scientific reports

OPEN

Whole exome sequencing of multi-regions reveals tumor heterogeneity in *Opisthorchis viverrini*-associated cholangiocarcinoma

Sirinya Sitthirak[1,3], Arporn Wangwiwatsin[1,3], Apinya Jusakul[3,4], Nisana Namwat[1,3], Poramate Klanrit[1,3], Hasaya Dokduang[3,6], Prakasit Sa-ngiamwibool[5], Attapol Titapun[2,3], Apiwat Jareanrat[2,3], Vasin Thanasukarn[2,3], Natcha Khuntikeo[2,3], Bin Tean Teh[7], Luke Boulter[8], Yoshinori Murakami[9] & Watcharin Loilome[1,3]✉

The study examines *Opisthorchis viverrini* (*OV*)-related cholangiocarcinoma (CCA), a serious malignancy common in Southeast Asia. Through multi-regional whole-exome sequencing of 52 tumor samples and 13 adjacent tissues from 13 patients, significant intratumoral heterogeneity (ITH) and inter-patient heterogeneity are shown. Chronic liver fluke infection induces a distinct mutational landscape, with 48–90% of mutations concentrated in each region of the tumor. The average mutation burden is 95 non-synonymous mutations per area, exceeding previous CCA investigations. Critical driver mutations in *TP53*, *SMAD4*, and other genes underscore their significance in pathogenesis. Mutational markers elucidate mechanisms including spontaneous deamination and impaired DNA repair. Unique mutation patterns distinguish *OV*-associated CCA from other variants. Chromosomal instability in patient K110 signifies aggressive tumor behavior and unfavorable prognosis. Targetable mutations such as *ERBB2* underscore the possibility for personalized therapeutics. These findings underscore the necessity for personalized strategies for treatment that target both trunk and branch mutations in endemic areas.

Keywords Cholangiocarcinoma, Multi-region sequencing, Tumor heterogeneity, Clonal evolution, Phylogenetic tree

Abbreviations

CCA	Cholangiocarcinoma
ITH	Intratumoral heterogeneity
WES	Whole exome sequencing
NGS	Next-generation sequencing
SNV	Single nucleotide variant
CNA	Copy number alteration
CNV	Copy number variation

[1]Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. [2]Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. [3]Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen 40002, Thailand. [4]Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand. [5]Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. [6]Faculty of Medicine, Mahasarakham University, Kantharawichai District, Mahasarakham 44000, Thailand. [7]National Cancer Centre Singapore, Duke-NUS Medical School, Singapore 169857, Singapore. [8]MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, Scotland, UK. [9]Department of Molecular Biology, Institute for Advanced Medical Sciences, Nippon Medical School, Tokyo, Japan. ✉email: watclo@kku.ac.th

Cholangiocarcinoma (CCA) is a disease that includes multiple types of tumors that have features similar to cholangiocyte development[1]. CCA can be categorized according to its anatomical location, which includes intrahepatic (iCCA), perihilar (pCCA), or distal (dCCA)[2]. In the past four decades, there has been a steady increase in the worldwide occurrence of CCA, making it the second most common liver cancer, behind hepatocellular carcinoma (HCC)[3–5]. CCA, the second most prevalent form of primary liver cancer, is uncommon worldwide but has a higher incidence in Southeast Asia, particularly in northeastern Thailand[6]. Although several risk factors have been proposed for CCA, the exact etiology of the disease remains uncertain. Primary sclerosing cholangitis is the most prominent risk factor in Western nations[7–9]. However, in Southeast Asian countries, the infection of human liver flukes, including *Opisthorchis viverrini* (*OV*) and *Clonorchis sinensis*, is strongly associated with the development of this disease[10]. CCA patients are often diagnosed at advanced stages, limiting treatment options and prognosis, with surgery often compromised by late detection. Consequently, developing new treatment options for CCA remains challenging[11,12].

ITH, which refers to the existence of several cell populations within a single tumor tissue sample, is a major factor in the ineffectiveness of anticancer therapies. ITH results in variation in the pace of tumor development, ability to invade surrounding tissues, susceptibility to drugs, and overall prognosis. These differences arise from changes in the genetic and phenotypic characteristics of the different cell populations inside the tumor[13]. Next-generation sequencing (NGS) technology has been widely employed in the investigation of tumor genome variation and has demonstrated significant potential in ITH research. The TRACERx lung research initiative, for example, exemplifies the utilization of NGS to rationalize samples from many regions of lung cancer tissues in 100 patients with early-stage Non-Small Cell Lung Cancer (NSCLC). This approach revealed extensive intratumoral heterogeneity (ITH) and copy number variation (CNV) in NSCLC patients, correlating ITH with prognosis and guiding future genomic medicine[14]. To date, the majority of tumor evolution studies have concentrated on cancers arising in Western patients which are driven by exposure to environmental carcinogens. There is currently very little knowledge as to how cancers driven by chronic parasitic infection evolve.

Recent advances in genomic profiling have identified key driver mutations such as *TP53*, *KRAS*, *IDH1/2*, and *FGFR* fusions in CCA, providing new insights into tumor biology and potential therapeutic targets[1]. These findings have facilitated the classification of molecular subtypes, improving precision medicine approaches for CCA management[15]. Gaining understanding of the variety of CCA could improve our understanding of how to manage this condition. A recent study found that chemotherapy treatment caused an increase in ITH in small cell lung cancer (SCLC), which in turn led to the development of numerous mechanisms of drug resistance[16]. The ITH status of patients with CCA who have not received chemotherapy remains undetermined due to the lack of accessible tumor samples.

The aim of this study was to examine the molecular diversity within surgically excised CCA tumors from patients in *OV*-endemic areas. The objective was to analyze the genomics profile and clonal evolution of 52 samples collected from 13 patients with CCA. This analysis could enhance our understanding of the evolutionary advancement of CCA in the setting of chronic infection.

Methods

Methodological framework for conducting the study

Figure 1 provides a concise summary of the primary procedure for this investigation. More precisely, following the extraction of DNA from both the tumor and adjacent tissues, a library preparation was created using the Agilent SureSelect V6 method and then analyzed using the NovaSeq 6000 platform. The sequencing data were aligned with the reference human genome to detect somatic single nucleotide variants (SNVs), somatic copy number alterations (SCNAs), and mutational signatures. Statistical analysis was conducted on somatic mutation data to deduce the subclone of individuals with CCA.

Patients and samples collection

Fresh frozen tissues were obtained from 13 resected liver specimens of patients who received curative resection at Srinagarind Hospital, Faculty of Medicine, Khon Kaen University. Most patients were residents of Northeastern Thailand, a region where *OV* infection is highly prevalent and recognized as a significant risk factor for CCA development[17]. This investigation acquired a total of 65 distinct sites, comprising four tumor regions and one adjacent area for each individual (Fig. 2). To investigate the diversity of CCA, we gathered data on individuals diagnosed with CCA from the Department of Pathology, Faculty of Medicine, Khon Kaen University. The study participants were individuals who had undergone surgical procedures and possessed tumors over 6 cm in size. All procedures were conducted in compliance with pertinent standards and regulations, including the Declaration of Helsinki and relevant national and institutional guidelines. The study received approval from the Human Research Ethics Committee of Khon Kaen University (Approval No. HE671356). Informed consent was secured from all participants and/or their legal guardians following a comprehensive explanation of the study. A total of 65 samples, comprising tumor tissue and adjacent non-cancerous tissue, were collected from 13 patients diagnosed with CCA. Simultaneously, we gathered comprehensive pathology and clinical data, encompassing patient prognosis. DNA obtained from tissues was utilized for whole-exome sequencing (WES).

DNA extraction and quantification

The genomic DNA was extracted from fresh frozen tumors and adjacent regions using a QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) following the instructions provided by the manufacturer. The genomic DNA was assessed for quality to determine the A260/280 and A260/A230 ratios using Nanodrop2000 (Thermo Fisher Scientific). As per the manufacturer's guidelines, the quantification of all DNA samples was performed using Qubit 3.0 and a dsDNA HS Assay Kit (Life Technologies).

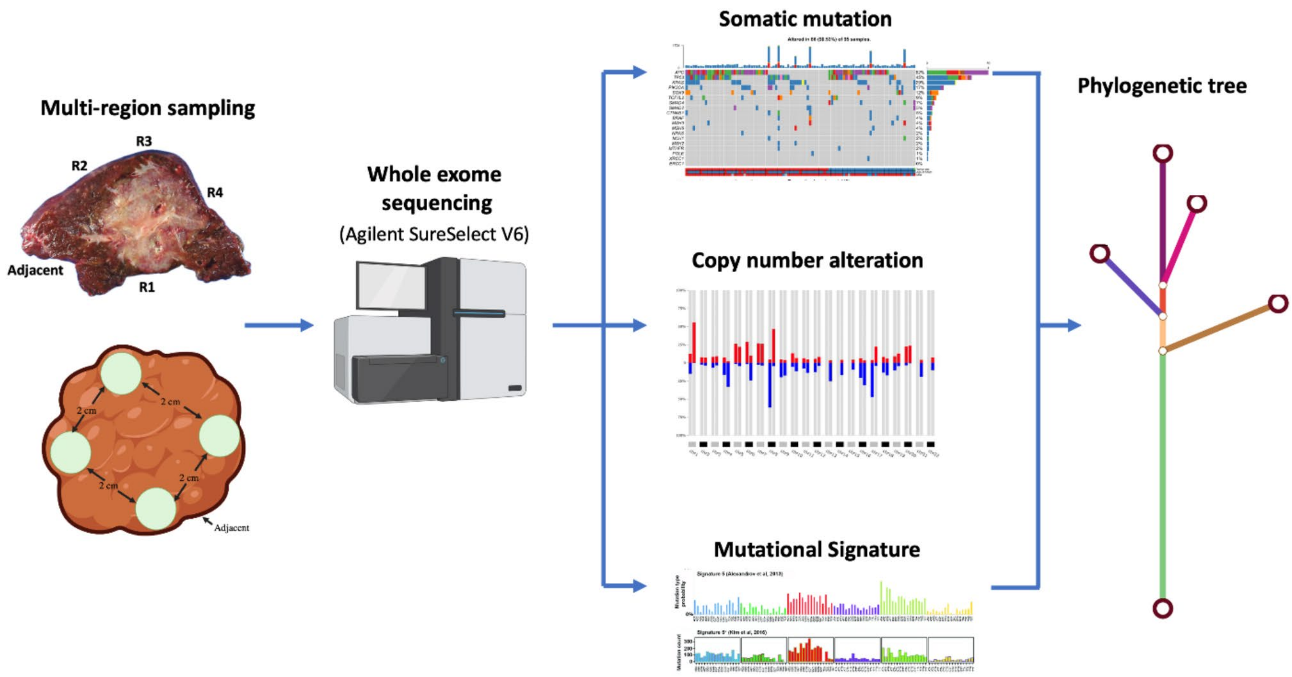

Fig. 1. Evaluating the genetic diversity in cholangiocarcinoma. The study design is visually shown in a schematic form, illustrating the investigation of ITH through the examination of multi-regions inside tumor tissues.

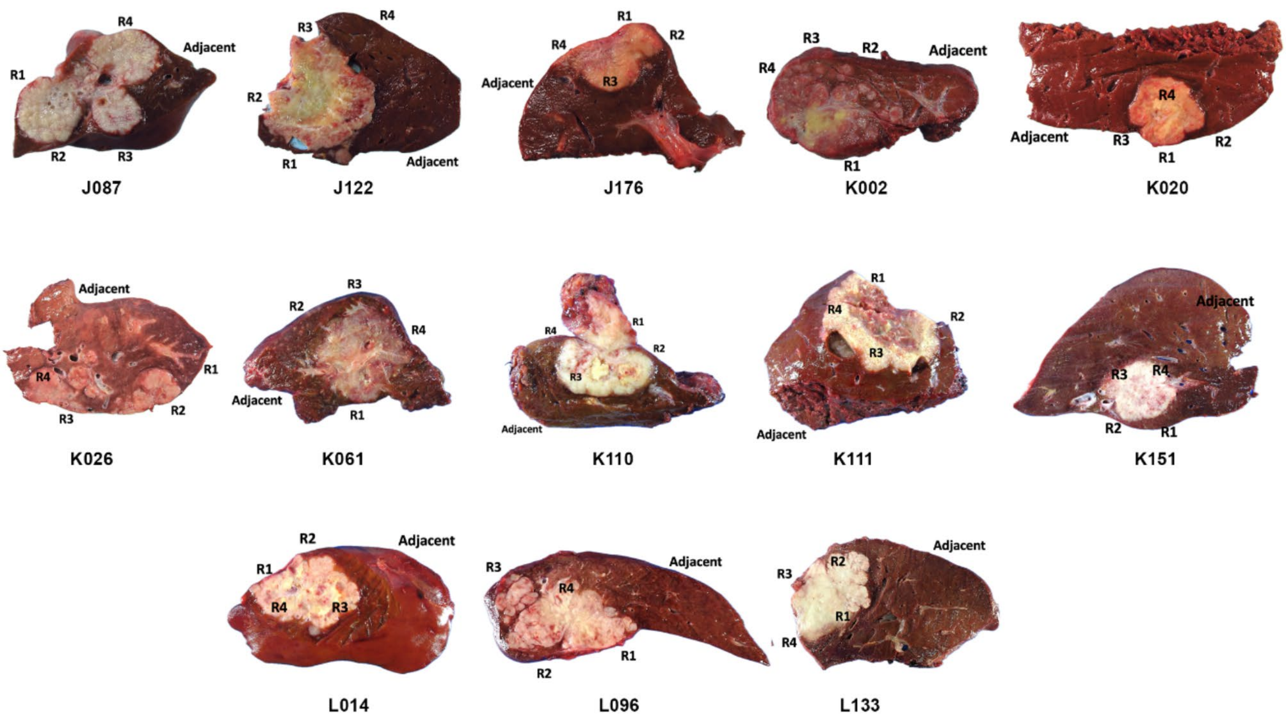

Fig. 2. We conducted a multi-regional sampling procedure on excised tumors. Each primary tumor was sampled from four distinct locations that were geographically separated, as well as one region close to the cancer. The distance between each sampling site and the tumor was greater than 2 cm.

Library preparation and multi-region exome sequencing
The Agilent SureSelect Human All Exon V6 (Agilent, USA) was used to build sequencing libraries, using the manufacturer's optimized technique. The NovaSeq 6000 platform (Illumina®, Novogene, Singapore) was used to undertake whole-exome sequencing, employing paired-end sequencing, at read length of 150 bp.

Multi-region whole-exome sequencing data processing
The paired-end sequencing reads were aligned to the reference human genome (build hg38) using the Burrows–Wheeler Aligner (BWA)[18]. The aligned data underwent further processing using the Genome Analysis Toolkit (GATK) programs[19], which involved eliminating duplicate reads resulting from PCR amplification, locally realigning around indels, and applying extra quality filtering. For somatic variant calling, we used Mutect2 from GATK (version 4.1.2) with default settings, matching it with adjacent tissues. To select the most reliable somatic calls, FilterMutectCalls with default parameters was applied on the output from Mutect2. The discovered mutations in the generated variation files were annotated and filtered using the FUNCOTATOR tool in GATK, with an additional filter applied using VAF > 0.1. In order to discover potential cancer driver genes, we compared mutations with the COSMIC cancer mutation census database[20], as well as genes that have been frequently mutated in prior CCA research investigations.

Somatic CNV identification and tumor purity estimation
CNVs were identified from the aligned whole-exome sequencing data using the CNVkit[21]. The CNVkit software was conducted using a batch command to analyze a total of 52 tumor samples together with their corresponding normal samples. Each input sample underwent median-centering of the data, followed by adjustments for read-depth biases. Scatter and heatmap plots were created to visually represent the discovered CNVs. The GISTIC2.0 algorithm was used to detect genomic areas that exhibited substantial amplification or deletion in the sample set[21,22]. The CNVs, which measures the extent of CNVs, was computed for each patient using a previously established technique, based on the discovered CNVs[23].

Mutational signature analysis
The Mutalisk[24] was employed to analyze the mutational signature in order to uncover the biological process in CCA. The cosine similarity score was used to compare the observed mutational spectrum across the 96 mutation classes with anticipated spectrum derived from the final combination of known signatures.

Intratumor heterogeneity (ITH) analysis
To assess mutational heterogeneity, we analyzed 13 patients with CCA. Pairwise comparisons were performed between different tumor regions within each patient, and mutations were classified into shared (present in multiple regions) and private (unique to a specific region) categories. For copy number alterations (CNAs), we evaluated heterogeneity across 13 patients by analyzing the CNA profiles of multiple tumor regions per patient. CNA segments with partial overlap between different tumor regions within the same patient were divided into short consecutive segments, generating a dataset that retained only overlapping or non-overlapping CN segments between any two tumor regions within a patient.

Cancer cell fraction and clone phylogeny estimate
The ABSOLUTE tool[25] was used to determine the proportion of cancer cells with each mutation. This technique considers the copy number data and tumor purity information. The PyClone-Vi programs was utilized to deduce the subclonal makeup and recreate the evolutionary history of each patient's tumor. It categorizes the genetic changes into several clusters that reflect various subpopulations within the larger population[26]. The input data for PyClone-Vi consisted of the cancer cell fractions determined by ABSOLUTE. These fractions were then transformed into a table of read counts for single nucleotide variants (SNVs) according to the instructions provided by the programs. Only single nucleotide variants (SNVs) that met specific requirements were used for the study to guarantee accurate calculation of subclones. These criteria included the SNVs being present on portions of the genome that have two copies (diploid heterozygous regions), having a minimum total read count of 20, a mutant read count higher than 2, and a variant allele frequency larger than 0.02. MesKit was executed to generate a subclonal phylogenetic tree and determine the frequency of each subclonal cluster[27]. These results were then visualized to facilitate comprehension.

Results
Patient characteristic
Patients who had resectable procedures at Srinagarind Hospital were included in this group as shown in Table 1. Samples of CCA patients were collected from 13 patients, with 4 tissue regions per patient totaling 52 samples, along with adjacent tissues. The patients ranged in age from 53 to 77 years, with an average age of 64 years. The cohort consisted of 7 males and 6 females. The majority of patients were from Northeastern Thailand, an *OV*-endemic area, where liver fluke infection is a major risk factor for CCA development[17]. Pathological staging of the tumors revealed that most patients (76.92%) were diagnosed at stage III, followed by stage II (23.08%), highlighting the late-stage detection of the disease. The post-operative survival period varied considerably, ranging from 54 to 1694 days, with a median survival of 755 days. Regarding biochemical markers, CA 19–9 levels varied significantly among patients, with some exceeding 1000 U/mL. Total bilirubin levels ranged from 0.4 to 4.6 mg/dL, while ALP, AST, and ALT levels also exhibited variability, reflecting differences in liver function and disease severity among the patients.

Patient ID	Gender	Age	Nationality	Place of residence	Stage	Survival (days)	CA 19–9 (U/mL)	Total Bilirubin (mg/dL)	ALP (U/L)	AST (U/L)	ALT (U/L)
J087	F	65	Thai	Northeastern	III	1694	8.6	0.4	117	23	25
J122	M	66	Thai	Northeastern	III	663	2	0.6	176	35	21
J176	M	57	Thai	Northeastern	II	322	> 1000	4.6	1631	139	95
K002	F	54	Thai	Northeastern	III	369	> 1000	0.6	122	59	40
K020	M	72	Thai	Northeastern	II	1448	14.98	0.4	50	21	16
K026	F	65	Thai	Northeastern	II	805	2	3.7	388	85	43
K061	M	77	Thai	Northeastern	III	591	> 1000	1.2	239	52	40
K110	F	58	Thai	Northeastern	III	120	2	0.6	177	21	15
K111	M	73	Thai	Northeastern	III	508	> 1000	0.9	78	27	22
K151	M	53	Thai	Northeastern	III	1157	< 0.6	0.8	82	48	54
L014	F	68	Thai	Northeastern	III	1121	263	0.2	81	20	9
L096	F	64	Thai	Northeastern	III	969	0.95	0.4	102	56	59
L133	M	62	Thai	Northeastern	III	54	422.2	1.1	442	76	42

Table 1. Patient characteristic.

Mutational landscape of 13 CCA patients using multi-region exome sequencing

A total of 65 multi-region samples, 52 from the tumor areas and 13 from the adjacent area, underwent WES with an average sequencing depth of 180X shown in supplementary Table S1. The germline DNA control was obtained from adjacent tissues at least 2 cm away from the tumor edge. This tissue was visually examined by independent cancer pathologists and found to be free of malignant cells. The tumor purity and tumor ploidy of different regions for each patient are presented in Supplementary Table S2. The WES analysis obtained an average sequencing depth of 180X for samples, with a majority of reads mapped on the exons. Supplementary Tables S3–S15 includes extensive data that specifically describe the nonsynonymous somatic mutations found in each region examined from the CCA patients. For each patient, we discovered an median of 95 mutations shown in Fig. 3A. (with a range of 36–154) from several regions. In the SNV class, the majority of SNVs are C > T, as seen in Fig. 3B. The results of the variant classification indicate that the majority of variant classifications are missense mutations, which contribute to 3902 mutations out of a total of 4810 mutations that were nonsynonymous as seen in Fig. 3C. The rest of the nonsynonymous mutations were 244 nonsense mutations, 281 splice site variations, and 298 deletions and 81 insertions (Fig. 3C).

An Oncoplot generated from the 52 regions across the 13 CCA patients revealed that the most frequently mutated genes were *TP53* (52%), *SMAD4* (23%), *TTN* (23%), *ARID1A* (21%), *ITIH5* (21%), and *ZFHX4* (21%). The frequency was calculated based on the number of mutations found per region relative to the total number of regions analyzed across all patients in the study. The comparative investigation of mutational profiles across patients demonstrated inter-patient heterogeneity, with each individual patient displaying unique mutation patterns. Furthermore, the comparison of mutational landscapes across different regions within the same patient uncovered ITH, with distinct mutational patterns evident among the tumor regions of a single individual (Fig. 3D).

Mutational signatures in multi-regional sequencing

Somatic mutations can be caused by various factors, such as exposure to external and internal mutagens, faulty DNA repair mechanisms, and errors in DNA replication. These mutation patterns can provide insights into the causes, prevention, and treatment of cancer. In order to gain a deeper understanding of the mutational processes involved in the progression of CCA, we analyzed the types of mutations that occur in CCA tumors using the Mutasik Map tool[24] to map mutations to the COSMIC signatures[28]. Mutational signature analysis performed on the CCA patients revealed that the predominant single base substitution (SBS) signatures observed in this cancer type were SBS1, SBS3, and SBS6 (Fig. 4). Specifically, SBS1 was associated with spontaneous deamination of 5-methylcytosine (a clock-like signature); SBS3 was linked to defective homologous recombination DNA damage repair; and SBS6 was indicative of defective DNA mismatch repair mechanisms based on their proposed etiologies. From the overall landscape of 52 mutational signatures, it is evident that each sample and each patient exhibit distinct proportions and compositions of mutational signatures. Consequently, the mutational signatures observed in each tumor sample reflect the unique interplay of these mutagenic forces, shaping the genomic landscape in a patient-specific manner. This finding underscores the inherent heterogeneity in the mutational processes underlying cancer development, which can be attributed to a multitude of factors, including exposure to various endogenous and exogenous mutagens, deficiencies in DNA repair mechanisms, and the intrinsic genomic instability of neoplastic cells.

The heterogeneity of CNA and chromosomal instability of CCA

Subsequently, we examined the variation in copy numbers. Although the significant level of ITH observed in somatic mutations, CNAs exhibited a lower degree of ITH. We detected deletions on chromosomes 5q, 9q, 10p, 13q, 14q, 17p, 18q, 19p, 21p, 21q and 22q (deletion: < = 1 copy number relative to ploidy) (Fig. 5A). This deletion harbored well-known tumor suppressors, such as *TP53* (17p13.1), *CDKN2A* (9p21.3), *ARID1A* (1p36.11) and *APC* (5p22.2). Previous genome sequencing investigations have identified *ARID1A* as a driver of intrahepatic

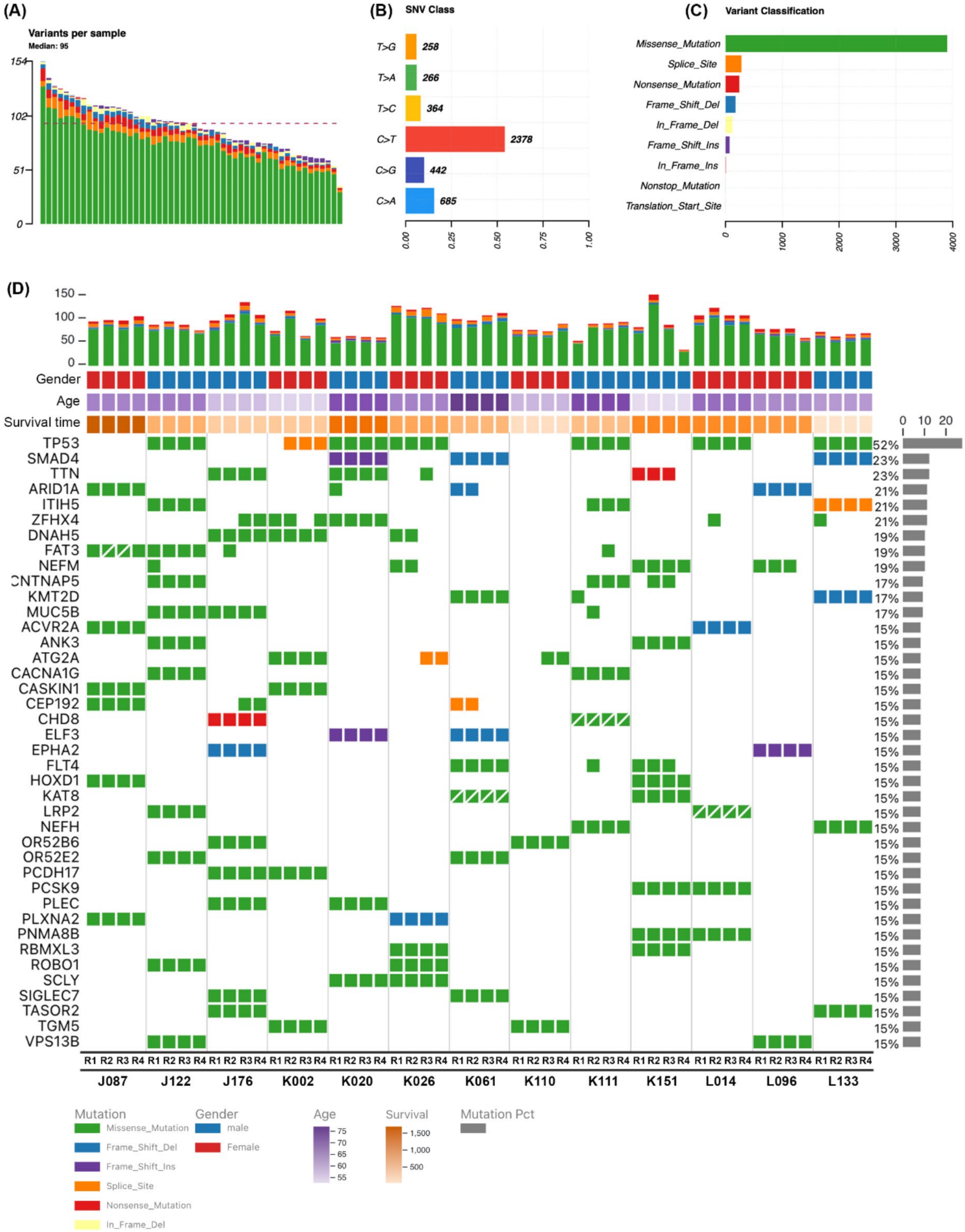

Fig. 3. Mutation spectrum of CCA. (**A**) Variant per sample, (**B**) Single Nucleotide Variant (SNV) class, (**C**) Variant classification, (**D**) The oncoplot displays the distribution of mutations detected in our samples among the most often altered genes in CCA. Each column in the dataset represents a single sample, with four sections per tumor. Each row corresponds to a separate gene. The presence of colored squares indicates the presence of altered genes, whereas the absence of squares indicates the absence of mutant genes. The various mutations are categorized by their variant type and shown by different colors: orange for splice site mutations, blue for frameshift deletions, green for missense mutations, red for nonsense mutations, and White diagonal for multi-hit mutations. Genes classified as "multi-hit" exhibit many mutations within the same region. The bar graph on the right displays the aggregate count and proportion of altered areas for each gene, out of a total of 52 regions. The colors of the bars correspond to different types of mutations. The upper graph displays the aggregate count and categorization (represented by distinct colors) of mutations in each tumor area.

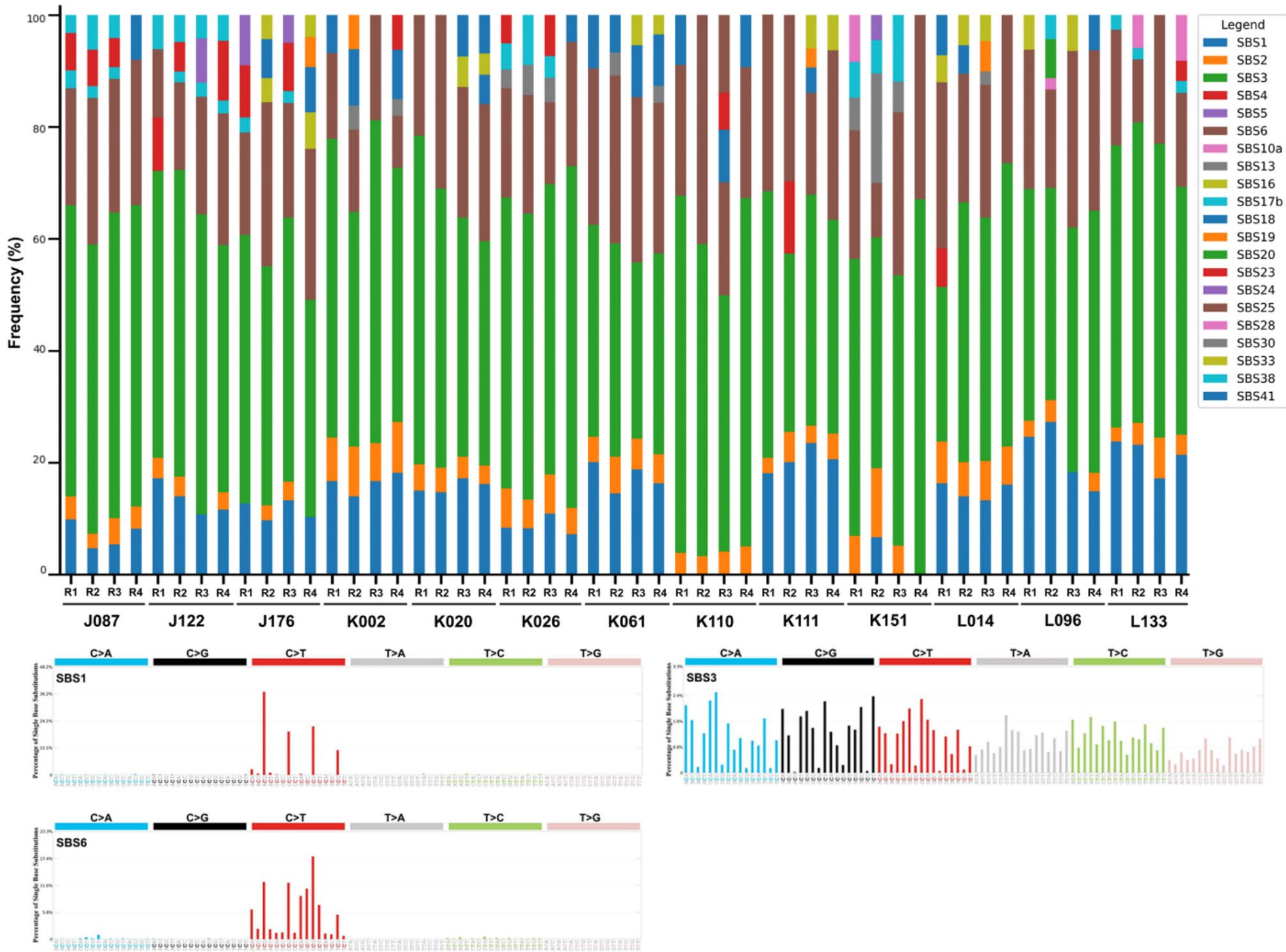

Fig. 4. Illustrates that multi-region whole-exome sequencing of CCA uncovers both mutational heterogeneity and distinct mutational signatures.

cholangiocarcinoma (iCCA)[29,30]. We detected significant amplification on chromosome 5p, 7p, 8q, and 19q, with amplification levels equal to or greater than twice the ploidy (Fig. 5A). Known oncogenes, such as *EGFP* (7p11.2), *TERT* (5p15.33), *STK* (19p13.3), *CDK6* (7q21.2), *MET* (7q31.2), *MYC* (8q24.21) and *HOXA9* (7p15.2) are identified within this region. Furthermore, we compared the somatic copy number alterations (SCNAs) in our group of patients with those in another group of patients in TCGA (The cancer genome atlas) (https://www.cancer.gov/tcga) with CCA using the default approach based on GISTIC 2.0 (Fig. 5B). Figure 5A and B show the CNA results. Figure 5A illustrates the distribution of copy number gains and losses across the chromosomes. Figure 5B presents the GISTIC analysis, which identifies significant regions of amplification and deletion. The combined results provide a comprehensive view of the genomic alterations, emphasizing the key regions that may play a crucial role in cancer development. PlotCNA of 13 patients was also generated to illustrate CNA profiles (Fig. 5C). The plotCNA function is capable of characterizing the CNA landscape across samples using copy number data obtained via segmentation techniques. Comparison of the PlotCNA results across patients revealed distinct CNA profiles for each individual, highlighting inter-patient heterogeneity in the patterns of CNA landscapes (Fig. 5C). The CNA analysis showed that patient K110 exhibited a markedly higher degree of copy number alterations compared to all other patients in the cohort (Fig. 5C).

Mutation analysis revealed varying degrees of intratumor heterogeneity (ITH) among the patients

The examination of the geographical spread of non-synonymous mutations in each patient revealed that the degree of ITH in patients with CCA exhibited notable variations (Fig. 6). Ubiquitous mutations were characterized as mutations that were found in all instances of the condition, and their prevalence varied from 9.4 to 51.8% across various individuals (Fig. 6). Obvious ITH was observed in all CCA patients. Notably, K026 and K151, with the percentages of ubiquitous mutations at 15.6% and 9.4% respectively, had fewer ubiquitous mutations. Thus, these data show that sequencing of the primary lesion alone, or any individual lesion, is insufficient to fully characterize the genomic landscape of CCA in certain patients.

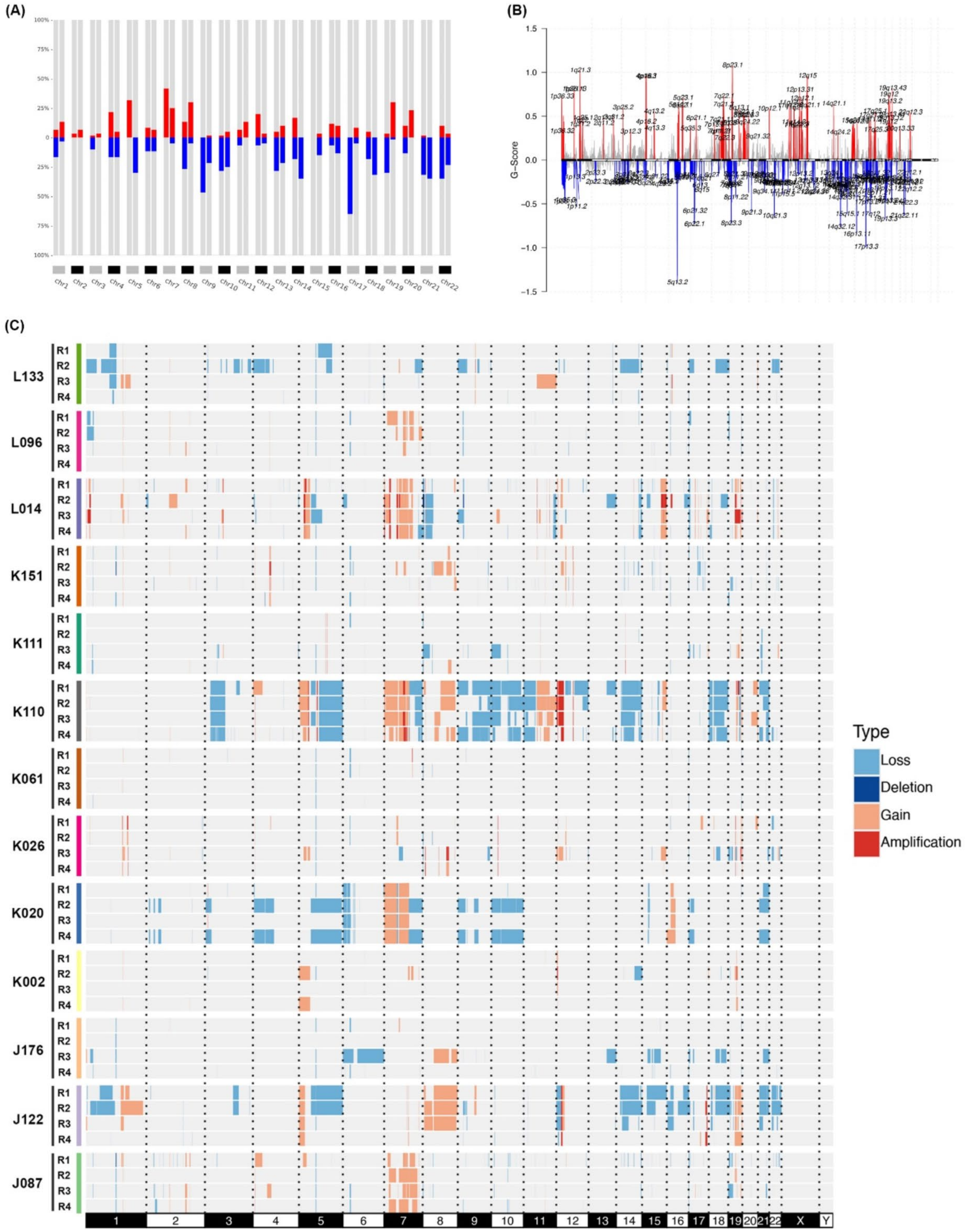

Fig. 5. (**A**) Perform CNV analysis by examining the frequencies of CNApp on both the p and q arms of each chromosome. Alteration frequency is defined as the proportion of areas that have undergone alteration, with red indicating gains and blue indicating losses. (**B**) Regions with cumulative copy number variations (CNVs). Deletions are depicted in the color blue, whereas amplifications are depicted in the color red. (**C**) PlotCNA displays the distribution of copy number alteration detected in our samples among the most often altered genes in CCA.

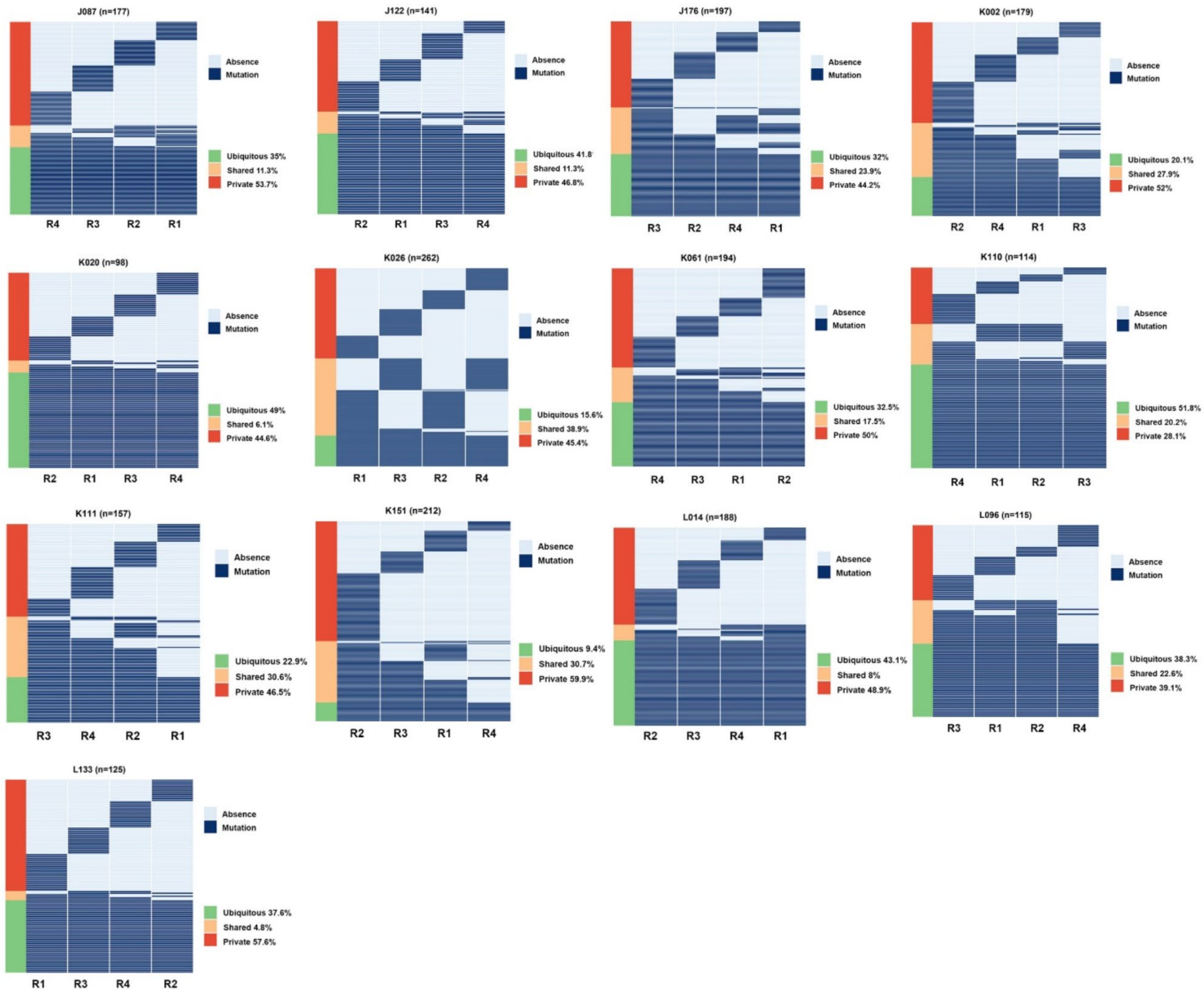

Fig. 6. The heatmap displays the spatial pattern of somatic mutations identified by exome sequencing. The presence of a mutation is shown by the dark blue colour, whereas the lack of a mutation is indicated by the light blue colour. The colour bar on the right side of the heatmap represents the classification of mutations based on whether they are ubiquitous, shared, or unique to the region (private).

CCA exhibited evidence of branched evolution

In order to examine the common evolutionary trends in CCA, we created a multi-region tree of tumors using somatic single-nucleotide variations (SNVs). Additionally, we utilized copy number alteration (CNA) data to reinforce the final phylogenetic trees. The visualization of branched evolution can be accolated through the construction of phylogenetic trees based on mutations (Fig. 7). It is understood that trunk mutations occur before branch mutations in the development of tumors[31–33]. The trees in K026 and K151 exhibited characteristics of abbreviated trunks and elongated branches, suggesting that multi-region metastasis occurred at an early stage. In contrast, the trees of other patients exhibited elongated trunks and abbreviated branches, suggesting that multi-region originated at a later stage in tumor progression. We identified possible driver mutations on the trees to accurately represent when they were acquired (Supplementary Table S16). Common driver mutations of CCA[15,34–36], such as those in *TP53*, *ARID1A*, *ERBB2* and *SMAD4* were often found in the trunks. However, in specific patients, driver mutations, including mutations in *KRAS*, *FAT3*, *KMT2C*, *KMT2D*, and *APC* were only found in certain lesions, suggesting that they were located on the branches and are likely acquired at a later stage in the development of disease and have significant impacts on the branching evolution of tumors. Therefore, the frequency of driving events in the tumors of certain individuals may have been overestimated when just the main tumor was subjected to sequencing.

Discussion

The evolutionary trajectory of tumors driven by chronic parasitic infection, such as those associated with *OV*, remains underexplored. This study provides a comprehensive, multi-region genomic analysis that reveals the intricate genetic diversity and evolutionary dynamics within CCA tumors linked to *OV* infection. Our

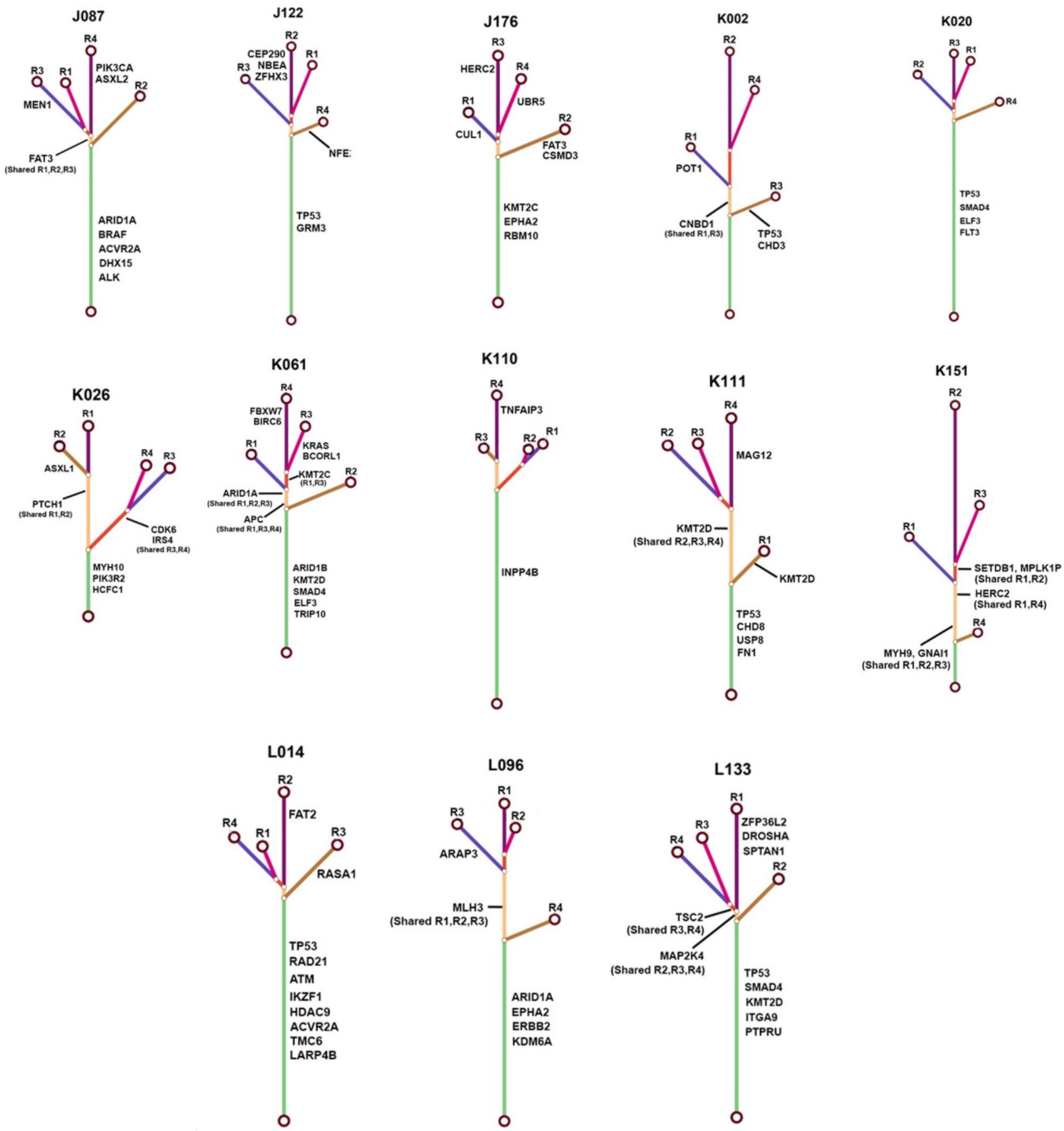

Fig. 7. Phylogenetic relationships among the tumor regions. The lengths of the branches are directly proportional to the number of somatic mutations that separate the places where the branches split. The marked genes in the branch received potential driver mutations.

examination across multiple tumor regions in 13 CCA patients demonstrates a high degree of ITH and inter-patient variability, highlighting the complex and unique mutational landscape associated with this infection-driven cancer.

Remarkably, between 48 and 90% of mutations identified through multi-region sequencing were heterogeneous, indicating they were not uniformly present across all examined regions. This extensive variability aligns closely with previous findings in intrahepatic cholangiocarcinoma (iCCA), such as those reported by Dong et al.[37]. Our analysis further reveals an average of 95 non-synonymous mutations per region, surpassing mutation counts reported in earlier CCA sequencing studies[15,34–36], thereby underscoring the heightened mutational burden in *OV*-associated CCA. In comparison, Walter et al.[38] observed private mutations across two regions per iCCA sample, with a mean ratio of private to non-synonymous mutations per sample at 12% (range:

0–57%). Our findings align with this study, demonstrating a comparable prevalence of private mutations across different tumor regions, thus reinforcing the profound heterogeneity present in CCA driven by chronic infection. Among the 52 regions analyzed from the 13 CCA patients, *TP53*, *SMAD4*, *ARID1A*, *ITIH5*, and *ZFHX4* emerged as the most commonly altered genes, indicating their central roles in the pathogenesis of CCA. Interestingly, our data revealed unique differences in both the implicated genes and their mutation frequencies when compared to previous research on liver fluke-associated CCAs. For instance, Ong et al.[34] reported high mutation rates in *TP53* (44.4%), *KRAS* (16.7%), *SMAD4* (16.7%), *MLL3* (14.8%), *ROBO2* (9.3%), *RNF43* (9.3%), *GNAS* (9.3%), and *PEG3* (5.6%). Results from our study, however, identified different mutation patterns, suggesting possible geographic or etiological variations in CCA driven by *OV*.

The mutational signature analysis of our CCA cohort revealed predominant single base substitution (SBS) signatures SBS1, SBS3, and SBS6 which are associated with spontaneous deamination (SBS1), defective homologous recombination repair (SBS3), and defective DNA mismatch repair (SBS6), respectively. These findings align with the work of Lin et al.[39], which identified similar SBS signatures in CCA, reinforcing the reproducibility and relevance of these mutational processes in this cancer type. By employing non-negative matrix factorization (NMF) analysis on whole exome sequencing (WES) data from intrahepatic cholangiocarcinoma (iCCA) patients, we extracted three distinct mutational signatures. One signature demonstrated a dominant pattern of C > T transitions, while another was strongly associated with aristolochic acid (AA) exposure, suggesting a substantial environmental influence on mutational processes within this cohort.

Crucially, our comparative analysis between *OV*-associated CCA and non-*OV* CCA reveals strikingly distinct mutational landscapes. *OV*-associated CCA is enriched with age-related and APOBEC-related mutagenesis, underscoring the impact of chronic liver fluke infection on tumorigenesis. The persistent inflammation and sustained damage to the biliary epithelium caused by *OV* infection result in unique genetic alterations, emphasizing the profound role of a chronic inflammatory environment in shaping the mutational profile of CCA. These findings are consistent with previous studies[40,41], and they underscore the necessity of accounting for infection-associated mutational signatures when studying CCA.

The identification of these distinct mutational signatures enhances our understanding of the molecular mechanisms driving tumorigenesis in CCA and opens avenues for personalized therapeutic strategies. The unique mutational landscapes, particularly in *OV*-associated cases, hold potential for tailored interventions that target the specific mutational processes at play in individual patients. Comprehensive mutational profiling thus becomes a pivotal tool in advancing personalized medicine for CCA, allowing clinicians to design treatments that directly address the underlying mutational drivers in each patient's tumor.

Our CNA analysis revealed that patient K110 exhibited markedly elevated levels of chromosomal alterations compared to all other patients in the cohort. This significant genomic instability may indicate a more aggressive tumor biology with potential implications for clinical outcomes. However, while patient K151 demonstrated the lowest survival rate within the cohort, we acknowledge that our study does not establish a direct causal relationship between chromosomal alterations and survival outcomes. Although prior findings in non-small cell lung cancer (NSCLC) by Jamal-Hanjani et al.[14] suggest that ITH driven by chromosomal instability is associated with an increased risk of recurrence and mortality, further validation in a larger, independent cohort is required to confirm the prognostic significance of chromosomal instability in CCA. Together, these observations suggest that chromosomal instability, as a driver of ITH, may serve as a valuable prognostic marker, guiding risk stratification and therapeutic decisions in CCA.

Our data provide promising evidence that specific driver mutations, such as ERBB2, may serve as actionable targets in CCA. Given the availability of ERBB2-targeted therapies, such as Trastuzumab, further clinical investigation into ERBB2-positive CCA could open new avenues for treatment. Additionally, other frequently observed driver mutations in genes like *TP53*, *SMAD4*, and *ARID1A*, which are typically found in the trunks of the phylogenetic trees, underscore their foundational role in tumor progression. This suggests potential therapeutic strategies that modulate pathways involving *TP53*, TGF-β signaling (*SMAD4*), and chromatin remodeling (*ARID1A*). In contrast, mutations such as *KRAS*, *FAT3*, *KMT2C*, *KMT2D*, and *APC* were identified on branches, representing later events in tumor evolution. These branch mutations contribute to subclonal diversity, highlighting the need for combination therapies that address both early clonal drivers and later subclonal mutations to effectively target tumor heterogeneity. Previous studies on ITH in CCA have shown that this type of cancer is marked by substantial genetic diversity within tumors. These studies indicate that different regions within the same tumor can harbor unique genetic mutations, contributing to the complexity of the disease and impacting treatment outcomes[42]. Our findings are consistent with the previous study by Dong and team, who found that branch evolution is the predominant pattern in iCCA, collectively shaped by parallel evolution and chromosome instability[37]. Together, these findings underscore the possibility of repurposing existing therapeutics for CCA patients based on their unique molecular profiles.

In summary, our findings emphasize the clinical relevance of both trunk and branch mutations, demonstrating how understanding the evolutionary trajectory of CCA can inform personalized treatment. Future research should focus on expanding patient cohorts to validate these pathways' roles, investigate the therapeutic potential of targeting trunk mutations, and explore combination strategies that address sub-clonal resistance mechanisms. This approach could enhance prognostic accuracy and optimize therapeutic outcomes for CCA patients, particularly in *OV*-endemic regions where this malignancy is most prevalent.

Data availability

The user provided the following information on the availability of data: The sequencing data may be accessed at NCBI under the project identifier PRJEB47824.

Received: 10 November 2024; Accepted: 19 March 2025

Published online: 29 March 2025

References

1. Banales, J. M. et al. Cholangiocarcinoma 2020: The next horizon in mechanisms and management. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 557–588. https://doi.org/10.1038/s41575-020-0310-z (2020).
2. Rizvi, S., Khan, S. A., Hallemeier, C. L., Kelley, R. K. & Gores, G. J. Cholangiocarcinoma—Evolving concepts and therapeutic strategies. *Nat. Rev. Clin. Oncol.* **15**, 95–111. https://doi.org/10.1038/nrclinonc.2017.157 (2018).
3. Saha, S. K., Zhu, A. X., Fuchs, C. S. & Brooks, G. A. Forty-year trends in cholangiocarcinoma incidence in the U.S.: Intrahepatic disease on the rise. *Oncologist* **21**, 594–599. https://doi.org/10.1634/theoncologist.2015-0446 (2016).
4. Khan, S. A. et al. Changing international trends in mortality rates for liver, biliary and pancreatic tumours. *J. Hepatol.* **37**, 806–813. https://doi.org/10.1016/s0168-8278(02)00297-0 (2002).
5. Taylor-Robinson, S. D. et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968–1998. *Gut* **48**, 816–820. https://doi.org/10.1136/gut.48.6.816 (2001).
6. Sripa, B. et al. Liver fluke induces cholangiocarcinoma. *PLoS Med.* **4**, e201. https://doi.org/10.1371/journal.pmed.0040201 (2007).
7. Catanzaro, E., Gringeri, E., Burra, P. & Gambato, M. Primary sclerosing cholangitis-associated cholangiocarcinoma: From pathogenesis to diagnostic and surveillance strategies. *Cancers (Basel)* **15**, 4947. https://doi.org/10.3390/cancers15204947 (2023).
8. Elvevi, A. et al. Clinical treatment of cholangiocarcinoma: An updated comprehensive review. *Ann. Hepatol.* **27**, 100737. https://doi.org/10.1016/j.aohep.2022.100737 (2022).
9. Kirstein, M. M. & Vogel, A. Epidemiology and risk factors of cholangiocarcinoma. *Visc. Med.* **32**, 395–400. https://doi.org/10.1159/000453013 (2016).
10. Sripa, B. & Pairojkul, C. Cholangiocarcinoma: Lessons from Thailand. *Curr. Opin. Gastroenterol.* **24**, 349–356. https://doi.org/10.1097/MOG.0b013e3282fbf9b3 (2008).
11. Anderson, C. D., Pinson, C. W., Berlin, J. & Chari, R. S. Diagnosis and treatment of cholangiocarcinoma. *Oncologist* **9**, 43–57. https://doi.org/10.1634/theoncologist.9-1-43 (2004).
12. Khan, S. A. et al. Guidelines for the diagnosis and treatment of cholangiocarcinoma: An update. *Gut* **61**, 1657–1669. https://doi.org/10.1136/gutjnl-2011-301748 (2012).
13. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell* **168**, 613–628. https://doi.org/10.1016/j.cell.2017.01.018 (2017).
14. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121. https://doi.org/10.1056/NEJMoa1616288 (2017).
15. Jusakul, A. et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov.* **7**, 1116–1135. https://doi.org/10.1158/2159-8290.CD-17-0368 (2017).
16. Simpson, K. L. et al. A biobank of small cell lung cancer CDX models elucidates inter- and intratumoral phenotypic heterogeneity. *Nat. Cancer* **1**, 437–451. https://doi.org/10.1038/s43018-020-0046-2 (2020).
17. Titapun, A. et al. High levels of serum IgG for *Opisthorchis viverrini* and CD44 expression predict worse prognosis for cholangiocarcinoma patients after curative resection. *Int. J. Gen. Med.* **14**, 2191–2204. https://doi.org/10.2147/IJGM.S306339 (2021).
18. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 (2009).
19. McKenna, A. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. https://doi.org/10.1101/gr.107524.110 (2010).
20. Tate, J. G. et al. COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947. https://doi.org/10.1093/nar/gky1015 (2019).
21. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873. https://doi.org/10.1371/journal.pcbi.1004873 (2016).
22. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, 41. https://doi.org/10.1186/gb-2011-12-4-r41 (2011).
23. Wolf, Y. et al. UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell* **179**, 219–235. https://doi.org/10.1016/j.cell.2019.08.032 (2019).
24. Lee, J. et al. Mutalisk: A web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Res.* **46**, W102–W108. https://doi.org/10.1093/nar/gky406 (2018).
25. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421. https://doi.org/10.1038/nbt.2203 (2012).
26. Gillis, S. & Roth, A. PyClone-VI: Scalable inference of clonal population structures using whole genome data. *BMC Bioinform.* **21**, 571. https://doi.org/10.1186/s12859-020-03919-2 (2020).
27. Liu, M. et al. MesKit: A tool kit for dissecting cancer evolution of multi-region tumor biopsies through somatic alterations. *Gigascience* https://doi.org/10.1093/gigascience/giab036 (2021).
28. Sondka, Z. et al. COSMIC: A curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.* **52**, D1210–D1217. https://doi.org/10.1093/nar/gkad986 (2024).
29. Jiao, Y. et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* **45**, 1470–1473. https://doi.org/10.1038/ng.2813 (2013).
30. Zou, S. et al. Mutational landscape of intrahepatic cholangiocarcinoma. *Nat. Commun.* **5**, 5696. https://doi.org/10.1038/ncomms6696 (2014).
31. Gerlinger, M. et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233. https://doi.org/10.1038/ng.2891 (2014).
32. de Bruin, E. C. et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256. https://doi.org/10.1126/science.1253462 (2014).
33. Zhang, J. et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259. https://doi.org/10.1126/science.1256930 (2014).
34. Ong, C. K. et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nat. Genet.* **44**, 690–693. https://doi.org/10.1038/ng.2273 (2012).
35. Nakamura, H. et al. Genomic spectra of biliary tract cancer. *Nat. Genet.* **47**, 1003–1010. https://doi.org/10.1038/ng.3375 (2015).
36. Chan-On, W. et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat. Genet.* **45**, 1474–1478. https://doi.org/10.1038/ng.2806 (2013).
37. Dong, L. Q. et al. Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma. *J. Hepatol.* **69**, 89–98. https://doi.org/10.1016/j.jhep.2018.02.029 (2018).
38. Walter, D. et al. Intratumoral heterogeneity of intrahepatic cholangiocarcinoma. *Oncotarget* **8**, 14957–14968. https://doi.org/10.18632/oncotarget.14844 (2017).
39. Lin, J. et al. Mutational spectrum and precision oncology for biliary tract carcinoma. *Theranostics* **11**, 4585–4598. https://doi.org/10.7150/thno.56539 (2021).

40. Loilome, W. et al. The hallmarks of liver fluke related cholangiocarcinoma: Insight into drug target possibility. *Recent Results Cancer Res.* **219**, 53–90. https://doi.org/10.1007/978-3-031-35166-2_4 (2023).
41. Labib, P. L., Goodchild, G. & Pereira, S. P. Molecular pathogenesis of cholangiocarcinoma. *BMC Cancer* **19**, 185. https://doi.org/10.1186/s12885-019-5391-0 (2019).
42. Chen, S. et al. Multiomic analysis reveals comprehensive tumor heterogeneity and distinct immune subtypes in multifocal intrahepatic cholangiocarcinoma. *Clin. Cancer Res.* **28**, 1896–1910. https://doi.org/10.1158/1078-0432.CCR-21-1157 (2022).

Author contributions

S.S. conducted the experiments, analyzed the data, created figures and tables, contributed to the writing of the manuscript, and approved the final version. A.W. analyzed the data, created visual representations, contributed to the writing, and gave final approval to the document. A.J. analyzed the data, prepared figures and tables, wrote or reviewed versions of the manuscript, and approved the final document. P.K. analyzed the data, contributed to the drafting of the article, and approved the final version. H.D. analyzed the data, contributed to drafting, and approved the final version. N.N. analyzed the data, contributed to writing and reviewing, and approved the manuscript. V.T. conducted data analysis, contributed to writing and revision, handled specimen inquiries and visualization, and approved the final draft. A.T. performed data analysis, prepared or revised versions of the manuscript, handled specimen inquiries and visualization, and approved the final version. A.Ja. conducted data analysis, contributed to writing and revision, handled specimen inquiries and visualization, and approved the final document. N.K. analyzed the data, contributed to writing and revision, managed specimen inquiries and visualization, and approved the document. P.S. conducted data analysis, created figures and tables, contributed to writing and reviewing, managed specimen inquiries and visualization, and approved the final text. T.B.T. contributed to writing and reviewing versions of the manuscript and approved the final text. L.B. was responsible for the conception and design of experiments, contributed to writing and reviewing, and approved the final text. Y.M. contributed to writing and reviewing the manuscript and approved the final text. W.L. oversaw the conception and design of experiments, secured funding, administered the project, executed experiments, analyzed data, created figures and tables, contributed to writing and reviewing, and approved the final text. All authors reviewed the manuscript.

Funding

The funding for this study was provided by the National Research Council of Thailand through the Fluke Free Thailand Project, the NSRF under the Basic Research Fund of Khon Kaen University and the IMSUT Joint Research Project, the Institute of Medical Science, University of Tokyo, Japan for Watcharin Loilome. Furthermore, Sirinya Sitthirak was granted financial assistance from the Invitation Research Grant (IN67067). Sirinya Sitthirak received a scholarship from the Cholangiocarcinoma Research Institute, specifically under Grant No. CARI01/2565. The funders were not involved in the study's design, data collection and analysis, decision to publish, or preparation of the publication.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The provided information pertains to the ethical approvals, including the authorizing authority and any associated reference numbers. The procedure for specimen collection and study received approval by the Ethic Committee of Human Research at Khon Kaen University (HE671356).

Consent to participate

All authors have given their agreement for the publishing of this work.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95142-3.

Correspondence and requests for materials should be addressed to W.L.