Molecular drivers, mutational sources, and preventative measures for cholangiocarcinoma

Jeff Mandell

Program in Computational Biology and Bioinformatics, Yale University

Advisor: Dr. Jeffrey Townsend, Elihu Professor of Biostatistics and Professor of Ecology & Evolutionary Biology, Yale School of Public Health, Yale University

Data and reproducible analysis

Introduction

Cholangiocarcinoma (CCA)—carcinoma of the bile ducts—has intrahepatic, perihilar, and distal subtypes based on proximity to the liver. Primary sclerosing cholangitis and inflammatory bowel disease are major risk factors for CCA, although around 50% of incidence is in patients with no known risk factors, and prognosis is typically poor[1]. To better understand the somatic mutational landscape of cholangiocarcinoma subtypes, we aggregated somatic variant calls from thirteen peer-reviewed publications with Yale-sequenced tumors to produce the largest data set to be assembled to date for cholangiocarcinoma sequencing, comprising 1,211 patient samples (**Table 1**).

Author or institution	Identifier	iCCA	pCCA	dCCA	eCCA
Borad	10.1371/journal.pgen.1004135	6	0	0	0
Chan-on	10.1038/ng.2806	2	0	0	2
Gao	10.1053/j.gastro.2014.01.062	6	0	0	0
Jiao	10.1038/ng.2813	31	0	0	0
Jusakul	10.1158/2159-8290.CD-17-0368	152	76	7	7
Memorial Sloan Kettering	ihch_msk_2021 (cBioPortal)	376	0	0	0
Nakamura	10.1038/ng.3375	122	41	29	0
Nepal	10.1002/hep.29764	12	0	0	0
Sia	10.1038/ncomms7087	8	0	0	0
TCGA	CHOL (release 33.1)	44	0	0	6
Wardell	10.1016/j.jhep.2018.01.009	44	24	28	0
Yale	Unpublished	7	1	6	4
Zhang	10.1038/s41467-022-30708-7	24	43	0	0
Zou	10.1038/ncomms6696	103	0	0	0
Totals		**937**	**185**	**70**	**19**

Table 1

Patient counts by data source and cholangiocarcinoma subtype: intrahepatic (iCCA), perihilar (pCCA), distal (dCCA), unspecified extrahepatic (eCCA).

Previous studies have typically identified CCA driver mutations by testing for statistical enrichment based on an assumption that all mutations are equally likely. However, mutation rate varies across the genome by orders of magnitude[2]; furthermore, due to differences in endogenous and exogenous mutagenic exposures, patterns of mutations are not equivalent across patients (i.e., some mutations are much more likely in some patients than others). The cancereffectsizeR software package provides methods to calculate sample- and site-specific mutation rates for single-base substitutions (SBS) in tumor sequencing data sets[3]. These mutation rates can be used under models of evolutionary genetic selection to estimate scaled selection coefficients (also called cancer effects), which quantify the contributions of mutations to proliferation and survival of cancer cell lineages. Cancer effect is a useful measure of cancer relevance: Variants of known cancer relevance have substantially higher cancer effects than most other variants (**Figure 1**), and cancer effect is a better predictor of cancer relevance than mutational prevalence or protein function impact scores[3].

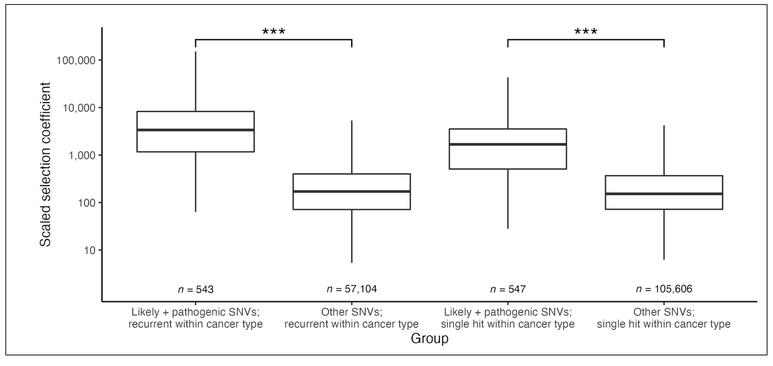

Figure 1

Box plots of the cancer effects of variants appearing in two or more patients across eight TCGA cohorts, separated by ClinVar somatic pathogenic annotation and recurrence within cohort. Estimates are specific to cancer type; variants appearing in multiple cohorts appear as multiple estimates. Reprinted from Mandell, Cannataro, and Townsend[3].

Results

We used cancereffectsizeR (v2.10.0) to calculate mutation rates and cancer effects in cholangiocarcinoma. We also conducted a bootstrapped mutational signature analysis to identify patterns of trinucleotide-context-specific substitution and associate them with mutagenic processes. Mutations in drivers *KRAS*, *IDH1*, *TP53,* and *ERBB2* had high cancer effects across subtypes; genes with subtype-specific high-effect mutations included *NRAS, BRAF,* and *ARAF* in intrahepatic, *NOTCH2NLA* in perihilar, and *SH2B2* in distal (**Figure 2**). Signatures associated with chemotherapeutic treatments contributed to mutation burden in all subtypes; the most substantial contribution in this category was SBS87 (thiopurine treatment), which was significantly enriched in the distal subtype (11.9% vs. 6.0% and 6.3% in other subtypes; **Figure 3**). Using our mutational signature attributions, we calculated the relative contribution of each mutational signature to cancer effect. Signatures associated with treatment (including SBS87) and environmental exposures (SBS42, haloalkane exposure; SBS24, aflatoxin exposure) contributed substantially to cancer effect in all subtypes (**Figure 4**).

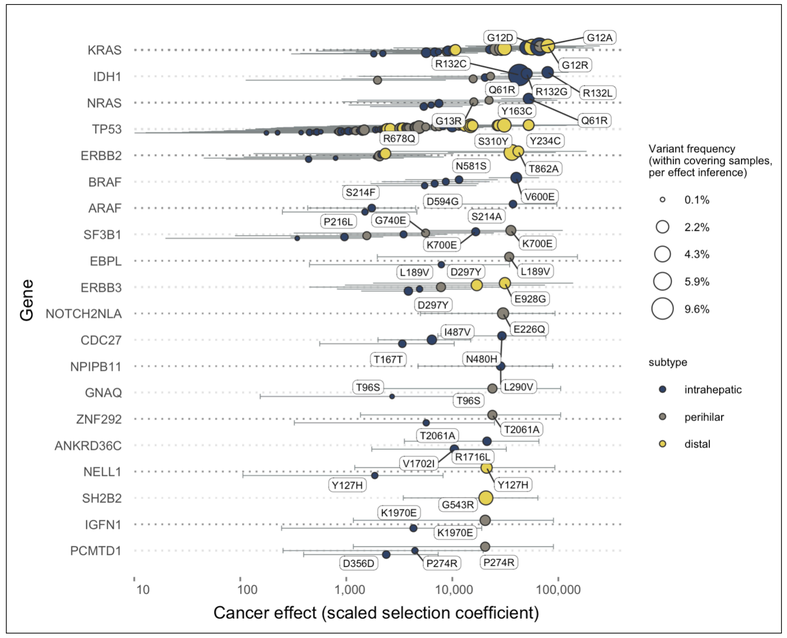

Figure 2

Cancer effects of recurrent coding substitutions calculated within cholangiocarcinoma subtypes. The top twenty genes (ordered by the highest-effect recurrent variant in each) are depicted.

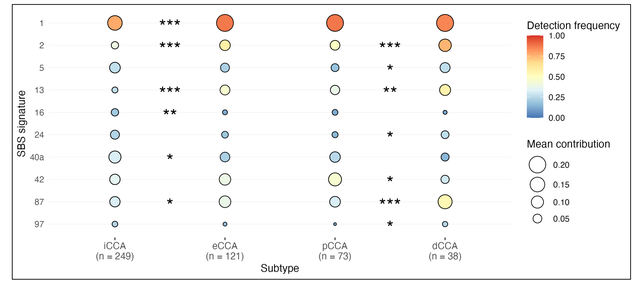

Figure 3

Mean fraction of SBS attributed to signatures (circle size) and frequency of signature detection (color) across one thousand bootstrapped signature extractions in exome and genome-sequenced samples with at least 50 non-recurrent SNVs. Asterisks indicate significantly different mean contributions (Mann-Whitney U test). *, $P < 0.05$; **, $P < .01$; ***, $P < .001$. Signatures with low detection and low mean contribution not shown.

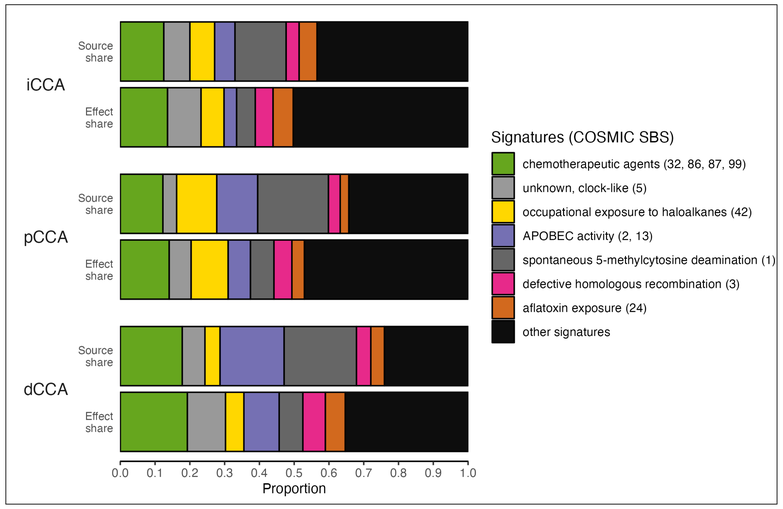

Figure 4

Proportional stacked bar plots of mutational source shares and cancer effect shares averaged by subtype across samples eligible for signature analysis. Cancer effects calculated using all samples in each subtype.

Discussion

This analysis produced a ranking of driver mutations in CCA subtypes by their estimated contributions to cancer initiation and progression. High-effect mutations, regardless of their frequency in cancer cohorts, can explain disease trajectories in the patients in which they are found, and they should be considered for further research and evaluated as drug targets.

Our mutational signature analysis quantified the contributions of various mutational sources to CCA and identified that a substantial proportion of disease-relevant CCA mutations come from actionable sources. Prolonged use of thiopurines and other chemotherapeutics used to treat cancers and autoimmune conditions have been previously associated with increased risk for cancer[4]. The present analysis suggests that the risk of promoting oncogenic progression in bile duct tissues should be considered when determining treatment strategies for inflammatory bowel disease, primary sclerosing cholangitis, and other gastrointestinal autoimmune conditions. The haloalkane exposure signature (SBS42) also emerged as a notable cancer effect source. This signature was originally discovered in a Japanese cohort of early-onset cholangiocarcinoma cases that was traced to occupational chemical exposures in the printing industry[5]. The presence of the signature in this broader data set suggests that efforts to reduce human exposure to haloalkanes may reduce CCA incidence. The same logic applies to detection and elimination of aflatoxin contamination in the food supply.

Future work will include analysis of co-occurrence patterns among high-effect mutations to identify distinct molecular disease profiles, which may or may not be shared among liver-proximity subtypes. Such profiles may be used to predict treatment response and survival and to guide treatment approaches. While somatic substitutions contribute to CCA disease progression via many biological pathways, other classes of genomic alterations also play a role in the mutational landscape. For example, intrahepatic CCA frequently carries *FGFR2* gene fusions. It is an ongoing project to develop methods for incorporating diverse classes of somatic alterations into somatic evolutionary models that could more completely represent the mutational landscapes of oncogenesis.

Methods

- Collected tumor sequencing data—a mix of whole-genome, whole-exome, and targeted gene (panel) sequencing—from thirteen sources. Where available, we used previously published somatic variant calls. Where not available, we aligned raw sequencing reads to the GRCh38 human genome build and called mutations using a GATK/Mutect2 workflow.
- Determined the targeted genomic regions for all exome and panel sequencing sources, in most cases by acquiring BED files from kit manufacturers. Out-of-coverage mutations were filtered out of analysis, as were mutations at sites of known germline variation or in low-complexity regions (with the exception of cancer-related variants annotated by COSMIC).
- Ran MutationalPatterns on bootstrapped mutation profiles to attribute mutations to COSMIC v3.4 mutational signatures. For each sample, average attributions were calculated across 1,000 bootstraps and compared among subtypes. Samples with fewer than 50 non-recurrent mutations (and all targeted sequencing samples) were excluded from signature analysis.
- Used cancereffectsizeR to estimate mutation rates and calculate subtype-specific cancer effects. We then attributed cancer effects to mutational signatures using the *mutational_signature_effects()* function.

References

1. Khan SA, Tavolari S, Brandi G, 2019. Cholangiocarcinoma: Epidemiology and risk factors. Liver Int. https://doi.org/10.1111/liv.14095
2. Lawrence, MS et al., 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. https://www.nature.com/articles/nature12213
3. Mandell, JD, Cannataro, VL., Townsend, J., 2023. Estimation of Neutral Mutation Rates and Quantification of Somatic Variant Selection Using cancereffectsizeR. Cancer Research. https://doi.org/10.1158/0008-5472.CAN-22-1508
4. Zheng KYC, Guo C-G, Wong IOL, et al., 2020. Risk of malignancies in patients with inflammatory bowel disease who used thiopurines as compared with other indications: a territory-wide study. Therapeutic Advances in Gastroenterology. https://doi.org/10.1177/1756284820967275
5. Mimaki S et al., 2016. Hypermutation and unique mutational signatures of occupational cholangiocarcinoma in printing workers exposed to haloalkanes. Carcinogenesis. https://doi.org/10.1093/carcin/bgw066