeutics, Mina Alpha, Boston Scientific, Exelixis, Bluejay and Captor Therapeutics. RM has received consulting and lecture fees from Servier and Roche and travel and education funding from MSD, Eli Lilly, Bayer, Roche, Astrazeneca. AL reports receiving consulting fees from Intercept Pharma, Alfa Sigma, Takeda, and Albireo Pharma, and speakers' fees from Gilead, Abbvie, MSD, Intercept Pharma, AlfaSigma, GSK and Incyte.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Patients with histologically proven advanced iCCA treated with first-line chemotherapy were retrospectively identified from the Modena Cancer Centre Biliary Tract Cancer Database, after review from the appropriate health research authorities (BILONG study protocol 465/18—reviewed by the Area Vasta Emilia Nord Ethics committee). Patients were deemed eligible if they presented with de novo advanced unresectable iCCA (ie, locally advanced or metastatic) and tissue from diagnostic liver biopsies were available. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Transcriptome data have been deposited in Gene Expression Omnibus (GEO) under this manuscript.

ORCID iDs
Guido Carpino http://orcid.org/0000-0001-8570-2519
Daniela Sia http://orcid.org/0000-0002-9000-611X
Diletta Overi http://orcid.org/0000-0003-3561-8903
Stefano Cascinu http://orcid.org/0000-0001-6289-7202
Dan Hogdall http://orcid.org/0000-0003-3832-2737
Enric Domingo http://orcid.org/0000-0003-4390-8767
Luke Boulter http://orcid.org/0000-0002-7954-6705
Ana Lleo http://orcid.org/0000-0002-0561-7902
Jesper B Andersen http://orcid.org/0000-0003-1760-5244
Chiara Braconi http://orcid.org/0000-0003-4835-1259

REFERENCES

1 Bertuccio P, Malvezzi M, Carioli G, *et al*. "Reply to: "global trends in mortality from Intrahepatic and extrahepatic cholangiocarcinoma"" *J Hepatol* 2019;71:1262–3.
2 Vithayathil M, Khan SA. Current epidemiology of cholangiocarcinoma in Western countries. *J Hepatol* 2022;77:1690–8.
3 Izquierdo-Sanchez L, Lamarca A, La Casta A, *et al*. Cholangiocarcinoma landscape in Europe: diagnostic, prognostic and therapeutic insights from the ENSCCA registry. *J Hepatol* 2022;76:1109–21.
4 Valle J, Wasan H, Palmer DH, *et al*. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Engl J Med* 2010;362:1273–81.
5 Oh D-Y, Ruth He A, Qin S, *et al*. Durvalumab plus gemcitabine and cisplatin in advanced biliary tract cancer. *NEJM Evidence* 2022;1.
6 Kelley RK, Ueno M, Yoo C, *et al*. Pembrolizumab in combination with gemcitabine and cisplatin compared with gemcitabine and cisplatin alone for patients with advanced biliary tract cancer (KEYNOTE-966): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* 2023;401:1853–65.
7 Lamarca A, Palmer DH, Wasan HS, *et al*. Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. *Lancet Oncol* 2021;22:690–701.
8 Primrose JN, Fox RP, Palmer DH, *et al*. Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study. *Lancet Oncol* 2019;20:663–73.
9 Bridgewater J, Lopes A, Palmer D, *et al*. Quality of life, long-term survivors and long-term outcome from the ABC-02 study. *Br J Cancer* 2016;114:965–71.
10 Moriwaki T, Yamamoto Y, Gosho M, *et al*. Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: a meta-analysis of randomised trials of first-line chemotherapy. *Br J Cancer* 2016;114:881–8.
11 Abou-Alfa GK, Sahai V, Hollebecque A, *et al*. Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. *Lancet Oncol* 2020;21:671–84.
12 Zhu AX, Macarulla T, Javle MM, *et al*. Final overall survival efficacy results of Ivosidenib for patients with advanced cholangiocarcinoma with IDH1 Mutation . *JAMA Oncol* 2021;7:1669.
13 Bekaii-Saab TS, Valle JW, Van Cutsem E, *et al*. FIGHT-302: first-line pemigatinib vs gemcitabine plus cisplatin for advanced cholangiocarcinoma with FGFR2 Rearrangements. *Future Oncol* 2020;16:2385–99.
14 Verdaguer H, Saurí T, Acosta DA, *et al*. ESMO scale for clinical actionability of molecular targets driving targeted treatment in patients with Cholangiocarcinoma. *Clin Cancer Res* 2022;28:1662–71.

15 Israel MA, Danziger N, McGregor KA, *et al*. Comparative genomic analysis of Intrahepatic cholangiocarcinoma: biopsy type, ancestry, and testing patterns. *Oncologist* 2021;26:787–96.
16 Wheeler DA, Takebe N, Hinoue T, *et al*. Molecular features of cancers exhibiting exceptional responses to treatment. *Cancer Cell* 2021;39:38–53.
17 Pleasance E, Bohm A, Williamson LM, *et al*. Whole-genome and transcriptome analysis enhances precision cancer treatment options. *Ann Oncol* 2022;33:939–49.
18 Caputo F, Gelsomino F, Spallanzani A, *et al*. Multicentre match-paired analysis of advanced biliary cancer long-term survivors: the BILONG study. *Clin Res Hepatol Gastroenterol* 2022;46:101955.
19 Job S, Rapoud D, Dos Santos A, *et al*. Identification of four immune subtypes characterized by distinct composition and functions of tumor microenvironment in intrahepatic cholangiocarcinoma. *Hepatology* 2020;72:965–81.
20 Martin-Serrano MA, Kepecs B, Torres-Martin M, *et al*. Novel Microenvironment-based classification of intrahepatic cholangiocarcinoma with therapeutic implications. *Gut* 2023;72:736–48.
21 O'Rourke CJ, Matter MS, Nepal C, *et al*. Identification of a pan-gamma-secretase inhibitor response signature for notch-driven cholangiocarcinoma. *Hepatology* 2020;71:196–213.
22 Carotenuto P, Fassan M, Pandolfo R, *et al*. Wnt signalling modulates transcribed-ultraconserved regions in hepatobiliary cancers. *Gut* 2017;66:1268–77.
23 Affo S, Nair A, Brundu F, *et al*. Promotion of cholangiocarcinoma growth by diverse cancer-associated fibroblast subpopulations. *Cancer Cell* 2021;39.
24 Lamarca A, Kapacee Z, Breeze M, *et al*. Molecular profiling in daily clinical practice: practicalities in advanced cholangiocarcinoma and other biliary tract cancers. *J Clin Med* 2020;9:2854.
25 Avan A, Caretti V, Funel N, *et al*. Crizotinib inhibits metabolic inactivation of gemcitabine in C-met-driven pancreatic carcinoma. *Cancer Res* 2013;73:6745–56.
26 Li W, Zhang X, Chen Y, *et al*. G-CSF is a key modulator of MDSC and could be a potential therapeutic target in colitis-associated colorectal cancers. *Protein Cell* 2016;7:130–40.
27 Halbrook CJ, Pontious C, Kovalenko I, *et al*. Macrophage-released pyrimidines inhibit gemcitabine therapy in pancreatic cancer. *Cell Metab* 2019;29:1390–9.
28 Alvisi G, Termanini A, Soldani C, *et al*. Multimodal single-cell profiling of intrahepatic cholangiocarcinoma defines hyperactivated Tregs as a potential therapeutic target. *J Hepatol* 2022;77:1359–72.
29 Li T, Tan Y-T, Chen Y-X, *et al*. Methionine deficiency facilitates antitumour immunity by altering M(6)A methylation of immune checkpoint transcripts. *Gut* 2023;72:501–11.
30 Boucher Y, Posada JM, Subudhi S, *et al*. Addition of losartan to FOLFIRINOX and chemoradiation reduces immunosuppression-associated genes, Tregs, and Foxp3+ cancer cells in locally advanced pancreatic cancer. *Clinical Cancer Research* 2023;29:1605–19.
31 Høgdall D, O'Rourke CJ, Dehlendorff C, *et al*. Serum Il6 as a prognostic biomarker and Il6R as a therapeutic target in biliary tract cancers. *Clin Cancer Res* 2020;26:5655–67.
32 Valle JW, Kelley RK, Nervi B, *et al*. Biliary tract cancer. *Lancet* 2021;397:428–44.
33 Sullivan KM, Jiang X, Guha P, *et al*. Blockade of interleukin 10 potentiates antitumour immune function in human colorectal cancer liver metastases. *Gut* 2023;72:325–37.
34 Andersen JB, Spee B, Blechacz BR, *et al*. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. *Gastroenterology* 2012;142:1021–31.
35 Nakamura H, Arai Y, Totoki Y, *et al*. Genomic spectra of biliary tract cancer. *Nat Genet* 2015;47:1003–10.
36 Jusakul A, Cutcutache I, Yong CH, *et al*. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov* 2017;7:1116–35.
37 Dong L, Lu D, Chen R, *et al*. Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma. *Cancer Cell* 2022;40:70–87.
38 Sia D, Hoshida Y, Villanueva A, *et al*. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* 2013;144:829–40.
39 Montal R, Sia D, Montironi C, *et al*. Molecular classification and therapeutic targets in extrahepatic cholangiocarcinoma. *J Hepatol* 2020;73:315–27.
40 Yoon JG, Kim MH, Jang M, *et al*. Molecular characterization of biliary tract cancer predicts chemotherapy and programmed death 1/programmed death-ligand 1 blockade responses. *Hepatology* 2021;74:1914–31.
41 Robinson DR, Wu Y-M, Lonigro RJ, *et al*. Integrative clinical genomics of metastatic cancer. *Nature* 2017;548:297–303.
42 Pleasance E, Titmuss E, Williamson L, *et al*. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. *Nat Cancer* 2020;1:452–68.
43 Bridgewater JA, Pugh SA, Maishman T, *et al*. Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis (new EPOC): long-term results of a multicentre, randomised, controlled, phase 3 trial. *Lancet Oncol* 2020;21:398–411.
44 Zeng Z, Zhao Y, Chen Q, *et al*. Hypoxic exosomal HIF-1Alpha-stabilizing CircZNF91 promotes chemoresistance of normoxic pancreatic cancer cells via enhancing Glycolysis. *Oncogene* 2021;40:5505–17.
45 Lin J, Xia L, Oyang L, *et al*. The POU2F1-ALDOA axis promotes the proliferation and chemoresistance of colon cancer cells by enhancing glycolysis and the pentose phosphate pathway activity. *Oncogene* 2022;41:1024–39.
46 Olson OC, Kim H, Quail DF, *et al*. Tumor-associated macrophages suppress the cytotoxic activity of antimitotic agents. *Cell Rep* 2017;19:101–13.
47 Yu J, Green MD, Li S, *et al*. Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. *Nat Med* 2021;27:152–64.
48 Kumagai S, Koyama S, Itahashi K, *et al*. Lactic acid promotes PD-1 expression in regulatory T cells in highly glycolytic tumor Microenvironments. *Cancer Cell* 2022;40:201–18.
49 Reinfeld BI, Madden MZ, Wolf MM, *et al*. Cell-programmed nutrient partitioning in the tumour microenvironment. *Nature* 2021;593:282–8.

Molecular portraits of patients with intrahepatic cholangiocarcinoma who diverge as rapid progressors or long survivors on chemotherapy

Colm J O'Rourke[1*], Massimiliano Salati[2,3*], Colin Rae C[4], Guido Carpino[5], Holly Leslie[4], Antonio Pea[4], Maria G Prete[4], LR Bonetti[6], F Amato[4], R Montal[7], R Upstill-Goddard[4], Colin Nixon[8], Paula Sanchon-Sanchez[4], Paolo Kunderfranco[9], Daniela Sia[10], Eugenio Gaudio[5], Diletta Overi[5], Stefano Cascinu[11], Dan Hogdall[1,12], Sian Pugh[13], Enric Domingo[14], John N Primrose[15], John Bridgewater[16], A Spallanzani[2], F Gelsomino[2], JM Llovet[17,18,19], Diego F. Calvisi[20], Luke Boulter[21,22], F Caputo[2], Ana Lleo[23,24], Nigel B Jamieson[4,22], G Luppi[2], M Dominici[2], Jesper B Andersen[1*#], Chiara Braconi[4, 22, 25*#]

[1]Biotech Research and Innovation Centre (BRIC), Department of Health and Medical Sciences, University of Copenhagen, Copenhagen N, Denmark.
[2]Division of Oncology, Department of Oncology and Hematology, University Hospital of Modena, Modena, Italy.
[3]PhD Program, Clinical and Experimental Medicine, University of Modena and Reggio Emilia, Modena, Italy.
[4]School of Cancer Sciences, University of Glasgow, Scotland, United Kingdom.
[5]Department of Anatomical, Histological, Forensic Medicine and Orthopaedic Sciences, Sapienza University of Rome, Rome, Italy.
[6]Division of Pathology, University of Modena and Reggio Emilia, Emilia-Romagna, Rome, Italy.
[7]Cancer Biomarkers Research Group, Department of Medical Oncology, Hospital Universitari Arnau de Vilanova-IRBLleida, Lleida, Catalonia, Spain.
[8]CRUK Beatson Cancer Research Institute, Glasgow, UK.
[9]Bioinformatics Unit, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.
[10]Liver Cancer Program, Divisions of Liver Diseases, Pathology Department and RM Transplant Institute, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA.
[11]Department of Medical Oncology, IRCCS San Raffaele Hospital, Milan, Italy.
[12]Department of Oncology, Herlev and Gentofte Hospital, Herlev, Copenhagen University Hospital, Copenhagen, Denmark.
[13]Addenbrooke's Hospital, Cambridge, United Kingdom.
[14]Department of Oncology, University of Oxford, Oxford, England, United Kingdom.
[15]Department of Surgery, University of Southampton, Southampton, United Kingdom.
[16]UCL Cancer Institute, London, United Kingdom.
[17]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clínic, University of Barcelona, Barcelona, Catalonia, Spain.
[18]Liver Cancer Program, Divisions of Liver Diseases, Pathology Department and RM Transplant Institute, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA.
[19]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.
[20]Institute of Pathology, University of Regensburg, 93053 Regensburg, Germany.

O'Rourke CJ, *et al. Gut* 2023;0:1–13. doi: 10.1136/gutjnl-2023-330748

[21]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.
[22]Cancer Research UK Scotland Centre, Glasgow-Edinburgh UK.
[23]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy.
[24]Internal Medicine and Hepatology Unit, Department of Gastroenterology, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.
[25]Beatson West of Scotland Cancer Centre, Glasgow, UK.
*Equal contribution, [#]corresponding authors

SUPPLEMENTARY INFORMATION

Total number of items: 12 supplementary figures, 11 supplementary tables

SUPPLEMENTARY MATERIALS & METHODS

RPLS COHORT

Patient selection

Patients with histologically proven advanced iCCA treated with first-line chemotherapy were retrospectively identified from the Modena Cancer Centre Biliary Tract Cancer Database, after review from the appropriate health research authorities (BILONG study protocol 465/18 - reviewed by the Area Vasta Emilia Nord Ethics committee). Patients were deemed eligible if they presented with *de novo* advanced unresectable iCCA (i.e. locally advanced or metastatic) and tissue from diagnostic liver biopsies were available. Patients with a diagnosis of mixed iCCA/HCC were excluded. Neither prior surgery nor liver-directed treatment were allowed. Clinical and laboratory data were retrieved retrospectively through electronic medical records review. The following baseline variables were collected and analysed before the commencement of first-line chemotherapy: age, sex, Eastern Cooperative Oncology Group (ECOG) performance status (PS), primary tumour site, disease status, location of metastatic sites. Overall survival (OS) was defined as the time from the first dose of first line chemotherapy to death. To meet the definition of rapid progressor (RP), patients had to have an OS ≤ 6 months which is below half the median survival time reported by the ABC-02 trial [1]. Long survivors (LS) were defined those patients who had OS ≥23 months since starting of chemotherapy, more than double the median survival time reported by the ABC-02 trial. The two patient subgroups were matched for major clinical features. None of these patients had

background liver cirrhosis. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

Human tissues

All patients underwent ultrasound-guided liver biopsy before the commencement of systemic treatment. Formalin Fixed Paraffin Embedded (FFPE) tissue slides were retrieved from liver biopsy. One slide was processed for targeted sequencing-based RNA expression analysis via TempO-Seq (Bioclavis, Glasgow, UK). One slide was used for macroscopic dissection of specific ROIs identified through a pathology review; tissue from each (region-of-interest) ROI was then subjected to targeted sequencing-based RNA expression analysis via TempO-Seq (Bioclavis, Glasgow, UK). A third consecutive slide was used for Digital Spatial Transcriptomic (see below).

Digital pathology

Hematoxylin-eosin slides were digitized using the APERIO platform (Leica Biosystems) at 20X of magnification. Each slide was analysed using the open-source software platform, QuPath (version 0.2.3) [2]. A region of interest (ROI) was annotated for each slide and the amount of tumoural tissue quantified. First, we characterised the amount of epithelial and stromal tissue components in each ROI by generating a random tree forest pixel classifier. Second, following cell detection on each ROI, a different random tree forest cell classifier was generated for each haematoxylin-eosin slide using cell features to classify cells into tumour, immune and stromal cells. To help the algorithm to perform an accurate classification, smoothed features at 25 radii were added and multiple rounds of cell classification review were performed.

Whole-transcriptome profiling by Tempo-seq

Whole transcriptome gene expression analysis was performed using the TempO-Seq Human Whole Transcriptome v2.0 panel (BioClavis, Glasgow, UK). Internal quality control was performed according to the following criteria: 1) number of mapped reads in positive RNA controls > 6 million [our study: 6,703,241]; 2) signal:noise ratio as the ratio between the total number of reads in the positive control and the total number of reads in the negative controls

>20 [our project 205;1]; 3) the percentage of mapped reads in positive controls > 80% [our project 97%]; 4) average reads/probe >250 [our project 471]. Process controls were run in replicate on each assay plate of samples to ensure quality metrics pass on a plate-wise level. High reproducibility of positive controls, and low signal in the negative control was observed in our project.

Sequencing libraries for targeted panels were generated; briefly in TempO-Seq, each Detector Oligo consisted of a sequence complementary to an mRNA target plus a universal primer binding site. They annealed in immediate juxtaposition to each other on the targeted RNA template such that they can be ligated together. Ligated detector oligos were PCR-amplified using a primer set (single-plex PCR reaction, with a single primer pair for each sample) that introduced both the adaptors required for sequencing and a sample-specific barcode. The barcode sequences flanked the target sequence and were inserted appropriately into the standard Illumina adaptors to permit standard dual-index sequencing of the barcodes and deconvolution of sample-specific reads from the sequencing data using the standard Illumina software. All the PCR-amplified and barcoded samples were pooled into a single library for sequencing. Sequencing reads were demultiplexed using the standard sequencing instrument software for each sample using the barcodes to give a FASTQ file for each.

TempO-Seq sequence files were analysed using the Tempo-SeqR software package. Each FASTQ file was aligned using the STAR algorithm to a pseudotranscriptome corresponding to the gene panel used in the assay. Data were normalized using DESeq2 [3]. Transcriptome data can be made available upon request following IRB approval.

GeoMx digital spatial profiling

NanoString GeoMx digital spatial profiling

To further characterise variation in transcriptomic expression between tumour-infiltrating myeloid cells in LS and RP patients, formalin-fixed paraffin-embedded (FFPE) sections of 12 cholangiocarcinoma biopsies were selected for analysis on the NanoString GeoMx Digital Spatial Profiler (DSP). This platform enables the characterisation of user-selected topographic Regions of Interest (ROI) from immunofluorescently (IF) stained FFPE tissue. The GeoMx instrument achieves RNA profiling in situ hybridization by employing DNA oligonucleotide

probes designed to bind mRNA targets. From 5′ to 3′, they comprise a 35- to 50-nucleotide target complementary sequence, an ultraviolet (UV) photocleavable linker and a 66-nucleotide indexing oligonucleotide sequence containing a unique molecular identifier (UMI), RNA ID sequence and primer binding sites. Up to 10 RNA detection probes were designed per target mRNA. In summary, the instrument employs UV light to cleave the UV-sensitive probes leading to release of the hybridized barcodes.

Slide preparation including hybridization of tissue with UV-photocleavable probes

The DSP procedure has previously been described in detail by Merritt et al [4] with our own protocol described in Fisher et al [5]. 5-µm FFPE sections of the 12 cholangiocarcinoma biopsies were mounted on positively charged Superfrost glass slides (ThermoFisher Scientific) and baked for 1hr at 60°C. The tissue was dewaxed, hydrated and treated with 1 µg/ml Proteinase K (ThermoFisher Scientific, AM2546) for 15 minutes before undergoing heat-induced epitope retrieval (HIER) on a Leica BOND Autostainer (pH 9.0, ER2 at 100°C) for 20 minutes. The slides were then stored until required in 1X PBS (PBS: Invitrogen, AM9625).

In-situ hybridization of RNA-directed DNA oligo probes (Immune Pathways Panel included 96 genes with the addition of CD58, ELK4, CD80, CD163, FOXO3, NanoString) was performed as per manufacturer's protocol. HybriSlip™ covers were applied prior to overnight incubation at 37°C for at least 16 hours (ThermoFisher). The following day, slides were washed twice with a 1:1 ratio of 100% deionized formamide (Ambion) and 4X SSC (Sigma Aldrich) at 37°C for 25 minutes.

The GeoMx DSP is capable of capturing four channels (FITC/525nm, Cy3/568nm, Texas Red/615nm and Cy5/666nm) for the detection of up to four customizable IF morphology markers for each tissue [4]. One channel (FITC/525nm) is reserved for the nuclear stain (SYTO13). The slides were blocked with Buffer W (Nanostring) for 30 minutes at RT before incubation. Immunofluorescence staining was performed using primary conjugated antibodies (PanCK (NanoString), CD68 (clone KP1, 1:200, Santa Cruz) and Ki67 (clone D2H10, 1:100, Cell Signaling Technologies)) and nucleic acid dye (SYTO 13 (NanoString)) as per the manufacturer protocol. Slides were then stored at 4°C in SSC before being loaded on the GeoMx DSP instrument for region of interest (ROI) selection and collection.

Region selection and collection

The whole slides were imaged at 20x magnification using the GeoMx DSP, with the integrated software suite used to select ROIs for downstream analysis based on the 4-plex immunofluorescence staining of SYTO 13, PanCK, CD68 and Ki67. Polygonal ROIs were selected primarily according to the high density of CD68+ cells. 30 ROIs were selected in total from across the 12 slides. The CD68 staining was then used to employ segmentation within each ROI to create an area of interest (AOI) from which DNA oligo probes would be cleaved and cell-type specific transcriptomic profiles obtained. After AOIs were selected, the GeoMx DSP employs an automatically controlled UV laser (385nm) to illuminate each AOI in turn, specifically cleaving barcodes within the CD68+ immunofluorescent (IF) segments but not in surrounding tissue segments (CD68-). A microcapillary collection system collected the liberated barcodes from each AOI and plated them into individual wells on a 96-well microtiter plate. This process was repeated in turn for each AOI before downstream processing using the NanoString MAX/FLEX nCounter system. The oligonucleotides were dried overnight and subsequently resuspended using 7 µl of DEPC-treated water.

nCounter hybridization assay for photocleaved oligo counting

The nCounter readout of GeoMx DSP-collected probes was performed according to manufacturer's protocol (NanoString, MAN-10089-08). In brief, samples were resuspended in dH20 prior overnight incubation (16–24 hours) with hybridization codes (Hyb Codes) at 65°C and heated lid (70°C). These Hyb Codes include reporter and capture probes to enable formation of a tripartite hybridization complex with the DNA oligo probes in the panel. Samples were then pooled by column into a 12-well strip tube before processing on NanoString's MAX/FLEX system, using the high sensitivity protocol (NanoString, MAN-10089-08).

Data processing and analysis

Data acquisition was performed by using the NanoString's Digital Analyzer (field of view, 555) and Digital Count Conversion (DDC) files were re-uploaded onto the GeoMx DSP for mapping of transcriptomic data to the spatial origin, where they underwent quality control, filtering normalization and background correction.

Bioinformatic analyses

RPLS signature identification

RPLS signature genes were identified as genes differentially expressed between RP and LS patients with ≥ 2-fold expression difference and unadjusted $p<0.05$ (Wilcoxon rank-sum test). As RP-high and LS-high signature genes significantly anti-correlated, a single RPLS score was assigned to each tissue sample, defined as: $([\log_2(\Sigma\text{RP-high genes}) - \log_2(\Sigma\text{LS-high genes})]_{\text{z-score}})$. To compare the predictive performance of the RPLS score against systemic features, it was included into multivariable Cox proportional hazards models with pre-treatment systemic features that differed between patient subgroups (platelets, ALP) or an optimized systemic signature (defined by AIC backwards elimination using all haematological and systemic features). The final formula derived for this optimized signature was: -211.622 + 41.2*Bili + 1.208*ALP.

Pathway analysis

Pathway over-representation analysis of RPLS signature genes was performed using ENRICHR against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [6] (including censoring of the non-relevant categories, “infectious disease” and “substance dependence”). Significantly over-represented pathways ($p<0.05$) were visualized using CytoScape (v3.9) [7]. Specifically, the EnrichmentMap application[8] was used to construct network maps in which each node represents a significant pathway and each line denotes shared genes between pathways (Jaccard coefficient above 0.375). Inter-related pathways were identified and biologically classified by Markov Cluster Algorithm (MCL) using the AutoAnnotate application (v1.3.3) [9].

Pathway expression analysis was performed using single sample gene set enrichment analysis (ssGSEA), as implemented in the ‘GSVA’ R package [10]. Prior to ssGSEA, transcriptome data were centered and scaled. KEGG [6] and Hallmarks [11] terms were downloaded from Molecular Signatures Database (MSigDB) v7.5.1 [12, 13] and used as input gene lists for ssGSEA. Only human- and cancer-relevant gene list categories were considered.

Transcription factor activities were predicted from transcriptome data using DoRothEA [14]. A gene was predicted to be under regulation of a specific transcription factor if it is a known transcription factor [14] and if its expression positively correlates with activity of the candidate transcription factor regulator.

Gemcitabine resistance genes and signatures

The expression of four genes involved in gemcitabine transport and metabolism that have been previously implicated in gemcitabine sensitivity [15, 16, 17] were investigated in our study: human equilibrative nucleoside transporter-1 (hENT1, encoded by *SLC29A1*); deoxycytidine kinase (*dCK*); ribonucleotide reductase catalytic subunit M1 (*RRM1*); ribonucleotide reductase catalytic subunit M1 (*RRM1*).

Three gene expression signatures previously reported to predict gemcitabine sensitivity were also evaluated in our study:

Mourragui et al. signature: This signature was identified based on *in vitro* sensitivity of 1,000 immortalized cancer cell lines to gemcitabine [18]. It consists of a weighted linear sum of expression of 1774 genes which was applied to each bulk tissue sample in the RPLS cohort.

Nicolle et al. signature: Also known as GemPred, this signature was identified based on *in vitro* sensitivity of 38 primary pancreatic adenocarcinoma (PDAC) cell lines to gemcitabine [19]. Whole-transcriptome profiles of RPLS samples were submitted to the GemPred web application (http://cit-apps.ligue-cancer.net/pancreatic_cancer/GemPred) in order to return a score per sample. Unreferenced values were used for further analysis.

Tiriac et al. signature: This signature was identified based on in vitro sensitivity of 77 PDAC organoids to gemcitabine [20]. It consists of 225 genes whose expression correlates (95 positively, 130 negatively) with gemcitabine sensitivity. These genes were combined into a weighted linear formula where each gene was assigned a coefficient proportional to its correlation coefficient, enabling assignment of an overall signature score to each of the bulk tissue samples in the RPLS cohort.

Digital cytometry

Cell type-specific deconvolution of the RPLS signature was performed in our RPLS cohort using CIBERSORTx [21]. Briefly, two public single cell RNA-sequencing (scRNA-seq) datasets from iCCA patients were downloaded from GEO (GSE125449 [22], GSE151530 [23]). These data were processed independently using the 'Seurat' package (v3) [24], including exclusion of outlier cells (mitochondrial content >0.05, gene counts <200 or 2500) and log normalization (scale factor of 10000). Six types of cells were commonly annotated by the original two scRNA-seq studies: B cells, cancer-associated fibroblasts (CAFs), T cells, tumour cells, tumour-associated myeloid cells (reannotated from cells originally reported as tumour-associated

macrophages), and tumour-associated endothelial cells (TECs). Only genes commonly detected in both datasets were used for gene expression classifier construction. Independently, a signature matrix was constructed for each scRNA-seq dataset in CIBERSORTx ('scRNA-Seq' as input data type with default settings). Each signature matrix was then applied independently to impute cell expression of the RPLS signature genes in our biopsy tissues (whole biopsies, tumour regions, tumour-associated stroma regions). This involved running the 'Impute Cell Expression' module of CIBERSORTx in 'Group Mode' with 'S-mode' batch correction and quantile normalization set as disabled. Each gene in the RPLS signature was only considered to originate from a specific cell type if this association was reproducible using both independent signature matrices (cell type-specific gene expression >1).

Immunogenomic and microenvironment analyses

Immune cell composition was predicted by cellular deconvolution using the xCell tool [25], implemented in the 'immunedeconv' package [26] with default settings. Anti-tumour immune activity was inferred using the cytolytic score, defined as the geometric mean of expression of granzyme A (*GZMA*) and perforin (*PRF1*) [27]. Cytokine activity was estimated using the CytoSig tool with default settings [28], implemented in Python on transcriptome data that was centralised and log2-transformed (including a pseudo-count of 1). Cancer testis antigen (CTA) expression was investigated using ssGSEA of CTA lists reported by CTdatabase (http://www.cta.lncc.br/) [29]. A gene expression-based microsatellite instability signature was calculated for RPLS samples using the Tumour Immune Dysfunction and Exclusion (TIDE) webtool [30]. A gene expression-based signature predictive of BRCA gene functionality and comprised of a weighted linear formula of 60 genes was used to assign a "BRCAness" score to each RPLS sample [31]. DNA repair signatures were obtained from Reactome [32]. Signatures associated with cancer-associated fibroblast subtypes detected in iCCA were investigated using ssGSEA with the gene lists reported in the original study [33].

IMMORTALIZED iCCA CELL LINES

Integration of cell line and RPLS tissue transcriptome data

Transcriptome data (RNA-seq) for 25 iCCA cell lines was downloaded from DepMap (22Q2) [34] quantified in $\log_2$(TPM+1) units. In total, 14,728 genes were commonly detected in the cell line dataset and our RPLS tumour core (TC) dataset. Datasets were combined together by

O'Rourke CJ, *et al. Gut* 2023;0:1–13. doi: 10.1136/gutjnl-2023-330748

performing quantile normalization ('PreProcess' package) followed by Combat batch correction ('sva' package) [35]. Removal of batch effects was confirmed by principal component analysis.

Classification of cell lines as LS- or RP-like

To develop a scoring method for RP- and LS-likeness, generalized linear models were run to identify genes individually capable of differentiating RP TCs from LS TCs in the adjusted dataset. In total, 441 significant genes ($p<0.05$) were identified. To generate an optimized classifier from these genes, least absolute shrinkage and selection operator (LASSO) regression was performed, resulting in the following classifier: -0.473260081235793 + (-0.20222060165566*CERS2) + (-0.0619806057536049*CHST9) + (-0.158248138046384*DCDC2) + (0.343556411268419*IGFBP4) + (-0.157768591678966*KPNA6) + (0.153671854832587*PDE12) + (0.041350560922029*PTPRU) + (-0.00927545715353832*SLC5A9) + (0.029840846853945*SPTLC3) + (0.105321982050334*STEAP2). This classifier was then applied to the adjusted cell line data to identify LS-like (< median; n=12) and RP-like (> median; n=13) cell lines.

Pharmacologic, genomic and transcriptomic comparison of LS- and RP-like cell lines

Gemcitabine sensitivity was compared between LS- (n=4) and RP-like (n=4) cell lines using previously reported high-throughput drug screening data from Saha and colleagues [36]. Differential expression of pathways and processes was compared between LS- and RP-like cell lines using ssGSEA, as earlier described (section 1.6.2). Hotspot mutation data were available for all 25 iCCA cells through DepMap (22Q2) [34]. Association of recurrent hotspot mutations (*IDH1*, *KCNA1*, *KIF4B*, *KRAS*, *PTEN*, *TBX5*, *TP53*) with LS- or RP-like cell lines was investigated by fisher's exact test for count data.

Identification of differential genetic dependencies

Genome-wide CRISPR inactivation screen data were available for 24 iCCA cell lines through DepMap (22Q2) [37], with gene effect scores inferenced by Chronos [38]. Pan-essential genes (n=2040) as reported by Dempster and colleagues [39] were filtered out from the dataset. To determine an appropriate cut-off value for a deleterious fitness effect, we calculated the

median Chronos score for each of the 2040 pan-essential genes across the 24 iCCA cell lines. We selected the largest of these median values as our negative fitness threshold (-0.0141220545). Therefore, non-essential genes were identified as enhanced survival dependencies if they met the following criteria: (1) mean Chronos score below the threshold (-0.0141220545) among cell lines in LS-like and/or RP-like subgroups; (2) significant difference in Chronos scores between LS-like and RP-like subgroups (Wilcoxon $p<0.05$). As true survival dependencies likely arise from multiple genes within common networks, we prioritized enhanced survival dependencies annotated within a common pathway (informed by ENRICHR; described in section 1.6.2). Dependency networks were visualized using CytoScape (StringDB score 0.75) and manually annotated based on MSigDB associations. The potential druggability of genes was investigated using The Drug Gene Interaction Database (DGIdb 4.0) [40].

ANALYSIS OF SINGLE CELL RNA-SEQUENCING DATASETS

Single cell RNA-sequencing dataset pre-processing and filtering

In this manuscript, 3 single cell RNA-sequencing (scRNA-seq) datasets were analysed:

GSE125449 [22]: Data were pre-processed as described in section 1.6.4. Cell type annotation was retained as previously reported by the authors, with the exception of tumour-associated macrophages which were renamed as tumour-associated myeloid cells. Exclusively considering samples where tumour cells were retained, the following cell numbers were included for analysis: 739 tumour cells, 354 cancer-associated fibroblasts, 386 tumour-associated myeloid cells, 2138 T cells (corresponding to 9 samples). A tumour-origin RPLS score (predicted by digital cytometry; section 1.6.4) was calculated for each tumour cell using the 'escape' package. Samples were ranked according to tumour-origin RPLS scores and classified as LS- or RP-like based on median expression.

GSE151530 [41]: Data were pre-processed as described in section 1.6.4. Cell type annotation was retained as previously reported by the authors, with the exception of tumour-associated macrophages which were renamed as tumour-associated myeloid cells. Exclusively considering samples where tumour cells were retained, the following cell numbers were included for analysis: 960 tumour cells, 314 cancer-associated fibroblasts, 478 tumour-associated myeloid cells, 1032 T cells (corresponding to 10 samples). A tumour-origin RPLS score (predicted by digital cytometry; section 1.6.4) was calculated for each tumour cell using

the 'escape' package. Samples were ranked according to tumour-origin RPLS scores and classified as LS- or RP-like based on median expression.

GSE171899 [42]: Data were pre-processed and cells were annotated by the original authors, as indicated in their manuscript. This dataset was exclusively comprised of immune (CD45+) cells and only cells originating from 3 patients who underwent adjuvant chemotherapy were considered: 39 cDC1 (BATF3+), 374 cDC2 (CD1C+), 185 lymphoid, 465 macrophages (ID3+), 365 macrophages (MARCO+), 200 macrophages (TREM2+), 535 monocytes (CD14+), 38 monocytes (FCGR3A). A myeloid-origin RPLS score (predicted by digital cytometry; section 1.6.4) was calculated for each immune cell using the 'escape' package.

Biological analysis of LS-like and RP-like iCCA in scRNA-seq datasets

Biological pathway and process expression were calculated in single cells using the 'escape' package (using MSigDB reference signature databases, as described in section 1.6). The following signatures associated with myeloid cell phenotypes and processes were also used: inflammatory and tolerogenic hepatic macrophages [43]; activated dendritic cells (aDC), immature dendritic cells (iDC), plasmocytoid dendritic cells (pDC) [25]; antigen processing and presenting machinery (APM score) [44]; MHC presentation (classical, class I) [32]; immunoinhibitors and immunostimulators [45]. Transcription factor activities and cytokine activities were also calculated for single cells as earlier described (section 1.6). Ligand:receptor (LR) interactions were identified using CellChat with default settings [46]. Only reproducible interactions (identified in GSE125449 and GSE151530) identified in one tumour subgroup but not the other (LS-like or RP-like) were considered as significant. The potential druggability of genes was investigated using The Drug Gene Interaction Database (DGIdb 4.0) [40]. Clinically actionable targets were identified according to Pharos (Tclin level) [47]. T cell subtypes were assigned using the 'ProjecTILs' package [48]. Immune-oncology targets of interest were defined by the The Cancer Research Institute (CRI) iAtlas consortium (https://isb-cgc.shinyapps.io/shiny-iatlas/) [49]. Metabolic flux inference was performed using METAFLUX[50] with default settings, involving calculation of metabolic reaction activity scores (cubic root normalised) per cell followed by median-based collapsing of activity scores per sample.

RESECTED CHOLANGIOCARCINOMA DATASETS

Transcriptome analyses

The following transcriptome datasets were analysed as part of our study:

Andersen cohort: This cohort is comprised of 138 resected iCCA and 37 resected pCCA (with survival data), originating from Australia, Europe (Belgium, France, Germany, Italy), and USA. Transcriptome data (humanRef-8v2 BeadChip and Human HT-12 v3 Expression BeadChip, Illumina) were generated by our group, as previously reported [51, 52]. Data were processed from raw IDAT files using background subtraction (iScan system, Illumina), followed by quantile normalization ('preprocessCore', R package). Data were quantified as normalized signal intensity (arbitrary units, 'a.u.').

Dong cohort: This cohort is comprised of 244 resected iCCA (with survival data), originating from China. Transcriptome data (RNA sequencing) were retrieved from the supplementary information of the original manuscript [53]. Data were quantified using $\log_2$(TPM+1), as reported by the original study authors.

GSE107943 cohort: This cohort is comprised of 30 resected iCCA (with classification as large duct- or small duct-type) from Korea. Transcriptome data (RNA sequencing) were retrieved from GEO (GSE107943) [54, 55]. Data were quantified using RPKM, as reported by the original study authors.

Job cohort: This cohort is comprised of 70 resected iCCA (with survival data), originating from France. Transcriptome data (Human Transcriptome Array 2.0, Affymetrix) were retrieved form ArrayExpress (E-MTAB-6389) [56]. Data were processed from raw CEL files using Robust Multichip Average (RMA) normalization, as implemented in the 'oligo' R package. Data were quantified as normalized signal intensity (arbitrary units, 'a.u.'). Genes with multiple probes were quantified as the median signal intensity across probes.

Jusakul cohort: This cohort is comprised of 81 resected iCCA (with survival data), originating from Asia (Korea, Singapore, Thailand), Brazil and Europe (France, Romania). Transcriptome data (HumanHT-12 V4.0 expression BeadChip) were retrieved from Gene Expression Omnibus (GEO; GSE89747) [57]. Data were extracted from raw IDAT files and underwent quantile normalization ('preprocessCore', R package). Data were quantified as normalized signal intensity (arbitrary units, 'a.u.').

Montal cohort: This cohort is comprised of 182 resected extrahepatic CCA (with survival data), originating from Europe (Spain, Switzerland) and USA. Transcriptome data (Human Genome

U219 Array, Affymetrix) were processed by the original study authors, as reported in the original manuscript [58]. Data were quantified as normalized signal intensity (arbitrary units, 'a.u.').

Nakamura cohort: This cohort is comprised of 111 resected iCCA (with survival data), originating from Japan. Transcriptome data (RNA sequencing) were retrieved from European Genome Phenome Archive (EGA; EGA00001000950) [59] and processed using standard bioinformatic pipelines, as previously reported by us [60]. Data were quantified using $\log_2$(RPKM+1).

Sia cohort: This cohort is comprised of 119 resected iCCA (with survival data), originating from Europe (Italy, Spain) and USA. Transcriptome data (Whole-Genome DASL HT assay, Illumina) were processed by the original study authors, as reported in the original manuscript [61]. Data were quantified as normalized signal intensity (arbitrary units, 'a.u.').

Genomic analyses

Genomic data were available for the Andersen, Dong and Nakamura cohorts, as follows:

Andersen cohort: Mutation data (TruSeq Amplicon Panel, Illumina) were available for 87% (120/138) of this cohort. Data were processed and analyzed as previously reported by our group [60]. *FGFR2* fusion data (quantitative polymerase chain reaction of known breakpoint regions) were also available for 80.4% (111/138) of the cohort [60].

Dong cohort: Mutation data (whole-exome sequencing) and *FGFR2* fusion data (RNA-seq) were available for 97.1% (237/244) and 100% (244/244) of this cohort, as originally reported by the study [53].

Nakamura cohort: Mutation data (whole-exome sequencing) and *FGFR2* fusion data (RNA-seq) were available for 100% (111/111) of this cohort [59], and these data were processed as originally reported by our group [60].

In each of these iCCA cohorts, cancer-relevant hotspot mutations were identified using the Cancer Hotspots database (https://www.cancerhotspots.org/) from Memorial Sloan Kettering Cancer Centre (MSKCC) [62].

ANIMAL MODELS

Changes in RPLS signature expression was evaluated in 14 iCCA-relevant mouse models relative to their study-specific controls. Human homologs were identified for mouse genes

O'Rourke CJ, *et al. Gut* 2023;0:1–13. doi: 10.1136/gutjnl-2023-330748

using the 'HMD_HumanPhenotype' file at Mouse Genome Informatics (MGI; informatics.jax.org) [63].

GSE66717: Transcriptome data (Affymetrix Mouse Gene 1.0 ST Array) were downloaded for *Pten*-/- (n=4), *Pten*-/- *Tgfbr2*-/- (n=4), and control (n=3) mice deposited in GSE66717 [64]. Raw CEL files were subjected to RMA normalization using the 'oligo' R package.

GSE140498: Transcriptome data (Affymetrix Mouse Gene 1.1 ST Array) were downloaded from GSE140498 [65] for the following samples: NEMO single knockout (n=3), NEMO/JNK1Δhepa double knockout (n=3), JNK(1/2)Δhepa double knockout with diethylnitrosamine (DEN) treatment (n=3), JNK(1/2)Δhepa double knockout with bile duct ligation (BDL; n=3), NEMO/JNK(1/2)Δhepa triple knockout (n=3), control mice with DEN treatment (n=3), control mice with bile duct ligation (n=3), and control mice (n=3).

GSE141511: Transcriptome data (RNA-seq) were downloaded for microdissected control (n=3), ductular proliferation (n=3), intraductal papillary neoplasia of the bile duct (IPNB; n=3) and tumour (n=3) tissues from cholangiocyte-directed KrasG12D mice with diet-induced inflammation (3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC)-treated) deposited in GSE141511 [66]. HTseq count files were merged using 'tximportData' and quantified in counts per million (CPM) using 'edgeR' [67].

AKT/YapS127A: Normalized transcriptome data (Affymetrix array) for this sleeping beauty model of iCCA [68] (n=3) and control mice (n=4) was kindly provided by Prof. Diego Calvisi (Universität Regensburg, Germany).

AKT/NICD: Transcriptome data (Mouse Gene Expression BeadChips, Illumina) for this sleeping beauty model of iCCA [69] (n=4) and control mice (n=3) were generated and analysed as part of our previous study [52].

Genetically engineered models of iCCA: Transcriptome data (RNA-seq) from five iCCA genetically engineered mouse models (GEMMs) harbouring cholangiocyte (CK19)-specific genomic alterations were kindly provided by Dr. Luke Boulter (University of Edinburgh, UK) [70]:, *Kras* & *Nf2* (n=3), *Kras* & *shTrp53* (n=3), *Kras* & *Tp53* (n=3), *Kras* & *Nf2* & *Tp53* (n=3), *Kras* & *Tp53* & *Plbx2* (n=3). Mouse gene IDs were batch annotated using informatics.jax.org.

OTHER CANCER DATASETS

Surgical HCC cohorts

TCGA-LIHC: This cohort is comprised of 359 resected HCC (histologically confirmed and survival data available), originating from USA and published as part of the The Cancer Genome Atlas [71]. Data were generated by RNA-seq and downloaded in RSEM normalized units from Broad GDAC Firehose (https://gdac.broadinstitute.org/). Clinical data were downloaded from cBioPortal (https://www.cbioportal.org/) [72].

GSE14520: This cohort is comprised of 178 resected HCC (histologically confirmed and survival data available). Transcriptome data (Affymetrix Human Genome U133A 2.0 Array) were retrieved from Gene Expression Omnibus (GSE14520) [73]. Data were processed from raw CEL files using Robust Multichip Average (RMA) normalization, as implemented in the 'oligo' R package. Data were quantified as normalized signal intensity (arbitrary units, 'a.u.'). Genes with multiple probes were quantified as the median signal intensity across probes.

Advanced basket cohorts

MET500 cohort: This basket cohort was comprised of 490 metastatic biopsies representative of 22 primary cancer groups [74]. Processed, normalized data (RNA-seq) were downloaded from the following weblink: https://met500.path.med.umich.edu/. Samples were stratified by metastatic sites: abdominal (n=32), liver (n=130), lung & respiratory (n=37), lymph node (n=114), other (n=177).

POG570 cohort: This basket cohort was comprised of 438 metastatic biopsies representative of 26 primary cancer groups [75]. Transcriptome data (RNA sequencing) were downloaded from https://www.bcgsc.ca/downloads/POG570/, followed by Ensembl ID mapping using 'biomaRt'. Data were quantified in TPM. Clinical and treatment information were also downloaded from the POG570 downloads repository: abdominal (n=54), liver (n=198), lung & respiratory (n=37), lymph node (n=87), other (n=62).

Metastatic colorectal cancer cohorts

New EPOC trial: Transcriptome data (normalized Affymetrix array, collapsed by median gene expression) for primary colorectal tumours (n=204) and resected liver metastases (n=145) from the phase III New EPOC trial [76] was kindly provided to us by the S:CORT consortium.

STATISTICAL ANALYSES

Statistical analyses were conducted in R version 4.0.3, unless otherwise stated. Heatmaps were generated using 'gplots'. Bar plots, violin plots, and XY plots were generated in 'ggplot2'. Kaplan-Meier survival analysis and Cox proportional hazards modelling were performed using 'survival' and 'survminer' packages. Forest plots were generated using the 'forestplot' package. Simpson's index was computed using the 'vegan' package. Unrooted dendrograms based on Manhattan distance and complete linkage were generated using 'cluster' and 'ape' packages. Data normality was investigated using Shapiro-Wilk test. Quantitative data were compared across multiple groups using Kruskal-Wallis or ANOVA statistics for non-normal and normal data, respectively. Qualitative data was compared between two groups using Pearson's Chi-squared test, except when one cell in a contingency table has an expected frequency below 5 in which case Fisher's exact test was applied. Correlation analysis was performed using Spearman or Pearson statistics for non-normal and normal data, respectively.

SUPPLEMENTAL FIGURES

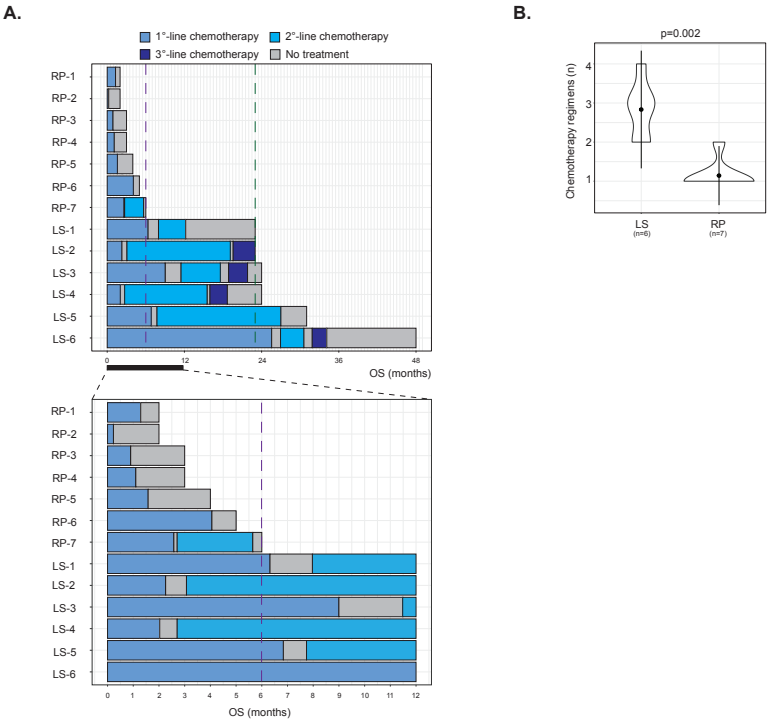


Figure S1: Patient treatment in the RPLS cohort. (A) Swimmer plot of patient management in the RPLS cohort, including a zoomed view on the first 12 months following initiation of

chemotherapy. (B) Comparison of number of chemotherapy regimens (Welch t-test) received by LS (n=6) and RP (n=7) patients.

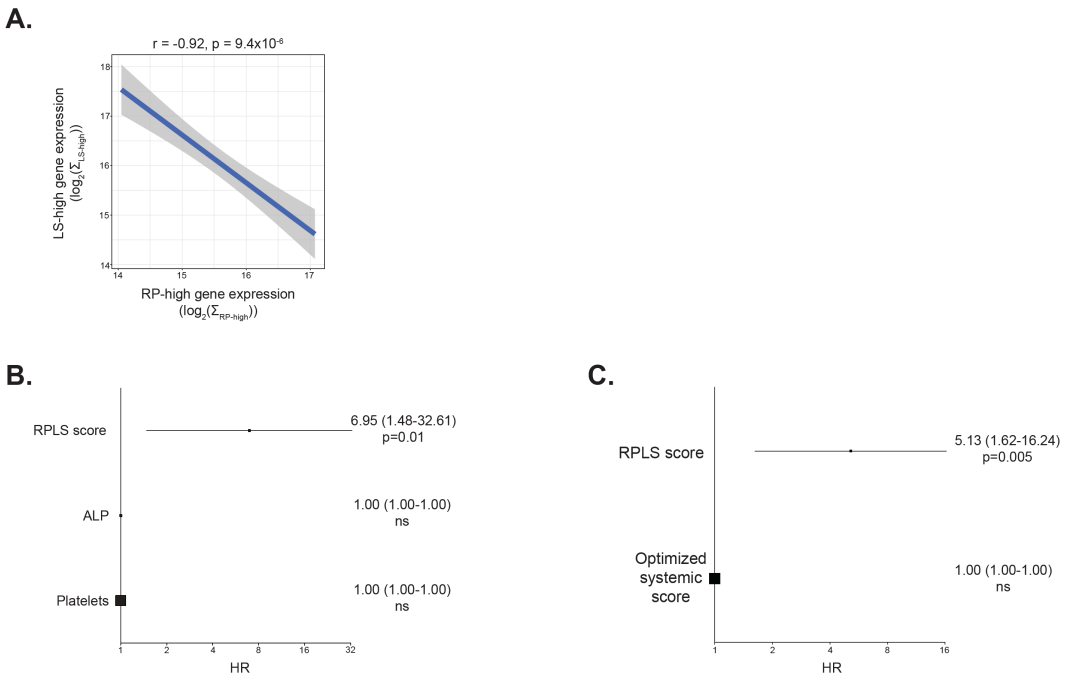


Figure S2: RPLS signature expression and predictive performance in the RPLS cohort. (A) Spearman correlation analysis of RP-high and LS-high gene expression. (B) Forest plot of Cox proportional hazards statistics derived from multivariable analysis of RPLS scores and systemic features differentially expressed between RP and LS patients. HR: hazard ratio; ns: not significant. (C) Forest plot of Cox proportional hazards statistics derived from multivariable analysis of RPLS scores and an optimized systemic signature (defined by AIC backwards elimination using all haematological and systemic features). The final formula derived for this optimized signature was: -211.622 + 41.2*Bili + 1.208*ALP.

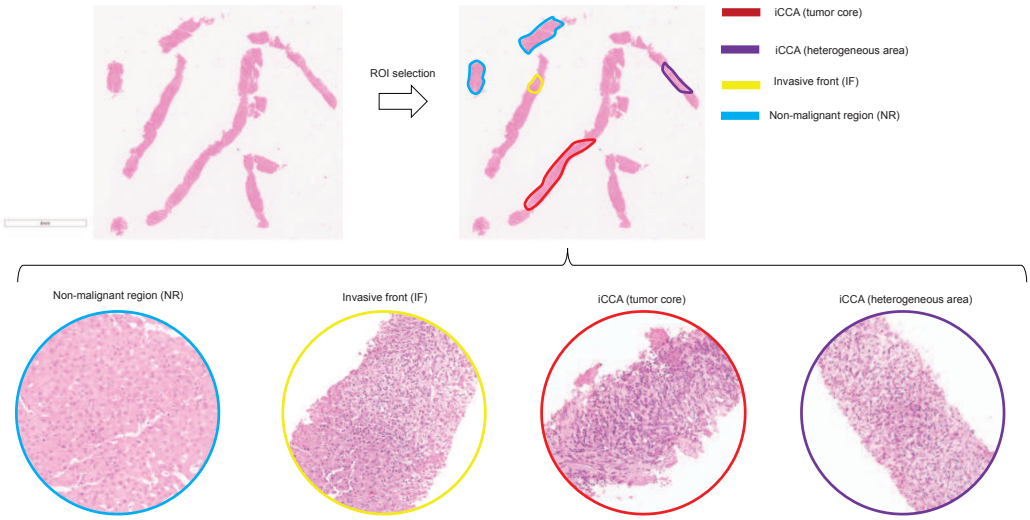


Figure S3: Representative markup of a diagnostic biopsy from the RPLS cohort. Target regions were geospatially macrodissected for whole transcriptome profiling (Tempo-seq).

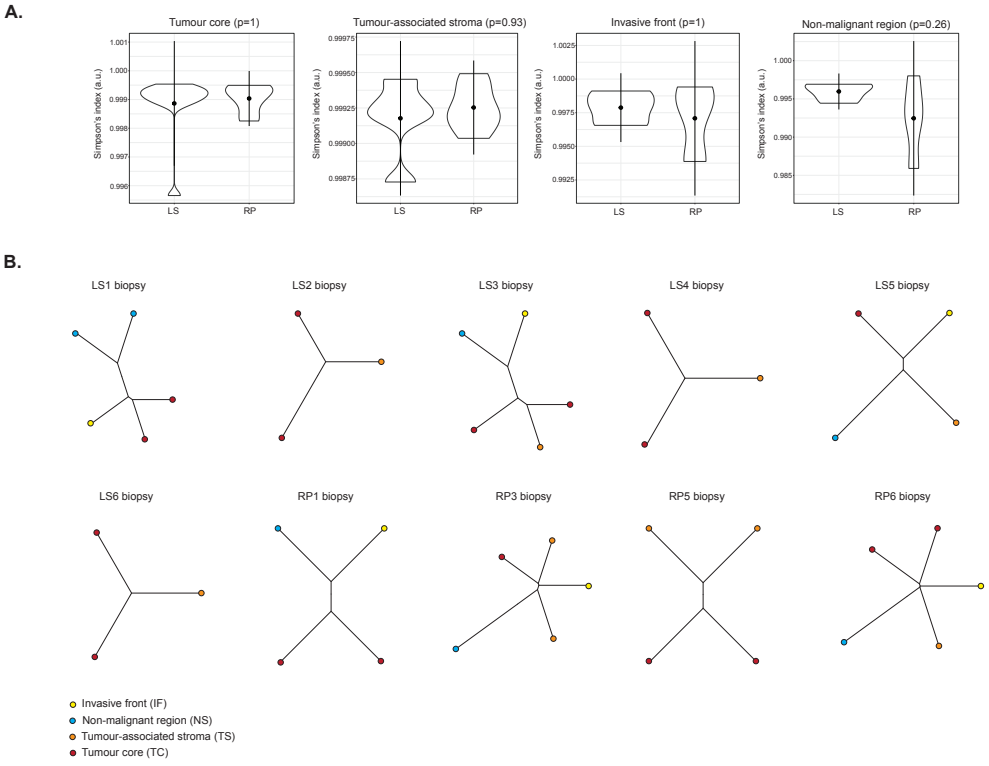


Figure S4: Inter- and intra-sample heterogeneity of macrodissected biopsy regions in the RPLS cohort. (A) Differential intra-sample transcriptomic heterogeneity (Wilcoxon test)

defined by Simpson's index between LS and RP biopsies: tumour cores (TCs; 11 LS, 9 RP), tumour stroma (TSs; 5 LS, 6 RP), invasive fronts (IFs; 3 LS, 3 RP) and non-malignant regions (NRs; 4 LS, 4 RP). (B) Phylotranscriptomic trees (Euclidean distance, Ward linkage) for individual biopsies with 3 or more geospatially macrodissected samples.

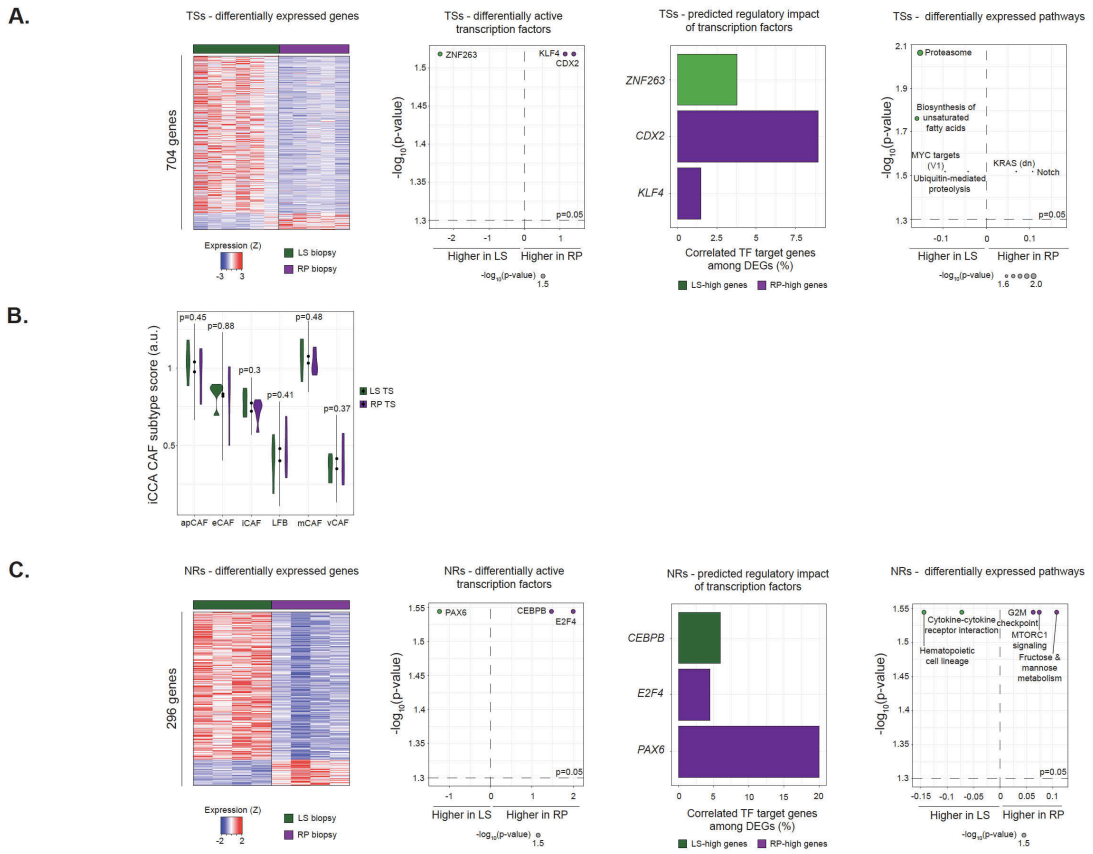


Figure S5: Transcriptomic profiles of tumour stroma and non-malignant regions of biopsies in the RPLS cohort. (A) Differentially expressed genes (min. 2-fold difference, p<0.05, Wilcoxon rank-sum test), differentially active transcription factors (p<0.05, Wilcoxon rank-sum test; DoRothEA), differentially expressed pathways (p<0.05, Wilcoxon rank-sum test; ssGSEA of KEGG and Hallmarks gene lists), and differentially active cytokines (p<0.05, Wilcoxon rank-sum test; CytoSig) between RP (n=6) and LS (n=5) tumour stroma. (B) Differential expression of iCCA cancer-associated fibroblast subtypes in tumour stroma. (C) Differentially expressed genes (min. 2-fold difference, p<0.05, Wilcoxon rank-sum test), differentially active transcription factors (p<0.05, Wilcoxon rank-sum test; DoRothEA), differentially expressed pathways (p<0.05, Wilcoxon rank-sum test; ssGSEA of KEGG and

Hallmarks gene lists), and differentially active cytokines (p<0.05, Wilcoxon rank-sum test; CytoSig) between RP (n=4) and LS (n=4) non-malignant regions.

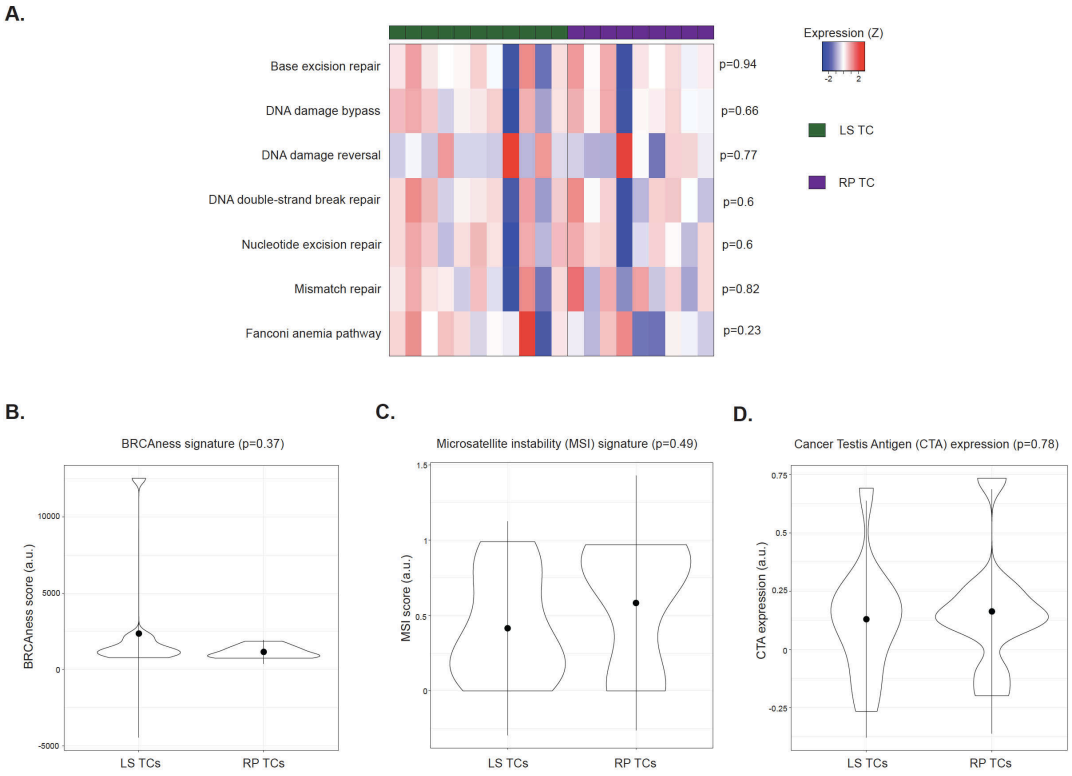


Figure S6: Genomic stability and immunogenicity of tumour cores in the RPLS cohort. Differential expression (Wilcoxon test) of signatures for (A) DNA repair processes (Reactome), (B) BRACness, (C) microsatellite instability, and (D) cancer testis antigen expression in LS and RP biopsies.

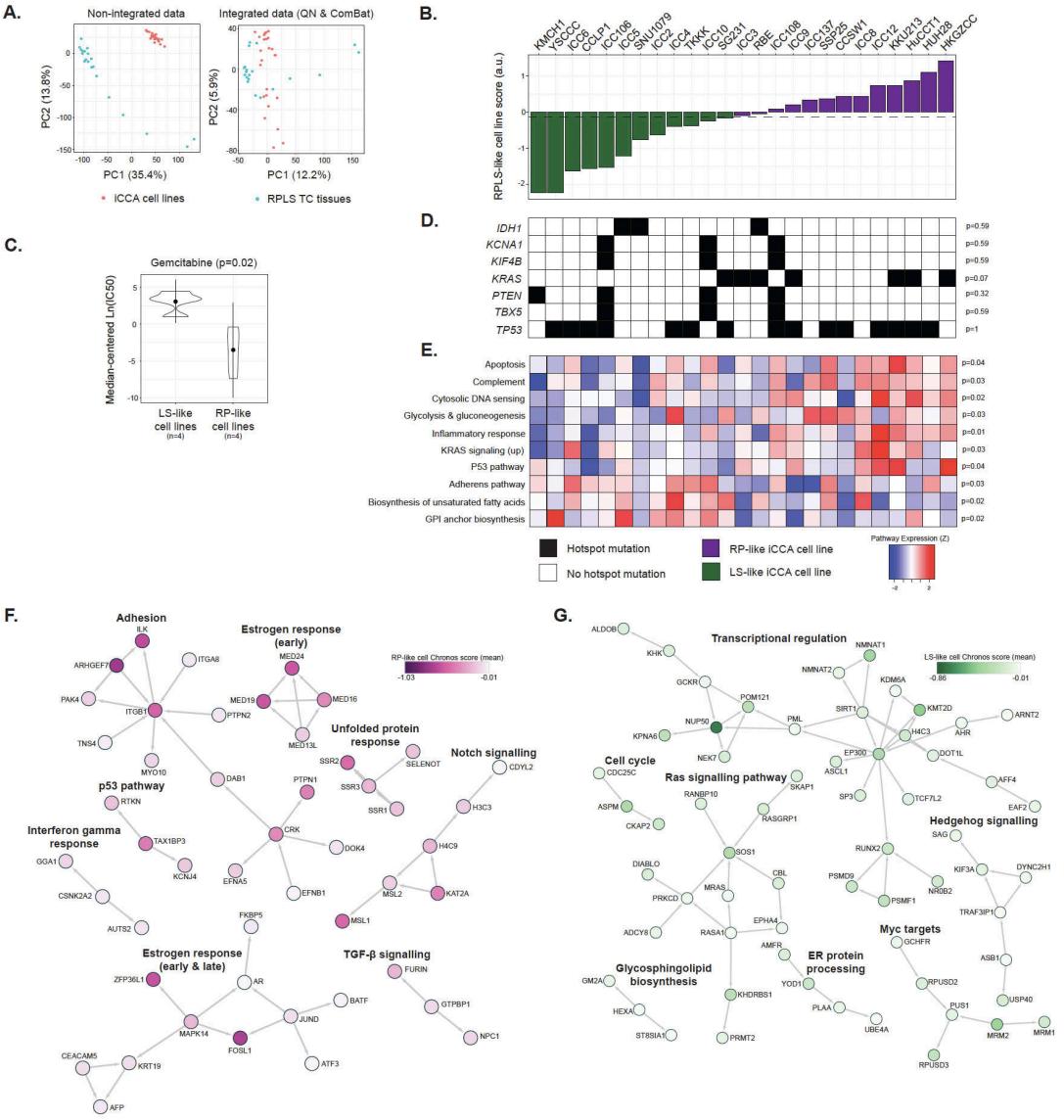


Figure S7: Evaluation of LS- and RP-like iCCA cell lines. (A) Principal component analysis of RPLS tumour core (TC) and iCCA cell line (DepMap) transcriptome profiles before and after data integration. PC: principal component; QN: quantile normalization. (B) Differential sensitivity of LS- and RP-like iCCA cell lines to gemcitabine using data from Saha et al. (Cancer Discovery 2016). IC50: half maximal inhibitory concentration; Ln: natural log. (C) Hotspot mutations in recurrently altered genes across LS- and RP-like cell lines (p-values from fisher's exact test).

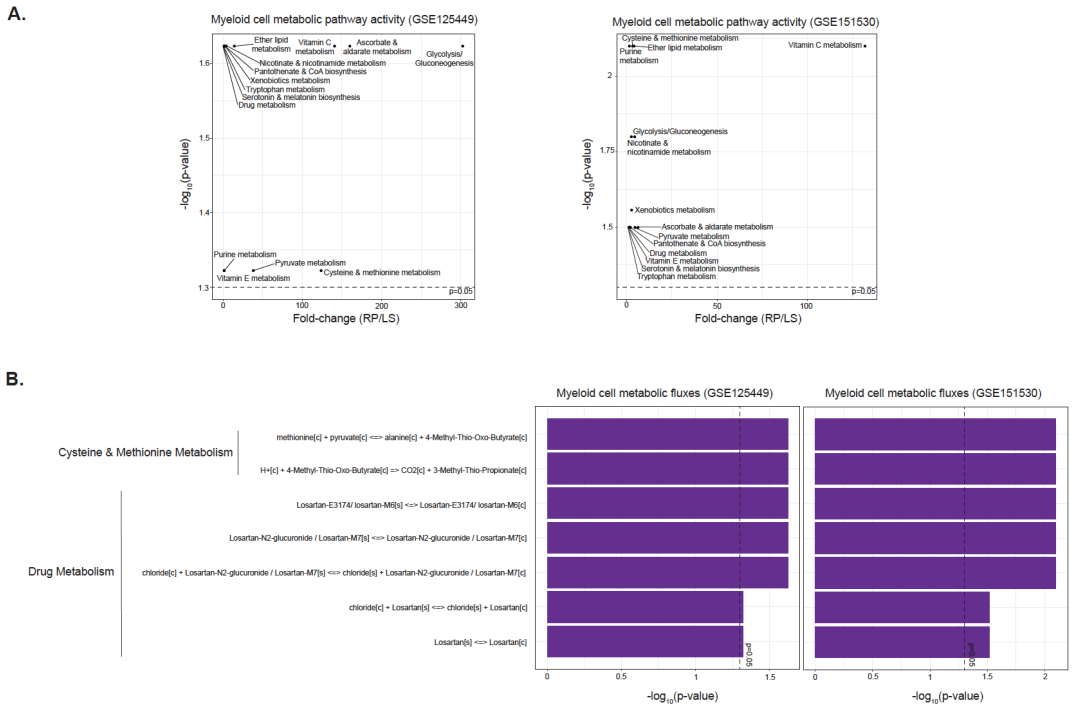


Figure S8: Metabolic flux inference in LS- and RP-like myeloid cells. (A) Differentially active metabolic pathways between myeloid cells in RP-like and LS-like tumours. Activity was inferred using METAFlux in two single cell RNA-sequencing datasets. Only pathways consistently differentially active (Wilcoxon p<0.05) in both datasets are shown. (B) Differentially active reactions (within differentially active pathways) between myeloid cells in RP-like and LS-like tumours. Reactions were pre-defined by BiGG Models and activity was inferred using METAflux in two single cell RNA-sequencing datasets. Only reactions consistently differentially active (Wilcoxon p<0.05) in both datasets are shown.

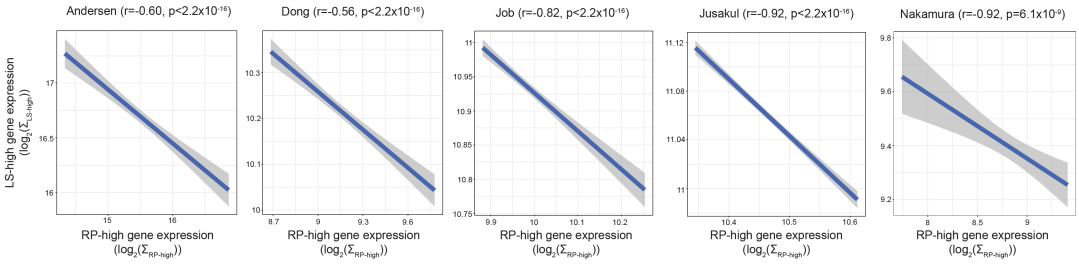


Figure S9: Correlation analysis of RP-high and LS-high gene expression in resected cohorts. Statistics were computed by Spearman's correlation statistics.

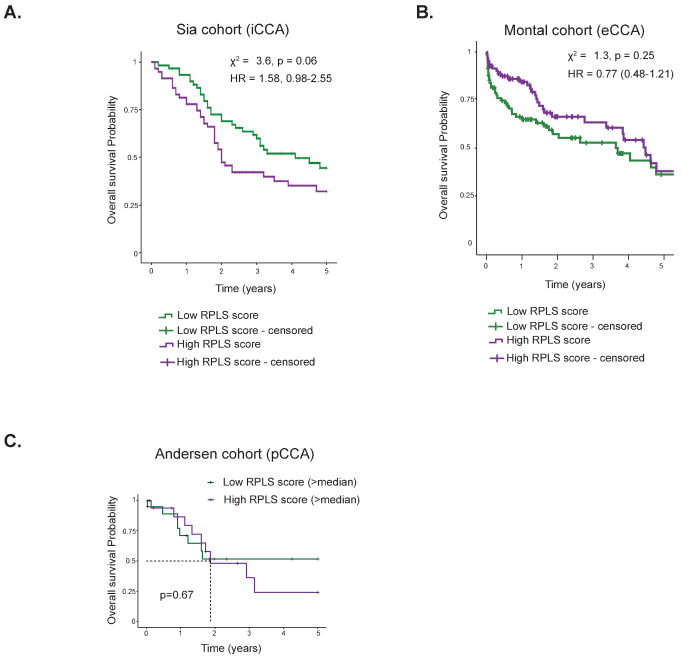


Figure S10: Prognostic associations of RPLS scores in additional resected CCA cohorts. Kaplan-Meier survival curves with log-rank statistics for (A) overall survival in the Sia iCCA cohort (n=149), (B) overall survival in the Montal eCCA cohort (n=182), and (C) overall survival in the Andersen pCCA cohort (n=37). Samples were stratified into high (>median) or low (<median) RPLS score groups.

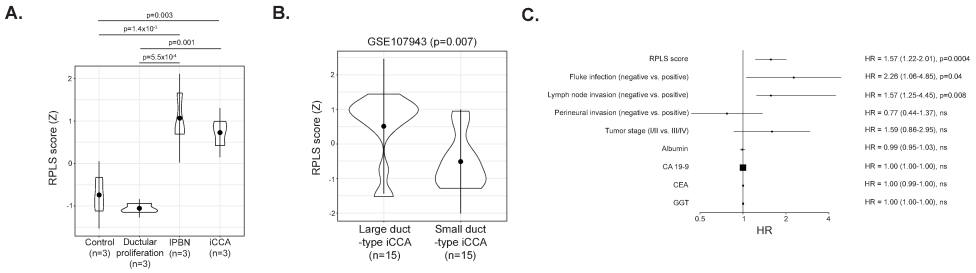


Figure S11: RPLS signature expression in mouse and human iCCA tissues. (A) Differential expression of the RPLS signature in microdissected tissue regions from mice with *Kras*^{G12D} expressing mice and diet-induced cholangitis. P-values were computed by Welch t-test. IPNB: intraductal papillary neoplasm of the bile duct. (B) Differential expression of the RPLS signature between small- (n=15) and large duct-type (n=15) iCCA. P-values were computed by Wilcoxon test. (C) Forest plot of Cox proportional hazards statistics derived from multivariable

analysis of RPLS scores and correlated clinicopathologic variables in the Dong cohort (n=244). γ-GT: γ-glutamyltransferase; CA 19-9: carbohydrate antigen 19-9; CEA: carcinoembryonic antigen; HR: hazard ratio; ns: not significant.

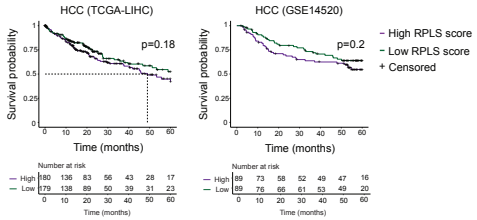


Figure S12: Prognostic associations of the RPLS signature in hepatocellular carcinoma. Kaplan-Meier survival curves with log-rank statistics for high (>median) and low (<median) RPLS signature expression in the TCGA-LIHC and GSE14520 cohorts of resected hepatocellular carcinoma patients.

SUPPLEMENTARY TABLE TITLES

Table S1: Baseline clinical characteristics of the RPLS cohort.

Table S2: Differential expression of the RPLS gene signature.

Table S3: Differential genetic dependencies of RP-like and LS-like iCCA cell lines based on genome-wide CRISPR inactivation screens.

Table S4: Differential pathway expression between RP-like and LS-like tumo**u**r cells in scRNA-seq from iCCA tissues.

Table S5: Differential activity of cytokines and transcription factors between cell types from RP- and LS-like tumo**u**rs.

Table S6: Ligand-receptor interactions specific to RP-like tumo**u**r cells.

Table S7: Druggability of ligand-stimulated receptors on RP-like tumo**u**r cells.

Table S8: Clinicopathologic associations of the RPLS signature across resected iCCA cohorts.

Table S9: Molecular pathobiological differences between liver metastatases and other metastases in advanced cancers.

O'Rourke CJ, *et al. Gut* 2023;0:1–13. doi: 10.1136/gutjnl-2023-330748

Table S10: Univariable and multivariable progression-free survival (PFS) analysis of the RPLS signature, clinicopathologic and genomic features in colorectal cancer liver metastases (New EPOC trial).

Table S11: Univariable and multivariable overall survival (OS) analysis of the RPLS signature, clinicopathologic and genomic features in colorectal cancer liver metastases (New EPOC trial).

SUPPLEMENTARY REFERENCES

1 Valle J, Wasan H, Palmer DH, Cunningham D, Anthoney A, Maraveyas A*, et al.* Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N Engl J Med 2010;**362**:1273-81.

2 Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD*, et al.* QuPath: Open source software for digital pathology image analysis. Sci Rep 2017;**7**:16878.

3 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;**15**:550.

4 Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G*, et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol 2020;**38**:586-99.

5 Fisher NC, Byrne RM, Leslie H, Wood C, Legrini A, Cameron AJ*, et al.* Biological Misinterpretation of Transcriptional Signatures in Tumor Samples Can Unknowingly Undermine Mechanistic Understanding and Faithful Alignment with Preclinical Data. Clin Cancer Res 2022;**28**:4056-69.

6 Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;**44**:D457-62.

7 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;**13**:2498-504.

8 Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;**5**:e13984.

9 Kucera M, Isserlin R, Arkhangorodsky A, Bader GD. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. F1000Res 2016;**5**:1717.

10 Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;**14**:7.

11 Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;**1**:417-25.

12 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;**102**:15545-50.

13 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;**27**:1739-40.

14 Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res 2019;**29**:1363-75.

15 Nakano Y, Tanno S, Koizumi K, Nishikawa T, Nakamura K, Minoguchi M*, et al.* Gemcitabine chemoresistance and molecular markers associated with gemcitabine transport and metabolism in human pancreatic cancer cells. Br J Cancer 2007;**96**:457-63.
16 Sebastiani V, Ricci F, Rubio-Viqueira B, Kulesza P, Yeo CJ, Hidalgo M*, et al.* Immunohistochemical and genetic evaluation of deoxycytidine kinase in pancreatic cancer: relationship to molecular mechanisms of gemcitabine resistance and survival. Clin Cancer Res 2006;**12**:2492-7.
17 Vos LJ, Yusuf D, Lui A, Abdelaziz Z, Ghosh S, Spratlin JL*, et al.* Predictive and Prognostic Properties of Human Equilibrative Nucleoside Transporter 1 Expression in Gemcitabine-Treated Pancreatobiliary Cancer: A Meta-Analysis. JCO Precis Oncol 2019;**3**:1-22.
18 Mourragui SMC, Loog M, Vis DJ, Moore K, Manjon AG, van de Wiel MA*, et al.* Predicting patient response with models trained on cell lines and patient-derived xenografts by nonlinear transfer learning. Proc Natl Acad Sci U S A 2021;**118**.
19 Nicolle R, Gayet O, Duconseil P, Vanbrugghe C, Roques J, Bigonnet M*, et al.* A transcriptomic signature to predict adjuvant gemcitabine sensitivity in pancreatic adenocarcinoma. Ann Oncol 2021;**32**:250-60.
20 Tiriac H, Belleau P, Engle DD, Plenker D, Deschenes A, Somerville TDD*, et al.* Organoid Profiling Identifies Common Responders to Chemotherapy in Pancreatic Cancer. Cancer Discov 2018;**8**:1112-29.
21 Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F*, et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;**37**:773-82.
22 Ma L, Hernandez MO, Zhao Y, Mehta M, Tran B, Kelly M*, et al.* Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. Cancer Cell 2019;**36**:418-30 e6.
23 Ma L, Wang L, Khatib SA, Chang CW, Heinrich S, Dominguez DA*, et al.* Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol 2021.
24 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd*, et al.* Comprehensive Integration of Single-Cell Data. Cell 2019;**177**:1888-902 e21.
25 Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;**18**:220.
26 Sturm G, Finotello F, List M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. Methods Mol Biol 2020;**2120**:223-32.
27 Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 2015;**160**:48-61.
28 Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R*, et al.* Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. Nat Methods 2021;**18**:1181-91.
29 Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, Cohen T*, et al.* CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res 2009;**37**:D816-9.
30 Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X*, et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;**24**:1550-8.
31 Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilas E, Francoeur N*, et al.* Gene expression profile of BRCAness that correlates with responsiveness to

chemotherapy and with outcome in patients with epithelial ovarian cancer. J Clin Oncol 2010;**28**:3555-61.
32 Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A*, et al.* The reactome pathway knowledgebase 2022. Nucleic Acids Res 2022;**50**:D687-D92.
33 Affo S, Nair A, Brundu F, Ravichandra A, Bhattacharjee S, Matsuda M*, et al.* Promotion of cholangiocarcinoma growth by diverse cancer-associated fibroblast subpopulations. Cancer Cell 2021;**39**:866-82 e11.
34 Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd*, et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature 2019;**569**:503-8.
35 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;**28**:882-3.
36 Saha SK, Gordan JD, Kleinstiver BP, Vu P, Najem MS, Yeo JC*, et al.* Isocitrate Dehydrogenase Mutations Confer Dasatinib Hypersensitivity and SRC Dependence in Intrahepatic Cholangiocarcinoma. Cancer Discov 2016;**6**:727-39.
37 Pacini C, Dempster JM, Boyle I, Goncalves E, Najgebauer H, Karakoc E*, et al.* Integrated cross-study datasets of genetic dependencies in cancer. Nat Commun 2021;**12**:1661.
38 Dempster JM, Boyle I, Vazquez F, Root DE, Boehm JS, Hahn WC*, et al.* Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. Genome Biol 2021;**22**:343.
39 Dempster JM, Pacini C, Pantel S, Behan FM, Green T, Krill-Burger J*, et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. Nat Commun 2019;**10**:5817.
40 Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ*, et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 2021;**49**:D1144-D51.
41 Ma L, Wang L, Khatib SA, Chang CW, Heinrich S, Dominguez DA*, et al.* Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol 2021;**75**:1397-408.
42 Alvisi G, Termanini A, Soldani C, Portale F, Carriero R, Pilipow K*, et al.* Multimodal single-cell profiling of intrahepatic cholangiocarcinoma defines hyperactivated Tregs as a potential therapeutic target. J Hepatol 2022;**77**:1359-72.
43 MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK*, et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun 2018;**9**:4383.
44 Wang S, He Z, Wang X, Li H, Liu XS. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. Elife 2019;**8**.
45 Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D*, et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep 2017;**18**:248-62.
46 Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH*, et al.* Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021;**12**:1088.
47 Sheils TK, Mathias SL, Kelleher KJ, Siramshetty VB, Nguyen DT, Bologa CG*, et al.* TCRD and Pharos 2021: mining the human proteome for disease biology. Nucleic Acids Res 2021;**49**:D1334-D46.

48 Andreatta M, Corria-Osorio J, Muller S, Cubas R, Coukos G, Carmona SJ. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. Nat Commun 2021;**12**:2965.
49 Eddy JA, Thorsson V, Lamb AE, Gibbs DL, Heimann C, Yu JX*, et al.* CRI iAtlas: an interactive portal for immuno-oncology research. F1000Res 2020;**9**:1028.
50 Huang Y, Mohanty V, Dede M, Tsai K, Daher M, Li L*, et al.* Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux. Nat Commun 2023;**14**:4883.
51 Andersen JB, Spee B, Blechacz BR, Avital I, Komuta M, Barbour A*, et al.* Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology 2012;**142**:1021-31 e15.
52 O'Rourke CJ, Matter MS, Nepal C, Caetano-Oliveira R, Ton PT, Factor VM*, et al.* Identification of a Pan-Gamma-Secretase Inhibitor Response Signature for Notch-Driven Cholangiocarcinoma. Hepatology 2020;**71**:196-213.
53 Dong L, Lu D, Chen R, Lin Y, Zhu H, Zhang Z*, et al.* Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma. Cancer Cell 2022;**40**:70-87 e15.
54 Ahn KS, O'Brien D, Kang YN, Mounajjed T, Kim YH, Kim TS*, et al.* Prognostic subclass of intrahepatic cholangiocarcinoma by integrative molecular-clinical analysis and potential targeted approach. Hepatol Int 2019;**13**:490-500.
55 Ahn KS, Kang KJ, Kim YH, Kim TS, Song BI, Kim HW*, et al.* Genetic features associated with (18)F-FDG uptake in intrahepatic cholangiocarcinoma. Ann Surg Treat Res 2019;**96**:153-61.
56 Job S, Rapoud D, Dos Santos A, Gonzalez P, Desterke C, Pascal G*, et al.* Identification of Four Immune Subtypes Characterized by Distinct Composition and Functions of Tumor Microenvironment in Intrahepatic Cholangiocarcinoma. Hepatology 2020;**72**:965-81.
57 Jusakul A, Cutcutache I, Yong CH, Lim JQ, Huang MN, Padmanabhan N*, et al.* Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma. Cancer Discov 2017;**7**:1116-35.
58 Montal R, Sia D, Montironi C, Leow WQ, Esteban-Fabro R, Pinyol R*, et al.* Molecular classification and therapeutic targets in extrahepatic cholangiocarcinoma. J Hepatol 2020;**73**:315-27.
59 Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A, Kato M*, et al.* Genomic spectra of biliary tract cancer. Nat Genet 2015;**47**:1003-10.
60 Nepal C, O'Rourke CJ, Oliveira D, Taranta A, Shema S, Gautam P*, et al.* Genomic perturbations reveal distinct regulatory networks in intrahepatic cholangiocarcinoma. Hepatology 2018;**68**:949-63.
61 Sia D, Hoshida Y, Villanueva A, Roayaie S, Ferrer J, Tabak B*, et al.* Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. Gastroenterology 2013;**144**:829-40.
62 Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P*, et al.* Accelerating Discovery of Functional Mutant Alleles in Cancer. Cancer Discov 2018;**8**:174-83.
63 Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Mouse Genome Database G. Mouse Genome Database (MGD) 2019. Nucleic Acids Res 2019;**47**:D801-D6.
64 Mu X, Pradere JP, Affo S, Dapito DH, Friedman R, Lefkovitch JH*, et al.* Epithelial Transforming Growth Factor-beta Signaling Does Not Contribute to Liver Fibrosis but Protects Mice From Cholangiocarcinoma. Gastroenterology 2016;**150**:720-33.

65 Cubero FJ, Mohamed MR, Woitok MM, Zhao G, Hatting M, Nevzorova YA*, et al.* Loss of c-Jun N-terminal Kinase 1 and 2 Function in Liver Epithelial Cells Triggers Biliary Hyperproliferation Resembling Cholangiocarcinoma. Hepatol Commun 2020;**4**:834-51.
66 Di-Luoffo M, Pirenne S, Saandi T, Loriot A, Gerard C, Dauguet N*, et al.* A novel mouse model of cholangiocarcinoma uncovers a role for Tensin-4 in tumor progression. Hepatology 2021.
67 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;**26**:139-40.
68 Zhang S, Song X, Cao D, Xu Z, Fan B, Che L*, et al.* Pan-mTOR inhibitor MLN0128 is effective against intrahepatic cholangiocarcinoma in mice. J Hepatol 2017;**67**:1194-203.
69 Fan B, Malato Y, Calvisi DF, Naqvi S, Razumilava N, Ribback S*, et al.* Cholangiocarcinomas can originate from hepatocytes in mice. J Clin Invest 2012;**122**:2911-5.
70 Younger NT, Wilson ML, Martinez Lyons A, Jarman EJ, Meynert AM, Grimes GR*, et al.* In Vivo Modeling of Patient Genetic Heterogeneity Identifies New Ways to Target Cholangiocarcinoma. Cancer Res 2022;**82**:1548-59.
71 Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas Research N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017;**169**:1327-41 e23.
72 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;**2**:401-4.
73 Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS*, et al.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010;**70**:10202-12.
74 Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J*, et al.* Integrative clinical genomics of metastatic cancer. Nature 2017;**548**:297-303.
75 Pleasance E, Bohm A, Williamson LM, Nelson JMT, Shen Y, Bonakdar M*, et al.* Whole genome and transcriptome analysis enhances precision cancer treatment options. Ann Oncol 2022.
76 Bridgewater JA, Pugh SA, Maishman T, Eminton Z, Mellor J, Whitehead A*, et al.* Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis (New EPOC): long-term results of a multicentre, randomised, controlled, phase 3 trial. Lancet Oncol 2020;**21**:398-411.

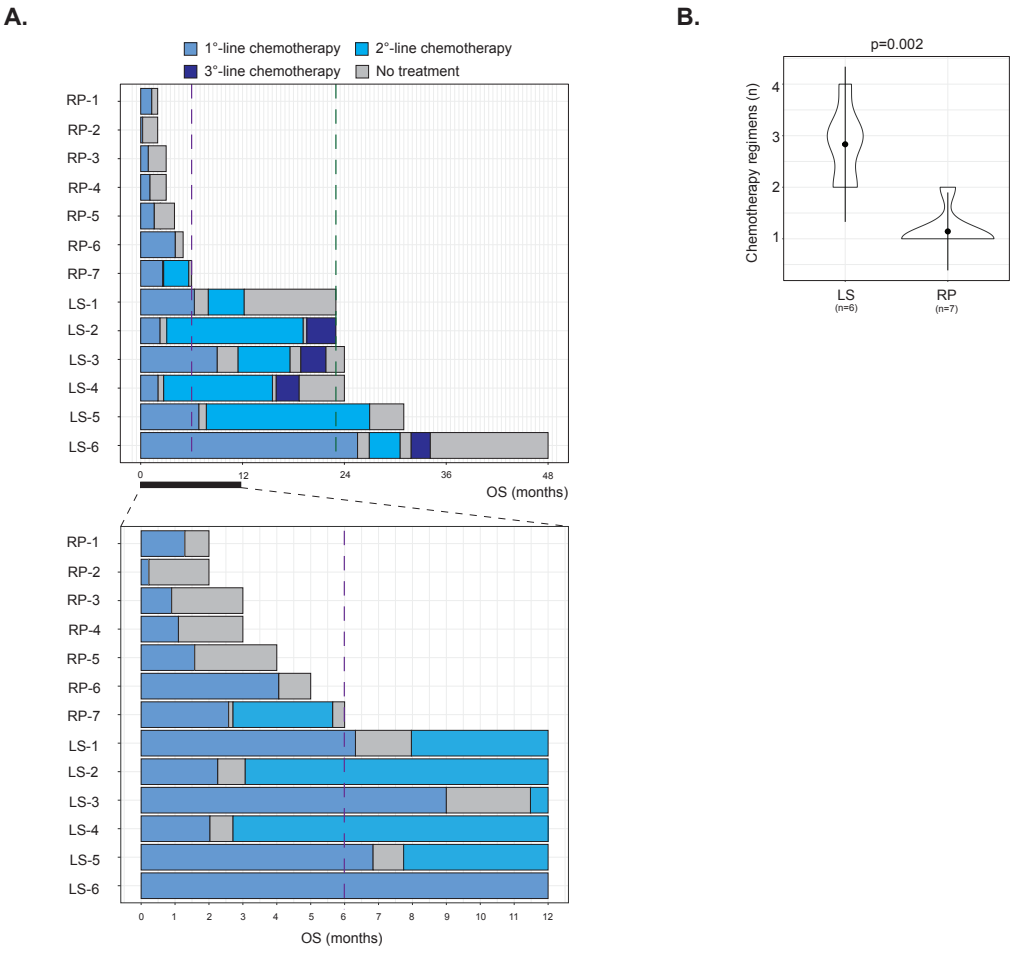

A.
1°-line chemotherapy
2°-line chemotherapy
3°-line chemotherapy
No treatment
RP-1
RP-2
RP-3
RP-4
RP-5
RP-6
RP-7
LS-1
LS-2
LS-3
LS-4
LS-5
LS-6
0
12
24
36
48
OS (months)
0
1
2
3
4
5
6
7
8
9
10
11
12
OS (months)
B.
p=0.002
Chemotherapy regimens (n)
1
2
3
4
LS
(n=6)
RP
(n=7)

FIGURE S1

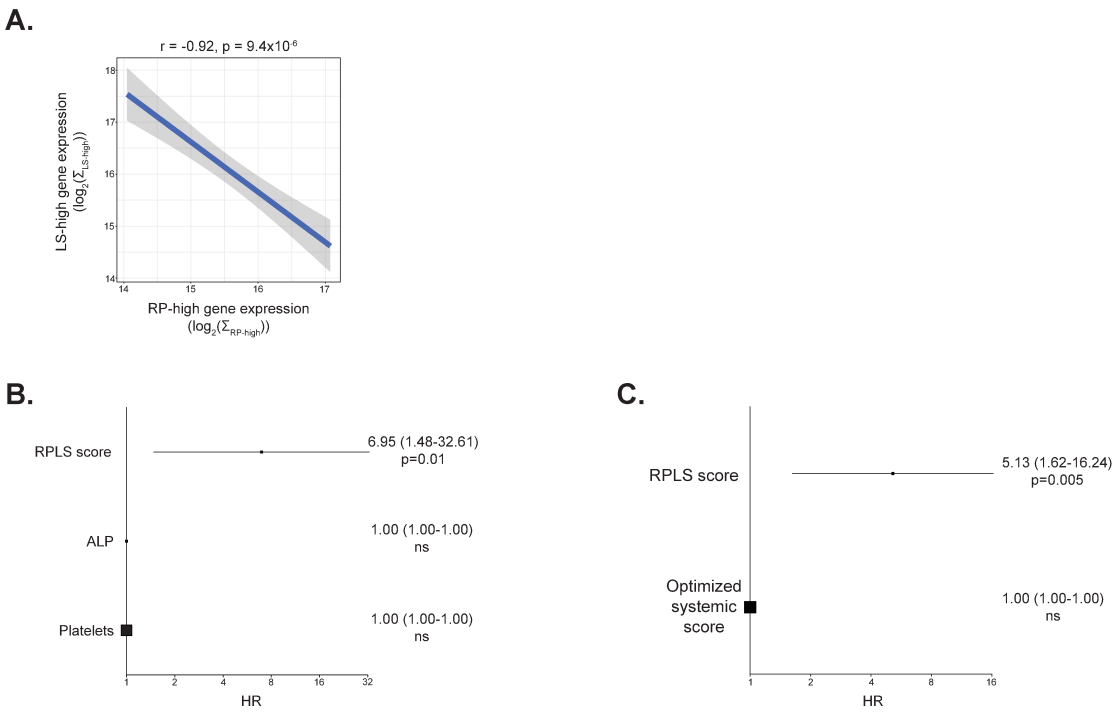

A.
r = -0.92, p = 9.4x10-6
LS-high gene expression
(log2(ΣLS-high))
RP-high gene expression
(log2(ΣRP-high))
B.
RPLS score
6.95 (1.48-32.61)
p=0.01
ALP
1.00 (1.00-1.00)
ns
Platelets
1.00 (1.00-1.00)
ns
HR
C.
RPLS score
5.13 (1.62-16.24)
p=0.005
Optimized systemic score
1.00 (1.00-1.00)
ns
HR

FIGURE S2

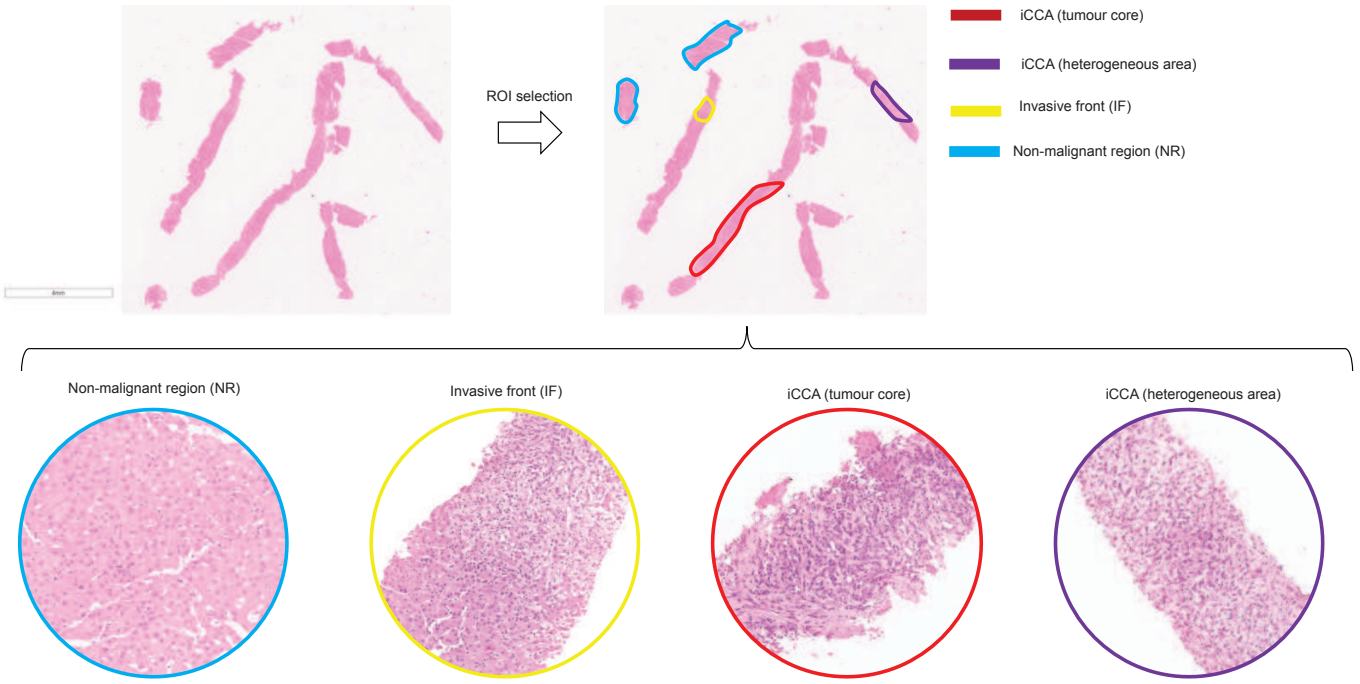
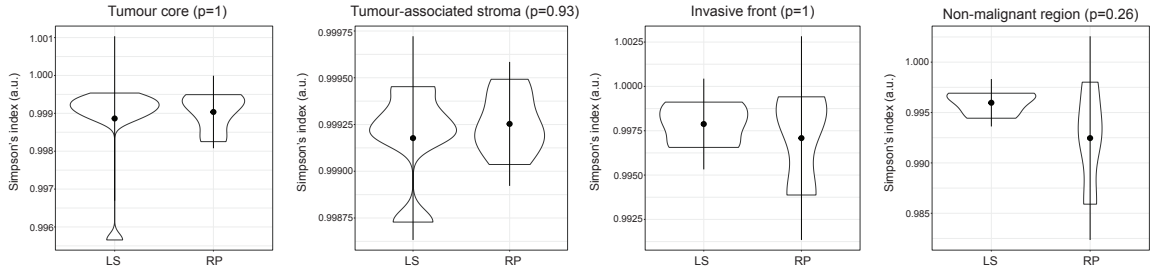

ROI selection
iCCA (tumour core)
iCCA (heterogeneous area)
Invasive front (IF)
Non-malignant region (NR)
Non-malignant region (NR)
Invasive front (IF)
iCCA (tumour core)
iCCA (heterogeneous area)

FIGURE S3

A.


Tumour core (p=1)
Tumour-associated stroma (p=0.93)
Invasive front (p=1)
Non-malignant region (p=0.26)
Simpson's index (a.u.)
LS
RP

B.

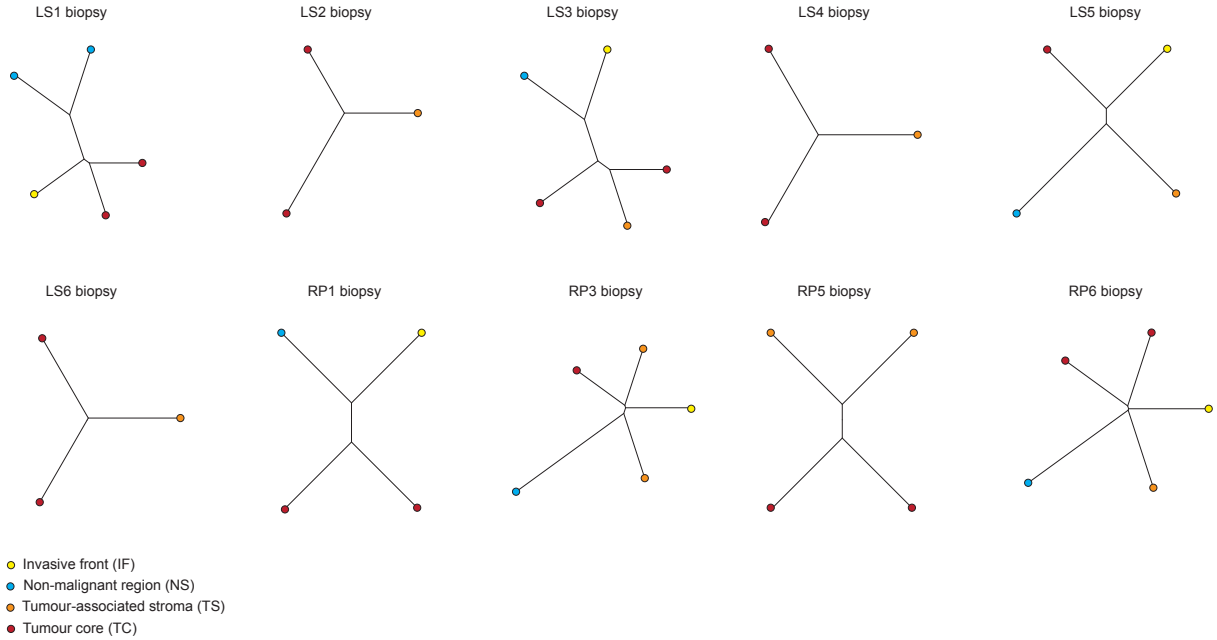

LS1 biopsy
LS2 biopsy
LS3 biopsy
LS4 biopsy
LS5 biopsy
LS6 biopsy
RP1 biopsy
RP3 biopsy
RP5 biopsy
RP6 biopsy
Invasive front (IF)
Non-malignant region (NS)
Tumour-associated stroma (TS)
Tumour core (TC)

FIGURE S4

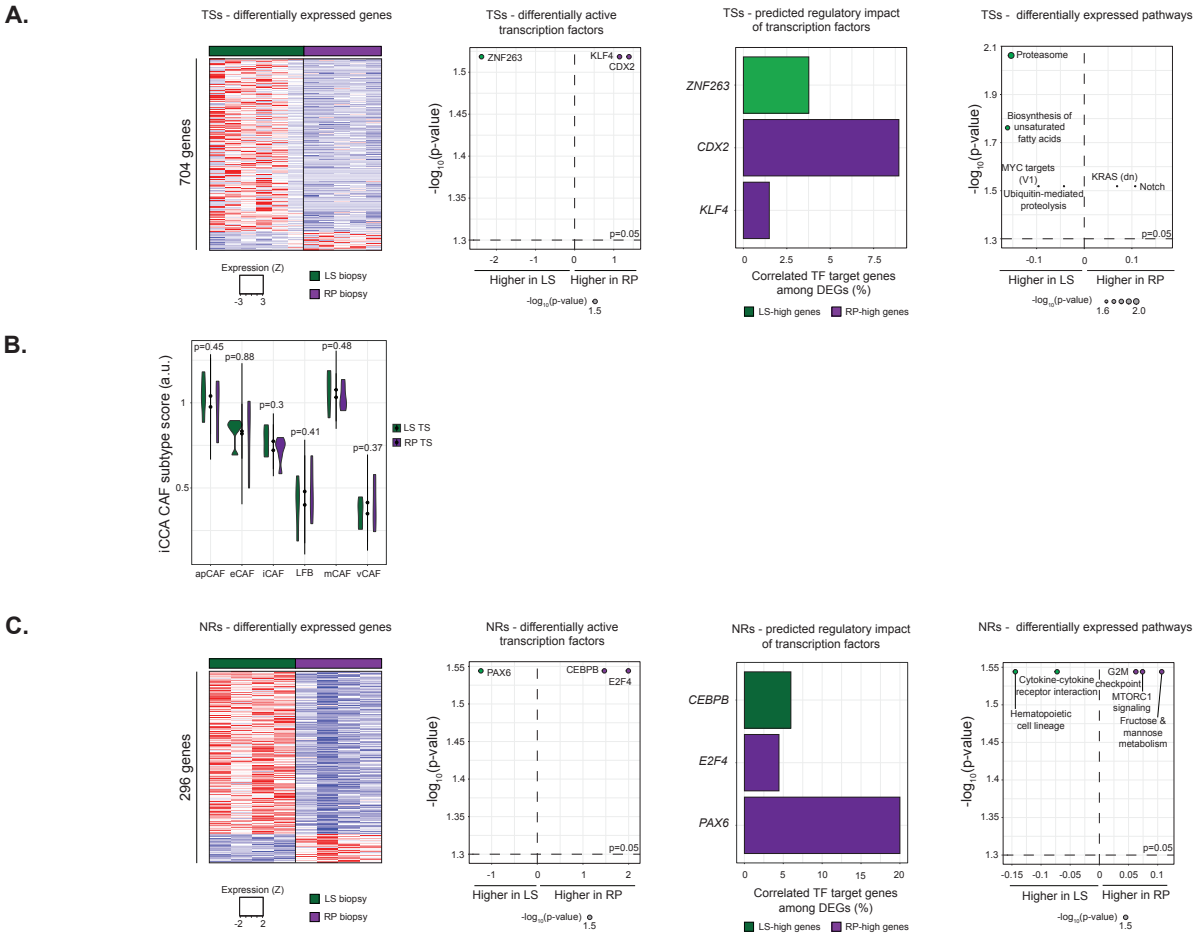

A.
TSs - differentially expressed genes
704 genes
Expression (Z)
LS biopsy
RP biopsy
TSs - differentially active transcription factors
ZNF263
KLF4
CDX2
p=0.05
Higher in LS
Higher in RP
TSs - predicted regulatory impact of transcription factors
ZNF263
CDX2
KLF4
Correlated TF target genes among DEGs (%)
LS-high genes
RP-high genes
TSs - differentially expressed pathways
Proteasome
Biosynthesis of unsaturated fatty acids
MYC targets (V1)
Ubiquitin-mediated proteolysis
KRAS (dn)
Notch
B.
iCCA CAF subtype score (a.u.)
p=0.45
p=0.88
p=0.3
p=0.41
p=0.48
p=0.37
LS TS
RP TS
apCAF
eCAF
iCAF
LFB
mCAF
vCAF
C.
NRs - differentially expressed genes
296 genes
Expression (Z)
LS biopsy
RP biopsy
NRs - differentially active transcription factors
PAX6
CEBPB
E2F4
p=0.05
Higher in LS
Higher in RP
NRs - predicted regulatory impact of transcription factors
CEBPB
E2F4
PAX6
Correlated TF target genes among DEGs (%)
LS-high genes
RP-high genes
NRs - differentially expressed pathways
Cytokine-cytokine receptor interaction
Hematopoietic cell lineage
G2M checkpoint
MTORC1 signaling
Fructose & mannose metabolism

FIGURE S5

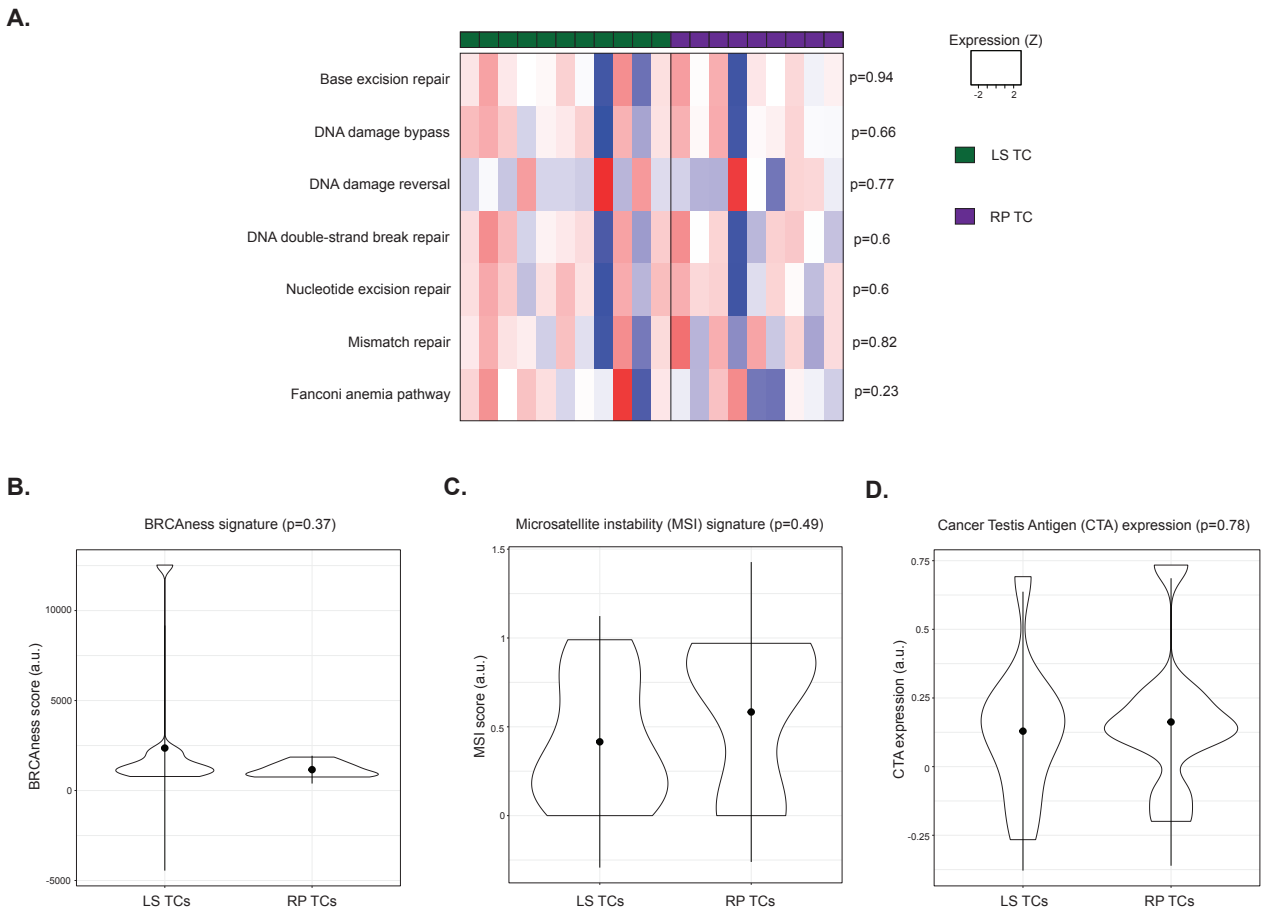
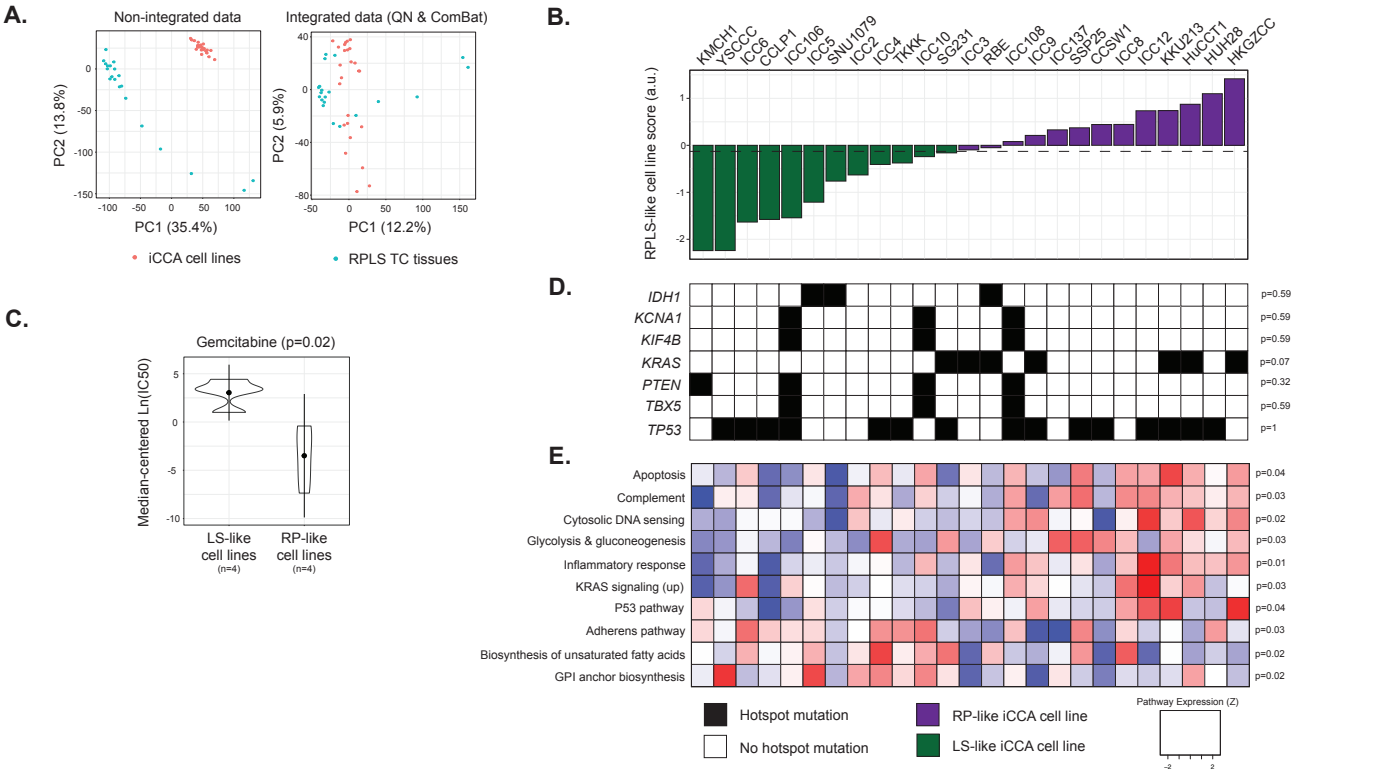

A.
Base excision repair
DNA damage bypass
DNA damage reversal
DNA double-strand break repair
Nucleotide excision repair
Mismatch repair
Fanconi anemia pathway
p=0.94
p=0.66
p=0.77
p=0.6
p=0.6
p=0.82
p=0.23
Expression (Z)
-2
2
LS TC
RP TC
B.
BRCAness signature (p=0.37)
BRCAness score (a.u.)
10000
5000
0
-5000
LS TCs
RP TCs
C.
Microsatellite instability (MSI) signature (p=0.49)
MSI score (a.u.)
1.5
1
0.5
0
LS TCs
RP TCs
D.
Cancer Testis Antigen (CTA) expression (p=0.78)
CTA expression (a.u.)
0.75
0.5
0.25
0
-0.25
LS TCs
RP TCs

FIGURE S6


A.
Non-integrated data
Integrated data (QN & ComBat)
PC2 (13.8%)
PC1 (35.4%)
PC2 (5.9%)
PC1 (12.2%)
iCCA cell lines
RPLS TC tissues
B.
KMCH1 YSCCC ICC6 CCLP1 ICC106 ICC5 SNU1079 ICC2 ICC4 TKKK ICC10 SG231 ICC3 RBE ICC108 ICC9 ICC137 SSP25 CCSW1 ICC8 ICC12 KKU213 HuCCT1 HUH28 HKGZCC
RPLS-like cell line score (a.u.)
C.
Gemcitabine (p=0.02)
Median-centered Ln(IC50)
LS-like cell lines (n=4)
RP-like cell lines (n=4)
D.
IDH1
KCNA1
KIF4B
KRAS
PTEN
TBX5
TP53
p=0.59
p=0.59
p=0.59
p=0.07
p=0.32
p=0.59
p=1
E.
Apoptosis
Complement
Cytosolic DNA sensing
Glycolysis & gluconeogenesis
Inflammatory response
KRAS signaling (up)
P53 pathway
Adherens pathway
Biosynthesis of unsaturated fatty acids
GPI anchor biosynthesis
p=0.04
p=0.03
p=0.02
p=0.03
p=0.01
p=0.03
p=0.04
p=0.03
p=0.02
p=0.02
Hotspot mutation
No hotspot mutation
RP-like iCCA cell line
LS-like iCCA cell line
Pathway Expression (Z)

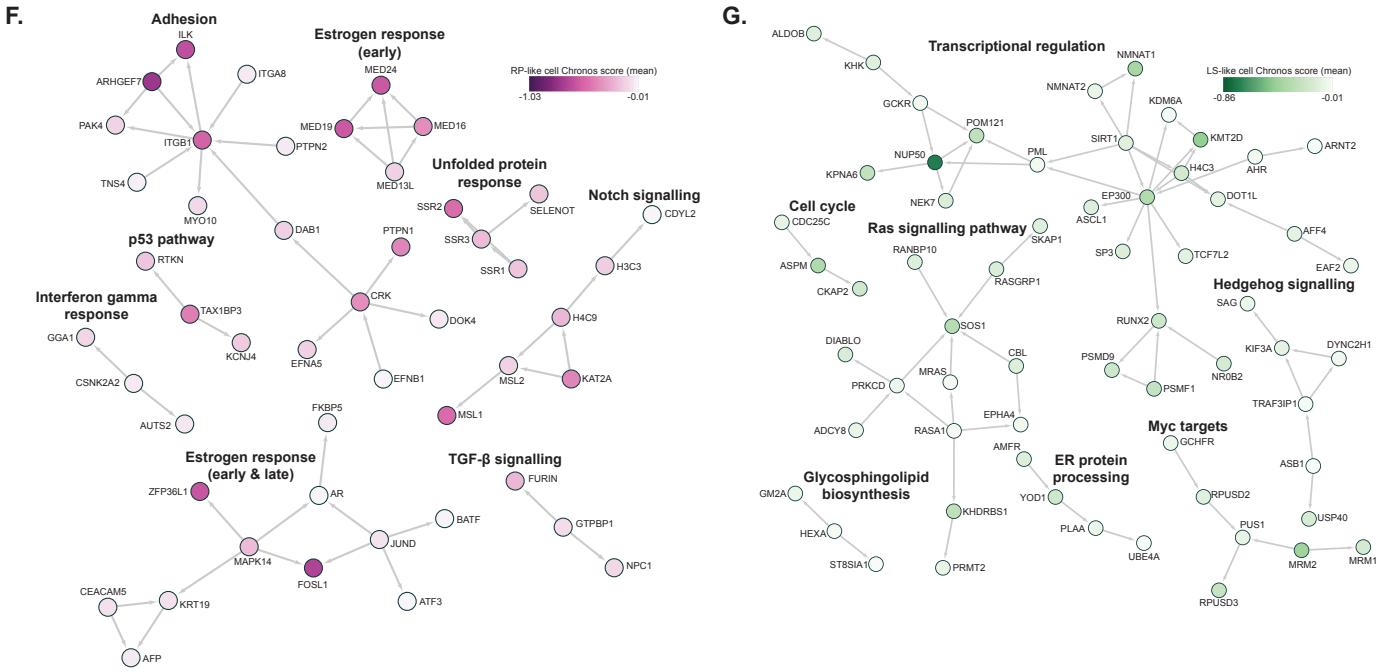
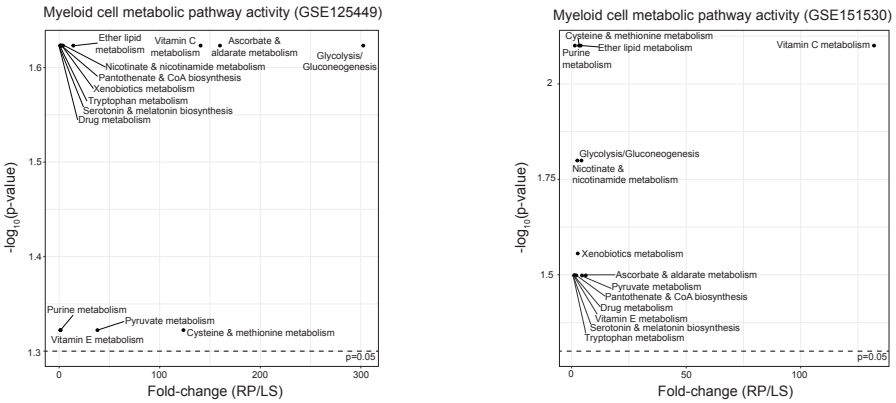

F.
Adhesion
Estrogen response (early)
Unfolded protein response
Notch signalling
p53 pathway
Interferon gamma response
Estrogen response (early & late)
TGF-β signalling
RP-like cell Chronos score (mean)
G.
Transcriptional regulation
Cell cycle
Ras signalling pathway
Hedgehog signalling
Glycosphingolipid biosynthesis
ER protein processing
Myc targets
LS-like cell Chronos score (mean)

FIGURE S7

A.


Myeloid cell metabolic pathway activity (GSE125449)
Ether lipid metabolism
Vitamin C metabolism
Ascorbate & aldarate metabolism
Glycolysis/ Gluconeogenesis
Nicotinate & nicotinamide metabolism
Pantothenate & CoA biosynthesis
Xenobiotics metabolism
Tryptophan metabolism
Serotonin & melatonin biosynthesis
Drug metabolism
Purine metabolism
Pyruvate metabolism
Vitamin E metabolism
Cysteine & methionine metabolism
p=0.05
-log10(p-value)
Fold-change (RP/LS)
Myeloid cell metabolic pathway activity (GSE151530)
Cysteine & methionine metabolism
Ether lipid metabolism
Purine metabolism
Vitamin C metabolism
Glycolysis/Gluconeogenesis
Nicotinate & nicotinamide metabolism
Xenobiotics metabolism
Ascorbate & aldarate metabolism
Pyruvate metabolism
Pantothenate & CoA biosynthesis
Drug metabolism
Vitamin E metabolism
Serotonin & melatonin biosynthesis
Tryptophan metabolism
p=0.05
-log10(p-value)
Fold-change (RP/LS)

B.

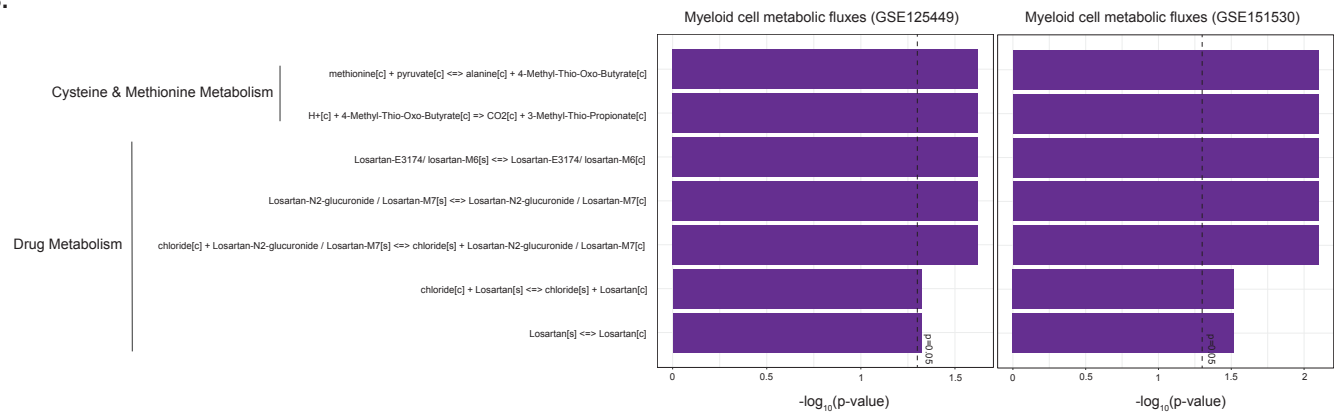

Myeloid cell metabolic fluxes (GSE125449)
Myeloid cell metabolic fluxes (GSE151530)
Cysteine & Methionine Metabolism
methionine[c] + pyruvate[c] <=> alanine[c] + 4-Methyl-Thio-Oxo-Butyrate[c]
H+[c] + 4-Methyl-Thio-Oxo-Butyrate[c] => CO2[c] + 3-Methyl-Thio-Propionate[c]
Drug Metabolism
Losartan-E3174/ losartan-M6[s] <=> Losartan-E3174/ losartan-M6[c]
Losartan-N2-glucuronide / Losartan-M7[s] <=> Losartan-N2-glucuronide / Losartan-M7[c]
chloride[c] + Losartan-N2-glucuronide / Losartan-M7[s] <=> chloride[s] + Losartan-N2-glucuronide / Losartan-M7[c]
chloride[c] + Losartan[s] <=> chloride[s] + Losartan[c]
Losartan[s] <=> Losartan[c]
p=0.05
-log10(p-value)
-log10(p-value)

FIGURE S8

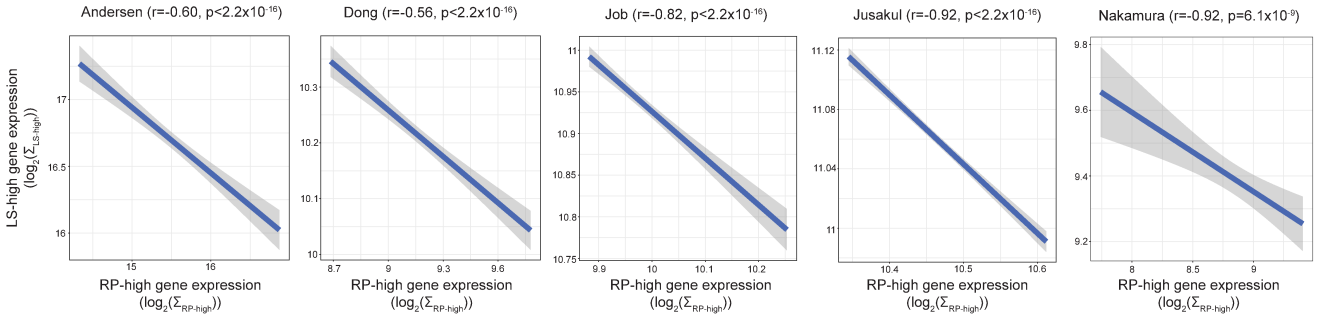
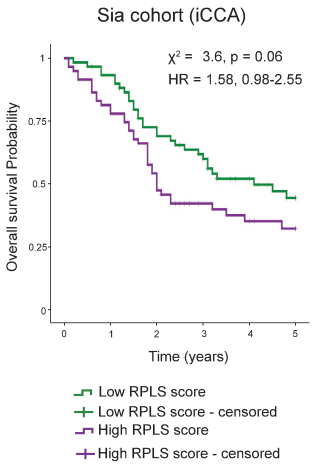
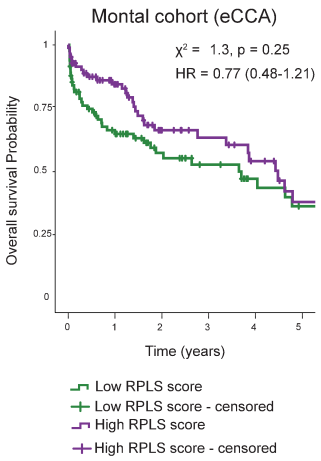

Andersen (r=-0.60, p<2.2x10-16)
Dong (r=-0.56, p<2.2x10-16)
Job (r=-0.82, p<2.2x10-16)
Jusakul (r=-0.92, p<2.2x10-16)
Nakamura (r=-0.92, p=6.1x10-9)
LS-high gene expression (log2(ΣLS-high))
RP-high gene expression (log2(ΣRP-high))

FIGURE S9

A.



B.



C.

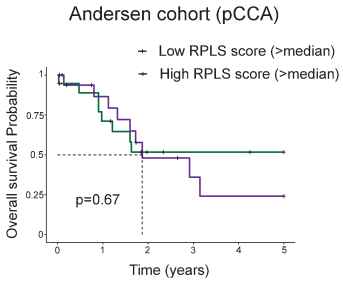
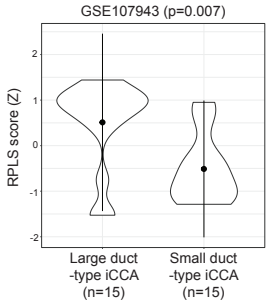


FIGURE S10

A.



B.

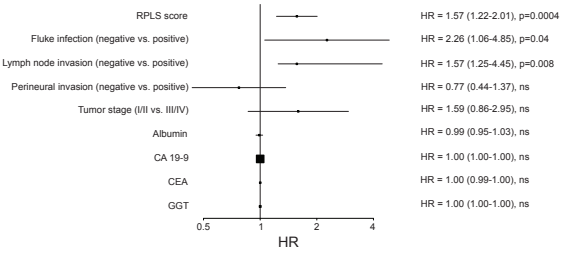


FIGURE S11

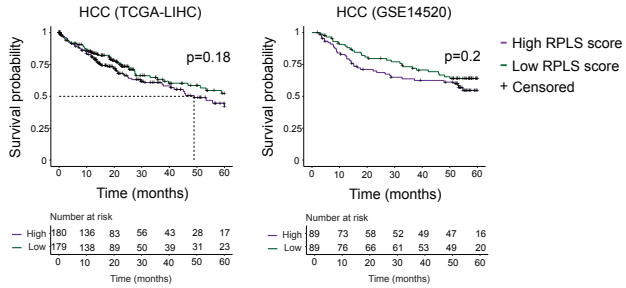

HCC (TCGA-LIHC)
HCC (GSE14520)
Survival probability
p=0.18
p=0.2
Time (months)
Number at risk
High 180 136 83 56 43 28 17
Low 179 138 89 50 39 31 23
High 89 73 58 52 49 47 16
Low 89 76 66 61 53 49 20
High RPLS score
Low RPLS score
Censored

FIGURE S12