Gallbladder adenocarcinomas undergo subclonal diversification and selection from precancerous lesions to metastatic tumors

Minsu Kang[1†], Hee Young Na[2†], Soomin Ahn[3]*, Ji-Won Kim[1,4]*, Sejoon Lee[5], Soyeon Ahn[6], Ju Hyun Lee[1], Jeonghwan Youk[1], Haesook T Kim[7], Kui-Jin Kim[8], Koung Jin Suh[1], Jun Suh Lee[9], Se Hyun Kim[1], Jin Won Kim[1], Yu Jung Kim[1], Keun-Wook Lee[1], Yoo-Seok Yoon[9], Jee Hyun Kim[1], Jin-Haeng Chung[2], Ho-Seong Han[9], Jong Seok Lee[1]

[1]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [2]Department of Pathology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [3]Department of Pathology and Translational Genomics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; [4]Genealogy Inc, Seoul, Republic of Korea; [5]Center for Precision Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [6]Medical Research Collaboration Center, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [7]Department of Data Science, Dana Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, United States; [8]Biomedical Research Institute, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [9]Department of Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

*For correspondence:
suminy317@gmail.com (SA);
jiwonkim@snubh.org (J-WK)

[†]These authors contributed equally to this work

Abstract We aimed to elucidate the evolutionary trajectories of gallbladder adenocarcinoma (GBAC) using multi-regional and longitudinal tumor samples. Using whole-exome sequencing data, we constructed phylogenetic trees in each patient and analyzed mutational signatures. A total of 11 patients including 2 rapid autopsy cases were enrolled. The most frequently altered gene in primary tumors was *ERBB2* and *TP53* (54.5%), followed by *FBXW7* (27.3%). Most mutations in frequently altered genes in primary tumors were detectable in concurrent precancerous lesions (biliary intraepithelial neoplasia [BilIN]), but a substantial proportion was subclonal. Subclonal diversity was common in BilIN (n=4). However, among subclones in BilIN, a certain subclone commonly shrank in concurrent primary tumors. In addition, selected subclones underwent linear and branching evolution, maintaining subclonal diversity. Combined analysis with metastatic tumors (n=11) identified branching evolution in nine patients (81.8%). Of these, eight patients (88.9%) had a total of 11 subclones expanded at least sevenfold during metastasis. These subclones harbored putative metastasis-driving mutations in cancer-related genes such as *SMAD4*, *ROBO1*, and *DICER1*. In mutational signature analysis, six mutational signatures were identified: 1, 3, 7, 13, 22, and 24 (cosine similarity >0.9). Signatures 1 (age) and 13 (APOBEC) decreased during metastasis while signatures 22 (aristolochic acid) and 24 (aflatoxin) were relatively highlighted. Subclonal diversity arose early in precancerous lesions and clonal selection was a common event during malignant transformation in GBAC. However, selected cancer clones continued to evolve and thus maintained subclonal diversity in metastatic tumors.

Editor's evaluation

This is the first dedicated study of clonal evolution in gallbladder cancer that involves precancerous, transformed primary and metastatic lesions. The main insights include the finding of subclonal diversity in precancerous lesions, a degree of bottlenecking during transformation but maintenance of some clonal complexity through metastases.

Introduction

Gallbladder adenocarcinoma (GBAC) is a malignant neoplasm that has a high incidence rate in Chile, India, Poland, Pakistan, Japan, and Korea (*Roa et al., 2022*; *Valle et al., 2021*; *Bray et al., 2018*; *Bridgewater et al., 2014*). Surgery is currently the only curative treatment modality for GBAC. However, because most patients are diagnosed at an advanced stage and thus inoperable, they receive palliative chemotherapy only. Therefore, the prognosis is poor with a median overall survival of only 11–15 months (*Roa et al., 2022*; *Valle et al., 2021*; *Valle et al., 2010*).

The recent advancement of massively parallel sequencing technology has enabled us to deeply understand the genome of a variety of cancers. In GBAC, tumor suppressor genes such as *TP53*, *ARID1A,* and *SMAD4* and oncogenes such as *ERBB2* (*HER2*), *ERBB3*, and *PIK3CA* are significantly mutated (*Roa et al., 2022*; *Nepal et al., 2021*; *Wardell et al., 2018*; *Li et al., 2019*; *Nakamura et al., 2015*; *Lin et al., 2021*; *Narayan et al., 2019*; *Li et al., 2014*). Of note, *ERBB2* amplification and overexpression occur in approximately 6.9–28.6% of GBAC (*Roa et al., 2022*; *Narayan et al., 2019*; *Nakazawa et al., 2005*; *Nam et al., 2016*) and may have therapeutic implications.

Cancer cells undergo clonal evolution by acquiring additional mutations and thus exhibit more aggressive phenotypes, including invasion and metastasis (*Turajlic et al., 2019*; *Greaves and Maley, 2012*; *Davis et al., 2017*). Several large-scale studies have provided evidence of clonal evolution in some cancer types, including lung and kidney cancers (*Jamal-Hanjani et al., 2017*; *Turajlic et al., 2018*). However, no study has analyzed the patterns of clonal evolution from the initiation of carcinogenesis to distant metastasis in patients with GBAC.

This study aims to analyze the clonal evolutionary trajectories during carcinogenesis and metastasis of GBAC using multi-regional and longitudinal specimens including precancerous lesions (biliary intraepithelial neoplasia [BilIN]), primary tumors, and metastatic tumors from patients who underwent biopsy, surgery, and rapid autopsy.

Table 1. Baseline characteristics of 11 patients with GBAC.

Patient ID*	Sex	Age at diagnosis	ECOG PS at diagnosis	Stage at diagnosis	Differentiation
GB-A1	F	70	1	IV	PD
GB-S1	F	66	1	IV	MD
GB-S2	F	66	1	IV	MD
GB-S3	F	75	1	III	MD
GB-A2	M	61	1	IV	MD
GB-S4	M	72	1	IV	MD
GB-S5	F	70	1	IV	PD
GB-S6	M	70	0	IV	MD
GB-S7	F	74	1	IV	PD
GB-S8	M	67	1	III	WD
GB-S9	M	59	0	IV	MD

*In patient ID, 'A' indicates autopsy cases whereas 'S' indicates surgery cases.

GBAC = gallbladder adenocarcinoma. ECOG PS = Eastern Cooperative Oncology Group performance status. M = male. F = female. PD = poorly differentiated. MD = moderately differentiated. WD = well-differentiated.
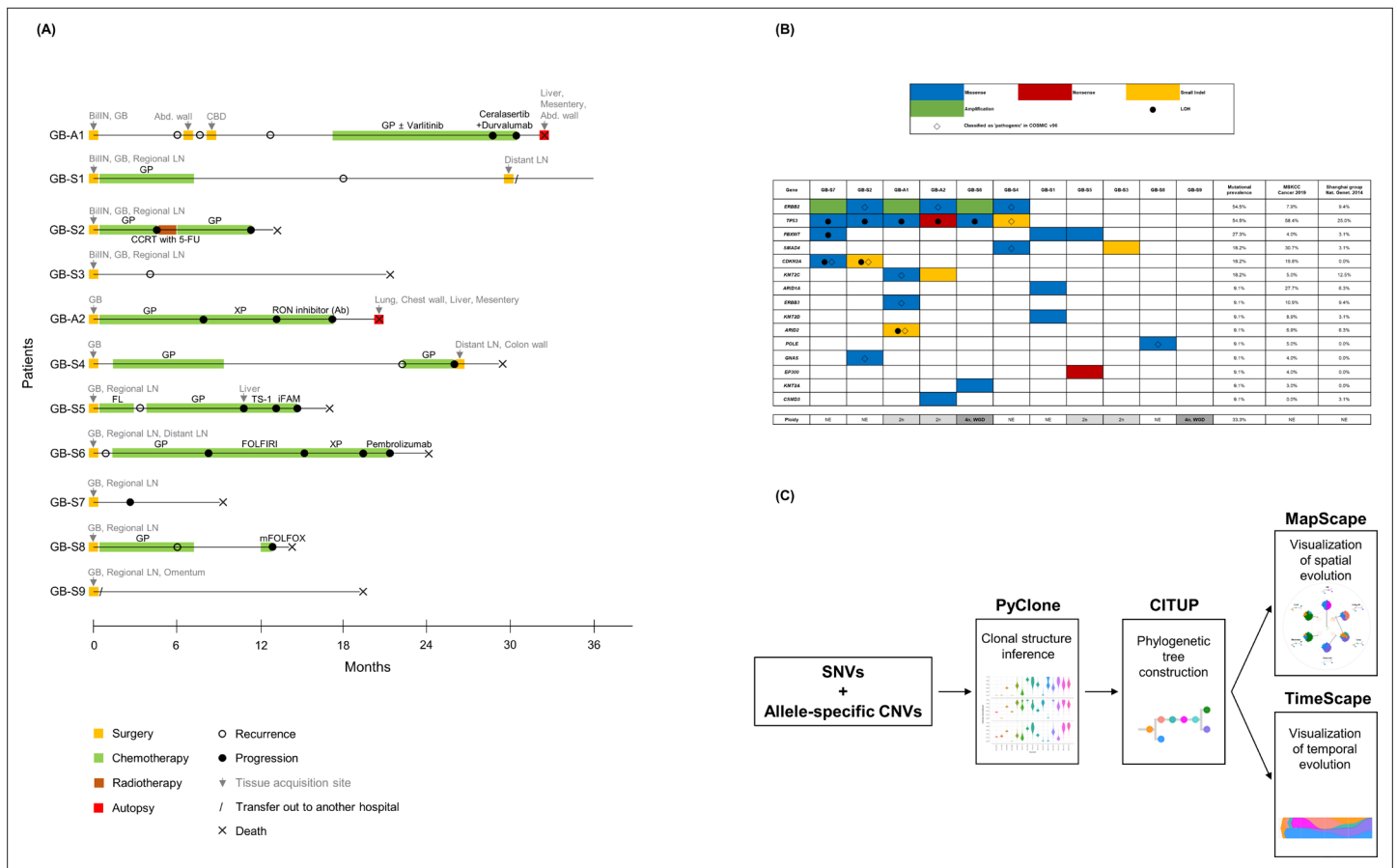
Figure 1. Clinical history of patients, mutational landscape, and study workflow. (**A**) Clinical history of 11 patients is summarized in the swimmer plot. (**B**) The mutational landscape of 11 primary tumors (GB) is visualized and compared with the two previous studies on gallbladder adenocarcinoma (GBAC). (**C**) The process of constructing clonal evolution trajectories using multiple tumor samples is shown in the workflow. Ab, antibody; Abd, abdominal; Adj, adjuvant; BilIN, biliary intraepithelial neoplasia; CBD, common bile duct; CCRT, concurrent chemoradiotherapy; CNVs, copy number variations; COSMIC, Catalogue Of Somatic Mutations In Cancer; FL, 5-fluorouracil + leucovorin; FOLFIRI, 5-fluorouracil + leucovorin + irinotecan; GB, gallbladder; GP, gemcitabine + cisplatin; iFAM, infusional 5-fluorouracil + doxorubicin + mitomycin-C; LOH, loss of heterozygosity; mFOLFOX, modified 5-fluorouracil + leucovorin + oxaliplatin; MSKCC, Memorial Sloan Kettering Cancer Center; NE, not evaluable; SNVs, single nucleotide variants; WGD, whole genome doubling; XP, capecitabine + cisplatin; 5-FU, 5-fluorouracil.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Inference of clonal structure by PyClone algorithm.

Figure supplement 2. BilIN and primary gallbladder adenocarcinoma (GBAC) of the GB-S7 patient presumed to be derived from different origins.

Results

Baseline characteristics of patients with GBAC

A total of 11 patients, including 2 rapid autopsy cases (GB-A1 and GB-A2) and 9 surgery cases (GB-S1 – 9), were enrolled in this study (*Table 1* and *Figure 1A*). There were 5 male and 6 female patients with a median age was 70 years (range, 59–75 years). Two patients had stage III and nine patients had IV disease at diagnosis. A total of 58 samples were analyzed, including 11 pairs of matched primary tumors and normal tissues, 6 concurrent BilIN, and 30 metastatic tumors. Among them, 15 samples obtained by rapid autopsy were fresh-frozen, and 43 samples obtained by surgery or biopsy were formalin-fixed paraffin-embedded (FFPE) (*Supplementary file 1A*). The median number of filtered somatic single nucleotide variants (SNVs) and small indels was 61 (range, 12–241). The number of metastatic tumors in each patient ranged from 1 to 11.
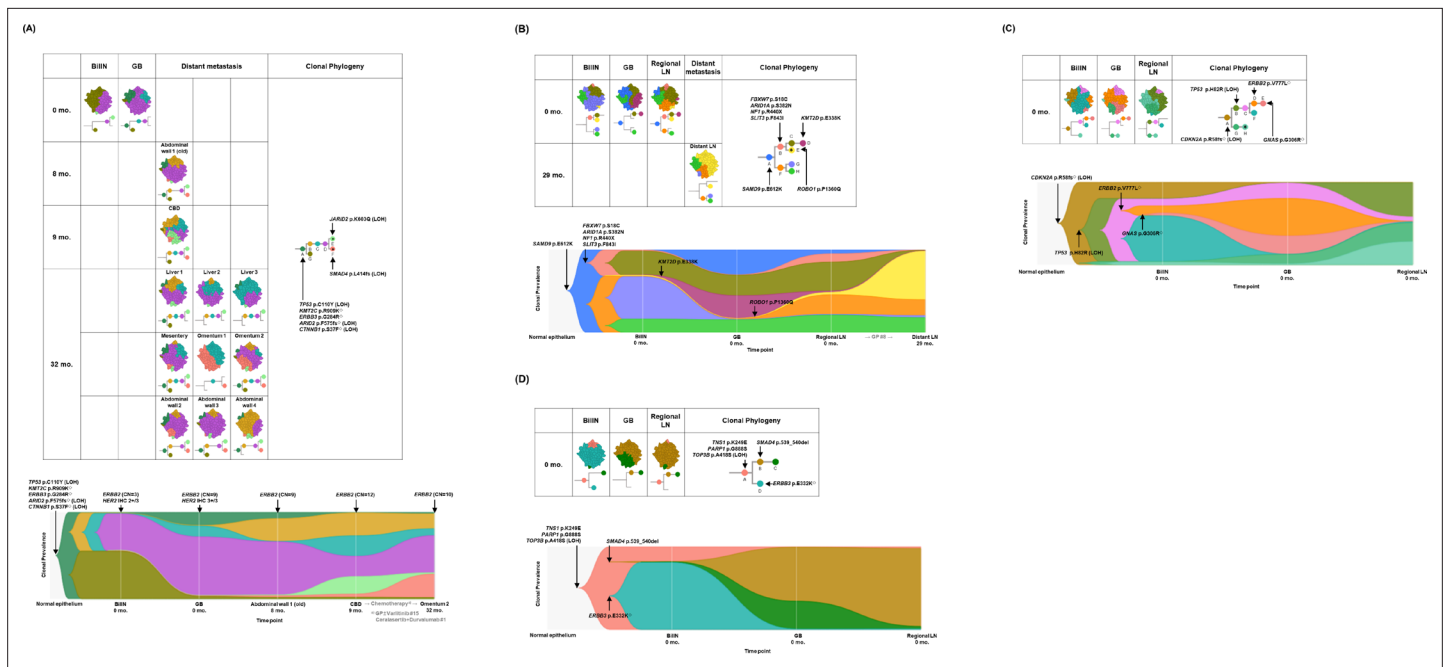
Figure 2. Spatial and temporal clonal evolution of four patients with gallbladder adenocarcinoma (GBAC) whose precancerous BilIN tissues were analyzed. (A–D) The most probable phylogenetic trees and MapScape and TimeScape results are visualized for GB-A1 (**A**), GB-S1 (**B**), GB-S2 (**C**), and GB-S3 (**D**). In MapScape visualization (table), the numbers in the row indicate the time taken from diagnosis to tissue acquisition in months, and the columns indicate whether the tissue is a precancerous lesion, primary tumor, or metastatic lesion. Colors represent distinct clones and clonal prevalences per site were proportional to the corresponding-colored area of the cellular aggregate representation. In TimeScape visualization (schema), clonal prevalences (vertical axis) were plotted across time points (horizontal axis) for each clone (colors). Asterisks (*) in the clonal phylogenetic tree denote subclones that constituted <5% in the primary tumor and expanded more than sevenfold in the metastatic tumor. Diamond symbol (◇) is used if the mutation in the figure is classified as 'pathogenic' in Catalogue Of Somatic Mutations In Cancer (COSMIC) v96. Notable events were marked with arrows. The time from diagnosis of GBAC to tissue acquisition was indicated under the sample name. Chemotherapy history was indicated in gray color, where '#' represents the number of chemotherapy cycles. BilIN, biliary intraepithelial neoplasia; CBD, common bile duct; CN, copy number; GB, gallbladder; GP, gemcitabine + cisplatin; LN, lymph node; LOH, loss of heterozygosity.

Mutational landscape and ploidy of primary tumors

The mutational landscape of the 11 primary tumors (GB) was analyzed and compared with previous literature (*Figure 1B*; *Narayan et al., 2019*; *Li et al., 2014*). The most frequently altered gene in the primary tumor was *ERBB2* and *TP53* (54.5%), followed by *FBXW7* (27.3%). Of the six *ERBB2* alterations, three were amplification and the other three were missense mutations classified as 'pathogenic' in the COSMIC (Catalogue Of Somatic Mutations In Cancer) v96 database. Among six *TP53* mutations found in six patients, five were accompanied by loss of heterozygosity (LOH).

Ploidy was analyzed in 15 tumors of six patients with purity >0.4 because ploidy estimation was inaccurate when tumor purity is ≤0.4 (*Favero et al., 2015*). In two patients (GB-S6 and GB-S9, 33.3%), whole genome doubling (WGD) was detected in both the primary and metastatic tumors (*Figure 1B*). In GB-S5 patient, WGD was found in distant metastasis, but not in primary GBAC.

Somatic mutations developed at the precancerous stage

Multi-regional distribution and longitudinal evolution of clones were analyzed using PyClone (*Figure 1—figure supplement 1* and *Supplementary file 2*; *Roth et al., 2014*) and CITUP (*Malikic et al., 2015*) and then visualized using MapScape and TimeScape (*Smith et al., 2017*; *Figure 1C*). Among six patients having concurrent BilIN tissues, two patients were excluded from the further analysis. One patient had low tumor purity of BilIN (0.03) and the other patient had different truncal mutations of BilIN and primary GBAC, suggesting different origins of the two tumors (*Figure 1—figure supplement 2*).

Most mutations in frequently altered genes in GBAC existed at the BilIN stage (10 of 13, 76.9%), but some of them were subclonal. In GB-A1 (*Figure 2A*), *TP53* C100Y with LOH, *KMT2C* R909K,

ERBB3 G284R*, ARID2* F575fs with LOH, and *CTNNB1* S37F with LOH were observed clonally in BilIN. In GB-S1 (**Figure 2B**), *SAMD9* E612K was clonal whereas *FBXW7* S18C*, ARID1A* S382N, and *NF1* R440X were subclonal. In GB-S2 (**Figure 2C**), considering the cellular prevalence of mutations, it is speculated that the mutations developed in the order of *CDKN2A* R58fs with LOH, *TP53* H82R with LOH, *ERBB2* V777L, and *GNAS* G306R during carcinogenesis. In GB-S3 (**Figure 2D**), *ERBB3* E332K was dominant in BilIN, while *SMAD4* 539_540del was not detected.

Subclonal diversity and 'selective sweep' phenomenon during the early stage of carcinogenesis

Branching evolution and subclonal diversity were commonly observed in the BilIN of the four patients (**Figure 2**). When compared with the concurrent primary tumors, one subclone commonly shrank in the primary tumors, while the other subclones that acquired additional mutations relatively expanded in the primary tumors, suggesting a selective sweep phenomenon (**Merlo et al., 2006**). Selected subclones underwent linear and branching evolution, and thus subclonal diversity was maintained after the BilIN stage. In GB-A1 (**Figure 2A**), clone A underwent branching evolution into B and G, and clone B linearly evolved into C and then D. Clone D, which acquired additional mutations, increased from 42.7% to 68.1% while clone G decreased from 55.7% to 2.8%. In GB-S1 (**Figure 2B**), clone D that acquired *KMT2D* E338K increased from 0.1% to 28.1%, while clone G decreased from 51.4% to 0%. In GB-S2 (**Figure 2C**), clone D acquired *ERBB2* V777L, clone E acquired *GNAS* G306R, and clone G increased from 9.0%, 3.5%, and 0% to 45.5%, 22.3%, and 11.4%, respectively. In contrast, clone F decreased from 55.8% to 0.6%. In GB-S3 (**Figure 2D**), clone B that acquired *SMAD4* 539_540del increased from 0.1% to 65.6%, while clone D containing *ERBB3* E332K decreased from 81.1% to 0.6%.

Table 2. List of subclones expanding during metastasis.

Patient ID	Subclone	No. of mutations	Putative driver mutations	Clonal prevalence in primary tumor	Clonal prevalence in metastasis	Metastatic organ
GB-A1	E	121	*JARID2* p.K603Q (LOH)	0.2%	20.5%	CBD
	F	54	*SMAD4* p.L414fs (LOH)	0.4%	49.8%	Omentum 1
					27.0%	Omentum 2
GB-S1	E	6	*ROBO1* p.P1360Q	0.0%	59.2%	Distant LN
GB-S2	H	6	–	1.2%	28.1%	Regional LN
GB-A2	F	14	*PRKCD* p.I153L	0.3%	73.1%	Liver
					68.3%	Mesentery
	G	3	*DICER1* p.T519A	2.8%	39.9%	Lung
					61.5%	Chest wall
GB-S4	F	32	*FBXW2* p.W450C	0.0%	34.1%	Colon wall
	H	36	–	0.0%	34.9%	Distant LN
GB-S5	B	28	*KIAA0100* p.F5S *CSMD2* p.E411K	1.2%	54.4%	Liver
GB-S6	B	7	*OSCP1* p.R351X	1.1%	60.2%	Lung
GB-S7	C	5	–	3.3%	24.7%	Regional LN

Putative driver mutations are indicated, and a full list of mutated genes is specified in **Supplementary file 2**.
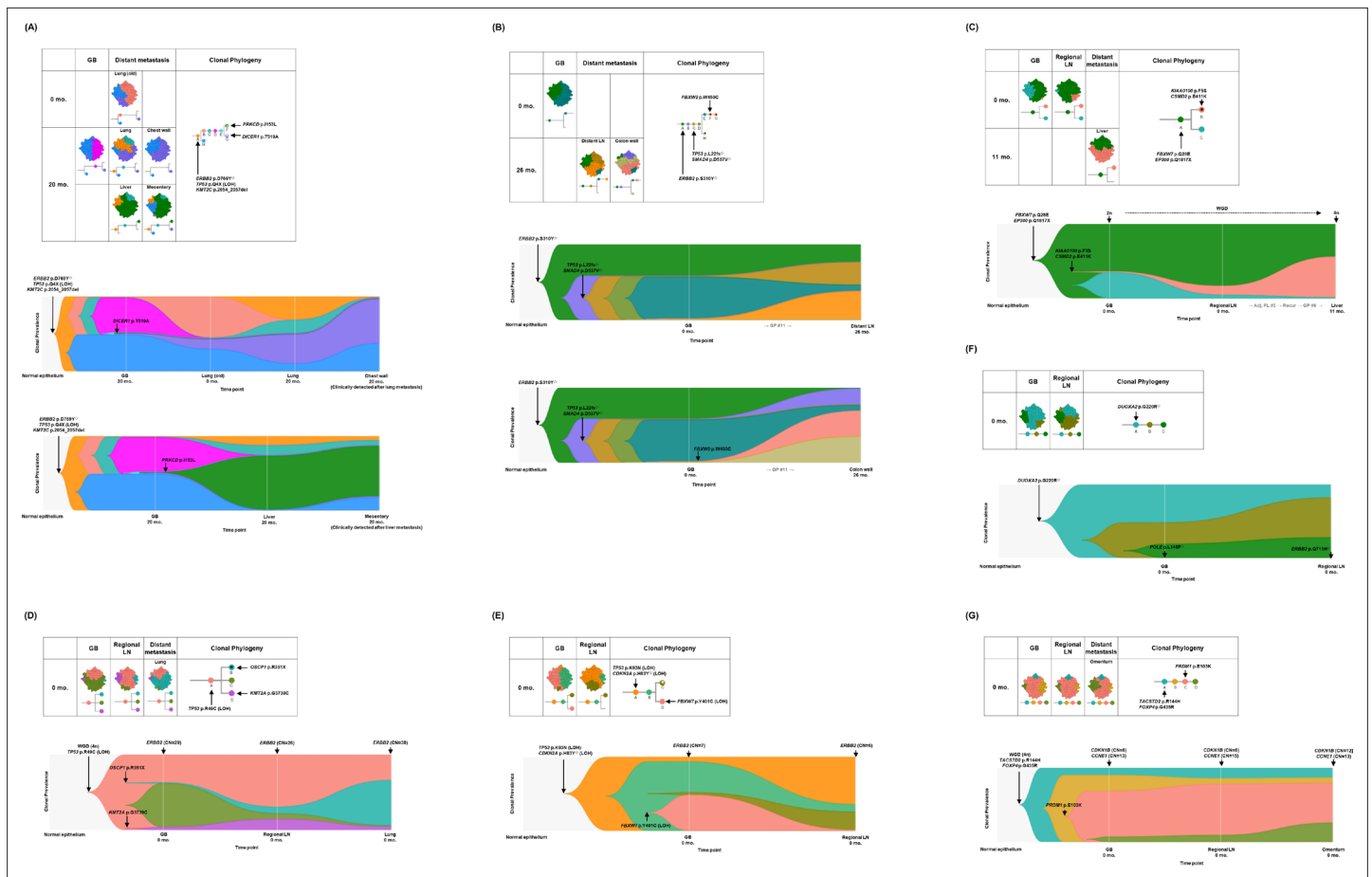GB = gallbladder. LN = lymph node. CBD = common bile duct.

Figure 3. Spatial and temporal clonal evolution of additional seven patients with gallbladder adenocarcinoma (GBAC). (**A–G**) The most probable phylogenetic trees and MapScape and TimeScape results are visualized for GB-A2 (**A**), GB-S4 (**B**), GB-S5 (**C**), GB-S6 (**D**), GB-S7 (**E**), GB-S8 (**F**), and GB-S9 (**G**). In MapScape visualization (table), the numbers in the row indicate the time taken from diagnosis to tissue acquisition in months, and the columns indicate whether the tissue is a precancerous lesion, primary tumor, or metastatic lesion. Colors represent distinct clones and clonal prevalences per site were proportional to the corresponding-colored area of the cellular aggregate representation. In TimeScape visualization (schema), clonal prevalences (vertical axis) were plotted across time points (horizontal axis) for each clone (colors). Asterisks (*) in the clonal phylogenetic tree denote subclones that constituted <5% in the primary tumor and expanded more than sevenfold in the metastatic tumor. Diamond symbol (◇) is used if the mutation in the figure is classified as 'pathogenic' in Catalogue Of Somatic Mutations In Cancer (COSMIC) v96. Notable events were marked with arrows. The time from diagnosis of GBAC to tissue acquisition was indicated under the sample name. Chemotherapy history was indicated in gray color, where '#' represents the number of chemotherapy cycles. Adj, adjuvant; BilIN, biliary intraepithelial neoplasia; CN, copy number; FL, 5-fluorouracil + leucovorin; GB, gallbladder; GP, gemcitabine + cisplatin; LN, lymph node; LOH, loss of heterozygosity; WGD, whole genome doubling.

Evolutionary trajectories and expansion of subclones during regional and distant metastasis

Combined analysis of regional and distant metastatic tumors revealed branching and linear evolution in nine patients (81.8%) and linear evolution only in two patients (18.2%). Of the nine patients with branching and linear evolution, eight (88.9%) had a total of 11 subclones expanded at least sevenfold in the regional or distant metastasis stage (*Table 2*). In GB-A1 (*Figure 2A*), clone E, which acquired *JARID2* (*Celik et al., 2018*) K603Q with LOH, increased from 0.2% to 20.5% during common bile duct metastasis. In addition, clone F, which acquired *SMAD4* (*Narayan et al., 2019*; *Jamal-Hanjani et al., 2017*; *Yoon et al., 2021*; *Zhao et al., 2018*) L414fs with LOH, expanded from 0.4% to 49.8% during omentum 1 metastasis and to 27.0% during omentum 2 metastasis. In GB-S1 (*Figure 2B*), clone B harboring *SLIT3* F843I mutation evolved into E by additionally acquiring *ROBO1* P1360Q mutation (*Gara et al., 2015*). In GB-A2 (*Figure 3A*), clone F, which acquired *PRKCD* (*Griner and Kazanietz, 2007*; *Yoshida, 2007*) I153L, expanded from 0.3% to 73.1% and 68.3% during metastasis to liver and mesentery, respectively. In addition, clone G acquired *DICER1* (*Foulkes et al., 2014*) T519A and

expanded from 2.8% to 39.9% and 61.5% during metastasis to the lung and chest wall, respectively. In the other five patients (*Figures 2C and 3B–E*), mutations in *FBXW2* (*Xu et al., 2017*), *KIAA0100* (*Gara et al., 2018*), *CSMD2* (*Gara et al., 2018*), and *OSCP1* (*Yi et al., 2017*) genes were observed.

In GB-S8 and GB-S9, only linear evolution was identified (*Figure 3F and G*). In GB-S9, amplification of cell cycle-related oncogenes *CDKN1B* and *CCNE1* was uniformly observed from primary GBAC to regional and distant metastatic tumors.

Polyclonal metastasis and intermetastatic heterogeneity

The metastatic lesions were uniformly polyclonal. In GB-A1, GB-A2, and GB-S4, which contained two or more distant metastatic lesions, the clonal composition of tissues obtained from the same or adjacent organs showed a similar tendency, while the clonal composition of anatomically distant organs was distinct from each other. In GB-A1 (*Figure 2A*), abdominal wall 1–4 did not contain clone C, and liver 1–3 did not contain clone F. In addition, omentum 1–2 had a high prevalence of clone F of over 27.0%. In GB-A2 (*Figure 3A*), as opposed to clone G, which was more prevalent in lung and chest wall lesions, clone F was more prevalent in liver and mesentery lesions. In clinical information of the GB-A2 patient, chest wall and mesentery metastases developed later than lung and liver metastases. In GB-S4 (*Figure 3B*), proportions of clones G and H were specifically high in distant LN and colon wall metastasis, respectively.

Mutational signatures during clonal evolution

We identified six mutational signatures (cosine similarity >0.9): signatures 1 (age), 3 (DNA double-strand break-repair), 7 (ultraviolet), 13 (APOBEC), 22 (aristolochic acid), and 24 (aflatoxin). Then, our results were compared with those of the Memorial Sloan Kettering Cancer Center (MSKCC) and Shanghai datasets (*Narayan et al., 2019*; *Li et al., 2014*). While MSKCC and Shanghai datasets consist merely of primary tumors, our dataset includes precancerous and metastatic lesions. Signatures 1, 3, and 13 were commonly dominant in all three datasets, while signatures 22 and 24 were exclusive in our dataset (*Figure 4A*). In our dataset, the limited number of SNVs and small indels per sample (median 61, range 12–241) made it difficult to compare among individual tumors (*Figure 4—figure supplement 1*). Therefore, we classified the mutations according to the timing of development during clonal evolution (*Figure 4B*): (1) early carcinogenesis (i.e., clone A in *Figures 2 and 3*), (2) late carcinogenesis (i.e., subclones which were not categorized in early carcinogenesis or metastasis), and (3) metastasis (i.e., subclones expanding during metastasis [*Table 2*]). At the metastasis phase, signatures 1 and 13 decreased while signatures 22 and 24 increased compared with early and late carcinogenesis. Then, the same analysis was performed according to the type of sample (*Figure 4—figure supplement 2*): (1) BilIN, (2) GB, (3) regional LN metastasis, and (4) distant metastasis. This mutational signature analysis was conducted using the published tool Mutalisk (*Lee et al., 2018*) and then validated with two additional tools, Signal (*Degasperi et al., 2020*) and MuSiCa (*Díaz-Gay et al., 2018*; *Supplementary file 1B* and *Figure 4—figure supplement 3*).

ERBB2 amplification during clonal evolution

In GB-A1, *ERBB2* copy number gain might have initiated from the precancerous stage and further progressed during the malignant transformation of BilIN. *ERBB2* copy number (*Figure 5A and B*) was 3 and 9 in concurrent BilIN and primary tumors, respectively. The increased copy number of *ERBB2* was maintained after distant metastasis (*Figure 2A*). *HER2* (=*ERBB2*) silver in situ hybridization (SISH) (*Figure 5C and D*) was conducted and *ERBB2* copy number per cell ranged from 1 to 5 in BilIN, and from 2 to 14 in primary GBAC. *ERBB2*/CEP17 ratio was 2.48 and 6.00 in BilIN and primary GBAC, respectively (*Figure 5E*). In addition, *HER2* immunohistochemistry (IHC) (*Figure 5F and G*) was carried out to evaluate whether the *ERBB2* amplification was correlated to *HER2* protein expression levels on the membrane of tumor cells. The *HER2* IHC results were 2+/3 in the BilIN and 3+/3 in primary GBAC. In GB-S6 and GB-S7 (*Figure 3D and E*), *ERBB2* amplification was uniformly identified from primary tumors to metastatic tumors.
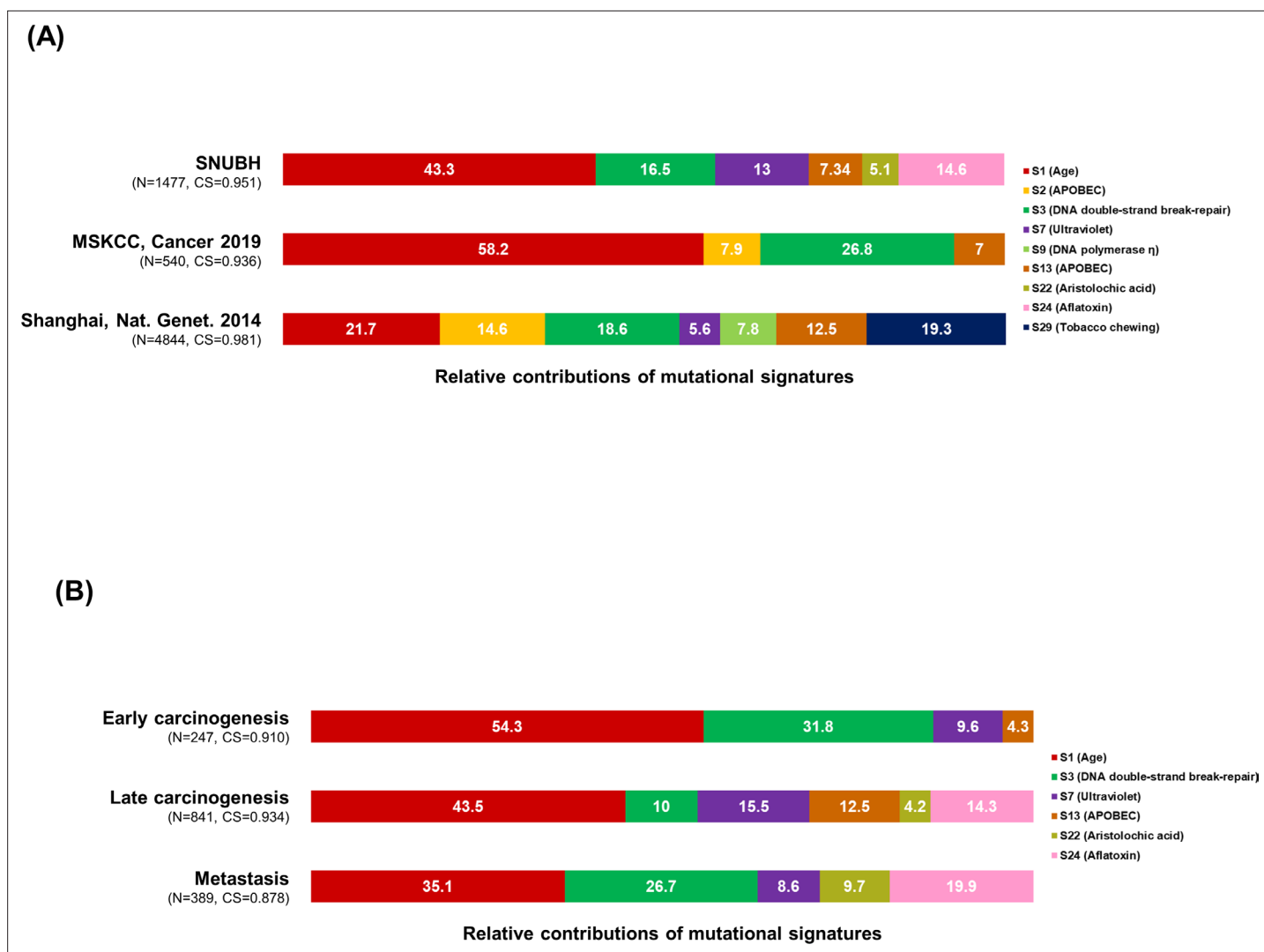
(A)

(B)

Figure 4. Mutational signature analysis. (**A–B**) The 100% stacked bar plots compare the proportions of known COSMIC Mutational Signatures v2 within our dataset and two public (MSKCC and Shanghai) datasets (**A**), and each category split according to the timing of development during clonal evolution (**B**). The total number of mutations (N) and cosine similarity (CS) values of each category were noted. BilIN, biliary intraepithelial neoplasia; COSMIC, Catalogue Of Somatic Mutations In Cancer; GB, gallbladder; LN, lymph node; MSKCC, Memorial Sloan Kettering Cancer Center.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Mutational signatures within each sample from 11 gallbladder adenocarcinoma (GBAC) patients.

Figure supplement 2. Mutational signatures within each category split according to the type of sample.

Figure supplement 3. Mutational signature analysis validated by two additional tools, Signal and MuSiCa.

Discussion

To the best of our knowledge, this is the first study to investigate clonal evolution from precancerous lesions to metastatic tumors in patients with GBAC. In this study, evolutionary trajectories of GBAC were inferred using multi-regional and longitudinal whole-exome sequencing (WES) data from precancerous lesions to primary and metastatic tumors. Based on these results, we derived comprehensive models of carcinogenesis and metastasis in GBAC.

In our analysis of carcinogenesis, we discovered three common themes. First, most mutations in frequently altered genes in primary GBAC are detected in concurrent BilIN (10 of 13, 76.9%), but a substantial proportion was subclonal. Second, branching evolution and subclonal diversity are commonly observed at the BilIN stage. Third, one subclone in BilIN commonly shrinks in the primary tumors, while the other subclones undergo linear and branching evolution, maintaining subclonal
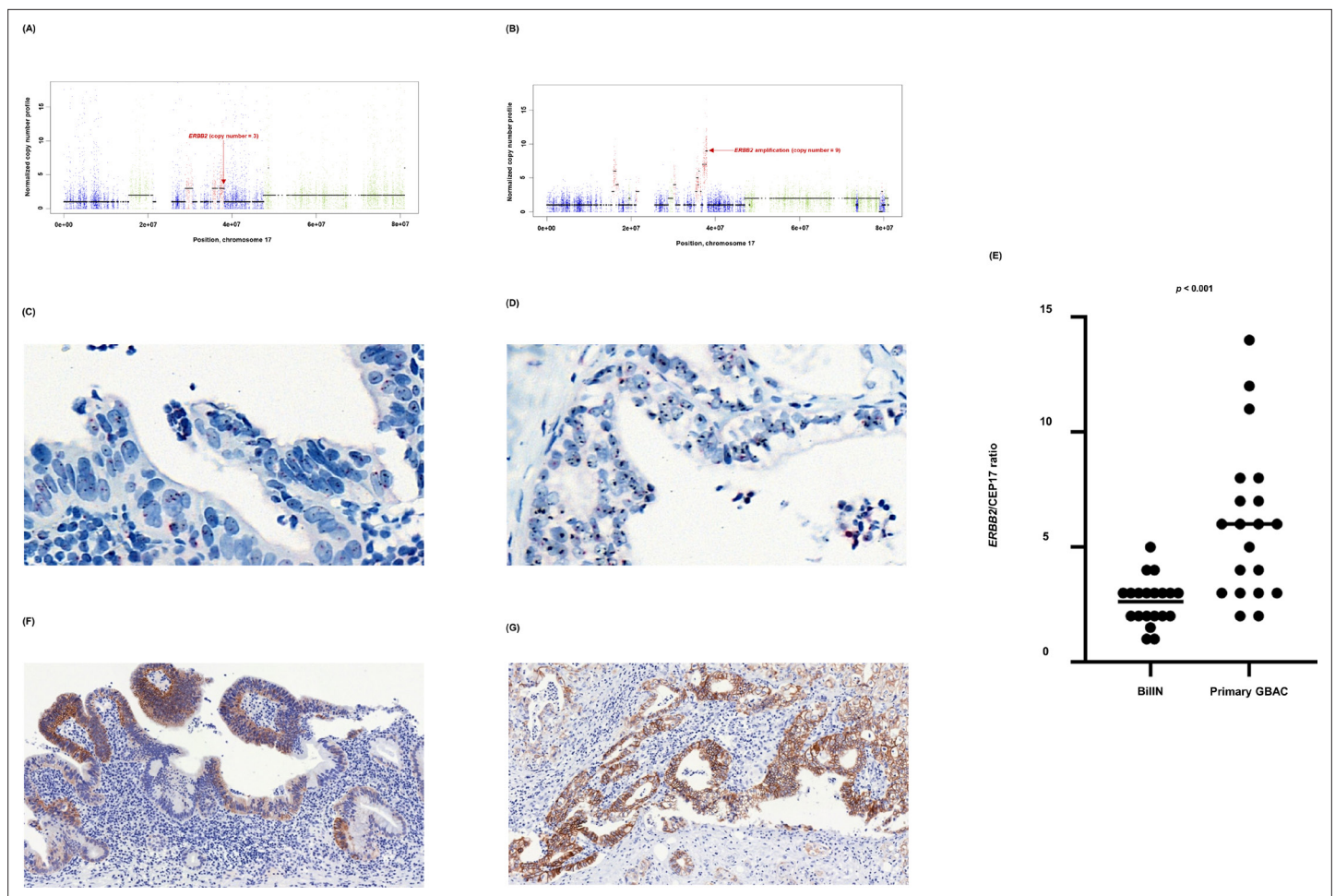
Figure 5. *ERBB2* copy number variation during neoplastic transformation of BilIN in GB-A1. (**A–G**) *ERBB2* gene amplification (**A and B**), *HER2* SISH (**C and D**), and *HER2* IHC (**F and G**) were compared between BilIN (**A, C**) and (**F**) and primary GBAC (**B**), (**D**), and (**G**) samples and the mean *ERBB2*/CEP17 ratio of BilIN and GB-A1 samples were compared by using the Wilcoxon rank-sum test (**E**). BilIN, biliary intraepithelial neoplasia; GBAC, gallbladder adenocarcinoma; IHC, immunohistochemistry; SISH, silver in situ hybridization.

diversity after the BilIN stage. A previous study in colorectal cancer by Vogelstein and colleagues demonstrated a stepwise carcinogenesis model from the precancerous lesion, adenoma, to invasive carcinoma by the accumulation of mutations, called the adenoma-carcinoma sequence (*Vogelstein et al., 1988*). In addition, recent studies on esophageal squamous cell carcinoma have reported that not only dysplasia but also histologically normal epithelia frequently harbor cancer-driving mutations (*Chen et al., 2017*; *Yokoyama et al., 2019*).

A recent study on GBAC (*Lin et al., 2021*) reported that *CTNNB1* mutation was frequently observed (5 out of 11) when BilIN and primary GBAC coexist. In addition, *ERBB2* alteration was found in 36.4%. In our study, *CTNNB1* S37F with LOH was observed in one (GB-A1) of four patients and 54.5% of patients had an *ERBB2* alteration. On the other hand, two previous studies on GBAC (*Narayan et al., 2019*; *Li et al., 2014*) reported the prevalence of *ERBB2* alteration as low as 7.9% and 9.4%, respectively. Race or lifestyle differences might have contributed to the difference.

In the analysis of metastasis, the following three phenomena are observed. First, subclonal expansion is frequent (8 of 11 patients, 72.7%) and some subclones expand substantially in metastatic tumors, leading to increased subclonal diversity. Previous studies suggest that subclonal diversity increases through branching evolution during progression and metastasis (*Turajlic and Swanton, 2016*; *Hong et al., 2015*; *Yachida et al., 2010*; *Minussi et al., 2021*). Second, metastases are polyclonal but metastatic lesions in one organ or adjacent organs show similar clonal compositions. Previously, it was thought that metastasis is initiated by the migration of a single cell

to another organ (*Nowell, 1976*). However, recent data suggest that polyclonal seeding occurs due to the migration of a cluster of cancer cells (*Turajlic et al., 2019*; *Cheung and Ewald, 2016*; *Ullah et al., 2018*; *Wei et al., 2017*). Third, we found the possibility of metastasis to metastasis spread. In recent studies on prostate and breast cancers, metastasis-to-metastasis spread was frequent (*Ullah et al., 2018*; *Gundem et al., 2015*). Our study found the possibility of metastasis-to-metastasis spread in GB-A2 (*Figure 3A*). Although intratumoral heterogeneity of primary GBAC may make it difficult to draw a strong conclusion, our data may support the possibility of metastasis-to-metastasis spread.

Of the 11 expanded subclones at the metastasis stage, we described putative driver mutations in eight subclones based on the previous literature (*Table 2*). For example, *SMAD4* mutations expanded during metastasis in GB-A1 and GB-S3 (*Figure 2A and D*) and have been associated with distant metastasis and poor prognosis in various cancers, including GBAC (*Narayan et al., 2019*; *Jamal-Hanjani et al., 2017*; *Yoon et al., 2021*; *Zhao et al., 2018*). In GB-S1 (*Figure 2B*), clone B containing *SLIT3* F843I evolves into E during metastasis by acquiring *ROBO1* P1360Q. The SLIT/ROBO pathway suppresses tumor progression by regulating invasion, migration, and apoptosis (*Gara et al., 2015*). In GB-A2 (*Figure 3A*), *DICER1* T519A is found in lung and chest wall metastases. The *DICER1* gene is associated with pleuropulmonary blastoma in children (*Foulkes et al., 2014*). In GB-S5 (*Figure 3C*), *KIAA0100* F5S and *CSMD2* E411K are found during metastasis. *KIAA0100* and *CSMD2* are frequently mutated during metastasis in adrenocortical carcinoma (*Gara et al., 2018*).

In mutational signature analysis, considering the evolutionary trajectories in cancers, we suggest that the timing of development during clonal evolution (*Figure 4B*) is better classification criterion than the type of sample (*Figure 4—figure supplement 2*). Our data indicate that the importance of signatures 1 and 13 decreased during metastasis while the roles of signatures 22 and 24 were relatively highlighted. Aristolochic acid is an ingredient of oriental herbal medicine (*Debelle et al., 2008*; *Hoang et al., 2013*). In addition, aflatoxin is known to be contained in soybean paste and soy sauce (*Ok et al., 2007*). Taken together, the two carcinogens might have little impact on the early stage of cancer development, but their impacts might be highlighted in overt cancer cells.

In this cancer precision medicine era, targeted sequencing data of a single specimen are not enough to determine whether the detected mutations are clonal or subclonal. This proof-of-concept study may enable us to deeply understand the clonal evolution in GBAC. Moreover, we found that some of the mutations were clonal while a substantial proportion was subclonal, which is usually not an effective druggable target. For example, if a drug targeting *ERBB2* p.V777L, a pathogenic mutation, is administered to GB-S2 patient (*Figure 2C*), the therapeutic effect will be limited in subclones without the *ERBB2* p.V777L mutation, especially from regional metastasis. Therefore, we believe that our study highlights the importance of precise genomic analysis of multi-regional and longitudinal samples in individual cancer patients. However, one caveat is that we cannot easily apply this to real-world patients because multi-regional and longitudinal tumor biopsies may not be feasible in most patients unless they underwent surgery and repeated biopsies. Recent studies using circulating tumor DNA have shown the possibility of easily detecting mutations involved in cancer development and progression (*Ignatiadis et al., 2021*). By detecting clonal mutations from the carcinogenesis stage in healthy individuals, we can diagnose GBAC at an early stage. In addition, by detecting subclonal mutations in patients with GBAC, we can monitor the expanding subclones during follow-up, which enables us to detect cancer progression earlier.

This study has several limitations. First, it is not possible to obtain samples through frequent biopsies whenever desired. Thus, tumor samples were not acquired according to their developmental sequence. Second, due to intratumoral heterogeneity, the clonal composition of a small piece of the tumor may not reflect that of the entire lesion (*Marusyk et al., 2012*). Third, as the number of analyzed samples was different among patients, an accessibility bias was inevitable – the more samples the patient has, the more clones we can identify.

In conclusion, subclonal diversity developed early in precancerous lesions and clonal selection was a common event during malignant transformation in GBAC. However, cancer clones continued to evolve in metastatic tumors and thus maintained subclonal diversity. Our novel approach may help us to understand the GBAC of individual patients and to move forward to precision medicine that enables early detection of carcinogenesis and metastasis, and effective targeted therapy in these patients.

Methods

Patients and tumor samples

Patients were eligible for this study if they were diagnosed with GBAC and received surgery between 2013 and 2018 at Seoul National University Bundang Hospital (SNUBH), Seongnam, Korea. Of the 272 patients who underwent surgery, 9 patients who were able to obtain ≥3 types of tissue from among BilIN, GBAC, regional LN metastasis, and distant metastasis and died at the time of analysis were enrolled in the study. In addition, two patients who donated their bodies for rapid autopsy were included in the study. Our board-certified pathologists (Prof. Soomin Ahn and Hee Young Na) identified and dissected BilIN and GBAC lesions from tissue slides. Patients' clinical information was obtained through retrospective medical record reviews. This study was approved by the Institutional Review Board (IRB) of SNUBH (IRB No. B-1902/522-303). Informed consent was waived because of the retrospective and anonymous nature of the study.

WES of GBAC

DNA was extracted from fresh-frozen or FFPE tissues. Library preparation and exome capture was carried out using Agilent SureSelect^{XT} Human All Exon V6 (Santa Clara, CA, USA). WES was conducted with a paired-end, 100 bp using Illumina NovaSeq 6000 (San Diego, CA, USA). The depth of coverage of tumors and normal control samples were at least 300× and 200×, respectively.

Analysis of WES data

WES data of GBAC and matched normal samples were analyzed using the Genomon2 pipeline (Institute of Medical Science, University of Tokyo, Tokyo, Japan; https://genomon.readthedocs.io/ja/latest/, accessed on February 1, 2022). In brief, sequencing reads from adapter-trimmed.fastq files were aligned to the human reference genome GRCh37 (hg19) without the 'chr' prefix using Burrows-Wheeler Aligner version 0.7.12, with default settings. SNVs and small indels were called by eliminating polymorphisms and sequencing errors and filtered by pre-specified criteria used in previous literature analyzed with the Genomon2 pipeline (*Yokoyama et al., 2019*; *Kakiuchi et al., 2020*; *Ochi et al., 2021*): (a) only exonic or splicing sites were included; (b) synonymous SNVs, unknown variants, or those without proper annotation were excluded; (c) polymorphisms in dbSNP 131 were excluded; (d) p-values <0.01 from Fisher's exact test were included; (e) simple repeat sequences were excluded; (f) strand ratio between positive-strand and negative-strand should not be 0 or 1 in tumor samples; (g) the number of variant reads should exceed 4 in tumor samples. For each patient, filtered variant lists of tumor samples were merged. Then, the merged list of target variants was manually called in each. bam file using bam-readcount version 0.8.0 (https://github.com/genome/bam-readcount; *Khanna et al., 2022a*; *Khanna et al., 2022b*, accessed on February 1, 2022) with Phred score and mapping quality of more than 30 and 60, respectively.

For samples with tumor purity >0.4, the ploidy of tumor cells was estimated using Sequenza version 3.0.0 to identify WGD (*Favero et al., 2015*). Copy number variations were analyzed using the Control-FREEC version 11.5 (*Boeva et al., 2012*). Gene amplification was defined as a copy number ≥6 (*Wolff et al., 2018*). In brief, the aligned.bam files were converted to.pileup.gz format using SAMtools version 1.9 (*Li, 2011*). The .pileup.gz files of data from GBAC and matched normal samples were analyzed using Control-FREEC with default settings. These datasets were then used to statistically infer clonal population structure using PyClone version 0.13.0 with the 'pyclone_beta_binomial' model (*Roth et al., 2014*). Clonal phylogeny was inferred by CITUP version 0.1.0 (*Malikic et al., 2015*) using cellular prevalence values of each cluster which were generated by PyClone. CITUP uses information from multiple samples and can infer clonal populations and their frequencies while satisfying phylogenetic constraints. To ensure accurate tree construction, clusters containing only one mutation were excluded from the input to CITUP if the mutation's role is unclear from previous literature. This filter removed 42 mutations from a total of 1577 mutations, representing less than 2.7% of all clustered mutations. In addition, in GB-A2, clusters which limited to only one organ were excluded from the analysis because the calculation for the phylogenetic tree using CITUP took more than 1 month when the number of clusters was ≥14. Analyzed results were visualized using MapScape for multi-regional specimens and TimeScape for longitudinal or putative longitudinal datasets, as appropriate (*Smith et al., 2017*). In all analysis steps, the data were adjusted for the tumor purity values of each tumor.

Mutational signature analysis was conducted using Mutalisk (*Lee et al., 2018*) and validated with Signal (*Degasperi et al., 2020*) and MuSiCa (*Díaz-Gay et al., 2018*). The COSMIC Mutational Signatures v2 (*Alexandrov et al., 2013*) was used as a reference. To compare with the other GBAC cohorts, we additionally analyzed two public datasets of GBAC from the MSKCC and the Shanghai group (*Narayan et al., 2019*; *Li et al., 2014*), which could be downloaded from the cBioPortal (https://www.cbioportal.org/, accessed on February 1, 2022; *Cerami et al., 2012*).

HER2 IHC and SISH

The histologic sections from individual FFPE tissues were deparaffinized and dehydrated. IHC and SISH analysis of *HER2*-positive cells was conducted by the board-certified pathologist using PATHWAY anti-*HER2*/neu antibody (4B5; rabbit monoclonal; Ventana Medical Systems, Tucson, AZ, USA) and a staining device (BenchMark XT, Ventana Medical Systems, Tuscon, AZ, USA), respectively, as previously described (*Koh et al., 2019*). Signals from 20 tumor cells were counted and a *HER2*/CEP17 ratio ≥2.0 was defined as *HER2* amplification (*Wolff et al., 2018*). Wilcoxon rank-sum test is used to compare the mean *HER2*/CEP17 ratio of BilIN versus primary tumor.

Additional information

Author contributions

Minsu Kang, Data curation, Software, Formal analysis, Validation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Hee Young Na, Resources, Data curation, Formal analysis, Validation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Soomin Ahn, Conceptualization, Resources, Formal analysis, Supervision, Validation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing; Ji-Won Kim, Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing; Sejoon Lee, Soyeon Ahn, Ju Hyun Lee, Jeonghwan Youk, Haesook T Kim, Formal analysis, Methodology, Writing – original draft, Writing – review and editing; Kui-Jin Kim, Resources, Methodology, Writing – original draft, Project administration, Writing – review and editing; Koung Jin Suh, Jun Suh Lee, Se Hyun Kim, Jin Won Kim, Yu Jung Kim, Keun-Wook Lee, Yoo-Seok Yoon, Jee Hyun Kim, Jin-Haeng Chung, Ho-Seong Han, Jong Seok Lee, Resources, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Minsu Kang ![ORCID] http://orcid.org/0000-0001-6491-2277
Soomin Ahn ![ORCID] http://orcid.org/0000-0002-1979-4010

Ji-Won Kim http://orcid.org/0000-0001-6426-9074
Se Hyun Kim http://orcid.org/0000-0002-2292-906X

Ethics

Human subjects: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Seoul National University Bundang Hospital (Number: B-1902/522-303). Informed consent was waived because of the retrospective and anonymous nature of the study.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.78636.sa1
Author response https://doi.org/10.7554/eLife.78636.sa2

Additional files

Supplementary files

• Supplementary file 1. Tables. (A) Baseline characteristics of samples are summarized. (B) Mutational signatures of our dataset are analyzed by three different tools, Mutalisk, Signal, and MuSiCa.

• Supplementary file 2. Detected somatic alternations in 11 patients with gallbladder adenocarcinoma (GBAC). Full list of mutations called in our cohort.

• MDAR checklist

Data availability

The raw sequence data underlying this manuscript are available as fastq files at the NCBI SRA database under the BioProject number PRJNA821382.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Kang M, Na HY, Ahn S, Kim J-W, Lee S, Ahn S, Lee JH, Youk J, Kim HT, Kim K-J, Suh KJ, Kim SH, Kim JW, Kim YJ, Lee KW, Yoon YS, Kim JH, Lee JS, Ho-Seong ZH, Jin-Haeng C	2022	Whole-exome sequencing of gallbladder carcinoma	https://www.ncbi.nlm.nih.gov/sra/PRJNA821382	NCBI Sequence Read Archive, PRJNA821382

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Narayan RR	2018	Gallbladder Cancer (MSK, Cancer 2018)	https://www.cbioportal.org/study?id=gbc_msk_2018	cBioPortal, Gallbladder-Cancer-(MSK)
Li M	2014	Gallbladder Carcinoma (Shanghai, Nat Genet 2014)	https://www.cbioportal.org/study?id=gbc_shanghai_2014	cBioPortal, Gallbladder-Carcinoma-(Shanghai)

References

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**:415–421. DOI: https://doi.org/10.1038/nature12477, PMID: 23945592

Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**:423–425. DOI: https://doi.org/10.1093/bioinformatics/btr670, PMID: 22155870

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA* **68**:394–424. DOI: https://doi.org/10.3322/caac.21492, PMID: 30207593

Bridgewater J, Galle PR, Khan SA, Llovet JM, Park J-W, Patel T, Pawlik TM, Gores GJ. 2014. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *Journal of Hepatology* **60**:1268–1289. DOI: https://doi.org/10.1016/j.jhep.2014.01.021, PMID: 24681130

Celik H, Koh WK, Kramer AC, Ostrander EL, Mallaney C, Fisher DAC, Xiang J, Wilson WC, Martens A, Kothari A, Fishberger G, Tycksen E, Karpova D, Duncavage EJ, Lee Y, Oh ST, Challen GA. 2018. Jarid2 functions as a tumor suppressor in myeloid neoplasms by repressing self-renewal in hematopoietic progenitor cells. *Cancer Cell* **34**:741–756.. DOI: https://doi.org/10.1016/j.ccell.2018.10.008, PMID: 30423295

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. 2012. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2**:401–404. DOI: https://doi.org/10.1158/2159-8290.CD-12-0095, PMID: 22588877

Chen X-X, Zhong Q, Liu Y, Yan S-M, Chen Z-H, Jin S-Z, Xia T-L, Li R-Y, Zhou A-J, Su Z, Huang Y-H, Huang Q-T, Huang L-Y, Zhang X, Zhao Y-N, Yun J-P, Wu Q-L, Lin D-X, Bai F, Zeng M-S. 2017. Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nature Communications* **8**:524. DOI: https://doi.org/10.1038/s41467-017-00650-0, PMID: 28900112

Cheung KJ, Ewald AJ. 2016. A collective route to metastasis: seeding by tumor cell clusters. Science **352**:167–169. DOI: https://doi.org/10.1126/science.aaf6546, PMID: 27124449

Davis A, Gao R, Navin N. 2017. Tumor evolution: linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta. Reviews on Cancer* **1867**:151–161. DOI: https://doi.org/10.1016/j.bbcan.2017.01.003, PMID: 28110020

Debelle FD, Vanherweghem JL, Nortier JL. 2008. Aristolochic acid nephropathy: a worldwide problem. *Kidney International* **74**:158–169. DOI: https://doi.org/10.1038/ki.2008.129, PMID: 18418355

Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, Morganella S, Nanda AS, Badja C, Koh G, Momen SE, Georgakopoulos-Soares I, Dias JML, Young J, Memari Y, Davies H, Nik-Zainal S. 2020. A practical framework and online tool for mutational signature analyses show Inter-tissue variation and driver dependencies. *Nature Cancer* **1**:249–263. DOI: https://doi.org/10.1038/s43018-020-0027-5, PMID: 32118208

Díaz-Gay M, Vila-Casadesús M, Franch-Expósito S, Hernández-Illán E, Lozano JJ, Castellví-Bel S. 2018. Mutational signatures in cancer (musica): a web application to implement mutational signatures analysis in cancer samples. *BMC Bioinformatics* **19**:224. DOI: https://doi.org/10.1186/s12859-018-2234-y, PMID: 29898651

Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC. 2015. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology* **26**:64–70. DOI: https://doi.org/10.1093/annonc/mdu479, PMID: 25319062

Foulkes WD, Priest JR, Duchaine TF. 2014. Dicer1: mutations, microRNAs and mechanisms. *Nature Reviews Cancer* **14**:662–672. DOI: https://doi.org/10.1038/nrc3802, PMID: 25176334

Gara RK, Kumari S, Ganju A, Yallapu MM, Jaggi M, Chauhan SC. 2015. Slit/robo pathway: a promising therapeutic target for cancer. *Drug Discovery Today* **20**:156–164. DOI: https://doi.org/10.1016/j.drudis.2014.09.008, PMID: 25245168

Gara SK, Lack J, Zhang L, Harris E, Cam M, Kebebew E. 2018. Metastatic adrenocortical carcinoma displays higher mutation rate and tumor heterogeneity than primary tumors. *Nature Communications* **9**:4172. DOI: https://doi.org/10.1038/s41467-018-06366-z, PMID: 30301885

Greaves M, Maley CC. 2012. Clonal evolution in cancer. *Nature* **481**:306–313. DOI: https://doi.org/10.1038/nature10762, PMID: 22258609

Griner EM, Kazanietz MG. 2007. Protein kinase C and other diacylglycerol effectors in cancer. *Nature Reviews. Cancer* **7**:281–294. DOI: https://doi.org/10.1038/nrc2110, PMID: 17384583

Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, Annala M, Kivinummi K, Goody V, Latimer C, O'Meara S, Dawson KJ, Isaacs W, Emmert-Buck MR, Nykter M, Foster C, Kote-Jarai Z, et al. 2015. The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**:353–357. DOI: https://doi.org/10.1038/nature14347, PMID: 25830880

Hoang ML, Chen CH, Sidorenko VS, He J, Dickman KG, Yun BH, Moriya M, Niknafs N, Douville C, Karchin R, Turesky RJ, Pu YS, Vogelstein B, Papadopoulos N, Grollman AP, Kinzler KW, Rosenquist TA. 2013. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Science Translational Medicine* **5**:197. DOI: https://doi.org/10.1126/scitranslmed.3006200, PMID: 23926200

Hong MKH, Macintyre G, Wedge DC, Van Loo P, Patel K, Lunke S, Alexandrov LB, Sloggett C, Cmero M, Marass F, Tsui D, Mangiola S, Lonie A, Naeem H, Sapre N, Phal PM, Kurganovs N, Chin X, Kerger M, Warren AY, et al. 2015. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nature Communications* **6**:6605. DOI: https://doi.org/10.1038/ncomms7605, PMID: 25827447

Ignatiadis M, Sledge GW, Jeffrey SS. 2021. Liquid biopsy enters the clinic-implementation issues and future challenges. *Nature Reviews. Clinical Oncology* **18**:297–312. DOI: https://doi.org/10.1038/s41571-020-00457-x, PMID: 33473219

Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R, Salm M, Horswell S, Escudero M, Matthews N, Rowan A, Chambers T, Moore DA, Turajlic S, Xu H, Lee S-M, et al. 2017. Tracking the evolution of non-small-cell lung cancer. *The New England Journal of Medicine* **376**:2109–2121. DOI: https://doi.org/10.1056/NEJMoa1616288, PMID: 28445112

Kakiuchi N, Yoshida K, Uchino M, Kihara T, Akaki K, Inoue Y, Kawada K, Nagayama S, Yokoyama A, Yamamoto S, Matsuura M, Horimatsu T, Hirano T, Goto N, Takeuchi Y, Ochi Y, Shiozawa Y, Kogure Y, Watatani Y, Fujii Y, et al. 2020. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **577**:260–265. DOI: https://doi.org/10.1038/s41586-019-1856-1, PMID: 31853061

Khanna A, Larson D, Srivatsan S, Mosior M, Abbott T, Kiwala S, Ley T, Duncavage E, Walter M, Walker J, Griffith O, Griffith M, Miller C. 2022a. bam-readcount. GitHub. https://github.com/genome/bam-readcount

Khanna A, Larson D, Srivatsan S, Mosior M, Abbott T, Kiwala S, Ley T, Duncavage E, Walter M, Walker J, Griffith O, Griffith M, Miller C. 2022b. Bam-readcount - rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software* **7**:3722. DOI: https://doi.org/10.21105/joss.03722

Koh J, Nam SK, Lee YW, Kim JW, Lee K-W, Ock C-Y, Oh D-Y, Ahn S-H, Kim H-H, Kang K-W, Kim WH, Lee H-Y, Lee HS. 2019. Trastuzumab specific epitope evaluation as a predictive and prognostic biomarker in gastric cancer patients. *Biomolecules* **9**:782. DOI: https://doi.org/10.3390/biom9120782, PMID: 31779184

Lee J, Lee AJ, Lee J-K, Park J, Kwon Y, Park S, Chun H, Ju YS, Hong D. 2018. Mutalisk: a web-based somatic mutation analyis toolkit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Research* **46**:W102–W108. DOI: https://doi.org/10.1093/nar/gky406, PMID: 29790943

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993. DOI: https://doi.org/10.1093/bioinformatics/btr509, PMID: 21903627

Li M, Zhang Z, Li X, Ye J, Wu X, Tan Z, Liu C, Shen B, Wang X-A, Wu W, Zhou D, Zhang D, Wang T, Liu B, Qu K, Ding Q, Weng H, Ding Q, Mu J, Shu Y, et al. 2014. Whole-Exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. *Nature Genetics* **46**:872–876. DOI: https://doi.org/10.1038/ng.3030, PMID: 24997986

Li M, Liu F, Zhang F, Zhou W, Jiang X, Yang Y, Qu K, Wang Y, Ma Q, Wang T, Bai L, Wang Z, Song X, Zhu Y, Yuan R, Gao Y, Liu Y, Jin Y, Li H, Xiang S, et al. 2019. Genomic ErbB2/ErbB3 mutations promote PD-L1-mediated immune escape in gallbladder cancer: a whole-exome sequencing analysis. *Gut* **68**:1024–1033. DOI: https://doi.org/10.1136/gutjnl-2018-316039, PMID: 29954840

Lin J, Peng X, Dong K, Long J, Guo X, Li H, Bai Y, Yang X, Wang D, Lu X, Mao Y, Sang X, Ji X, Zhao H, Liang H. 2021. Genomic characterization of co-existing neoplasia and carcinoma lesions reveals distinct evolutionary paths of gallbladder cancer. *Nature Communications* **12**:4753. DOI: https://doi.org/10.1038/s41467-021-25012-9, PMID: 34362903

Malikic S, McPherson AW, Donmez N, Sahinalp CS. 2015. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics* **31**:1349–1356. DOI: https://doi.org/10.1093/bioinformatics/btv003, PMID: 25568283

Marusyk A, Almendro V, Polyak K. 2012. Intra-Tumour heterogeneity: a looking glass for cancer? *Nature Reviews. Cancer* **12**:323–334. DOI: https://doi.org/10.1038/nrc3261, PMID: 22513401

Merlo LMF, Pepper JW, Reid BJ, Maley CC. 2006. Cancer as an evolutionary and ecological process. *Nature Reviews. Cancer* **6**:924–935. DOI: https://doi.org/10.1038/nrc2013, PMID: 17109012

Minussi DC, Nicholson MD, Ye H, Davis A, Wang K, Baker T, Tarabichi M, Sei E, Du H, Rabbani M, Peng C, Hu M, Bai S, Lin Y-W, Schalck A, Multani A, Ma J, McDonald TO, Casasent A, Barrera A, et al. 2021. Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**:302–308. DOI: https://doi.org/10.1038/s41586-021-03357-x, PMID: 33762732

Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A, Kato M, Hama N, Hosoda F, Urushidate T, Ohashi S, Hiraoka N, Ojima H, Shimada K, Okusaka T, Kosuge T, Miyagawa S, Shibata T. 2015. Genomic spectra of biliary tract cancer. *Nature Genetics* **47**:1003–1010. DOI: https://doi.org/10.1038/ng.3375, PMID: 26258846

Nakazawa K, Dobashi Y, Suzuki S, Fujii H, Takeda Y, Ooi A. 2005. Amplification and overexpression of c-erbB-2, epidermal growth factor receptor, and c-Met in biliary tract cancers. *The Journal of Pathology* **206**:356–365. DOI: https://doi.org/10.1002/path.1779, PMID: 15892172

Nam A-R, Kim J-W, Cha Y, Ha H, Park JE, Bang J-H, Jin MH, Lee K-H, Kim T-Y, Han S-W, Im S-A, Kim T-Y, Oh D-Y, Bang Y-J. 2016. Therapeutic implication of HER2 in advanced biliary tract cancer. *Oncotarget* **7**:58007–58021. DOI: https://doi.org/10.18632/oncotarget.11157, PMID: 27517322

Narayan RR, Creasy JM, Goldman DA, Gönen M, Kandoth C, Kundra R, Solit DB, Askan G, Klimstra DS, Basturk O, Allen PJ, Balachandran VP, D'Angelica MI, DeMatteo RP, Drebin JA, Kingham TP, Simpson AL, Abou-Alfa GK, Harding JJ, O'Reilly EM, et al. 2019. Regional differences in gallbladder cancer pathogenesis: insights from a multi-institutional comparison of tumor mutations. *Cancer* **125**:575–585. DOI: https://doi.org/10.1002/cncr.31850, PMID: 30427539

Nepal C, Zhu B, O'Rourke CJ, Bhatt DK, Lee D, Song L, Wang D, Van Dyke AL, Choo-Wosoba H, Liu Z, Hildesheim A, Goldstein AM, Dean M, LaFuente-Barquero J, Lawrence S, Mutreja K, Olanich ME, Lorenzo Bermejo J, CGR Exome Studies Group, Ferreccio C, et al. 2021. Integrative molecular characterisation of gallbladder cancer reveals micro-environment-associated subtypes. *Journal of Hepatology* **74**:1132–1144. DOI: https://doi.org/10.1016/j.jhep.2020.11.033, PMID: 33276026

Nowell PC. 1976. The clonal evolution of tumor cell populations. *Science* **194**:23–28. DOI: https://doi.org/10.1126/science.959840, PMID: 959840

Ochi Y, Yoshida K, Huang Y-J, Kuo M-C, Nannya Y, Sasaki K, Mitani K, Hosoya N, Hiramoto N, Ishikawa T, Branford S, Shanmuganathan N, Ohyashiki K, Takahashi N, Takaku T, Tsuchiya S, Kanemura N, Nakamura N, Ueda Y, Yoshihara S, et al. 2021. Clonal evolution and clinical implications of genetic abnormalities in blastic transformation of chronic myeloid leukaemia. *Nature Communications* **12**:2833. DOI: https://doi.org/10.1038/s41467-021-23097-w, PMID: 33990592

Ok HE, Kim HJ, Shim WB, Lee H, Bae D-H, Chung D-H, Chun HS. 2007. Natural occurrence of aflatoxin B1 in marketed foods and risk estimates of dietary exposure in Koreans. *Journal of Food Protection* **70**:2824–2828. DOI: https://doi.org/10.4315/0362-028x-70.12.2824, PMID: 18095437

Roa JC, García P, Kapoor VK, Maithel SK, Javle M, KoshiolJ. 2022. Gallbladder cancer. *Nature Reviews. Disease Primers* **8**:70. DOI: https://doi.org/10.1038/s41572-022-00398-y, PMID: 36302789

Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A, Shah SP. 2014. PyClone: statistical inference of clonal population structure in cancer. *Nature Methods* **11**:396–398. DOI: https://doi.org/10.1038/nmeth.2883, PMID: 24633410

Smith MA, Nielsen CB, Chan FC, McPherson A, Roth A, Farahani H, Machev D, Steif A, Shah SP. 2017. E-scape: interactive visualization of single-cell phylogenetics and cancer evolution. *Nature Methods* **14**:549–550. DOI: https://doi.org/10.1038/nmeth.4303, PMID: 28557980

Turajlic S, Swanton C. 2016. Metastasis as an evolutionary process. Science **352**:169–175. DOI: https://doi.org/10.1126/science.aaf2784, PMID: 27124450

Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, Nicol D, O'Brien T, Larkin J, Horswell S, Stares M, Au L, Jamal-Hanjani M, Challacombe B, Chandra A, Hazell S, Eichler-Jonsson C, Soultati A, Chowdhury S, Rudman S, et al. 2018. Tracking cancer evolution reveals constrained routes to metastases: tracerx renal. *Cell* **173**:581–594.. DOI: https://doi.org/10.1016/j.cell.2018.03.057, PMID: 29656895

Turajlic S, Sottoriva A, Graham T, Swanton C. 2019. Resolving genetic heterogeneity in cancer. *Nature Reviews. Genetics* **20**:404–416. DOI: https://doi.org/10.1038/s41576-019-0114-6, PMID: 30918367

Ullah I, Karthik G-M, Alkodsi A, Kjällquist U, Stålhammar G, Lövrot J, Martinez N-F, Lagergren J, Hautaniemi S, Hartman J, Bergh J. 2018. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *The Journal of Clinical Investigation* **128**:1355–1370. DOI: https://doi.org/10.1172/JCI96149, PMID: 29480816

Valle J, Wasan H, Palmer DH, Cunningham D, Anthoney A, Maraveyas A, Madhusudan S, Iveson T, Hughes S, Pereira SP, Roughton M, Bridgewater J, ABC-02 Trial Investigators. 2010. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *The New England Journal of Medicine* **362**:1273–1281. DOI: https://doi.org/10.1056/NEJMoa0908721, PMID: 20375404

Valle JW, Kelley RK, Nervi B, Oh DY, Zhu AX. 2021. Biliary tract cancer. *Lancet* **397**:428–444. DOI: https://doi.org/10.1016/S0140-6736(21)00153-7, PMID: 33516341

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. 1988. Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine* **319**:525–532. DOI: https://doi.org/10.1056/NEJM198809013190901, PMID: 2841597

Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, Polak P, Kim J, Hatanaka Y, Maejima K, Lawlor RT, Nakanishi Y, Mitsuhashi T, Fujimoto A, Furuta M, Ruzzenente A, Conci S, Oosawa A, Sasaki-Oku A, Nakano K, et al. 2018. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *Journal of Hepatology* **68**:959–969. DOI: https://doi.org/10.1016/j.jhep.2018.01.009, PMID: 29360550

Wei Q, Ye Z, Zhong X, Li L, Wang C, Myers RE, Palazzo JP, Fortuna D, Yan A, Waldman SA, Chen X, Posey JA, Basu-Mallick A, Jiang BH, Hou L, Shu J, Sun Y, Xing J, Li B, Yang H. 2017. Multiregion whole-exome sequencing of matched primary and metastatic tumors revealed genomic heterogeneity and suggested polyclonal seeding in colorectal cancer metastasis. *Annals of Oncology* **28**:2135–2141. DOI: https://doi.org/10.1093/annonc/mdx278, PMID: 28911083

Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, Jenkins RB, Press MF, Spears PA, Vance GH, Viale G, McShane LM, Dowsett M. 2018. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of clinical oncology/college of American pathologists clinical practice guideline focused update. *Archives of Pathology & Laboratory Medicine* **142**:1364–1382. DOI: https://doi.org/10.5858/arpa.2018-0902-SA, PMID: 29846104

Xu J, Zhou W, Yang F, Chen G, Li H, Zhao Y, Liu P, Li H, Tan M, Xiong X, Sun Y. 2017. The β-trcp-FBXW2-SKP2 axis regulates lung cancer cell growth with FBXW2 acting as a tumour suppressor. *Nature Communications* **8**:14002. DOI: https://doi.org/10.1038/ncomms14002, PMID: 28090088

Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**:1114–1117. DOI: https://doi.org/10.1038/nature09515, PMID: 20981102

Yi M, Yang J, Li W, Li X, Xiong W, McCarthy JB, Li G, Xiang B. 2017. The NOR1/OSCP1 proteins in cancer: from epigenetic silencing to functional characterization of a novel tumor suppressor. *Journal of Cancer* **8**:626–635. DOI: https://doi.org/10.7150/jca.17579, PMID: 28367242

Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, Shiozawa Y, Sato Y, Aoki K, Kim SK, Fujii Y, Yoshida K, Kataoka K, Nakagawa MM, Inoue Y, Hirano T, Shiraishi Y, Chiba K, Tanaka H, Sanada M, et al. 2019. Age-Related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**:312–317. DOI: https://doi.org/10.1038/s41586-018-0811-x, PMID: 30602793

Yoon JG, Kim MH, Jang M, Kim H, Hwang HK, Kang CM, Lee WJ, Kang B, Lee CK, Lee MG, Chung HC, Choi HJ, Park YN. 2021. Molecular characterization of biliary tract cancer predicts chemotherapy and programmed death 1/programmed death-ligand 1 blockade responses. *Hepatology* **74**:1914–1931. DOI: https://doi.org/10.1002/hep.31862, PMID: 33884649

Yoshida K. 2007. Pkcdelta signaling: mechanisms of DNA damage response and apoptosis. *Cellular Signalling* **19**:892–901. DOI: https://doi.org/10.1016/j.cellsig.2007.01.027, PMID: 17336499

Zhao M, Mishra L, Deng CX. 2018. The role of TGF-β/SMAD4 signaling in cancer. *International Journal of Biological Sciences* **14**:111–123. DOI: https://doi.org/10.7150/ijbs.23230, PMID: 29483830