Defining incidence and complications of fibrolamellar liver cancer through tiered computational analysis of multiple clinical data types.

John D. Gordan (john.gordan@ucsf.edu)  
UCSF

Travis Zack  
University of California, San Francisco  
https://orcid.org/0000-0002-1620-6455

Alan Venook  
UCSF

Jennifer Wild  
University of California, San Francisco

Atul Butte  
University of California, San Francisco

Samantha Maisel  
University of California, Los Angeles

Amin Yaqubie  
Memorial Sloan Kettering Cancer Center

Ghassan Abou-Alfa  
Memorial Sloan Kettering Cancer Center

Jennifer J Knox  
Princess Margaret Cancer Centre

Robert Mayer  
Dana-Farber Cancer Institute

Allison O'Neill  
Dana-Farber Cancer Institute

Brief Communication

Keywords:

Posted Date: September 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2082118/v1
Abstract

Importance:

The incidence and biochemical consequences of rare tumor subtypes are often hard to study. Fibrolamellar liver cancer (FLC) is one such rare malignancy affecting adolescents and young adults. To better characterize the incidence and biochemical consequences of this disease, we combined a comprehensive analysis of electronic medical record and national payee data and find that FLC incidence is likely 6–8 times higher than previous estimates. By deploying unsupervised learning modes on clinical laboratory data on patients with hyperammonemia, we find that FLC-associated hyperammonemia mirrors metabolic dysregulation in Urea Cycle Disorders. We demonstrate that advanced computational analysis of rich clinical datasets can provide key clinical and biochemical insights into rare cancers.

Introduction

FLC is a rare, primary liver cancer predominantly found in adolescents and young adults (AYA) without underlying liver disease\(^1\). Previously considered a morphological variant of hepatocellular carcinoma (HCC), FLC is now understood to be a separate entity with distinct molecular, histological, demographic, and clinical characteristics\(^2,3\). While systemic treatments approved for HCC are sometimes used for FLC, the significant molecular and mechanistic differences between the diseases suggests that each should have a distinct treatment paradigm\(^4\).

The most comprehensive SEER-based study of FLC reported an annual incidence of 0.02 per 100,000\(^5,6\). This rate is roughly 5 to 20-fold lower than might be expected based on clinical practice observations, which suggest that FLC incidence is approximately 1% that of HCC\(^7-9\). Also, SEER data show a bimodal distribution in the age of FLC patients, with peaks at 15–19 years and 70–74 years, whereas disease experts find no other evidence for a major cluster of FLC patients above the AYA age range\(^4,10,11\). These older patients may instead have a recently defined variant of conventional HCC, marked by BAP1 mutation and lacking the FLC-associated DNAJB1-PRKACA fusion\(^12\). This suggests that underdiagnosis and misclassification both impact the registry data available for FLC through SEER, and may hinder further attempts at precision-guided treatment of this rare diagnosis.

Even though most FLC patients have normal underlying liver parenchyma, hyperammonemia is a frequent complication of the disease. Rather than liver insufficiency, hyperammonemia in FLC is thought to be a paraneoplastic syndrome associated with Urea Cycle metabolism dysfunction\(^13\), though this mechanism remains unproven.

Applying a multimodal approach to comprehensive medical records from a large medical center, combined with a large insurance claim dataset, we have better elucidated national incidence of FLC and a potential unique biochemical mechanism for hyperammonemia in these patients.
Results & Conclusions

Improved estimation of FLC incidence rates through application of biological knowledge of disease to large Payer dataset

To model the national incidence of FLC, we identified all patients with ICD-10-CM coding of “liver cell carcinoma” (LC, C22.0) or “other specified carcinoma of the liver” (C22.7) in the Komodo dataset. To enrich this cohort for FLC patients, we excluded patients with any of 102 ICD codes associated with chronic liver inflammation (Table S1, Figure 1A). We confirmed 33 cases of FLC out of 4300 patients with ICD diagnoses of LC (0.6% of overall cases; Figure 1A, Supplementary Methods) within UCSF’s EMR between 2012 and 2021 through direct validation of clinical and pathology notes. Only 4 of these patients had ICD codes from our exclusion criteria, with “Other Liver Cancer”, hepatoblastoma, bile duct carcinoma or sarcomas of the liver, occurring in 3 out of the 4 patients.

All UCSF FLC patients had an age at diagnosis below 50 years (mean 25.5 ± 8.3) whereas 87% of UCSF HCC patients were diagnosed at ages over 50 years (mean 61.96 ± 13.9). FLC accounted for 21% of LC cases in patients under 50 years old and 50% in patients under 30 years old without ICD-documented chronic liver inflammatory conditions. We used a Bayesian statistical method to map the relative rates of FLC and conventional HCC in the UCSF medical record onto the Komodo data set to predict national incidence based on age and clinical covariates. After scaling, we calculated to an annual FLC incidence of 602 +/- 110 cases (Fig. 1D), about 0.185/100,000 individuals within the US, 9x the previous estimate using SEER data.

Hyperammonemia

Prospective analysis in 2 recent FLC clinical trials showed 14.3-50% of tested patients with advanced FLC have elevated serum ammonia at baseline, with 27.3-71.3% having hyperammonemia at end of treatment (Fig. 2A). Within the Komodo dataset, 372 or 8% of patients coded with LC without inflammatory liver disease also had a diagnosis code for high ammonia levels (Fig. 2B). Analysis of FLC patients with extreme hyperammonemia has suggested an association with acquired urea cycle deficiency (UCD)\(^{19,20}\). We hypothesized that patients with hyperammonemia would partition into different metabolic “states” based on their clinical laboratory values. Using methods commonly used to identify similar partitions in cell lineages through single cell RNA-seq analysis, we further hypothesized that patients with hyperammonemia and FLC would partition differently from those with hyperammonemia due to cHCC. Of 3066 patient encounters with elevated ammonia levels and a CMP in the UCSF EMR, 19 were from FLC patients, 295 from cHCC patients and 442 from UCD patients. We identified the natural partitions of lab values with unsupervised clustering (Fig. 2C; Table S3). These clusters matched clinical metabolic states of hepatic dysfunction (Fig. 2E, Supplementary Fig. S1). Whereas HCC patients were distributed across
these clusters, FLC and UCD patients were enriched in cluster 1 (FDR-corrected $p = 0.044, 1.4e^{-11}$ respectively) and 7 (corrected $p = 0.038, 0.001$; Fig. 2D). The combined enrichment significance within these clusters was $p=0.0005$ and $p=3.0e^{-10}$ for FLC and UCD encounters respectively. Cluster 7 was characterized by significantly higher ammonia levels than other clusters in the dataset (Fig. 2E). Cluster 1 was remarkable for relatively normal CMP, lacking reductions in sodium and chloride compared to other hyperammonemia clusters.

Conclusions

Rare malignancies account for 20% of cancer incidence in the United States yet are poorly understood because of the challenges in assembling accurate patient cohorts and the frequent lack of specific ICD coding. Here we present a comprehensive analysis of large EMR and payer databases to define FLC incidence, finding that FLC is likely under-represented in SEER and under-diagnosed in young patients with liver cancer. Furthermore, our highly stringent ICD-10 code exclusions removed 10% of FLC patients from the UCSF cohort, thus establishing a lower bound for FLC incidence.

Our prospective analysis also identified a higher level of hyperammonemia in FLC patients than currently recognized in clinical practice, suggesting hyperammonemia is an underappreciated source of co-morbidity in these patients. We further demonstrate a computational method to assess the mechanism of this paraneoplastic complication, using unsupervised clustering to highlight the similarities between patients with urea cycle disorders and hyperammonemia due to FLC. Our findings demonstrate the utility of analyzing biochemical phenotypes directly from patient laboratory results. This workflow can serve as a model for developing biochemical insights into disease biology from patient laboratory data and is applicable to rare cancers and other diseases.

There are some limitations with this approach. Firstly, the FLC to HCC ratio at UCSF may differ from that of the broader oncology community given UCSF’s role as a tertiary care center. The interplay between the selection bias at a tertiary center and the potential under-diagnosis or failure to diagnose FLC at less experienced centers remains unknown. Another limitation is the single institution focus of this study.

Comprehensive understanding of rare diseases requires accurate identification of patients – a challenge often compounded by a lack of disease-specific ICD coding. This problem can be self-perpetuating, as imprecise billing data can lead to under-estimation of incidence, which decreases prioritization of disease-specific ICD code creation. Our findings suggest significantly higher national FLC incidence than previous estimates. Furthermore, our approach can capture incidence, detailed clinical information, and
potential systemic issues regarding misdiagnosis or underdiagnosis. Integrated analysis of rich computational data sources thus complements patient-led research findings.

**Methods**

*Identification and analysis of FLC patients within the UCSF dataset.*

For patients with ICD-10 codes for liver cancer between 2012 and 2021, we identified 102 ICD codes for liver conditions that predispose for HCC (Table S1), as well as 4 ICD codes for non-HCC liver cancers. Complete UCSF records from 2011–2021, including diagnosis codes and subspeciality and pathology notes, were collected, the latter analyzed using term extraction to identify patients with possible FLC diagnoses. Patients such identified were confirmed by clinician review (see supplementary methods).

*Estimating nation-wide FLC incidence.*

An age-stratified patient cohort was constructed in Komodo's Healthcare Map\(^{14}\) using the ICD-10-CM inclusion/exclusion criteria verified in the UCSF dataset. We used Bayesian inference to estimate national incidence with uncertainty. A Monte Carlo Simulation\(^{15}\) was applied to UCSF’s proportion of FLC to HCC patients between 0 and 50 years old in 10-year bins, assuming a binomial generative model. Based on aggregate patient counts within the Komodo dataset, we extrapolated the expected value and variance of this distribution to estimate national incidence.

**Identification and Characterization of hyperammonemia patients at UCSF**

We identified all clinical encounters where ammonia levels were drawn at UCSF between 2011 and 2021 and collected complete metabolic panel (CMP) data and maximum ammonia levels from each. Using the 3066 encounters with ammonia $\geq 50$ and a complete metabolic panel, we computed the neighborhood graph using a Minkowski distance metric followed by Leiden clustering.\(^{16}\) Enrichment of HCC, FLC, and urea cycle disorders within specific clusters was determined by Fisher’s exact test.

**Analysis of hyperammonemia in FLC patients on clinical trial**

Baseline and on-treatment ammonia levels were collected for FLC patients participating in the NCT01642186 and NCT02234986 trials\(^{17,18}\) and are reported here. Demographic details shown in Table S2.

**Declarations**

**Acknowledgements:** This work was supported by Chan Zuckerberg Initiative, the Fibrolamellar Cancer Foundation, and the Katie Moore Foundation. Dr. Gordan is supported by a Burroughs Wellcome Fund.
References


13. Surjan, R. C., dos Santos, E. S., Basseres, T., Makdissi, F. F. & Machado, M. A. A proposed physiopathological pathway to hyperammonemic encephalopathy in a non-cirrhotic patient with


**Figures**
Bayesian inference of annual FLC incidence: A. Analytical workflow: A subpopulation of patients with the “Liver Cancer” ICD10 code but lacking other significant co-morbidities was created in the Komodo Healthcare Map, a national billing database. In parallel, the age-specific distribution of FLC vs. cHCC diagnoses was determined in the UCSF EMR using chart-validate FLC diagnoses. This distribution was then applied to the Komodo dataset to generate an overall annual incidence. B. Step-by-step breakdown
of the identification of the initial Liver Cancer cohort, showing cohort scale and changes. C. Relative proportion of FLC patients per age cohort on the 2018 incident population. D. Empirically derived incidence distribution from our Monte Carlo Simulation, illustrating confidence in the estimated annual cases.

Figure 2

Characterizing FLC-associated hyperammonemia. A. Incidence of hyperammonemia at baseline and end of treatment (EOT) for patients participating in two successive FLC clinical trials. Prospective assessment at screening was instituted partially through NCT01642186, and performed from initiation of NCT02234986. Patients were included in baseline and EOT cohorts if a value had been collected at that time. B. Relative prevalence of hyperammonemia diagnostic coding in all liver carcinoma patients (e.g. including hepatitis diagnostic codes) vs. FLC proxy population, showing lower rates of detected hyperammonemia in the non-cirrhotic cohort. C. Umap of entire concurrent complete metabolic profile results in all patients with laboratory evidence of hyperammonemia at UCSF, demonstrating subgroups of distinct clinical phenotypes. D. Umap from C, with only UCD (small dots) and FLC (large dots) shown,
illustrating enrichment in a subset of patient clusters. E. Violin plots of individual lab values within each cluster from C; color shows median values while plot morphology indicates relative distribution.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1Wilcoxonrankedsumsignificancebylabvalueandcluster.pdf
- SupplementaryMethods.docx
- TableS1CMSandICDcodelists.xlsx
- TableS2FLCpatientcomorbidities.csv
- TableS3EncountercountsbyclusterandDiagnosis.csv