Exploring the links between genomic alterations and patient outcomes in biliary tract cancer

Dr Valerie Elizabeth Crolley

Research Department of Oncology

University College London Cancer Institute

Faculty of Medical Sciences

A thesis submitted for the degree of

Doctor of Philosophy

University College London

PhD Supervisors:

Prof John Bridgewater

Prof Javier Herrero

Dr Crispin Hiley

Declaration

I, Valerie Elizabeth Crolley, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Valerie Crolley

5th November 2024

Abstract

Biliary tract cancers (BTC) are rare but aggressive cancers of the biliary tract. Despite recent developments and newly licenced treatments, the outcomes for most BTC patients remains poor.

The BILCAP clinical trial established adjuvant capecitabine as the standard of care treatment after BTC resection. Translational work from this trial involved collecting archived fixed-formalin tissue from consented BILCAP patients and carrying out low-pass whole genome (lp-WGS), targeted gene (TGS) and RNA sequencing (RNA-seq) for copy number (CN), mutation and gene-fusion analysis.

Nearly all the alterations investigated did not significantly affect recurrence risk (PFS) or overall survival (OS), including FGFR2 fusions (OS hazard ratio (HR) 0.86 p=0.6, PFS HR 0.87 p=0.6). However, the presence of amplified EGFR (CN ≥ 4) significantly decreased both OS (HR 2.54 p=0.04) and PFS (HR 2.38 p=0.03).

In a comparison with RNA-seq data from I3O-MC-JSBF, a phase II clinical trial of patients with locally-advanced or metastatic BTC, the tumour microenvironment (TME) of patients with early-stage BTC had a significantly higher proportion of regulatory T-cells but significantly lower proportions of CD4+ T-cells, dendritic cells and neutrophils.

The BILCAP cohort shows a wide variety of driver and potentially targetable mutations in unselected BTC patients, comparable to similar datasets. Patients with EGFR amplification had significantly reduced OS and PFS, indicating that EGFR amplification may be an important indicator in determining prognosis and could provide an attractive target for future targeted anti-cancer therapy in BTC.

Overall, the TME shows differences between early-stage and metastatic BTCs, with early-stage cancers having a more suppressive TME and advanced BTCs having a more inflammatory TME. This potentially explains the success of using immunotherapy in locally-advanced and metastatic BTC (as seen in the TOPAZ-1 and KEYNOTE-966 clinical trials), but may have implications as to the potential success of immunotherapy in the early-stage and adjuvant clinical settings.

Impact Statement

This project marks one of the first investigations into the genomic landscape of biliary tract cancer, and how it affects patient outcomes. While looking into the genomic landscape is not completely novel, nowhere else has it been possible to investigate it along with linked high quality clinical data available from a clinical trial.

Using data from the phase III BILCAP clinical trial is also novel; not only is this one of the largest phase III clinical trials in biliary tract cancer, all of the patients enrolled on this trial were treated before targeted treatments for biliary tract cancer became available. Any attempt to recreate the analysis from this project with a newer clinical trial would have information obscured by the fact that patients with a targetable alteration who received a targeted treatment did better than they otherwise would have. The patients on BILCAP did not have this opportunity, so the outcomes associated with varying genomic alterations seen in this study are therefore unaffected by the effect of targeted treatments.

Although the entire BILCAP dataset is not yet available to be analysed, the results so far have been of significant importance to those studying biliary tract cancer. This project is the first work to demonstrate with significance that patients with FGFR2 fusions do not have improved outcomes compared to those without. Demonstrating this answers an important question posed by NICE in the licencing of pemigatinib, a FGFR2 fusion inhibitor, for use in the treatment of FGFR2 fusion positive biliary tract cancer. Smaller studies had potentially indicated that biliary tract cancer patients with FGFR2 fusions have improved outcomes compared to those without, potentially indicating that drugs to target FGFR2 fusions (such as pemigatinib) had less of an effect on improving patient outcomes than it initially appeared. However, as this project shows that FGFR2 fusions do not affect outcomes, any benefit for patients given FGFR2 fusion inhibitors must be due to the FGFR2 fusion inhibitors themselves, which is of great importance and interest to pharmaceutical companies developing novel FGFR2 inhibitor drugs for biliary tract cancer. Similar findings in

this project for other targetable alterations are also of interest to the pharmaceutical industry.

Demonstrating the poorer outcomes for patients with EGFR amplification not only potentially demonstrates the first prognostic biomarker for biliary tract cancer, but also a potential treatment option. Patients with early-stage biliary tract cancer having a ‘colder’ tumour microenvironment may also have implications that adjuvant immunotherapy may be less effective than initially hypothesised, or when compared to its use in locally advanced or metastatic disease.

This project, and the projects started during the course of this PhD are novel in the field of biliary tract cancer research. The work carried out here already has significant implications for the development of new treatments and biomarkers for biliary tract cancer, and the continuation of the work from this PhD and the projects started during it will one day lead to improved outcomes for patients with biliary tract cancer.

UCL Research Paper Declaration Form

1. **For a research manuscript that has already been published**

   a) **What is the title of the manuscript?**

   Investigating alterations in cancer driver genes and other potentially targetable mutations in patients with intrahepatic cholangiocarcinoma (iCCA) treated on the randomised phase III multicentre BILCAP clinical trial.

   b) **Please include a link to or doi for the work**

   10.1200/JCO.2023.41.16_suppl.4019

   c) **Where was the work published?**

   Journal of Clinical Oncology

   d) **Who published the work?** (e.g. OUP)

   American Society of Clinical Oncology

   e) **When was the work published?**

   2023

   f) **List the manuscript's authors in the order they appear on the publication**

   Valerie Elizabeth Crolley, Rachel Guest, Andrew David Beggs, Eleanor Jaynes, Steve Thorn, Javier Herrero, Ian Tomlinson, Juan W. Valle, John Neil Primrose, John A Bridgewater.

   g) **Was the work peer reviewed?**

   Yes

   h) **Have you retained the copyright?**

   Yes

   i) **Was an earlier form of the manuscript uploaded to a preprint server?**

   No

   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   ☒

   *I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For multi-authored work, please give a statement of contribution covering all authors**

VEC wrote and presented the abstract, AB and JH assisted with the bioinformatics analysis, RG, EJ and SH assisted with data analysis from the BILCAP trial and IT, JWV, JNP, and JAB assisted with the study design.

3. **In which chapter(s) of your thesis can this material be found?**

   Chapters 2, 3 and 4.

4. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

   Candidate

   Valerie Crolley

   Date:

   5th November 2024

   Supervisor/ Senior Author (where appropriate)

   John Bridgewater

   Date

   5th November 2024

Acknowledgements

First of all, I would like to thank my supervisor, John Bridgwater, for choosing to work with me on this project. His advice and support throughout have been utterly invaluable, and I truly look forward to continuing to work with him in the future. I would also like to thank Javier Herrero for his bioinformatics support and advice. Finally, I would like to thank Crispin Hiley for his support, especially at the beginning of my PhD.

I would like to thank all of the members of the Bill Lyons Informatic Centre (BLIC) for providing support throughout this PhD, particularly Lucia Conde for providing technical support. Chuling Ding was a good friend throughout this project, as well as Mark Tran and Stephen Henderson.

I would also like to thank my other collaborators on the BILCAP project; in particular John Primrose, Andrew Beggs, Ian Tomlinson, Eleanor Jayne, Peter Fletcher, Rachel Guest, Steve Thorn and Juan Valle, without whom this project would simply not have been possible.

I would also like to thank all of the other members of ENS-CCA, CCA-UK, AMMF and the wider biliary tract cancer community for being so kind, friendly and welcoming. In particular, I would like to thank Chiara Braconi, Fede Nichetti, Pilar Acedo Nunez, Shahid Khan, Timothy Kendall and Helen Moremont.

I would also like to thank the members of UCL, Jazz Pharmaceuticals (formally Zymeworks), Eli Lilly and Genentech for their help in setting up projects and gaining access to data. This includes, but is not limited to Kavita Shah, Sam Margurat, Sharmika Basu, Megan Kane, Temire Bob-Egbe, Tracey Dorsel and Isabel Brown Freihofer.

I would also like to thank the members of the PEACE project for helping to set up that part of the project, with particular thanks to Cristina Naceur-Lombardelli, Hayley Bridger and Vinaya Gunasri.

This PhD would not have been possible without the funding provided by Huw Evans, so I cannot thank him and his family enough for giving me this opportunity. I would also like to thank Alice Pink and Alice Clasper for all of their help in making this possible.

Last but not least I would like to thank my friends, and in particular Nick, Laurie, Felix and Anjui for listening to my rants and rambling about this project over the years. And finally (and most importantly), I would like to thank my mama, Jacqueline Crolley, without whom I would have nothing and be nothing.

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

AMMF	The Alan Morement Memorial Fund
ASCO	American Society of Clinical Oncology
BAM	Binary alignment map
BILCAP	Capecitabine compared with observation in resected biliary tract cancer
BLIC	Bill Lyons Informatics Centre
BTC	Biliary tract cancer
BWA	Burrows-Wheeler aligner
CCA	Cholangiocarcinoma
CCNE1	Gene encoding G1/S-specific cyclin-E1
cfDNA	Cell-free DNA
ctDNA	Circulating tumour DNA
dCCA	Distal cholangiocarcinoma
CT	Computed tomography
DNA	Deoxyribonucleic acid
EGA	European Genome-Phenome Archive
EGFR	Epidermal growth factor receptor
EORTC	European Organisation for Research and Treatment of Cancer
ERBB2	Gene encoding the HER2 protein
ESCAT	ESMO Scale for Clinical Actionability of molecular Targets
FASTQ	Text file containing genomic sequence data
FFPE	Formalin-Fixed paraffin-embedded
FGFR	Fibroblast growth factor receptor
FGFRi	Fibroblast growth factor receptor inhibitor
FISH	Fluorescence in situ hybridization
FOLFOX	Folinic acid, fluorouracil and oxaliplatin
GATK	Genome Analysis Toolkit (from the Broad Institute)
GBC	Gallbladder cancer
Gem/Cis	Gemcitabine and cisplatin
GI	Gastrointestinal tract
HER2	Human epidermal growth factor receptor 2
HPC	High-performance computing/computer
iCCA	Intrahepatic cholangiocarcinoma
ICF	Informed consent form
ICGC	International Cancer Genome Consortium
IHC	Immunohistochemistry
lp-WGS	Low pass whole genome sequencing
MAF	Mutation annotation format
MEK	Mitogen-activated protein kinase

MMR-D	Mismatch repair deficiency
MRI	Magnetic resonance imaging
MSI	Microsatellite instability
mTOR	Mammalian target of rapamycin
NGS	Next-generation sequencing
ORR	Overall response rate
OS	Overall survival
pCCA	Perihilar cholangiocarcinoma
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
PEACE	Posthumous Evaluation of Advanced Cancer Environment
PFS	Progression-free survival
PI3K	Phosphoinositide 3-kinase
PSC	Primary sclerosing cholangitis
RECIST	Response evaluation criteria in solid tumours
RNA	Ribonucleic acid
RNA-Seq	Next-generation sequencing of ribonucleic acid (RNA)
SACT	Systemic anti-cancer therapy
SoC	Standard of care treatment
TIGIT	T cell immunoreceptor with Ig and ITIM domains
T-reg	Regulatory T-cell
TKI	Tyrosine kinase inhibitor
TME	Tumour microenvironment
UCL	University College London
UK	United Kingdom
VCF	Variant call format
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Chapter 1: Introduction

1.1: What is biliary tract cancer?

Biliary tract cancers are rare but aggressive cancers of the biliary tree of the liver. This includes cancers of the bile ducts of the liver, called cholangiocarcinoma, as well as cancer of the gallbladder. Cholangiocarcinoma can be fully subdivided based on its anatomical location into intrahepatic, perihilar and distal cholangiocarcinoma (1). Intrahepatic cholangiocarcinoma arises from the second order bile ducts and above, perihilar cholangiocarcinomas arise from the second order bile ducts to the insertion of the cystic duct and distal cholangiocarcinomas arise from any of the more distal bile ducts (1). Distal and perihilar cholangiocarcinomas are also collectively referred to as extrahepatic cholangiocarcinomas.

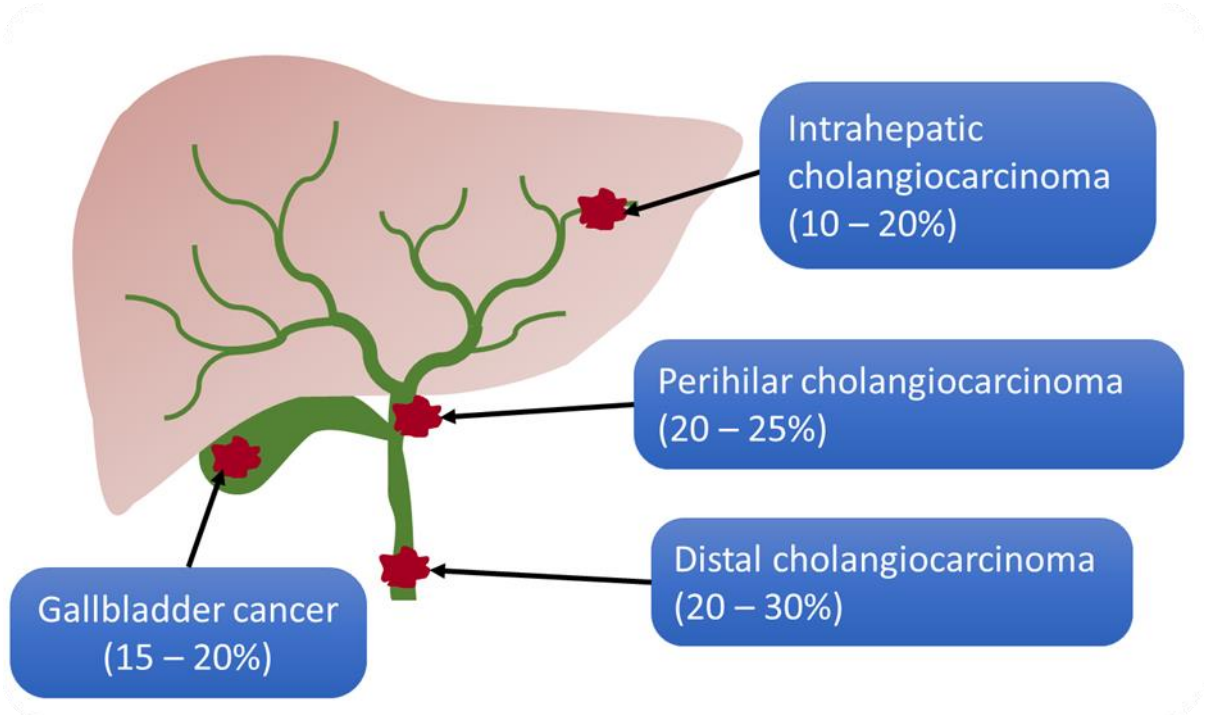


Figure 1: This diagram shows the different anatomical subtypes of biliary tract cancer, and the proportion of overall biliary tract cancers that make up each subtype (2). Redrawn by the author from (1).

The vast majority of biliary tract cancers are adenocarcinomas, and mainly arise from the mature cholangiocytes lining the biliary tree and gallbladder (1, 3).

Several risk factors have been linked to the development of biliary tract cancer, some of which are shared by all subtypes and some of which are specific to certain subtypes (1). Common causes include smoking, high alcohol intake as well as infection with hepatitis B and C (4). Gallbladder cancer is also highly linked to the presence of gallstones (1), whereas cholangiocarcinoma is highly linked to primary sclerosing cholangitis (5), bile duct stones and the presence of choledochal cysts (4).

Although biliary tract cancers are relatively rare in most developed countries (6), biliary tract cancers are very common in some parts of the world, such as south-east Asia (1, 6). The very high incidence of cholangiocarcinoma in parts of south-east Asia, including China, Thailand, Laos, Cambodia, Vietnam and South Korea is very likely to be due to the endemic liver flukes present in this part of the world. Liver flukes are small trematodes, and two species, *Clonorchis sinensis* and *Opisthorchis viverrine,* have been conclusively proven to be a major factor in the development of biliary tract cancer (1, 7). These liver flukes have a life cycle involving water snails and freshwater fish and are ingested when eating raw or lightly smoked fish. After ingestion the flukes live in the bile ducts for many years, causing localised inflammation and damage which leads to the development of cholangiocarcinoma (7).

Rates of gallbladder cancer have been falling in the Western world due to the increased number of cholecystectomies (1, 6), but the incidence of cholangiocarcinoma, and in particular intrahepatic cholangiocarcinoma, has been increasing steadily over the past 30 years (8–12) for reasons that are still not fully understood.

The different anatomical subtypes of biliary tract cancer not only differ in their potential aetiology, but also appear to differ at a genomic level (1, 13–17). Mutations in *KRAS* are common in all of the subtypes (14–16), while mutations in *IDH*, *EPHA2*, and *BAP1* along with *FGFR2* fusions are more common in intrahepatic

cholangiocarcinomas (15, 16). Extrahepatic cholangiocarcinomas are more likely to have fusions in *PRKACA* and *PRKACB* (14–16), with gallbladder cancers more likely to have mutations in *ELF3* (17).

1.2: The current standard of care for the treatment of biliary tract cancer

Unlike many other major cancer types, improvements in the prognosis of patients with biliary tract cancer has been slow and the prognosis is very poor for the majority of patients. A study looking at outcomes in the UK showed that nearly half of patients diagnosed with cholangiocarcinoma did not receive any form of intervention or systemic treatment for their cancer (18). Patients often present late, with 50% of biliary tract cancer patients presenting with symptoms for the first time via A&E, while overall 19% of all cancers initially present through an A&E admission (18). Only 21% of patients had early stage 1-2 cancer (which is generally potentially resectable and curable). However, even in patients who have early-stage cancer and resectable disease most patients relapse with recurrent or metastatic disease. When patients have fully resected disease (R0 resection), the median overall survival (OS) is 33.4 months, with 43.3% of patients alive 5 years after surgery (19). However, if the resection margins are not clear and microscopic residual disease remains (R1 resection), the outcomes are far poorer, with a median overall survival of 24.7 months and only 13.7% of patients alive 5 years after surgery (19).

1.2.1: Treatment of early-stage biliary tract cancer

For the minority of patients with early-stage and resectable disease at presentation, the current standard of care is surgery followed by adjuvant chemotherapy. While neoadjuvant (chemotherapy before surgery) treatment has been hypothesised to increase the rates of R0 resections and therefore reduce the rate of relapse after

surgery, there have been relatively few studies investigating this and, at present, no phase III clinical trials.

Adjuvant chemotherapy (chemotherapy after surgery) on the other hand is relatively well studied. The standard of care treatment in the UK, Europe and most of the western world is treatment with 6 cycles of adjuvant capecitabine chemotherapy. This is based on the results of the BILCAP clinical trial (20), which compared capecitabine chemotherapy with observation, which was the standard of care treatment at the time. This study is discussed in more detail below, but patients receiving adjuvant chemotherapy had an average of 6.9 months more relapse free survival compared to those on observation (24.4 vs 17.5 months) with an average of 17 months improvement in overall survival (53 months in the chemotherapy arm vs 36 months in the observation arm) (20). A similar level of improvement was seen when using adjuvant S-1 (a combination of tegafur (a prodrug of fluorouracil), gimeracil, and oteracil potassium) instead of gemcitabine and cisplatin chemotherapy vs observation in the adjuvant setting, as seen in the JCOG1202 / ASCOT clinical trial (21), so S-1 is often used for adjuvant treatment in east Asia.

There have been several studies looking into other treatment combinations in the adjuvant setting, but none have demonstrated an improvement against the current standard of care. These include the BCAT clinical trial comparing single agent gemcitabine to observation (22), the PRODIGE-12 study comparing gemcitabine and oxaliplatin versus observation (23, 24), and the STAMP clinical trial comparing adjuvant gemcitabine and cisplatin to capecitabine (25).

1.2.2: Treatment of locally advanced or metastatic biliary tract cancer

As mentioned above, the majority of patients diagnosed with biliary tract cancer are diagnosed with metastatic disease or have disease that is too locally advanced to be fully removed in an operation. These patients require systemic therapy, and the current standard of care in the first line is combination chemotherapy and

immunotherapy with gemcitabine, cisplatin and durvalumab. The results of the TOPAZ-1 clinical trial showed that 24.9% of patients were alive at 24 months after treatment with cisplatin, gemcitabine and durvalumab compared to 10.4% of patients who only received cisplatin and gemcitabine (26). The results of TOPAZ-1 have caused cisplatin, gemcitabine and durvalumab to become the first-line standard of care treatment in the UK. Combination chemo-immunotherapy using pembrolizumab in combination with gemcitabine and cisplatin has also been shown to improve overall survival when used in the KEYNOTE-966 clinical trial (27), but this combination has not yet been approved for use in the UK by NICE. Gemcitabine and cisplatin combination chemotherapy is still sometimes used as the first-line standard of care treatment based on results from the ABC-02 clinical trial, which showed a benefit when using gemcitabine and cisplatin versus single agent gemcitabine (28). Gemcitabine and S-1 might also be used in east Asia, as gemcitabine and S-1 has been shown to be non-inferior to gemcitabine and cisplatin in Japanese patients (29). Single agent gemcitabine is also used with patients who are not fit enough for more intensive therapies (for example, patients with a performance status of 2) or those who have serious co-morbidities but still fit enough for some systemic therapy. Although rare, some biliary tract cancers have high levels of microsatellite instability (MSI high), and as with all MSI high gastrointestinal cancers these patients can be treated with single agent nivolumab.

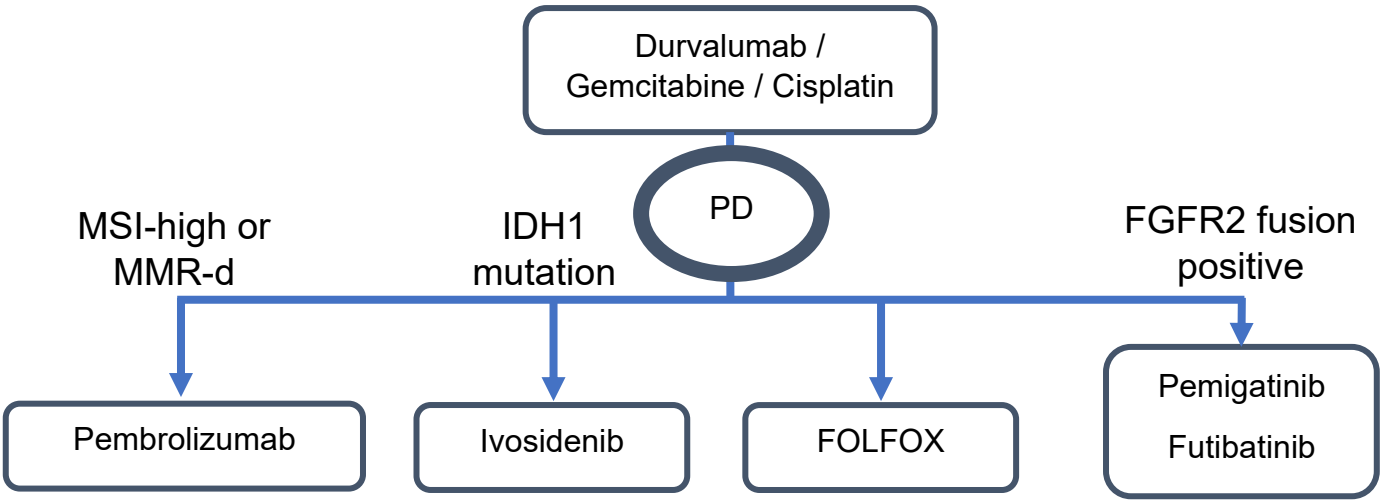


Figure 2: This diagram shows the second-line treatments currently available for locally advanced or metastatic biliary tract cancer in the UK.

The current second-line treatment available for all biliary tract cancer patients is FOLFOX, a combination chemotherapy of leucovorin calcium (folinic acid), fluorouracil, and oxaliplatin. The ABC-06 clinical trial showed a 6-week survival benefit for patients who received FOLFOX rather than best supportive care in patients who had progressed after treatment with first line chemotherapy (6.2 months versus 5.3 months) (30).

However, while treatment with FOLFOX is available to all patients with biliary tract cancer after progression, there are targeted treatment options available for patients with specific genomic mutations and alterations. Fibroblast growth factor 2 (*FGFR2*) fusion mutations are present in around 15% of patients with biliary tract cancer, and patients with this alteration are eligible in the UK to receive treatment with pemigatinib, a drug which specifically targets this alteration (31). Licensing was based on the results of the phase II FIGHT-202 clinical trial, which showed that FGFR2 fusion positive patients had an average progression-free survival of 6.9 months and an average overall survival of 21.1 months when receiving pemigatinib having progressed on previous chemotherapy (32, 33). A second FGFR2 inhibitor called futibatinib is also NICE approved in the UK (34, 35), based on the results of the phase II FOENIX-CCA2 clinical trial, which showed a median progression-free survival of 9.0 months and median overall survival of 21.7 months (36, 37). Futibatinib has a similar side effect profile to pemigatinib, but also has efficacy in

patients who have progressed after treatment with other first line FGFR2 inhibitors, although this is dependent on the mechanism of resistance (2, 38).

Other targeted options are also available, including ivosidenib which targets the R132 mutation in *IDH1*. This treatment is licenced in the UK for patients with an R132 *IDH1* mutation who have progressed on first-line standard of care treatment, based on the results of the phase III ClarIDHy clinical trial (39, 40). This study compared ivosidenib to best supportive care in patients who had received and progressed after first-line chemotherapy. ClarIDHy showed that patients receiving ivosidenib had a median progression-free survival of 2.7 months compared to 1.4 months for patients receiving best supportive care (39), leading to ivosidenib becoming a treatment option for these patients. Patient with tumours that are MSI-high or have deficiencies in mismatch repair genes (MMR-d) are also eligible to receive up to 2 years of treatment with pembrolizumab, a checkpoint inhibitor immunotherapy, based on the results of the KEYNOTE-158 clinical trial (41).

Clinical trials involving other targets have also been carried out in patients with locally advanced or metastatic disease. The Rare Oncology Agnostic Research (ROAR) clinical trial was a phase II basket trial of patients with locally advanced or metastatic tumours with V600E *BRAF* mutations (42). 43 patients with biliary tract cancer were recruited into the ROAR trial, and half of these patients (20 / 43 (47%, 95% CI 31–62)) were independently assessed as having responded to dabrafenib and trametinib *BRAF* targeted treatment (43), so *BRAF* targeted treatment is forming part of the biomarker targeted SAFIR-ABC-10 clinical trial (see section 1.3.1: SAFIR-ABC-10). The SUMMIT basket trial investigated patients with *HER2* mutation positive locally advanced or metastatic tumours (44), and the results in patients with biliary tract cancer showed that neratinib was well tolerated and had moderate anti-tumour activity with an overall response rate of 16% (95% CI 4.5–36.1%) (44). However, the investigators suggested that neratinib may be more effective in combination treatment. The phase II KCSG-HB19–14 clinical trial of patients with *HER2* amplified locally advanced or metastatic biliary tract cancers who had progressed on treatment with gemcitabine and cisplatin showed that a combination of FOLFOX and trastuzumab had a disease control rate of 79.4% (95% CI 62.9–

89.9) and an overall response rate of 29.4% (95% CI 16.7–46.3) (45). These results have led to neratinib and trastuzumab being used as a treatment for *HER2* mutation positive tumours in the SAFIR-ABC-10 clinical trial (see section 1.3.1: SAFIR-ABC-10). *HER2* amplified biliary tract cancers were also investigated in the HERIZON-BTC-01 phase II clinical trial using the bispecific anti-HER2 antibody zanidatamab. 41.3% (95% CI 30.4-52.8) of patients showed an objective overall response, and 18 (21%) patients were still receiving treatment at the 24-month data cutoff (46). Zanidatamab is therefore also being used in the SAFIR-ABC-10 clinical trial in patients with *HER2* amplified tumours (see section 1.3.1: SAFIR-ABC-10).

1.3: Current and upcoming clinical trials in biliary tract cancer

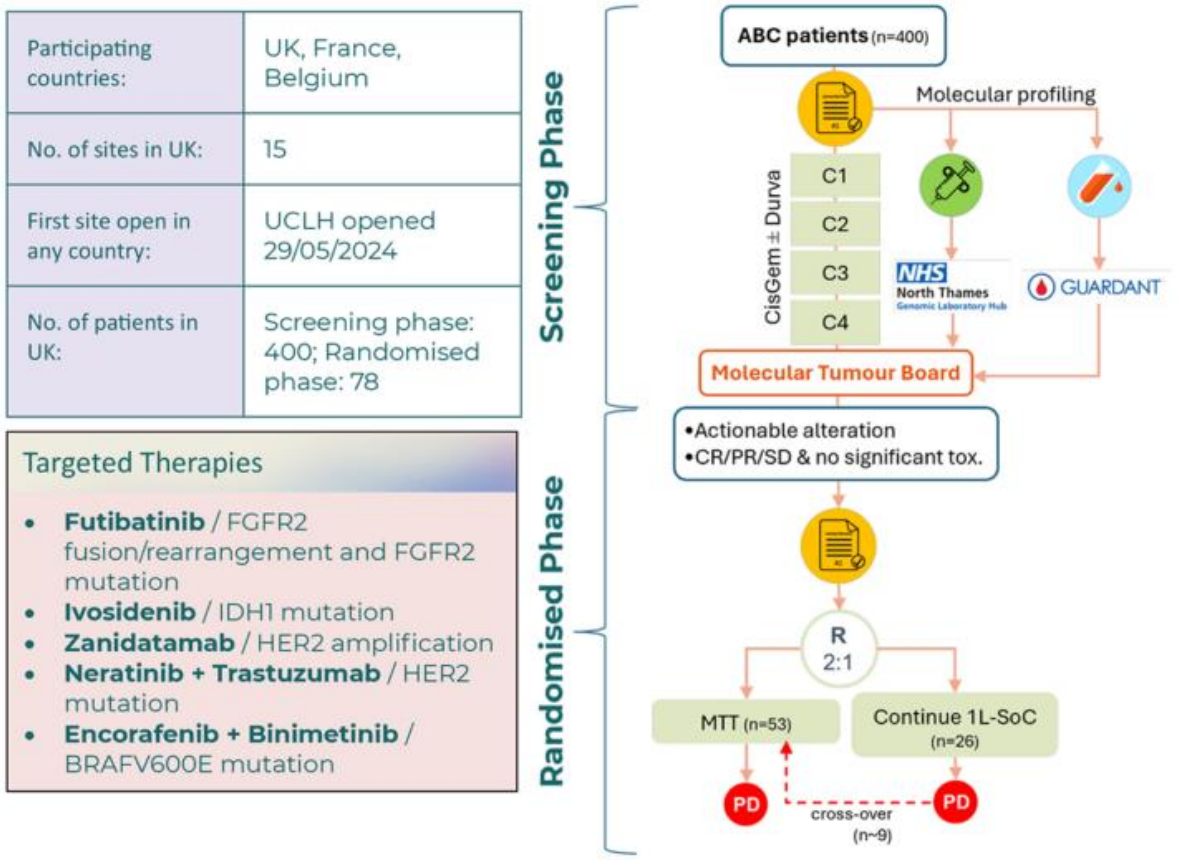
Although the number of treatments available for biliary tract cancer has increased massively in the last few years, there are still several patient groups who lack treatment options. Patients with locally advanced or metastatic biliary tract cancers with genomic alterations which have targeted treatment options currently have to receive first-line systemic therapy with chemo-immunotherapy to receive targeted therapy in the UK. There are several clinical trials now being run and set up to investigate first-line targeted therapy for biliary tract cancer, and the biggest of these is SAFIR-ABC-10.

1.3.1: SAFIR-ABC-10

The SAFIR-ABC-10 clinical trial is designed to investigate the efficacy of targeted anti-cancer treatments for the first line treatment of biliary tract cancers. As discussed earlier, the current first-line standard of care treatment for biliary tract cancer is combination chemo-immunotherapy with durvalumab, cisplatin and gemcitabine, which can be very poorly tolerated and is not suitable for many newly diagnosed patients. More and more targeted treatment options are becoming

available, but nearly all the studies to date which show efficacy have been in the second-line setting. There is a need to investigate the efficacy of targeted treatments in the first-line setting, but as biliary tract cancers are uncommon it is very difficult to recruit the number of patients necessary to individual clinical trials, which has led to the discontinuation of several clinical trials such as PROOF 301 (47, 48) and FIGHT-207 (49).

SAFIR-ABC-10 has therefore been developed as a way to investigate the efficacy of multiple targeted treatment options for the first-line treatment of biliary tract cancer in a single clinical trial (see Figure 3).



Key: CisGem ± Durva – cisplatin, gemcitabine and durvalumab treatment, C1 – cycle 1, CR – complete response, MTT – molecularly targeted treatment, PR – partial response, SD – stable disease, SoC – standard of care.

Figure 3: This diagram shows the design and details of the SAFIR-ABC-10 clinical trial. Diagram courtesy of AMMF (50)

The main set up of the trial means that all of the patients recruited into the trial will have genetic profiling of their tumours to look for potentially targetable options, which include *IDH1* mutations, *HER2* amplification, *HER2* mutations, *FGFR* fusions and *BRAF* V600E mutations. While this is ongoing, all patients will receive 4 cycles of cisplatin and gemcitabine +/- durvalumab. If a patient is shown to have an actionable alteration, the patients are then randomised in a 2:1 manner to receive either a targeted treatment (the exact treatment depending on the alteration found) or 2 further cycles of chemotherapy +/- immunotherapy. Patients who progress on chemo-immunotherapy and have a targetable alteration are allowed to cross over to a targeted treatment. Patients recruited to SAFIR-ABC-10 who do not have a targetable mutation will receive the SEVILLA trial protocol (ivonescimab vs FOLFOX or second-line standard of care).

To detect a hazard ratio (HR) of 0.60, which represents an increase in progression-free survival from 6 to 10 months with 80% power and a two-sided significance level of 0.05 and 2:1 randomisation, 132 progression-free survival events are required. Assuming a 36-month recruitment period with 12-months of further follow up, 151 patients should be enough to gain the required number of progression-free survival events, which increases to 159 to compensate for a 5% drop-out rate (51). In order to obtain the required number of progression-free events, around 800 patients will need to be enrolled and screened on the trial. Based on the ABC-02 (28) and BINGO (52) clinical trials, around 70% of patients had their disease controlled by first-line chemotherapy without adverse events, representing around 560 (70%) of the initial 800 patients. It is technically difficult to obtain molecular profile data from biliary tract cancer tissue as biliary tract cancer tissue can be difficult to obtain, so it is likely that 20% of tumour samples would be unsuitable for analysis, meaning that around 448 (80%) of the 560 patients would have samples suitable for molecular analysis. The MOSCATO clinical trial (53, 54) and other published data have given an indication on the incidence of various molecular alterations in biliary tract cancer, so the authors estimate that around 35% of profiled patients will have a ESCAT (ESMO Scale for Clinical Actionability of molecular Targets) I or II profile that is targetable by one of the treatments in ABC-10, which would lead to 159 eligible patients, the number needed for the statistical analysis.

1.3.2: Other open or recent clinical trials

Other targeted treatments were also being investigated in the first line setting. FIGHT-302 was a phase III clinical trial comparing pemigatinib to gemcitabine and cisplatin in the first-line treatment of patients with locally advanced or metastatic biliary tract cancers with FGFR2 rearrangements (55). FOENIX-CCA3 was similar phase III clinical trial, comparing futibatinib with gemcitabine and cisplatin for the first-line treatment for patients with locally advanced or metastatic biliary tract cancer with FGFR2 fusions (56). Unfortunately, these studies closed early because of recruitment difficulties, low incidence of the targetable alteration, the time taken to obtain a fusion result and the addition of durvalumab to standard of care treatment. Clarification on whether futibatinib should be used at a dose of 16mg or 20mg is also being investigated as part of the FOENIX-CCA4 clinical trial (57).

Other trials are looking into different options for patients with early-stage, resectable cancers. ACTICCA-1 is a phase III clinical trial aiming to compare adjuvant gemcitabine and cisplatin chemotherapy with either capecitabine or observation in patients with resected biliary tract cancer (58); this trial has recently closed, and the results are due to be published soon. Another adjuvant clinical trial is also due to open soon; the ARTEMIDE-Biliary-01 clinical trial is a phase III clinical trial that will compare combination rilvegostomig (a bispecific anti-PD-1 and TIGIT antibody) and chemotherapy with placebo and chemotherapy in patients with resected early-stage biliary tract cancer.

1.4: Translational analysis of genomic data from clinical trials

1.4.1: Introduction and rationale

The main part of this project was the analysis of genomic data from clinical trials and associating genomic changes with patient characteristics. Unlike more common tumour types, such as lung and prostate cancer, there is much less genomic information available from patients with biliary tract cancers. Even in the genomic data from the 100,000 Genomes project carried out by Genomics England (59), there are fewer than 50 samples from patients with biliary tract cancers. While there have been landmark studies looking at the overall genetic landscape of patients with biliary tract cancers, in particular the papers by Nakamura *et al.* (16) and Jusakul *et al.* (15) there, as of yet, hadn't been a study investigating the association between genomic changes and patient outcomes. This was having an impact on the development of new treatments for biliary tract cancers; in the documentation from NICE on the approval of pemigatinib (an FGFR inhibitor), it was noted that the improvement in survival in patients given pemigatinib might not solely be due to the new treatment, as it was possible that patients with biliary tract cancers who had FGFR2 fusions had better outcomes than those who did not (31).

The overall plan for this research was therefore to collect a dataset of patients with biliary tract cancer for whom overall outcomes were available, and find any possible associations between genomic alterations and both relapse-free (PFS) and overall survival (OS). Ideally this would have to be a retrospective dataset, as patients with actionable mutations diagnosed more recently would have better outcomes than those without, as they would have access to targeted treatments that would improve their outcomes compared to those without targeted options. The dataset that best fits this is from the BILCAP clinical trial; a phase III clinical trial comparing adjuvant capecitabine to best supportive care in patients with early-stage biliary tract cancer, as seen on the CONSORT diagram in Figure 4 (20).

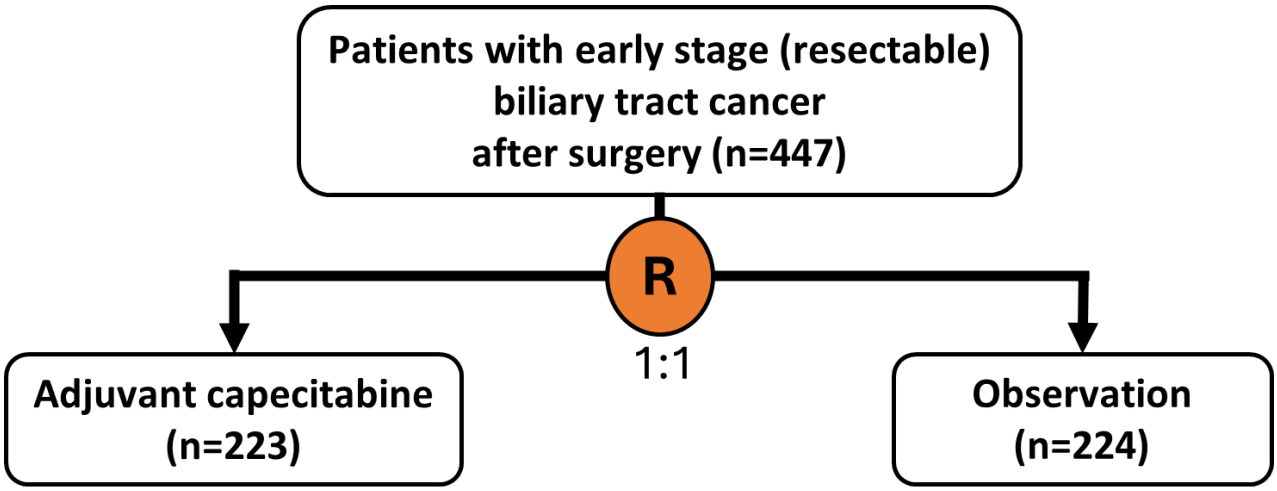


Figure 4: This CONSORT diagram shows the overall structure of the phase III BILCAP clinical trial.

This trial closed before the first targeted treatments became available, so patients who had recurrent or metastatic disease would not have had access to these treatments. Recruitment for the trial closed in 2014, so patients who would have relapsed would have relapsed by now, and enough time has passed that the full overall survival is known for these patients. Also, the fact that this is an adjuvant study means that it is more likely that sufficient tissue for genomic analysis would be available from the surgical resection samples.

The plan was that tissue from patients on the BILCAP clinical trial would be taken and DNA and RNA extracted, as per Figure 5.

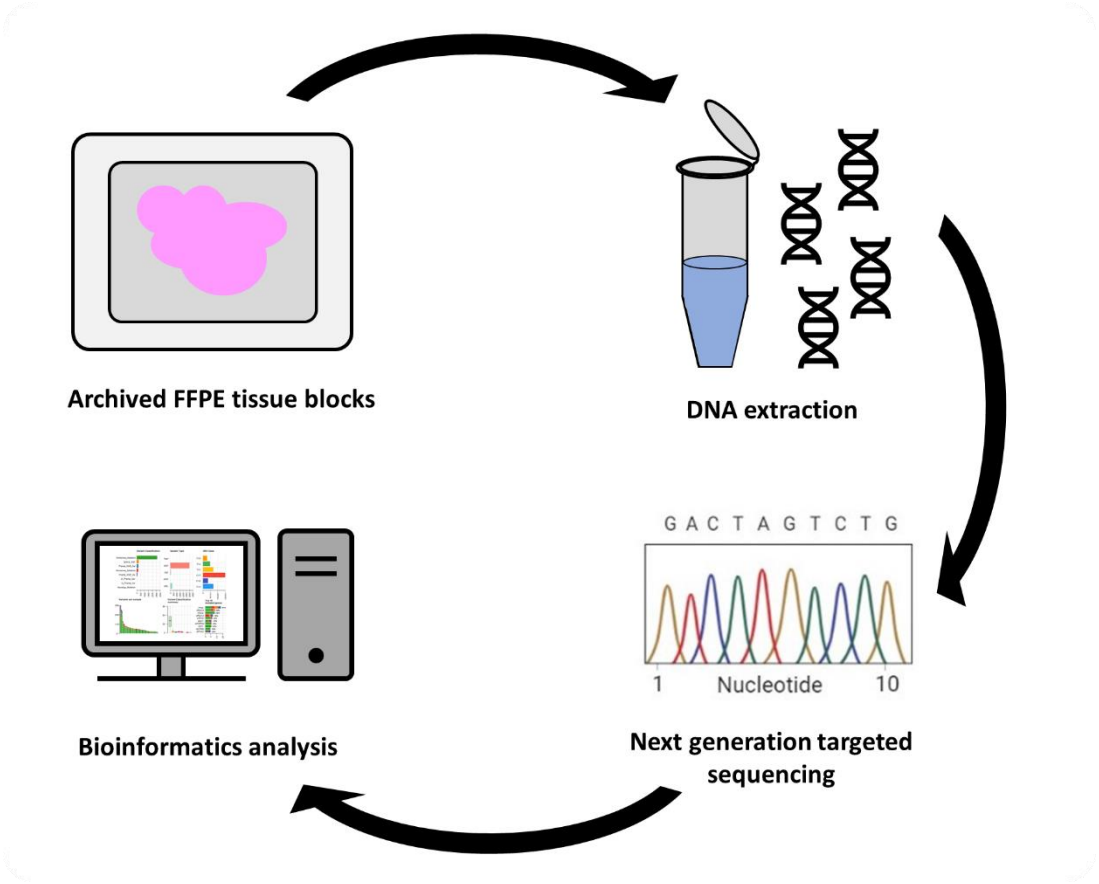


Figure 5: This diagram shows how the tissue samples from patients in the BILCAP clinical trial was processed and analysed.

Due to budget concerns and issues with sequencing older FFPE samples, not every type of genomic analysis could be performed; for example, gene expression arrays are very unreliable in this type of sample. However, targeted gene sequencing (for mutation analysis), low-pass whole genome sequencing (lp-WGS) (for copy number analysis) and RNA-seq (for fusion analysis and tumour microenvironment analysis) could all be performed on the samples. Together with the information on patient outcomes gathered from the clinical trial data, any possible associations between genomic alterations and patient outcomes could be found from this data.

However, the BILCAP data would not be the only dataset analysed during the course of this project. While the DNA and RNA was being extracted from the BILCAP samples, bioinformatics pipelines for DNA and RNA analysis would need to be created; previously analysed datasets would need to be gathered and re-analysed to create the pipelines required, so that the bioinformatics analysis could be completed

more easily as the raw BILCAP data became available. It would also allow for validation of the results from the BILCAP analysis; if the proportions of mutations and other alterations seen in the BILCAP data was massively different to that seen in the analysis of other biliary tract cancer datasets, it would show that there were likely to be issues with the bioinformatics pipeline being developed that would need to be addressed.

1.4.2: Plans for data analysis and dataset selection

The first step was to identify the datasets to be used in developing the bioinformatics pipelines for this project. The first and most obvious dataset to be used was the cholangiocarcinoma dataset from the paper by Jusakul *et al.* (15). This dataset is archived on the European Genome-Phenome Archive (EGA) under study number EGAS00001001653 (60). This dataset consists of three datasets from the original paper; EGAD00001003834, which consists of WGS FASTQ data from 12 patients with cholangiocarcinoma, EGAD00001001988, which consists of WGS FASTQ data from 59 patients with cholangiocarcinoma (and includes samples from Nakamura *et al.*'s 2015 paper (16)) and EGAD00001001994, which consists of 376 targeted sequencing data FASTQ files from 188 patients. Clinical data on was also available from the supplementary data of the paper for the patients whose tissue was included in this study. This included information on age, gender, cholangiocarcinoma subtype (extrahepatic, intrahepatic or perihilar), histology, TNM staging at diagnosis, length of overall survival, whether the patient had died as well as information on possible tumour aetiologies (whether or not it was a liver fluke related cancer and whether the patient had primary sclerosing cholangitis (PSC), hepatitis B or hepatitis C) (15). This data included matched tumour and normal FASTQ files, taken from each patient's tumour when it was resected. The data is generally of very high depth and quality, so it was a good dataset to start working on to build the analysis pipelines.

However, the Jusakul dataset does not fully cover the tumour types seen in the BILCAP dataset. It does not include any patients with primary gallbladder cancer and does not include any patients with metastatic biliary tract cancer at presentation

(which is also not included in the BILCAP dataset, but would be interesting to use as a point of comparison). Archived data from patients with gallbladder cancer was harder to find; while many papers have been published discussing the genomics of biliary tract cancers, older papers often do not have the raw sequencing data easily accessible. Changes in the rules of most publishers mean that newer papers will nearly all have their data archived, although gaining access can be difficult when it relies on getting permission from a single person whose email may or may not up to date. Even if the sequencing data is available, comprehensive clinical data may not be available. Nearly all datasets will include data on patients age and gender, but information such as tumour subtype, histology, staging and survival is much harder to come by.

Many papers were investigated, but most were excluded as they either had relatively few patients (<100) or had limited clinical data available (14, 61–74). Papers from Zou *et al.* 2014 (75) and Jiao *et al.* 2013 (76) both contained data from patients with intrahepatic cholangiocarcinomas, but access to the raw data was difficult to obtain and there was plenty of data available from patients with intrahepatic cholangiocarcinomas from the Jusakul *et al.* dataset. One large dataset of patients with gallbladder cancer was available from Li *et al.* 2014 (77), but it was not possible to get access to this archived dataset as the person labelled as being the contact for accessing the data did not respond to attempted communication.

Fortunately, a paper had recently been published which contained a large enough number of patients with gallbladder cancer. Pandey *et al.* studied a large cohort of patients will gallbladder cancer and high-risk non-cancerous gallbladder conditions (such as gallstones and gallbladder polyps) in their paper from 2020, and the data was archived on the EGA under study number EGAS00001003004 (17, 78). The archived data included three datasets; EGAD00001004853, which included 392 WES FASTQ samples, EGAD00001004854, which included 120 RNA-seq FASTQ samples and EGAD00001004855, which included 361 WGS FASTQ samples. In this project, the WES and WGS samples were used, which included matched tumour and normal samples. Clinical data was also available for these patients, and included age, gender, histology, and TNM staging (17). Unfortunately, although the paper

mentions overall survival and progression-free survival, this data is not available in the clinical data provided in the supplementary data of the paper. Attempts made to contact the authors to see if this information could be made available were unsuccessful. Fortunately, access to this data was not via the authors but was through the sponsors of the work, Genentech (California, USA), which is a subsidiary of Roche. Gaining access to this data took a significant amount of time, but as this dataset is large and contains high quality data it was worthwhile to use this data to help create the pipelines required.

1.4.3: Analysis of patients with metastatic cholangiocarcinoma and associations with patient outcome (I3O-MC-JSBF dataset)

The BILCAP clinical trial recruited patients with early-stage, fully resectable biliary tract cancers, and the data from the Jusakul *et al.* paper was likewise nearly all from patients with early-stage biliary tract cancers (15, 20). Data from patients who had locally advanced or metastatic disease at presentation is much rarer and more difficult to obtain. A large part of this is practical difficulty; although most patients with biliary tract cancers are diagnosed at a locally advanced or metastatic stage, nearly half of all patients do not undergo any form of cancer directed treatment (18, 79). Obtaining tissue samples even for the diagnosis of biliary tract cancer can be difficult, with cytology brushings from endoscopic retrograde cholangio-pancreatography (ERCP) being a common source of tissue for diagnosis. However, this generally provides small amounts of localised cells, which may be enough for a cancer diagnosis but is often not adequate for analysis of potential targetable genomic changes such as *FGFR2* fusions and *IDH1* mutations. Having any tissue left over from those investigations for storage in a tissue bank for further genomic investigations is near unheard of, and most clinical trials of patients with locally advanced or metastatic biliary tract cancer do not collect tissue beyond diagnosis, or store tissue for sequencing in the longer term.

However, one clinical trial, I3O-MC-JSBF, was slightly different. This study, as seen in the modified CONSORT diagram seen in Figure 6, recruited patients with locally advanced or metastatic biliary tract cancers who had not had any previous treatment, and assigned them to receive either intravenous (Arm A) or oral (Arm B) treatment. In each arm, patients were randomised 2:1 to receive either gemcitabine + cisplatin + ramucirumab or gemcitabine + cisplatin + placebo (Arm A) or gemcitabine + cisplatin + merestinib or gemcitabine + cisplatin + placebo (Arm B). No crossover was allowed, and the trial recruited around 300 patients in total.

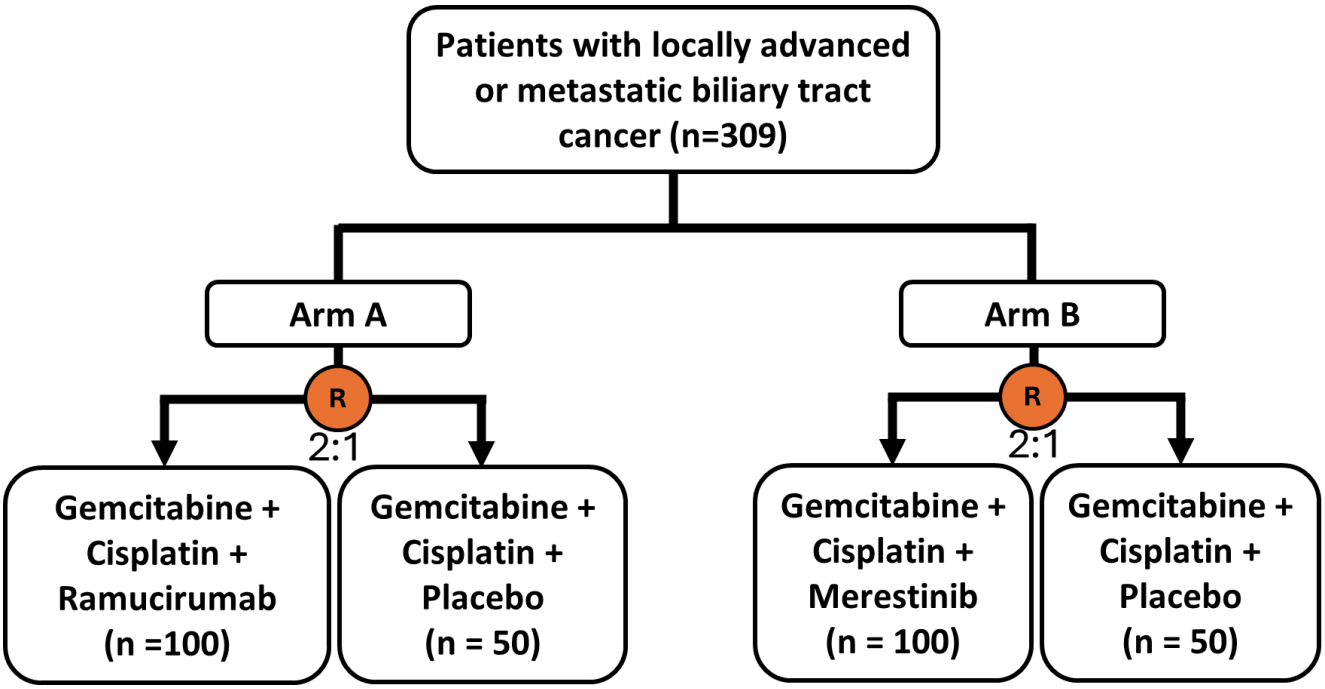


Figure 6: This simplified CONSORT diagram shows the overall design of the phase II I3O-MC-JSBF clinical trial.

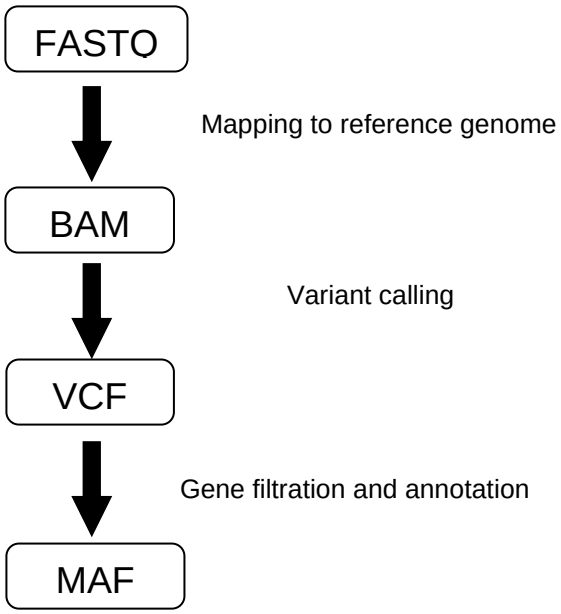
Unfortunately, the trial was negative, as the addition of ramicirumab or merestinib did not improve either progression-free or overall survival when compared to gemcitabine + cisplatin alone (80, 81). However, as part of enrolment onto the trial all patients had a fresh tissue biopsy taken from the primary tumour. These tissue samples were then processed so that high quality bulk RNA-seq data was available from 129 patients.

As of yet, no one has directly compared the genomic landscape of patients with early-stage tumours and those from tumours which were locally advanced or metastatic at presentation. Although the differences in DNA and RNA sequencing makes direct comparisons slightly less accurate than if it had been DNA compared to DNA or RNA compared to RNA, comparing the data from patients from BILCAP with patients enrolled onto I3O-MC-JSBF would start to show if there were any major differences between patients with tumours at different stages of development.

Chapter 2: Methods

2.1: Pipeline development and re-analysis of a previously published dataset of patients with early-stage cholangiocarcinoma (Jusakul *et al.* dataset)

The datasets obtained were all FASTQ files, of either matched WGS, WES, RNA-seq or targeted sequencing DNA. This data would need to be extensively processed before it could be analysed, and before any associations could be made with clinical characteristics. In order to process the data, the FASTQ files would need to be aligned to a standard genome, variants would need to be called to find tumour specific mutations and these mutations would then need to be annotated to determine which genes the mutations were in, and whether or not these would have any functional consequences.



Key: FASTQ - Text file containing genomic sequence data, BAM – binary alignment map, VCF – variant call format, MAF – mutation annotation format.

Figure 7: This diagram shows the stages in analysing genomic data from raw FASTQ files to annotated MAF files for use in data visualisation.

.

2.2: Alignment of FASTQ files to produce BAM files

The first step in analysing the raw genomic data was to align the raw FASTQ files to a reference genome to produce BAM files. FASTQ files are text files of the sequenced DNA fragments produced during library preparation, and also contain quality scores on how likely it is that each base called is true, and not an error. In order to build these fragments into a complete genome, they need to be aligned together. This can be done to create a *de novo* genome, but it is much faster and more accurate if the fragments are aligned using the help of a reference genome for the organism in question. As the human genome has a well curated reference genome (current build hg38) (82), the FASTQ files for this project were aligned with the human reference genome.

There are several bioinformatics tools which can be used to align FASTQ files to form an aligned whole genome file, which is most commonly stored in binary format for space saving reasons and is known as a BAM (binary alignment map). This file type contains all the aligned FASTQ sequences, information on their position in the aligned genome and the quality information score from the FASTQ files. BAM files are often the files used for many downstream analyses, including variant calling (for mutation analysis) and copy number analysis. One of the most widely used tools is called BWA-mem, which is one of the algorithms from the Burrows-Wheeler Aligner and is particularly good at aligning sequences with low divergence to a large reference genome, which is the situation when aligning FASTQ files to the human genome (83), and so this tool was used for alignment in this project.

However, aligning FASTQ files to a reference is not the only step that needs to happen to produce the most accurate and useful BAM files possible. The reads from the FASTQ file need to be read so that low quality reads can be appropriately discarded, the FASTQ reads need to have any adaptor sequences from the library preparation step removed (trimming) and the final BAM files need to have duplicate reads flagged so that they can be removed or ignored as needed by downstream analysis tools. There are bioinformatics tools that can perform each of these steps, but it can be tedious to run each step separately, and doing so is more prone to

introducing errors, as well as making each step harder to reproduce in the future. Fortunately, there is a pipeline available which can perform all these steps in a self-contained and reproducible manner.

nf-core/sarek is a workflow designed to detect variants from whole genome or targeted sequencing data (84, 85). It contains multiple tools, including BWA-mem for alignment and other tools for quality control, trimming and marking duplicates. This tool is built using Nextflow, which allows for different tools to be combined into a single pipeline where the tools will run one after the other with only a single set of input data and commands (86). Nextflow based pipelines are extremely useful when running on a high-performance computing (HPC) environment, as they can be configured to run quickly and resource efficiently on specific HPCs. Lucia Conde, part of the development team for nf-core/sarek works at the Bill Lyons Informatics Centre (BLIC) at UCL, and so has ensured that nf-core/sarek is configured to run on UCL's HPC, called Myriad. This part of the project therefore used nf-core/sarek (version 2.6.1 for this part of the project) to align the FASTQ files to form BAM files for downstream analysis. Due to the size of the input FASTQ files (easily reaching 17 GB for a set of paired FASTQ files) and the number of samples, it took well over a week just to align each sample in the Jusakul datasets (although samples were run in parallel in batches), but each were successfully aligned to form a BAM file for each matched tumour and normal sample.

```
#!/bin/bash -l

#$ -l h_rt=48:00:0

#$ -l mem=12G

#$ -N EGAD00001003834_mapping

#$ -wd /home/regmvcr/Scratch/workspace/EGAD00001003834

#$ -pe smp 12

#$ -t 1-12


number=$SGE_TASK_ID

paramfile=/home/regmvcr/Scratch/jobscripts/EGAD00001003834/EGAD00001003834_sample
_names.txt

SAMPLE="` sed -n ${number}p $paramfile | awk '{print $1}'` "


#nfcore/sarek script for patient "$SAMPLE"


shopt -s expand_aliases


#Make directory (if not already made)


mkdir /home/regmvcr/Scratch/workspace/EGAD00001003834/"$SAMPLE"

cd /home/regmvcr/Scratch/workspace/EGAD00001003834/"$SAMPLE"


#Load dependencies on Myriad


module load blic-modules

module load nfcore_sarek/2.6.1


export
TOWER_ACCESS_TOKEN=eyJ0aWQiOiAzNTExfS4yYzJkNTBiZDA4NmFmYjQxOWNiZDkwZjNiYz
U0YzJjYzA2MDZjYWY2

export NXF_VER=20.10.0
```

Figure 8: This is an example of one of the Unix bash scripts used to align the raw FATQ files into BAM for the Jusakul et al dataset.

2.3: Variant calling, annotation and visualisation of called mutations

Once the FASTQ files had been aligned to form BAM files, the next step was to find any tumour specific mutations in each sample. As this dataset contains samples that have a matched tumour and normal tissue for each patient, this increases the accuracy of calling tumour specific variants as identified variants can be compared against both the reference genome as well as the patient's normal genome so germline variants can also be excluded.

There are several variant calling bioinformatics tools available, including Mutect2 (87, 88), Manta (89), Strelka (90–92) and Varscan2 (93–95). Mutect2, Manta and Strelka are all included in nf-core/sarek, so the project started with using these tools as part of the pipeline. After initial experimentation with the variant calling part of the nf-core/sarek pipeline, it appeared that the most reliable variant callers were Manta and Strelka2, with the results of the Manta variant caller fed into Strelka as per GATK best practices (92).

Once variants are called, the next step is to annotate the called variants to show which gene is affected by each variant, and what affect that would have on the encoded protein. Again, there are multiple tools available as part of the nf-core/sarek pipeline, including VEP (96) and SNFeff (97). VEP was chosen over SNPeff at this point due to improved reliability in nf-core/sarek and ease of use with tools used for other steps.

However, while the nf-core/sarek pipeline can output annotated variant calling files (that is, a list of somatic mutations present in a tumour genome but not in a matched normal sample, labelled with the gene where each mutation lies) this is a text based format, and so it is not easy to visualise the results obtained. A further tool would need to be used, and after a search the easiest to use and gain the most informative output was determined to be a tool called Maftools (98). This tool uses the R programming language to take analysed genomic data, and produce outputs such as oncoplots which can clearly depict which mutations are present, how common each

mutation is in a group of patients and which genes these mutations are in. This type of plot can also integrate information on copy number variation of different genes, and clinical information on the patients whose samples are included in the plot. However, Maftools requires a specific file type, called a MAF, as its input and tools are needed to convert the VCF output from the nf-core/sarek pipeline to the MAF output required. A search showed that there were two main tools which could be used: either Funcotator (99) or vcf2maf (100). In the end, the decision was made to use vcf2maf, as it was easier to use with VEP and easier to set up and run on Myriad. Funcotator runs better with the GATK pipeline using Mutect2, which wasn't being used at this point.

2.4: Initial filtration of called variants

However, the initial oncoplots from the re-analysis of the 12-sample Jusakul *et al.* dataset had significant issues, as seen in Figure 9. Far too many variants were being called for each sample, indicating that that there has likely been an issue with filtering out mutations that are likely to be due to artifacts in the sequencing and analysing process rather than being actual alterations present in the sample. These artefactual alterations and mutations need to be removed and filtered from the results, so that only mutations and alterations likely to be relevant to biliary tract cancer and true mutations rather than artefacts are included. This process is known as variant filtration, and steps were taken to filter out the most obvious artefacts.

The first method to filter out these artefacts was to use bcftools and VEP to filter the results of the nf-core/sarek pipeline to remove some of the artefacts before the output is converted to MAF using vcf2maf. BCFtools (derived from the widely used SAMtools suite of tools for bioinformatics and genomic analysis (101)) was used to filter out variants detected by the variant caller but that don't appear to be 'true' variants, and for example are more likely due to artifacts when sequencing. Only variants which passed internal quality control checks throughout the variant calling pipeline are marked as 'PASS' in the VCF information, and BCFtools was used to filter out the called mutations lacking this label. VEP was then used (using the

filter_vep) command to filter out any variants not labelled as ‘SOMATIC’, so that only the mutations in the COSMIC (Catalogue Of Somatic Mutations In Cancer) list of somatic mutations were included in the final filtered VCF files. Importing data into MAFtools using the read_maf command also has an extra filtering step, as it removes all variants which do not produce a non-synonymous mutation. This step can also be configured to remove FLAGS, genes which are frequently found to be mutated in exome sequencing but have no known pathogenic impact (102), and the top 20 of these genes can be removed when using the read_maf tool.

2.5: New variant calling pipeline and improved variant filtration

The results from the initial pipeline, as seen in Figure 11 and Figure 12, shows that expected variants were being called, but that extra artefactual variants weren’t being fully filtered out correctly despite the filtration steps being added to the pipeline. It was at this point that a shortcoming of the nf-core/sarek pipeline became apparent. While using a Nextflow pipeline such as nf-core/sarek has the benefit of being more straightforward to run and more easily reproducible than running each step individually, as discussed earlier, the main drawback to this is that the pipeline is relatively inflexible. In order to make a fully functioning pipeline for this project, extra variant filtration steps are clearly needed. While the nf-core/sarek pipeline does include some of these steps, it is relatively inflexible to control the details of these steps and, for example, add extra filtering steps in the middle of the pipeline as required. It was therefore apparent that instead of using the nf-core/sarek pipeline from beginning to end, the results of the alignment step would need to be taken and run through a separate variant calling step, one which allowed for greater control over variant filtration.

After another investigation of the variant callers available, the variant caller decided upon was Mutect2 (87). This was available as part of the nf-core/sarek pipeline but getting it to function properly as part of that pipeline was incredibly difficult and would not have allowed for the separate filtration steps required. Mutect2 was used for several reasons: firstly, as part of the GATK best practices pathway it is both well-

known and updated frequently, so any issues that arose when using it are more likely to have a known solution or to be promptly fixed. Secondly, it also fits as part of a larger set of tools, so there are separate filtration steps available, including FilterMutectCalls and its associated setup, which can be configured separately and specifically to whichever dataset is being worked on. Finally, Mutect2 as part of GATK works with Funcotator (99) which allows for relatively straightforward annotation and conversion of VCF to MAF output using a single tool. While nf-core/sarek would still be used for alignment (which requires relatively little adjustment to run properly), using Mutect2 and Funcotator separately would allow for more flexibility in variant calling and filtration, and is the way that many bioinformaticians use nf-core/sarek in practice.

2.6: Mutation re-analysis of the WGS and WES data from Pandey *et al.*

Work on the Jusakul *et al.* dataset meant that a pipeline had been created for the analysis of mutations from whole genome or targeted sequencing data. Jusakul *et al.*'s dataset only included patients with cholangiocarcinoma, so the Pandey *et al*. dataset was used to include patients with gallbladder cancer as part of the background dataset (17). The Pandey *et al.* dataset (EGAS00001003004) as mentioned earlier consists of three datasets, WGS (EGAD00001004855), WES (EGAD00001004853) and RNA-seq (EGAD00001004854), of which the WGS and WES were used in this project (17). Patients who were part of this dataset but who did not have gallbladder adenocarcinoma were excluded from further analysis, so a total of 143 WGS and 150 WES samples were analysed.

The pipeline developed, using nf-core/sarek for alignment with Mutect2, Funcotator and Maftools used for variant calling, annotation and visualisation respectively, was successfully run on the low-pass WGS and WES samples, as discussed in section 3.1.2: Re-analysis of a previously published dataset of patients with gallbladder cancer (Pandey *et al.* dataset).

2.7: Copy number analysis pipeline development

Once the pipeline for mutation analysis had been created, the next step was to start looking at copy number analysis. This would require the use of different tools, as the pipeline used so far does not work for copy number analysis.

As with mutation analysis, there are multiple tools available for copy number analysis. The two main tools investigated as part of this project were ASCAT (103) and control-FREEC (104), both of which are available as part of the nf-core/sarek pipeline. ASCAT is very well known and widely used, but the version used in nf-core/sarek only works with paired tumour/normal whole genome sequencing, while control-FREEC as part of nf-core/sarek can also work with whole exome or targeted sequencing data.

Starting with using ASCAT as part of nf-core/sarek, the tool did not initially run on the lp-WGS data. Eventually, after updating the nf-core/sarek pipeline to a newer slightly modified version with the help of Lucia Conda (she exposed the –ascat_penalty parameter so that it could be edited as part of the nf-core/sarek pipeline in version 2.6.1, the version of nf-core/sarek being used at this time) results from ASCAT started to become available, as seen in Figure 17.

However, while the tool ran as expected, the output of the tool was often unable to provide a result for the ploidy of the samples run using the pipeline. Attempting to modify the input parameters of nf-core/sarek ASCAT further did not improve the number of samples being run successfully using the pipeline. The next step was therefore to try a different tool for copy number analysis; control-FREEC, the other copy number analysis tool that was part of nf-core/sarek.

However, running control-FREEC through nf-core/sarek was very difficult. Nextflow pipelines, including nf-core/sarek can be very processing heavy and require large amounts of storage space for temporary, intermediate and reference files used while the pipeline is running. control-FREEC was even more difficult to run than normal;

running a single ~2GB lp-WGS sample took over 3 weeks of processing time and used up nearly 25TB of space on the HPC. It is not clear why it took so long and took so much space, as limited optimisation can only explain so much, but it was clear that this would take far too much time and too much space to be used for all the other Jusakul *et al.,* Pandey *et al.* and BILCAP data that this tool would be used on.

Fortunately, around this time the ASCAT algorithm was updated. ASCAT would originally only run on matched tumour-normal whole genome samples and would not produce sensible results from whole exome or targeted sequencing samples (103). However, ASCAT version 3.0 was updated to allow for copy number analysis of whole exome and whole genome matched tumour-normal data (105). Attempts to run the newer version of ASCT through an updated version of nf-core/sarek were again unsuccessful, so instead the files were run using the ASCAT R package (version 3.1.1) on the HPC. This was successful once the correct GC correction files were made available by members of the Van-Loo laboratory.

As well as graphical data, ASCAT also produces information on the changes in copy number of specific segments of the genome, which can be used in downstream analyses to determine the increase or decrease in the copy number of specific genes. A tool called bedtools intersect (106) was used to annotate the copy number data produced using ASCAT to find whether segments containing specific genes were amplified or deleted, and the information on the copy number status of individual genes was used in the downstream analysis and correlation with patient outcomes.

2.8: Analysis of patients with early-stage biliary tract cancer and associations with clinical characteristics and patient outcomes (BILCAP dataset)

Once the raw sequencing data from the analysis of tissue from the BILCAP clinical trial became available, the next stage of the project was to analyse it, using the

pipelines developed by working on the Jusakul *et al.* and Pandey *et al.* datasets. The tissue from consented patients enrolled in BILCAP was stored at the University of Southampton and was extracted and sequenced by the team at the University of Birmingham.

However, when the tissue was analysed, it became clear that the data available from the patients in BILCAP would be very different to that from the Jusakul *et al.* and Pandey *et al.* datasets. BILCAP recruited patients from 2006 to 2014, so the oldest samples were well over ten years old before the DNA and RNA could be extracted and processed. The samples were all also FFPE samples, which can lead to significant DNA and RNA degradation. This means that instead of matched WGS or WES data, the BILCAP data would be unpaired RNA-seq, lp-WGS and targeted gene sequenced data. While some of the pipelines already developed could be used to analyse the BILCAP data, they would have to be significantly modified to be used in the BILCAP data.

The targeted sequencing data was analysed through the proprietary Qiagen pipeline, using Mutect2. However, using the previously developed ASCAT pipeline was not going to work using this data, as ASCAT requires matched tumour and normal data, even in the new version which can run on whole exome or targeted sequencing data rather than whole genome data (105). This meant that a different tool had to be used for copy number analysis, and the tool which ended up being used is called ichorCNA (107), which specialises in DNA which is low quality or highly fragmented, such as in the older tissue samples used in the BILCAP data analysis. The output from ichorCNA was mainly text based and did not show which genes were in the regions shown to be amplified. The data on which segments of the genome had copy number alterations in each sample were therefore combined with data on which genes are at which co-ordinates in the whole genome using a tool called bedtools (106), specifically the bedtools intersect tool, which could calculate which genes were in each copy number altered segment.

As bulk RNA-seq data was available from the patients in the BILCAP cohort, another factor that could be investigated was the tumour microenvironment, and the types and quantities of different immune system cells present in each tumour. Initially this was due to be carried out using a tool called CIBERSORTx (108) in R, but difficulties in gaining access to the reference files meant that gaining useful output from that tool was impossible. Instead, after investigating several other commonly used tools, the tool used was quanTIseq (109) (in particular the R package based on this tool called quantiseqr), which was able to deconvolute the RNA-seq data for each patient and identify the proportion of different immune system cells present in each patient's sample. Figure 21 shows the results of running quantiseqr on the first 45 patient cohort, and using R it was straightforward to identify which samples have numbers of immune system cells at different cut-off points, for example having a fraction of CD-4 T-cells > 0.05.

Data on gene fusions, which is highly important for known pathogenic genes in biliary tract cancers such as FGFR2, could also be calculated using the bulk RNA-seq data available. This was initially calculated using a tool called FusionCatcher, which is designed to find somatic fusion genes in RNA-seq data (110). However, FusionCatcher can be relatively insensitive and miss fusions, especially fusions with unusual partners. Instead, a different tool called Kallisto was used to determine gene fusions (111). Kallisto has the opposite issue, and in order to ensure that all fusions were picked up the tool was set to be very sensitive to possible gene fusions, and therefore more were picked up using this tool in this workflow than might otherwise be picked up using other tools.

The data on which genes were altered in each patient, whether via changes in copy number or mutation as well as information on the tumour microenvironment, was then collated to gain an overall picture of which genes were altered. Data from the clinical trial on individual patients' relapse-free and overall survival was then used to produce Kaplan Meier curves (using the survival R package) to see if any specific alteration had a significant effect on survival. Forest plots were calculated using a database of genomic alterations and tumour microenvironment changes separated by patient, using the survival, survminer and forestmodel R packages (112–115).

2.7: Analysis of the tumour microenvironment in the BILCAP and I30-MC-JSBF clinical trials

Comparisons between early-stage and locally advanced or metastatic biliary tract cancers started with an investigation into the tumour microenvironment. As bulk RNA-seq data was available for both patients from BILCAP (as seen in section 3.2.2: BILCAP2 and analysis of the first patients from the BILCAP clinical trial and patients from I3O-MC-JSBF), a direct comparison between the two would be more straightforward. Raw RNA-seq data from the BILCAP and I3O-MC-JSBF patients were first aligned and quantified using STAR-SALMON, using nf-core RNA-seq (116). The output from this was then analysed using QuanTIseqR (109), which deconvoluted the RNA-seq data and allowed for the quantification of the number of different types of immune system cells present in each tumour sample.

Chapter 3: Results

3.1: Pipeline development and re-analysis of the Jusakul *et al.* and Pandey *et al.* datasets

3.1.1: Re-analysis of a previously published dataset of patients with cholangiocarcinoma (Jusakul *et al.* dataset) and first filtration steps

The first results of the re-analysis of the Jusakul *et al.* dataset started to become available as the pipeline was being set up. The pipeline was initially set up to run on the 12-sample whole genome subset of the data from Jusakul *et al.* (EGA dataset EGAD00001003834); using a smaller dataset initially would show if the pipeline would run to completion without issue and produce sensible looking output without taking too much processing time.

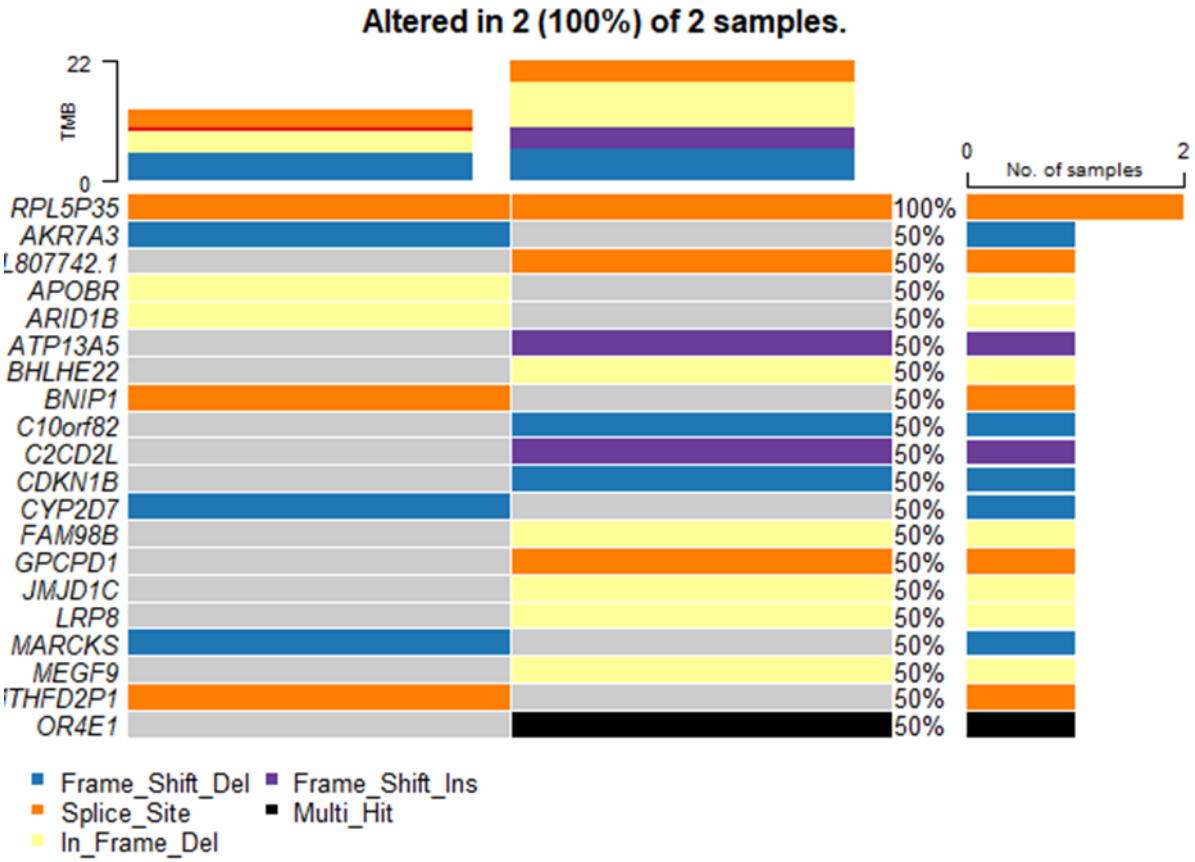


Figure 9: This figure shows the first oncoplot produced using data from 2 samples from the EGAD00001003834 dataset, CCA_RO_1 and CCA_SG_17 (the two samples from the dataset chosen were chosen at random).

Figure 9 shows the first oncoplot of the re-analysis produced using the nf-core/sarek pipeline, using the alignment, variant calling and annotation steps from the pipeline before visualising the data using maftools. At this point it appeared that the pipeline was working as expected, with different types of mutations being called from each of the samples analysed. The next step therefore was to run the pipeline on the entire 12 sample subset, the oncoplot for which is shown in Figure 10.

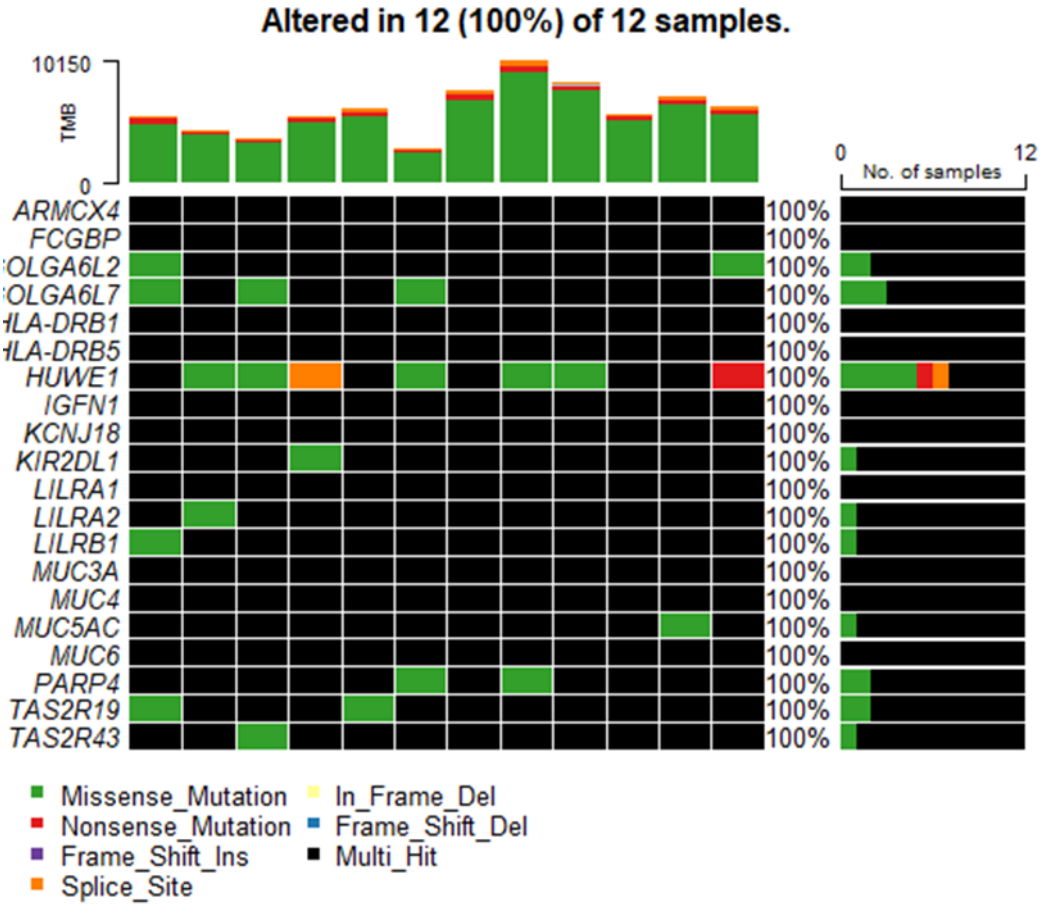


Figure 10: This is the first oncoplot of the 12-sample subset of whole genome sequencing data from the Jusakul et al. *dataset. This oncoplot shows the top 20 most mutated genes among the 12 samples.*

Figure 10 shows the first oncoplot made for the 12-sample subset of the Jusakul *et al.* data and shows a wide variety of mutated genes in each sample. However, looking more closely at the results, it is clear that the output of the pipeline is not quite what would be expected. Many of the labelled mutated genes, including MUC4

and ARMCX4 are not particularly relevant in cancer; while these genes are often mutated, they are rarely drivers of carcinogenesis and are far more likely to be passenger mutations. Also, another issue is that nearly all the genes are labelled as muti-hit mutations; that is, that there is more than one mutation in each gene. While this is not uncommon, the sheer number of multi-hit mutated genes in this sample is too high to be truly plausible, especially in the setting of cholangiocarcinoma.

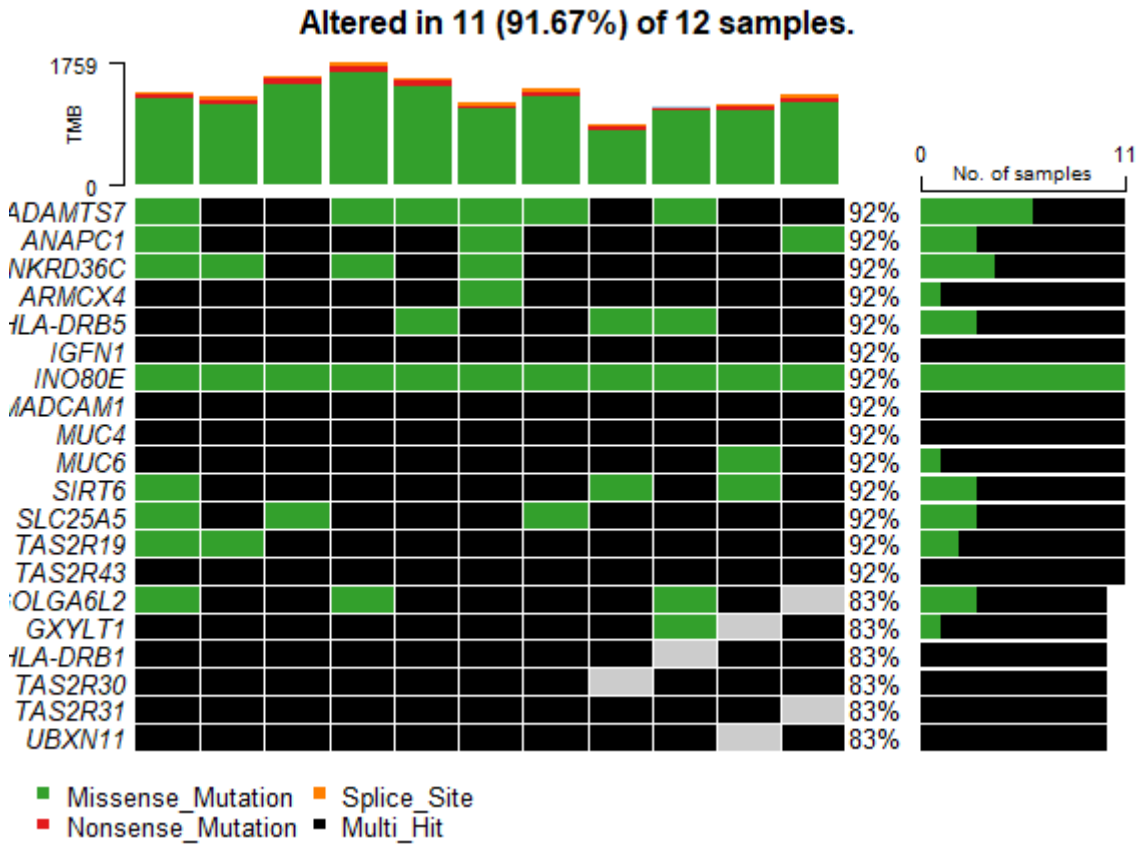


Figure 11: An oncoplot of the 20 most mutated genes in the 12-sample whole genome sequence subset from Jusakul et al. *This was created after the input VCF files had been filtered for 'PASS' and 'SOMATIC' and the top 20 FLAG genes had been removed.*

Figure 11 shows the oncoplot created with the 'PASS' and 'SOMATIC' filters applied, and FLAG gene results removed, as discussed in section 2.4: Initial filtration of called variants. While the number of multi-hit gene results have decreased, there are still far more than expected for this tumour type in this dataset indicating that there are still issues with the variant filtration. However, when looking at the sample in more detail, it became clear that the pipeline was picking up expected mutations, as seen in Figure 12.

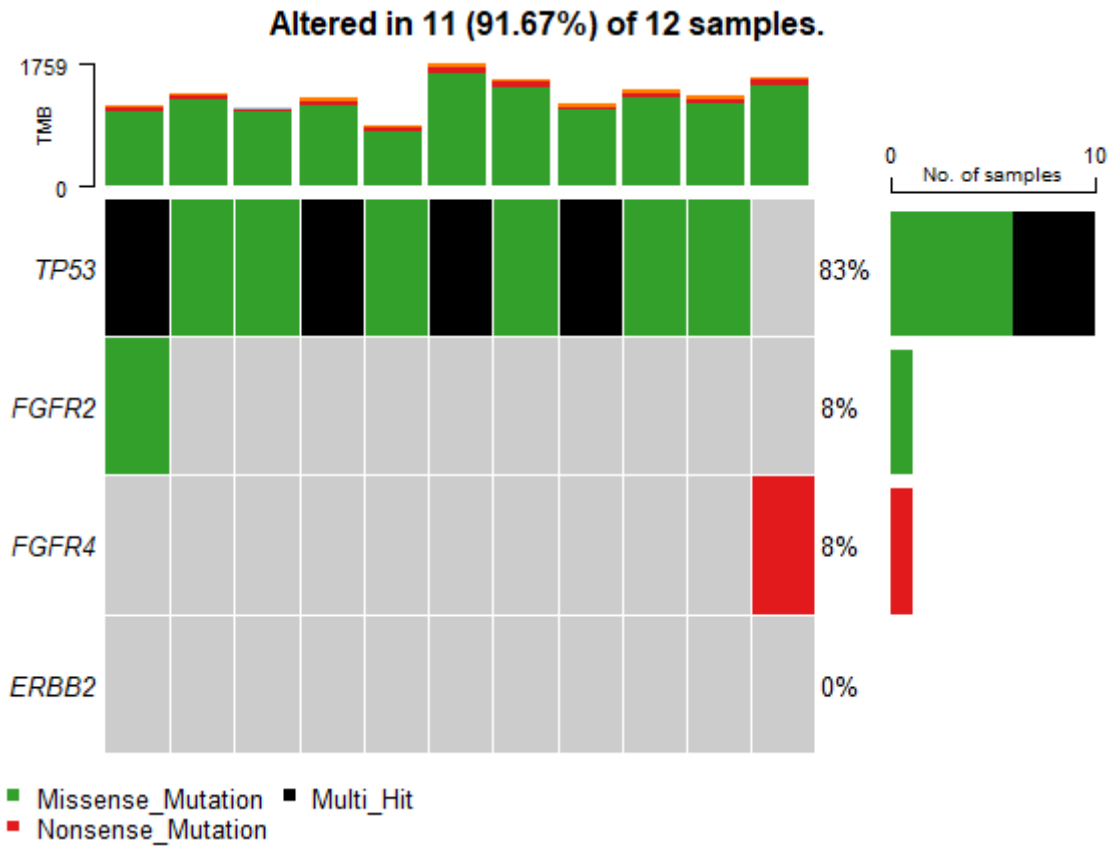


Figure 12: This figure shows an oncoplot for the 12-sample whole genome sequence subset from Jusakul et al. *This figure focuses on mutations in several genes of specific clinical interest, including* TP53, FGFR2, FGFR4 *and* ERBB2.

Figure 12 is an oncoplot showing the mutations in a few genes of interest; *TP53, FGFR2, FGFR4* and *ERBB2*. The results here are close to what would be expected from the previously published data (15); nearly all the samples had a mutation in *TP53,* while none had a mutation in *ERBB2* (which is expected, as *ERBB2* is far more affected by amplification than mutation in biliary tract cancer (15, 17)). 17% (2/12) of samples had an *FGFR* mutation, which is a similar proportion to that found in previous studies (15).

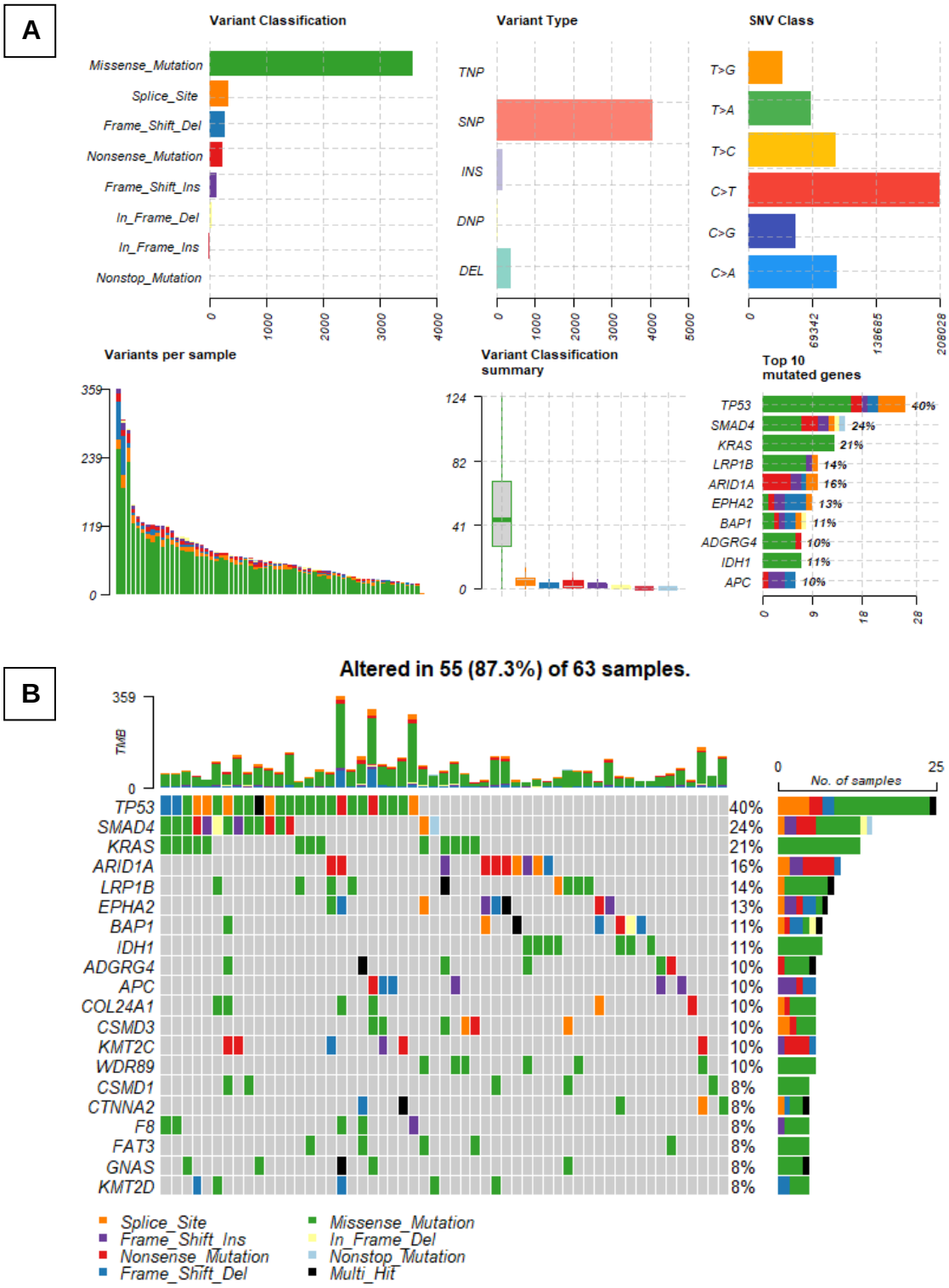


Figure 13: This summary of the MAF files (A) and oncoplot (B) show the mutations present in the 71 whole genome sequencing samples from the Jusakul et al. *dataset. The samples were analysed using nf-core/sarek for alignment with Mutect2, Funcotator and Maftools used for variant calling, annotation and visualisation respectively.*

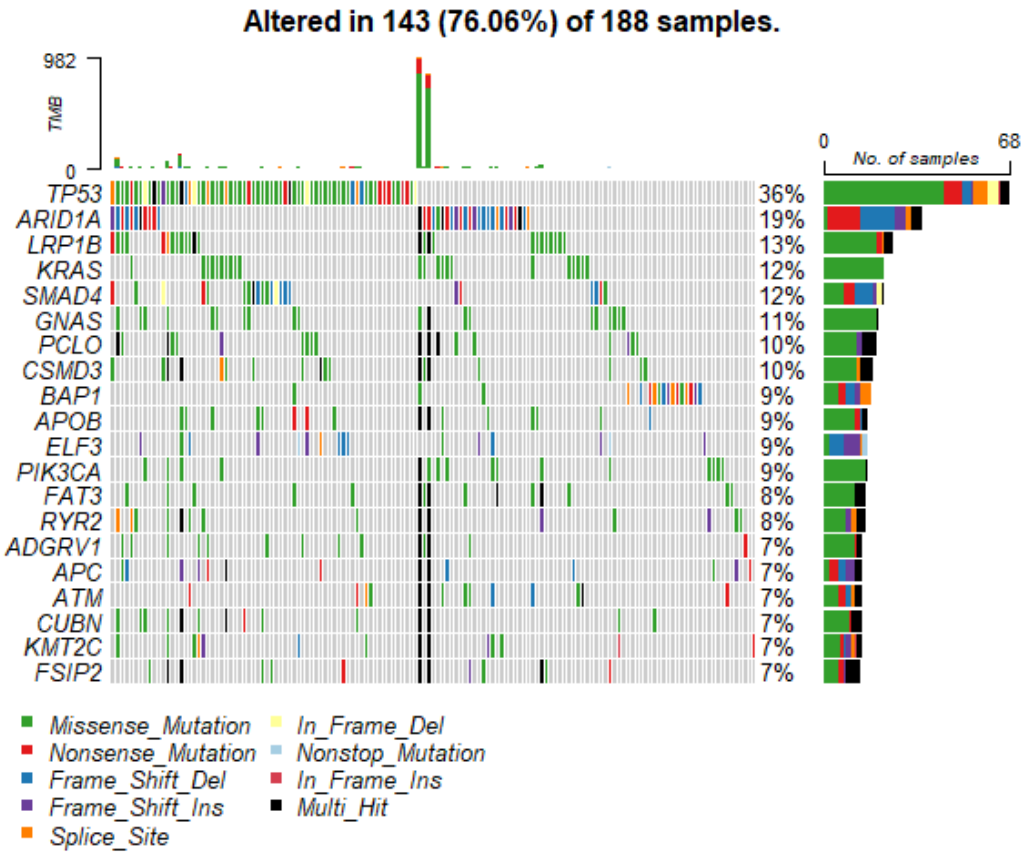


Figure 14: This oncoplot shows the 188-sample targeted sequence subset from the Jusakul et al. *dataset, created using nf-core/sarek for alignment with Mutect2, Funcotator and Maftools used for variant calling, annotation and visualisation respectively.*

Figure 13 and Figure 14 both show oncoplots from the re-analysis of the Jusakul *et al.* dataset once further filtering of the called variants had been put into place. Figure 13 shows the results of the re-analysis of the 71 whole genome sequencing subset (datasets EGAD00001003834 and EGAD00001001988) and Figure 14 shows the results of the re-analysis of the 188 targeted sequencing subset (EGAD00001001994), using the pipeline developed during this project. The mutations found and their proportions using the developed pipeline are very close to those found in Jusakul ***et al.'s*** paper; although they are not identical as the full dataset from the original paper was not available and slightly different tools were used in this analysis compared to the original. However, as the results are very similar it indicates that the developed pipeline ran as expected to find the results expected in the re-analysis of this dataset. The result of running this pipeline indicate that this pipeline could successfully be used in the analysis of other datasets.

3.1.2: Re-analysis of a previously published dataset of patients with gallbladder cancer (Pandey *et al.* dataset)

The first step in the analysis of the data from Pandey *et al.* was to run the pipeline developed while working on the Jusakul *et al.* dataset. The first dataset run was the WGS dataset; the initial FASTQ files were aligned using nf-core/sarek, variants were called using Mutect2 and the VCFs annotated using Functotator to produce the oncoplot seen in Figure 15.

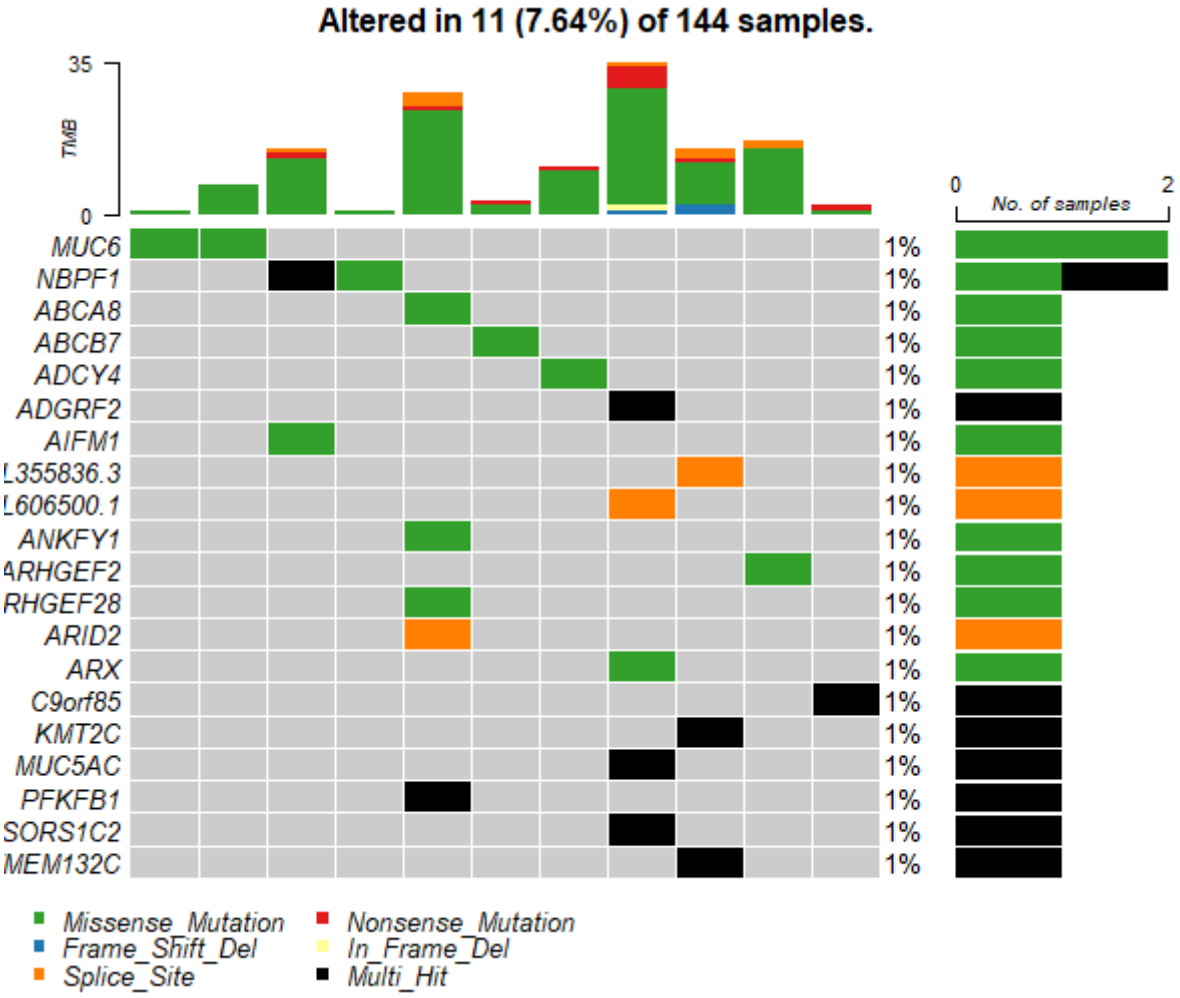


Figure 15: This is the oncoplot generated from the whole genome sequencing data from the Pandey et al. *gallbladder cancer dataset.*

However, the mutations seen in Figure 15 are not as expected, with genes expected to be frequently mutated in gallbladder cancer apparently not present in this sample. This was initially perplexing, until it became apparent that the whole genome data in

this dataset was in fact low-pass whole genome data. The depth of low-pass whole genome data is much less than normal whole genome or whole exome data (0.1x – 1x instead of >10-50^{+}x), which makes it unsuitable for mutation analysis as gaps in the sequencing are likely to miss mutations. Low pass WGS is more suited to copy number analysis, which is discussed in more detail below.

As the whole genome data was unsuitable for mutation analysis, the whole exome data was examined instead. The same pipeline created before, using nf-core/sarek, Mutect2 and Functator, was used to create the oncoplot seen in Figure 16.

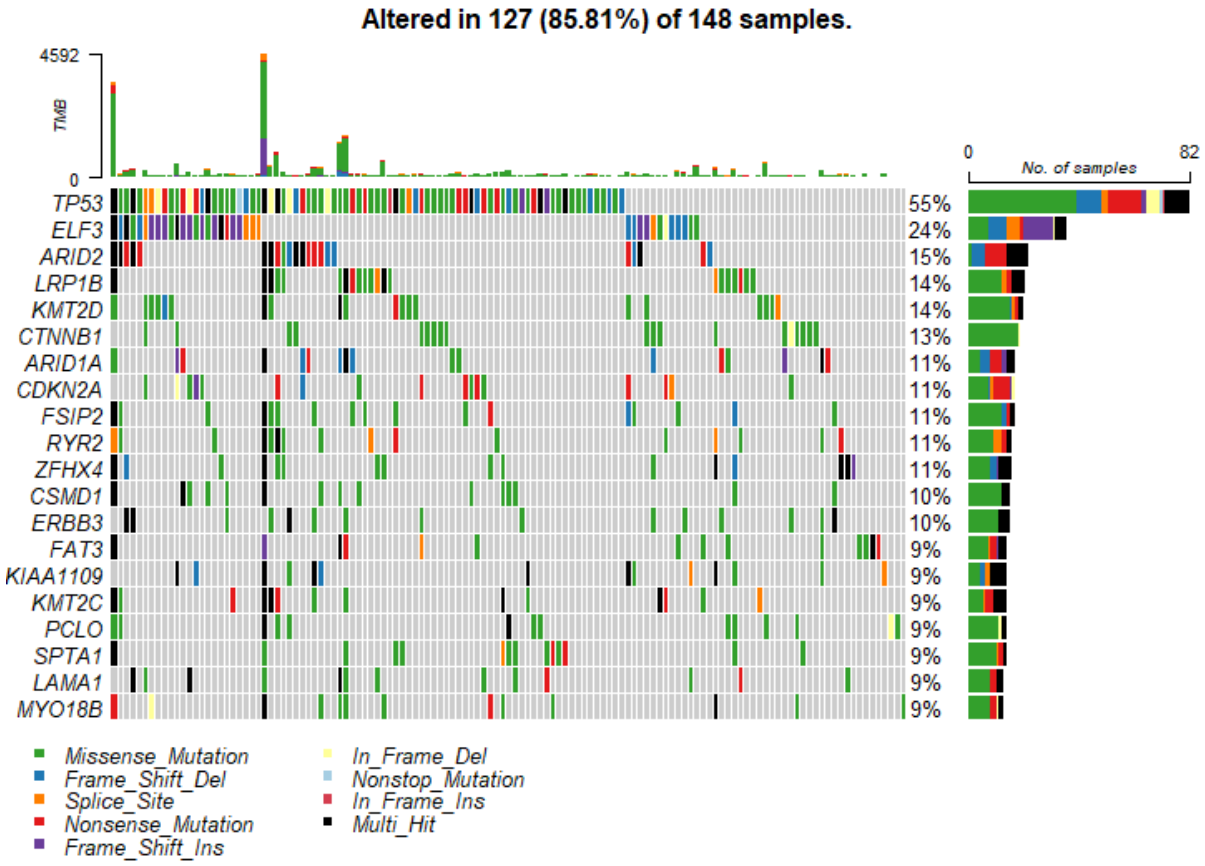


Figure 16: This is the oncoplot generated from a re-analysis of the whole exome sequencing data from the Pandey et al. *gallbladder cancer dataset, using the pipeline developed in the re-analysis of the Jusakul et al. dataset.*

Figure 16 is an oncoplot showing the 20 most commonly mutated genes in the Pandey *et al.* dataset, produced by a re-analysis of the Pandey *et al,* whole exome dataset using the pipeline developed in the re-analysis of the Jusakul *et al.* dataset

as mentioned above. These results are very similar to those in the published data from Pandey *et al.* with the differences in the tools used taken into account, again demonstrating that the pipeline is working as expected.

3.1.3: Re-analysis of Pandey *et al.* dataset to create a pipeline for copy number analysis

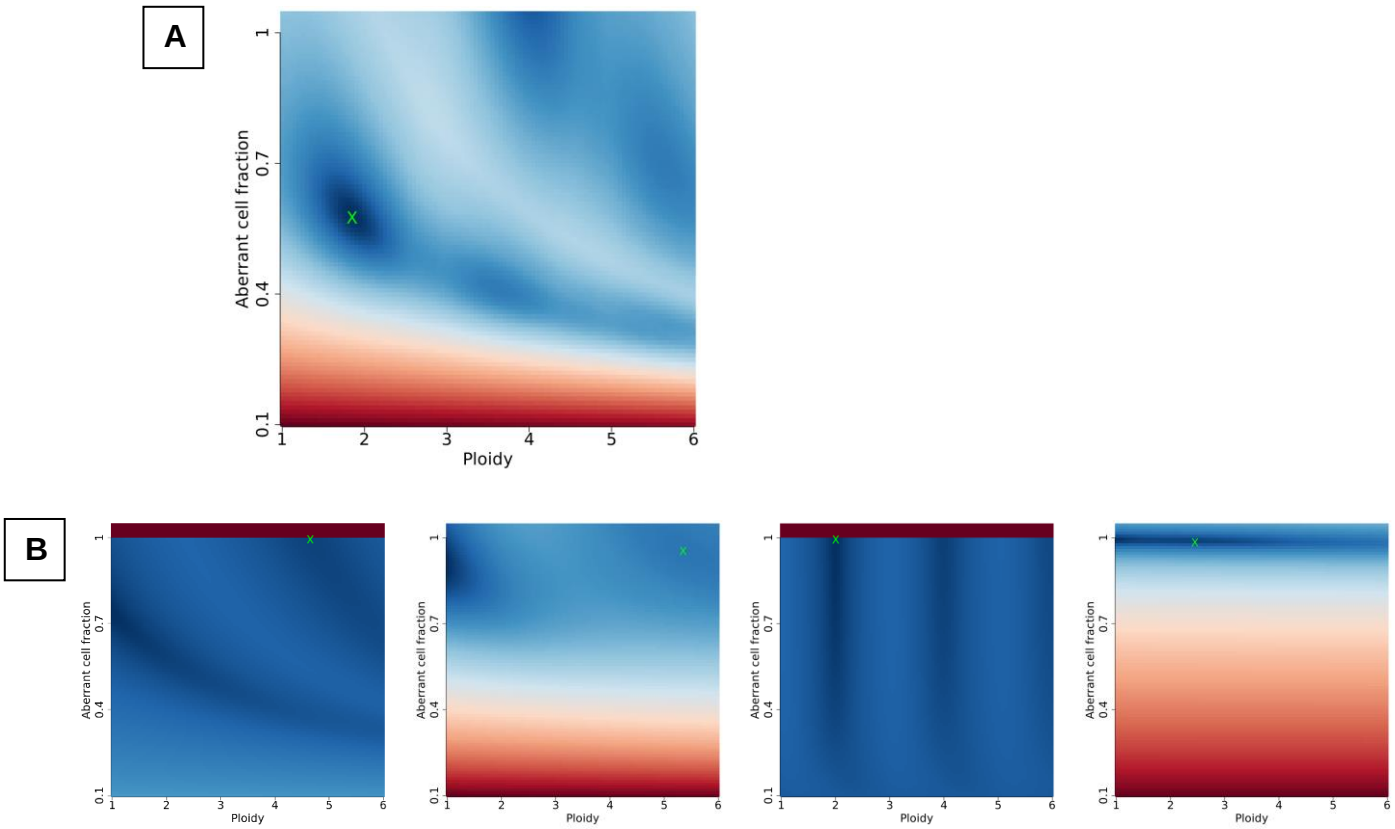


Figure 17: This figure shows examples of some of the sunrise plots produced using nf-core/sarek version 2.6.1 on the lp-WGS gallbladder cancer data from Pandey et al. *The sunrise plots should look like the example in (A), where the true ploidy is at the highest density area of the plot. However, the vast majority of the plots looked like the samples in (B) where it is impossible to calculate the true ploidy of the samples.*

Initially, nf-core/sarek was used to re-analyse the lp-WGS data from the Pandey *et al.* dataset, using a tool called ASCAT as part of the pipeline. As seen in Figure 17, while some of the calculations worked to produce the ploidy of the samples, the vast majority did not produce sensible results. The sunrise plots should have an area of highest density which ASCAT can then use to calculate the true ploidy of the original

tumour sample; a successful example of which is seen in Figure 17 [A]. Unfortunately, the vast majority of the sunrise plots produced looked like the examples in Figure 17 [B], where there are no areas of highest density from which ASCAT can calculate the ploidy of the sample.

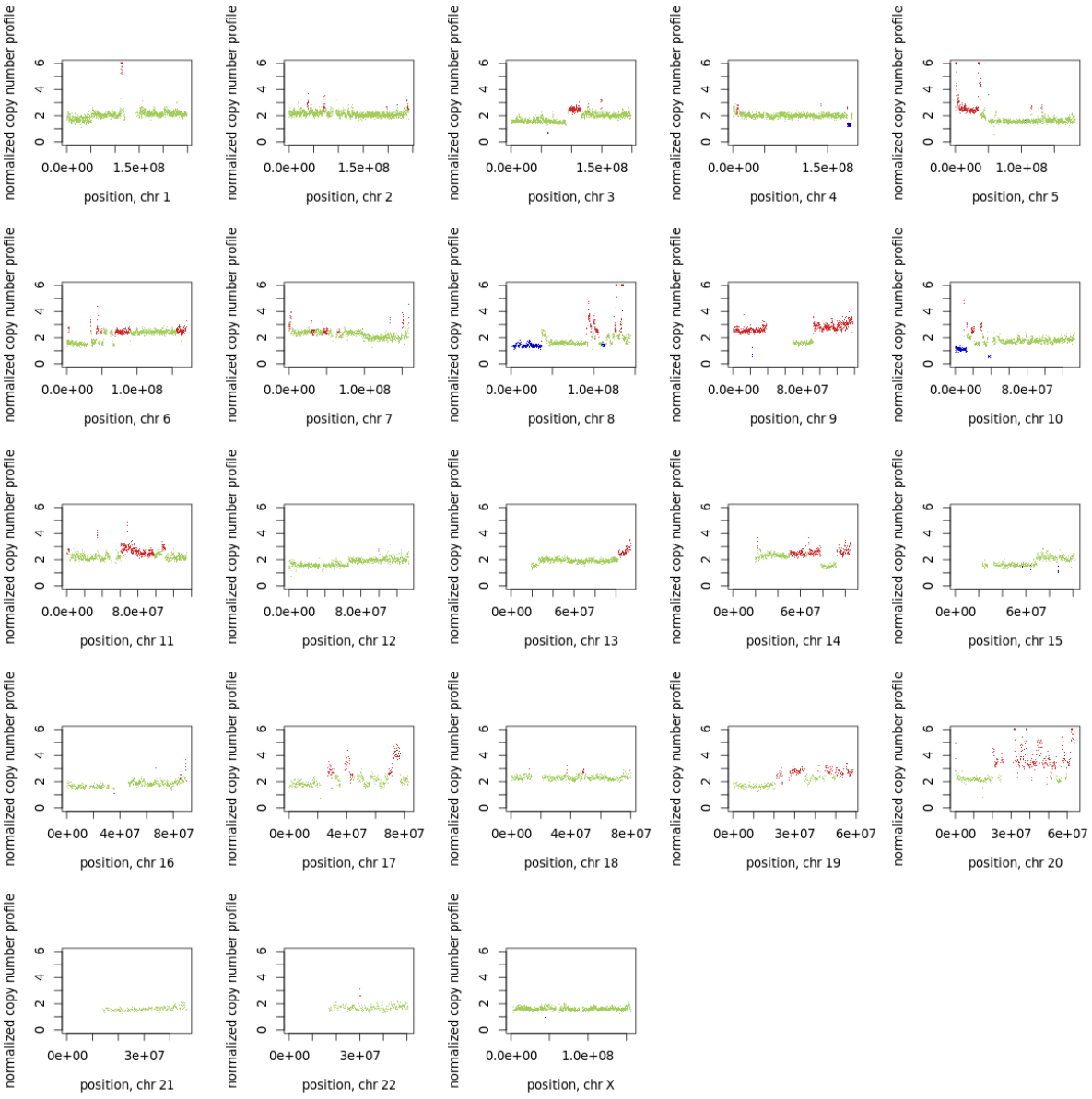


Figure 18: This figure shows the results of copy number analysis run using nf-core/sarek control-FREEC on one of the lp-WGS gallbladder cancer samples from Pandey et al.*'s dataset. Each chromosome is shown as its own plot, with areas of copy number gain shown in red and areas of copy number loss shown in blue.*

Since ASCAT via nf-core/sarek was often unable to calculate results for sample ploidy, and therefore copy number, a different tool called control-FREEC was instead

used via nf-core/sarek. As seen in Figure 18, control-FREEC was able to produce sensible looking copy number data in graphical form with areas of copy number gain and loss clearly visible on the output. However, as discussed in section 2.7: Copy number analysis pipeline development, the storage and processing time required for using control-FREEC via nf-core/sarek on a single sample meant that it would have been impossible to scale up to analyse entire datasets. Instead, an updated version of the R package version of ASCAT was used to re-analyse the lp-WGS data, the results of which are seen in Figure 19.

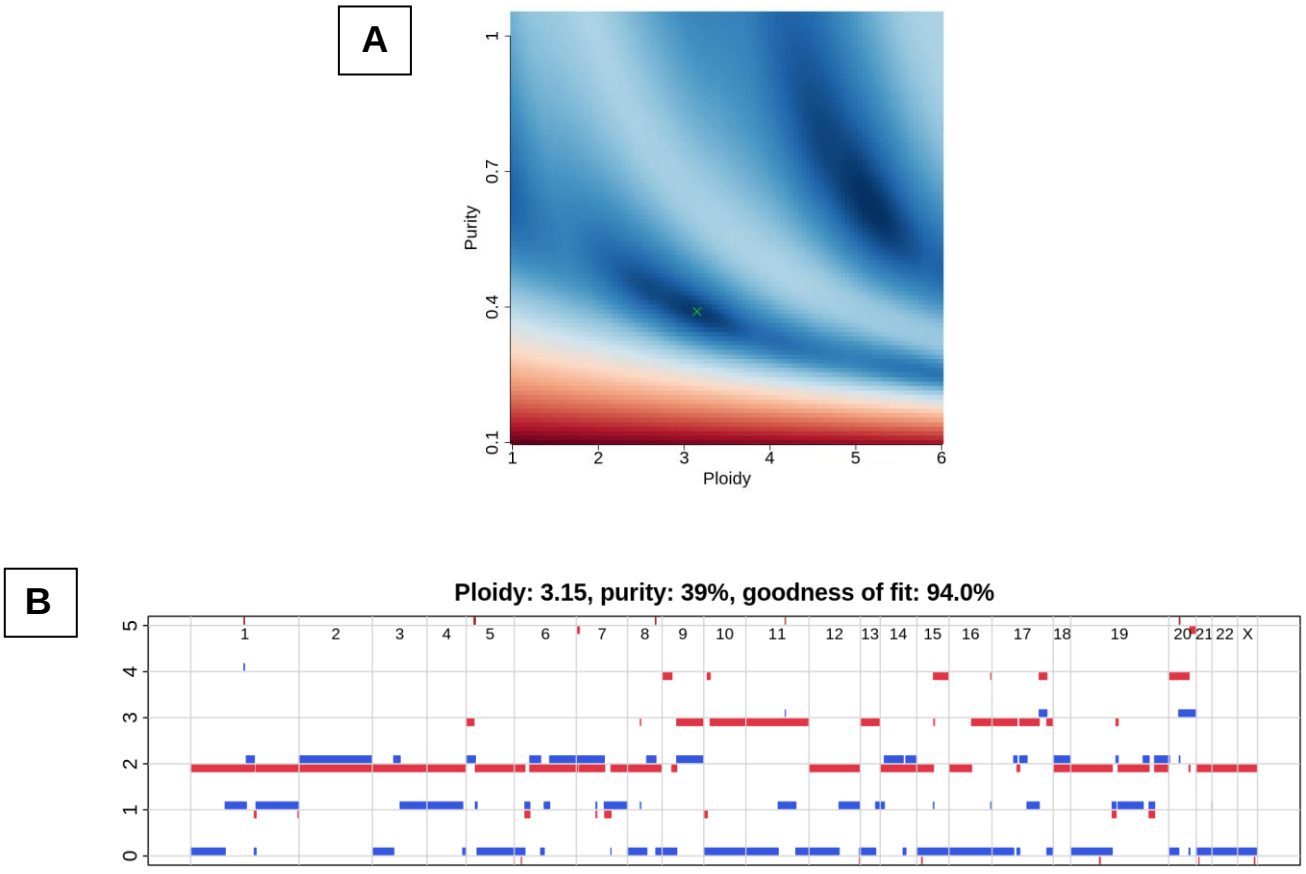


Figure 19: This figure shows a sunrise plot (A) and an ASCAT profile (B) for one of the whole exome gallbladder cancer samples, showing the results of a successful run of ASCAT. In this particular sample, the overall ploidy of the sample is 3, with areas of significant genome amplification.

Figure 19 shows the results of a single sample analysis using the updated version of ASCAT. In this randomly selected single lp-WGS sample from the Pandey *et al.* dataset, ASCAT shows that the sample had an overall ploidy of 3, with areas of significant genome amplification.

3.2: Analysis of data from the BILCAP clinical trial

3.2.1: BILCAP1 and initial analysis of the BILCAP data

The BILCAP data did not become available all at once. As discussed in section 1.4.1: Introduction and rationale, the trial recruited a total of 447 patients (20), but not all the patients had sufficient tissue available for analysis and not every patient had consented to further analysis of their tissue (although the vast majority did).

The first sequencing data from the BILCAP cohort consisted of 46 patients, 12 of whom had gallbladder cancer, 14 had intrahepatic cholangiocarcinoma, 13 had perihilar cholangiocarcinoma and 7 had distal cholangiocarcinoma. The raw targeted sequencing data had been analysed by the team at Birmingham using a proprietary pipeline from Qiagen (Hilden, Germany) which is based on the GATK Mutect2 tool. The output from this pipeline was then analysed and visualised using maftools, as seen in Figure 20.

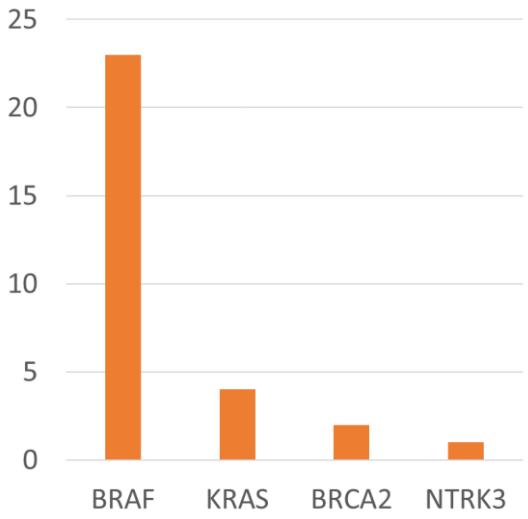


Figure 20: This figure shows a bar chart of the most frequently mutated genes found in the first part of the BILCAP dataset analysed.

When compared to the results of the Jusakul *et al.* and Pandey *et al.* datasets, these results made little sense. While *BRAF* mutations are present in biliary tract cancer, they are normally present at a frequency of <5% (15). The initial BILCAP dataset showed that the rate of *BRAF* mutations was close to 50%, which was highly

unusual. There were also few signs of the *IDH1, ERBB2* and *FGFR2* alterations which are much more common in biliary tract cancers (15, 17).

These results continued to make little sense with the analysis of further samples from the BILCAP clinical trial (as seen below), so these samples were excluded from the final analysis. In the end, it turned out that this was the correct thing to do. There had been a mix-up in the extraction and sequencing process, and so the samples labelled as ‘BILCAP1’ were in fact samples from patients with colorectal cancer. These results were therefore excluded, and not included in any further analysis of the BILCAP cohort of cholangiocarcinoma patients.

3.2.2: BILCAP2 and analysis of the first patients from the BILCAP clinical trial

As the results of the analysis of the first patients from BILCAP made little sense, when data from the next batch of patients from BILCAP became available they were analysed on their own, and not combined with any previous data. The new cohort of patients from the BILCAP trial consisted of data from 45 patients with intrahepatic cholangiocarcinoma. All 45 of these samples successfully underwent low-pass whole genome sequencing and RNA-seq, and 39 of the 45 samples successfully underwent targeted gene sequencing. In this cohort, the median age was 61 years (95% CI 57.8 – 64.2), 55.6% were female and 23 received capecitabine (51.1%).

Figure 21 shows the results of the quantiseqr analysis of the first 45-patient cohort and demonstrates the proportion of immune system cells in each sample, thus showing the tumour microenvironment (TME) in each sample. Interestingly, some samples have very high numbers of regulatory T-cells, possibly indicating that these tumours had a ‘cold’ TME and would likely not have responded well to immunotherapy.

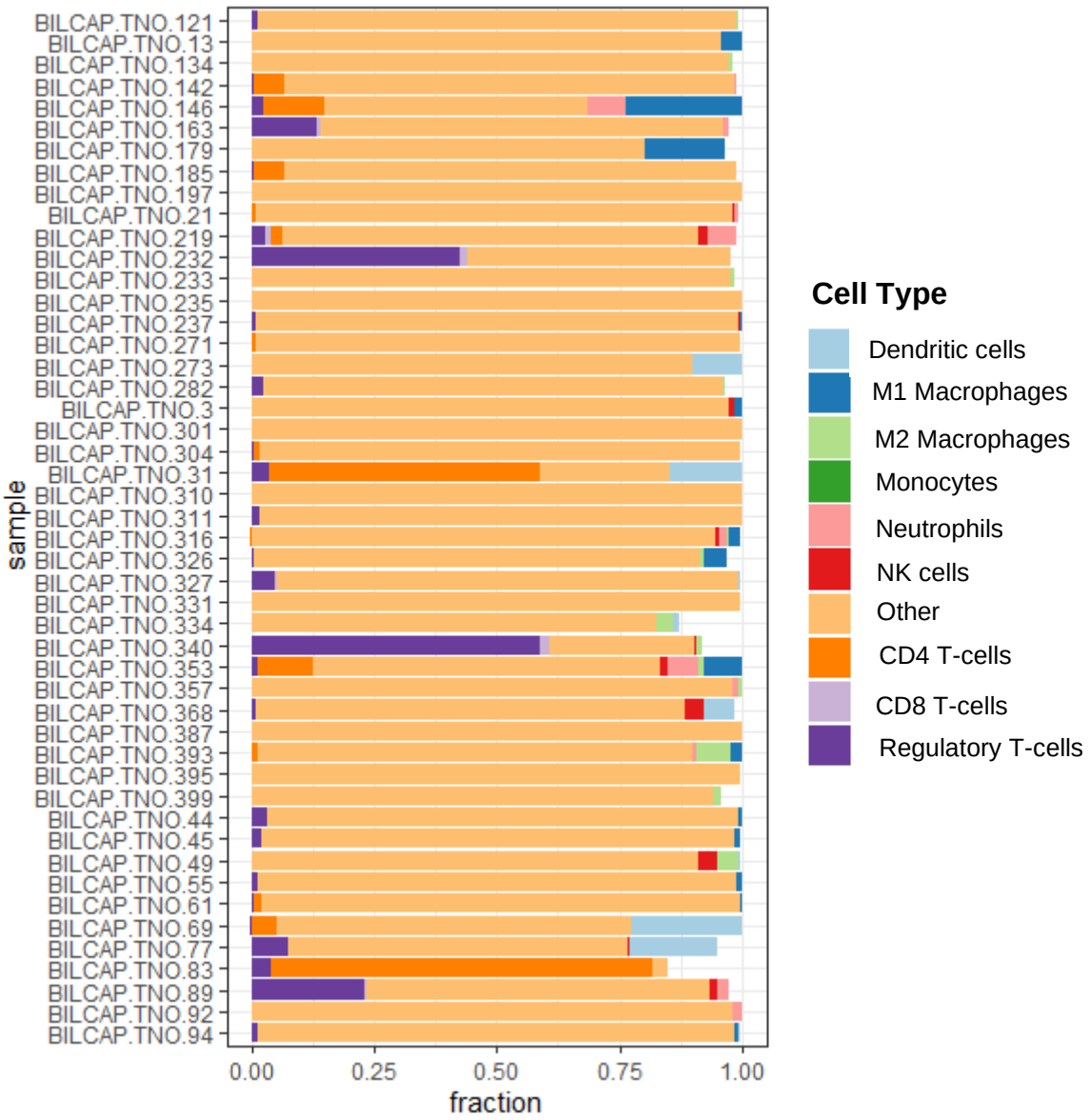


Figure 21: This is a plot produced using quantiseqr, which shows the proportion of different immune system cells present in samples from patients in the first 45 patient cohort.

Table 1: This table shows the proportion of patients with various gene alterations in the initial 45 intrahepatic cholangiocarcinoma cohort.

Alteration	**Total number**	**Percentage altered (%)**
FGFR2 gene fusions	9	20.0
NTRK1 fusions	3	6.7
FGFR1 fusions	3	6.7
FGFR3 fusions	2	4.4
ROS1 mutations	12	33.3
MET mutations	10	27.8
ALK mutations	7	19.4
IDH1 pathogenic mutations	4	11.1
FGFR2 pathogenic mutations	1	2.7
NTRK1 amplification	9	20.0
ERBB2 amplification	8	17.8
MET amplification	7	15.6
Regulatory T-cell fraction > 0.2	3	6.7
CD4 T-cell fraction > 0.2	3	6.7

As seen in Table 1, *FGFR2* gene fusions were present in 9 patients (20.0%), as were fusions in *NTRK1* (n=3, 6.7%), *FGFR1* (n=3, 6.7%), *FGFR3* (n=2, 4.4%) and *FGFR4* (n=2, 4.4%). These results are slightly higher than expected, and are likely due to the particularly permissive settings used in analysing fusions using Kallisto in order to ensure that no fusions with unusual fusion partners were missed.

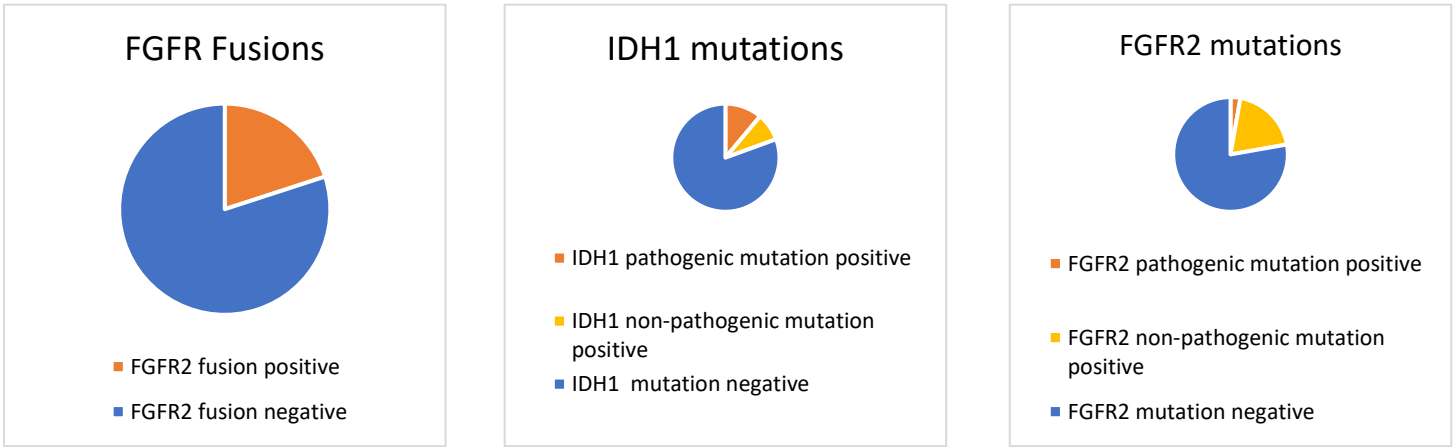


Figure 22: These pie charts show the proportion of the 45-patient cohort who has FGFR2 fusions, IDH1 *mutations and* FGFR2 *mutations.*

Commonly mutated driver genes included *ROS1* (n=12, 33.3%), *MET* (n=10, 27.8%) and *ALK* (n=7, 19.4%) with known pathogenic variants seen in *IDH1* (n=4, 11.1%;

total number of mutations=7, 19.4%), *BRAF* (n=2, 5.6%; total n=6, 16.7%), *FGFR2* (n=1, 2.7%; total n=8, 22.2%), *FGFR3* (n=1, 2.7%, total n=6, 16.7%), *IDH2* (n=1, 2.7%; total n=6, 16.7%) and *EGFR* (n=1, 2.7%; total n=4, 11.1%). Commonly amplified (CN ≥ 4, ploidy < 3) genes included *NTRK1* (n=9, 20.0%), *ERBB2* (n=8, 17.8%), and *MET* (n=7, 15.6%).

The pie charts in Figure 22 show the proportion of patients in the cohort with *FGFR2* fusions, *IDH1* mutations and *FGFR2* mutations. Pathogenic mutations in *FGFR2* included F276C and pathogenic *IDH1* mutations included 2 R132G mutations, and one mutation each in R132S and R132L.

The overall proportion and types of alterations seen in this cohort was similar to that seen in previously published data (15, 17). However, what is novel about this dataset is that information on patient outcomes was also available, such as relapse-free and overall survival of the patients. The overall picture was that most of the alterations investigated in this cohort did not significantly affect recurrence risk (PFS) or overall survival (OS), including *FGFR2* fusions (OS HR 1.23, p=0.695; recurrence HR 1.32 p=0.555) (see Figure 23).

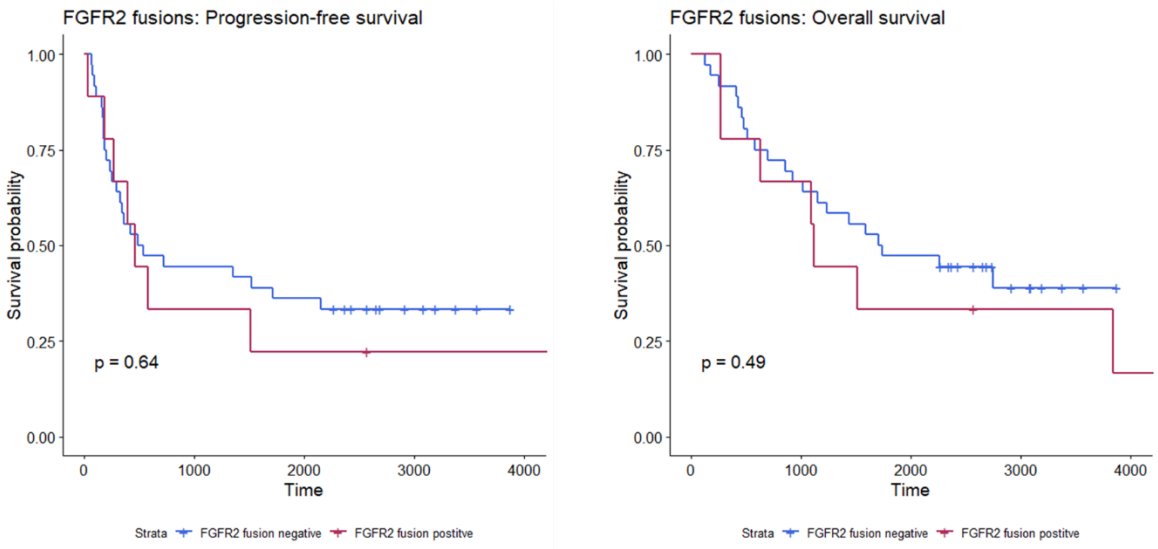


Figure 23: These Kaplan Meier curves demonstrate that the presence of an FGFR2 *fusion did not significantly affect either relapse-free or overall survival.*

However, the presence of a *FGFR3* fusion gene significantly reduced OS (OS HR 6.57, p=0.0091; recurrence HR 3.71, p=0.0734), and having ≥ 4 copies of either *NTRK1* (OS HR 3.55, p=0.0027; recurrence HR 3.48, p=0.0019) or *MET* (OS HR 6.06, p<0.001; recurrence HR 6.05 p<0.001) significantly reduced OS and increased the risk of recurrence. Although the overall numbers of patients were relatively small, the differences in survival with patients with *MET* amplification was particularly striking (Figure 24).

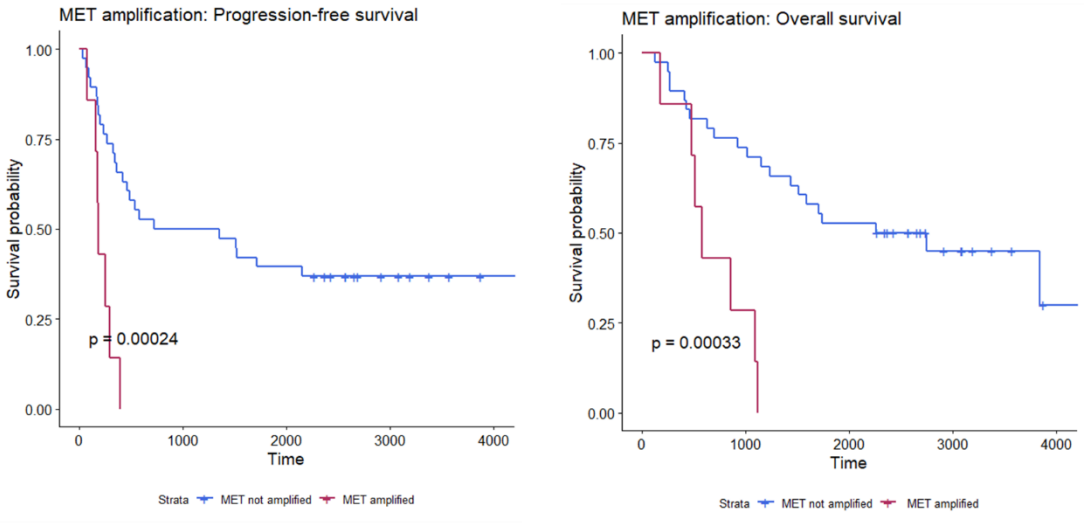


Figure 24: These Kalan Meier curves show the differences in relapse-free and overall survival in patients with and without amplified MET *(≥4 copies), where patients with amplified* MET *did significantly worse in both overall and relapse-free survival than those without.*

3.2.3: BILCAP3 and analysis of different biliary tract cancer subtypes

As more samples from patients from the BILCAP clinical trial became available, more sequencing data became available and more patterns began to emerge between the number and type of genomic alterations seen and their effect on patient outcome, including relapse-free and overall survival.

After the initial 45 intrahepatic cholangiocarcinoma patient cohort was available, the results of which are seen in section 3.2.2: BILCAP2 and analysis of the first patients from the BILCAP clinical trial, the next cohort consisted of an additional 53 patients, leading to a new total cohort of 98 patients. Unlike the first group, this cohort

consisted of different anatomical subtypes of biliary tract cancer, with 47 having intrahepatic cholangiocarcinoma, 47 with gallbladder cancer, 2 with distal cholangiocarcinoma and 2 with perihilar cholangiocarcinoma. All 98 patient samples underwent RNA-seq, with 95 of 98 undergoing low-pass WGS and 39 of 98 undergoing targeted sequencing. Unfortunately, issues with the Qiagen sequencing kits for the targeted sequencing meant that the numbers of patients successfully undergoing targeted sequencing lagged behind the samples undergoing lp-WGS or RNA-seq.

Again, the overall types of genomic alterations found were similar to the smaller cohort and to previously published data. 62 (63.3%) patients were female, and 48 (49.0%) received adjuvant capecitabine. *FGFR2* gene fusions were present in 24 patients (24.5%), as were fusions in *NTRK1* (n=4, 4.1%), *FGFR1* (n=4, 4.1%), *FGFR3* (n=2, 2.0%) and *FGFR4* (n=2, 2.0%). Known pathogenic mutations were seen in *IDH1*(n=4, 10.3%, total number of mutations=8, 20.5%), *IDH2* (n=1, 2.6%, total n=6, 15.4%), and *FGFR2* (n=1, 2.6%, total n=7, 17.9%). Commonly amplified (copy number ≥ 4) genes included *NTRK1* (n=28, 29.5%), *ERBB2* (n=27, 28.4%) and *MDM2* (n=20, 21.1%) with *MYC* (n=19, 20.0%), *EGFR* (n=16, 16.8%) and *MET* (n=15, 15.8%) also amplified.

With a larger cohort of patients, the evidence became even stronger that the majority of the genomic alterations did not affect relapse-free (Figure 25) or overall survival (Figure 26).

Overall survival

Variable	N	Hazard ratio	p
FGFR2 fusions	24 / 98 (24.5%)	0.94 (0.50, 1.75)	0.83
NTRK1 fusions	4 / 98 (4.1%)	0.30 (0.04, 2.20)	0.24
FGFR1 fusions	4 / 98 (4.1%)	1.13 (0.34, 3.74)	0.84
FGFR3 fusions	2 / 98 (2.0%)	5.84 (1.32, 25.81)	0.02
Regulatory T-cell fraction > 0.05	17 / 98 (17.3%)	0.60 (0.29, 1.24)	0.17
CD4 T-cell fraction > 0.05	12 / 98 (12.2%)	0.95 (0.44, 2.03)	0.89

Variable	N	Hazard ratio	p
Known pathogenic IDH1 mutations	4 / 39 (10.3%)	0.55 (0.12, 2.42)	0.4

Variable	N	Hazard ratio	p
NTRK1 amplification	28 / 95 (29.5%)	0.98 (0.48, 1.98)	0.94
ERBB2 amplification	27 / 95 (28.4%)	0.99 (0.53, 1.85)	0.97
MET amplification	15 / 95 (15.8%)	0.42 (0.11, 1.52)	0.19
EGFR amplification	16 / 95 (16.8%)	4.84 (1.47, 15.93)	0.01
MDM2 amplification	20 / 95 (21.1%)	1.24 (0.63, 2.43)	0.53

Figure 26: This forest plot shows which gene alterations have a significant effect on overall survival in the 98-patient cohort.

Relapse-free survival

Variable	N	Hazard ratio	p
FGFR2 fusions	24 / 98 (24.5%)	0.91 (0.50, 1.67)	0.76
NTRK1 fusions	4 / 98 (4.1%)	0.27 (0.04, 1.99)	0.20
FGFR1 fusions	4 / 98 (4.1%)	1.24 (0.38, 4.04)	0.72
FGFR3 fusions	2 / 98 (2.0%)	4.11 (0.95, 17.71)	0.06
Regulatory T-cell fraction > 0.05	17 / 98 (17.3%)	0.37 (0.16, 0.88)	0.02
CD4 T-cell fraction > 0.05	12 / 98 (12.2%)	0.91 (0.41, 2.03)	0.82

Variable	N	Hazard ratio	p
Known pathogenic IDH1 mutations	4 / 39 (10.3%)	0.53 (0.13, 2.25)	0.4

Variable	N	Hazard ratio	p
NTRK1 amplification	28 / 95 (29.5%)	0.76 (0.37, 1.58)	0.46
ERBB2 amplification	27 / 95 (28.4%)	1.05 (0.57, 1.91)	0.88
MET amplification	15 / 95 (15.8%)	0.73 (0.23, 2.33)	0.60
EGFR amplification	16 / 95 (16.8%)	3.22 (1.11, 9.35)	0.03
MDM2 amplification	20 / 95 (21.1%)	1.47 (0.78, 2.75)	0.23

Figure 25: This forest plot shows which gene alterations have a significant effect on relapse-free survival in the 98-patient cohort.

Again, *FGFR2* fusions had no significant effect on relapse-free or overall survival, as seen in the Kaplan Meier curves in Figure 27.

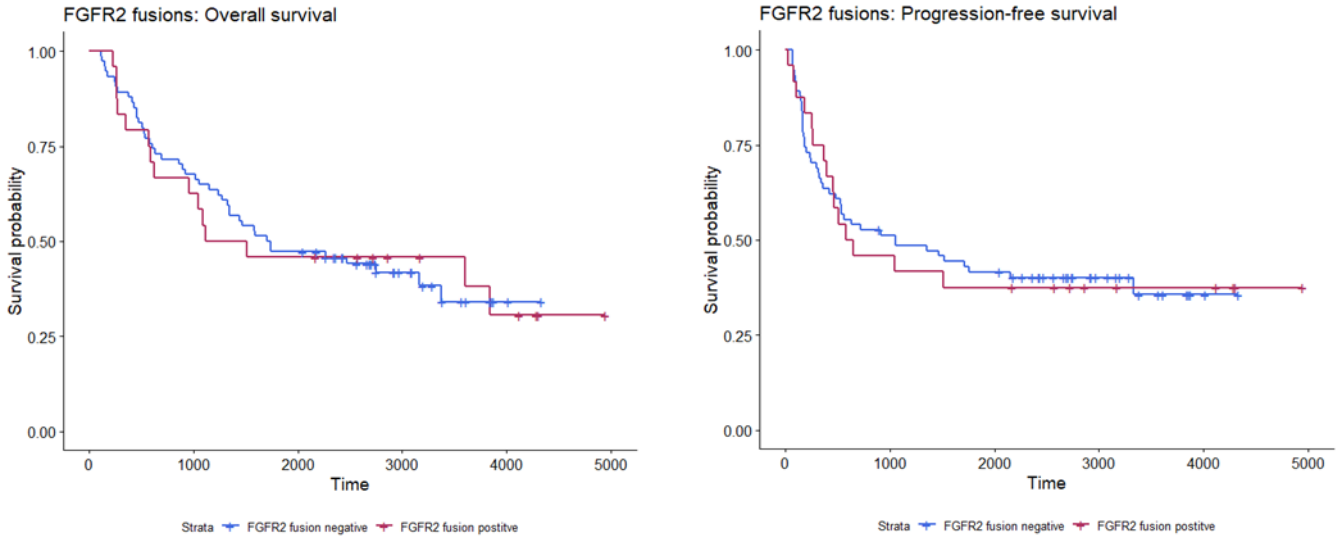


Figure 27: These Kaplan Meier curves from the 98-patient cohort show that FGFR2 fusions have no significant effect on relapse-free or overall survival.

However, while the associations between *FGFR3* fusions and *MET* amplification were no longer significant in the larger cohort (partly due to their relatively low frequency in the larger cohort), a new alteration started to show a significant effect. Patients with amplification of *EGFR* had significantly reduced both overall and relapse-free survival, as seen in the Kaplan Meier curves in Figure 28.

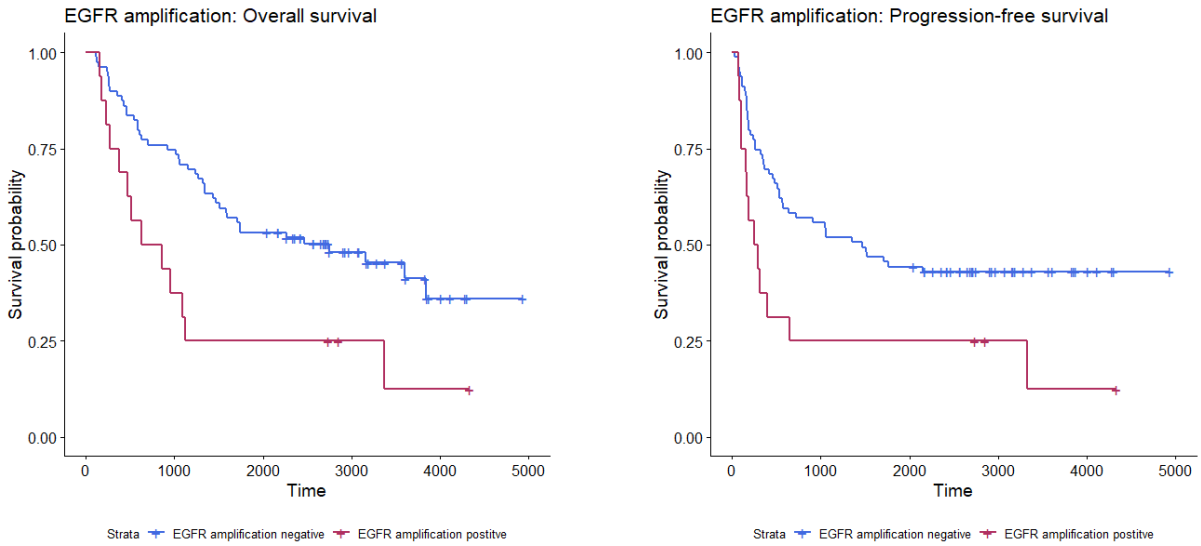


Figure 28: These Kaplan Meier curves show that amplified EGFR *(i.e. CN ≥4) decrease both relapse-free and overall survival when present in the tumours of patients with biliary tract cancer.*

3.2.4: BILCAP4 and the analysis of the majority of the BILCAP cohort

The previous analyses only analysed part of the BILCAP cohort; data from more patients became available in batches over time and the entire cohort reanalysed each time to see if there were any new patterns or associations in the data. With the next batch of data available, genomic data was available for a total of 204 patients; 50 had intrahepatic cholangiocarcinoma (iCCA), 47 had gallbladder cancer (GBC), 96 had distal cholangiocarcinoma (dCCA) and 11 had perihilar cholangiocarcinoma (pCCA) (Figure 29).

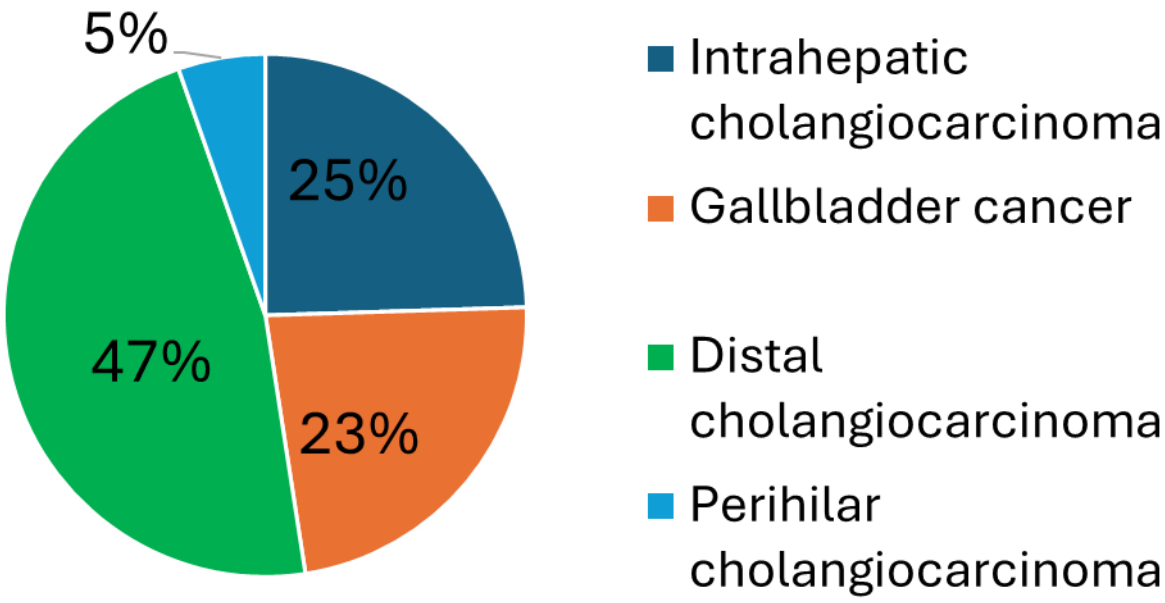


Figure 29: This pie chart shows the proportion of each of the biliary tract cancer subtypes present in the 204-patient subset of patients from the BILCAP clinical trial.

As the number of patients in the cohort increased and the number of patients in each biliary tract cancer subtype increased, it became increasingly clear that there are clear genomic differences between each subtype. Different subtypes of biliary tract cancer had significant differences in the number and type of alterations seen, as shown in Table 2. Although none of the changes are yet to show any differences in survival between different subtypes, this is being re-analysed as more data becomes available, in particular data on perihilar cholangiocarcinoma, which is still underrepresented in this dataset.

Table 2: This table shows the proportion of some of the most common genomic alterations in each of the biliary tract cancer subtypes found in the BILCAP patient cohort.

Anatomical Subtype	**Intrahepatic CCA**	**Perihilar CCA**	**Gallbladder cancer**	**Distal CCA**
FGFR2 fusions	9	0	14	4
NTRK1 fusions	3	0	1	0
FGFR1 fusions	3	0	1	2
FGFR3 fusions	2	0	0	1
Treg fraction > 0.05	6	2	11	24
CD4 fraction+ > 0.05	6	0	6	1
NTRK1 amplification	14	3	13	24
ERBB2 amplification	17	7	11	56
MET amplification	7	0	8	9
EGFR amplification	7	0	9	2
MDM2 amplification	12	5	9	42

Overall, the number and types of genomic alterations seen in the larger cohort, as seen in Table 3 mirrored that in the previous smaller cohort, as well as in previously published studies such Jusakul *et al.* (15).

Table 3: This table shows the number of different potentially targetable genomic alterations in data from 204 patients enrolled on the BILCAP clinical trial.

Alteration	**Total number (n)**	**Percentage altered (%)**
FGFR2 fusions	28	**24.5**
NTRK1 fusions	4	2.0
FGFR1 fusions	6	2.9
FGFR3 fusions	3	1.5
IDH1 pathogenic mutations	4	10.3
IDH2 pathogenic mutations	1	2.6
FGFR2 pathogenic mutations	1	2.6
NTRK1 amplification	54	26.9
ERBB2 amplification	91	45.3
MDM2 amplification	68	33.8
MYC amplification	27	13.4
EGFR amplification	18	9.0
MET amplification	24	11.9
Regulatory T-cell fraction > 0.05	43	21.1
CD4 T-cell fraction > 0.05	13	6.4

Of note, the overall number of gene amplifications appears to be very high in this dataset. For example, Jusakul *et al.* found the rate of *EGFR* amplification to be around 3% (15), whereas it is around 9% is this dataset. The rates of *ERBB2* and *MDM2* amplification are also higher in this dataset than in previously published datasets. The reason behind this is almost certainly due to the categorisation of 'amplification' during this project. Levels of amplification measured using methods such as IHC have been well characterised as they are used for clinical and diagnostic purposes. However, as amplification at a genomic level is not yet used in clinical practice there is no formal agreement as to what constitutes an amplified gene at a genomic level. A copy number level of ≥4 was used in this project to denote amplification in order to capture all gain of copy number of ≥2 and see if any level of amplification has an effect on survival and relapse, even if this level of amplification would likely be too small to be picked up on IHC (or other methods of amplification) and may not necessarily indicate increased expression at a protein level.

Looking into the amplification in more detail shows that low level amplification is common in all of the amplified genes seen in this cohort, as demonstrated in Figure 30 and Figure 31 below.

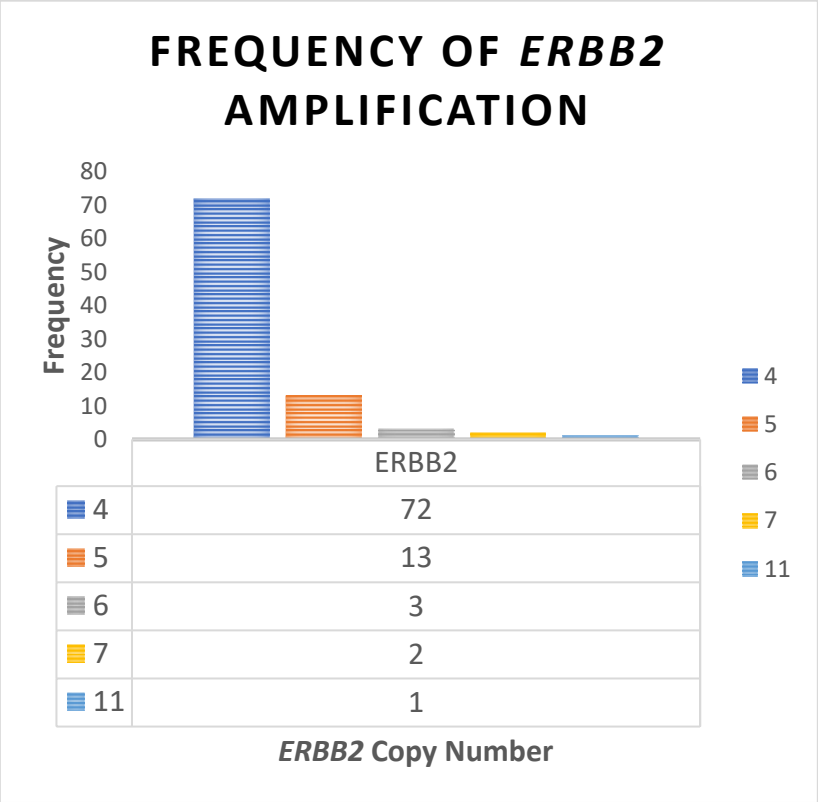


Figure 30: This figure shows the frequency of samples with an ERBB2 *amplification and copy number ≥4 in the 204-patient BILCAP cohort.*

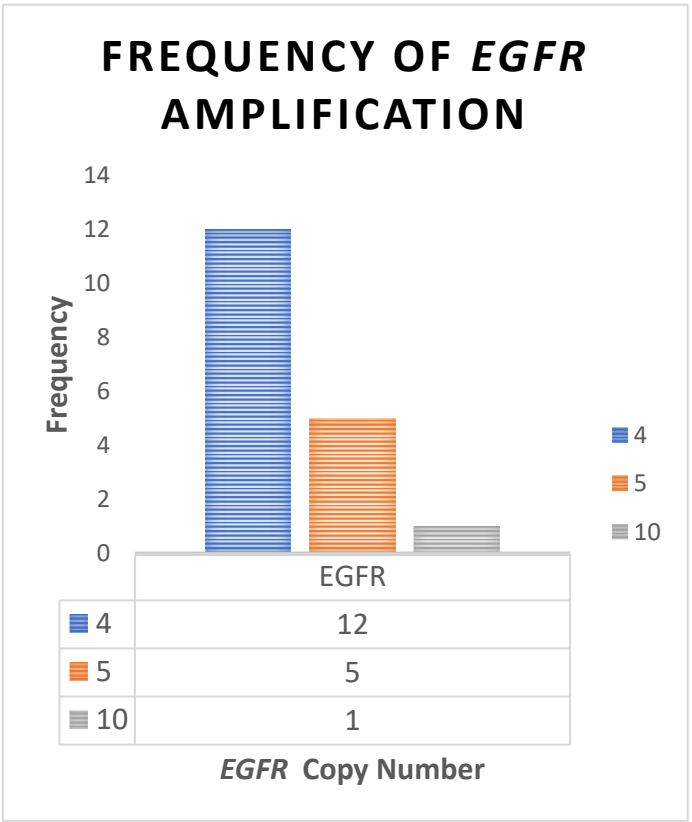


Figure 31: This figure shows the frequency of samples with an EGFR *amplification and copy number ≥4 in the 204-patient BILCAP cohort.*

Most of the amplification is a copy number of 4, which may not be detectable using protein-based methods to detect increased expression and may not be correlated with increased expression detectable using RNA based methods. When only considering a higher cut-off for copy number the numbers align much more closely with data from previous publications; for example, when considering amplified *EGFR,* the percentage of samples with copy number ≥5 is 6 / 204 or 2.9%, which matches the 3% rate seen in the Jusakul *et al.* dataset (15). A similar calculation for *ERBB2* shows that 19 / 204 or 9.3% of the BILCAP dataset had a copy number ≥5; this is slightly higher than the 3.9 – 8.5% seen in the Jusakul *et al.* dataset, but expected as the Jusakul *et al.* dataset did not include any patients with gallbladder cancer, which are known to have higher rates of *ERBB2* amplification (15, 17). Similar findings are seen in the other amplified genes investigated during this project (data not shown).

The association between overall and relapse-free survival was again examined in this larger cohort, to see if there were any associations between the genomic alterations seen and patient outcomes. As seen in the forest plots in Figure 33 and Figure 32, again, most of the genomic alterations do not have any effect. The association between *MET* amplification and poor outcome is no longer significant, but it may be the case that it is only true for intrahepatic cholangiocarcinomas, but more patient samples are needed before this can be proven. Using higher cut-offs for amplification (e.g. copy number ≥5) also did not have an effect on the significance of the alteration on relapse-free and overall survival.

Relapse-free Survival

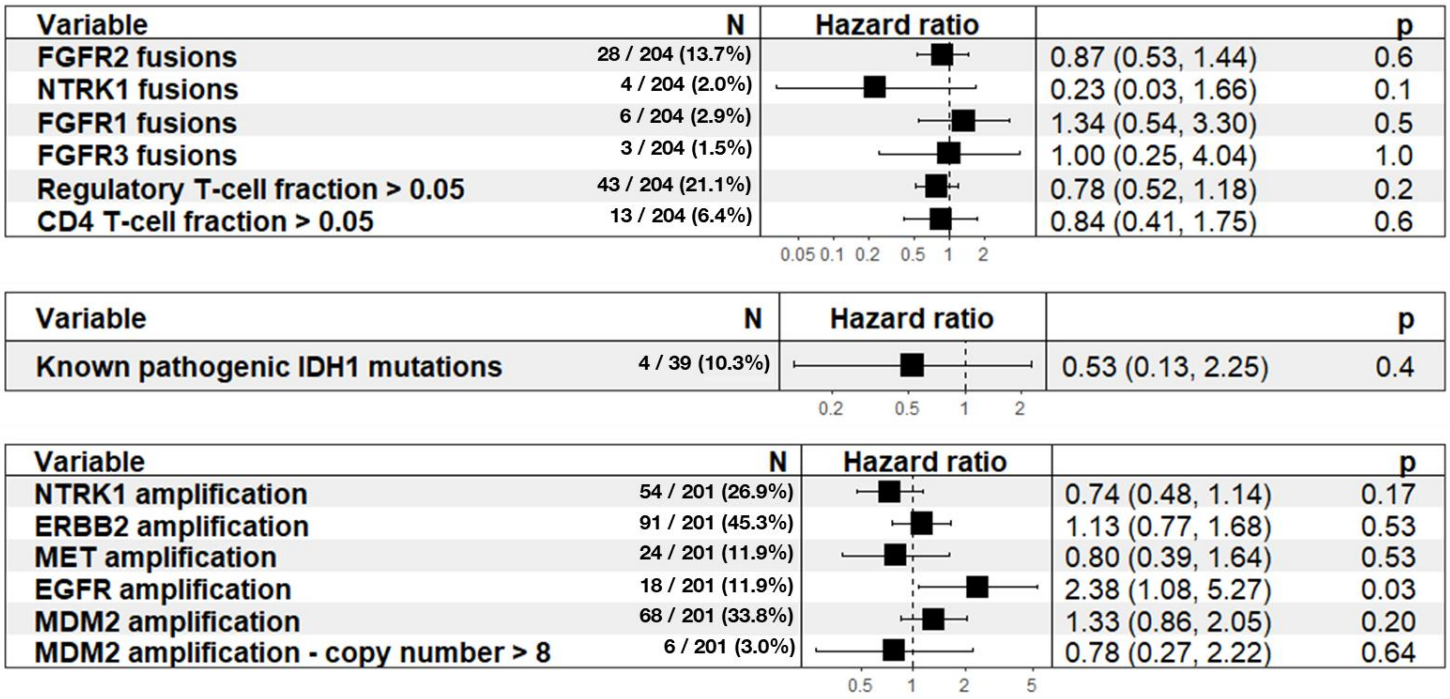


Figure 33: This forest plot demonstrates whether there are any associations between different genomic alterations and relapse-free survival in the 204 patient BILCAP cohort.

Overall Survival

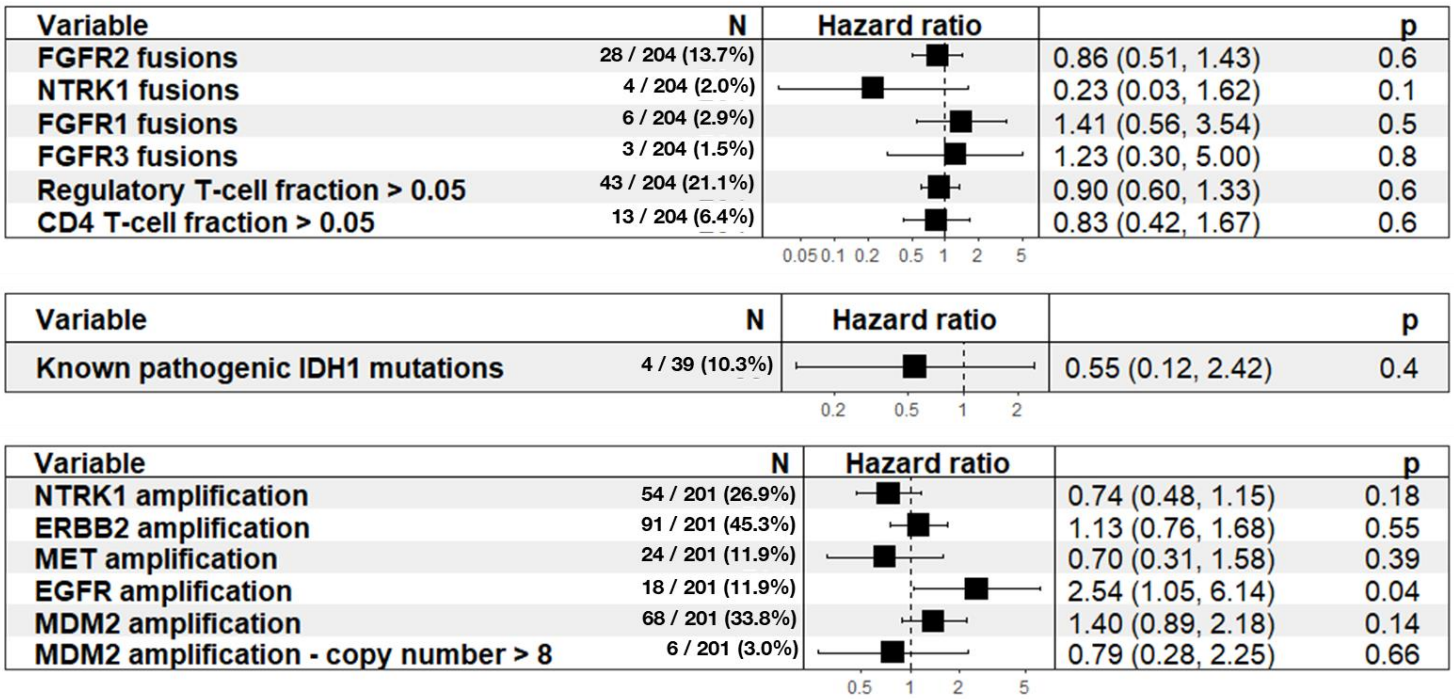


Figure 32: This forest plot demonstrates whether there are any associations between different genomic alterations and overall survival in the 204 patient BILCAP cohort.

FGFR2 fusions were, again, shown to have no impact on overall or relapse-free survival as seen in Figure 34.

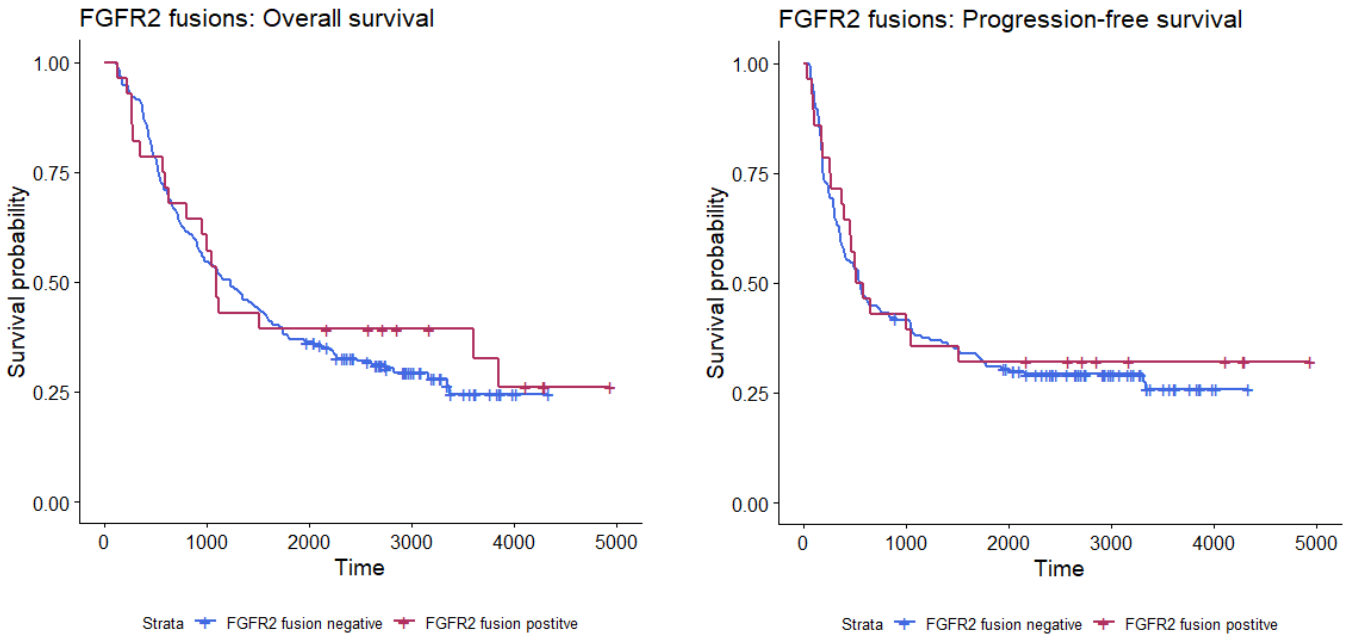


Figure 34: These Kaplan Meier curves show that there is no significant difference in overall and relapse-free survival in patients with FGFR2 *fusions.*

However, one genomic alteration did show an effect on outcomes, as shown in the Kaplan Meier curves in Figure 35.

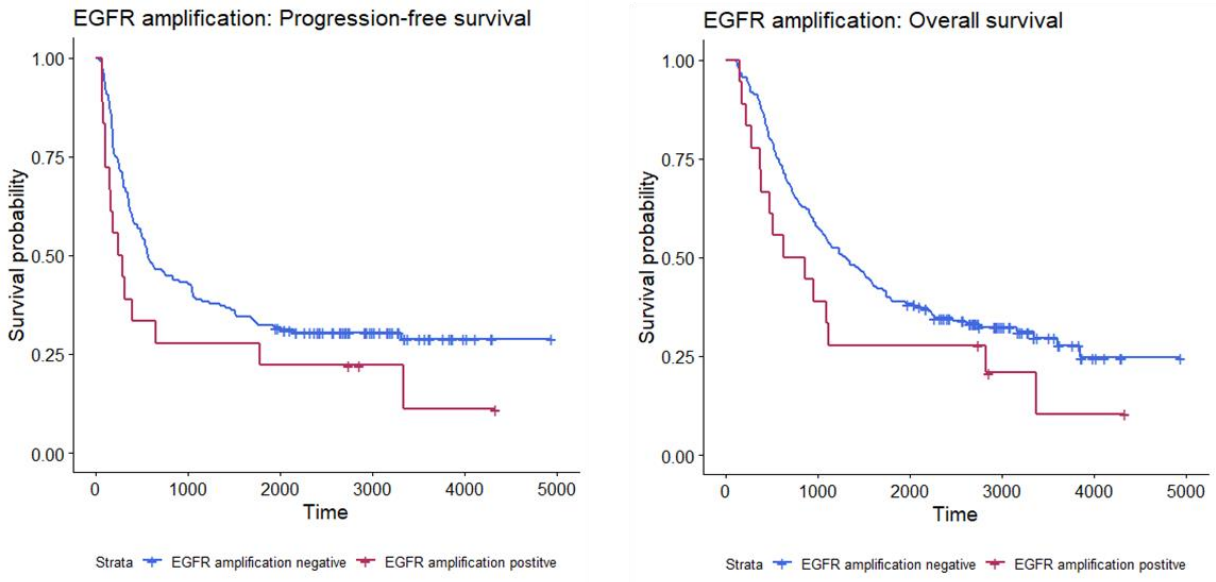


Figure 35: These Kaplan Meier curves show that the presence of EGFR *amplification (CN ≥4) significantly reduces both relapse-free and overall survival.*

EGFR amplification, which has been defined in this study as having ≥4 copies of the *EGFR* gene, significantly reduced both relapse-free and overall survival.

3.2.5: Comparison of the tumour microenvironment between early-stage (BILCAP) and locally advanced or metastatic (I3O-MC-JSBF) biliary tract cancers

RNA-seq data was available from both the BILCAP and the I30-MC-JSBF clinical trials, allowing for the comparison of the tumour microenvironments between samples in each dataset. The results of this analysis were unexpected, as there were clear and significant differences between the tumour microenvironments of early-stage and locally advanced or metastatic biliary tract cancers, as seen in Figure 36.

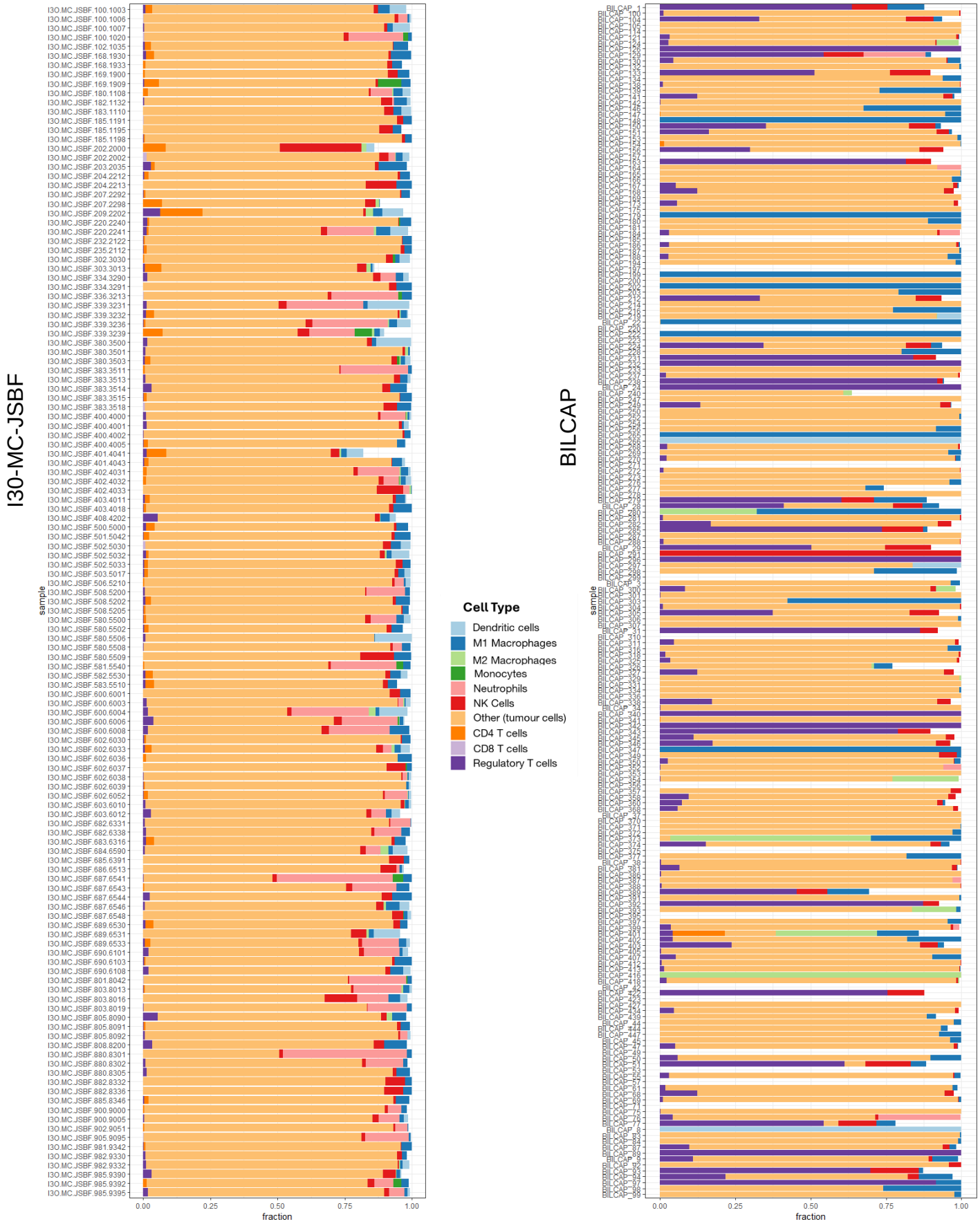


Figure 36: These plots show the different cells making up the tumour microenvironment in locally advanced or metastatic cancers (from I3O-MC-JSBF, left) versus early-stage biliary tract cancer (from BILCAP, right)

What was immediately apparent in a comparison of the plots is that the tumour microenvironments were significantly different between early-stage and locally advanced or metastatic tumours. Some of the differences are due to the quality of the data; as mentioned earlier, the BILCAP samples were older FFPE samples and the overall tumour content of the samples is lower than that in samples from the I3O-MC-JSBF clinical trial. Some samples labelled as ‘tumour’ in the original histopathological analysis may have had a high content of necrotic tissue or pus, which would have increased the levels of immune cells compared to ‘other’ cells, which is the label that quantiseqr applies to tumour cells in the sample. However, even with this taken into account, and extreme levels of T cell infiltration accounted for, the differences between early-stage and locally advanced or metastatic cancers are marked.

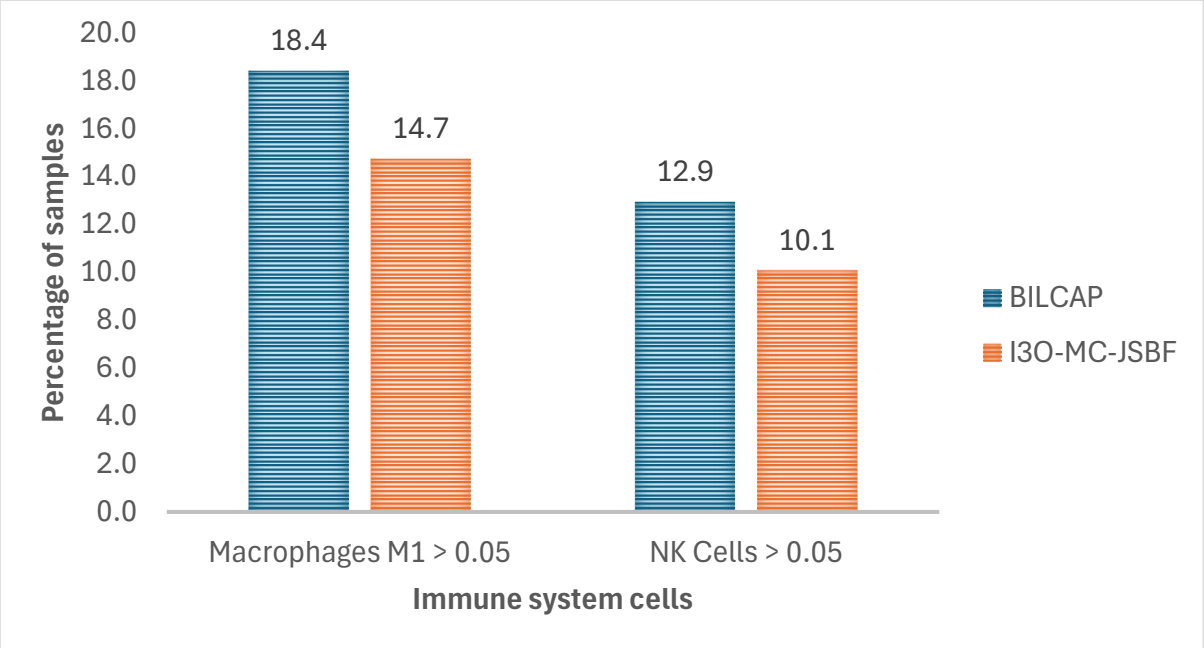


Figure 37: This graph shows that there is little difference between the proportion of M1 macrophages and NK cells in tumour samples from patients from the BILCAP and I30-MC-JSBF clinical trials.

Some of the proportions of immune system cells were the same between early-stage and locally advanced or metastatic biliary tract cancers; as seen in Figure 37, the proportion of M1 macrophages and NK cells were very similar between the two groups of patients.

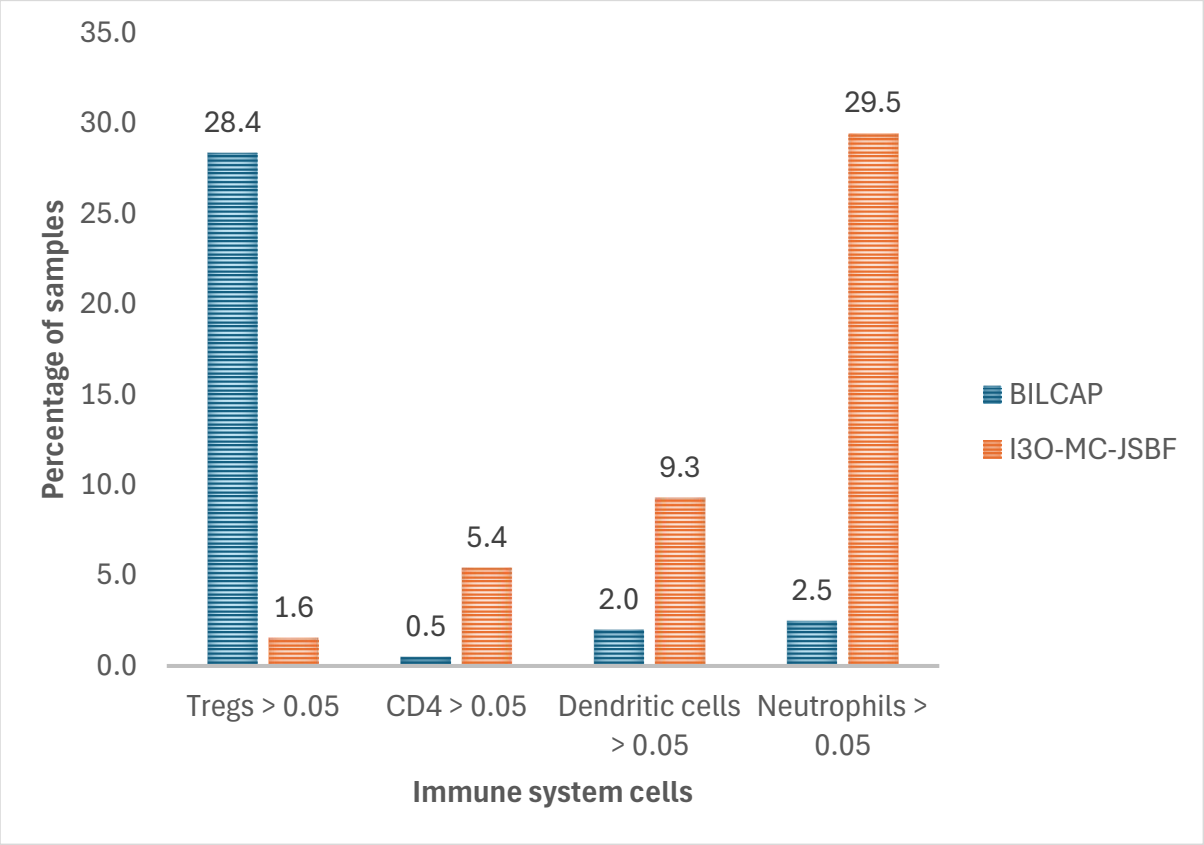


Figure 38: This graph shows that there are significant differences between the proportions of regulatory T-cells, CD4+ T-cells, dendritic cells and neutrophils from tumour samples from patients from the BILCAP and I30-MC-JSBF clinical trials.

However, the proportion of other immune system cells showed significant differences as seen in Figure 38. Samples from the BILCAP clinical trial, or early-stage patients, had a significantly higher proportion of regulatory T-cells than samples from I30-MC-JSBF. Likewise, samples from I30-MC-JSBF with locally advanced or metastatic patients had higher proportions of CD4+ T cells, dendritic cells and neutrophils than samples from BILCAP.

Chapter 4: Discussion

4.1: Other projects

4.1.1: Analysis of patients receiving an anti-HER2 antibody (zanidatamab) and the association between genomic alterations in *ERBB2* and response to treatment

4.1.1.1: Rationale and background

Another project that was set up during the course of this PhD was a collaboration with a pharmaceutical company to investigate the genomic landscape of patients receiving a novel bispecific anti-HER2 antibody called zanidatamab (117, 118). Zanidatamab (ZW25) is a HER2 bispecific monoclonal antibody which binds to the extracellular domains of HER2 also targeted by trastuzumab and pertuzumab (119).

An initial phase I study (NCT02892123) demonstrated that zanidatamab was well tolerated, with an overall response rate of 40% when compared to trastuzumab and pertuzumab (119), and produced durable responses, even in biliary tract cancer patients who had received several prior lines of therapy (117). Phase II clinical trials using zanidatamab in biliary tract cancer, including HERIZON-BTC-01 (NCT04466891), have shown that zanidatamab provided meaningful clinical benefit to patients with HER2 amplified unresectable, locally advanced or metastatic biliary tract cancer, with an acceptable safety profile (46).

Determining whether a tumour has abnormal HER2 expression is critical for selecting which patients should receive zanidatamab, to ensure that patients who may respond receive treatment and that patients who are unlikely to respond are not treated unnecessarily. Currently, tumour HER2 positivity is determined by immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH); however, there may be alternative methods of determining whether a patient's tumour has altered HER2 expression. This project aims to determine which of these methods is

best in predicting response to zanidatamab, as it may differ from other HER2 directed therapies.

Potential alternatives to IHC and FISH for determining HER2 positivity are being considered in other HER2 dependent cancers, such as gastric cancer and breast cancer. These alternatives involve examining genomic changes in the tumours; including investigating changes in *ERBB2* copy number (120), mutations in the *ERBB2* gene (121) and expression of HER2 (122). Investigating these genomic changes also allows for the examination of co-expressed alterations which can affect the impact of HER2 inhibition. This includes novel mutations which may be acquired during treatment, for example, co-occurring mutations in genes such as *PIK3CA* (123) and *CCNE1* (124).

4.1.1.2: Project plan and design

The initial plan for this project was proposed to Zymeworks (USA), who initially developed and manufactured zanidatamab. The project as initially proposed would have collected fresh frozen tissue samples from both tumour and normal tissues from patients with metastatic biliary tract cancer, prior to treatment with zanidatamab, and recruited patients would be HER2 positive according IHC +/- FISH, as per the current standard of care (Figure 39). Each patient investigated on this study must have received at least 3 cycles of zanidatamab and have had at least one tumour measurement imaging performed after starting treatment. All patients must also be eligible to receive zanidatamab, and able to give informed consent to have their tissues used in this study. Patients must also have information on progression-free survival and overall survival for their information to be used in this study.

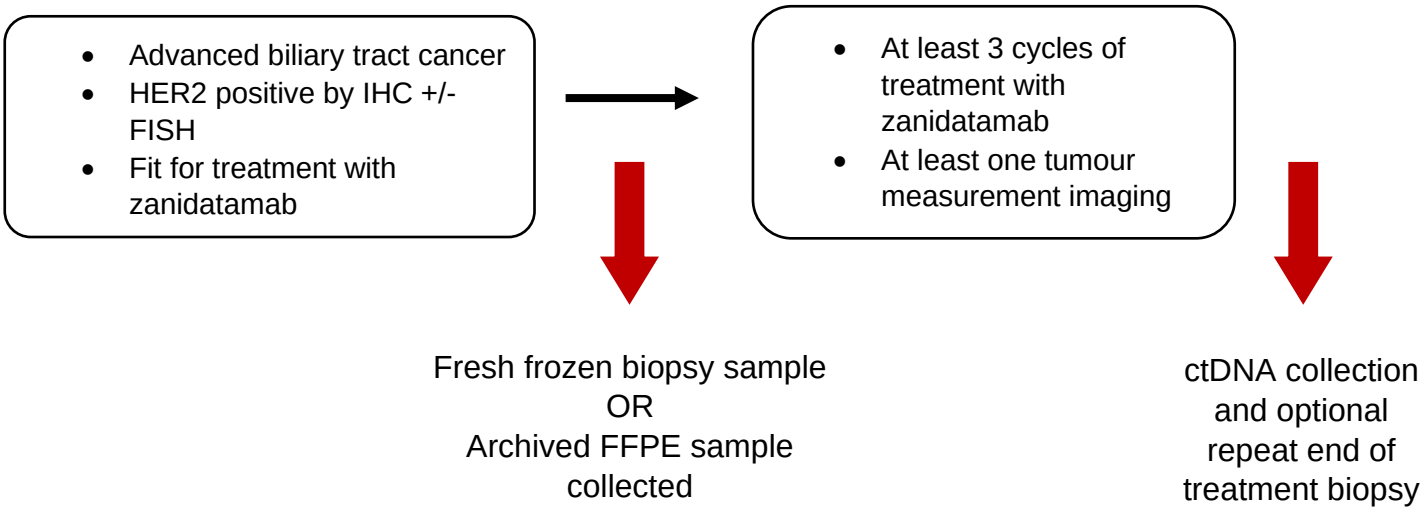


Figure 39: A diagram showing the initial proposed plan for the translational zanidatamab study proposed to Zymeworks.

The normal and tumour tissue samples collected at the start of treatment would be processed for RNA-seq and whole exome sequencing (WES). RNA-seq analysis would allow for the analysis of *ERBB2* expression and the detection of any fusion *ERBB2* genes, while whole exome sequencing would allow for the discovery of mutations in *ERBB2*, as well as alterations in copy number which might also affect HER2. Both RNA-seq and WES would also show any mutations or amplifications in other genes, which could have an association with whether patients respond to treatment with zanidatamab. Levels of mutations, gene expression and copy number alterations will then be correlated with best overall response, progression-free survival and overall survival, to see if any specific genomic alterations correlate with specific clinical outcomes (Figure 39). Using RNA-Seq and WES will not determine regulatory variants (which would require bisulphite sequencing) and genomic translations not present as gene fusions detectable using RNA-Seq, but changes in expression and mutations most likely to be present related to *ERBB2* will be detected by RNA-Seq and WES.

Patients receiving zanidatamab treatment in the UK on the SAFIR-ABC-10 study and on compassionate use programmes will be approached and consented for academic investigation as part of the extant UCL Biobank Ethics approval. All patients consented for this study will receive and have the opportunity to read a patient information sheet specific to this study before informed consent is obtained.

As this is not a clinical trial to investigate response to treatment, there is no specific minimum sample size required. However, as more patient samples are analysed the conclusions of the study will become more accurate and reliable. This study will start with samples from 10 patients (20 samples total, with a tumour and normal tissue sample from each patient).

The collection and analysis of samples will be split into two parts.

The project will begin with the recruitment of patients and the collection of tissue and blood samples, and data on outcome and response to treatment with zanidatamab will also begin to be collected at this time. Guardant will be responsible for analysis of the collected ctDNA at the beginning and end of treatment. The tissue samples (either fresh frozen or FFPE) will initially be analysed using WES by an external comany, and this data will be analysed at UCL using their HPC.

In the second part of the study, the tissue samples will be analysed using RNA-seq, and any available end of treatment biopsy samples will be sent for WES and RNA-seq.

Throughout the duration of this study, further whole blood samples will be taken and stored for potential future circulating tumour RNA analysis.

4.1.1.3: Initial project discussions with Zymeworks and Jazz Pharmaceuticals

This project was initially developed with Zymeworks (USA), who were the pharmaceutical company responsible for developing and manufacturing zanidatamab. The proposal was made as an investigator-initiated proposal for research to the company, and the research plan was presented to the research team at Zymeworks, who were interested in the proposal and were keen on further development.

After developing the outline of the project, the next part was to determine the funding for the project. Analysis of the data would be carried out at UCL, so the costs (aside from paying a relatively small amount for long-term archival data storage of the final analysis data) would be minimal. However, the Biliary Tract Cancer Research team at UCL does not have any formal laboratory space or a dedicated wet-lab research team, so an external group would have to be hired to extract the DNA and RNA from the samples, prepare the libraries for sequencing and sequence the RNA and DNA. Ideally, this would all be done in-house at the UCL Cancer Institute, but at the time this project was being developed the UCL team did not have a large amount of experience extracting DNA and RNA from FFPE samples and did not have a formal protocol for this. Instead, external companies such as Novogene and Azenta were approached and provided quotes for the extraction, preparation and sequencing of the samples, which informed how much budget would be required to carry out the study.

4.1.1.4: Current status of the project

The manufacturing and development of zanidatamab have now been taken over from Zymeworks by a different company, Jazz Pharmaceuticals (USA). Fortunately, the team has also transferred over, and so planning for this project continues. A provisional budget has been agreed spanning two financial years to provide for both DNA and RNA sequencing, and the project trial protocol, informed consent form (ICF) and an application for formal ethical approval are all in progress.

4.1.2: Analysis of the genomic landscape and tumour evolution of a patient with gallbladder cancer through the PEACE project

4.1.2.1: Project design and initiation

The final project set up during the course of this PhD was a project to investigate the genomic landscape and tumour evolution in a patient with biliary tract cancer. To date, there has not yet been a study of a patient with biliary tract cancer where multiple areas of the primary cancer and multiple metastatic deposits from a single patient were all analysed together, allowing for the analysis of tumour evolution and differences between the primary and metastatic sites. Similar studies have been done in lung cancer (125) and melanoma (126), so it would be interesting to compare and see similar phylogenetic mapping in biliary tract cancers.

However, biliary tract cancers are relatively uncommon, so it would be difficult to find a patient where it would be possible to take samples for sequencing from multiple metastatic sites. Fortunately, a study called PEACE (Posthumous Evaluation of Advanced Cancer Environment) was being run at UCL (clinical trial number NCT03004755) (127). This study recruited patients with different types of metastatic cancer and asked them to consent to having samples of their cancer taken after their death. When a patient who has consented to this trial dies, they undergo an autopsy where tissue samples are taken from multiple regions of the primary tumour, from metastatic deposits identified on cross-sectional imaging (such as CT scans) performed before death and from normal uninvolved tissue. These tissue samples are stored as fresh frozen and FFPE, and slides from the samples are cut, stained, and digitally photographed. The tissue can then be sent for sequencing for different projects as required.

4.1.2.2: Data collection, plans for sample processing and sequencing

One patient with biliary tract cancer was recruited into PEACE and gave her consent for her tissue samples to be used in future analysis. This patient was a 53-year-old lady with metastatic gallbladder cancer (adenocarcinoma). Initially diagnosed with

the cancer a year before her death, she had stable disease after treatment with 8 cycles of gemcitabine and cisplatin chemotherapy but soon progressed and died around a year after diagnosis. She had a past medical history which included ulcerative colitis, primary sclerosing cholangitis and has previously had a Dukes Stage A T2 colon cancer removed. The autopsy carried out after death took samples of whole bloods from four different areas of the primary gallbladder cancer and from metastatic deposits from the liver and pylorus. All of these samples were stored both as fresh-frozen and FFPE samples, and stained H&E slides were produced. The FFPE tumour block from the previous colon cancer was also available and was stored for use in this project.

The plan for analysis of these tissues will start with analysis of the H&E stained slides from each of the sampled regions and the colon cancer, which will be carried out by a histopathology colleague at UCL. This will allow for the identification of the best tissue blocks to send for RNA and DNA extraction, as there would be little use in sending blocks which contain relatively little tumour content. If tumour content is an issue throughout the blocks, it may also be possible to perform microdissection on the blocks to ensure that the tissue sent for DNA and RNA extraction has as high a tumour content as possible. Normal tissue will also be sent from the gallbladder for matched RNA analysis, as well as normal tissue either from the gallbladder or from an uninvolved organ (such as lung) for matched DNA analysis. Ordinarily whole blood would be used for this, but very little was available after the autopsy, and it may not be sufficient for analysis.

Once the best samples for analysis have been identified, the next steps will be to send the samples for DNA and RNA extraction. DNA and RNA extraction from FFPE samples continue to improve in reliability, but it is likely that the fresh-frozen samples will be sent, together with the FFPE colon cancer sample. The current plan is that the extraction and sequencing will be done in-house by the UCL genomics team, and that DNA and RNA extraction will be performed on all of the tumour samples for WES and RNA-seq. Normal gallbladder tissue will also undergo RNA extraction and RNA-seq, while other uninvolved tissue will undergo WES. This will allow for the greatest amount of information to be obtained from the least amount of sequenced

tissue. Matched tumour, normal and metastatic deposit WES will allow for the analysis of mutations and copy number alterations, while RNA-seq will allow for the analysis of differential gene expression, tumour microenvironment analysis and gene fusion identification. Together, this information will allow for the use of tools such as MAPping SubClonal Events (MAPSCE) (128) for phylogenetic mapping, which has not yet been performed on samples from patients with biliary tract cancer.

At present, the patient samples have been collected, identified and are ready for histopathology analysis. Once the best samples for sequencing have been identified, the samples will be sent for DNA andRNA extraction and sequencing and the bioinformatics analysis will begin.

4.2: Summary and novelty of the findings

This work focuses on the analysis of genomic data from the BILCAP clinical trial and is the first piece of work to demonstrate the link between specific genomic alterations and patient outcomes in patients with biliary tract cancers.

In particular, the results showing that *FGFR2* fusions have no effect on patient outcomes is particularly significant. This was first shown in the smaller BILCAP dataset, but even in the larger datasets there is strong evidence that *FGFR2* fusions have no effect. The reason why this specific negative result is so significant is that it is particularly important for the development of *FGFR2* fusion targeting drugs and in applications for licensing. During the licensing of pemigatnib by NICE, one of the main concerns which almost led to the drug not being approved was that there was no evidence that the improved outcomes seen after treatment with pemigatinib were purely due to the drug (31). Smaller studies had potentially indicated that patients with *FGF2* fusions actually had better outcomes (129), so the results of this study indicating that there was actually no difference is very important to show that any improved outcomes seen in clinical trials are purely due to the treatment, and not the underlying genomic changes. While this is most evident in *FGFR2* fusion positive

patients give the issues raised in licencing, the results of this project showing that the vast majority of targetable alterations do not affect survival will have wider implications for other targeted treatments, in that it shows that any positive effects seen on outcomes with the use of those drugs is purely due to the new treatments.

As well as directly answering questions raised about the efficacy of *FGFR2* inhibitors and whether or not improvements in outcomes were due to the medication or the alterations themselves, the results of this project have also has raised potential targets for treatments and prognostic testing.

The results demonstrating that *EGFR* amplification was significantly linked to poor outcomes was especially unexpected, and further research is needed into investigating the possible mechanisms behind the poorer outcomes. *EGFR* targeted treatments have previously been investigated as potential treatment options, including studies using panitumumab and GEMOX (gemcitabine and oxaliplatin) (130), gemcitabine and oxaliplatin with or without cetuximab (52) or gemcitabine and cetuximab (131). However, none of these studies showed an improvement in outcomes with EGFR targeted treatment, but targeting *EGFR* amplified patients rather than *EGFR* mutated patients, or combining *EGFR* targeted treatments with chemo-immunotherapy may show more positive results in clinical trials.

Even if *EGFR* amplification does not prove to be amenable for targeting for treatment, it may prove to be an important biomarker. At present, there are no proven biomarkers in biliary tract cancers to illustrate which cancers are more aggressive and likely to recur quickly after resection. Surgery for the resection of biliary tract cancer is major surgery, with prolonged recovery times even with the best surgical techniques in use. It is becoming more and more recognised that some patients will develop a recurrence of their cancer very quickly, within months to a year, despite optimal surgery (132). Of particular note, it is interesting that *EGFR* amplification remains a significant risk factor with a copy number of 4, which is a level on amplification that is likely too low to be measured using other methods of RNA and protein expression such as IHC. It may be that *EGFR* amplification,

together with other factors could be used as a method of identifying those patients at risk of relapsing very quickly after surgery. A test could be developed that could prevent these patients from undergoing major and ultimately futile surgery, and possibly encourage the use of other systemic treatments at an earlier stage in these patients to improve their overall outcomes.

The results of the tumour microenvironment analysis and comparison between data from the BILCAP and I30-MC-JSBF clinical trials possibly have major implications for the neoadjuvant and adjuvant treatment of biliary tract cancer. Immunotherapy has been shown to be effective in locally advanced or metastatic biliary tract cancer, through the results of the TOPAZ-1 clinical trial (26). The results of this tumour microenvironment analysis may go some way in explaining why this is the case; the samples from patients with locally advanced or metastatic biliary tract cancers had higher proportions of CD4 T-cells, dendritic cells and neutrophils, all of which point to the tumours being more inflammatory in nature with a ‘hot’ immune environment. Tumours with these types of alterations respond well to immunotherapy, as was seen in TOPAZ-1. However, the early-stage tumours have far fewer pro-inflammatory immune system cells and large numbers of immunosuppressive regulatory T-cells. This implies that the immune environment in these tumours is more ‘cold’, and implies that these tumours may respond far less well to immunotherapy than locally advanced or metastatic biliary tract cancers. The ARTEMIDE-Biliary01 clinical trial is investigating the use of immunotherapy in the adjuvant setting, comparing adjuvant chemotherapy versus chemotherapy and rilvegostomig, a bispecific anti-PD-1 and TIGIT antibody (133); and it will be very interesting to see if the implications of this tumour microenvironment analysis are demonstrated in the results from that trial, and in other similar upcoming trials.

Although the main work on BILCAP is not yet complete, once the full available dataset has been analysed with low-pass WGS, RNA-seq and targeted sequencing, the work will be formally written up for journal publication.

While work on the zantidatamab anti-HER2 study and the PEACE project are ongoing, both projects represent some of the first investigations of their kind into these fields. The use of anti-HER2 targeted treatment in biliary tract cancer is still at a very early stage and not yet in widespread use, but the success seen when using anti-HER2 targeted treatment in biliary tract cancer both in clinical trials and in compassionate access use indicates that this is very likely to become much more widespread in the future. As with all targeted treatment, looking into the mechanisms of both primary and acquired resistance is of utmost importance when trying to ensure that as many patients as possible benefit from this treatment, and the design of the zanidatamab study means that this can be investigated using real-world patient data and tissue, which is the clearest and most accurate method of investigation.

With regards to the PEACE project, to date it does not appear that the intra-tumoural differences in DNA and RNA have been investigated in biliary tract cancer. It also does not appear that there has been any work into mapping the phylogenetic trees of tumours within a single patient with biliary tract cancer, so the results of the study will be both unique and of great interest to those interested in the genomic landscape of biliary tract cancer.

4.3: Limitations and future work

The main issue with this work as currently presented in this thesis is that it is still incomplete, but work on the zanidatamab and PEACE projects are ongoing. These projects will form the basis of an application for a Clinician Scientist position in the upcoming months.

4.3.1 BILCAP

Currently, the main plan is to obtain more data from the BILCAP cohort. A total of 447 patients were recruited to BILCAP, and around 300 or so patients have samples

which can be used for genomic analysis, which means that there are around 100 more patients whose samples need to be analysed. The proportion of subtypes of the analysed samples is also very uneven, with relatively few perihilar cholangiocarcinomas, so more perihilar cholangiocarcinoma samples need to be extracted and sequenced for analysis. Unfortunately, there were significant issues with the Qiagen kit being used for targeted sequencing. This issue has now been resolved, but the targeted sequencing data was not available during the course of this PhD.

The dataset as a whole needs to be more complete before the genomic analysis of BILCAP can be considered complete, and this would be clear to any reviewer of any papers from the current dataset. However, new data has very recently become available, and more data is coming as more and more of the samples are sequenced, and the results will be formally published once the dataset is complete.

A major and ongoing issue with the BILCAP dataset is the quality of the data. BILCAP recruited patients from 2006 to 2014 (20), so the newest tissue samples are now over 10 years old. When the trial was being designed no-one could have been aware of plans to perform next-generation sequencing on the tissue samples, so the tissue samples were stored long-term in paraffin and formalin to best preserve the tissue for future histopathological analysis. However, fixed formalin (FFPE) samples, especially the older samples such as those from the BILCAP clinical trial are notoriously difficult to use for next generation sequencing. Newer DNA and RNA extraction techniques and sequencing are making it easier to obtain data from older FFPE samples, an improved bioinformatics techniques make it easier to filter out the ‘noise’ present in older samples, and to compensate for the sequencing errors seen after formalin fixation. However, issues with low tumour content in some of the samples remains an issue. Newer clinical trials avoid this issue by using fresh or fresh-frozen tissue for genomic analysis, but in order to have the rich historical clinical and outcome data from BILCAP using the less-than-ideal FFPE tissue for genomic analysis is unavoidable. Issues and sequencing errors from using FFPE samples could also be ameliorated by using matched tumour and normal DNA,

allowing for improved filtration of patient germline genomic alterations as well as the artefactual issues caused by formalin fixation. This tissue to carry this out is available, but sequencing matched normal tissue would have doubled the cost of the sequencing and limited the overall number of samples from different patients. It is possible that funding for sequencing matched normal tissues may be possible in the future, and this will improve the accuracy of detecting genomic alterations on an individual patient level.

There are also more experiments that can be run on the current BILCAP data and can be expanded once more of the data is available. Differential gene analysis and a comparison of genomic differences and outcomes between patients who did or did not receive adjuvant chemotherapy was initially due to be carried out by a fellow collaborative laboratory. However, this work will now be carried out as part of this larger project once the genomic data from all of the different anatomical subtypes of biliary tract cancer becomes available. RNAseq data will be used for the differential gene analysis, and the targeted seq and low-pass WGS data will be used to investigate mutations and copy number differences between patients who received adjuvant capecitabine and relapsed and those who received adjuvant capecitabine and did not relapse. It will be interesting to see if there are any genomic changes or patterns of genomic changes that are associated with being more or less likely to have a recurring or metastatic tumour after treatment with adjuvant capecitabine.

Analysis of microsatellite instability (MSI), an important prognostic factor for response to immunotherapy (134) can also be carried out on non-matched tumour DNA using the MSIsensor-pro tool, which requires targeted sequenced DNA (135). Once the targeted sequencing data has been analysed for more samples, it will be possible to analyse the effects of MSI on relapse-free and overall survival without the receipt of immunotherapy confounding the results. The targeted sequencing DNA will also be used to analyse the effect of co-occurring mutations on the risk of relapse and overall survival. At present data from this form of analysis would be very limited due to how few samples have targeted sequencing data available for mutation

analysis, but once more of the targeted sequencing data it available it will be possible to run this analysis.

The BILCAP samples also include histopathology slides, and the analysis of these is also forming part of the larger overall project to analyse data from the BILCAP clinical trial. While TME analysis is possible using RNA-seq data, as seen above, spatial transcriptomics data would provide more detailed information on a subset of samples. Plans are being made to collaborate with another laboratory team at UCL to use GeoMx (136) to analyse spatial transcriptomics and the TME in far more detail than what is possible with RNA-seq deconvolution. These slides may also form parts of collaborative projects with even more research groups, including a team in Thailand looking to compare fluke-associated cholangiocarcinoma with non-fluke-associated cholangiocarcinoma.

4.3.2: I3O-MC-JSBF

The data from I3O-MC-JSBF also requires further analysis. The pipelines built during this project are for the analysis of DNA sequencing data, but as the I3O-MC-JSBF data is RNA-seq data new pipelines are being built. These will allow for the analysis of mutations, fusions and copy number from the I3O-MC-JSBF data as an analysis into biliary tract cancers that are locally advanced or metastatic at presentation. This data will also be used to compare the genomic landscape of early-stage biliary tract cancer with biliary tract cancers that are locally advanced or metastatic at clinical presentation. The TME analysis already performed is the beginning of this, but this will be extended to also look at mutations, copy number alterations and fusions once the I3O-MC-JSBF data has been processed. RNA-seq data from the STAMP and JCOG1202-ASCOT clinical trials, both of which are early-stage clinical trials of adjuvant treatment in biliary tract cancer, will also shortly become available and will be focus of future translational work. A comparison between the TME of samples from these trials and those from I3O-MC-JSBF will show if the differences seen are due to any issues with the BILCAP trial samples themselves, or if they represent a true trend in differences between the TME of early-stage and locally advanced or metastatic tumours.

4.4: Conclusion

This work represents several novel areas of work in the field of biliary tract cancer. The current results of the work looking into the translational bioinformatics of the BILCAP clinical trial represent a unique opportunity to see how genomic alterations affect patient outcome in a patient set who were unable to receive any targeted treatment, meaning that any differences in outcome must be due to the alterations themselves. This has already proven that FGFR2 fusions in and of themselves do not affect outcomes, meaning that the improved outcomes seen in patients given FGFR2 inhibitors much be due to the treatment, which is groundbreaking and demonstrates the value of these targeted treatments. The indication that EGFR amplification reduced both relapse-free and overall survival indicates that it may be a possible measurable prognostic marker, and it may also a potential target for future treatment. As more of the BILCAP data becomes available and is analysed, more information that could lead to the development of future studies and clinical trials will become available. The other projects on zanidatamab and PEACE will also lead to increased knowledge of the resistance mechanisms of targeted treatment and information on the development and spread of biliary tract cancer, so this work as a whole will hopefully lead to further future improvements in the outcomes of patients with biliary tract cancer.

References

1. Banales JM, Marin JJG, Lamarca A, Rodrigues PM, Khan SA, Roberts LR, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol [Internet]. Springer US; 2020; 17(9):557–88. Available from: http://dx.doi.org/10.1038/s41575-020-0310-z.

2. Crolley VE, Bridgewater J. Futibatinib for the Treatment of Previously Treated, Unresectable, Locally Advanced or Metastatic Intrahepatic Cholangiocarcinoma with Fibroblast Growth Factor Receptor 2 (FGFR2) Gene Fusions or Other Rearrangements. touchREVIEWS in Oncology & Haematology [Internet]. 2024 [cited 2024 Aug 8]; 20(1):12–9. Available from: https://touchoncology.com/gastrointestinal-cancers/journal-articles/futibatinib-for-the-treatment-of-previously-treated-unresectable-locally-advanced-or-metastatic-intrahepatic-cholangiocarcinoma-with-fibroblast-growth-factor-receptor-2-fgfr2-gene-fusions-or-other/#article-information.

3. Moeini A, Haber PK, Sia D. Cell of origin in biliary tract cancers and clinical implications. JHEP Reports [Internet]. 2021; 3(2):100226. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2589555921000021.

4. Clements O, Eliahoo J, Kim JU, Taylor-Robinson SD, Khan SA. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: A systematic review and meta-analysis. J Hepatol. 2020; 72(1):95–103.

5. Petrick JL, Yang B, Altekruse SF, Dyke AL Van, Koshiol J, Graubard BI, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: A population-based study in SEER-Medicare. PLoS One. Public Library of Science; 2017; 12(10).

6. Valle JW, Kelley RK, Nervi B, Oh DY, Zhu AX. Biliary tract cancer. The Lancet [Internet]. Elsevier Ltd; 2021; 397(10272):428–44. Available from: http://dx.doi.org/10.1016/S0140-6736(21)00153-7.

7. Qian MB, Chen YD, Liang S, Yang GJ, Zhou XN. The global epidemiology of clonorchiasis and its relation with cholangiocarcinoma. Infectious Diseases of Poverty. BioMed Central Ltd.; 2012.

8. Patel T. Increasing incidence and mortality of primary intrahepatic cholangiocarcinoma in the United States. Hepatology. 2001; 33(6):1353–7.

9. Florio AA, Ferlay J, Znaor A, Ruggieri D, Alvarez CS, Laversanne M, et al. Global trends in intrahepatic and extrahepatic cholangiocarcinoma incidence from 1993 to 2012. Cancer. 2020; 126(11):2666–78.

10. Saha SK, Zhu AX, Fuchs CS, Brooks GA. Forty-Year Trends in Cholangiocarcinoma Incidence in the U.S.: Intrahepatic Disease on the Rise. Oncologist. 2016; 21(5):594–9.

11. West J, Wood H, Logan RFA, Quinn M, Aithal GP. Trends in the incidence of primary liver and biliary tract cancers in England and Wales 1971-2001. Br J Cancer. 2006; 94(11):1751–8.

12. Taylor-Robinson SD, Toledano MB, Arora S, Keegan TJ, Hargreaves S, Beck A, et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968-1998. Gut. 2001; 48(6):816–20.

13. Andersen JB, Spee B, Blechacz BR, Avital I, Komuta M, Barbour A, et al. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology. W.B. Saunders; 2012; 142(4).

14. Chan-On W, Nairismägi ML, Ong CK, Lim WK, Dima S, Pairojkul C, et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. Nat Genet. 2013; 45(12):1474–8.

15. Jusakul A, Cutcutache I, Yong CH, Lim JQ, Huang MN, Padmanabhan N, et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. Cancer Discov. 2017; 7(10):1116–35.

16. Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A, Kato M, et al. Genomic spectra of biliary tract cancer. Nat Genet [Internet]. Nature Publishing Group; 2015; 47(9):1003–10. Available from: http://dx.doi.org/10.1038/ng.3375.

17. Pandey A, Stawiski EW, Durinck S, Gowda H, Goldstein LD, Barbhuiya MA, et al. Integrated genomic analysis reveals mutated ELF3 as a potential gallbladder cancer vaccine candidate. Nat Commun [Internet]. Springer US; 2020; 11(1):1–13. Available from: http://dx.doi.org/10.1038/s41467-020-17880-4.

18. AMMF. Rethink Liver Cancer White Paper. [date unknown].

19. Izquierdo-Sanchez L, Lamarca A, Casta A La, Buettner S, Utpatel K, Klümpen HJ, et al. Cholangiocarcinoma landscape in Europe: Diagnostic, prognostic and therapeutic insights from the ENSCCA Registry. J Hepatol. Elsevier B.V.; 2022; 76(5):1109–21.

20. Primrose JN, Neoptolemos J, Palmer DH, Malik HZ, Prasad R, Mirza D, et al. Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study. Lancet Oncol. 2019; 20(5):663–73.

21. Nakachi K, Konishi M, Ikeda M, Mizusawa J, Eba J, Okusaka T, et al. A randomized Phase III trial of adjuvant S-1 therapy vs. observation alone in resected biliary tract cancer: Japan Clinical Oncology Group Study (JCOG1202, ASCOT). Jpn J Clin Oncol. Oxford University Press; 2018; 48(4):392–5.

22. Ebata T, Hirano S, Konishi M, Uesaka K, Tsuchiya Y, Ohtsuka M, et al. Randomized clinical trial of adjuvant gemcitabine chemotherapy versus

observation in resected bile duct cancer. British Journal of Surgery. John Wiley and Sons Ltd; 2018; 105(3):192–202.

23. Edeline J, Benabdelghani M, Bertaut A, Watelet J, Hammel P, Joly J-P, et al. Gemcitabine and Oxaliplatin Chemotherapy or Surveillance in Resected Biliary Tract Cancer (PRODIGE 12-ACCORD 18-UNICANCER GI): A Randomized Phase III Study. J Clin Oncol [Internet]. 2019; 37(8):658–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30707660.

24. Edeline J, Hirano S, Bertaut A, Konishi M, Benabdelghani M, Uesaka K, et al. Individual patient data meta-analysis of adjuvant gemcitabine-based chemotherapy for biliary tract cancer: combined analysis of the BCAT and PRODIGE-12 studies. Eur J Cancer. Elsevier Ltd; 2022; 164:80–7.

25. Jeong H, Kim KP, Jeong JH, Hwang DW, Lee JH, Kim KH, et al. Adjuvant gemcitabine plus cisplatin versus capecitabine in node-positive extrahepatic cholangiocarcinoma: The STAMP randomized trial. Hepatology. Wolters Kluwer Medknow Publications; 2023; 77(5):1540–9.

26. Oh D-Y, Ruth He A, Qin S, Chen L-T, Okusaka T, Vogel A, et al. Durvalumab plus Gemcitabine and Cisplatin in Advanced Biliary Tract Cancer. NEJM Evidence. Massachusetts Medical Society; 2022; 1(8).

27. Kelley RK, Ueno M, Yoo C, Finn RS, Furuse J, Ren Z, et al. Pembrolizumab in combination with gemcitabine and cisplatin compared with gemcitabine and cisplatin alone for patients with advanced biliary tract cancer (KEYNOTE-966): a randomised, double-blind, placebo-controlled, phase 3 trial. The Lancet. Elsevier B.V.; 2023; 401(10391):1853–65.

28. Valle J, Wasan H, Palmer DH, Cunningham D, Anthoney A, Maraveyas A, et al. Cisplatin plus Gemcitabine versus Gemcitabine for Biliary Tract Cancer. New England Journal of Medicine [Internet]. 2010; 362(14):1273–81. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0908721.

29. Chiang NJ, Chen MH, Yang SH, Hsu C, Yen CJ, Tsou HH, et al. Multicentre, phase II study of gemcitabine and S-1 in patients with advanced biliary tract cancer: TG1308 study. Liver International. Blackwell Publishing Ltd; 2020; 40(10):2535–43.

30. Lamarca A, Palmer DH, Wasan HS, Ross PJ, Ma YT, Arora A, et al. Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. Lancet Oncol [Internet]. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.; 2021; 22(5):690–701. Available from: http://dx.doi.org/10.1016/S1470-2045(21)00027-9.

31. NICE. Pemigatinib for treating relapsed or refractory advanced cholangiocarcinoma with FGFR2 fusion or rearrangement [Internet]. 2021. Available from: www.nice.org.uk/guidance/ta722.

32. Abou-Alfa GK, Sahai V, Hollebecque A, Vaccaro G, Melisi D, Al-Rajabi R, et al. Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. Lancet Oncol. 2020; 21(5):671–84.

33. Vogel A, Sahai V, Hollebecque A, Vaccaro G, Melisi D, Al-Rajabi R, et al. Pemigatinib for previously treated locally advanced or metastatic cholangiocarcinoma: Final results from FIGHT-202. Annals of Oncology. Elsevier BV; 2022; 33:S379.

34. Syed YY. Futibatinib: First Approval. Drugs. Adis; 2022; 82(18):1737–43.

35. European Medicines Agency (EMA). Futibatinib: Summary of Product Characteristics (SPC) [Internet]. 2023 [cited 2023 Nov 30]. Available from: https://www.ema.europa.eu/en/documents/product-information/lytgobi-epar-product-information_en.pdf.

36. Goyal L, Meric-Bernstam F, Hollebecque A, Valle JW, Morizane C, Karasic TB, et al. Futibatinib for FGFR2 -Rearranged Intrahepatic Cholangiocarcinoma. New England Journal of Medicine. Massachusetts Medical Society; 2023; 388(3):228–39.

37. Goyal L, Meric-Bernstam F, Hollebecque A, Morizane C, Valle JW, Karasic TB, et al. Updated results of the FOENIX-CCA2 trial: Efficacy and safety of futibatinib in intrahepatic cholangiocarcinoma (iCCA) harboring FGFR2 fusions/rearrangements. . Journal of Clinical Oncology. 2022; 40(16_suppl):4009–4009.

38. Goyal L, Kongpetch S, Crolley VE, Bridgewater J. Targeting FGFR inhibition in cholangiocarcinoma. Cancer Treat Rev [Internet]. Elsevier Ltd; 2021; 95(February):102170. Available from: https://doi.org/10.1016/j.ctrv.2021.102170.

39. Abou-Alfa GK, Macarulla T, Javle MM, Kelley RK, Lubner SJ, Adeva J, et al. Ivosidenib in IDH1-mutant, chemotherapy-refractory cholangiocarcinoma (ClarIDHy): a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol. 2020; 21(6):796–807.

40. Zhu AX, Macarulla T, Javle MM, Kelley RK, Lubner SJ, Adeva J, et al. Final Overall Survival Efficacy Results of Ivosidenib for Patients with Advanced Cholangiocarcinoma with IDH1 Mutation: The Phase 3 Randomized Clinical ClarIDHy Trial. JAMA Oncol. American Medical Association; 2021; 7(11):1669–77.

41. Marabelle A, Le DT, Ascierto PA, Giacomo AM Di, Jesus-Acosta A De, Delord J-P, et al. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. J Clin Oncol [Internet]. 2020; 38(1):1–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31682550.

42. Subbiah V, Lassen U, Élez E, Italiano A, Curigliano G, Javle M, et al. Dabrafenib plus trametinib in patients with BRAFV600E-mutated biliary tract cancer (ROAR): a phase 2, open-label, single-arm, multicentre basket trial. Lancet Oncol. Lancet Publishing Group; 2020; 21(9):1234–43.

43. Subbiah V, Kreitman RJ, Wainberg ZA, Gazzah A, Lassen U, Stein A, et al. Dabrafenib plus trametinib in BRAFV600E-mutated rare cancers: the phase 2 ROAR trial. Nat Med. Nature Research; 2023; 29(5):1103–12.

44. Harding JJ, Piha-Paul SA, Shah RH, Cleary JM, Quinn DI, Brana I, et al. Targeting HER2 mutation–positive advanced biliary tract cancers with neratinib: Final results from the phase 2 SUMMIT basket trial. Journal of Clinical Oncology [Internet]. Wolters Kluwer; 2024; 40(16_suppl):4079. Available from: https://doi.org/10.1200/JCO.2022.40.16_suppl.4079.

45. Lee C kun, Chon HJ, Cheon J, Lee MA, Im HS, Jang JS, et al. Trastuzumab plus FOLFOX for HER2-positive biliary tract cancer refractory to gemcitabine and cisplatin: a multi-institutional phase 2 trial of the Korean Cancer Study Group (KCSG-HB19–14). Lancet Gastroenterol Hepatol. Elsevier Ltd; 2023; 8(1):56–65.

46. Harding JJ, Fan J, Oh DY, Choi HJ, Kim JW, Chang HM, et al. Zanidatamab for HER2-amplified, unresectable, locally advanced or metastatic biliary tract cancer (HERIZON-BTC-01): a multicentre, single-arm, phase 2b study. Lancet Oncol. Elsevier Ltd; 2023; 24(7):772–82.

47. Abou-Alfa GK, Borbath I, Goyal L, Lamarca A, Macarulla T, Oh D-Y, et al. PROOF 301: A multicenter, open-label, randomized, phase 3 trial of infigratinib versus gemcitabine plus cisplatin in patients with advanced cholangiocarcinoma with an FGFR2 gene fusion/rearrangement. Journal of Clinical Oncology [Internet]. Wolters Kluwer; 2022; 40(16_suppl):TPS4171–TPS4171. Available from: https://doi.org/10.1200/JCO.2022.40.16_suppl.TPS4171.

48. Abou-Alfa GK, Borbath I, Roychowdhury S, Goyal L, Lamarca A, Macarulla T, et al. PROOF 301: Results of an early discontinued randomized phase 3 trial of the oral FGFR inhibitor infigratinib vs. gemcitabine plus cisplatin in patients with advanced cholangiocarcinoma (CCA) with an FGFR2 gene fusion/rearrangement. Journal of Clinical Oncology [Internet]. 2024; 42(3_suppl):516–516. Available from: https://ascopubs.org/doi/10.1200/JCO.2024.42.3_suppl.516.

49. Féliz L, Asatiani E, Lihou C, Zhen H, Silverman I, Ahnert JR. Abstract A077: FIGHT-207: Phase 2 study of pemigatinib in patients with previously treated, locally advanced/metastatic or unresectable solid tumor malignancies harboring activating fibroblast growth factor receptor (FGFR) gene alterations. Mol Cancer Ther. 2019; 18(12_Supplement):A077–A077.

50. AMMF. AMMF SAFIR ABC-10 [Internet]. [cited 2024 Sep 2]. Available from: https://ammf.org.uk/safir-abc-10/.

51. Malka D, Borbath I, Bridgewater J, Lopes A, Couch D, Dick J, et al. SAFIR-ABC-10 – Precision Medicine Clinical Trial Protocol. 2024.

52. Malka D, Cervera P, Foulon S, Trarbach T, la Fouchardière C de, Boucher E, et al. Gemcitabine and oxaliplatin with or without cetuximab in advanced biliary-tract cancer (BINGO): A randomised, open-label, non-comparative phase 2 trial. Lancet Oncol. Lancet Publishing Group; 2014; 15(8):819–28.

53. Massard C, Michiels S, Ferté C, Deley MC Le, Lacroix L, Hollebecque A, et al. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: Results of the MOSCATO 01 trial. Cancer Discov. American Association for Cancer Research Inc.; 2017; 7(6):586–95.

54. Verlingue L, Malka D, Allorant A, Massard C, Ferté C, Lacroix L, et al. Precision medicine for patients with advanced biliary tract cancers: An effective strategy within the prospective MOSCATO-01 trial. Eur J Cancer. Elsevier Ltd; 2017; 87:122–30.

55. Bekaii-Saab TS, Valle JW, Cutsem E Van, Rimassa L, Furuse J, Ioka T, et al. FIGHT-302: First-line pemigatinib vs gemcitabine plus cisplatin for advanced cholangiocarcinoma with FGFR2 rearrangements. Future Oncology. Future Medicine Ltd.; 2020.

56. Borad MJ, Bridgewater JA, Morizane C, Shroff RT, Oh D-Y, Moehler MH, et al. A phase III study of futibatinib (TAS-120) versus gemcitabine-cisplatin (gem-cis) chemotherapy as first-line (1L) treatment for patients (pts) with advanced (adv) cholangiocarcinoma (CCA) harboring fibroblast growth factor receptor 2 (FGFR2) gene rearrangements (FOENIX-CCA3). Journal of Clinical Oncology. 2020; 38(4_suppl):TPS600–TPS600.

57. Macarulla T, Mizuno T, Brandi G, Li J, Chen M-H, Kang JH, et al. FOENIX-CCA4: A phase 2 study of futibatinib 20 mg and 16 mg in patients with advanced cholangiocarcinoma (CCA) and fibroblast growth factor receptor 2 (*FGFR2*) fusions/rearrangements. Journal of Clinical Oncology. 2024; 42(3_suppl):TPS572–TPS572.

58. Stein A, Arnold D, Bridgewater J, Goldstein D, Jensen LH, Klümpen HJ, et al. Adjuvant chemotherapy with gemcitabine and cisplatin compared to observation after curative intent resection of cholangiocarcinoma and muscle invasive gallbladder carcinoma (ACTICCA-1 trial) - a randomized, multidisciplinary, multinational phase III trial. BMC Cancer. BioMed Central Ltd.; 2015; 15(1).

59. Genomics England. Cancer in the 100,000 Genomes Project [Internet]. [cited 2024 Mar 21]. Available from: https://www.genomicsengland.co.uk/initiatives/100000-genomes-project/cancer.

60. European Genome-Phenome Archive. Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma -

European Genome-Phenome Archive [Internet]. [cited 2024 Mar 26]. Available from: https://ega-archive.org/studies/EGAS00001001653.

61. Goeppert B, Toth R, Singer S, Albrecht T, Lipka DB, Lutsik P, et al. Integrative Analysis Defines Distinct Prognostic Subgroups of Intrahepatic Cholangiocarcinoma. 2091 Hepatology. 2019; 69(5).

62. Verlingue L, Hollebecque A, Boige V, Ducreux M, Malka D, Ferté C. Matching genomic molecular aberrations with molecular targeted agents: Are biliary tract cancers an ideal playground? European Journal of Cancer. Elsevier Ltd; 2017.

63. Zhou SL, Xin HY, Sun RQ, Zhou ZJ, Hu ZQ, Luo C Bin, et al. Association of KRAS Variant Subtypes with Survival and Recurrence in Patients with Surgically Treated Intrahepatic Cholangiocarcinoma. JAMA Surg. American Medical Association; 2022; 157(1):59–65.

64. Xue L, Guo C, Zhang K, Jiang H, Pang F, Dou Y, et al. Comprehensive molecular profiling of extrahepatic cholangiocarcinoma in Chinese population and potential targets for clinical practice. Hepatobiliary Surg Nutr. AME Publishing Company; 2019; 8(6):615–22.

65. Fujimoto A, Furuta M, Shiraishi Y, Gotoh K, Kawakami Y, Arihiro K, et al. Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. Nat Commun. Nature Publishing Group; 2015; 6.

66. Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, et al. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. J Hepatol. Elsevier B.V.; 2018; 68(5):959–69.

67. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. Nat Genet. Nature Research; 2020; 52(8):859–64.

68. Wang T, Drill E, Vakiani E, Pak LM, Boerner T, Askan G, et al. Distinct histomorphological features are associated with IDH1 mutation in intrahepatic cholangiocarcinoma. Hum Pathol. W.B. Saunders; 2019; 91:19–25.

69. Ong CK, Subimerb C, Pairojkul C, Wongkham S, Cutcutache I, Yu W, et al. Exome sequencing of liver flukeg-associated cholangiocarcinoma. Nat Genet. 2012; 44(6):690–3.

70. Sia D, Losic B, Moeini A, Cabellos L, Hao K, Revill K, et al. Massive parallel sequencing uncovers actionable FGFR2-PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. Nat Commun. Nature Publishing Group; 2015; 6.

71. Nan L, Wang C, Wang J, Zhang S, Bo X, Wang Y, et al. ARID1A Downregulation Predicts High PD-L1 Expression and Worse Clinical Outcome in Patients With Gallbladder Cancer. Front Oncol. Frontiers Media S.A.; 2022; 12.

72. Andersen JB. Molecular pathogenesis of intrahepatic cholangiocarcinoma. J Hepatobiliary Pancreat Sci. Blackwell Publishing Asia; 2015; 22(2):101–13.

73. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, et al. Erratum: Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles (Cell Reports (2017) 18(11) (2780–2794) (S2211124717302140) (10.1016/j.celrep.2017.02.033)). Cell Reports. Elsevier B.V.; 2017.

74. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. Cell Rep. Elsevier B.V.; 2017; 18(11):2780–94.

75. Zou S, Li J, Zhou H, Frech C, Jiang X, Chu JSC, et al. Mutational landscape of intrahepatic cholangiocarcinoma. Nat Commun. Nature Publishing Group; 2014; 5(May).

76. Jiao Y, Pawlik TM, Anders RA, Selaru FM, Streppel MM, Lucas DJ, et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. Nat Genet. Nature Publishing Group; 2013; 45(12):1470–3.

77. Li M, Zhang Z, Li X, Ye J, Wu X, Tan Z, et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. Nat Genet [Internet]. Nature Publishing Group; 2014; 46(8):872–6. Available from: http://dx.doi.org/10.1038/ng.3030.

78. European Genome-Phenome Archive. Genentech study of gallbladder cancer - European Genome-Phenome Archive [Internet]. [cited 2024 Mar 26]. Available from: https://ega-archive.org/studies/EGAS00001003004.

79. Jose S, Zalin-Miller A, Knott C, Paley L, Tataru D, Morement H, et al. Cohort study to assess geographical variation in cholangiocarcinoma treatment in England. World J Gastrointest Oncol. Baishideng Publishing Group Inc; 2023; 15(12):2077–92.

80. Girvan AC, Zhang W, Liu J, Tang C, Ebert PJ, Aggarwal A, et al. Addition of ramucirumab or merestinib to standard first-line chemotherapy for locally advanced or metastatic biliary tract cancer: a randomised, double-blind, multicentre, phase 2 study. Articles Lancet Oncol [Internet]. 2021. Available from: www.thelancet.com/oncology.

81. Valle JW, Vogel A, Denlinger CS, He AR, Bai LY, Orlova R, et al. Addition of ramucirumab or merestinib to standard first-line chemotherapy for locally advanced or metastatic biliary tract cancer: a randomised, double-blind, multicentre, phase 2 study. Lancet Oncol [Internet]. Elsevier Ltd; 2021; 22(10):1468–82. Available from: http://dx.doi.org/10.1016/S1470-2045(21)00409-5.

82. The Human Genome Project. Genome assembly GRCh38. NCBI [Internet]. 2013 [cited 2024 Nov 4]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/.

83. lh3. Burrows-Wheeler Aligner [Internet]. [cited 2024 Apr 2]. Available from: https://bio-bwa.sourceforge.net/.

84. Nystedt B, Garcia M, Juhos S, Larsson M, Olason PI, Martin M, et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res. F1000 Research Ltd; 2020; 9.

85. nf-core. nf-core/sarek [Internet]. [cited 2024 Apr 2]. Available from: https://nf-co.re/sarek/3.4.0.

86. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol [Internet]. Nature Research; 2020; 38(3):276–8. Available from: https://www.nature.com/articles/s41587-020-0439-x.

87. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2 [Internet]. [date unknown]. Available from: https://doi.org/10.1101/861054.

88. Chen Z, Yuan Y, Chen X, Chen J, Lin S, Li X, et al. Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. Sci Rep. Nature Research; 2020; 10(1).

89. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016; 32(8):1220–2.

90. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012; 28(14):1811–7.

91. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods [Internet]. Springer US; 2018; 15(8):591–4. Available from: http://dx.doi.org/10.1038/s41592-018-0051-x.

92. Illumina. Strelka User Guide. GitHub [Internet]. 2021 [cited 2021 Dec 14]. Available from: https://github.com/Illumina/strelka/blob/v2.9.x/docs/userGuide/README.md.

93. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009; 25(17):2283–5.

94. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res [Internet]. 2012; 22(3):568–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22300766.

95. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. Curr Protoc Bioinformatics [Internet]. 2013; 44(3):15.4.1-17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25553206.

96. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol [Internet]. Genome Biology; 2016; 17(1):1–14. Available from: http://dx.doi.org/10.1186/s13059-016-0974-4.

97. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). Taylor and Francis Inc.; 2012; 6(2):80–92.

98. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018; 28(11):1747–56.

99. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: Cancer variant annotation tool. Hum Mutat. 2015; 36(4):E2423–9.

100. Mskcc. vcf2maf. GitHub [Internet]. 2021 [cited 2021 Dec 14]. Available from: https://github.com/mskcc/vcf2maf.

101. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. Oxford University Press; 2021; 10(2):1–4.

102. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJY, Karnebeek C Van, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014; 7(1):1–14.

103. Raine KM, Loo P Van, Wedge DC, Jones D, Menzies A, Butler AP, et al. ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. Curr Protoc Bioinformatics [Internet]. 2016; 56:15.9.1-15.9.17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27930809.

104. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012; 28(3):423–5.

105. VanLoo lab. ASCAT Github [Internet]. [cited 2024 Apr 4]. Available from: https://github.com/VanLoo-lab/ascat.

106. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2.

107. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high

concordance with metastatic tumors. Nat Commun. Nature Publishing Group; 2017; 8(1).

108. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. Nature Research; 2019; 37(7):773–82.

109. Finotello F, Mayer C, Plattner C, Laschober G, Rieder Di, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med. BioMed Central Ltd.; 2019; 11(1).

110. Nicorici D, Satalan MS, Edgren H, Kangaspeska S, Murumägi A, Kallioniemi O, et al. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data [Internet]. 2014. Available from: http://code.google.com/p/fusioncatcher/.

111. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol [Internet]. 2016; 34(5):525–7. Available from: https://doi.org/10.1038/nbt.3519.

112. Therneau TM. A Package for Survival Analysis in R [Internet]. 2024. Available from: https://CRAN.R-project.org/package=survival.

113. Terry M. Therneau, Patricia M. Grambsch. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

114. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using “ggplot2” [Internet]. 2021. Available from: https://CRAN.R-project.org/package=survminer.

115. Kennedy N. forestmodel: Forest Plots from Regression Models [Internet]. 2020. Available from: https://CRAN.R-project.org/package=forestmodel.

116. Harshil Patel, Phil Ewels, Alexander Peltzer, Jonathan Manning, Olga Botvinnik, Gregor Sturm, et al. nf-core/rnaseq: nf-core/rnaseq v3.14.0 - Hassium Honey Badger. 2024.

117. Meric-Bernstam F, Hanna DL, El-Khoueiry AB, Kang Y-K, Oh D-Y, Chaves JM, et al. Zanidatamab (ZW25) in HER2-positive biliary tract cancers (BTCs): Results from a phase I study. Journal of Clinical Oncology [Internet]. Wolters Kluwer; 2021; 39(3_suppl):299. Available from: https://doi.org/10.1200/JCO.2021.39.3_suppl.299.

118. Lee KW, Im Y-H, Lee KS, Cho JY, Oh D-Y, Chung HCC, et al. Zanidatamab, an anti-HER2 bispecific antibody, plus chemotherapy with/without tislelizumab as first-line treatment for patients with advanced HER2-positive breast cancer or gastric/gastroesophageal junction adenocarcinoma: A phase 1B/2 trial-in-progress. Journal of Clinical Oncology [Internet]. Wolters Kluwer; 2021; 39(15_suppl):TPS2656–TPS2656. Available from: https://doi.org/10.1200/JCO.2021.39.15_suppl.TPS2656.

119. Kam AE, Masood A, Shroff RT. Current and emerging therapies for advanced biliary tract cancers. Lancet Gastroenterol Hepatol [Internet]. Elsevier Ltd; 2021; 6(11):956–69. Available from: http://dx.doi.org/10.1016/S2468-1253(21)00171-0.

120. Gomez-Martin C, Plaza JC, Pazo-Cid R, Salud A, Pons F, Fonseca P, et al. Level of HER2 gene amplification predicts response and overall survival in her2-positive advanced gastric cancer treated with trastuzumab. Journal of Clinical Oncology. 2013; 31(35):4445–52.

121. Cocco E, Javier Carmona F, Razavi P, Won HH, Cai Y, Rossi V, et al. Neratinib is effective in breast tumors bearing both amplification and mutation of ERBB2 (HER2). Sci Signal [Internet]. 2018; 11(551). Available from: http://www.ncbi.nlm.nih.gov/pubmed/30301790.

122. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun [Internet]. Nature Publishing Group; 2017; 8(May):1–12. Available from: http://dx.doi.org/10.1038/ncomms15081.

123. Mezynski MJ, Farrelly AM, Cremona M, Carr A, Morgan C, Workman J, et al. Targeting the PI3K and MAPK pathways to improve response to HER2-targeted therapies in HER2-positive gastric cancer. J Transl Med [Internet]. BioMed Central; 2021; 19(1):1–16. Available from: https://doi.org/10.1186/s12967-021-02842-1.

124. Lee JY, Hong M, Kim ST, Park SH, Kang WK, Kim KM, et al. The impact of concomitant genomic alterations on treatment outcome for trastuzumab therapy in HER2-positive gastric cancer. Sci Rep. 2015; 5:1–6.

125. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. Nature Publishing Group; 2017; 545(7655):446–51.

126. Spain L, Coulton A, Lobon I, Rowan A, Schnidrig D, Shepherd STC, et al. Late-Stage Metastatic Melanoma Emerges through a Diversity of Evolutionary Pathways. Cancer Discov. American Association for Cancer Research Inc.; 2023; 13(6):1364–85.

127. University College London. The PEACE (Posthumous Evaluation of Advanced Cancer Environment) Study (PEACE) [Internet]. [cited 2024 Mar 21]. Available from: https://clinicaltrials.gov/study/NCT03004755.

128. Tran HS. Integrating multi-omics data by mapping subclonal events on tumour evolutionary trees [Internet]. 2023 [cited 2024 Mar 21]. Available from: https://discovery.ucl.ac.uk/id/eprint/10180204/.

129. Graham RP, Barr Fritcher EG, Pestova E, Schulz J, Sitailo LA, Vasmatzis G, et al. Fibroblast growth factor receptor 2 translocations in intrahepatic cholangiocarcinoma. Hum Pathol. W.B. Saunders; 2014; 45(8):1630–8.

130. Peraldo-Neia C, Cavalloni G, Fenocchio E, Cagnazzo C, Gammaitoni L, Cereda S, et al. Prognostic and predictive role of EGFR pathway alterations in biliary cancer patients treated with chemotherapy and anti-EGFR. PLoS One. Public Library of Science; 2018; 13(1).

131. Borbath I, Ceratti A, Verslype C, Demols A, Delaunoit T, Laurent S, et al. Combination of gemcitabine and cetuximab in patients with advanced cholangiocarcinoma: a phase II study of the Belgian Group of Digestive Oncology. Annals of Oncology [Internet]. 2013; 24:2824–9. Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005390.pub2/.

132. Francesca Ratti. Perihilar cholangiocarcimoma: can we avoid futile surgery? ALPS 2024. Madonna Di Campiglio; 2024.

133. Fan J, Bekaii-Saab TS, Aldrighetti LA, Bridgewater JA, Ferrone CR, Harding JJ, et al. A phase 3, randomized study of adjuvant rilvegostomig plus chemotherapy in resected biliary tract cancer: ARTEMIDE-Biliary01. 2024.

134. Palmeri M, Mehnert J, Silk AW, Jabbour SK, Ganesan S, Popli P, et al. Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. ESMO Open. Elsevier B.V.; 2022; 7(1).

135. Jia P, Yang X, Guo L, Liu B, Lin J, Liang H, et al. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. Genomics Proteomics Bioinformatics. Beijing Genomics Institute; 2020; 18(1):65–71.

136. Hernandez S, Lazcano R, Serrano A, Powell S, Kostousov L, Mehta J, et al. Challenges and Opportunities for Immunoprofiling Using a Spatial High-Plex Technology: The NanoString GeoMx® Digital Spatial Profiler. Frontiers in Oncology. Frontiers Media S.A.; 2022.

Appendix

Published abstracts

The impact of alterations in cancer driver genes and other potentially targetable mutations on progression and overall survival in patients with biliary tract cancer treated on the randomised phase III multicentre BILCAP clinical trial.

Journal of Clinical Oncology, 2023 Volume 41, Number 16_suppl
https://doi.org/10.1200/JCO.2023.41.16_suppl.4019

V.E. Crolley[1], R. Guest[2], A.D. Beggs[3], E. Jaynes[4], S. Thorn[5], J. Herrero[1], I. Tomlinson[6], J.W. Valle[7], J.N. Primrose[8] & J.A. Bridgewater[9]

[1] Cancer Institute, University College London, London, UK; [2] Clinical Surgery, University of Edinburgh, Edinburgh, UK; [3] Institute of Cancer & Genomic Science, The University of Birmingham - Medical School, Birmingham, UK; [4] Pathology, University Hospital Southampton NHS Foundation Trust, Southampton, UK; [5] Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK; [6] Nuffield Department of Medicine, University of Oxford, Oxford, UK; [7] Medical Oncology Department, The Christie NHS Foundation Trust, Manchester, UK; [8] Hepaticopancreatobiliary Surgery, Southampton General Hospital, Southampton, UK; [9] UCL Cancer Institute, UCL Cancer Institute – Paul O'Gorman Building, London, UK

Background

The BILCAP clinical trial established adjuvant capecitabine as the current standard of care treatment after biliary tract cancer resection.

Methods

Translational work from this clinical trial involved collecting archived fixed formalin tissue from consented BILCAP patients and carrying out low-pass whole genome (lp-WGS), targeted gene (TGS) and RNA sequencing (RNA-seq) for copy number (CN), mutation and gene-fusion analysis. In total, 98 patients underwent RNA-seq, 95 of 98 underwent lp-WGS and 39 of 98 underwent TGS.

Results

47 patients had intrahepatic cholangiocarcinoma, 47 patients had gallbladder cancer, 2 patients had perihilar cholangiocarcinoma, and 2 patients had distal cholangiocarcinoma. 62 (63.3%) patients were female, and 48 (49.0%) received adjuvant capecitabine. FGFR2 gene fusions were present in 24 patients (24.5%), as were fusions in NTRK1 (n=4, 4.1%), FGFR1 (n=4, 4.1%), FGFR3 (n=2, 2.0%) and FGFR4 (n=2, 2.0%). Known pathogenic mutations were seen in IDH1(n=4, 10.3%, total number of mutations=8, 20.5%), IDH2 (n=1, 2.6%, total n=6, 15.4%), and FGFR2 (n=1, 2.6%, total n=7, 17.9%). Commonly amplified (CN ≥ 4) genes included NTRK1 (n=28, 29.5%), ERBB2 (n=27, 28.4%) and MDM2 (n=20, 21.1%) with MYC (n=19, 20.0%), EGFR (n=16, 16.8%) and MET (n=15, 15.8%) also amplified. Nearly all the alterations investigated did not significantly affect recurrence risk (PFS) or overall survival (OS), including FGFR2 fusions (OS hazard ratio (HR) 1.11 p=0.762,

PFS HR 1.10 p=0.763). However, the presence of amplified EGFR (CN ≥ 4) significantly decreased both OS (HR 5.40 p=0.01) and PFS (HR 3.44 p=0.04).

Conclusions

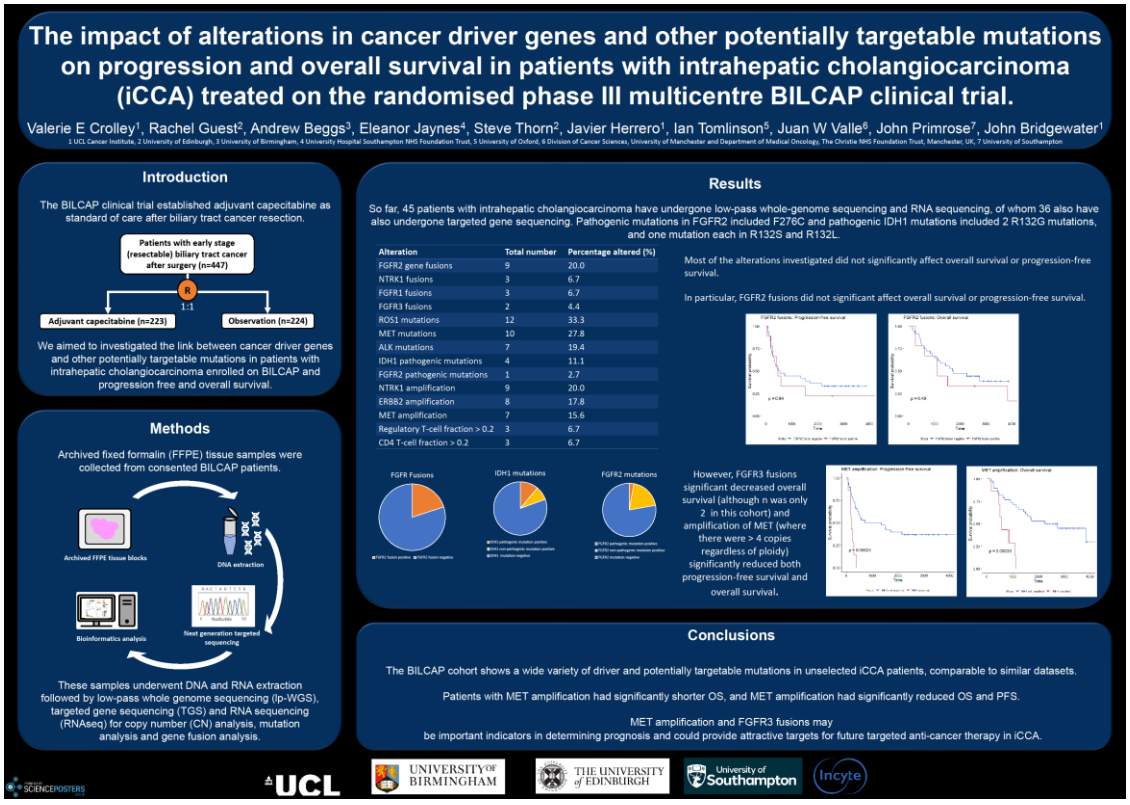
The BILCAP cohort shows a wide variety of driver and potentially targetable mutations in unselected biliary tract cancer patients, comparable to similar datasets. Of note, patients with EGFR amplification had significantly reduced OS and PFS. This indicates that EGFR amplification may be an important indicator in determining prognosis and could provide an attractive target for future targeted anti-cancer therapy in biliary tract cancer.

Clinical trial identification

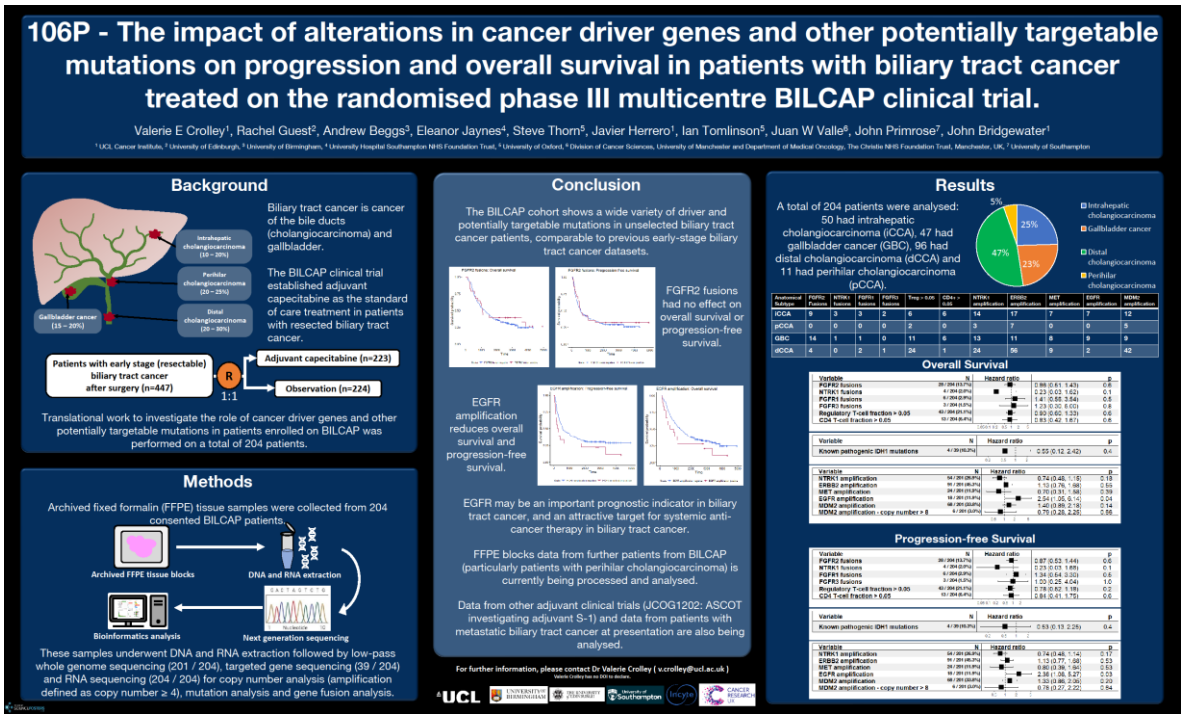
EudraCT 2005-003318-13.

Poster presentations

AMMF European Cholangiocarcinoma Conference 2023



European Society for Medical Oncology (ESMO) Congress 2023



Comparison of the tumour microenvironment between early-stage and metastatic biliary tract cancers: analysis of RNAseq data from the BILCAP and I3O-MC-JSBF clinical trials

Valerie E Crolley[1,2], Rachel Guest[3], Andrew Beggs[4], Eleanor Jaynes[5], Steve Thorn[3], Javier Herrero[1], Ian Tomlinson[6], Juan W Valle[7], John Primrose[8], John Bridgewater[1]

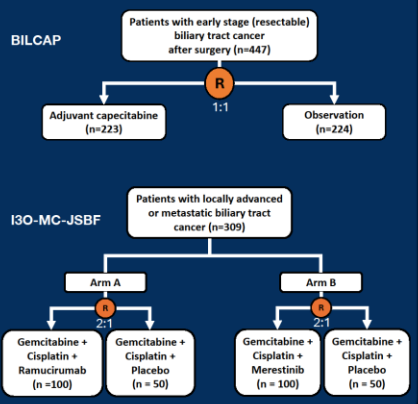
1 UCL Cancer Institute, 2 St Bartholomew's Hospital, 3 University of Edinburgh, 4 University of Birmingham, 5 University Hospital Southampton NHS Foundation Trust, 6 University of Oxford, 7 The Cholangiocarcinoma Foundation, 8 University of Southampton

Background

The importance of the tumour microenvironment is becoming increasingly recognised in tumorigenesis, and in determining whether tumours become invasive and form metastases.

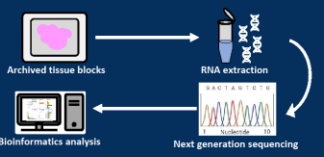
This project aimed to investigate any differences in the tumour microenvironment of biliary tract cancers at different stages of tumour development.

To do this, RNAseq data was collected from patients enrolled in two clinical trials: the phase III BILCAP clinical trial of early-stage biliary tract cancer patients and the phase II I3O-MC-JSBF clinical trial of patients with advanced or metastatic biliary tract cancer.



Methods

RNA was extracted from archived tumour samples from 201 patients enrolled the BILCAP clinical trial and from 129 patients enrolled on the I3O-MC-JSBF clinical trial.

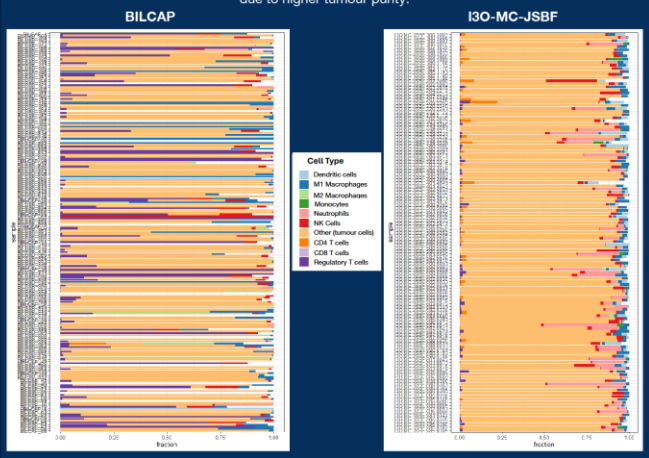


The bulk RNA was then sequenced using RNAseq, and analysed using the nf-core RNAseq STAR-salmon tool.

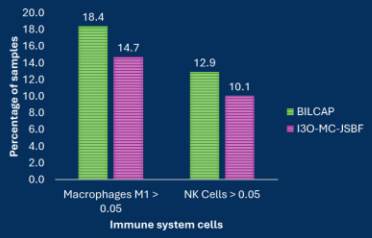
QuanTIseqR (using immundeconv) was then used to determine the proportion of each type of immune cell in the tumour microenvironment of each patient sample.

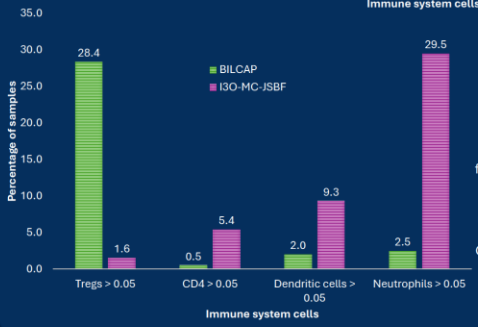
Results

Overall, the proportion of immune system cells was lower in the I3O-MC-JSBF samples, likely due to higher tumour purity.



The proportion of M1 macrophages was similar between the early-stage and locally advanced / metastatic datasets, as was the proportion of NK cells.





However, there were significant differences in the proportion of other immune system cells.

Early-stage cancers had far more regulatory T-cells, while the tumours of locally advanced and metastatic cancer patients had higher numbers of CD4 T-cells, dendritic cells and neutrophils.

Conclusions

Overall, the tumour microenvironment shows differences between early-stage and locally advanced or metastatic biliary tract cancer.

Early-stage biliary tract cancers had a more immunosuppressive tumour microenvironment with higher numbers of regulatory T cells.

In comparison, locally advanced or metastatic biliary tract cancers had a more inflammatory tumour microenvironment, with higher numbers of CD4 T cells and neutrophils.

These results may help to explain the success in using immunotherapy in locally advanced and metastatic biliary tract cancers, as seen in the TOPAZ-1 and KEYNOTE-922 clinical trials.

In conclusion, these results may also have significant implications as to the potential success of immunotherapy in early-stage biliary tract cancer in the neoadjuvant and adjuvant clinical settings.

For further details and information please contact: v.crolley@ucl.ac.uk