

Natural history of a parasite-induced biliary cancer

Thomas Crellen^{1,2,3*}, Francesca Vita^{4,5}, Chiara Braconi^{4,6,7}, Paiboon Sithithaworn^{8,9}, and T. Déirdre Hollingsworth²

¹School of Biodiversity One Health and Veterinary Medicine, Graham Kerr Building, University of Glasgow, 82 Hillhead Street, Glasgow G12 8QQ, UK; ²Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK; ³Parasites and Microbes Program, Wellcome Sanger Institute, Hinxton CB10 1RQ, UK; ⁴School of Cancer Sciences, Wolfson Wohl Cancer Research Centre, University of Glasgow, Garscube Estate, Switchback Road, Glasgow G61 1QH, UK; ⁵Department of Oncology, University of Turin, Turin 10043, Italy; ⁶Beatson West of Scotland Cancer Centre, 1053 Great Western Road, Glasgow G12 0YN, UK; ⁷CRUK Scotland Cancer Centre, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK; ⁸Department of Parasitology, Faculty of Medicine, Khon Kaen University, 123 Thanon Mittraphap, Khon Kaen 40002, Thailand; ⁹Cholangiocarcinoma Research Institute, Khon Kaen University, 123 Thanon Mittraphap, Khon Kaen 40002, Thailand

*Corresponding author: thomas.crellen@glasgow.ac.uk

1 Pathogens are major drivers of cancer globally. Quantifying the relationship between infection and carcinogenesis is therefore crucial for developing
2 preventative programs. The foodborne trematode *Opisthorchis viverrini* is a primary cause of biliary cancer (cholangiocarcinoma) and infects 12
3 million people in Southeast Asia. In tumours from patients exposed to *O. viverrini* we found that the earliest chromosomal amplification of driver
4 genes occurred at 30 years old on average, two to four decades before cancer diagnosis, and disproportionately contained *FGFR2*, *TP53* and
5 *PTEN* genes. We then fitted transmission models to parasitological data from Thailand spanning 27 years ($n = 11,517$) finding that, for people born
6 between 1960–1989, first exposure occurred at two years old and by 30 years individuals had been cumulatively infected with a median of 72 worms.
7 Trematodes are long-lived and our analysis quantifies the average lifespan of *O. viverrini* as 13 years (90% credible interval [CrI] 6–23 years) within
8 human hosts. Finally, we calculate the lifetime probability of diagnosis with cholangiocarcinoma as 1.2% (90% CrI 1.0–1.4%) given prior exposure to
9 the parasite. Overall, our study demonstrates how pathogen exposure drives patterns of cancer within human populations.

Keywords: Cholangiocarcinoma | Foodborne trematodiasis | Cancer genomics | Somatic evolution | Epidemiology | Neglected tropical diseases

1 Introduction

2 Infectious organisms are a major contributor to the burden
3 of human cancers globally (1), thus challenging the classic
4 distinction between ‘communicable’ and ‘non-communicable’
5 diseases. Of the eleven pathogens recognised as direct human
6 carcinogens; three are parasitic trematodes*†. Infection with
7 the foodborne liver flukes *Opisthorchis viverrini* and *Clonorchis*
8 *sinensis* cause cancer of the bile duct (cholangiocarcinoma),
9 whilst the waterborne blood fluke *Schistosoma haematobium*
10 causes cancer of the bladder (squamous cell carcinoma). The
11 aetiologies resulting from infection with these parasites are
12 classified as neglected tropical diseases, owing to the historical
13 neglect of research into both the causative agents and the
14 afflicted populations (2). The pathology arising from infection
15 with parasitic worms (also known as helminths) is typically
16 chronic making it difficult to quantify the impact of any single
17 helminth species over decades of exposure and in populations
18 with multiple co-infections (3, 4).

19 The liver fluke *O. viverrini* has a foodborne transmission
20 route and is acquired from eating raw or lightly fermented
21 freshwater fish, which is a traditional component of the diet in
22 regions of Southeast Asia (5). The fluke has a complex lifecycle,
23 which involves asexual reproduction within freshwater snails
24 and encysts as a mammalian-infective stage (metacercariae)
25 in cyprinid fish. Despite public health programs to control the
26 parasite, an estimated 12.4 million people were infected with
27 *O. viverrini* across Thailand, Laos, Cambodia and Vietnam in
28 2018 (6). While the prevalence has shown gradual declines in
29 Thailand due to periodic parasite control programs during the
30 second half of the twentieth century (7), progress has recently
31 slowed (8, 9). In Cambodia and Laos, by contrast, there is
32 evidence for increased transmission over the past two decades
33 (6, 10, 11). Liver fluke endemic countries have the highest
34 incidence of cholangiocarcinoma globally and cases of hepatic
35 and biliary cancers in these regions are disproportionately at-
36 tributable to cholangiocarcinoma, rather than hepatocellular

37 carcinoma (12, 13). If an early diagnosis is made then curative
38 surgical resection is possible in a minority of cases (14). Pa-
39 tients generally present late, however, and prognosis is poor
40 with a median survival time of <1 year following diagnosis and
41 the standard of care; systemic chemotherapy with cisplatin
42 plus gemcitabine (15, 16). The mechanisms through which
43 flukes induce cholangiocarcinoma is a combination of mechani-
44 cal damage, inflammation of the biliary epithelium, and the
45 secretion of proteins; in particular the peptide granulin (17).

46 Given the poor prognosis for cholangiocarcinoma and the
47 preventable nature of parasite infection, there is a strong moti-
48 vation to understand the link between liver fluke exposure and
49 carcinogenesis in humans (3, 18). Prior to the onset of driver
50 mutations, anthelmintic treatment and reducing parasite ex-
51 posure should be prioritised as public health interventions,
52 whereas following the onset of irreversible malignancies the
53 priority for interventions shifts to ultrasound screening for liver
54 pathology and early referral for surgery (19).

55 Quantifying the epidemiological relationship between liver
56 fluke infection and the resulting biliary pathology is challenging,
57 however, due to i) the dynamic nature of parasite transmission;
58 ii) exposure being cumulative over time, as humans do not gen-
59 erate protective immunity to reinfection; iii) the long lifespan
60 of adult worms within the human host, which is unknown for *O.*
61 *viverrini*; and iv) the decades-long temporal lag between initial
62 parasite infection and cases of cholangiocarcinoma (3, 20).

63 In this study we infer the timing of the earliest driver
64 mutations for fluke-induced cholangiocarcinoma using compu-
65 tational methods which characterise the evolution of tumours
66 from a single biopsy (21, 22). We then define the age of first ex-
67 posure to liver fluke by fitting dynamic transmission models to
68 parasitological survey data spanning several decades. Finally,
69 we estimate the lifetime probability of diagnosis with cholan-
70 giocarcinoma given infection with *O. viverrini*. By combining
71 evolutionary cancer genomics with epidemiological analysis, we
72 obtain unique insights into the relationship between pathogen
73 exposure and tumourigenesis, with implications for evidence-
74 based disease control (18).

* <https://monographs.iarc.fr/wp-content/uploads/2018/06/mono61.pdf>

† <https://monographs.iarc.fr/wp-content/uploads/2018/06/mono100B.pdf>

75 Results

76 **Cholangiocarcinoma tumour genomes.** We obtained paired
77 tumour and normal whole-genome sequences from cholangio-
78 carcinoma patients who were previously infected with liver
79 fluke and were treated at a large public hospital in North-
80 east Thailand (23). We used a bioinformatics pipeline to call
81 somatic single nucleotide variants (SNVs), copy number alter-
82 ations, and inferred the clonal status of these mutations (see
83 Methods). The age of the patients at surgery ranged from 37–
84 79 years (median 57 years), 50% of patients were female, and
85 all were born prior to 1980 (23). After mapping reads, variant
86 calling, and filtering (see Methods) we obtained 2,349–27,821
87 (median 10,360) somatic SNVs and 268–14,230 (median 1,382)
88 somatic indels per-tumour. The overall ploidy (chromosomal
89 copies) per tumour ranged from 1.2–3.7 (median 2.0).

90 **Evolution of cholangiocarcinoma tumours.** We estimated the
91 timing of driver mutations in 43 genes implicated in cholan-
92 giocarcinoma development (24–26) using evolutionary models
93 which time the emergence of somatic single nucleotide variants
94 (SNVs) occurring on amplified sections of chromosome (21, 22).
95 We first assessed whether any tumours had been subject to
96 whole-genome duplication events based on the correlated tim-
97 ings of chromosomal amplifications across the genome (21) and
98 concluded that this had occurred in three tumours (see Fig.
99 1A). Across all tumours, the copy number was disproportion-
100 ately higher in chromosomes 7 and 17. In 17/22 tumours with
101 sufficient ploidy and tumour purity, we inferred the timing of
102 focal chromosomal amplifications in potential driver genes for
103 cholangiocarcinoma (see Methods). Overall the majority of
104 amplifications (166/287; 58%) occurred later in ‘chronological
105 time’ (0.75 or later), with a smaller proportion (46/287; 16%)
106 occurring earlier in the lifespan of the tumour (before 0.5), as
107 shown in Fig. 1B.

108 **Timing of amplified driver mutations.** We determined the ear-
109 liest amplified driver gene for each tumour sample, and classified
110 these as either ‘clonal’ (occurring on the most recent common
111 cell lineage of the tumour) or ‘subclonal’ (a subsequent clonal
112 expansion within the tumour which has not risen to fixation)
113 (27, 28). To calculate the age at which these amplifications
114 occurred in patients, we multiplied the chronological time es-
115 timate with the age at surgery (see Methods). The earliest
116 clonal amplification of driver genes occurred at a median age
117 of 30 years and with an interquartile range (IQR) of 20–43
118 years (Fig. 1C). The earliest subclonal amplifications occurred
119 at a median age of 33 years (IQR 29–41 years). A variety
120 of genes were the earliest amplified, with the most frequently
121 occurring being the tumour suppressor *PTEN* in the clonal
122 lineages (earliest in two tumours and amplified in 7/17 tu-
123 mours), while the signalling gene *NRAS* was the most common
124 in subclonal lineages (earliest in two tumours and amplified in
125 7/17 tumours) see Fig. 1B.

126 To determine whether certain genes are disproportionately
127 likely to be amplified early or late within the lifespan of the
128 tumour, we applied a generalised linear model (see Methods)
129 to the chronological age estimates for each of the amplifica-
130 tion events, while controlling for host sex, tumour anatomical
131 subtype (extrahepatic or perihilar) and clonality (clonal or
132 subclonal). We excluded nine amplifications with fewer than
133 eight somatic SNVs, giving 278 amplification events from 17
134 tumours for our analysis. The model estimates per gene are
135 shown by chronological age in Fig. 1D. Overall, the driver

136 genes which were disproportionately found to be amplified early
137 were the fibroblast growth factor receptor *FGFR2*, *PTEN* and
138 the tumour suppressor *TP53* (these genes were amplified in
139 7 tumours), while the tumour suppressor genes *BAP1* and
140 *PBRM1* (both amplified in 5 tumours), and the receptor gene
141 *FGFR1* (amplified in 4 tumours) were preferentially amplified
142 late. Our findings add support to previous studies which have
143 noted the early clonal amplification of *TP53*, in particular, as
144 a driver across a range of cancer types (21).

145 **Liver fluke transmission.** As fluke-induced pathology of the bil-
146 iary tract is chronic, there is a need to consider prior exposure
147 to the parasite at the individual or population level which can
148 be estimated from historical parasitological surveys (3). We
149 therefore collated epidemiological surveys from Northeast Thai-
150 land where diagnostic observations of *O. viverrini* infection
151 intensity (worm burdens or faecal egg counts) were available
152 by host age. Our analysis uses data from 4,056 individuals
153 obtained from seven surveys conducted between 1980–1989,
154 prior to the onset of large-scale control programs against liver
155 fluke, and 7,448 individuals from five surveys conducted be-
156 tween 1994–2017, which are summarised in Table 1. We fitted
157 a mechanistic parasite transmission model to these data which
158 considers that the parasite burden at a given age is a result
159 of flukes infecting the host with an age-variable transmission
160 rate, referred to as the ‘force of infection’, and parasites exit
161 the host due the spontaneous death rate of adult worms (20).
162 Full model details are provided in Methods.

163 **Exposure to liver fluke in human populations.** The average age
164 at which a person born between 1960–1989 was first infected
165 with a single *O. viverrini* fluke is 22 months (90% prediction
166 interval [PI] 19–25 months). The average age of first infection
167 for a person born after 1990 rises to 8 years (90% PI 6.5–9.5
168 years), though a smaller proportion of the population ultimately
169 becomes exposed during their lifetime; 33% post-intervention
170 compared with 88% pre-intervention. We define exposure to *O.*
171 *viverrini* as the cumulative number of adult flukes acquired by
172 a person over time, which is given by the area under the force of
173 infection curve (Equation 6). In endemic regions of Thailand in
174 the 1980s by ages 10, 20, and 30 years the median number of *O.*
175 *viverrini* flukes acquired was 12 (90% PI 8–16), 39 (90% PI 27–
176 53), and 72 (90% PI 48–103) respectively (Fig. 2C). Following
177 public health interventions in the 1990s (7), we estimate that
178 the *O. viverrini* force of infection dropped substantially, with
179 an almost 40-fold reduction in the age-dependent transmission
180 rate, resulting in the majority of individuals born after 1990
181 remaining uninfected by age 30 (median worm burden is zero).
182 The distribution of *O. viverrini* parasites among human hosts
183 shows high variation, as is characteristic for helminths (29),
184 and for the top decile of the most heavily infected hosts the
185 cumulative burden is ≥ 693 worms by 30 years of age in the
186 pre-intervention period (90% PI 523–931) and ≥ 14 worms in
187 the post-intervention period (90% PI 10–21). These findings
188 add weight to the reported declines in parasite transmission in
189 Thailand following control programs (6–8), though our analysis
190 is the first to quantify this effect in terms of worm burden, the
191 age of first exposure, and the cumulative proportion exposed.

192 **Lifespan of the adult worm.** The lifespan of helminths in hu-
193 man infections is considered to be in the order of one year
194 to three decades, although this is rarely quantified (30, 31).
195 We estimate the lifespan of *O. viverrini* within our force of
196 infection model as 12.9 years (90% CrI 6.2–23.1 years), which

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

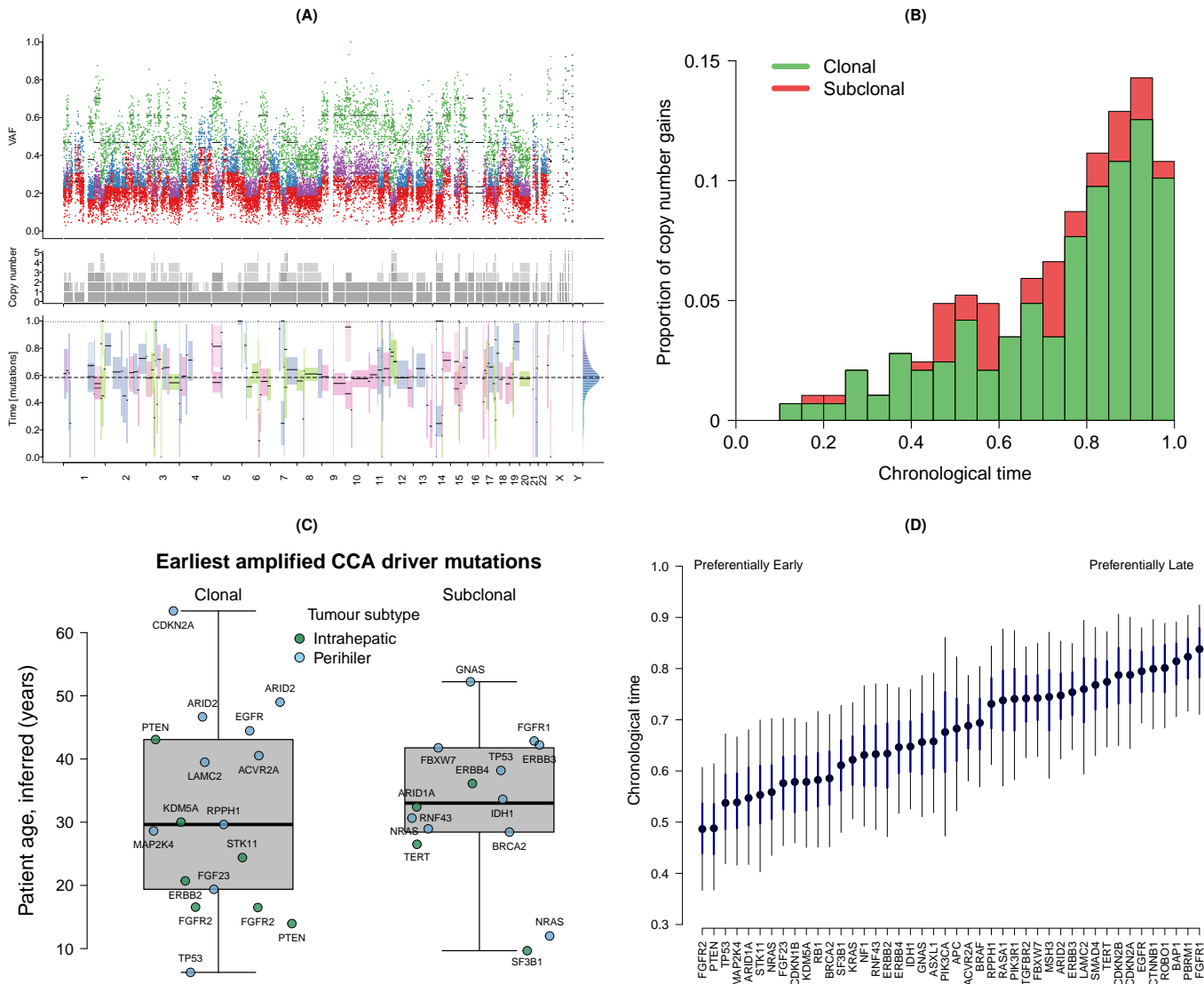


Fig. 1. Evolution of cholangiocarcinoma tumours. A) Subclonal lineage reconstruction for cholangiocarcinoma tumour CCA_TH19 using MutationTimer (21). The upper plot shows the variant allele frequency (VAF) for somatic SNVs which are coloured by clonal state; clonal (early) = green, clonal (late) = purple, clonal (unknown) = blue, subclonal = red. The middle plot shows the inferred copy number frequency (ploidy) by chromosome. The lower plot shows the inferred timings of copy number gains. For this tumour, the timings of copy number gains are correlated across chromosomes, indicating a whole-genome duplication event at a 'mutation time' of 0.6. B) Histogram showing the timing of 287 clonal and subclonal amplification events for 44 driver mutations in 17 cholangiocarcinoma tumours inferred by AmplificationTimer (22). C) The earliest amplified clonal and subclonal driver mutations in cholangiocarcinoma (CCA) tumours (22) multiplied by the patient age at surgery (23) to give the inferred age for each amplification event. Each point represents the first amplified driver gene per patient, labelled with the gene name, and the points are coloured by tumour anatomical subtype (intrahepatic or perihilar). The overlaid box and whisker plots shows the median and interquartile range. D) Estimates of the chronological time of amplification by gene. Results are shown from a generalised linear model applied to estimates from AmplificationTimer (see Methods). The black points give the posterior median, the thick blue line gives the 50% credible interval and the thin black line the 95% credible interval.

197 is defined as the average time for half of adult stage parasites
 198 to die within the human host in the absence of anthelmintic
 199 treatment (Equation 3, see Methods). Our posterior distribu-
 200 tion for parasite mortality is long tailed (Fig. 2) with the
 201 top five percent of adult *O. viverrini* lifespans ≥ 25.7 years.
 202 This finding is consistent with case reports of the related liver
 203 fluke *C. sinensis* from travellers, which is known to persist in
 204 human hosts for 26 years (32). We note that the posterior
 205 distribution of the worm mortality parameter is correlated
 206 with a force of infection parameter, and therefore we encourage
 207 use of the interval for *O. viverrini* adult worm lifespan (6–23
 208 years), rather than the point estimate, in future analyses.

209 **Latent and induction periods of cholangiocarcinoma.** Bringing
 210 together the evidence presented thus far, we hypothesise that
 211 people born between 1960–1989 in Northeast Thailand first
 212 became infected with liver fluke by age two and driver muta-
 213 tions for biliary cancer occurred around age 30. We sought to
 214 validate our hypothesis using cancer registry data (33). The
 215 age distribution of 10,737 cases of cholangiocarcinoma diag-
 216 nosed in Northeast Thailand between 1985–2009 are shown in
 217 Figure 3A. The median age of diagnosis was 59 years between
 218 1985–1997 and 63 years between 1997–2009. Using a time-to-
 219 event analysis, which accounts for interval censored data, we
 220 estimated the age of driver mutation and the subsequent age
 221 of cholangiocarcinoma as sequential gamma distributions (see
 222 Methods). The induction period (time from initial parasite

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

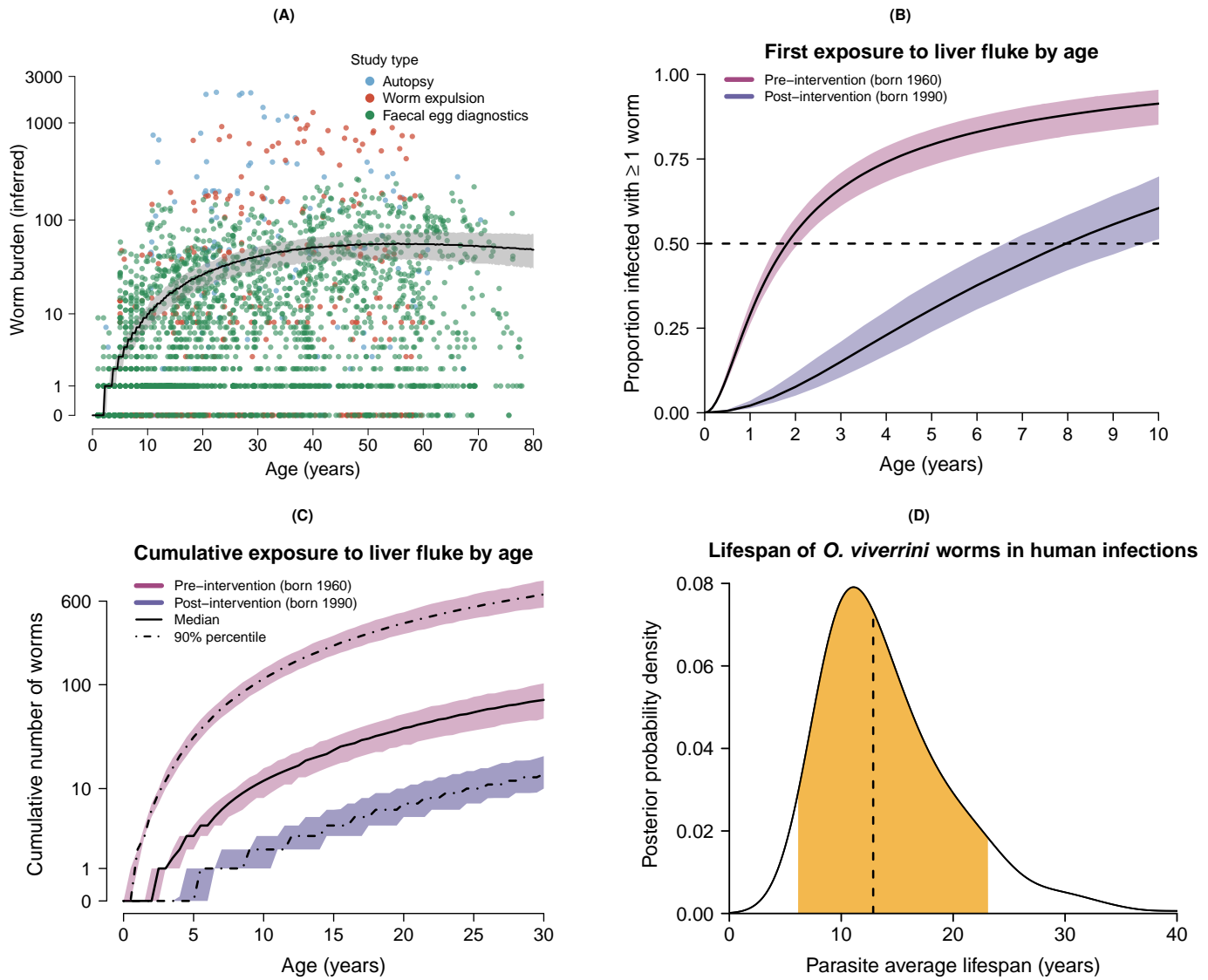


Fig. 2. Epidemiological data on the carcinogenic liver fluke *Opisthorchis viverrini* in Thailand. A) Worm burdens of *O. viverrini* by participant age, inferred from multiple surveys conducted between 1980–1989 prior to large-scale interventions. The data used for model fitting were counts of i) adult worms obtained from autopsy, or ii) worm expulsion, or iii) parasite eggs in faeces (data shown from S1–S6, see Table 1). The fitted black line shows the simulated median worm burden by host age, with the shaded area giving the 90% prediction interval (model fitted to data from S1–S7; see Table 1). The y-axis is log transformed. B) Proportion of exposed population with at least one parasite in the pre-intervention (1980–1989; S1–S7) and post-intervention periods (1994–2017; S8–S12 see Table 1). Solid lines are central estimates and shaded areas give the 90% credible interval (CrI). Where the solid lines cross the dashed line, this gives the age at which half of the exposed population become first infected, which is 22 months (90% CrI 19–25 months) of age in the pre-intervention period and 8 years of age (90% CrI 7–9.5 years) in the post-intervention period. C) Cumulative exposure to *O. viverrini* by age (Equation 6). The solid line shows the simulated median for the pre-intervention period, while the dot-dash lines show the cumulative worm burden for the upper 90% population percentile. Shaded areas give the 90% prediction interval. The y-axis is log transformed. D) Average lifespan of adult *O. viverrini* worms in human hosts. The posterior probability density is shown here for $\log(2)/\sigma$; where σ is the spontaneous death rate of adult worms (Equation 3). Dashed line gives the posterior median (13 years). The filled area gives the 90% CrI (6–23 years).

223 infection to driver mutation) was estimated as 28 years (90%
 224 CrI 16–39 years) and the latent period (time from driver muta-
 225 tion to cancer diagnosis) was estimated as 32 years (90%
 226 CrI 21–44 years). These distributions are shown in Fig. 3B.
 227 The incidence data therefore supports our estimate of the time
 228 to first amplified mutation at thirty years, given exposure
 229 by age two. The latent period estimate from registry data
 230 is longer than the 22 years indicated by our earlier analysis
 231 of the tumour genomes (time between first amplified driver
 232 mutation and biliary surgery), which likely reflects an older
 233 average age of diagnosis in the registry data compared to our,
 234 much smaller, sample of patients with sequenced biliary tu-
 235 mours (23, 33). Prior research on the induction and latent

236 periods for fluke-induced cholangiocarcinoma in humans has
 237 been limited, though American veterans from the Vietnam
 238 War had an elevated incidence of biliary cancer five or more
 239 decades after potential exposure to *O. viverrini* between 1965–
 240 1971 (34). While the association between military service in
 241 Vietnam and fluke-induced cholangiocarcinoma is contentious
 242 (35), the reported time between parasite exposure and cancer
 243 for United States veterans (induction plus latent period) is
 244 consistent with our findings.

245 **Lifetime probability of cholangiocarcinoma.** We estimate the
 246 lifetime probability of acquiring cholangiocarcinoma by fitting
 247 a survival model (see Methods) to age-incidence cancer registry

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

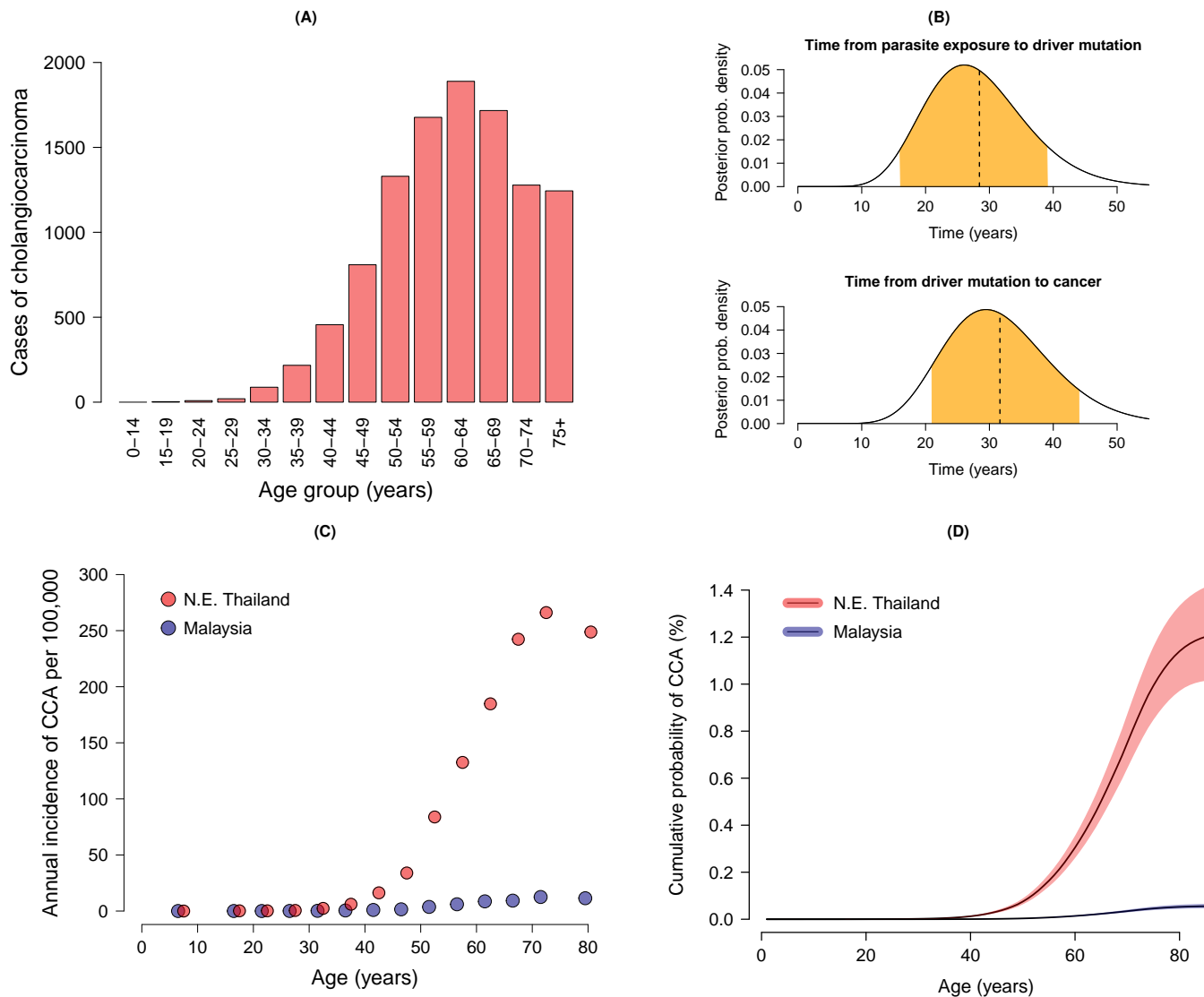


Fig. 3. Incidence of cholangiocarcinoma (CCA) at the population level. A) The age distribution of 10,737 cases of cholangiocarcinoma from Northeast Thailand between 1985–2009 (33). B) Time-to-event distributions for the induction period (from parasite exposure to driver mutation) and the latent period (from driver mutation to cancer diagnosis). The plots show the posterior probability distributions, where the dashed line gives the median and the shaded orange area shows the 90% credible interval (CrI). The induction period was estimated as a median of 28 years and the latent period as 32 years. C) Annual incidence of cholangiocarcinoma per 100,000 people by age for Northeast (N.E.) Thailand, where the carcinogenic liver fluke *O. viverrini* is endemic, and Malaysia which is non-endemic for *O. viverrini*. D) Cumulative lifetime probability of diagnosis with cholangiocarcinoma for the two populations, N.E. Thailand and Malaysia, at age α , conditional on survival to age α . The model output is shown from a survival analysis (Equation 17, see Methods), where the solid black lines give the posterior median and the shaded areas the 90% CrI. Probability given as a percentage (%).

248 data from Northeast Thailand between 1998–2009 (33) and
 249 Malaysia between 2007–2009 (36), which neighbours Thailand
 250 but is non-endemic for liver fluke (37). The underlying data are
 251 shown in Fig. 3C as an annual population incidence per 100,000
 252 population and the survival model predictions are shown in Fig.
 253 3D. Our results show that by 75 years of age the cumulative
 254 probability of diagnosis with cholangiocarcinoma, conditional
 255 on the person surviving to that age, is 0.046% in Malaysia
 256 (90% CrI 0.039–0.054%) and 0.99% in Northeast Thailand
 257 (90% CrI 0.85–1.1%), which is twenty-fold higher. Under the
 258 assumptions that 90% of excess cancer cases in Northeast
 259 Thailand are attributable to infection with liver fluke (38) and
 260 the total proportion of the population exposed to the parasite
 261 between 1960–1989 was 88% (estimated above, see Fig. 2B.),
 262 we calculate the probability of developing cholangiocarcinoma,

263 given prior infection with *O. viverrini*, as 1.0% (90% CrI 0.87–
 264 1.2%) by 75 years of age and 1.2% (90% CrI 1.0–1.4%) by
 265 85 years. While a previous report suggested a higher lifetime
 266 probability of 5% (39), the data and method used to produce
 267 this estimate are unclear. Calculating the lifetime probability
 268 of cancer given infection depends on the population exposed
 269 to *O. viverrini*. Prevalence surveys taken at a single moment
 270 in time will underestimate this proportion (38, 40), and hence
 271 overestimate the relative probability of cancer among infected
 272 people. The probability of developing cholangiocarcinoma
 273 covaries with the infection intensity of liver fluke (41). Given
 274 improved longitudinal data, the lifetime probability of biliary
 275 cancer should be stratified by the cumulative worm burden in
 276 future analyses.

277 Discussion

278 Our study uncovers many crucial, yet previously unknown, epi-
279 demiological processes on the pathway from parasite infection
280 to carcinogenesis in humans. The aim of such research is to
281 inform control programs to reduce both transmission of liver
282 fluke and subsequent cases of cancer (18, 19). Given the find-
283 ings presented here, that the age of first fluke infection is prior
284 to ten years of age and the first amplified driver mutations for
285 biliary cancer are on average at 30 years of age, we recommend
286 that interventions should be targeted to younger individuals
287 in at-risk regions, such as school-based deworming with the
288 anthelmintic praziquantel (42). Relatively few studies have
289 investigated the impact of anthelmintic treatment on prevent-
290 ing or reversing biliary damage in humans (43–45), though the
291 current evidence suggests a beneficial effect of deworming even
292 in the context of repeated reinfection. The long temporal lag
293 of four to seven decades between first parasite exposure and
294 diagnosis with cholangiocarcinoma (see Fig. 2B and Fig. 3)
295 means that changes to *O. viverrini* transmission intensity will
296 take many years to influence the incidence of cholangiocarci-
297 noma at the population level (19). Future randomised clinical
298 trials investigating the impact of deworming with praziquantel
299 on the risk of biliary cancer could be considered unethical
300 given the need to withhold or restrict anthelmintic treatment
301 for controls (18). Dynamic simulations have a useful role to
302 play, therefore, in estimating the magnitude of interventions
303 over long time periods (46). Our results provide a platform
304 for biomarker discovery by highlighting early driver genes and
305 we argue that screening for biliary malignancies should start
306 from thirty years of age. Furthermore, the long latent period
307 provides a window for early therapeutic interventions (47).

308 While there are relatively few studies investigating the epi-
309 demiological relationship between infection and carcinogenesis,
310 when compared against other pathogen-driven cancers our
311 estimates of the induction and latent periods for fluke-induced
312 cholangiocarcinoma appear to be longer. For human papillo-
313 mavirus (HPV), the time from infection to high-grade cervical
314 intraepithelial neoplasia (CIN2/3) has been estimated from reg-
315 istry data as three years, and the time from CIN2/3 to cancer
316 23.5 years (48). The time from infection with hepatitis B to
317 hepatocellular carcinoma is reported as 25–30 years in a review
318 (49), though we are unaware of any supporting evidence. Mech-
319 anisms for a potentially longer tumorigenesis from *O. viverrini*
320 infection, compared with other carcinogenic pathogens, have
321 not been investigated. We note that the impact of fluke in-
322 fection on biliary pathology is dose-dependent (41), and so
323 depends on the distribution of parasites within an endemic
324 community (40). There may also be evolutionary reasons for a
325 slower accumulation of pathology from helminths, given their
326 long lifespans (30, 50).

327 As the data used here are primarily from Northeast Thailand
328 our estimates of epidemiological parameters are most applicable
329 to this region. Large-scale ultrasound screening programs for
330 liver disease in Thailand have successfully diagnosed thousands
331 of cholangiocarcinoma and pre-cancerous cases since 2015 (51).
332 However, the enhanced screening will likely lead to short-term
333 increases in the reported incidence of cholangiocarcinoma, thus
334 further complicating our understanding of the relationship
335 between parasite infection and cancer (19). Extending our
336 analyses to Laos and Cambodia, where the burden of parasitic
337 disease is greater (6, 10) and the capacity for healthcare systems
338 to diagnose and treat biliary cancer more limited (52), remains
339 a priority for future research.

340 Methods and Materials

341 **Cholangiocarcinoma whole-genome sequences.** We accessed
342 paired tumour-normal whole-genome sequences from 22 in-
343 dividuals in Northeast Thailand with previous exposure to
344 liver fluke infection. Tumour tissue was obtained from patients
345 during surgical resection of the biliary tract at Srinagarind Hos-
346 pital in Khon Kaen, Thailand (23). Normal somatic genomes
347 were obtained from patient blood samples. Tumours are classi-
348 fied according to their anatomical location on the biliary tree;
349 namely intrahepatic cholangiocarcinoma (within the hepatic
350 ducts), perihilar cholangiocarcinoma (between second order
351 bile ducts and the cystic duct insertion) or distal cholangiocar-
352 cinoma (below the cystic duct). Here our samples consisted
353 of eight perihilar and 14 intrahepatic tumours. The age at
354 surgery ranged from 37–79 years (median 57 years) and 11/22
355 (50%) of patients were female.

356 **Cancer Genomics.** Starting with fastq files, we trimmed adap-
357 tor sequences from 150bp illumina paired-end reads using
358 TrimGalore v.0.4.4 (53) and mapped the reads against the hu-
359 man reference (GRCh37) using bwa mem v.0.7.17 with default
360 parameters (54). The resulting BAM files were then sorted
361 and indexed with samtools v.1.9. Duplicates were marked and
362 removed using GATK v.4.1.4.1 (55) and base quality scores
363 re-calibrated for tumour sequences with ICGC PCAWG con-
364 sensus vcf files as known variants for single nucleotide variants
365 (SNVs) and indels (56). Examining the depth of mapped reads,
366 we took the output from samtools depth -a (including zeroes)
367 and binned the mean depth within 1Mb segments (Fig. 1A).
368 Overall we found that the mean sequencing depth averaged
369 across the whole genomes ranged from 35x–73x per-sample
370 (median 57x) for normal genomes and 45x–72x per-sample
371 (median 54x) for tumour genomes.

372 We called somatic SNVs and indels using GATK Mutect2
373 with a panel of normals provided by the Broad Institute (57),
374 and with additional filters to remove secondary and supple-
375 mentary reads. Prior to filtering, the number of somatic SNVs
376 ranged from 109,339–292,886 per sample (median 131,057).
377 We calculated the fraction of reads in the normal with tumour
378 contamination using the GATK tool CalculateContamination
379 in combination with 4.7 million common germline alleles (MAF
380 0.01–0.20) derived from diverse Asian populations in Singapore
381 (58). This revealed that the contamination in normal samples
382 was low with <0.7% of reads coming from cross-sample con-
383 tamination. Using the data on contamination, we filtered the
384 Mutect2 variants calls, leaving 2,349–27,821 (median 10,360)
385 somatic SNVs and 268–14,230 (median 1,382) somatic indels
386 per-tumour. The number of somatic variants called in our
387 samples are consistent with those observed for other biliary
388 cancers (59).

389 **Subclonal reconstruction.** We estimated tumour copy number
390 using the Battenberg algorithm (27), with reference data from
391 the 1000 Genomes Project. The estimated fraction of tumour
392 cells (rather than normal tissue) in our cholangiocarcinoma
393 genomes, also known as the tumour purity or cellularity, varied
394 substantially between samples (range 10–90%; median 60%).
395 The overall ploidy per tumour ranged from 1.2–3.7 (median
396 2.0). We then phased somatic variants and assigned them to
397 subclonal lineages (28) using dpclust3p and dpclust (60),
398 implemented in R v.4.3.2. The number of clonal and subclonal
399 lineages, which variants were assigned to, varied from 2–5 per
400 tumour (median 3).

Timing of driver mutations. We applied algorithms to estimate the chronological timing of driver mutations in amplified regions; MutationTimeR (21) and AmplificationTimeR (22), which were implemented in R v.4.3.3. The MutationTimeR algorithm uses a panel of 371 known driver mutations identified by the PCAWG consortium (59). We used the temporal correlation of copy number gains from MutationTimeR to identify if tumours had undergone whole-genome duplication (21). For the AmplificationTimeR analysis, we focused on timing the amplification of 43 driver mutations which have been identified as important in early-stage cholangiocarcinoma (25, 26) or previously detected in these tumours (23). The estimates for the chronological time of amplifications were calibrated using C>T mutations at CpG sites, which have been established to have clock-like properties (61). The age in years at which driver gene amplifications occurred was calculated as the product of the bootstrapped chronological time estimates from AmplificationTimeR (22), which fall in the interval [0, 1], and the patient's age at surgery (23).

Earliest and latest amplified genes. We used a generalised linear model to determine which genes were amplified disproportionately early or late within the lifespan of the tumour. We modelled the chronological time estimates for gene g ($y_{t,g}$), estimated by AmplificationTimeR (22) with a minimum of 10 mutations, which fall in the interval [0,1] using a beta distribution parameterised by a mean μ_g and precision κ

$$y_{t,g} \sim \text{Beta_Proportion}(\mu_g, \kappa), \quad [1]$$

where the mean is a transformed linear function of a gene-specific intercept (α_g) plus covariates (x) and slopes (β). The three binary explanatory variables are the sex of the patient, whether the tumour is intrahepatic or perihilar, and whether the amplification is clonal or subclonal

$$\mu_g = \text{logit}^{-1}(\alpha_g + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3). \quad [2]$$

The model was fitted in a Bayesian framework using the stan language v.2.34.1 (62) implemented with cmdstanr v.0.7.1 (63) in R v.4.3.3; see details on parameter estimation below.

Liver fluke surveys. We obtained data from epidemiological surveys on the liver fluke *O. viverrini* in Thailand between 1980–2017. The parasitological observations in these surveys consisted of i) adult worms obtained from liver dissection at autopsy, ii) adult worms recovered through expulsion following anthelmintic treatment, or iii) parasite egg counts obtained by faecal examination, see Table 1. We contacted the study authors to obtain parasitological data by host age. Where these were unavailable, we simulated individual-level data from summary tables which contained the sample size, mean, and variance for parasitological observations (worm burdens or faecal egg counts) aggregated by age group using a Pearson type I distribution.

Parasite transmission model. We developed a mechanistic model that incorporates key aspects of parasite ecology (20, 30) fitted to individual-level data on adult worm burdens or faecal egg counts (Table 1). Our process model characterises the mean worm burden at the population level (M) by host years of age (a) as an immigration–death process

$$\frac{dM}{da} = \lambda(a) - M\sigma, \quad [3]$$

Survey & Method	Province	District	Year(s)	n^1	Age Range
S1 Autopsy (64)	Khon Kaen	Multiple ²	1982–9	159	2–78
S2 Expulsion (65)	Khon Kaen	Ban Nam	1987	33	15–56
S3 Expulsion (66)	Kalasin	Huai Mek	1989	373	5–60
S4 Faecal egg (67)	Khon Kaen	Chonnabot	1980	1651	1–72
S5 Faecal egg (68)	Multiple	Rural ³	1981–3	433	1–65
S6 Faecal egg (68)	Multiple	Urban ³	1981–3	126	1–70
S7 Prevalence (69)	Khon Kaen	Nong Wai	1981	1284	1–65
S8 Faecal egg (7)	Multiple	Multiple	1994	65	1–70
S9 Prevalence (7)	Multiple	Multiple	1994	1912	1–70
S10 Prevalence (70)	Ratchasima	Multiple	2010–1	1168	5–90
S11 Prevalence (71)	Multiple	Multiple	2013	3916	15–70
S12 Prevalence (72)	Khon Kaen	Khon Kaen	2016–7	387	11–91

Table 1. Cross-sectional surveys analysed in this study investigating the relationship between host age and *Opisthorchis viverrini* worm burden, faecal egg counts, or prevalence by faecal egg diagnostic (Fig. 2A). Surveys were conducted in Northeast (N.E.) Thailand either before the onset of liver fluke control programs (S1–S7; pre-intervention), or following a national control program in the early 1990s (S8–S12; post-intervention). Age range of the participants is shown in years. ¹Sample size of human participants in survey. ²Autopsy cases were from N.E. Thailand, with the majority from Khon Kaen province. ³Hospital-based study where patients from N.E. Thailand were recruited and classified as originating from either rural or urban communities.

where σ gives the spontaneous death rate of adult worms in the absence of anthelmintic treatment. The expected lifespan of adult *O. viverrini* parasites is therefore given as the time taken for half of adult worms to die in the absence of anthelmintic treatment; $\log(2)/\sigma$. As there is evidence for age-dependent reinfection rates (73), we model the *O. viverrini* force of infection as a function of host age

$$\lambda(a) = \eta a e^{-\beta a}. \quad [4]$$

The dynamic model (Equations 3 and 4) has the following analytical solution for worm burden by age a ,

$$M(a) = \frac{\eta e^{-a\sigma}}{(\beta - \sigma)^2} \left(1 + e^{a(\sigma - \beta)} (a(\sigma - \beta) - 1) \right). \quad [5]$$

The cumulative parasite exposure (δ) by age a is given by the definite integral of the force of infection (Equation 4)

$$\delta(a) = \int_0^a \eta a e^{-\beta a} da. \quad [6]$$

Observation model for survey data. The true number of *O. viverrini* adult worms per individual i of age group a ($x_{i,a}$) follows the negative binomial distribution (NB), which takes the form of a gamma–Poisson mixture model parameterised with a mean worm burden $M(a)$ (Equations 3 and 5) and age-dependent dispersion k_a (40). Values of k_a are estimated for each age-group and are themselves normally distributed with an overall mean of μ_k and a standard deviation of σ_k . During autopsy surveys, adult *O. viverrini* were carefully removed from cross-sections of liver and worm recovery is likely close to 100%, therefore we consider that the true worm count for each individual is equal to the recovered worms in autopsy studies ($x_{i,a} \equiv w_{i,a} \mid r = 1$),

$$\begin{aligned} Pr(X_{i,a} = x_{i,a} \mid M(a), k_a) &= \\ NB(x_{i,a} \mid M(a), k_a). & \end{aligned} \quad [7]$$

In surveys where *O. viverrini* flukes were obtained by expulsion, participants were treated with the anthelmintic praziquantel followed by a saline purgative, which is known to result in imperfect subsequent recovery of from faeces (74). Therefore, we allow the true worm burden to be greater than, or equal to, the observed number of worms for individuals in expulsion studies ($x_{i,a} \geq w_{i,a}$) and the probability of observing w_i worms given a true count of x_i is a binomial sampling process with probability of worm recovery $r = 0.44$ (40),

$$Pr(W_{i,a} = w_{i,a} | M(a), k_a, r) = \sum_{x=w_{i,a}}^{\infty} Binom(w_{i,a} | x, r) \cdot NB(x | M(a), k_a), \quad [8]$$

where *Binom* refers to a standard binomial probability distribution. In surveys where the outcome variables are eggs per gram of stool (y), if at least one egg is observed for individual i of age group a ($y_{i,a} \geq 1$), we relate this to the expected egg count for that individual using a negative binomial error distribution, where the mean is given as a density-dependent function of the true worm burden; $\pi(x) = (\Lambda x)^\gamma$ and the dispersion is given by parameter h , which has previously been estimated for *O. viverrini* as 0.4 (40),

$$Pr(Y_{i,a} = y_{i,a} | M(a), k_a, \Lambda, \gamma, h) = \sum_{x=1}^{\infty} NB(y_{i,a} | \pi(x), h) \cdot NB(x | M(a), k_a). \quad [9]$$

Where zero eggs are observed ($y_{i,a} = 0$), we consider the individual diagnostic sensitivity as a saturating function of the worm burden, $se(x) = x/(x + b)$, where the parameter b has been previously estimated for *O. viverrini* as 1.7 (40),

$$Pr(Y_{i,a} = 0 | M(a), k_a, b) = \sum_{x=0}^{\infty} \left(1 - \frac{x}{x + b}\right) \cdot NB(x | M(a), k_a). \quad [10]$$

The population level sensitivity (S) for faecal egg diagnostics is a function of the worm burden distribution at the population level (40),

$$S(M, k) = \sum_{x=1}^{\infty} se(x) \cdot \binom{x+k-1}{x} \frac{\left(\frac{k}{M+k}\right)^k \left(\frac{M}{M+k}\right)^x}{p(M, k)}, \quad [11]$$

where $p(M, k)$ indicates the true prevalence and is given by

$$p(M, k) = 1 - \left(\frac{k}{M+k}\right)^k. \quad [12]$$

Given an assumed diagnostic specificity of one, the observed prevalence p' for age group a is therefore related to the true prevalence with the following expression,

$$p' = p(M, k) \cdot S(M, k) + (1 - p(M, k)) \cdot (1 - sp), \quad [13]$$

where sp gives the faecal egg diagnostic specificity, which is assumed here to be 1. For surveys where only the prevalence is given by faecal egg diagnostic, we represent this as individual-level positive or negative outcomes; $z_{i,a} \in \{0, 1\}$. The probability for the binary diagnostic observations is therefore given by a Bernoulli distribution

$$Pr(Z_{i,a} = z_{i,a} | M(a), k_a, b) = \begin{cases} p'_a & \text{if } z_{i,a} = 1 \\ 1 - p'_a & \text{if } z_{i,a} = 0. \end{cases} \quad [14]$$

Epidemiological parameter estimation. Model fitting was performed in a Bayesian framework using the stan language v.2.34.1 (62) implemented with cmdstanr v.0.7.1 (63) in R v.4.3.3. Parameters were assigned weakly or moderately informative prior distributions based on the results from a previous analysis (40) for the pre-intervention data (S1–S7, see Table 1). For the post-intervention analysis (S8–S12), several model parameter values were taken directly from the pre-intervention analysis posterior distributions; including the parameters relating worm burdens to egg counts (Λ, γ) and the worm mortality rate (σ), which was taken as the highest percentile of the posterior estimate ($\sigma = 0.116$), corresponding to a mean worm lifespan of 6 years, to account for higher parasite mortality resulting from periodic anthelmintic treatment. Each model was run with four parallel chains with a burn-in of 1700 iterations per chain and a total of 1000 iterations. The Gelman–Rubin diagnostic $\hat{r} \leq 1.01$ and effective sample size >500 were used to diagnose successful Markov chain convergence. Results are presented as credible intervals (CrI) of parameter posterior distribution or prediction intervals (PI) from simulations.

Induction and latent period. We performed a time-to-event analysis to validate the induction and latent periods estimated by the earlier evolutionary cancer analysis using population-level data. We obtained the age distribution of 10,737 cases of cholangiocarcinoma diagnosed at Srinagarind Hospital between 1985–2009 (33) which were grouped into age classes. We consider that for each cancer case at age a there is an unknown age at which a driver mutation was obtained m , where $m < a$. Values of m are drawn from a gamma distribution with mean μ_{mut} and shape parameter α_{mut} . Following the mutation at age m there is a latent period before cancer diagnosis at age a , this latent period is also gamma distributed with mean μ_{cca} and shape α_{cca} . The probability of developing cholangiocarcinoma at age a , ϕ_a is therefore

$$\phi_a = \sum_{m=1}^{a-1} Gamma(m | \mu_{mut}, \alpha_{mut}) \cdot Gamma(a - m | \mu_{cca}, \alpha_{cca}). \quad [15]$$

Cholangiocarcinoma cases are reported by age group with a lower bound of l and an upper bound of u in years. The likelihood \mathcal{L} for each observation is therefore optimised on the interval probability, given by

$$\mathcal{L} = 1 - \prod_{a=l}^u (1 - \phi_a). \quad [16]$$

Probability of acquiring cholangiocarcinoma. We adopt a binomial regression framework with the probability of contracting cholangiocarcinoma at age a in years given by

$$\pi_a = \text{logit}^{-1}(\alpha_a + \beta x_f), \quad [17]$$

where α is an age dependent intercept and β is a coefficient multiplied by a binary variable $x_f \in \{0, 1\}$ indicating whether the cases of cholangiocarcinoma come from a region endemic for liver fluke ($x_f = 1$ for Thailand and $x_f = 0$ for Malaysia). Values of α are drawn from a multivariate normal distribution with a covariance matrix which accounts for temporal autocorrelation between years.

Counts of cholangiocarcinoma cases c are reported by age group with a lower bound l and an upper bound u in years

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

and are therefore interval censored. Cases are drawn from an underlying population of size N . We can therefore express a binomial likelihood function for each interval,

$$Pr(C_i^u = c_i^u | N_i^u, \theta_i^u) = \text{Binom}(c_i^u | N_i^u, \theta_i^u), \quad [18]$$

where the interval probability of cancer θ_i^u is related to the yearly probability

$$\theta_i^u = 1 - \prod_{a=l}^u (1 - \pi_a). \quad [19]$$

Data on cholangiocarcinoma cases in Thailand come from the cancer registry at Srinagarind Hospital in Khon Kaen (33). For this analysis we used counts of cholangiocarcinoma cases between 1998–2009 along with corresponding demographic data (N) for Khon Kaen in the same period (75). Cancer case and demographic data for Malaysia between 2007–2009 were obtained from the World Health Organization Mortality Database (36), using version 10 ICD codes C22 (malignant neoplasm of liver and intrahepatic bile ducts) and C24 (malignant neoplasm of other and unspecified parts of biliary tract). We assumed that half of these biliary cancers were attributable to cholangiocarcinoma (17) and, given the very low survival rate, that mortality is a valid approximation for incidence (15, 16).

Acknowledgements. We thank Dr. Stefan Dentro and Prof. Moritz Gerstung for providing additional code relating to the PCAWG consortium analysis to calculate mutation times of 2,778 tumour genomes. We appreciate the comments from Dr. Arporn Wangwiwatsin and Prof. Watcharin Loilome on a poster relating to this manuscript. We further thank Dr. Jessica Clark and Dr. Paul Johnson for discussions on the survival analysis. Dr. Claudio Nunes-Alves also provided valuable feedback on the manuscript.

Access to whole-genome paired cancer and normal sequence data was approved by the International Cancer Genome Consortium Data Access Compliance Office (reference: DACO-63). We acknowledge the important role of Prof. Apinya Jusakul, Prof. Patrick Tan and colleagues for originally generating this sequence data.

Access to the SG10K_Pilot data was approved by the National Precision Medicine Data Access Committee in Singapore (Application no. SG10KP00043). We thank the 'SG10K_Pilot Investigators' for providing the 'SG10K_Pilot data' (EGAD00001005337). The data from the 'SG10K_Pilot Study' reported here were obtained from EGA. This manuscript was not prepared in collaboration with the 'SG10K_Pilot Study' and does not necessarily reflect the opinions or views of the 'SG10K_Pilot Study'.

We used data from the World Health Organization (WHO) Mortality Database for biliary cancer incidence. The analyses, interpretations and conclusions shown in this article are attributable solely to the authors and not to WHO, which is responsible only for the provision of the original information.

This research was funded in whole, or in part, by the Wellcome Trust [Grant number 215919/B/19/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

C.B. acknowledges that this study was performed under the collaborative umbrella of the European Network for the Study of Cholangiocarcinoma (EURO-CHOLANGIO-NET) supported by COST action (CA) 18122 and the Precision-BTC network supported by CA22125. F.V. acknowledges funding from a short term grant from EURO-CHOLANGIO-NET CA18122.

P.S. acknowledges funding from the National Research Council of Thailand as part of the Fluke-Free Thailand program.

Ethical Statement. As an analysis of previously published and anonymous human data, this study met the criteria for exemption from ethical review at the Universities of Oxford and Glasgow.

Data and Code Accessibility. This study uses previously published datasets. Code to reproduce the analysis is available at <https://github.com/tc13/ov-cca-models/>.

Authors' Contributions. T.C. conceptualised the study, T.C. developed methodology, T.C. analysed epidemiological data, T.C. and F.V. analysed cancer genomic data, P.S. provided resources, C.B. provided supervision for cancer genomics, T.D.H. provided supervision for epidemiological modelling, T.C. wrote the manuscript.

Competing Interests. C.B. received honoraria as speaker (AstraZeneca, Incyte) and consultant (Incyte, Servier, Boehringer Ingelheim, AstraZeneca, Jazz Therapeutics, Tahio), received research funds (Avacta, Medannex, Servier) and her spouse is an employee of AstraZeneca. The other authors declare no competing interests.

References

1. Catherine de Martel, Damien Georges, Freddie Bray, Jacques Ferlay, and Gary M Clifford. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet global health*, 8(2):e180–e190, 2020.
2. Peter J Hotez. *Forgotten people, forgotten diseases: the neglected tropical diseases and their impact on global health and development*. John Wiley & Sons, 2021.
3. Anna Borlase, Joaquin M Prada, and Thomas Crellen. Modelling morbidity for neglected tropical diseases: the long and winding road from cumulative exposure to long-term pathology. *Philosophical Transactions of the Royal Society B*, 378(1887):20220279, 2023.
4. Sergi Alonso, Moses Arinaitwe, Alon Atuhaire, Andrina Barungi Nankasi, Joaquin M Prada, Emma McIntosh, and Poppy HL Lamberton. The short-term impact of schistosoma mansoni infection on health-related quality of life: implications for current elimination policies. *Proceedings B*, 291(2024):20240449, 2024.
5. Carl Grundy-Warr, Ross H Andrews, Paiboon Sithithaworn, Trevor N Petney, Banchop Sripa, Luxana Laithavewat, and Alan D Ziegler. Raw attitudes, wetland cultures, life-cycles: socio-cultural dynamics relating to opisthorchis viverrini in the mekong basin. *Parasitology international*, 61(1):65–70, 2012.
6. Ting-Ting Zhao, Yi-Jing Feng, Pham Ngoc Doanh, Somphou Sayasone, Virak Khieu, Chhosak Nithikathkul, Men-Bao Qian, Yuan-Tao Hao, and Ying-Si Lai. Model-based spatial-temporal mapping of opisthorchiasis in endemic countries of southeast asia. *elife*, 10:e59755, 2021.
7. P Jongsuksuntigul and T Imsomboon. The impact of a decade long opisthorchiasis control program in northeastern thailand. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 28(3):551–557, 1997.
8. Apiporn Suwannatnai, Prasert Saichua, and Melissa Haswell. Epidemiology of Opisthorchis viverrini infection. *Advances in Parasitology*, 101:41–67, 2018.
9. Oranard Wattanawong, Sopon Iamsirithaworn, Thongroo Kophachon, Worayuth Nak-Ai, Ampas Wisetmorat, Thitima Wongsaraj, Paron Dekumyong, Chhosak Nithikathkul, Apiporn T Suwannatnai, and Banchob Sripa. Current status of helminthiasis in Thailand: A cross-sectional, nationwide survey, 2019. *Acta Tropica*, 223:106082, 2021.
10. Virak Khieu, Thomas Fürst, Kazuko Miyamoto, Tai-Soon Yong, Jong-Yil Chai, Rekol Huy, Sinuon Muth, and Peter Odermatt. Is opisthorchis viverrini emerging in cambodia? *Advances in parasitology*, 103:31–73, 2019.
11. Bong-Kwang Jung, Sooji Hong, Taehee Chang, Jaeun Cho, Seungwon Ryoo, Keon Hoon Lee, Jeongyu Lee, Woon-Mok Sohn, Sung-Jong Hong, Virak Khieu, et al. High endemicity of opisthorchis viverrini infection among people in northern cambodia confirmed by adult worm expulsion. *Scientific Reports*, 13(1):9654, 2023.
12. V Vatanasapt, N Martin, H Sriplung, K Chindavijjak, S Sontipong, H Sriamporn, DM Parkin, and J Ferlay. Cancer incidence in thailand, 1988-1991. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 4(5):475–483, 1995.
13. Andrea A Florio, Jacques Ferlay, Ariana Znaor, David Ruggieri, Christian S Alvarez, Mathieu Laversanne, Freddie Bray, Katherine A McGlynn, and Jessica L Petrick. Global incidence and trends in intra-and extrahepatic cholangiocarcinoma from 1993 to 2012. *Cancer*, 126(11):2666, 2020.
14. William R Jarnagin, Yuman Fong, Ronald P DeMatteo, Mithat Gonen, Edmund C Burke, Jessica Bodniewicz, Miranda Youssef, David Klimstra, and Leslie H Blumgart. Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma. *Annals of surgery*, 234(4):507–519, 2001.
15. Juan Valle, Harpreet Wasan, Daniel H Palmer, David Cunningham, Alan Anthoney, Anthony Maraveyas, Srinivasan Madhusudan, Tim Iveson, Sharon Hughes, Stephen P Pereira, et al. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *New England Journal of Medicine*, 362(14):1273–1281, 2010.
16. Sumera Rizvi, Shahid A Khan, Christopher L Hallemeier, Robin K Kelley, and Gregory J Gores. Cholangiocarcinoma—evolving concepts and therapeutic strategies. *Nature reviews Clinical oncology*, 15(2):95–111, 2018.
17. Banchob Sripa, Paul J Brindley, Jason Mulvenna, Thewarach Laha, Michael J Smout, Eimorn Mairiang, Jeffrey M Bethony, and Alex Loukas. The tumorigenic liver fluke Opisthorchis viverrini—multiple pathways to cancer. *Trends in Parasitology*, 28(10):395–407, 2012.
18. Thomas Crellen, Paiboon Sithithaworn, Opal Pitaksakulrat, Narong Khuntikeo, Graham F Medley, and T Déirdre Hollingsworth. Towards evidence-based control of Opisthorchis viverrini. *Trends in Parasitology*, 37(5):370–380, 2021.
19. Narong Khuntikeo, Bandit Thinkhamrop, Thomas Crellen, Chatanun Eamudomkarn, Trevor N Petney, Ross H Andrews, and Paiboon Sithithaworn. Epidemiology and control of opisthorchis viverrini infection: Implications for cholangiocarcinoma prevention. In *Liver Fluke, Opisthorchis viverrini Related Cholangiocarcinoma: Liver Fluke Related Cholangiocarcinoma*, pages 27–52. Springer, 2023.
20. GF Medley and DA Bundy. Dynamic modeling of epidemiologic patterns of schistosomiasis morbidity. *The American journal of tropical medicine and hygiene*, 55(5 Suppl):149–158, 1996.
21. Moritz Gerstung, Clemency Jolly, Ignaty Leshchiner, Stefan C Dentro, Santiago Gonzalez, Daniel Rosebrock, Thomas J Mitchell, Yulia Rubanova, Pavana Anur, Kaixian Yu, et al. The evolutionary history of 2,658 cancers. *Nature*, 578(7793):122–128, 2020.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

- 731 22. G Maria Jakobsdottir, Stefan C Dento, Robert G Bristow, and David C Wedge. Amplification-
732 timer: An R package for timing sequential amplification events. *Bioinformatics*, page btae281,
733 2024.
- 734 23. Apinya Jusakul, Ioana Cutcutache, Chern Han Yong, Jing Quan Lim, Mi Ni Huang, Nisha
735 Padmanabhan, Vishwa Nellore, Sarinya Kongpetch, Alvin Wei Tian Ng, Ley Moy Ng, et al.
736 Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarci-
737 noma. *Cancer discovery*, 7(10):1116–1135, 2017.
- 738 24. Benjamin Goepfert, Trine Folseraas, Stephanie Roessler, Matthias Kloor, Anna-Lena Volckmar,
739 Volker Endris, Ivo Buchhalter, Albrecht Stenzinger, Krzysztof Grzyb, Marit M Grimsrud, et al.
740 Genomic characterization of cholangiocarcinoma in primary sclerosing cholangitis reveals
741 therapeutic opportunities. *Hepatology*, 72(4):1253–1266, 2020.
- 742 25. Benjamin Goepfert, Damian Stichel, Reka Toth, Sarah Fritzsche, Moritz Anton Loeffler,
743 Anna Melissa Schlitter, Olaf Neumann, Yassen Assenov, Monika Nadja Vogel, Arianeb Mehrabi,
744 et al. Integrative analysis reveals early and distinct genetic and epigenetic changes in intraductal
745 papillary and tubulopapillary cholangiocarcinogenesis. *Gut*, 71(2):391–401, 2022.
- 746 26. Khaa Hoo Ong, Yao-Yu Hsieh, Hong-Yue Lai, Ding-Ping Sun, Tzu-Ju Chen, STEVEN KUAN-
747 HUANG, Yu-Feng Tian, Chia-Lin Chou, Yow-Ling Shiue, Hung-Chang Wu, et al. Lamc2 is
748 a potential prognostic biomarker for cholangiocarcinoma. *Oncology Letters*, 26(6):1–14, 2023.
- 749 27. Serena Nik-Zainal, Peter Van Loo, David C Wedge, Ludmil B Alexandrov, Christopher D
750 Greenman, King Wai Lau, Keiran Raine, David Jones, John Marshall, Manasa Ramakrishna,
751 et al. The life history of 21 breast cancers. *Cell*, 149(5):994–1007, 2012.
- 752 28. Stefan C Dento, David C Wedge, and Peter Van Loo. Principles of reconstructing the subclonal
753 architecture of cancers. *Cold Spring Harbor perspectives in medicine*, 7(8):a26625, 2017.
- 754 29. DJ Shaw, BT Grenfell, and AP Dobson. Patterns of macroparasite aggregation in wildlife host
755 populations. *Parasitology*, 117(6):597–610, 1998.
- 756 30. Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*.
757 Oxford University Press, 1991.
- 758 31. Jae Hoon Lim. Liver flukes: the malady neglected. *Korean journal of radiology*, 12(3):269–279,
759 2011.
- 760 32. HD Attwood and ST Chou. The longevity of *Clonorchis sinensis*. *Pathology*, 10(2):153–156,
761 1978.
- 762 33. Supot Kamsa-ard, Surapon Wiangnon, Krittika Suwanrungrung, Supanee Promthet, Narong
763 Khuntikeo, Siriporn Kamsa-ard, Suwannee Mahaweerawat, et al. Trends in liver cancer
764 incidence between 1985 and 2009, khon kaen, thailand: cholangiocarcinoma. *Asian Pac J*
765 *Cancer Prev*, 12(9):2209–2213, 2011.
- 766 34. George Psevdos, Florence M Ford, and Sung-Tae Hong. Screening US vietnam veterans for
767 liver fluke exposure 5 decades after the end of the war. *Infectious Diseases in Clinical Practice*,
768 26(4):208–210, 2018.
- 769 35. Theodore E Nash, David Sullivan, Edward Mitre, Eric Garges, Victoria J Davey, Gary Roselle,
770 Stephen J Davies, and Peter D Rumm. Comments on "Screening US Vietnam veterans for liver
771 fluke exposure 5 decades after the end of the war". *Infectious Diseases in Clinical Practice*, 26
772 (4):240–241, 2018.
- 773 36. World Health Organization. website, February 2024. URL [https://www.who.int/data/
774 data-collection-tools/who-mortality-database](https://www.who.int/data/data-collection-tools/who-mortality-database). WHO Mortality Database.
- 775 37. Banchob Sripa, Apiporn T Suwannatnai, Sompou Sayasone, Dung Trung Do, Virak Khieu, and
776 Yichao Yang. Current status of human liver fluke infections in the greater mekong subregion.
777 *Acta tropica*, 224:106133, 2021.
- 778 38. D Maxwell Parkin, Petcharin Srivatanakul, Myriam Khlai, Dhiraphol Chenvidhya, Pornarong
779 Chotiwan, Somchai Insiripong, Kristan A L'abbé, and Christopher P Wild. Liver cancer in
780 thailand. i. a case-control study of cholangiocarcinoma. *International journal of cancer*, 48(3):
781 323–328, 1991.
- 782 39. Nopparat Songserm, Supanee Promthet, Paiboon Sithithaworn, Chamsai Pientong, Tipaya
783 Ekaklaksananan, Peechanika Chopjitt, and Donald Maxwell Parkin. Risk factors for cholangio-
784 carcinoma in high-risk area of thailand: role of lifestyle, diet and methylenetetrahydrofolate
785 reductase polymorphisms. *Cancer epidemiology*, 36(2):e89–e94, 2012.
- 786 40. Thomas Crellen, Melissa Haswell, Paiboon Sithithaworn, Sompou Sayasone, Peter Odermatt,
787 Poppy HL Lambertson, Simon EF Spencer, and T Déirdre Hollingsworth. Diagnosis of helminths
788 depends on worm fecundity and the distribution of parasites within hosts. *Proceedings of the*
789 *Royal Society B*, 290(1991):20222204, 2023.
- 790 41. Melissa R Haswell-Elkins, Soisungwan Satarug, Mitsuhiro Tsuda, Eimorn Mairiang, Hiroyasu
791 Esumi, Paiboon Sithithaworn, Pisaln Mairiang, Minoru Saitoh, Puangrat Yongvanit, and David B
792 Elkins. Liver fluke infection and cholangiocarcinoma: model of endogenous nitric oxide
793 and extragastric nitrosation in human carcinogenesis. *Mutation Research/Fundamental and*
794 *Molecular Mechanisms of Mutagenesis*, 305(2):241–252, 1994.
- 795 42. Joanne P Webster, David H Molyneux, Peter J Hotez, and Alan Fenwick. The contribution of
796 mass drug administration to global health: past, present and future. *Philosophical Transactions*
797 *of the Royal Society B: Biological Sciences*, 369(1645):20130434, 2014.
- 798 43. Eimorn Mairiang, Melissa R Haswell-Elkins, Pisaln Mairiang, Paiboon Sithithaworn, and David B
799 Elkins. Reversal of biliary tract abnormalities associated with *Opisthorchis viverrini* infection
800 following praziquantel treatment. *Transactions of the Royal Society of Tropical Medicine and*
801 *Hygiene*, 87(2):194–197, 1993.
- 802 44. S Pungpak, C Viravan, B Radomyos, K Chalermrut, C Yemput, W Ploksawasadi, M Ho,
803 T Harinasuta, and D Bunnag. *Opisthorchis viverrini* infection in thailand: studies on the
804 morbidity of the infection and resolution following praziquantel treatment. *The American journal*
805 *of tropical medicine and hygiene*, 56(3):311–314, 1997.
- 806 45. Eimorn Mairiang, Thewarach Laha, Sasithorn Kaewkes, Alex Loukas, Jeffrey Bethony, Paul J
807 Brindley, and Banchob Sripa. Hepatobiliary morbidities detected by ultrasonography in
808 *Opisthorchis viverrini*-infected patients before and after praziquantel treatment: a five-year
809 follow up study. *Acta tropica*, 217:105853, 2021.
- 810 46. Sue J Goldie, Daniel Grima, Michele Kohli, Thomas C Wright, Milton Weinstein, and Eduardo
811 Franco. A comprehensive natural history model of hpv infection and cervical cancer to estimate
812 the clinical impact of a prophylactic hpv-16/18 vaccine. *International journal of cancer*, 106(6):
813 896–904, 2003.
- 814 47. Do-Youn Oh, Aiwu Ruth He, Shukui Qin, Li-Tzong Chen, Takuji Okusaka, Arndt Dugel,
815 Jin-Won Kim, Thatthan Suksoombooncharoen, Myung Ah Lee, Masayuki Kitano, et al. Durval-
816 umab plus gemcitabine and cisplatin in advanced biliary tract cancer. *NEJM evidence*, 1(8):
817 EVID0a2200015, 2022.
- 818 48. Margaretha A Vink, Johannes A Bogaards, Folkert J van Kemenade, Hester E de Melker,
819 Chris JLM Meijer, and Johannes Berkhof. Clinical progression of high-grade cervical intraepi-
820 thelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored
821 national registry data. *American journal of epidemiology*, 178(7):1161–1169, 2013.
- 822 49. William M Lee. Hepatitis b virus infection. *New England journal of medicine*, 337(24):1733–
823 1745, 1997.
- 824 50. SA Frank and P Schmid-Hempel. Mechanisms of pathogenesis and the evolution of parasite
825 virulence. *Journal of evolutionary biology*, 21(2):396–404, 2008.
- 826 51. Kavin Thinkhamro, Narong Khuntikeo, Nittaya Chamadol, Apiporn T Suwannatnai, Surachai
827 Phimha, and Matthew Kelly. Associations between ultrasound screening findings and cholan-
828 giocarcinoma diagnosis in an at-risk population. *Scientific Reports*, 12(1):13513, 2022.
- 829 52. Pheonepong Ayé Soukhathammavong, Virasack Rajpho, Khampheng Phongluxa, Youthanav-
830 anhan Vonghachack, Jan Hattendorf, Bouasay Hongvanthong, Oroth Rasaphon, Banchob Sripa,
831 Kongsap Akkhavong, Christoph Hatz, et al. Subtle to severe hepatobiliary morbidity in
832 opisthorchis viverrini endemic settings in southern laos. *Acta tropica*, 141:303–309, 2015.
- 833 53. Felix Krueger. URL <https://github.com/FelixKrueger/TrimGalore>. Github.
- 834 54. Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows–wheeler
835 transform. *bioinformatics*, 25(14):1754–1760, 2009.
- 836 55. Geraldine A Van der Auwera and Brian D O'Connor. *Genomics in the cloud: using Docker,*
837 *GATK, and WDL in Terra*. O'Reilly Media, 2020.
- 838 56. PCAWG. URL https://dcc.icgc.org/releases/PCAWG/consensus_snv_indel. DACO.
- 839 57. Broad Institute. URL <https://console.cloud.google.com/storage/browser/gatk-best-practices>.
840 GATK Best Practices.
- 841 58. Degang Wu, Jinzhuang Dou, Xiaoran Chai, Claire Bellis, Andreas Wilim, Chih Chuan Shih,
842 Wendy Wei Jia Soon, Nicolas Bertin, Clarabelle Bitong Lin, Chiea Chuen Khor, et al. Large-
843 scale whole-genome sequencing of three diverse asian populations in singapore. *Cell*, 179(3):
844 736–749, 2019.
- 845 59. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis
846 of whole genomes. *Nature*, 578(7793):82–93, 2020.
- 847 60. Niccolo Bolli, Hervé Avet-Loiseau, David C Wedge, Peter Van Loo, Ludmil B Alexandrov,
848 Inigo Martincorena, Kevin J Dawson, Francesco Iorio, Serena Nik-Zainal, Graham R Bignell,
849 et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature*
850 *communications*, 5(1):2997, 2014.
- 851 61. Zane Koch, Adam Li, Daniel S Evans, Steven Cummings, and Trey Ideker. Somatic mutation
852 as an explanation for epigenetic aging. *bioRxiv*, 2023.
- 853 62. Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael
854 Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic
855 programming language. *Journal of Statistical Software*, 76(1), 2017.
- 856 63. Jonah Gabry, Rok Češnovar, and Andrew Johnson. *cmdstanr: R Interface to 'CmdStan'*, 2023.
857 <https://mc-stan.org/cmdstanr/>, <https://discourse.mc-stan.org>.
- 858 64. Paiboon Sithithaworn, Smarn Tesana, Vichit Pipitgool, Sasithorn Kaewkes, Kasemsri Thaik-
859 lar, Chavalit Pairojkul, Bunchob Sripa, Anucha Paupairoj, Pisit Sanpitak, and Chanarong
860 Aranyanat. Quantitative post-mortem study of *Opisthorchis viverrini* in man in North-East
861 Thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 85(6):765–768,
862 1991.
- 863 65. Roger J Ramsay, Paiboon Sithithaworn, Paul Prociw, Douglas E Moorhouse, and Chingchai
864 Methaphat. Density-dependent fecundity of *Opisthorchis viverrini* in humans, based on faecal
865 recovery of flukes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83(2):
866 241–242, 1989.
- 867 66. MR Haswell-Elkins, DB Elkins, Paiboon Sithithaworn, Phattara Treesarawat, and Sasithorn
868 Kaewkes. Distribution patterns of *Opisthorchis viverrini* within a human community. *Parasitology*,
869 103(1):97–101, 1991.
- 870 67. ES Upatham, V Viyanant, S Kurathong, J Rojborwonwitaya, WY Brockelman, S Ardsungnoen,
871 P Lee, and S Vajrasthira. Relationship between prevalence and intensity of *Opisthorchis viverrini*
872 infection, and clinical symptoms and signs in a rural community in north-east thailand. *Bulletin*
873 *of the World Health Organization*, 62(3):451, 1984.
- 874 68. Sucha Kurathong, Pravit Lerdrerasirikul, Virasak Wongpaatoon, Chutima Pramoolsinsap, and
875 E Suchart Upatham. *Opisthorchis viverrini* infection in rural and urban communities in northeast
876 thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81(3):411–414,
877 1987.
- 878 69. S Sornmani, FP Schelp, P Vivatanasesth, W Patihatakorn, P Impand, P Sitabutra, P Worasan,
879 and S Preuksaraj. A pilot project for controlling o. viverrini infection in nong wai, northeast
880 thailand, by applying praziquantel and other measures. *Arzneimittel-forschung*, 34(9B):1231–
881 1234, 1984.
- 882 70. Soraya J Kaewpitoon, Ratana Rujirakul, and Natthawut Kaewpitoon. Prevalence of *Opisthorchis*
883 *viverrini* infection in nakhon ratchasima province, northeast thailand. *Asian Pacific Journal of*
884 *Cancer Prevention*, 13(10):5245–5249, 2012.
- 885 71. Kesorn Thaeuwongiew, Seri Singthong, Saowalux Kutchamart, Sasithorn Tangsawad, Su-
886 pannee Promthet, Supan Sailugkum, and Narong Wongba. Prevalence and risk factors for
887 *Opisthorchis viverrini* infections in upper northeast thailand. *Asian Pacific Journal of Cancer*
888 *Prevention*, 15(16):6609–6612, 2014.
- 889 72. Pokkamol Laoraksawang, Oranuch Sanpool, Rutchanee Rodpai, Tongjit Thanchomrang,
890 Wanida Kanarkard, Wanchai Maleewong, Rattaphol Kraiklang, and Pewpan M Intapan. Cur-
891 rent high prevalences of strongyloides stercoralis and *Opisthorchis viverrini* infections in rural
892 communities in northeast thailand and associated risk factors. *BMC Public Health*, 18:1–11,
893 2018.
- 894 73. ES Upatham, V Viyanant, WY Brockelman, S Kurathong, P Lee, and R Kraengraeng. Rate of re-
895 infection by *Opisthorchis viverrini* in an endemic northeast Thai community after chemotherapy.
896 *International Journal for Parasitology*, 18(5):643–649, 1988.
- 897 74. Chenghua Shen, Jae-hwan Kim, Jeong-Keun Lee, Young Mee Bae, Min-Ho Choi, Jin-Kyoung
898 Oh, Min Kyung Lim, Hai-Rim Shin, and Sung-Tae Hong. Collection of *Clonorchis sinensis* adult
899 worms from infected humans after praziquantel treatment. *The Korean Journal of Parasitology*,
900 45(2):149, 2007.
- 901 75. Thomas Brinkhoff. website, February 2024. URL <http://www.citypopulation.de>. City Population.