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

Minsu Kang1,*, Hee Young Na2,*, Soomin Ahn3,†, Ji-Won Kim1,4,†, Sejoon Lee5, Soyeon Ahn6, Ju Hyun Lee1, Jeonghwan Youk1, Haesook T. Kim7, Kui-Jin Kim8, Koung Jin Suh1, Jun Suh Lee9, Se Hyun Kim1, Jin Won Kim1, Yu Jung Kim1, Keun-Wook Lee1, Yoo-Seok Yoon9, Jee Hyun Kim1, Jin-Haeng Chung2, Ho-Seong Han9, and Jong Seok Lee1

*These authors contributed equally to this work.

1Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

2Department of Pathology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

3Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea


5Center for Precision Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

6Medical Research Collaboration Center, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea
†Co-corresponding authors: Ji-Won Kim, M.D., Ph.D., Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82 Gumi-ro-173-beon-gil, Bundang-gu, Seongnam 13620, Korea; Tel: +82-31-787-7084; Fax: +82-31-787-4098; Email: jiwonkim@snubh.org

Soomin Ahn, M.D., Ph.D., Department of Pathology and Translational Genomics, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea; Tel: +82-2-3410-2800; Fax: +82-2-3410-0025; Email: suminy317@gmail.com
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 9 patients (81.8%). Of these, 8 patients (88.9%) had a total of 11 subclones expanded at least 7-fold during metastasis. These subclones harbored putative metastasis-driving mutations in cancer-related genes such as SMAD4, ROBO1, and DICER1. In mutational signature analysis, 6 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.

Keywords: Gallbladder adenocarcinoma, Clonal evolution, Carcinogenesis, Metastasis
Introduction

Gallbladder adenocarcinoma (GBAC) is a malignant neoplasm that has a high incidence rate in Chile, India, Poland, Pakistan, Japan, and Korea (1-4). 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 (1, 2, 5).

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 (1, 6-12). Of note, ERBB2 amplification and overexpression occur in approximately 6.9-28.6% of GBAC (1, 11, 13, 14) and may have therapeutic implications.

Cancer cells undergo clonal evolution by acquiring additional mutations and thus exhibit more aggressive phenotypes, including invasion and metastasis (15-17). Several large-scale studies have provided evidence of clonal evolution in some cancer types, including lung and kidney cancers (18, 19). 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.
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.

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) (11, 12). 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 6 patients, five were accompanied by loss of heterozygosity (LOH).

Ploidy was analyzed in 15 tumors of 6 patients with purity > 0.4 because ploidy estimation was inaccurate when tumor purity is ≤ 0.4 (20). In 2 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) (21) and CITUP (22) and then visualized using MapScape and TimeScape (23) (Figure 1C). Among 6 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
Branching evolution and subclonal diversity were commonly observed in the BilIN of the 4 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 (24). 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%.

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 9 patients (81.8%) and linear evolution only in 2 patients (18.2%). Of the 9 patients with branching and linear evolution, eight (88.9%) had a total of 11 subclones expanded at least 7-fold in the regional or distant metastasis stage (Table 2). In GB-A1 (Figure 2A), clone E, which acquired JARID2 (25) K603Q with LOH, increased from 0.2%
to 20.5% during common bile duct metastasis. In addition, clone F, which acquired SMAD4
(11, 18, 26, 27) 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 (28). In GB-A2 (Figure 3A), clone F, which acquired PRKCD (29, 30) I153L,
expanded from 0.3% to 73.1% and 68.3% during metastasis to liver and mesentery,
respectively. In addition, clone G acquired DICER1 (31) T519A and expanded from 2.8% to
39.9% and 61.5% during metastasis to the lung and chest wall, respectively. In the other 5
patients (Figures 2C and 3B-E), mutations in FBXW2 (32), KIAA0100 (33), CSMD2 (33),
and OSCP1 (34) genes were observed.

In GB-S8 and GB-S9, only linear evolution was identified (Figures 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 (11, 12). 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 (35) and then validated with two additional tools, Signal (36) and MuSiCa (37) (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 BillN 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) (Figures 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 BillN and 3+/3 in primary GBAC. In GB-S6 and GB-S7 (Figures 3D and E), *ERBB2* amplification was uniformly identified from primary tumors to metastatic tumors.
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 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 (38). 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 (39, 40).

A recent study on GBAC (10) 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 1 (GB-A1) of 4 patients and 54.5% of patients had an ERBB2 alteration. On the other hand, two previous studies on GBAC (11, 12) 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 (41-44). 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 (45). However, recent data suggest that polyclonal seeding occurs due to the migration of a cluster of cancer cells (15, 46-48). Third, we found the possibility of metastasis to metastasis spread. In recent studies on prostate and breast cancers, metastasis-to-metastasis spread was frequent (47, 49). 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 8 subclones based on the previous literature (Table 2). For example, SMAD4 mutations expanded during metastasis in GB-A1 and GB-S3 (Figures 2A and D) and have been associated with distant metastasis and poor prognosis in various cancers, including GBAC (11, 18, 26, 27). 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 (28). 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 (31). 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 (33).
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 (50, 51). In addition, aflatoxin is known to be contained in soybean paste and soy sauce (52). 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 (53). 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 (54). 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, 2 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 SureSelectXT 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 Feb. 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 (40, 55, 56): (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, accessed on Feb. 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 (20). Copy number variations (CNVs) were analyzed using the Control-FREEC version 11.5 (57). Gene amplification was defined as a copy number ≥ 6 (58). In brief, the aligned .bam files were converted to .pileup.gz format using SAMtools version 1.9 (59). 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 (21). Clonal phylogeny was inferred by CITUP.
version 0.1.0 (22) 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 1,577 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 one month when the number of clusters was $\geq 14$. Analyzed results were visualized using MapScape for multi-regional specimens and TimeScape for longitudinal or putative longitudinal datasets, as appropriate (23). In all analysis steps, the data were adjusted for the tumor purity values of each tumor.

Mutational signature analysis was conducted using Mutalisk (35) and validated with Signal (36) and MuSiCa (37). The COSMIC Mutational Signatures v2 (60) 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 (11, 12), which could be downloaded from the cBioPortal (https://www.cbioportal.org/, accessed on Feb. 1, 2022) (61).

**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 (62). Signals from 20
tumor cells were counted and a \( \text{HER2/CEP17 ratio} \geq 2.0 \) was defined as \( \text{HER2} \)
amplification (58). Wilcoxon rank-sum test is used to compare the mean \( \text{HER2/CEP17 ratio} \) of BilIN versus primary tumor.
Acknowledgments

This study was supported by a grant from Seoul National University Bundang Hospital Research Fund (No. 16-2021-001) and the Small Grant for Exploratory Research (SGER) program (NRF-2018R1D1A1A02086240) of the National Research Foundation (NRF), Korea. The authors would like to express the deepest respect to the two patients who donated their bodies for this study after death.

Competing interests statement

All other authors declare no competing interests.

Data availability statement

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


Table 1. Baseline characteristics of 11 patients with GBAC.

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

* 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.
Table 2. List of subclones expanding during metastasis.

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

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 legends

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 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.

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

(A-K) The Clonal population structure was statistically inferred for GB-A1 (A), GB-S1 (B), GB-S2 (C), GB-S3 (D), GB-A2 (E), GB-S4 (F), GB-S5 (G), GB-S6 (H), GB-S7 (I), GB-S8 (J), and GB-S9 (K) using PyClone. See Supplementary File 2 to check the matching information for the PyClone clusters and the CITUP clones.

BilIN, biliary intraepithelial neoplasia; GB, gallbladder; CBD, common bile duct; LN, lymph node.
Figure 1—figure supplement 2. BiILN and primary GBAC of the GB-S7 patient presumed to be derived from different origins.

(A) H&E staining of BiILN. (B) H&E staining of GBAC. (C) Variant allele frequency of mutated genes in BiILN and primary GBAC.

BiILN, biliary intraepithelial neoplasia; H&E, hematoxylin, and eosin.

Figure 2. Spatial and temporal clonal evolution of 4 patients with GBAC whose precancerous BiILN 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 7-fold in the metastatic tumor. Diamond symbol (◇) is used if the mutation in the figure is classified as 'pathogenic' in 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.

BiILN, biliary intraepithelial neoplasia; CBD, common bile duct; CN, copy number; GB,
gallbladder; GP, gemcitabine + cisplatin; LN, lymph node; LOH, loss of heterozygosity.

**Figure 3. Spatial and temporal clonal evolution of additional 7 patients with 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 7-fold in the metastatic tumor. Diamond symbol (◇) is used if the mutation in the figure is classified as 'pathogenic' in 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.

**Figure 4. Mutational signature analysis.**

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.
The 100% stacked bar plots compare the proportions of known COSMIC Mutational Signatures v2 within our dataset and two public (MSKCC and Shangai) 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.

Figure 4—figure supplement 1. Mutational signatures within each sample from 11 GBAC patients.

The 100% stacked bar plots comparing the proportions of known COSMIC mutational signatures v2 within each sample from 11 GBAC patients.

BilIN, biliary intraepithelial neoplasia; CBD, common bile duct; COSMIC, catalogue of somatic mutations in cancer; GB, gallbladder; LN, lymph node.

Figure 4—figure supplement 2. Mutational signatures within each category split according to the type of sample.

The 100% stacked bar plots comparing the proportions of known COSMIC mutational signatures v2 within each category split according to the type of sample. 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.
Figure 4—figure supplement 3. Mutational signature analysis validated by two additional tools, Signal and MuSiCa.

The 100% stacked bar plots comparing the proportions of known COSMIC Mutational Signatures v2 within each category split according to the timing of development during clonal evolution. The total number of mutations (N) and cosine similarity (CS) values of each category were noted.

COSMIC, catalogue of somatic mutations in cancer.

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.
Supplementary File legends

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 GBAC.

Full list of mutations called in our cohort.
Figure 4(A)

SNUBH
(N=1477, CS=0.951)

MSKCC, Cancer 2019
(N=540, CS=0.936)

Shanghai, Nat. Genet. 2014
(N=4844, CS=0.981)

Relative contributions of mutational signatures

Figure 4(B)

Early carcinogenesis
(N=247, CS=0.910)

Late carcinogenesis
(N=841, CS=0.934)

Metastasis
(N=389, CS=0.878)

Relative contributions of mutational signatures
Figure 4—figure supplement 1
Figure 4—figure supplement 2

BillN
(N=168, CS=0.860)

GB
(N=831, CS=0.928)

Regional LN metastasis
(N=436, CS=0.946)

Distant metastasis
(N=769, CS=0.932)

Relative contributions of mutational signatures