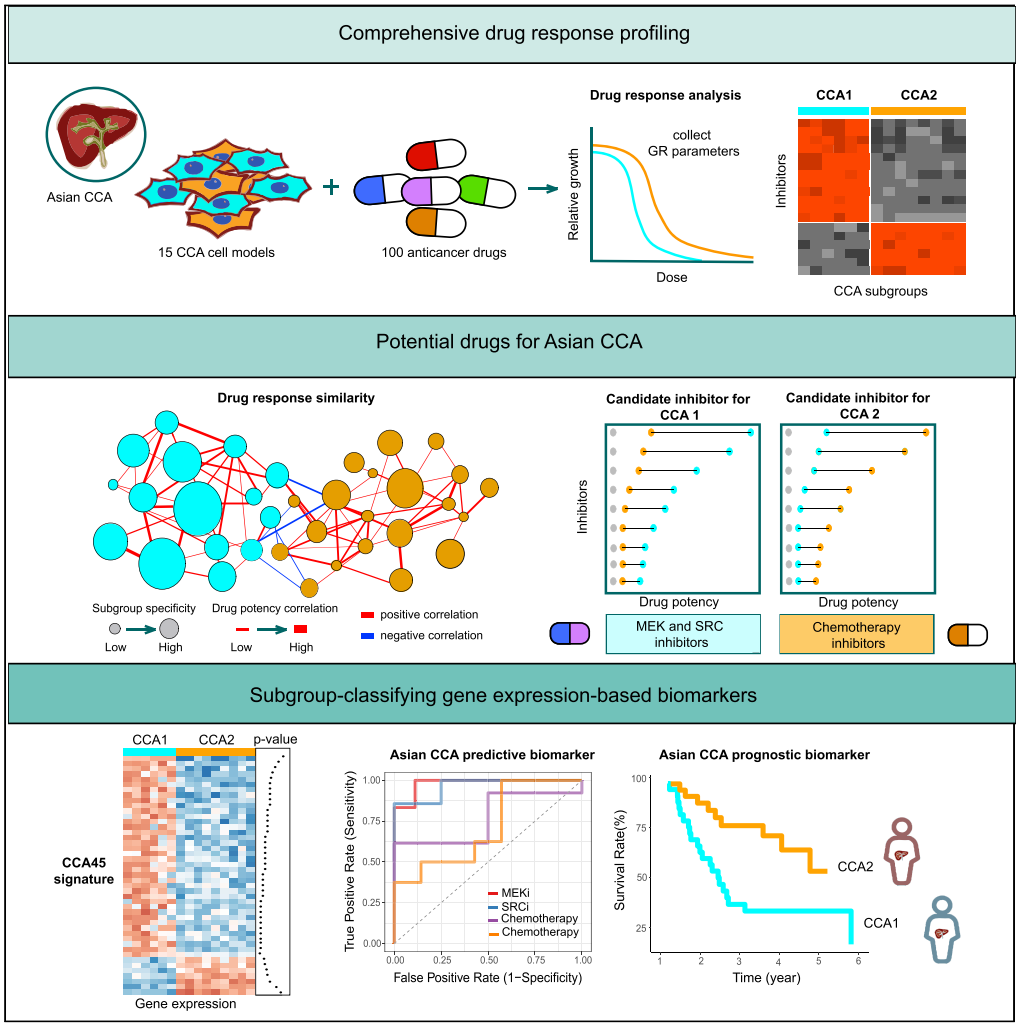
iScience

CellPress
OPEN ACCESS

Article

Comprehensive drug response profiling and pan-omic analysis identified therapeutic candidates and prognostic biomarkers for Asian cholangiocarcinoma



Supawan Jamnongsong, Patipark Kueanjinda, Pongsakorn Buraphat, ..., Seiji Okada, Siwanon Jirawatnotai, Somponnat Sampattavanich

somponnat.sam@mahidol.edu

Highlights

Drug-response profiling separated Asian CCA cell lines into CCA1 and CCA2 subgroups

Drug potency and subgroup specificity enabled therapeutic candidate shortlist

Pan-omic profiling revealed the molecular uniqueness of CCA1 and CCA2 subgroups

CCA45 gene signature predicted poor prognosis in Asian CCA patients

Jamnongsong et al., iScience 25, 105182
October 21, 2022
https://doi.org/10.1016/j.isci.2022.105182

iScience

CellPress
OPEN ACCESS

Article

Comprehensive drug response profiling and pan-omic analysis identified therapeutic candidates and prognostic biomarkers for Asian cholangiocarcinoma

Supawan Jamnongsong,[1,5] Patipark Kueanjinda,[1,2,5] Pongsakorn Buraphat,[1] Phuwanat Sakornsakolpat,[1] Kulthida Vaeteewoottacharn,[3] Seiji Okada,[4] Siwanon Jirawatnotai,[1] and Somponnat Sampattavanich[1,6,*]

SUMMARY

Cholangiocarcinoma (CCA) is rare cancer with the highest incidence in Eastern and Southeast Asian countries. Advanced CCA patients rely on chemotherapeutic regimens that offer unsatisfied clinical outcomes. We developed a comprehensive drug response profiling to investigate potential new drugs using CCA cell lines from Thai and Japanese patients against 100 approved anti-cancer drugs. We identified two major CCA subgroups that displayed unique molecular pathways from our integrative pan-omic and ligand-induced pathway activation analyses. MEK and Src inhibitors specifically killed the CCA1 subgroup without causing cytotoxicity to the normal cholangiocyte. Next, we developed the CCA45 signature to classify CCA patients based on their transcriptomic data. Our CCA45 signature could accurately predict prognosis, especially for Asian CCA patients. Our study provides a comprehensive public resource for drug repurposing in CCA and introduces analytical strategies for prioritizing cancer therapeutic agents for other rare cancer.

INTRODUCTION

Recent progress in precision medicine has made cancer treatment safer and more efficient. Together with the development of molecularly specific therapeutic regimens, regulatory bodies have begun to approve anti-cancer drug use based on specific molecular biomarkers instead of the tumor locations. Such advancement encourages an immediate application of approved targeted therapies in different cancer types via abbreviated drug repurposing trials. Unfortunately, many follow-up clinical studies showed that drug sensitivity in different cancer types did not strictly follow the simple single-biomarker drug indication. Rare cancer types suffer even more from the limited resources in conducting clinical studies and the more time-consuming process of recruiting patients, all contributing to their unsurprisingly limited treatment options.

Cholangiocarcinoma (CCA) is considered one of the rarest cancer types worldwide, although it can be more commonly found in Southeast and East Asia (Sripa et al., 2007). Over the past few decades, treatment of advanced CCA has depended solely on Gemcitabine-based chemotherapy, either as monotherapy or as a combination regimen (i.e., Gemcitabine plus Cisplatin or Gemcitabine plus 5-fluorouracil) (Benson et al., 2021; Ducreux et al., 1998; Valle et al., 2016). Clinical studies reveal that chemotherapeutic options only show a small benefit in approximately 24% of liver fluke (*Opisthorchis viverrini*)-associated CCA patients (Butthongkomvong et al., 2013). Among the drug responder subgroup, the median survival time was only 7–12 months (Butthongkomvong et al., 2013;Valle et al., 2010). Specifically for Thai CCA patients whose diseases are commonly associated with liver fluke infection, the overall median survival was poorer at only four months (with only 6% receiving curative resection; 15% expired before receiving any treatment) (Luvira et al., 2016). The poor clinical outcome of CCA patients urges immediate research to identify more efficient treatment options and more accurate subgroup-classifying biomarkers.

To understand the therapeutic response heterogeneity in CCA, prior research efforts focused on characterizing the molecular differences for the different CCA patient subgroups. Exome sequencing was first

[1]Siriraj Center of Research Excellence for Systems Pharmacology, Department of Pharmacology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

[2]Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand

[3]Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

[4]Division of Hematopoiesis, Joint Research Center for Human Retrovirus Infection, Kumamoto University, Kumamoto 860-0811, Japan

[5]These authors contributed equally

[6]Lead contact

*Correspondence: somponnat.sam@mahidol.edu

https://doi.org/10.1016/j.isci.2022.105182

applied to investigate the mutational landscape of liver fluke-associated CCA (Ong et al., 2012). The study found that *TP53*, *SMAD4*, and *MLL3* mutations were common in liver fluke-associated CCA, whereas the mutations of *IDH1*, *IDH2*, and *BAP1* were only observed in the non-liver fluke-associated CCA. Mutations of *ARID1A*, *RNF43*, and *KRAS* can be observed in both subgroups. A comparative follow-up study of CCA tissue specimens identified novel CCA driver genes enriched in the liver-fluke associated CCA, including *ERBB2* amplification and *BRCA1/2* mutations. In contrast, Caucasian CCA patients harbored recurrent mutations of *IDH1/2* and *BAP1* or the fusion of *FGFR* (Jusakul et al., 2017). The Thailand Initiative in Genomics and Expression Research for Liver Cancer (TIGER-LC) consortium later performed an in-depth molecular comparison of the Asian intrahepatic cholangiocarcinoma (iCCA) and hepatocellular carcinoma (Chaisaingmongkol et al., 2017). In addition to revealing that Asian iCCA and HCC shared common mutations, including *TP53*, *ARID1A*, and *ARID2*, the study identified *PLK1* and *ETT2* as important subtype-specific biomarkers of iCCA patients shown to be predictive of poorer prognosis (Chaisaingmongkol et al., 2017). These prior studies brought about profound knowledge of CCA etiology. They contributed to the early clinical trials to assess the potential use of novel targeted therapies such as Selumetinib, MK2206, Lapatinib, and Sorafenib in advanced CCA patients (Ahn and Bekaii-Saab, 2017). Although some therapeutic agents demonstrated anti-tumor activity in a small patient subset, their overall clinical outcomes were somewhat disappointing, offering the progression-free survival of 1–4 months and an overall response rate of ~10%, a similar overall outcome to chemotherapeutic agents. Such unexpectedly poor clinical outcomes arose mainly because of the non-selective enrollment of these prior clinical studies. More recent clinical trials began to adopt biomarker-led designs (Pellino et al., 2018). Still, assessing all subgroup-classifying biomarkers takes a long time and significant financial investment using the conventional clinical studies-based exploration.

Although the genomic uniqueness of the CCA subgroup has been demonstrated in different studies, there has never been a comprehensive effort to identify novel therapeutic options for Thai CCA patients. Toward this goal, we assembled a panel of 15 CCA cell lines of Asian ethnic groups, consisting of CCA cells derived from Thai patients (from *O. viverrini* endemic areas) and Japanese patients. We used these cell lines for our comprehensive drug response profiling and pan-omic assays for subgroup analysis and biomarker discovery. In addition to successfully identifying two major drug response-based subgroups, we developed shortlists of potential anti-cancer drugs accompanied by the corresponding subgroup-classifying biomarkers. We hope this study will promote further investigation towards developing more effective drugs for CCA patients, making our data publicly available for academic use at https://sisp.shinyapps.io/AsianCCAbrowser/.

RESULTS

The genetic alterations and transcriptomic profiles of our Asian CCA cell line panel

To identify novel drug candidates for Asian CCA, we assembled 15 CCA cell lines and one cholangiocyte cell line for a comprehensive drug response profiling and biomarker discovery using mutation analysis and basal transcriptomics (Figure 1A). Our cell line panel consists of 9 Thai and 6 Japanese cell lines. Specifically, KKU-100, KKU-213A, KKU-213B, and KKU-213C are confirmed to be *O. viverrini*-associated cell lines with the proven observation of *O. viverrini* in their original tissue (Sripa et al., 2005, 2020). These cell lines were derived from 3 different tissue origins: intrahepatic CCA (iCCA: KKU-055, KKU-213A, KKU-213B, KKU-213C, HuCCA-1, HuCCT-1, HuH-28, SSP-25, RBE, and YSCCC), perihilar CCA (pCCA: KKU-100) and extrahepatic CCA (eCCA: TFK-1) (Figure 1B). Three CCA cell lines (KKK-D068, KKK-D131, and KKK-D138) were newly established from the patient-derived xenograft of iCCA (Vaeteewoottacharn et al., 2019). KKU-213A, KKU-213B, and KKU-213C served as important intra-patient heterogeneity controls in our study, derived from different locations of a Thai iCCA patient (Sripa et al., 2020). An immortalized cholangiocyte cell line, MMNK-1, was used as our normal cholangiocyte reference.

Using a comprehensive panel sequencing of 548 cancer-related genes, we first compared the genomic variants of our CCA cell lines. Mutated genes are associated with key biological functions, namely chromatin remodeling: *KMT2C* or *MLL3* (100%), *ARID1A* (33%), *SMARCA4* (27%), *BAP1* (13%) and *PBRM1* (7%); MAPK signaling: *KRAS* (40%), *ERBB2* (27%), *MAP2K2* (27%); Notch signaling *NOTCH2* (87%), *NOTCH3*(20%), *NCOR1* (20%), genome instability *TP53* (80%); WNT signaling *PEG3* (60%); TGF-β signaling *SMAD4* (27%); DNA repair *BLM* (27%); Oxidative stress response *KEAP1* (20%) and PI3K signaling *PTEN* (27%) (Figure 1B). Similar to the previous reports (Dong et al., 2018; Saha et al., 2016), we confirmed that RBE carries both *IDH1* mutation (R132S) and *KRAS* mutation (G12V) whereas KKU-213A and HuCCT-1 carry only *KRAS*

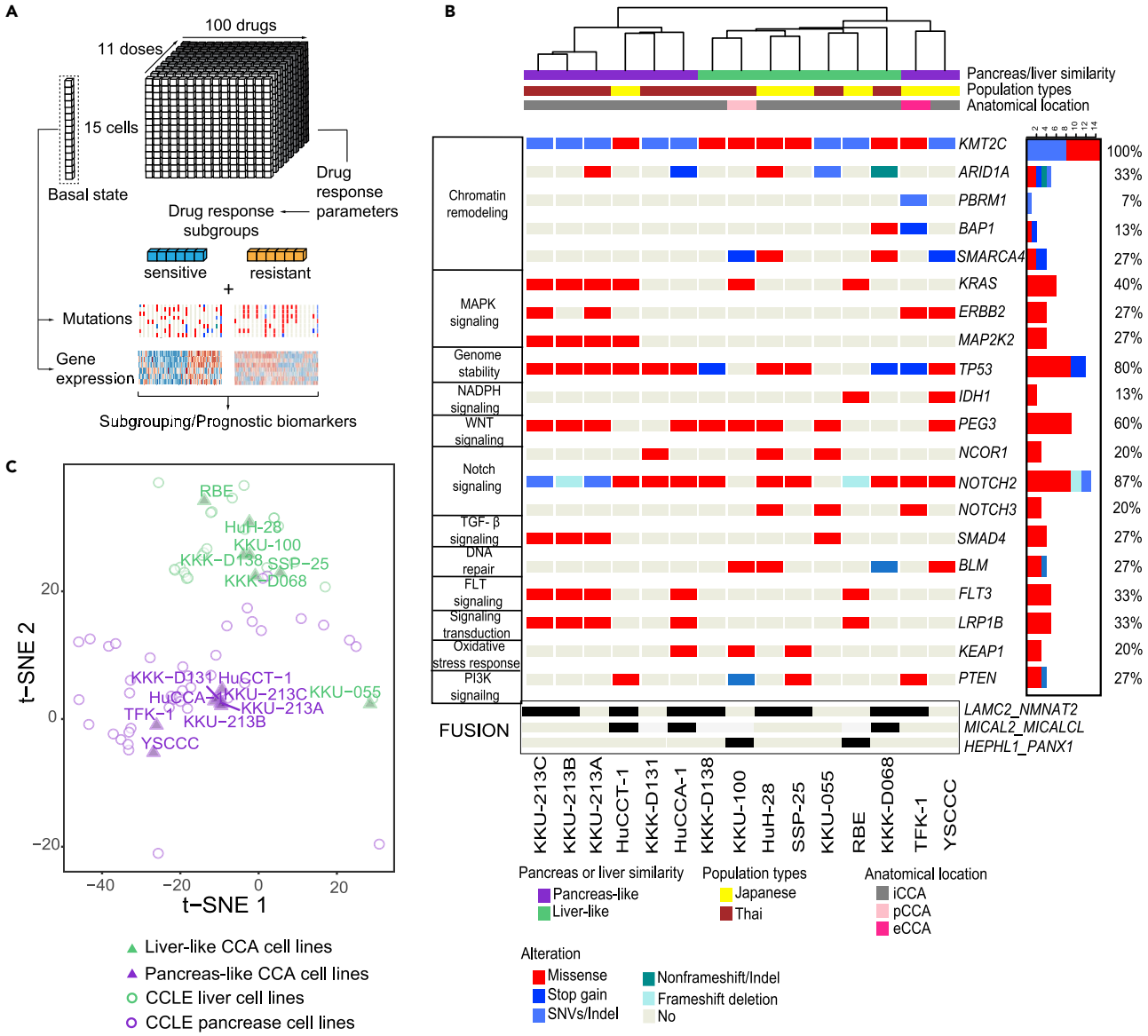


Figure 1. Building a library of pan-omic analysis and comprehensive drug response profiles of Asian CCA cell lines

(A) Overview of the pan-omic model in cholangiocarcinoma (CCA) cell lines. 9 Thai CCA cell lines and 6 Japanese CCA cell lines were profiled for comprehensive drug response against 100 anti-cancer compounds. Genetic alternation analysis and basal transcriptomics were performed using 548-gene cancer panel sequencing and RNA sequencing. Association of drug response and pan-omic data were then analyzed to develop predictive and subgroup-classifying biomarkers.

(B) Common genetic alterations and fusion genes in CCA cell lines. The genomic alterations were grouped according to their biological functions, showing their frequency as the percentage of cell lines with each specific alternation. The top-three fusion genes were also shown in the bottom panel. See also Table S2 for the complete list of fusion genes. Cell lines were clustered based on the similarity of genomic alternations and annotated by their ethnic origins (Japanese or Thai), anatomical locations (intrahepatic: iCCA, perihilar: pCCA, and extrahepatic: eCCA), and their similarity to the pancreas (pancreas-like) or liver (liver-like) transcriptomic profiles.

(C) The t-SNE plot of basal gene expression profiles of 15 CCA cell lines (filled triangles) compared with the transcriptomic profiles of pancreatic and liver cancer cell lines in the CCLE database (unfilled circles). Cell lines with transcriptome profiles similar to the liver and pancreatic cancers are labeled in green and purple, respectively.

mutation (KKU-213A: G13C, HuCCT-1: G12D). Overall, the observed mutational profiles in our CCA cell lines were consistent with those from clinical specimens, highlighting the presence of *TP53*, *KMT2C*, *ARID1A*, *and KRAS* mutations as common genomic variants in both CCA and HCC (Chaisaingmongkol et al., 2017; Chan-On et al., 2013; Jusakul et al., 2017; Ong et al., 2012).

We next profiled the basal transcriptomics of our CCA cell lines and compared them against the 934 cell lines across different cancer types in the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012). We found that our CCA cell lines shared transcriptomic characteristics with those from pancreas and liver origins. Specifically, KKU-213A, KKU-213B, KKU-213C, KKK-D131, HuCCT-1, HuCCA-1, TFK-1, and YSCCC (purple triangles) showed high similarity to pancreatic cancer cell lines (purple circles) (Figure 1B&C and Figure S1). In contrast, KKK-D068, KKK-D138, KKU-100, HuH28, KKU-055, RBE, and SSP-25 (green triangles) were more similar to the liver ones (green circles) (Figures 1B&C and Figure S1). See also Table S1 for the molecular subtypes of CCA. This finding is consistent with prior reports showing that CCA could exhibit characteristics of either hepatocellular carcinoma (Chaisaingmongkol et al., 2017; Farshidfar et al., 2017) or pancreatic cancer (Ong et al., 2012). Of interest, a large cluster of the pancreas-like cell lines (KKU-213A, KKU-213B, KKU-213C, KKK-D131, and HuCCA-1) were from Thai patients (except HuCCT-1). This observation agrees with the prior observation that cancer tissues from Thai CCA patients demonstrate similar pathological and molecular profiles to those of pancreatic ductal carcinoma (Pairojkul, 2014).

We also performed gene fusion analysis of our cell lines using their transcriptomic sequencing results. Overall, we confirmed the concordance of our gene fusion profiles with the previously reported cancer-associated transcript fusions (Yoshihara et al., 2015) (gene fusion panel in Figure 1B and Table S2). The top-three fusion genes in our CCA cell lines include *LAMC2-NMNAT2*, *MICAL2-MICALCL*, and *HEPLH1-PANX1*. The FGFR fusion genes, commonly observed in Caucasian CCA tissue (Farshidfar et al., 2017), were not observed in any Asian CCA cell lines. The above results demonstrated that our CCA cell line collection represents genomic characteristics and disease diversity similar to those surveyed from prior clinical reports in Asian CCA patients.

Building comprehensive drug response profiles for Asian CCA

To determine potential new therapeutic agents for Asian CCA, we applied a high-throughput drug screening of recently developed anti-cancer drugs against cell lines in our Asian CCA cell line panel. Our compound library comprises 17 conventional chemotherapies and 83 targeted therapies, covering receptor tyrosine kinases, intracellular kinases, epigenetic modulators, and other novel drug targets such as proteasome inhibitors and PARP inhibitors. Because our CCA cell lines exhibited differential context-dependent growth rates (Figure S2A), we quantified the drug efficacy and potency using the newly developed growth rate inhibition matrices, GR_{max} and GR_{50} (Hafner et al., 2016). The measurement of GR_{max} helped us to compare drug efficacy across cell lines, separating cytostatic ($GR_{max} > 0$) from cytotoxic ones ($GR_{max} < 0$). While almost all conventional chemotherapies (blue annotation) exhibited cytotoxic effects as expected, targeted therapies (black annotation) showed broader drug efficacy (Figure 2A). It is also important to note the independent relationship between GR_{50} and GR_{max} across all CCA cell lines (Figure 2A; compare the ordering of GR_{max} in green against GR_{50} in orange). Our result confirmed the similar observation in breast cancer cell lines that drug potency and drug efficacy provide non-redundant pharmacological responses (Hafner et al., 2019). We then compared the level of drug efficacy against the uniformity of drug potency across cell lines (as quantified by IQR of drug potency $\log_{10}(GR_{50})$) using standard-of-care treatment as our reference (namely Cisplatin, Gemcitabine, and 5-FU). For example, MEK inhibitors GR_{50} (TAK733, PD0325901, and Selumetinib; purple circles) exhibited a more heterogeneous drug response across cell lines in comparison with EGFR inhibitors (Gefitinib, Lapatinib, and Afatinib; orange circles) that showed more uniform drug potency (Figure 2B). We also observe differential drug response characteristics between the different drug options within the same drug class. Afatinib, an irreversible EGFR inhibitor, showed a more cytotoxic effect than its reversible counterparts such as Gefitinib, Lapatinib, and Erlotinib (Figure S2B, left panel). Likewise, different CDK4/6 inhibitors exhibited differing drug potency and efficacy (Figure S2B, right panel), reflecting their unique polypharmacological profiles (Hafner et al., 2019). These observations highlight the extent and applicability of our comprehensive drug response database. All previously explained datasets are available for further academic use at https://sisp.shinyapps.io/AsianCCAbrowser/.

An in-depth investigation of drug response similarity reveals unique CCA subgroups and their corresponding therapeutic candidates

We next attempted to classify subgroups of CCA cell lines based on the similarity of drug response among all therapeutic agents. Out of the 100 compounds tested, 77 drugs show measurable drug potency across the different CCA cell lines (excluding drugs without responses in any CCA cell lines). Using the unsupervised hierarchical clustering (Figure S3A; only drug pairs with |Spearman's correlation score| > 0.6 are used

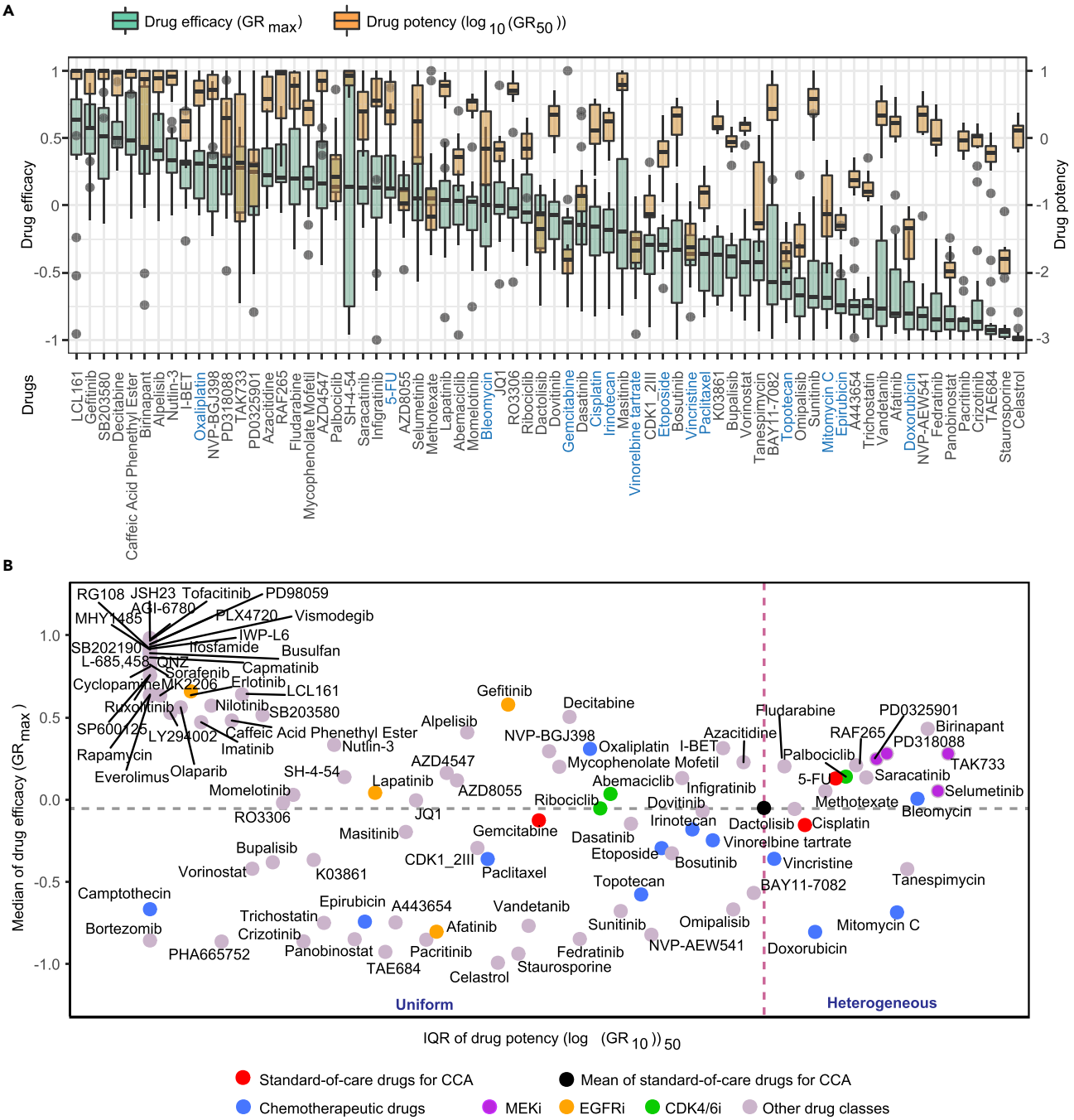


Figure 2. Comparison of growth rate inhibition metrics across all drug candidates

(A) Comparison of drug response parameters (growth rate-normalized drug potency [$\log_{10}(GR_{50})$] and efficacy [GR_{max}]) for the 41 compounds (chemotherapeutic agents in blue and targeted therapeutic agents in black) whose IQR is greater than 0.35 across all CCA cell lines. The distribution of drug potency parameters was ranked based on the averaged GR_{max}. Boxes range from the first to the third quartile, while bold lines mark medians. Upper and lower points mark the position of outliers.

(B) Anti-cancer drug response classification using growth rate inhibition metrics (GR). In addition to their cytostatic ($GR_{max} > 0$) and cytotoxic status ($GR_{max} < 0$), candidate drugs were classified based on the IQR of $\log_{10}(GR_{50})$ as exhibiting heterogeneous (e.g., MEKi, purple dots) or uniform (e.g., EGFRi, yellow dots) drug potency. The mean IQR of standard-of-care chemotherapeutic agents (Cisplatin, Gemcitabine, and 5-FU shown in red dots) was used as the cutoff reference (black dot). Some drug classes, such as CDK4/6i, exhibited both heterogeneous (Palbociclib) and uniform (Ribociclib, Abemaciclib) drug potency.

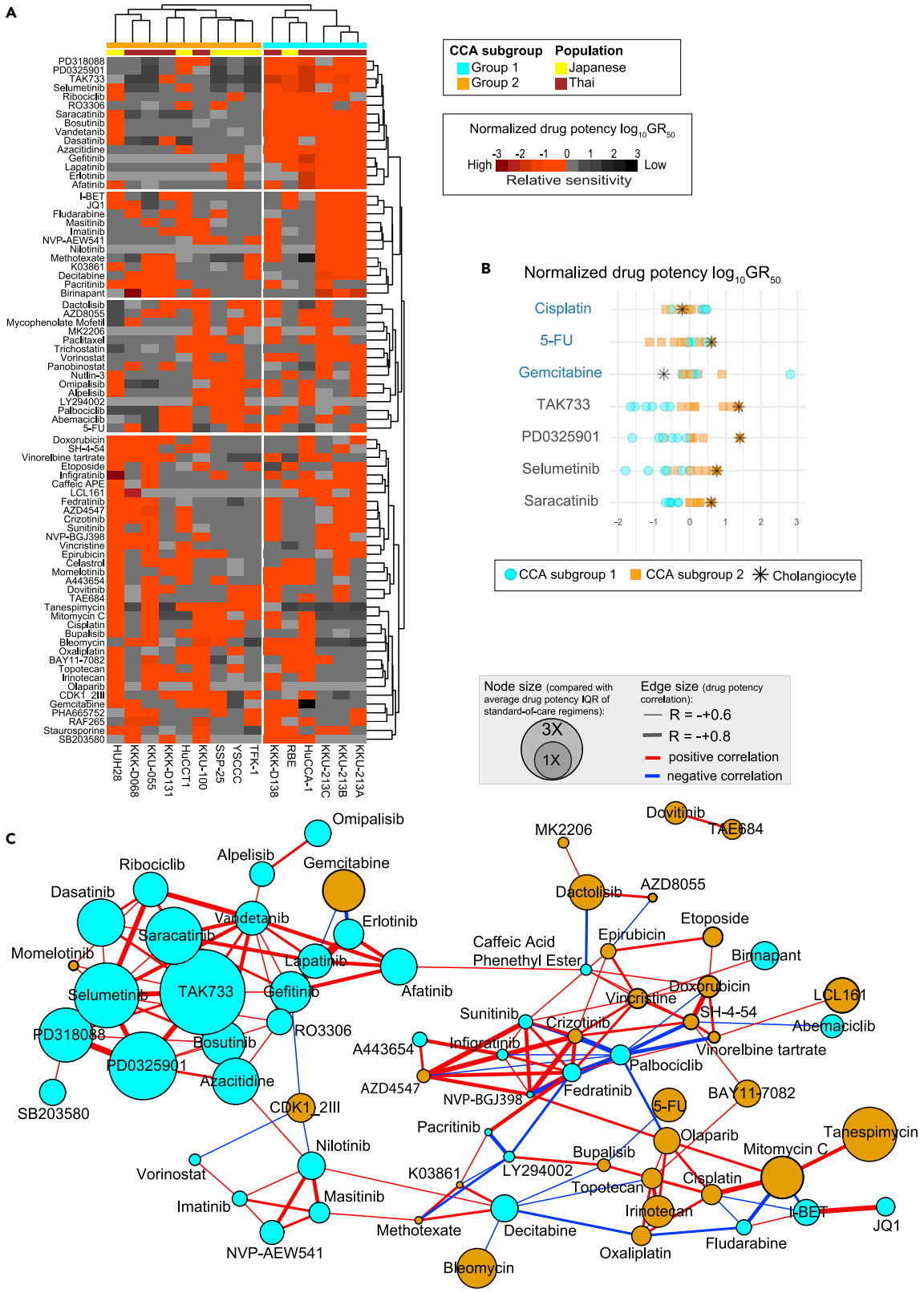

A
CCA subgroup
Group 1
Group 2
Population
Japanese
Thai
Normalized drug potency log10GR50
High
Low
Relative sensitivity
PD318088
PD0325901
TAK733
Selumetinib
Ribociclib
RO3306
Saracatinib
Bosutinib
Vandetanib
Dasatinib
Azacitidine
Gefitinib
Lapatinib
Erlotinib
Afatinib
I-BET
JQ1
Fludarabine
Masitinib
Imatinib
NVP-AEW541
Nilotinib
Methotexate
K03861
Decitabine
Pacritinib
Birinapant
Dactolisib
AZD8055
Mycophenolate Mofetil
MK2206
Paclitaxel
Trichostatin
Vorinostat
Panobinostat
Nutlin-3
Omipalisib
Alpelisib
LY294002
Palbociclib
Abemaciclib
5-FU
Doxorubicin
SH-4-54
Vinorelbine tartrate
Etoposide
Infigratinib
Caffeic APE
LCL161
Fedratinib
AZD4547
Crizotinib
Sunitinib
NVP-BGJ398
Vincristine
Epirubicin
Celastrol
Momelotinib
A443654
Dovitinib
TAE684
Tanespimycin
Mitomycin C
Cisplatin
Bupalisib
Bleomycin
Oxaliplatin
BAY11-7082
Topotecan
Irinotecan
Olaparib
CDK1_2III
Gemcitabine
PHA665752
RAF265
Staurosporine
SB203580
HUH28
KKK-D068
KKU-055
KKK-D131
HuCCT1
KKU-100
SSP-25
YSCCC
TFK-1
KKK-D138
RBE
HuCCA-1
KKU-213C
KKU-213B
KKU-213A
B
Normalized drug potency log10GR50
Cisplatin
5-FU
Gemcitabine
TAK733
PD0325901
Selumetinib
Saracatinib
-2
-1
0
1
2
3
CCA subgroup 1
CCA subgroup 2
Cholangiocyte
Node size (compared with average drug potency IQR of standard-of-care regimens):
3X
1X
Edge size (drug potency correlation):
R = -+0.6
R = -+0.8
positive correlation
negative correlation
C
Omipalisib
Alpelisib
Gemcitabine
Ribociclib
Dasatinib
Vandetanib
Erlotinib
Saracatinib
Momelotinib
Lapatinib
Afatinib
Selumetinib
TAK733
Gefitinib
RO3306
PD318088
Bosutinib
PD0325901
Azacitidine
SB203580
CDK1_2III
Nilotinib
Vorinostat
Masitinib
Imatinib
NVP-AEW541
Dovitinib
TAE684
MK2206
Dactolisib
AZD8055
Etoposide
Epirubicin
Caffeic Acid Phenethyl Ester
Birinapant
Doxorubicin
Vincristine
LCL161
SH-4-54
Sunitinib
Crizotinib
Abemaciclib
Infigratinib
Vinorelbine tartrate
A443654
Palbociclib
AZD4547
Fedratinib
BAY11-7082
NVP-BGJ398
5-FU
Pacritinib
Olaparib
Tanespimycin
Mitomycin C
Bupalisib
K03861
LY294002
Topotecan
Cisplatin
Irinotecan
I-BET
JQ1
Methotexate
Decitabine
Fludarabine
Oxaliplatin
Bleomycin

Figure 3. Subgroup classification of CCA cell lines based on their comprehensive drug response profiles

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.isci.2022.105182.

(A) Heatmap showing normalized drug potency (median-centered $\log_{10}(GR_{50})$) of the 15 CCA cell lines against different anti-cancer drugs. Cell lines and drug candidates were clustered based on Spearman's rank-order correlation, showing only 68 drugs with an absolute value of correlation score greater than 0.6 ($|R| > 0.6$) (see also Figure S3A for complete information). Cell lines that exhibit the normalized drug potency below the group median were illustrated in increasing red-colored intensities to highlight their relatively more sensitive drug response (increasing black-colored intensities depicts the relatively more resistant cell lines). The two major CCA clusters were annotated in cyan (subgroup 1) and orange (subgroup 2), each marked by its ethnic origin (Thai in brown and Japanese in yellow).

(B) Comparison of normalized drug potency between CCA1 and CCA2 cell lines of example drug candidates (standard-of-care chemo drugs in blue and targeted therapeutic drugs in black). CCA1 and CC2 cell lines were highlighted in cyan and orange, respectively. Drug potency of the immortalized human cholangiocyte cell line (MMNK-1) in asterisks was shown for comparison.

(C) Network diagram showing drug response similarity and CCA subgroup specificity among 68 drug candidates. Drugs that offer specificity to either CCA1 or CCA2 subgroups are illustrated in cyan or orange circles, respectively. The node size represents the subgroup specificity score defined by the normalized subgroup distance away from the population drug potency (see Figure S3C for the complete details). The size of each edge represents Spearman's rank-order correlation coefficient (r) between the connected drug pair, with its color representing either negative (blue) or positive (red) correlation pattern.

for further analysis), we could divide these cell lines into two major subgroups, namely the CCA drug-response subgroup 1 (CCA1) and subgroup 2 (CCA2) (Figure 3A). The CCA1 subgroup consisted of 5 Thai CCA (KKU-213A, KKU-213B, KKU-213C, HuCCA-1, KKK-D138) and 1 Japanese CCA (RBE) cell lines (Figure 3A, cyan annotation), sharing high sensitivity to most pro-growth inhibitors such as MEK inhibitors, Src-abl inhibitors, EGFR inhibitors, and some CDK inhibitors. The CCA2 subgroup consists of 4 Thai CCA (KKU-100, KKK-D131, KKU-055, KKK-D068) and 5 Japanese CCA cell lines (TFK-1, YSCCC, SSP-25, HuCCT-1, HuH-28) (Figure 3A, orange annotation), showing poorer sensitivity to the previously mentioned inhibitors but stronger sensitivity to most chemotherapeutic agents. Standard-of-care chemotherapies (e.g., Cisplatin, Gemcitabine, and 5-FU) showed small drug potency differences between the two CCA subgroups (less than 3-fold in mean GR50). For Gemcitabine and Cisplatin, the drug potency of the MMNK-1 cholangiocyte cell line was indistinguishable from those of the two CCA subgroups (Figure 3B, compare black cross marks to the orange and cyan data points). On the other hand, MEK and Src inhibitors (e.g., TAK733, PD0329501, Selumetinib, and Saracatinib) showed wide separation between the two CCA subgroups and limited toxicity to the MMNK-1 cell line (Figure 3B).

We next attempted to understand better the relationship between drug response similarity and their preferential response towards either CCA subgroups. We systematically illustrated these using a force-directed network graph based on the Fruchterman-Reingold algorithm (Fruchterman and Reingold, 1991) (Figure 3C). In particular, the network edge represents the pairwise correlation of drug potency (red and blue edges implicate positive and negative correlation, respectively, whereas the edge thickness shows the magnitude of the correlation score). In addition to observing expected drug response relationships, such as the positive correlation of Selumetinib and TAK733 (Pearson's R of 0.85; Figure S3B, left panel), we also identified novel drug response relationships. The negative correlation of drug response between Erlotinib and Gemcitabine was unexpected because of their unknown shared molecular mechanisms of action (Pearson's R of −0.75; Figure S3B, right panel). To score subgroup specificity, we compared the drug potency of each CCA subgroup to that of the population mean, normalizing this distance by the mean of the population drug-potency IQR from the three standard-of-care chemotherapies (Figure S3C). The magnitude of such a score was then used to illustrate the node size in the network diagram. Based on the previously described approach, we identified 38 drugs that showed specificity toward CCA1 cell lines (cyan nodes; here onwards referred to as CCA1-specific drugs), whereas the other 30 drugs showed specificity toward the CCA2 subgroup (orange nodes; or CCA2-specific drugs) (Figure 3C). The CCA1-specific drugs include MEK inhibitors (TAK733, PD0325901, Selumetinib, PD318088), Src-Abl inhibitors (Saracatinib, Dasatinib, Bosutinib), EGFR inhibitors (Gefitinib, Afatinib, and Lapatinib), CDK4/6 inhibitors (Ribociclib, Palbociclib, and Abemaciclib), and one epigenetic modulator (Azacitidine). The CCA2-specific drugs include thirteen conventional chemotherapeutic agents with a few targeted therapeutics, namely HSP90 inhibitor (Tanespimycin), CDK1/2 inhibitor (CDK1/2III), and PARP inhibitor (Olaparib).

Because significant drug potency does not necessarily imply satisfactory drug efficacy, we investigated whether the previously identified subgroup-specific drug candidates exhibit acceptable drug efficacy in their preferred CCA subgroups. Specifically, we quantified the correlation of GR_{max} and GR_{50} for each identified subgroup-specific drug candidate (Figure 4). To our surprise, we found that the majority of CCA1-specific drugs showed a strong correlation between drug potency and drug efficacy (Figures 4A and 4B, upper panels; red and blue bubbles infer positive and negative correlation scores, respectively, and drugs

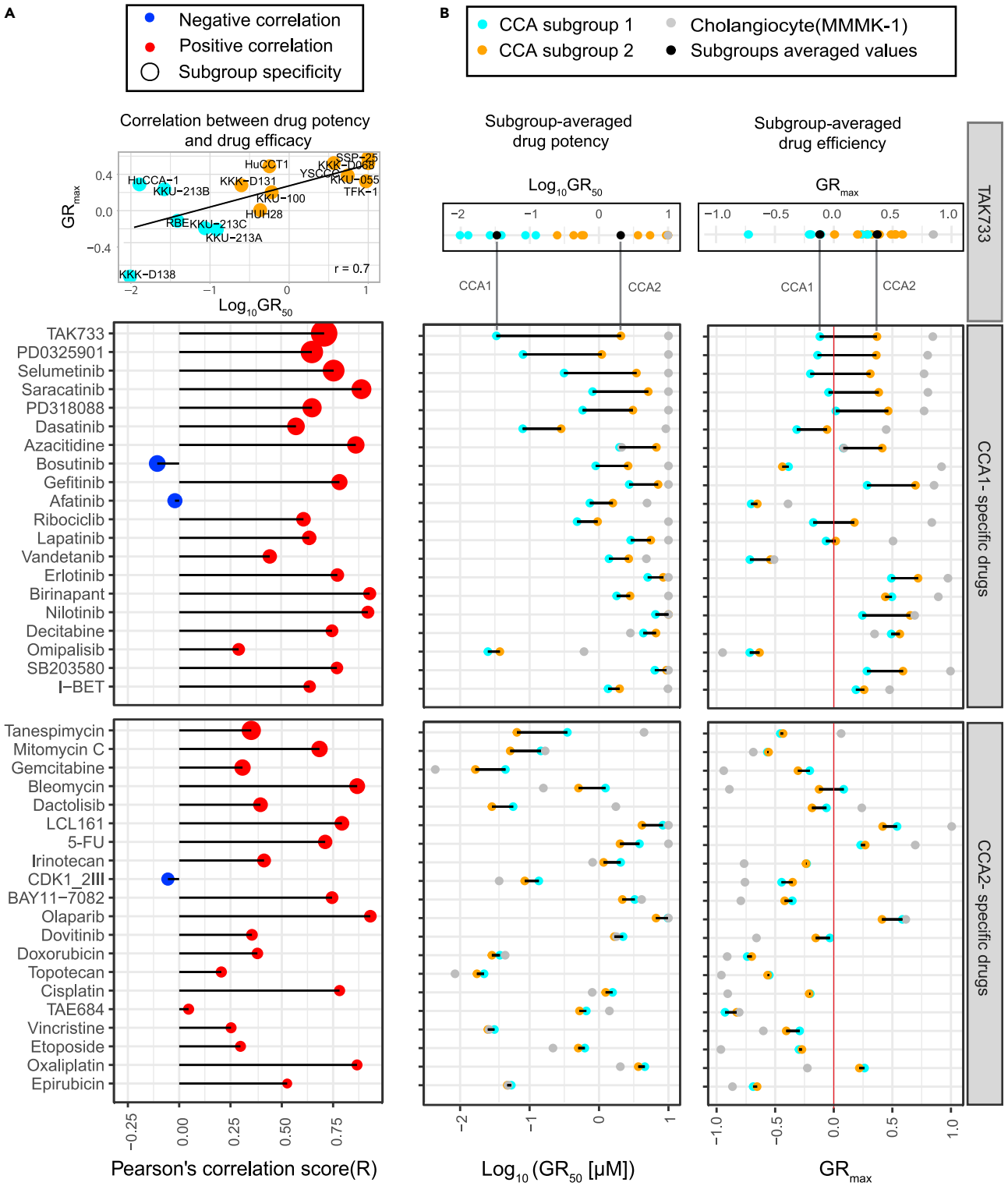


Figure 4. Correlation between drug efficacy and drug potency of subgroup-specific drug candidates

(A) Correlation between drug efficacy (GR_{max}) and drug potency ($log_{10}(GR_{50})$) for both CCA1-specific (middle panel) and CCA2-specific (bottom panel) drug candidates. Drug candidates for each CCA subgroup are ranked based on the absolute values of Pearson's correlation scores. The top panel shows an example positive correlation between drug efficacy and drug potency from TAK733.

Figure 4. *Continued*

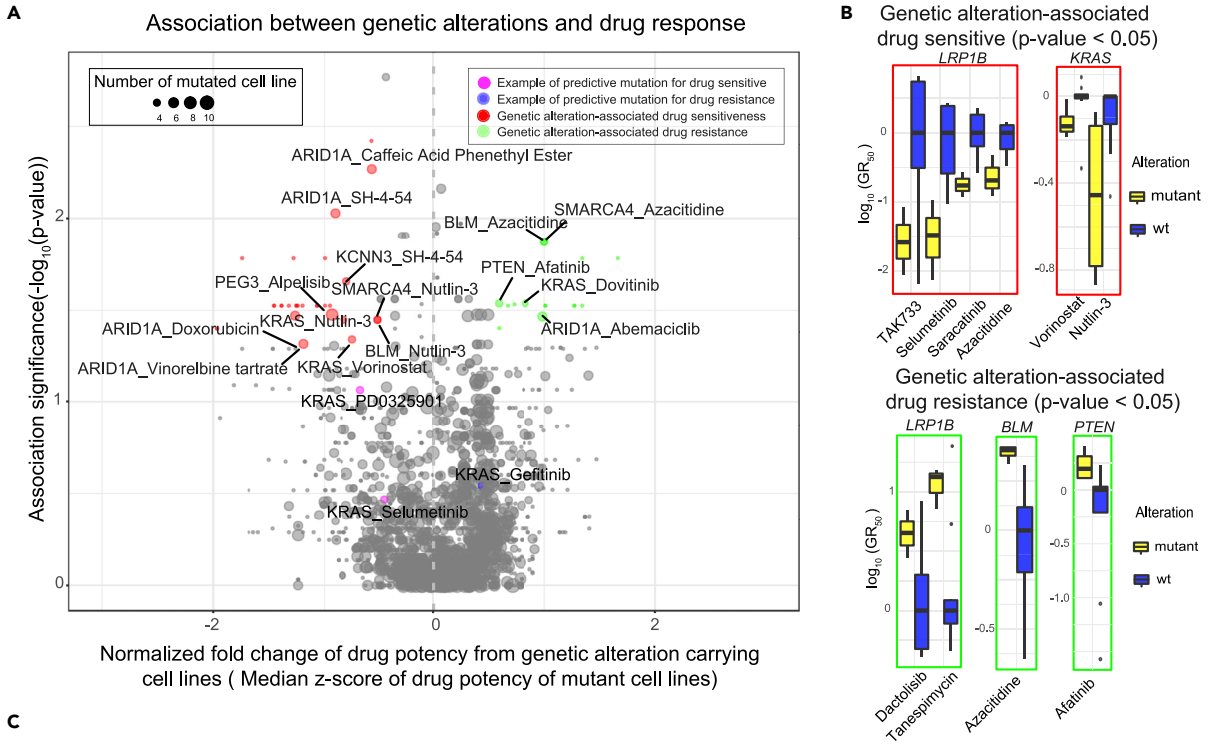
(B) Comparison of drug potency (left panel) and drug efficacy (right panel) between the two CCA subgroups. The mean values of these parameters from either CCA1- or CCA2-subgroup cell lines are shown in either cyan or orange dots, respectively. Gray dots represent the equivalent values for the immortalized human cholangiocyte cell line (MMNK-1).

are ranked by the subgroup specificity score which is shown by the bubble size). MEK inhibitors (TAK733, PD0325901, and Selumetinib) as the top-three CCA1-specific drug candidates showed clear separation of drug potency ($\log_{10}(GR_{50})$) between subgroups. They also exhibited a higher cytotoxic effect against CCA1 cell lines (mean $GR_{max} \leq 0$ for CCA1 vs. mean $GR_{max} \sim 0.5$ for CCA2) (Figure 4B). The other CCA1-specific drugs that showed a strong correlation between drug potency and drug efficacy with selective cytotoxic effects against CCA1 cell lines include Saracatinib, Dasatinib, Ribociclib, and Lapatinib. Of interest, CCA1-specific drug candidates also exhibited negligible pharmacological effects against the MMNK1 normal cholangiocyte cell line, implying reasonable therapeutic indices. In contrast, CCA2-specific drugs showed a weaker correlation between drug efficacy and drug potency. Few CCA2-specific drugs showed robust cytotoxicity against CCA2 cell lines (Figures 4A and 4B; lower panels). The top-5 CCA2-specific drugs (Tanespimycin, Mitomycin C, Gemcitabine, Bleomycin, and Dactolisib) showed narrow separations between the two CCA subgroups and significant cytotoxicity against the MMNK1 cholangiocyte cell line (Figures 4A and 4B lower panels). Tanespimycin showed the highest specificity against CCA2 cell lines with negligible toxicity against the MMNK1 cell line. However, both CCA subgroups showed indistinguishable cytotoxic effects (GR_{max} below 0 for both CCA subgroups) (Figures 4A and 4B lower panels). These results indicated that our novel analytical approach enables systematic identification of potentially new drug candidates for Asian CCA. Our comprehensive analysis of drug response, using both drug potency and drug efficacy and the newly developed subgroup specificity score, was able to identify selective therapeutic candidates for both CCA subgroups.

Identification of drug response-associated mutations

We were next interested in developing biomarkers that can be used to classify CCA patient subgroups in the clinical setting. We reasoned that our drug response-based CCA subgroups, if truthful, must also correspond with the molecular machinery that underlies the unique drug response between the two CCA subgroups. We first examined whether the observed drug sensitivity, as quantified by drug potency, would be associated with any mutations found across all CCA cell lines (Figure 5A). All chosen drug response-associated mutations must show more substantial association scores than the previously known mutations, such as the associations of *KRAS* mutations to MEK inhibitor sensitivity (Figure 5A, pink circles) or Gefitinib resistance (Figure 5A, blue circles) (Figure 5A) (Iorio et al., 2016). Among all mutations from the 189 genes that passed our filtering criteria (see more details in the STAR methods session), the mutations of only 15 genes showed a significant association with drug response ($|\text{median Z-score of } \log_{10}(GR_{50})| > 0.5$ with p-value < 0.05 using Wilcoxon test; Figure 5A). These results were consistent with another parallel approach where we scored the significance of drug response-associated mutations based on the difference in drug potency between the wild-type and the mutation-carrying cell lines (Figure S4). Although not anticipated, many of the detected drug response-associated mutations in CCA cell lines were reported previously in other cancer types. For example, *KRAS* mutations were shown to be predictive of Nutlin-3 sensitivity (*KRAS* G12V) in ovarian carcinoma and Vorinostat sensitivity (*KRAS* G12D) in lung cancer (Crane et al., 2015; Ma et al., 2013). Likewise, *PTEN* mutations were predictive of resistance to EGFR inhibitors in non-small cell lung cancer (Gkountakos et al., 2019). On the other hand, some drug response-associated mutations were novel to our study. The tumor suppressor gene *LRP1B* was found to be associated with sensitivity to several kinase inhibitors (including TAK733, Selumetinib, Saracatinib) and an epigenetic inhibitor (Azacitidine) as well as with resistance to a PI3K/mTOR inhibitor (Dactolisib) and an Hsp90 molecular chaperone inhibitor (Tanespimycin) (Figure 5B). This result highlights our approach's uniqueness in identifying drug response-associated mutations unbiasedly, both known and novel relationships.

We next asked whether any identified drug response-associated mutations were observed uniquely in either CCA subgroup. In general, we found both subgroup-specific and common drug response-associated mutations. In particular, *LRP1B* and *FLT3* mutations were detected only in CCA1 cell lines (except KKK-D138), whereas *BLM*, *SMARCA4*, *SLC19A1*, *NCOR1*, *NOTCH3*, and *PTEN* were observed only in CCA2 cell lines (Figure 5C). Mutations of *KRAS*, *PEG3*, *ARID1A*, *KCNN3*, *KEAP1*, *MSH6*, and *ERBB2* could be observed in both CCA subgroups. For example, *ARID1A* mutations were found to be associated with both drug sensitivity to Caffeic acid phenethyl ester (NF-κB inhibitor), SH-4-54 (STAT3/5 inhibitor), Doxorubicin (topoisomerase II inhibitor), and Vinorelbine tartrate (mitotic spindle inhibitor) as well as with drug



C

Genetic alteration-associated drug sensitive (p-value < 0.05)

CCA subgroup 1

No. of mutation	Genes	% found in Asian CCA patients	Drugs
3	*FLT3*	NA	Azacitidine Saracatinib Selumetinib TAK733
	LRP1B	8.2%	Azacitidine Saracatinib Selumetinib TAK733

CCA subgroup 2

No. of mutation	Genes	% found in Asian CCA patients	Drugs
4	*BLM*	NA	Nutlin-3
	SMARCA4	4.3%	Nutlin-3
3	*SLC19A1*	NA	Bleomycin
	NCOR1	2.9%	Dovitinib Irinotecan Olaparib Pacritinib
	NOTCH3	NA	Olaparib Mitomycin C Tanespimycin

CCA both subgroups

No. of mutation	Genes	% found in Asian CCA patients	Drugs
4 3	*PEG3*	3.8%	Alpelisib
3 2	*ARID1A*	18.8%	Caffeic Acid Phenethyl Ester Doxorubicin SH-4-54 Vinorelbine tartate
2 2	*KRAS*	16.1%	Nutlin-3 Vorinostat
	KCNN3	NA	SH-4-54
2 1	*KEAP1*	NA	Bleomycin LY294002
	MSH6	1.68%	Palbociclib

Genetic alteration-associated drug resistance (p-value < 0.05)

CCA subgroup 1

No. of mutation	Genes	% found in Asian CCA patients	Drugs
3	*FLT3*	NA	Dactolisib Tanespimycin
	LRP1B	8.2%	Dactolisib Tanespimycin

CCA subgroup 2

No. of mutation	Genes	% found in Asian CCA patients	Drugs
4	*BLM*	NA	Azacitidine
	PTEN	3.8%	Afatinib
	SMARCA4	4.3%	Azacitidine

CCA both groups

No. of mutation	Genes	% found in Asian CCA patients	Drugs
3 2	*ARID1A*	18.8%	Abemaciclib
2 2	*KRAS*	16.1%	Dovitinib
2 1	*ERBB2*	3.6%	Bleomycin Staurosporine
	KEAP1	NA	Pacritinib
	MSH6	1.68%	Infigratinib Crizotinib Vinorelbine tartate

Figure 5. Association of commonly observed mutations and drug response in CCA cell lines

(A) Volcano plot of all identified drug response-mutation association pairs. The significant association pairs represent those with the median Z-score of $\log_{10}(GR_{50}) > 0.5$ and p-value < 0.05 using the Wilcoxon test, with red and green bubbles showing drug sensitivity- and drug resistance-associated mutations. The bubble size represents the number of cell lines that carry such mutations. Previously identified KRAS-associated drug response relationships (drug sensitivity in purple and drug resistance in blue) are shown here as references. See also Figure S4 alternative significance scoring approach.

(B) Comparison of drug potency (GR_{50}) between the wild-type (blue) and the mutation-carrying (yellow) cell lines. Example drug sensitivity-associated mutations include *LRP1B* for TAK733, Selumetinib, Saracatinib, Azacitidine, and *KRAS* for Vorinostat and Nutlin-3. Example drug resistance-associated mutations include *LRP1B* for Dactolisib and Tanespimycin, *BLM* for Azacitidine, and *PTEN* for Afatinib. Boxes range from the first to the third quartile, while bold lines mark medians. Upper and lower points mark the position of outliers.

(C) Unique drug response-associated mutations of different CCA subgroups. The circle size represents the number of cell lines, showing those from CCA subgroup 1 in cyan and CCA subgroup 2 in orange. The Mutations previously reported in the Asian CCA cohort (Jusakul et al., 2017) were also labeled with their prevalence (from total patients of 416).

resistance to Abemaciclib (CDK4/6 inhibitor). We found that 9 of the 15 drug response-associated mutations were genomic alterations reported in prior clinical cohorts of CCA patients (Jusakul et al., 2017) (Figure 5C), implying that these mutations may be clinically translatable.

Comparison of ligand-induced pathway activation between the two CCA subgroups

The ligand-induced activation of cancer-associated signaling cascades is another biomarker set that was shown to be predictive of drug sensitivity (Jones et al., 2017; Niepel et al., 2014). Having hypothesized that the two CCA subgroups must rely on unique signaling cascades, we compared the ligand-induced pathway activity between the two CCA subgroups. Specifically, the phosphorylation of ERK1/2, p70/S6K, AKT, STAT1, and the p65/RelA translocation were quantified using the immunofluorescent staining technique after the stimulation with common pathway-specific ligands including EGF, IGF1, Anisomycin, TNFα, and IFNγ. The ligand-induced protein activity of each representative node was then quantified based on the differential area-under-the-curve (AUC) across the different time points compared to the non-stimulated cells (Figure 6A and S5A). Using the principal component analysis (PCA), we could show that the overall signal activation between the CCA1 and CCA2 cell lines is quantitatively distinguishable (Figure 6B). We then divided these ligand-induced pathway activities into five subgroups using hierarchical clustering (Figure S5B). Of interest, the ligand response of AKT activation in cluster 2 primarily contributed to PC1, whereas the ligand response of ERK activation in cluster 1 contributed to PC2. To further examine the preference for ligand-induced pathway activation, we applied the logistic regression analysis to the signaling node activities. The result showed that PI3K/AKT pathway activation (pAKT) was preferable in CCA2 subgroup cell lines, with a substantial significance score (odds ratio = 3.07×10^{-5}; p-value = 9.71×10^{-3}). On the other hand, activation of other signaling proteins was observed more commonly in CCA1 subgroup cell lines (Figure 6C). Notably, although all ligand types could activate AKT in CCA2 cell lines, only pro-growth ligands (IGF1, Anisomycin, and EGF) could activate ERK in CCA1 cell lines (Figure 6D). Other observed CCA1 subgroup-specific pathway activations included nuclear translocation of RelA in response to IFNγ treatment and phosphorylation of S6K in response to EGF and IGF1 treatment (Figure S5C). These results suggest that AKT activity is more preferentially activated in CCA2 cell lines, regardless of ligand types.

Prior studies have shown that targeting cancer with MAPK-targeting drugs may not be effective in MAPK-hyperactivated cancer because of feedback activation (Kitai et al., 2016; Merchant et al., 2017). To test this hypothesis, we measured changes of ERK phosphorylation in both CCA subgroups (KKU-213C and KKK-D138 cell lines from CCA1 subgroup and KKU-100, and SSP-25 from CCA2 subgroup) after treatment with MEKi PD-0325901 for 3 h at low (1.25 μM) and high (5 μM) doses or vehicle control in medium containing 10% FBS. The results showed that MEK inhibitor treatment significantly reduced ERK phosphorylation in both the KKU-213C (with *KRAS* mutation, with p-value = 0.0066) and KKK-D138 cells (with wild-type *KRAS*, p-value = 0.0022) (Figure S5D). For the CCA2 subgroup, MEK inhibitor successfully suppressed ERK phosphorylation in the KKU-100 cell line more efficiently (with *KRAS* mutation, p value = 0.0019) than in the SSP25 cell line (with wild-type *KRAS*, p value = 0.059) (Figure S5D). These results imply that CCA1 cell lines may be more sensitive to MEK inhibitors perhaps because of their addiction to the MAPK pathway. On the other hand, MAPK pathway in the CCA2 cell lines, even though are partly active, are not the main growth driver.

Previous studies have shown that the ligand-specific pathway activation can strongly depend on the mutational status of cancer cells (Li et al., 2017; Niepel et al., 2014; Sampattavanich et al., 2018). To extend this

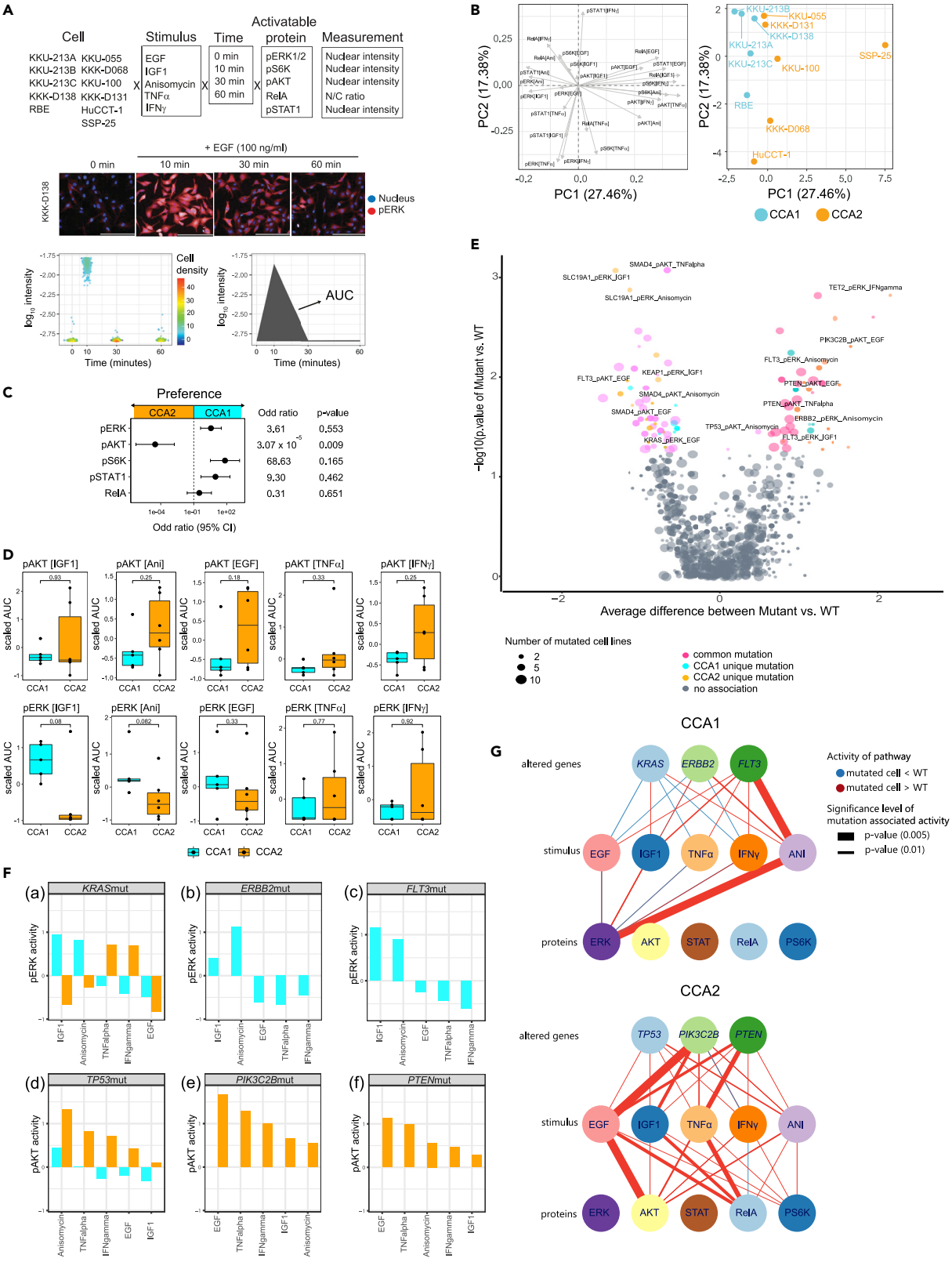

A
Cell
KKU-213A KKU-055
KKU-213B KKK-D068
KKU-213C KKU-100
KKK-D138 KKK-D131
RBE HuCCT-1
SSP-25
Stimulus
EGF
IGF1
Anisomycin
TNFα
IFNγ
Time
0 min
10 min
30 min
60 min
Activatable protein
pERK1/2
pS6K
pAKT
RelA
pSTAT1
Measurement
Nuclear intensity
Nuclear intensity
Nuclear intensity
N/C ratio
Nuclear intensity
+ EGF (100 ng/ml)
0 min
10 min
30 min
60 min
KKK-D138
Nucleus
pERK
Cell density
AUC
log10 intensity
Time (minutes)
B
PC2 (17.38%)
PC1 (27.46%)
KKU-213B
KKU-055
KKK-D131
KKK-D138
KKU-213A
KKU-213C
KKU-100
SSP-25
RBE
KKK-D068
HuCCT-1
CCA1
CCA2
C
Preference
CCA2
CCA1
Odd ratio
p-value
pERK 3.61 0.553
pAKT 3.07 x 10-5 0.009
pS6K 68.63 0.165
pSTAT1 9.30 0.462
RelA 0.31 0.651
Odd ratio (95% CI)
D
pAKT [IGF1]
pAKT [Ani]
pAKT [EGF]
pAKT [TNFα]
pAKT [IFNγ]
pERK [IGF1]
pERK [Ani]
pERK [EGF]
pERK [TNFα]
pERK [IFNγ]
scaled AUC
CCA1
CCA2
E
SMAD4_pAKT_TNFalpha
SLC19A1_pERK_IGF1
SLC19A1_pERK_Anisomycin
TET2_pERK_IFNgamma
PIK3C2B_pAKT_EGF
FLT3_pERK_Anisomycin
FLT3_pAKT_EGF
KEAP1_pERK_IGF1
PTEN_pAKT_EGF
SMAD4_pAKT_Anisomycin
PTEN_pAKT_TNFalpha
SMAD4_pAKT_EGF
ERBB2_pERK_Anisomycin
TP53_pAKT_Anisomycin
FLT3_pERK_IGF1
KRAS_pERK_EGF
−log10(p.value of Mutant vs. WT
Average difference between Mutant vs. WT
Number of mutated cell lines
2
5
10
common mutation
CCA1 unique mutation
CCA2 unique mutation
no association
F
(a) KRASmut
(b) ERBB2mut
(c) FLT3mut
(d) TP53mut
(e) PIK3C2Bmut
(f) PTENmut
pERK activity
pAKT activity
IGF1
Anisomycin
TNFalpha
IFNgamma
EGF
CCA1
CCA2
G
CCA1
altered genes
KRAS
ERBB2
FLT3
stimulus
EGF
IGF1
TNFα
IFNγ
ANI
proteins
ERK
AKT
STAT
RelA
PS6K
Activity of pathway
mutated cell < WT
mutated cell > WT
Significance level of mutation associated activity
p-value (0.005)
p-value (0.01)
CCA2
altered genes
TP53
PIK3C2B
PTEN
stimulus
EGF
IGF1
TNFα
IFNγ
ANI
proteins
ERK
AKT
STAT
RelA
PS6K

Figure 6. Comparison of the ligand-mediated pathway activation between the two CCA subgroups

(A) Experimental workflow for determining the ligand-mediated pathway activation. Cell lines were exposed to 100 ng/mL of EGF, IGF1, Anisomycin, TNFα, and IFNγ for 10, 30, and 60 min. Phosphorylation of ERK, AKT, S6K, and STAT1 and the nuclear translocation of the RelA (the p65 subunit of NF-κB) were measured. Temporal responses were then converted to AUC for quantifying the activity of each pathway.

(B) Principal component analysis of the ligand-mediated activity of pERK, pS6K, pSTAT1, pAKT, and RelA. Loadings of ligand responses (left) and the PC1 and PC2 scores were compared between the CCA1 (cyan) and CCA2 cell lines (orange) (right).

(C) Analysis of the association between CCA subgroups and signal activation. The odds ratio (probability of being CCA1 over the probability of being CCA2) was calculated using the log of the coefficient derived from the generalized linear model used to assess the relationship between CCA subtypes (CCA1 or CCA2) and ligand-mediated activation of key proteins (pERK, pS6K, pSTAT1, pAKT, or RelA). A p-value < 0.05 indicates that significant ligand-activated signaling pathway. Error bars represent the interquartile range of the odds ratio.

(D) Comparison of ligand-specific pathway activity between the two CCA subgroups. Data between two subgroups were compared using the Wilcoxon rank sum test. Boxes range from the first to the third quartile, while lines inside the box mark medians. Upper and lower points mark the position of outliers.

(E) Association of the mutation status and the ligand-mediated pathway activity bias. The average difference of protein activity between the mutated and the wild-type cell lines (x-axis) was plotted against the significant score (y-axis: $-\log_{10}$(p-value) from the Wilcoxon rank sum test; showing only those with p-value < 0.05). The bubble size represents the number of mutated cell lines. Significant mutations observed in both CCA subgroups are displayed in pink, while the CCA1-specific or CCA2-specific mutations are shown in cyan and orange, respectively.

(F) Example mutation-associated pathway activity bias for common (*KRAS* and *TP53*), CCA1-specific (*ERBB2* and *FLT3*), and CCA2-specific (*PIK3C2B* and *PTEN*) mutations. The mutation-associated activity bias for ERK and AKT pathways was shown across all ligands and compared between the two CCA subgroups. See also Figure S5E for similar results of other signaling proteins.

(G) Summary network diagrams of the mutation-associated pathway activity bias of CCA1 and CCA2 subgroups. Nodes represent mutated genes (top panel), ligands (middle panel), and signaling proteins (bottom panel). Each edge represents the identified mutation-ligand-protein association group. The edge color highlights the direction of the activity bias (red: mutant> wild-type vs. blue: mutant< wild-type), and the edge size shows the significance score of the association.

investigation in our work, we analyzed the relationship between mutations and pathway activity in each CCA subgroup (Figure 6E). CCA cell lines from both subgroups with the *KRAS* mutation exhibited less ERK activity than the wild-type cell lines, even after EGF stimulation (Figure 6Fa), indicating that *KRAS* mutation desensitized the ligand-induced ERK activity. Of interest, only the CCA1 cell lines with *KRAS* mutation had elevated ERK activity in the presence of IGF1 or Anisomycin. On the contrary, the CCA2 cell lines with *KRAS* mutation showed elevated ERK activity with TNFα or IFNγ stimulation. These results suggest a possible unique cross-talk relationship between the inflammatory cytokine-related pathway and the ERK signaling pathway in either CCA subgroups. Apart from *KRAS*, other mutations such as *ERBB2* and *FLT3* also affected ERK activity, especially in CCA1 subgroup (Figures 6FB and 6C). In the presence of IGF1 and Anisomycin, CCA1 cell lines with either *ERBB2* or *FLT3* mutations showed elevated ERK activity. Similar mutation-induced pathway activity changes were also seen in CCA2 cell lines. Specifically, mutations of *TP53*, *PIK3C2B*, *KEAP1* and *PTEN* elevated AKT activity across all ligands in CCA2 cell lines, most pronouncedly with EGF and TNFα. At the same time, only minor changes were observed in CCA1 cell lines (Figures 6FD and 6F and S5E).

We constructed network diagrams to gain further insight into the association of mutation and ligand-induced pathway activity between the two CCA subgroups. Three node layers represent different mutated genes, stimulating ligands, and pathway proteins (Figure 6G). The network edges represent the relative activity of the assigned signaling proteins between cell lines that harbor the mutations and the wild-type genes (red edges showed mutated > wild-type whereas blue edges showed mutated < wild-type). *FLT3* mutation showed the highest significance score for CCA1 subgroup, inducing elevated ERK activation with Anisomycin treatment. On the other hand, *PI3KC2B* mutation showed the highest significance score for CCA2 subgroup, elevating AKT activation with EGF treatment. Collectively, our pathway activation analysis showed that the two CCA subgroups exhibit unique ligand-induced pathway activation with a distinct association to the stimulating ligands and the underlying mutations. These results substantiate a context-dependent nature of signaling cascade activity, making it less robust as subgroup-classifying biomarkers.

Development of CCA45 signature for the classification of CCA drug response-based subgroups

We next attempted to investigate the association of basal gene expression levels with our drug response-based CCA subgroups. Gene expression-based biomarkers have been developed by scoring the correlation between drug sensitivity and the expression of individual genes (Rees et al., 2016). However, such approach may suffer from identifying the context-dependent non-causal biomarkers that do not underlie the observed drug response phenotypes. Recently, a novel approach has been proposed to identify truly

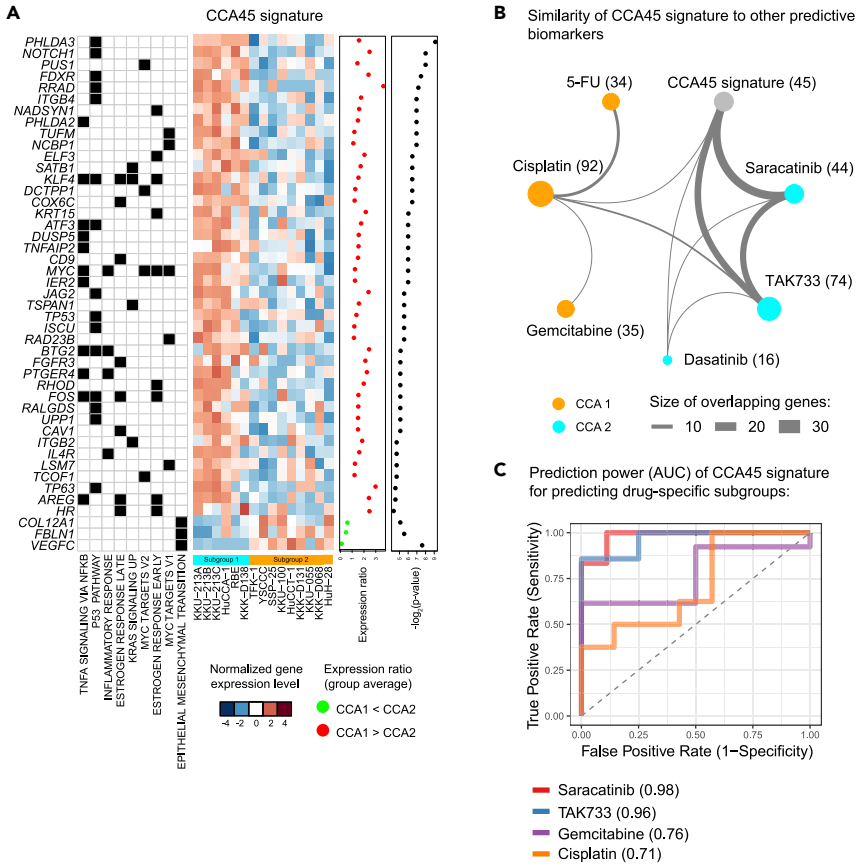


Figure 7. CCA45 signature and its accuracy in predicting drug response

(A) The different gene expression levels of the CCA45 signature across all CCA cell lines. Each of the CCA45 signature genes is annotated with its associated cancer hallmarks, the ratio of the mean expression level between CCA1 and CCA2 subgroups, and the Mann-Whitney U-test significant level. See also Figure S6 for the detailed development of the CCA45 signature.

(B) Network diagram showing the overlapped gene sets between the CCA45 signature (gray node) and the other drug response-predictive biomarkers, both for the CCA1-specific drugs (Saracatinib, TAK733, Dasatinib; cyan nodes), as well as the CCA2-specific drugs (5-FU, Cisplatin, Gemcitabine; orange nodes). The node size represents the number of biomarkers (shown in parentheses), and the edge thickness depicts the number of overlapped genes.

(C) Receiver operating characteristic (ROC) curve for predicting the CCA45 signature using sensitivity to Saracatinib, TAK733, Gemcitabine, and Cisplatin. The area-under-the-curve (AUC) is shown in parenthesis.

causal biomarkers by utilizing pre-established hallmark gene sets (Targonski et al., 2019) and gene set enrichment analysis (GSEA) (Subramanian et al., 2005). Using this approach and the 50 cancer hallmark gene sets from the Molecular Signatures Database (MSigDB) as our reference gene sets (Liberzon et al., 2015; Subramanian et al., 2005), we attempted to identify gene sets capable of classifying our drug response-based CCA subgroups. As a result, we identified nine significantly enriched hallmark gene sets (p-value < 0.05) that represented the underlying mechanisms of drug response in CCA1 subgroup (Figure S6A). The hallmarks consisted of eight positively enriched gene sets (NES >0; TNFα signaling via NF-κB, P53 pathway, inflammatory response, estrogen response late, KRAS signaling up, MYC targets, early estrogen response gene sets) and one negatively enriched gene set (NES <0; epithelial-mesenchymal transition (EMT) gene set).

Next, we attempted to identify gene biomarkers that could classify our CCA subgroups from the pool of leading-edge genes from the nine significantly enriched hallmark gene sets (Figure S6B). A carefully chosen criterion of significance level (ROC AUC >0.8, Mann-Whitney's p value < 0.05) was applied to narrow down the list of candidate biomarkers previously selected by the classification model. As a result, we managed to identify 45 genes from the leading-edge gene list of the enriched hallmarks that could classify CCA cell lines into the drug response subgroups with high accuracy. Specifically, forty-two biomarkers were

upregulated in CCA1 cell lines, and three (solely from the EMT gene set) were upregulated in CCA2 cell lines. Hereafter, we referred to these 45 biomarker gene sets as the CCA45 signature (Figure 7A). In particular, *PHLDA3*, *NOTCH1*, *PUS1*, *FDXR*, and *RRAD* were the highest enriched genes in CCA1 cell lines (Mann-Whitney's p value < 0.05), and most of them (4/5) were associated with the TP53 signaling pathway. On the other hand, *COL12A1*, *VEGFC*, and *FBLN1* were associated solely with the EMT hallmark and are significantly up-regulated in the CCA2 cell lines. Of interest, our finding is consistent with previous reports showing enrichment of TP53-related pathway in most liver fluke-associated CCA patients (Chaisaingmongkol et al., 2017; Jusakul et al., 2017; Ong et al., 2012).

To examine the CCA45 signature specificity in classifying the CCA drug response-based subgroups, we compared its ability to predict drug response subgroup (CCA1 versus CCA2) and organ similarity (liver vs. pancreas) to the performance of using whole transcriptomic data as a predictor. Using the estimated goodness of prediction (q^2) from the partial least squares regression (PLSR) model, we showed that the CCA45 signature predicted CCA1 drug response-based subgroup with high accuracy (q^2 = 0.896, p value < 0.05), whereas the whole transcriptomic data could not (Figure S6C). In addition, the CCA45 signature was restricted to classifying CCA drug response-based subgroups but failed to predict for organ similarity. These results, therefore, confirm the specificity of the CCA45 signature for the classification of the CCA subgroups based on drug response.

In addition to the classification of CCA subgroups, we asked whether the CCA45 signature also holds predictive values for the sensitivity of any subgroup-specific drug candidates. First, we needed to sub-categorize CCA cell lines into drug-sensitive and drug-resistant subgroups for all individual drug candidates (Figure S6D). Specifically, a cut-off drug potency was determined using kernel density estimation for each therapeutic agent to separate cell lines into the two drug response subgroups (see more details in the STAR Methods section and Table S3 for complete results). Next, we applied a similar GSEA-based approach to determine the leading-edge genes for the drug sensitivity prediction of all therapeutic candidates (Table S4). Of interest, we found that the predictive biomarkers for CCA1-specific drugs shared many overlapped biomarkers with the CCA45 signature. Specifically, the predictive biomarkers for MEK inhibitors (Saracatinib and TAK733; blue circles) shared the most overlapped biomarkers with the CCA45 signature, whereas the predictive biomarkers for standard-chemotherapies (Gemcitabine, Cisplatin, and 5-FU; orange circles) showed only a few overlapped genes (Figure 7B). When we attempted to apply the CCA45 signature to predicting drug sensitivity of these key therapeutic candidates, prediction accuracy of the CCA45 signature performed well for Saracatinib and TAK733 (AUC >0.95) but poorly for stand-of-care chemotherapies (AUC <0.8) (Figure 7C). Taken together, our analyses showed that the newly developed CCA45 signature can accurately classify CCA drug response-based subgroups and predict drug sensitivity, especially for the CCA1-specific drug candidates.

CCA45 signature implicates poorer prognosis in Asian CCA patients

The classification of CCA by their molecular etiologies has been explored in prior research work. A critical study by Jusakul and colleagues (Jusakul et al., 2017) utilized a pan-omics approach to study CCA patients' molecular uniqueness from Asian and Caucasian populations. By characterizing mutations, copy number variation, gene expression, and epigenetic modifications, the team identified four unique molecular clusters, highlighting *ERBB2* amplification and *TP53* mutations as common characteristics in liver fluke-positive CCAs. Despite a comprehensive analysis of CCA molecular landscapes, the potential therapeutic options for such four molecular clusters were not characterized. Motivated by such limitation, we attempted to examine whether our findings of subgroup-specific drug candidates and the developed subgrouping biomarker could potentially be translated for clinical practice. First, we evaluated the similarity between our CCA45 signature and the four previously reported CCA molecular clusters. Using the unsupervised subclass mapping technique (SubMap) (Hoshida et al., 2007), we observed a high concordance between our CCA1 subgroup and the CCA molecular cluster type I (Bonferroni-adjusted p-value = 7.9×10^{-3}, Figure S7A). We then attempted to classify CCA patients from clinical cohort (Jusakul et al., 2017) into our drug response-based subgroups using the CCA45 signature. Specifically, we combined the normalized gene expression data from CCA patients and our CCA cell line panel and performed the unsupervised hierarchical clustering. The result revealed that CCA patients could be grouped with CCA1 and CCA2 cell lines. Based on this clustering, we could assign individual CCA patients to our drug response-based subgroups (Figure 8A). Of interest, the patients with the characteristics of CCA1 and CCA2 subgroups were composed mainly of the CCA molecular clusters type I and type IV, respectively. We then used the ridge regression to

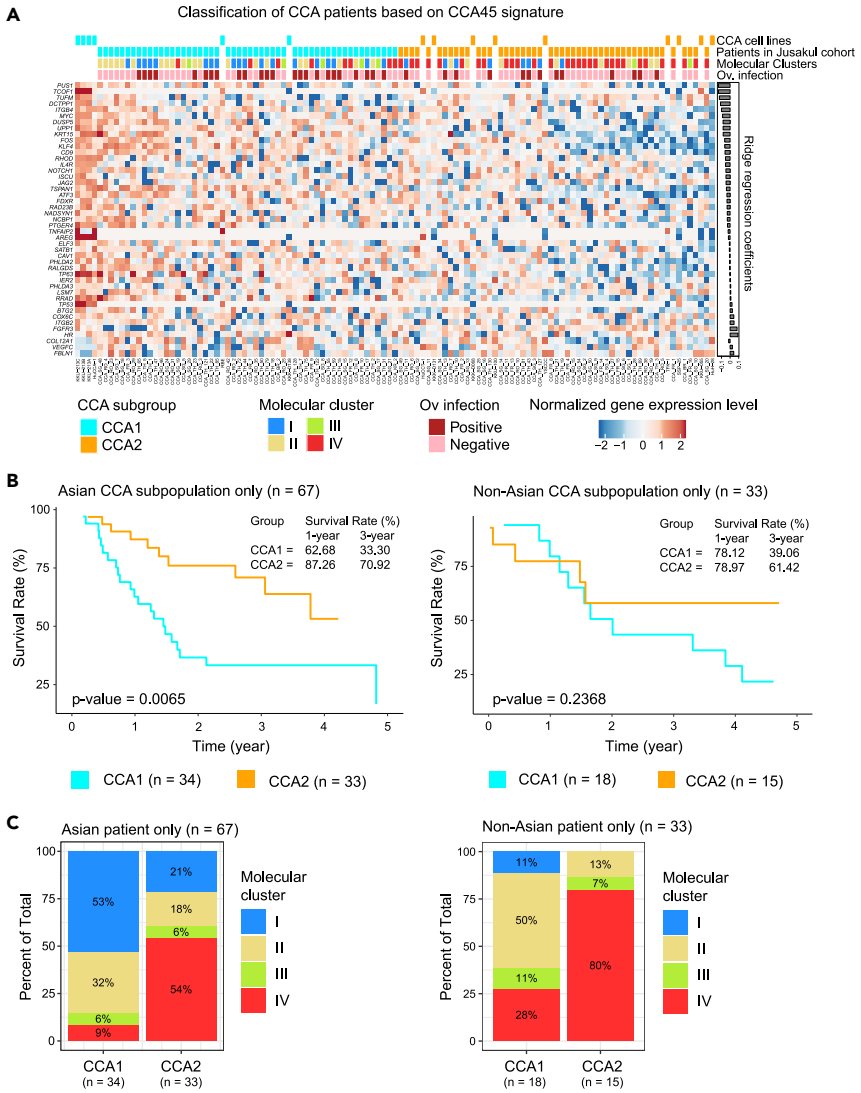


Figure 8. Prognostic value of the CCA45 signature for survival prediction in the Jusakul cohort

(A) Classification of CCA patients in the Jusakul cohort (n = 100) using the CCA45 signature. All samples are clustered using Euclidean distance and Ward2 algorithm. Genes are arranged according to their coefficient values from the ridge regression model. The CCA cell lines are annotated based on their drug response-based subgroups from Figure 3A. CCA patients are annotated as CCA1 and CCA2 subgroups based on the hierarchical clustering, with their corresponding molecular clusters and *O. viverrini* infection status from Jusakul et al. study (Jusakul et al., 2017).

(B) Survival analysis of Asian CCA patients (n = 67, left panel) and non-Asian CCA patients (n = 33, right panel) between CCA1 (cyan) and CCA2 (orange) subgroups. Kaplan-Meier plot is shown, and log-rank statistics are used to determine the significance level of median survival times when comparing CCA1 and CCA2 subgroups.

(C) Fractions of the molecular clusters within CCA1 and CCA2 subgroups among the Asian (left panel) and non-Asian (right panel) CCA patients.

estimate the weight coefficients of each biomarker (Figure 8A right panel; Table S5) and developed a weighted sum model (WSM) score for the classification of CCA patients (WSM score >0 for the CCA1 subgroup and WSM score <= 0 for the CCA2 subgroup). The clustering landscape of CCA patient and cell lines was best visualized when the data was dimensionally reduced using PCA and the two patient subgroups were separated using support vector machine (SVM) (Figure S7B). Indeed, the SVM-based prediction model could predict drug response subgroups of CCA patients with high accuracy compared to the hierarchical clustering approach (AUC = 0.89) (Figure S7C).

Having confirmed that our drug response-based subgrouping can be applied in the clinical setting, we next examined whether the developed CCA45 signature also holds any prognostic values. To assess this, we utilized the reported survival data from the same clinical cohort to compare the survival time of different patient subgroups. First, CCA1 patients showed a median survival time of 1.58 years, whereas CCA2 patients had indeterminable survival time of at least four years (log-rank p-value = 0.004, Figure S7D). In addition, CCA1 and CC2 patients showed 1-year survival rates of 67.94% and 84.52%, and 3-year survival rates of 35.38% and 67.88%, respectively. When we performed subgroup analysis on different patient ethnicity, we found that the prognosis prediction using the CCA45 signature was more accurate in the Asian CCA subgroup (p-value = 0.0065) than that of the non-Asian subgroup (p-value = 0.2368) (Figure 8B). Of interest, when we limited our analysis to only Thai CCA patients, CCA1 patients still showed a poorer 1-year survival rate than CCA2 ones (50% and 83.33%, respectively). However, their overall survival times were not significantly different (Figure S7E, log-rank p-value = 0.0884).

Having found that the CCA1 and CCA2 patients were composed mainly of the molecular cluster type I and IV, we inspected the distribution of molecular clusters in each drug response-based subgroup. Across all patients, we found that patients with molecular cluster types I and II were most frequently observed in the CCA1 patient subgroups (Figure S7F). When we limited the analysis only to the Asian CCA patients, the proportion of molecular cluster type I patients was substantially higher than other cluster types in the CCA1 subgroup (53%) (Figure 8C, left panel). For the CCA2 subgroup, the molecular cluster type IV patients were found to be the majority contributor (>50%) regardless of their ethnicity (Figures 8C and S7F). Although the median survival time between the two CCA drug response-based subgroups within the Thai CCA subpopulation was not significantly different, most CCA1 patients were composed of the molecular cluster types I and II (Figure S7G). Collectively, we demonstrated that the CCA45 signature also holds prognostic values, especially for Asian CCA patients. Nonetheless, further clinical studies are needed to confirm the efficacy of our proposed subgroup-specific drug candidates and the developed biomarker panel in more diverse patient subgroups.

DISCUSSION

Advanced CCA patients suffer from limited drug options like other rare cancer types, mainly relying on standard chemotherapeutic agents with poor clinical outcomes. To determine if a newly developed anti-cancer drug could be applicable for CCA patients, past efforts relied on non-selective trials with unsatisfactory clinical outcomes (probability of success (POS) for clinical studies without and with biomarkers of 1.6% vs. 10.7%) (Wong et al., 2019). Our study reported an unbiased high-throughput approach to repurposing anti-cancer drug candidates for the commonly underrepresented Asian CCA. We took advantage of the unique CCA cell lines from Thai CCA patients (from *O. viverrini* infection endemic areas) and the Japanese CCA cell lines from the Japanese cell banks to establish a comprehensive library of drug responses against 100 approved anti-cancer drugs. By carefully investigating the pairwise correlation of drug potency across all cell lines, we were able to subgroup Asian CCA cell lines based on their drug response similarity and assign to each CCA subgroup suitable drug candidates. For the first time, we identified a subset of Asian CCA patients that demonstrated our newly developed gene expression signature called CCA45 and showed that this patient subgroup could benefit from targeted therapy such as MEK and Src inhibitors. Our approach to identifying drug candidates based on 1) subgroup specificity score, 2) the comparison of drug potency and efficacy between subgroups, and 3) the estimation of the therapeutic index by profiling toxicity against the normal cell line can serve as an accelerated strategy for identifying specific drug choices for neglected subgroups of rare cancer beyond CCA. To encourage further exploration and clinical translation of our work, we made all data publicly available for academic use at https://sisp.shinyapps.io/AsianCCAbrowser/.

Our study further elucidated the molecular uniqueness of Asian CCA subgroups. We identified 15 mutations associated with either drug sensitivity or drug resistance using the accompanying mutational profiling. Some mutations were uniquely identified in either CCA subgroups. For example, the genetic alterations of *LRP1B* and *FLT3* genes were only observed in CCA1 cell lines. *LRP1B* mutations were associated with platinum drug resistance in Asian metastatic CCA patients (Xu et al., 2020). *FLT3* mutations were also reported to be associated with Azacitidine sensitivity in acute myeloid leukemia (AML) (Zhao et al., 2019). We hypothesize that some identified drug response-associated mutations may act as oncogenic drivers and render CCA cells sensitive to novel drug candidates. This phenomenon, also known as

"acquired vulnerability," has opened new therapeutic opportunities in other cancer types (Nijhawan et al., 2012; Wang et al., 2018).

Considering the heterogeneous nature of CCA pathological specimens (Bragazzi et al., 2018), our newly developed CCA45 gene expression signature can make the classification of CCA patients more robust than using the mutation-centric approach in conventional studies. Although the CCA45 signature was generated from the comprehensive drug response profiling in cell lines, we showed that CCA1 subgroup exhibits a similar transcriptomic profile to the CCA molecular cluster type I, commonly observed among the liver fluke-positive CCA patients (Jusakul et al., 2017). The CCA45 signature successfully classified individual CCA patients into our drug response-based subgroups. We also showed that the CCA45 signature holds prognostic values, especially for the Asian CCA subpopulation. This effort is consistent with the recent adoption of transcriptomics-based subgroup-classifying biomarkers in the clinical diagnosis of breast cancer (Smid et al., 2016) and colorectal cancer (Guinney et al., 2015), substantiating the use of gene expression as a more robust approach for cancer classification. We encourage a further collaborative effort to develop the gene expression-based consensus molecular subgroups in CCA. This effort can be made possible by enabling data sharing from large CCA cohorts with diverse patient ethnicity; a common problem shared across all rare cancer types. The accuracy of subgroup classification could also be further improved with more CCA subgroup-representing cell lines. Specifically for Thai CCA cell lines, we are expanding our cell line collection using the successful establishment protocol via a patient-derived xenograft mouse model (Vaeteewoottacharn et al., 2019). Following the success of tumoroid development from iCCA patients (Broutier et al., 2017; Saito, 2019), we are also establishing *ex vivo* avatar models from Thai CCA patients and will make these models available for further in-depth preclinical studies among the international research community.

MEK and Src inhibitors were the best drug candidates for CCA1 subgroup, showing selective cytotoxicity against CCA1 cell lines with minimal toxicity against the MMNK1 cholangiocyte cell line. Some of these drugs have already been investigated in biomarker-guided clinical trials but showed poor outcomes and heterogeneous responses. In a phase II study of Selumetinib in patients with metastatic biliary cancers, the absence of ERK phosphorylation was associated with a lack of response. However, the median overall survival was only 9.8 months (Bekaii-Saab et al., 2011). A similar drug screening study in the Caucasian iCCA cell lines identified the association of the *IDH1* mutations with Dasatinib sensitivity (Saha et al., 2016). This finding led to a follow-up phase-II clinical study in CCA patients with *IDH* mutant (NCT02428855). More recently, a cohort study of the ROAR basket trial reported an improved efficacy in patients with biliary cancer with BRAF V600E mutation treated with Dabrafenib and Trametinib combination (median overall survival of 11.3 months) (Subbiah et al., 2020). Understanding how different subgroups of CCA patients could respond to regimens that target proteins within the same signaling cascade, regardless of their mutational profiles, will help us translate these insights into more patients.

Our study was also the first to report the ligand-mediated pathway activation in CCA. Cognate ligands of different signaling cascades exhibit distinct activation patterns of their respective pathways for the different CCA subgroups. We discovered that IGF1 enhanced ERK signaling pathway in the CCA1 subgroup, consistent with a prior finding (Andersen et al., 2012). Similarly, Anisomycin and the cognate ligand EGF also activated the ERK signaling pathway to a lesser extent. In CCA2 subgroup, however, the same set of ligands was observed to activate the AKT signaling pathway. Our findings support the existence of ERK-AKT signaling pathway cross-talk in CCA, in which activation of one or both pathways promotes cell proliferation and anti-cancer drug resistance (Ewald et al., 2014; Yoon et al., 2011). Our study found that CCA1 subgroup was more sensitive to MEK inhibitors than CCA2 subgroup that showed higher AKT activity. This finding is consistent with previous research indicating that tyrosine kinase inhibitor-sensitive CCA had low AKT activation. AKT phosphorylation in iCCA with abundant surface EGFR has also been reported, which is controversial (Schmitz et al., 2007). Despite the rarity of EGFR mutations, subgroups of CCA with an identical amount of EGFR but distinct AKT or ERK signaling activity can be distinguished.

Some of the disparities in drug responses between our two CCA subgroups may be attributable to genetic background differences. Our MEK inhibitor-sensitive CCA1 subgroup associated with *KRAS* mutation and IGF1-induced ERK activity highlighted previous findings that *KRAS* was required and may bridge the ERK

and AKT signaling pathways (Molina-Arcas et al., 2013; Simpson et al., 2017). In addition, we reported that the *FLT3*mutation was uniquely associated with the CCA1 subgroup and that ERK activity varied based on the ligand type. Of interest, *FLT3* mutation-harboring cancer cells have been reported to exhibit dynamic ERK signaling pathway activation (Bruner et al., 2017; McCubrey et al., 2008). Additional research utilizing a live-cell kinase translocation reporter may be necessary to determine whether the ERK activity observed in the CCA1 subgroup is also dynamic (Regot et al., 2014).

In contrast, our CCA2 subgroup showed a consistent AKT response to pro-growth (IGF1, EGF) and pro-inflammatory (TNFα and IFNγ) ligands. The *TP53*, *PIK3C2B*, *KEAP1*, and *PTEN* mutations in the CCA2 subgroup may act as constitutive activators of the AKT pathway, consistent with the findings in other cancer types. For instance, *TP53* mutation led to a significantly activated AKT pathway (Yue et al., 2020). *PTEN* mutation was shown to induce basal AKT phosphorylation (Niepel et al., 2014). *PIK3KC2B* mutation could also activate the AKT pathway in non-small cell lung cancer (Kind et al., 2017). Finally, the inactivation of the *FoxO3-KEAP1* in CCA accounts for the constitutive activation of *NRF2* that can result in the activation of the AKT pathway (Guan et al., 2016).

In conclusion, we demonstrated that a comprehensive drug response profiling of representative cell lines and the accompanying omics data could accelerate the identification of drug candidates for a rare understudied cancer like CCA. The biomarker developed based on our comprehensive drug response profiling proves to be more directly applicable to personalized drug choice selection and the prediction of patient prognosis. Finally, we confirm that transcriptomics-based biomarkers improve the robustness of subgroup stratification in CCA patients.

Limitations of the study

Our study is undoubtedly limited by the small number of CCA cell lines. Although we successfully showed using clinical data from different ethnic backgrounds that both the CCA1 and CCA2 signatures exist in clinical settings, broadening our analysis using larger number of cell line collections and cancer models other than 2D cell culture can further validate our prioritized drug lists. Our group is currently working to produce more supportive evidence using 3D spheroid culture models. We are also expanding our collection of patient-derived xenograft models from CCA patients, which can later be used to validate the choice of repurposable drugs such as MEK and Src inhibitors for patients with the CCA1 signature. The prognostic value of CCA45 signature can also be further validated with access to a broader clinico-genomic database of CCA patients. In particular, we are interested in validating the use of CCA45 signature as the molecular inclusion criteria from clinical trials that utilized MEK and Src inhibitors in CCA patients.

Significance

Our library of drug response profiling and pan-omic data offers novel treatment options for Asian liver fluke-associated CCA with subgroup-classifying biomarkers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and reagents
- METHOD DETAILS
  - DNA preparation and cancer gene panel sequencing
  - RNA preparation, sequencing, and transcriptome profiling
  - Drug response profiling and comparison and network analysis
  - Drug sensitivity determination, pathway analysis, and drug response biomarkers prediction
  - Association between drug sensitivity and genomic alteration
  - Ligand response profiling and analysis

- Clinical data selection and normalization
- Subclass mapping
- Prediction of drug response subtype in CCA cell lines and patients
- Prognostic evaluation of the biomarkers

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105182.

ACKNOWLEDGMENTS

This work was supported by the National Science and Technology Development Agency of Thailand (NSTDA, P-15-50208), the Advanced Research in Pharmacology Fund, Siriraj Foundation (D003421), and Mahidol University (Basic Research Fund: fiscal year 2021). S.S., S.Ji were supported by the Chalermphrakiat Grant, Faculty of Medicine, Siriraj Hospital, Mahidol University. P.K. was supported by the National Research Council of Thailand.

AUTHOR CONTRIBUTIONS

Conceptualization, S.S., S.Ji., S.Ja., and P.K.; Methodology, S.S., S.Ja., P.K., and P.B.; Investigation, S.Ja., P.K.; Formal Analysis, S.Ja., P.K., P.B., and P.S.; Writing-Original Draft, S.S., S.Ja., and P.K.; Writing – Review and Editing, S.S., S.Ja., P.K., S.Ji., S.O., and K.V.; Visualization, S.Ja., P.K., and P.B.; Funding Acquisition, S.S., S.Ji., and K.V.; Resources, S.S., S.Ji., K.V., and S.O.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 6, 2022
Revised: August 11, 2022
Accepted: September 20, 2022
Published: October 21, 2022

REFERENCES

Ahn, D.H., and Bekaii-Saab, T. (2017). Biliary cancer: intrahepatic cholangiocarcinoma vs. extrahepatic cholangiocarcinoma vs. gallbladder cancers: classification and therapeutic implications. J. Gastrointest. Oncol. *8*, 293–301. https://doi.org/10.21037/jgo.2016.10.01.

Andersen, J.B., Spee, B., Blechacz, B.R., Avital, I., Komuta, M., Barbour, A., Conner, E.A., Gillen, M.C., Roskams, T., Roberts, L.R., et al. (2012). Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology *142*, 1021–1031.e15. https://doi.org/10.1053/j.gastro.2011.12.005.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607. https://doi.org/10.1038/nature11003.

Bekaii-Saab, T., Phelps, M.A., Li, X., Saji, M., Goff, L., Kauh, J.S.W., O'Neil, B.H., Balsom, S., Balint, C., Liersemann, R., et al. (2011). Multi-institutional phase II study of selumetinib in patients with metastatic biliary cancers. J. Clin. Oncol. *29*, 2357–2363. https://doi.org/10.1200/JCO.2010.33.9473.

Benson, A.B., D'Angelica, M.I., Abbott, D.E., Anaya, D.A., Anders, R., Are, C., Bachini, M., Borad, M., Brown, D., Burgoyne, A., et al. (2021). Hepatobiliary cancers, version 2.2021, NCCN clinical practice guidelines in oncology. J. Natl. Compr. Canc. Netw. *19*, 541–565. https://doi.org/10.6004/jnccn.2021.0022.

Bragazzi, M.C., Ridola, L., Safarikia, S., Matteo, S.D., Costantini, D., Nevi, L., and Cardinale, V. (2018). New insights into cholangiocarcinoma: multiple stems and related cell lineages of origin. Ann. Gastroenterol. *31*, 42–55. https://doi.org/10.20524/aog.2017.0209.

Broutier, L., Mastrogiovanni, G., Verstegen, M.M., Francies, H.E., Gavarró, L.M., Bradshaw, C.R., Allen, G.E., Arnes-Benito, R., Sidorova, O., Gaspersz, M.P., et al. (2017). Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. Nat. Med. *23*, 1424–1435. https://doi.org/10.1038/nm.4438.

Bruner, J.K., Ma, H.S., Li, L., Qin, A.C.R., Rudek, M.A., Jones, R.J., Levis, M.J., Pratz, K.W., Pratilas, C.A., and Small, D. (2017). Adaptation to TKI treatment reactivates ERK signaling in tyrosine kinase-driven leukemias and other malignancies. Cancer Res. *77*, 5554–5563. https://doi.org/10.1158/0008-5472.CAN-16-2593.

Butthongkomvong, K., Sirachainan, E., Jhankumpha, S., Kumdang, S., and Sukhontharot, O.U. (2013). Treatment outcome of palliative chemotherapy in inoperable cholangiocarcinoma in Thailand. Asian Pac. J. Cancer Prev. *14*, 3565–3568. https://doi.org/10.7314/apjcp.2013.14.6.3565.

Chaisaingmongkol, J., Budhu, A., Dang, H., Rabibhadana, S., Pupacdi, B., Kwon, S.M., Forgues, M., Pomyen, Y., Bhudhisawasdi, V., Lertprasertsuke, N., et al. (2017). Common molecular subtypes among asian hepatocellular carcinoma and cholangiocarcinoma. Cancer Cell *32*, 57–70.e3. https://doi.org/10.1016/j.ccell.2017.05.009.

Chan-On, W., Nairismägi, M.L., Ong, C.K., Lim, W.K., Dima, S., Pairojkul, C., Lim, K.H., McPherson, J.R., Cutcutache, I., Heng, H.L., et al.

(2013). Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. Nat. Genet. *45*, 1474–1478. https://doi.org/10.1038/ng.2806.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Clark, N.A., Hafner, M., Kouril, M., Williams, E.H., Muhlich, J.L., Pilarczyk, M., Niepel, M., Sorger, P.K., and Medvedovic, M. (2017). GRcalculator: an online tool for calculating and mining dose–response data. BMC cancer *17* (1), 1–11. https://doi.org/10.1186/s12885-017-3689-3.

Crane, E.K., Kwan, S.Y., Izaguirre, D.I., Tsang, Y.T.M., Mullany, L.K., Zu, Z., Richards, J.S., Gershenson, D.M., and Wong, K.K. (2015). Nutlin-3a: a potential therapeutic opportunity for TP53 wild-type ovarian carcinomas. PLoS One *10*, e0135101. https://doi.org/10.1371/journal.pone.0135101.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, complex systems *1695* (5), 1–9. https://igraph.org.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dong, M., Liu, X., Evert, K., Utpatel, K., Peters, M., Zhang, S., Xu, Z., Che, L., Cigliano, A., Ribback, S., et al. (2018). Efficacy of MEK inhibition in a K-Ras-driven cholangiocarcinoma preclinical model. Cell Death Dis. *9*, 31. https://doi.org/10.1038/s41419-017-0183-4.

Ducreux, M., Rougier, P., Fandi, A., Clavero-Fabri, M.C., Villing, A.L., Fassone, F., Fandi, L., Zarba, J., and Armand, J.P. (1998). Effective treatment of advanced biliary tract carcinoma using 5-fluorouracil continuous infusion with cisplatin. Ann. Oncol. *9*, 653–656.

Ewald, F., Nörz, D., Grottke, A., Hofmann, B.T., Nashan, B., and Jücker, M. (2014). Dual Inhibition of PI3K-AKT-mTOR- and RAF-MEK-ERK-signaling is synergistic in cholangiocarcinoma and reverses acquired resistance to MEK-inhibitors. Invest. New Drugs *32*, 1144–1154. https://doi.org/10.1007/s10637-014-0149-7.

Farshidfar, F., Zheng, S., Gingras, M.C., Newton, Y., Shih, J., Robertson, A.G., Hinoue, T., Hoadley, K.A., Gibb, E.A., Roszik, J., et al. (2017). Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. Cell Rep. *18*, 2780–2794. https://doi.org/10.1016/j.celrep.2017.02.033.

Fruchterman, T.M.J., and Reingold, E.M. (1991). Graph drawing by force-directed placement. Softw:. Pract. Exper. *21*, 1129–1164. https://doi.org/10.1002/spe.4380211102.

Gkountakos, A., Sartori, G., Falcone, I., Piro, G., Ciuffreda, L., Carbone, C., Tortora, G., Scarpa, A., Bria, E., Milella, M., et al. (2019). PTEN in lung cancer: dealing with the problem, building on new knowledge and turning the game around. Cancers *11*, 1141. https://doi.org/10.3390/cancers11081141.

Guan, L., Zhang, L., Gong, Z., Hou, X., Xu, Y., Feng, X., Wang, H., and You, H. (2016). FoxO3 inactivation promotes human cholangiocarcinoma tumorigenesis and chemoresistance through Keap1-Nrf2 signaling. Hepatology *63*, 1914–1927. https://doi.org/10.1002/hep.28496.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356. https://doi.org/10.1038/nm.3967.

Hafner, M., Mills, C.E., Subramanian, K., Chen, C., Chung, M., Boswell, S.A., Everley, R.A., Liu, C., Walmsley, C.S., Juric, D., and Sorger, P.K. (2019). Multiomics profiling establishes the polypharmacology of FDA-approved CDK4/6 inhibitors and the potential for differential clinical activity. Cell Chem. Biol. *26*, 1067–1080.e8. https://doi.org/10.1016/j.chembiol.2019.05.005.

Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods *13*, 521–527. https://doi.org/10.1038/nmeth.3853.

Hoshida, Y., Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2007). Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One *2*, e1195. https://doi.org/10.1371/journal.pone.0001195.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A landscape of pharmacogenomic interactions in cancer. Cell *166*, 740–754. https://doi.org/10.1016/j.cell.2016.06.017.

Jones, D.S., Jenney, A.P., Swantek, J.L., Burke, J.M., Lauffenburger, D.A., and Sorger, P.K. (2017). Profiling drugs for rheumatoid arthritis that inhibit synovial fibroblast activation. Nat. Chem. Biol. *13*, 38–45. https://doi.org/10.1038/nchembio.2211.

Jusakul, A., Cutcutache, I., Yong, C.H., Lim, J.Q., Huang, M.N., Padmanabhan, N., Nellore, V., Kongpetch, S., Ng, A.W.T., Ng, L.M., et al. (2017). Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. Cancer Discov. *7*, 1116–1135. https://doi.org/10.1158/2159-8290.CD-17-0368.

Kassambara, A., Kosinski, M., and Biecek, P. (2020). Drawing survival curves using 'ggplot2' [r package survminerversion 0.4.8]. https://rpkgs.datanovia.com/survminer/.

Kind, M., Klukowska-Rötzler, J., Berezowska, S., Arcaro, A., and Charles, R.P. (2017). Questioning the role of selected somatic PIK3C2B mutations in squamous non-small cell lung cancer oncogenesis. PLoS One *12*, e0187308. https://doi.org/10.1371/journal.pone.0187308.

Kitai, H., Ebi, H., Tomida, S., Floros, K.V., Kotani, H., Adachi, Y., Oizumi, S., Nishimura, M., Faber, A.C., and Yano, S. (2016). Epithelial-to-Mesenchymal transition defines feedback activation of receptor tyrosine kinase signaling induced by MEK inhibition in KRAS-mutant lung cancer. Cancer Discov. *6*, 754–769. https://doi.org/10.1158/2159-8290.CD-15-1377.

Kuhn, M. (2008). Building predictive models inRUsing thecaretPackage. J. Stat. Softw. *28*, 5. https://doi.org/10.18637/jss.v028.i05.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883. https://doi.org/10.1093/bioinformatics/bts034.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., Vellano, C.P., Roebuck, P., Yu, Q., Eterovic, A.K., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. Cancer Cell *31*, 225–239. https://doi.org/10.1016/j.ccell.2017.01.005.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108. https://doi.org/10.1093/nar/gkt214.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

Luvira, V., Nilprapha, K., Bhudhisawasdi, V., Pugkhem, A., Chamadol, N., and Kamsa-ard, S. (2016). Cholangiocarcinoma patient outcome in northeastern Thailand: single-center prospective study. Asian Pac. J. Cancer Prev. *17*, 401–406. https://doi.org/10.7314/apjcp.2016.17.1.401.

Ma, T., Galimberti, F., Erkmen, C.P., Memoli, V., Chinyengetere, F., Sempere, L., Beumer, J.H., Anyang, B.N., Nugent, W., Johnstone, D., et al. (2013). Comparing histone deacetylase inhibitor responses in genetically engineered mouse lung cancer models and a window of opportunity trial in patients with lung cancer. Mol. Cancer Ther. *12*, 1545–1555. https://doi.org/10.1158/1535-7163.MCT-12-0933.

Maaten, L.v.d., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

McCubrey, J.A., Steelman, L.S., Abrams, S.L., Bertrand, F.E., Ludwig, D.E., Bäsecke, J., Libra, M., Stivala, F., Milella, M., Tafuri, A., et al. (2008). Targeting survival cascades induced by activation of Ras/Raf/MEK/ERK, PI3K/PTEN/Akt/mTOR and Jak/STAT pathways for effective leukemia therapy. Leukemia *22*, 708–722. https://doi.org/10.1038/leu.2008.27.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M.G.F., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N., et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput. Biol. *7*, e1001138. https://doi.org/10.1371/journal.pcbi.1001138.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al.

(2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol. *16*, e2005970. https://doi.org/10.1371/journal.pbio.2005970.

Merchant, M., Moffat, J., Schaefer, G., Chan, J., Wang, X., Orr, C., Cheng, J., Hunsaker, T., Shao, L., Wang, S.J., et al. (2017). Combined MEK and ERK inhibition overcomes therapy-mediated pathway reactivation in RAS mutant tumors. PLoS One *12*, e0185862. https://doi.org/10.1371/journal.pone.0185862.

Molina-Arcas, M., Hancock, D.C., Sheridan, C., Kumar, M.S., and Downward, J. (2013). Coordinate direct input of both KRAS and IGF1 receptor to activation of PI3 kinase in KRAS-mutant lung cancer. Cancer Discov. *3*, 548–563. https://doi.org/10.1158/2159-8290.CD-12-0446.

Niepel, M., Hafner, M., Pace, E.A., Chung, M., Chai, D.H., Zhou, L., Muhlich, J.L., Schoeberl, B., and Sorger, P.K. (2014). Analysis of growth factor signaling in genetically diverse breast cancer lines. BMC Biol. *12*, 20. https://doi.org/10.1186/1741-7007-12-20.

Nijhawan, D., Zack, T.I., Ren, Y., Strickland, M.R., Lamothe, R., Schumacher, S.E., Tsherniak, A., Besche, H.C., Rosenbluh, J., Shehata, S., et al. (2012). Cancer vulnerabilities unveiled by genomic loss. Cell *150*, 842–854. https://doi.org/10.1016/j.cell.2012.07.023.

Ong, C.K., Subimerb, C., Pairojkul, C., Wongkham, S., Cutcutache, I., Yu, W., McPherson, J.R., Allen, G.E., Ng, C.C.Y., Wong, B.H., et al. (2012). Exome sequencing of liver fluke-associated cholangiocarcinoma. Nat. Genet. *44*, 690–693. https://doi.org/10.1038/ng.2273.

Pairojkul, C. (2014). Liver fluke and cholangiocarcinoma in Thailand. Pathology *46*, S24.

Pellino, A., Loupakis, F., Cadamuro, M., Dadduzio, V., Fassan, M., Guido, M., Cillo, U., Indraccolo, S., and Fabris, L. (2018). Precision medicine in cholangiocarcinoma. Transl. Gastroenterol. Hepatol. *3*, 40. https://doi.org/10.21037/tgh.2018.07.02.

R Development Core Team. (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.r-project.org/.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116. https://doi.org/10.1038/nchembio.1986.

Regot, S., Hughey, J.J., Bajar, B.T., Carrasco, S., and Covert, M.W. (2014). High-sensitivity measurements of multiple kinase activities in live single cells. Cell *157*, 1724–1734. https://doi.org/10.1016/j.cell.2014.04.039.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nature genetics *38* (5), 500–501. https://doi.org/10.1038/ng0506-500.

Saha, S.K., Gordan, J.D., Kleinstiver, B.P., Vu, P., Najem, M.S., Yeo, J.C., Shi, L., Kato, Y., Levin, R.S., Webber, J.T., et al. (2016). Isocitrate dehydrogenase mutations confer Dasatinib hypersensitivity and SRC dependence in intrahepatic cholangiocarcinoma. Cancer Discov. *6*, 727–739. https://doi.org/10.1158/2159-8290.CD-15-1442.

Saito, Y. (2019). Establishment of an organoid bank of biliary tract and pancreatic cancers and its application for personalized therapy and future treatment. J. Gastroenterol. Hepatol. *34*, 1906–1910. https://doi.org/10.1111/jgh.14773.

Sampattavanich, S., Steiert, B., Kramer, B.A., Gyori, B.M., Albeck, J.G., and Sorger, P.K. (2018). Encoding growth factor identity in the temporal dynamics of FOXO3 under the combinatorial control of ERK and AKT kinases. Cell Syst. *6*, 664–678.e9. https://doi.org/10.1016/j.cels.2018.05.004.

Schmitz, K.J., Lang, H., Wohlschlaeger, J., Sotiropoulos, G.C., Reis, H., Schmid, K.W., and Baba, H.A. (2007). AKT and ERK1/2 signaling in intrahepatic cholangiocarcinoma. World J. Gastroenterol. *13*, 6470–6477. https://doi.org/10.3748/wjg.v13.i48.6470.

Shin, H., Shannon, C.P., Fishbane, N., Ruan, J., Zhou, M., Balshaw, R., Wilson-McManus, J.E., Ng, R.T., McManus, B.M., and Tebbutt, S.J.; PROOF Centre of Excellence Team (2014). Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. PLoS One *9*, e91041. https://doi.org/10.1371/journal.pone.0091041.

Simpson, A., Petnga, W., Macaulay, V.M., Weyer-Czernilofsky, U., and Bogenrieder, T. (2017). Insulin-like growth factor (IGF) pathway targeting in cancer: role of the IGF Axis and opportunities for future combination studies. Target. Oncol. *12*, 571–597. https://doi.org/10.1007/s11523-017-0514-5.

Smid, M., Rodríguez-González, F.G., Sieuwerts, A.M., Salgado, R., Prager-Van der Smissen, W.J.C., Vlugt-Daane, M.v.d., van Galen, A., Nik-Zainal, S., Staaf, J., Brinkman, A.B., et al. (2016). Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nat. Commun. *7*, 12910. https://doi.org/10.1038/ncomms12910.

Sripa, B., Kaewkes, S., Sithithaworn, P., Mairiang, E., Laha, T., Smout, M., Pairojkul, C., Bhudhisawasdi, V., Tesana, S., Thinkamrop, B., et al. (2007). Liver fluke induces cholangiocarcinoma. PLoS Med. *4*, e201. https://doi.org/10.1371/journal.pmed.0040201.

Sripa, B., Leungwattanawanit, S., Nitta, T., Wongkham, C., Bhudhisawasdi, V., Puapairoj, A., Sripa, C., and Miwa, M. (2005). Establishment and characterization of an opisthorchiasis-associated cholangiocarcinoma cell line (KKU-100). World J. Gastroenterol. *11*, 3392–3397. https://doi.org/10.3748/wjg.v11.i22.3392.

Sripa, B., Seubwai, W., Vaeteewoottacharn, K., Sawanyawisuth, K., Silsirivanit, A., Kaewkong, W., Muisuk, K., Dana, P., Phoomak, C., Lert-Itthiporn, W., et al. (2020). Functional and genetic characterization of three cell lines derived from a single tumor of an Opisthorchis viverrini-associated cholangiocarcinoma patient. Hum. Cell *33*, 695–708. https://doi.org/10.1007/s13577-020-00334-w.

Subbiah, V., Lassen, U., Élez, E., Italiano, A., Curigliano, G., Javle, M., de Braud, F., Prager, G.W., Greil, R., Stein, A., et al. (2020). Dabrafenib plus trametinib in patients with BRAF(V600E)-mutated biliary tract cancer (ROAR): a phase 2, open-label, single-arm, multicentre basket trial. Lancet Oncol. *21*, 1234–1243. https://doi.org/10.1016/S1470-2045(20)30321-1.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

Targonski, C.A., Shearer, C.A., Shealy, B.T., Smith, M.C., and Feltus, F.A. (2019). Uncovering biomarker genes with enriched classification potential from Hallmark gene sets. Sci. Rep. *9*, 9747. https://doi.org/10.1038/s41598-019-46059-1.

Therneau, T.M. (2022). A package for survival analysis in R. https://cran.r-project.org/web/packages/survival/.

Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model (Springer). https://doi.org/10.1007/978-1-4757-3294-8.

Thompson, J.A., Tan, J., and Greene, C.S. (2016). Cross-platform normalization of microarray and RNA-seq data for machine learning applications. PeerJ *4*, e1621. https://doi.org/10.7717/peerj.1621.

Vaeteewoottacharn, K., Pairojkul, C., Kariya, R., Muisuk, K., Imtawil, K., Chamgramol, Y., Bhudhisawasdi, V., Khuntikeo, N., Pugkhem, A., Saeseow, O.T., et al. (2019). Establishment of highly transplantable cholangiocarcinoma cell lines from a patient-derived xenograft mouse model. Cells *8*, 496. https://doi.org/10.3390/cells8050496.

Valle, J., Wasan, H., Palmer, D.H., Cunningham, D., Anthoney, A., Maraveyas, A., Madhusudan, S., Iveson, T., Hughes, S., Pereira, S.P., et al. (2010). Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N. Engl. J. Med. *362*, 1273–1281. https://doi.org/10.1056/NEJMoa0908721.

Valle, J.W., Borbath, I., Khan, S.A., Huguet, F., Gruenberger, T., ESMO Guidelines Committee, et al.. (2016). Biliary cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. *27*, v28–v37. https://doi.org/10.1093/annonc/mdw324.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

Wang, L., Leite de Oliveira, R., Huijberts, S., Bosdriesz, E., Pencheva, N., Brunen, D., Bosma, A., Song, J.Y., Zevenhoven, J., Los-de Vries, G.T.,

et al. (2018). An acquired vulnerability of drug-resistant melanoma with therapeutic potential. Cell *173*, 1413–1425.e14. https://doi.org/10.1016/j.cell.2018.04.012.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

Williams, C.R., Baccarella, A., Parrish, J.Z., and Kim, C.C. (2016). Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC Bioinf. *17*, 103. https://doi.org/10.1186/s12859-016-0956-2.

Wong, C.H., Siah, K.W., and Lo, A.W. (2019). Estimation of clinical trial success rates and related parameters. Biostatistics *20*, 273–286. https://doi.org/10.1093/biostatistics/kxx069.

Xu, S.-F., Guo, Y., Zhang, X., Zhu, X.-D., Fan, N., Zhang, Z.-L., Ren, G.-B., Rao, W., and Zang, Y.-J. (2020). Somatic mutation profiling of intrahepatic cholangiocarcinoma: comparison between primary and metastasis tumor tissues. J. Oncol. *2020*, 5675020. https://doi.org/10.1155/2020/5675020.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961. https://doi.org/10.1093/nar/gks1111.

Yoon, H., Min, J.K., Lee, J.W., Kim, D.G., and Hong, H.J. (2011). Acquisition of chemoresistance in intrahepatic cholangiocarcinoma cells by activation of AKT and extracellular signal-regulated kinase (ERK) 1/2. Biochem. Biophys. Res. Commun. *405*, 333–337. https://doi.org/10.1016/j.bbrc.2010.11.130.

Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., and Verhaak, R.G.W. (2015). The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene *34*, 4845–4854. https://doi.org/10.1038/onc.2014.406.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

Yue, X., Wu, F., Li, Y., Liu, J., Boateng, M., Mandava, K., Zhang, C., Feng, Z., Gao, J., and Hu, W. (2020). Gain of function mutant p53 protein activates AKT through the Rac1 signaling to promote tumorigenesis. Cell Cycle *19*, 1338–1351. https://doi.org/10.1080/15384101.2020.1749790.

Zhao, J., Song, Y., and Liu, D. (2019). Gilteritinib: a novel FLT3 inhibitor for acute myeloid leukemia. Biomark. Res. *7*, 19. https://doi.org/10.1186/s40364-019-0170-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) rabbit mAb	Cell Signaling Technology	Cat #9101; RRID:AB_331646
phospho-Akt (Ser473) (D9E) XP rabbit mAb	Cell Signaling Technology	Cat #4060; RRID:AB_2315049
phospho-p70 S6 kinase (Thr389)(108D2) rabbit mAb	Cell Signaling Technology	Cat #9234; RRID:AB_2269803
phospho-Stat1 (Tyr701) (58D6) rabbit mAb	Cell Signaling Technology	Cat #9167; RRID:AB_561284
NF-κB p65 (D14E12) XP rabbit mAb	Cell Signaling Technology	Cat #8242; RRID:AB_10859369
Anti-rabbit IgG (H + L), F(ab')2 fragment (Alexa Fluor 647 conjugate)	Cell Signaling Technology	Cat #4414; RRID:AB_10693544
Chemicals, peptides, and recombinant proteins		
DMEM	GIBCO	Cat #11995040
RPMI	GIBCO	Cat #11875085
F12	GIBCO	Cat #11765047
Trypsin	GIBCO	Cat #25200056
DMSO	Sigma-Aldrich	Cat #D8418
DMF	Sigma-Aldrich	Cat #227056
DAPI	Invitrogen	Cat #D1306
GENEzol	Geneaid	Cat #GZR100
A443654	MedChemExpress	Cat #HY-10425
Capmatinib	Novartis	Cat #HY-13404
Alpelisib (BYL719)	Novartis	Cat #S1105
LCL161	Novartis	Cat #HY-15518
LEE011	Novartis	Cat #HY-15777
Infigratinib	Novartis	Cat #HY-13311
(+)-JQ1	Selleck Chemicals	Cat #S7110
17-AAG (Tenespimycin)	Selleck Chemicals	Cat #S1141
5-Fluorouracil	Selleck Chemicals	Cat #S1209
Abemaciclib	Selleck Chemicals	Cat #S7158
Afatinib	Selleck Chemicals	Cat #S1011
AGI-6780	Selleck Chemicals	Cat #S7241
Anisomycin	Selleck Chemicals	Cat #S7409
Azacitidine	Selleck Chemicals	Cat #S1782
AZD0530 (Saracatinib)	Selleck Chemicals	Cat #S1006
AZD4547	Selleck Chemicals	Cat #S2801
AZD6244 (Selumetinib)	Selleck Chemicals	Cat #S1008
AZD8055	Selleck Chemicals	Cat #S1555
BAY11-7082	Selleck Chemicals	Cat #S2913
Dactolisib	Selleck Chemicals	Cat #S1009
BGJ398	Selleck Chemicals	Cat #S2183
Birinapant	Selleck Chemicals	Cat #S7015

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bleomycin	Selleck Chemicals	Cat #S1214
Bortezomib (PS-341)	Selleck Chemicals	Cat #S1013
Bosutinib	Selleck Chemicals	Cat #S1014
Buparlisib (BKM120)	Selleck Chemicals	Cat #S2247
Busulfan	Selleck Chemicals	Cat #S1692
Caffeic Acid Phenethyl Ester	Selleck Chemicals	Cat #S7414
Camptothecin	Selleck Chemicals	Cat #S1288
CDK1/2 inhibitor III	Santa Cruz	Cat #CAS443798
Celastrol	Selleck Chemicals	Cat #S1290
Cisplatin	Selleck Chemicals	Cat #S1166
Cyclopamine	Selleck Chemicals	Cat #S1146
Dasatinib	Selleck Chemicals	Cat #S1021
Decitabine	Selleck Chemicals	Cat #S1200
Doxorubicin	Selleck Chemicals	Cat #S1208
Epirubicin	Selleck Chemicals	Cat #S1223
Erlotinib	Selleck Chemicals	Cat #S7786
Etoposide	Selleck Chemicals	Cat #S1225
Everolimus(RAD001)	Selleck Chemicals	Cat #S1120
Fedratinib	Selleck Chemicals	Cat #S2736
Fludarabine	Selleck Chemicals	Cat #S1491
Gefitinib	Selleck Chemicals	Cat #S1025
Gemcitabine	Selleck Chemicals	Cat #S1714
I-BET151 (GSK1210151A)	Selleck Chemicals	Cat #S2780
Ifosfamide	Selleck Chemicals	Cat #S1302
Imatinib (STI571)	Selleck Chemicals	Cat #S2475
Irinotecan	Selleck Chemicals	Cat #S2217
IWP-L6	Selleck Chemicals	Cat #S7301
JSH-23	Selleck Chemicals	Cat #S7351
L-685458	Selleck Chemicals	Cat #S7673
K03861	Selleck Chemicals	Cat #S8100
Lapatinib	Selleck Chemicals	Cat #S2111
LY294002	Selleck Chemicals	Cat #S1105
Masitinib	Selleck Chemicals	Cat #S1064
Methotrexate	Selleck Chemicals	Cat #S1210
MHY1485	Selleck Chemicals	Cat #S7811
Mitomycin C	Selleck Chemicals	Cat #S8146
MK2206	Selleck Chemicals	Cat #S1078
Momelotinib (CYT387)	Selleck Chemicals	Cat #S2219
Mycophenolate Mofetil	Selleck Chemicals	Cat #S1501
Nilotinib (AMN-107)	Selleck Chemicals	Cat #S1033
Nutlin-3	Selleck Chemicals	Cat #S1061
NVP-AEW541	Selleck Chemicals	Cat #S1034
Olaparib	Selleck Chemicals	Cat #S1060
Omipalisib (GSK2126458)	Selleck Chemicals	Cat #S2658
Oxaliplatin	Selleck Chemicals	Cat #S1224

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Paclitaxel	Selleck Chemicals	Cat #S1150
Pacritinib (SB1518)	Selleck Chemicals	Cat #S8057
Panobinostat (LBH589)	Selleck Chemicals	Cat #S1030
PD0325901	Selleck Chemicals	Cat #S1036
PD0332991 (Palbociclib)	Selleck Chemicals	Cat #S1579
PD318088	Selleck Chemicals	Cat #S1568
PD98059	Selleck Chemicals	Cat #S1177
PF234066 (Crizotinib)	Selleck Chemicals	Cat #S1068
PHA-665752	Selleck Chemicals	Cat #S1070
PLX4720	Selleck Chemicals	Cat #S1152
QNZ (EVP4593)	Selleck Chemicals	Cat #S490
RAF265 (CHIR-265)	Selleck Chemicals	Cat #S2161
Rapamycin(Sirolimus)	Selleck Chemicals	Cat #S1039
RG108	Selleck Chemicals	Cat #S2821
RO3306	Selleck Chemicals	Cat #S7149
Ruxolitinib(INCB018424)	Selleck Chemicals	Cat #S1378
SB202190 (FHPI)	Selleck Chemicals	Cat #S1077
SB203580	Selleck Chemicals	Cat #S1076
SH-4-54	Selleck Chemicals	Cat #S7337
Sorafenib	Selleck Chemicals	Cat #S7397
SP600125	Selleck Chemicals	Cat #S1460
Staurosporine	Selleck Chemicals	Cat #S1421
Sunitinib	Selleck Chemicals	Cat #S7781
TAE684	Selleck Chemicals	Cat #S1108
TAK733	Selleck Chemicals	Cat #S2617
TKI258, Dovitinib	Selleck Chemicals	Cat #S1018
Tofacitinib	Selleck Chemicals	Cat #S2789
Topotecan HCL	Selleck Chemicals	Cat #S1231
Trichostatin	Selleck Chemicals	Cat #S1045
Vincristine	Selleck Chemicals	Cat #S1241
Vinorelbine Tartrate	Selleck Chemicals	Cat #S4269
Vismodegib	Selleck Chemicals	Cat #S1082
Vorinostat	Selleck Chemicals	Cat #S1047
Vandetanib	Selleck Chemicals	Cat #S1046
Whole Cell Stain Green	Thermo Scientific	Cat #8403301
DAPI (4′,6-diamidino-2-phenylindole, dihydrochloride)	Invitrogen	Cat #D1306
Critical commercial assays		
Odyssey Blocking Buffer (PBS)	LI-COR	Cat #927–40000
QIAamp DNA Mini Kit	QIAGEN	Cat #51304
PureLink™ RNA Mini Kit	Invitrogen	Cat # 12183025
Qubit™ dsDNA BR Assay Kit	Invitrogen	Cat # Q32850
Qubit™ RNA BR Assay Kit	Invitrogen	Cat # Q10211
NEBNext® Ultra™ Directional RNA Library Prep Kit	New England BioLabs	Cat # E7760

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
RNA-seq data batch1 - 15 CCA cell lines - MMNK-1, MCF-7, T-47D and ZR-75-1	This paper This paper	GEO: GSE124623 GEO: GSE125034
RNA-seq data batch2 - 15 CCA cell lines	This paper	GEO: GSE125035
Selected 550 cancer driver gene mutationdata - 15 CCA cell lines	This paper	SRA: PRJNA543619
Jusakul cohort	Jusakul et al. (2017)	GEO: GSE89749
source code	Github	https://github.com/sisyspharm/ccaexplorer
Experimental models: Cell lines		
Human: HuCCA-1	JRCB cell bank	JCRB1657; RRID CVCL_M255
Human: HuCCT-1	JRCB cell bank	JCRB0425; RRID CVCL_0324
Human: HuH-28	JRCB cell bank	JCRB0426; RRID CVCL_2955
Human: KKK-D068	JRCB cell bank	JCRB1775; RRID CVCL_XD18
Human: KKK-D131	JRCB cell bank	JCRB1777; RRID CVCL_XD19
Human: KKK-D138	JRCB cell bank	JCRB1779; RRID CVCL_XD20
Human: KKU-055	JRCB cell bank	JCRB1551; RRID CVCL_M258
Human: KKU-100	JRCB cell bank	JCRB1568; RRID CVCL_3996
Human: KKU-213C	JRCB cell bank	JCRB1561; RRID CVCL_M260
Human: KKU-213A	JRCB cell bank	JCRB1557; RRID CVCL_M261
Human: KKU-213B	JRCB cell bank	JCRB1556; RRID CVCL_M264
Human: MCF-7	ATCC	ATCC® HTB-22™; RRID CVCL_0031
Human: MMNK-1	JRCB cell bank	JCRB1554; RRID CVCL_M266
Human: RBE	RIKEN cell bank	RCB1292; RRID CVCL_4896
Human: SSP-25	RIKEN cell bank	RCB1293; RRID CVCL_4902
Human: T-47D	ATCC	ATCC® HTB-133™; RRID CVCL_0553
Human: TFK-1	RIKEN cell bank	RCB2537; RRID CVCL_2214
Human: YSCCC	RIKEN cell bank	RCB1549; RRID CVCL_3629
Human: ZR-75-1	ATCC	ATCC® CRL-1500; RRID CVCL_0588
Software and algorithms		
Analysis of NGS data	Basepair software	https://www.basepairtech.com
annovar April 2018	Wang et al. (2010)	http://annovar.openbioinformatics.org/en/latest/
Bioconductor package GRmetrics (v 1.16.0)	(Clark et al., 2017)	http://bioconductor.org/packages/release/bioc/html/GRmetrics.html
Bioconductor package sva (v3.8)	(Leek et al., 2012)	https://bioconductor.org/packages/release/bioc/html/sva.html
CellProfiler (v3.1.8)	(McQuin et al., 2018)	https://doi.org/10.1371/journal.pbio.2005970
defuse (v0.8.1)	(McPherson et al., 2011)	https://github.com/amcpherson/defuse
fastp (v0.19.4)	(Chen et al., 2018)	https://github.com/OpenGene/fastp
GenePattern 2.0	(Reich et al., 2006)	https://doi.org/10.1038/ng0506-500
R (v4.0.4)	R Foundation for Statistical Computing	http://www.r-project.org
R package caret (v6.0–82)	(Kuhn, 2008)	http://www.cran.r-project.org
R package clusterProfiler (v3.18.1)	(Yu et al., 2012)	http://www.cran.r-project.org
R package igraph (v1.2.6)	(Csardi and Nepusz, 2006)	http://www.cran.r-project.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R package survminer (v0.4.8)	(Kassambara et al., 2020)	http://www.cran.r-project.org
R package survival (v3.3.1)	(Therneau, 2000, 2022)	http://www.cran.r-project.org
R package stats	(R Development Core Team, 2021)	http://www.cran.r-project.org
R package TDM (v1.3.2)	(Thompson et al., 2016)	http://www.cran.r-project.org
R package tidyverse (v1.3.0)	(Wickham et al., 2019)	http://www.cran.r-project.org
sambamba (v0.6.6)	(Tarasov et al., 2015)	http://lomereiter.github.io/sambamba/
samtools (v1.6)	(Li et al., 2009)	http://samtools.sourceforge.net/
STAR (v2.5.3a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
subread (v1.6.2)	(Liao et al., 2013)	http://subread.sourceforge.net/

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the lead contact, Somponnat Sampattavanich (somponnat.sam@mahidol.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- RNA-seq data have been deposited at Gene Expression Omnibus (GEO) and selected 548 cancer driver gene mutation raw data have been deposited at GeneBank and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at Github and is publicly available as of the date of publication, which is listed in the key resources table.
- The app portal can be accessed at https://sisp.shinyapps.io/AsianCCAbrowser/.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact on request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and reagents

The different cell lines were obtained from the following sources: HuCCA-1 (JCRB1657), HuCCT-1 (JCRB0425), HuH-28 (JCRB0426), KKU-055 (JCRB1551), KKU-100 (JCRB1568), KKU-213A (JCRB1557), KKU-213B (JCRB1556), KKU-213C (JCRB1561), KKK-D068 (JCRB1775), KKK-D131 (JCRB1777), KKK-D138 (JCRB1779), RBE (RCB1292), SSP-25 (RCB1293), TFK-1 (RCB2537), YSCCC (RCB1549), MCF-7(ATCC® HTB-22™), MMNK-1 (JCRB1554), T-47D(ATCC® HTB-133™), ZR-75-1(ATCC® CRL-1500). All cell lines used in this study were tested mycoplasma-free by the PCR method. The sex of cell lines are as follows; Male: HuCCA-1, HuCCT-1, KKU-055, KKU-213A, KKU-213B, KKU-213C, KKK-D068, KKK-D068, TFK-1, and Female: HuH-28, KKU-100, KKK-D138, RBE, SSP-25, YSCCC, MCF-7, T-47D, and ZR-75-1. We cultured KKU-055, KKU-213A, KKU-213B, KKU-213C, KKK-D068, KKK-D131, KKK-D138 HuCCT-1, HuH-28, and MCF-7 in DMEM (Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin (Gibco). Six cell lines (RBE, SSP-25, T47-D, TFK-1, YSCCC, and ZR75-1) were cultured in RPMI (Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin (Gibco). KKU-100 HuCCA-1 and MMNK-1 were grown in Ham-F12 (Gibco) supplemented with 10%FBS and 1% penicillin streptomycin (Gibco).

METHOD DETAILS

DNA preparation and cancer gene panel sequencing

The genomic DNA was extracted using the QIamp DNA extraction kit (Qiagen). Using Qubit dsDNA HS assay kit and agarose gel electrophoresis, extracted DNA was quantified for extraction yield and quality. Target sequencing of DNA samples was performed using the NovoPM^{TM} cancer panels (NovogeneAIT

Singapore Genomics Pte Ltd) comprising the coding region of 548 cancer-related genes and 21 non-coding regions (paired-end 150bp with depth >1000X). GATK best practice was used for calling single nucleotide polymorphisms (SNPs) and small insertions and deletions (small InDels), and ANNOVAR was used for variant annotation (Wang et al., 2010). Variant filtering was performed to exclude low-quality data (variant depth >50, excluding non-synonymous variants, frequency in ExAC, 1000Genomes $\geq$0.1, ESP <0.1, and an allelic fraction >0.25).

RNA preparation, sequencing, and transcriptome profiling

RNA sequencing was performed for all cell lines, including 15 CCA cell lines, MMNK-1, and breast cancer cell lines (MCF-7, T-47D, and ZR-75-1). The CCA cell lines and MMNK-1 were sequenced twice, whereas the breast cancer cell lines were sequenced once. RNA was extracted using Trizol, purified by PureLink™RNA isolation kit (Invitrogen), and treated with PureLink DNase set (Invitrogen) to remove DNase. The amount of RNA for each sample was measured using the Qubit RNA HS assay kit (Invitrogen). The RNA integrity was analyzed using Agilent 2100 Bioanalyzer (Agilent Technologies). All the samples had high RNA integrity (RNA integrity number (RIN) > 8). RNA sequencing libraries were prepared with Illumina-compatible NEBNext® Ultra™ Directional RNA Library Prep Kit (New England BioLabs) and sequenced at Genotypic Technology Pvt. Ltd. In brief, 1 μg RNA was subjected to fragmentation and priming. The first-strand synthesis was performed in the presence of Actinomycin D (Gibco, Life Technologies), followed by the second-strand synthesis. After cDNA purification and adapter ligation, the purified sequencing libraries were quality-checked using Qubit Fluorometer (Thermo Fisher Scientific) Agilent 2200 Tapestation. Transcriptome sequencing was performed using Illumina Hiseq2000 (Illumina) with 150 × 2 paired-end chemistry reads and sequenced at 20–25 million reads per sample.

Sequencing results were aligned to the human reference (hg19) using the STAR method version v2.5.3a (Dobin et al., 2013) at the Basepair website (https://www.basepairtech.com/). Other pipelines used for the analysis of the transcripts included 'sambamba' (v0.6.6) (Tarasov et al., 2015), 'samtools' (v1.6) (Li et al., 2009), 'fastp' (v0.19.4) (Chen et al., 2018), and 'subread' (v1.6.2) (Liao et al., 2013). Low-quality base pairs with a Phred quality score of less than 10 were removed (McPherson et al., 2011). The 'deFuse' algorithm (v0.8.1) (Williams et al., 2016) was used to identify fusion transcripts with the hg19 reference genome. The concordance of fusion genes was calculated from those found in both technical replicates.

Transcription per million (TPM) for each sample was used for gene expression analysis. Genes with a relatively low expression (median of TPM = 0) across all cells were removed and transformed to log scale [log2 (TPM+1)]. Batch effect removal was performed on two different sequencing batches of CCA cell line using the 'sva' package (Leek et al., 2012). The corrected dataset was then assessed using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) (Figure S8A). Spearman's correlations of each technical replicate after batch effect removal were shown in Figure S8B (median 0.88, IQR = [0.84–0.88], minimum = 0.81, maximum = 0.89), which were comparable with a previous study of RNA sequencing (Shin et al., 2014)

To compare the basal gene expressions of our CCA cell lines against other cells from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), we transformed expression levels to normalized z-score. As a result, there were 12,058 genes in common from both data sets. We then applied t-SNE (Maaten and Hinton, 2008) to visualize the cell lines. K-nearest neighbor clustering was used to examine the hepatic/pancreatic similarity between our CCA cell lines and the CCLE cells. In addition, we selected the top 1,000 genes by variance and performed hierarchical clustering using Spearman's correlation (Figure S1).

Drug response profiling and comparison and network analysis

The responses of 15 CCA cell lines to 100 compounds were collected using an automated high-content imaging system. Briefly, cell lines were plated in two biological replicates in 96-well plates (Corning cat#3904) using the complete medium described previously. After 24 h, all cells were at 70% confluency and subjected to treatment with a specific compound for another 72 h. Each drug was prepared using a two-fold serial dilution method to generate 11 concentrations ranging from 10 nM to 10 μM, and 0.5% (v/v) of solvent was used as a control. After 72 h of treatment, cells were fixed in 4% paraformaldehyde and washed with PBS containing 0.1% of Tween 20. Nuclei were stained in DAPI (1:1000 dilution, Invitrogen) for 1 h and washed with PBS twice. Next, images of nuclei were collected using a 10× objective lens on

Operetta HCS (Perkin Elmer), followed by nucleus segmentation and counting using Columbus software (Perkin Elmer).

To minimize the effect of heterogeneous cell division rate across cell lines, we adopted the growth rate inhibition (GR) parameters, such as GR_{50} (the concentration that inhibits growth by 50%) and GR_{max} (the effect at the highest tested dose). These parameters were calculated using the 'Grmetrics' package from Bioconductor for R (Hafner et al., 2016). When the underlying fitted parameters were beyond our dose range, we set the highest and lowest concentrations to 10 μM and 10 nM, respectively. We transformed GR50 to median-centered log10(GR50) across all 15 CCA cell lines to find drug response similarities between cell lines. We then compared each drug pairwise using Spearman's correlation function 'cor' from the 'stats' R package.

We determined a therapeutic index network to showcase a drug relationship and determined the potential drug candidates for each CCA subgroup using the 'iGraph' (v1.2.6) (Csardi and Nepusz, 2006) package from R. Based on the correlation matrix of drug response similarity, the size of drug nodes in the network was calculated by subtracting mean GR_{50} of the drug in each CCA subgroups with its population mean of GR_{50}, followed by normalization by the mean IQR of the three conventional standard-of-care drugs (gemcitabine, Cisplatin, and 5FU; see Figure S3C). The drug nodes were colored as follows: cyan for CCA1 and orange for CCA2 subgroups. The edge thickness represented the similarity level (Spearman's rank-order correlation value greater than 0.6) of the drug response pattern between each drug node pair. Red and blue edge colors represented positive and negative correlations, respectively. The landscape of the network was generated using a force-directed Fruchterman-Reingold algorithm.

Drug sensitivity determination, pathway analysis, and drug response biomarkers prediction

We used hierarchical clustering for grouping CCA subgroups based on drug potency (GR_{50}). To identify the drug sensitivity group of cells for each drug, we used median-centered GR_{50} and applied kernel density to estimate the intersection between two Gaussian kernels using the 'density' function from the 'stats' package for R. As a result, we could assign to the cells drug-sensitive or resistant label. Any drugs whose variance was less than the 20^{th} percentile of the median-centered GR_{50} across all cell lines were excluded from further analysis, resulting in 80 remaining drugs.

After determining drug response subgroups, we began selecting candidate biomarkers from the leading-edge gene list of each significantly enriched cancer hallmark gene set in each drug response subgroup. To overcome the small sample size bias, we generated 100 sets of discovery-validation pairs. The discovery sets consisted of six cell lines; three were randomly selected from the drug-sensitive group and the other three from the drug-resistant group. The remaining cell lines from each group were combined and used as the validation set. Next, ten randomly selected discovery sets were tested in different models (random forest, generalized linear model, and stochastic boosting) and evaluated for prediction performance in the drug response subgroup of the models. We found that the random forest model (AUC = 0.875; Figure S6) outperformed other models (data not shown) and subsequently was selected as our primary classification tool. For the biomarker selection process, we first built a control model using the 'trainControl' function from the 'caret' package (v 6.0–82) (Kuhn, 2008) for R using the leave-one-out cross-validation technique. The training model was then constructed from a discovery set and evaluated against its validation set for prediction performance. The classification performance of the biomarkers was reported as an area-under-curve of receiver operation characteristic (AUC ROC). For candidate biomarkers, any genes with AUC ROC above our selection criteria (ROC AUC >0.8 and VIP >80) were chosen. In addition, we observed variable Importance in Projection (VIP) scores of the biomarkers and found that they ranged from 80 to 100 (data not shown), suggesting their contribution to drug sensitivity prediction. Finally, the biomarkers were further shortlisted based on their expression level and whether they were differentially expressed in the two subgroups (Mann-Whitney p values < 0.05).

To identify enriched hallmark gene sets in each CCA cell line, we converted the expression levels (in TPM) of genes from a particular cell line to the fold-change of gene expression levels against the population-median expression levels of the same genes. These genes were pre-ranked in descending order and analyzed for the gene set enrichment analysis (GSEA) (Subramanian et al., 2005) against 50 cancer hallmark gene sets downloaded from Broad Institute's Molecular Signatures Database (MSigDB gene set collection version 6.2) (Liberzon et al., 2015), and comprehensively profiled using the 'clusterProfiler' R package (v 3.18.1)

(Yu et al., 2012). Significantly enriched hallmark gene sets were determined using the following criteria: Benjamini-Hochberg (BH)-adjusted p value < 0.05 and FDR q-value < 0.25. If the BH-adjusted p value cutoff <0.05 did not hold any gene sets, we adjusted the BH-adjusted p value to 0.1. We first converted the median expression levels to fold-change levels of drug-sensitive over the drug-resistant subgroup (gene-wise) to identify the enriched hallmark gene sets for the CCA45 signature. A similar approach was also used to identify enriched hallmark gene sets for the predictive biomarkers of the individual drugs.

Association between drug sensitivity and genomic alteration

To explore how genomic altered genes affected drug sensitivity in Asian CCA, we analyzed the association of 189 genomic-altered genes and drug responses of CCA cell lines to 100 anti-cancer drugs. Any drug with a standard deviation lower than zero and two CCA cell lines, KKU-213B and KKU-213C, were excluded from this analysis. For each drug, we log-transformed GR_{50} values to z-score and separated the cells into two groups with or without mutation of the particular gene. In the mutant group, we calculated the median of the z-score. The mutation is annotated as a drug resistance-associated mutation if the median value is positive. In contrast, if the median value is negative, the mutation is drug sensitive-associated mutation. Known mutation and drug response association from the Genomics of Drug Sensitivity in Cancer (GDSC) database was used as a reference (Yang et al., 2013). Each mutation-drug pair was compared against the 25^{th} and 75^{th} percentile of the population's drug potency. The Wilcoxon rank sum test was used to evaluate the significance of the association to either drug sensitivity or drug resistance. The drug-mutation association pairs with p-value < 0.05 and the median z-score of drug potency in mutation-containing cell lines >0.5 (for drug resistance-associated mutations) or < −0.5 (for drug sensitivity-associated mutations) were considered significant.

Ligand response profiling and analysis

Eleven CCA cell lines were seeded at 10,000 cells/well in 96-well plates (Corning cat#3904) using the complete medium described previously. After 24 h, when all cells were at least 70% confluency, they were treated with 100 ng/mL of EGF, IGF1, Anisomycin, TNFα, or IFNγ for 10, 30, and 60 min, whereas untreated cells were used as a baseline. After the treatment was finished, the medium was completely removed from the cells, and they were fixed with 100 μL of 4% paraformaldehyde for 30 min, followed by washing twice with PBS. The cell membrane was permeabilized using 100 μL of absolute methanol for 1 h at −20°C, followed by washing twice with PBS. Cells were incubated with 100 μL of blocking buffer (Li-COR Odyssey Blocking Buffer PBS, cat#927–40000) for 1 h at room temperature, then washed twice with PBS. Diluted antibodies in blocking buffer (1:1000 dilution) were added to the cells and incubated in the dark for 1 h at room temperature, followed by washing twice with PBS containing 0.05% of Tween 20. Cells were further stained using the Whole Cell Stain Green (1:1000 dilution; Thermo Scientific, cat#8403201), and nuclei were stained using DAPI (1:1000 dilution, Invitrogen) for 1 h, following washing twice with PBS. The fluorescence images of proteins of interest, whole cells, and nuclei were collected using a 10× objective lens on Operetta HCS (Perkin Elmer). The images were analyzed for cell segmentation and intensity measurement using a customized analysis pipeline in CellProfiler software (v 3.1.8) (McQuin et al., 2018).

For ligand response analysis, the raw intensity of proteins of interest at the single-cell level was converted to a $\log_{10}$ value and projected as a heat-density plot. The median intensity level at each time point and the baseline were then calculated and further used to calculate area-under-the-curve (AUC), where positive AUC indicates response above baseline and negative AUC shows response below the baseline. Logistic regression was implemented using the 'glm' function from the 'stats' R package to determine the association between the ligand-induced pathway activation and CCA subgroups. Similarly, the principal component analysis was performed by implementing the 'prcomp' function from the 'stats' R package.

We performed analysis of the association between CCA subgroups and signal activation using odds ratio (probability of being CCA1 over probability of being CCA2) that is calculated using the log of the coefficient derived from the generalized linear model used to assess the relationship between CCA subtypes (CCA1 or CCA2) and ligand-mediated activation of key proteins (pERK, pS6K, pSTAT1, pAKT, or RelA). A p value < 0.05 indicates a significant ligand-activated signaling pathway. Error bars represent the interquartile range of the odds ratio.

We performed a systematic analysis by combining 189 genomic-altered genes of CCA cell lines with ligand responsiveness to gain insights into explaining the different signaling cascades between two CCA

subtypes. AUC of ligand response levels was normalized within each ligand and each signaling protein across all CCA cell lines using the scale function in 'stat' R package. We then combined and averaged the scaled AUC of KKU-213A, KKU-213B, and KKU-213C to avoid bias in the calculation. The differences between CCA cell lines with unique mutations were scored using the Wilcoxon rank sum test. We determined the mean difference by calculating the difference score, which is the mean difference between the means of testing groups. All figures were drawn using tidyverse (v1.3.0) (Wickham et al., 2019) package run in R (v 4.0.4) and RStudio (v 1.4.1103) software.

To examine suppression efficiency of MEK inhibitor (PD-0325901) on ERK phosphorylation in CCA cells, we selected KKU-213C (with mutant *KRAS*) and KKK-D138 (with wild-type *KRAS*) cell lines as representatives of CCA1 subgroup and KKU-100 (with mutant *KRAS*) and SSP-25 (with wild-type *KRAS*) as representatives of CCA2 subgroup. Cells were plated at 3,000 cells per well in 384-well plates (Perkin Elmer cat#6057308) in complete medium (with 10% FBS) as previously mentioned. After 24 h, the cells were treated with high dose (5 μM) and low dose (1.25 μM) of MEK inhibitor (PD-0325901) and 0.5% (v/v) DMSO as vehicle control for 3 h. Then, the cells were removed of culture medium, fixed, and stained with phospho-p44/42 (pERK1/2) antibodies to detect ERK activity and DAPI to detect nuclei. Immunofluorescent images of the cells were acquired by high-content imager using 10× objective lens on Operetta HCS (Perkin Elmer). The images were segmented and nuclear intensity of pERK in cells was quantified using the pipeline in CellProfiler software (v 3.1.8) as described previously. The nuclear intensity levels of pERK were averaged per well and normalized to the intensity from the controls and reported as relative intensity of nuclear pERK1/2. The one-way ANOVA test was performed to determine the statistical significance of the difference in relative pERK intensity between MEKi treated cell lines and DMSO control.

Clinical data selection and normalization

Clinical data containing patient survival time, molecular clusters, and tumor gene expression profiles were collected from the Gene Expression Omnibus database (GSE89749) (Jusakul et al., 2017). Only patients with assigned molecular clusters were included in further analysis. To enable a cross-platform comparison of RNA sequencing data from our CCA cell lines and clinical samples, we normalized the gene expression levels using the training distribution matching (TDM) R package (v 1.3.2) (Thompson et al., 2016).

Subclass mapping

To find the association between subclasses observed in our CCA drug response subtypes from CCA cell lines and the molecular clusters from CCA patients, we used the unsupervised subclass similarity evaluation tool, SubMap, from Broad Institute's GenePattern (Hoshida et al., 2007; Subramanian et al., 2005). Briefly, we prepared two data inputs: a) gene expression and b) subtype data. In the gene expression data, any genes with low expression levels (TPM = 0 or less than 10th percentile) across 50% of the samples in each data set were omitted, leaving 9,369 and 12,790 genes remaining in both patient and cell line data sets, respectively. In subtypes data, we labeled molecular clusters I to IV to samples in the patient dataset and CCA drug response subtypes 1 and 2 in our CCA cell line dataset, generating 4 × 2 subclass pairs. For subclass mapping analysis, we followed the procedure described previously (Hoshida et al., 2007). The nominal p-value was calculated to compare the significance of biomarker enrichment in the up-regulated end of the ranking list using GSEA. This process was repeated for every possible pairwise combination of subclasses. The significance of the association between subclasses was then estimated using the Fisher inverse chi-square statistic (F). As a result, a subclass association matrix (SA) was reported with Bonferroni-adjusted p-values. Any pairs with a p-value less than 0.05 were considered significantly similar.

Prediction of drug response subtype in CCA cell lines and patients

We assembled the gene expression data of the 45 genes from the CCA45 signature from the cell lines and the patient cohort to assign the drug response-based subgroups in CCA patients. Unsupervised hierarchical clustering of both datasets was then performed using the Euclidean distance and the Ward.D2 method. The coefficient values of the CCA45 signatures were determined using ridge regression. We calculated the weighted sum model score (WSM score) to help assign the drug-response based CCA subgroups for the individual patients, using the following equation:

$$P_i^{WSM-score} = \sum_{j=0}^{n} w_j p_{ij}$$

where P_i is the WSM score of patient i, w_j is the coefficient for j gene, and p_{ij} is the expression level of gene j of patient i. Applying these coefficient values to the generalized linear model, we could predict drug response subtypes of 15 CCA cell lines with ROC AUC = 1.0 and CCA patients with ROC AUC = 0.89.

The landscape of CCA drug response subtypes (including data from cell lines and patients) was projected based on the dimensional reduction of their gene expression levels using PCA. We identified the boundary line dividing CCA1 and CCA2 zones using a supervised, non-linear support vector machine (SVM).

Prognostic evaluation of the biomarkers

To evaluate the prognostic property of CCA drug response biomarkers, we performed a survival analysis using clinical data of CCA patients. Patients with missing survival time, vital status, and/or molecular clusters were removed from this analysis, leaving 100 patients. Significant differences between survival times of CCA1 and CCA2 subgroups of patients were calculated using the log-rank test. The p-value<0.05 indicated a significant difference. The analysis was performed using 'survival' (v3.3.1) (Therneau, 2000, 2022) and 'survminer' (v0.4.9) (Kassambara et al., 2020) R packages to estimate survival time, 1-year and 3-year survival rates of patient groups.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were analyzed using R (http://www.R-project.org/, version 4.0.4) (R Development Core Team, 2021). Statistical tests for each analysis can be found listed in each figure legend. We use spearman correlation to compare the drug response similarity (Figure 3C). Differences between groups were examined by a non-parametric Wilcoxon rank-sum test or Mann-Whitney U-test where appropriate, (Figure 5A–5C, 6D and 6E, and 7A). The association between CCA subgroups and ligand-induced signal activation using odds ratio (Figure 6C). The one-way ANOVA test was performed to determine the statistical significance of the difference in relative pERK intensity between MEKi treated cell lines and DMSO control (Figure S5D). The significance of the association between subclasses observed in our CCA drug response subtypes from CCA cell lines and the molecular clusters from CCA patients was calculated using the Fisher inverse chi-square statistic (F). A subclass association matrix (SA) was examined with Bonferroni-adjusted p values (Figure S7A). Any pairs with a p value less than 0.05 were considered significantly similar. Significant differences between survival times of CCA1 and CCA2 subgroups of patients were calculated using the log-rank test (Figure 8B and S7D and S7E). In all analyses, $p < 0.05$ were considered to be significant.