Frequency of *FLCN* Loss of Function Variants and Birt-Hogg-Dubé-Associated Phenotypes in a Healthcare System Population

Juliann M. Savatt^{1,2}, Hermela Shimelis¹, Andres Moreno-De-Luca^{1,2,3}, Natasha T. Strande^{1,2}, Matthew T. Oetjens¹, David H. Ledbetter^{1,2}, Christa Lese Martin^{1,2}, Scott M. Myers¹, Brenda M. Finucane¹

Affiliations

1. Autism & Developmental Medicine Institute, Geisinger, Lewisburg, Pennsylvania
2. Genomic Medicine Institute, Geisinger, Danville, Pennsylvania
3. Department of Radiology, Geisinger, Danville, Pennsylvania

Corresponding Author

Juliann M. Savatt

jmsavatt@geisinger.edu

Abstract

Purpose: Current penetrance estimates of Birt-Hogg-Dubé (BHD)-associated cutaneous, pulmonary, and renal manifestations are based on clinically ascertained families. In a healthcare system population, we used a genetics-first approach to estimate the prevalence of pathogenic/likely pathogenic (P/LP) variants in the *FLCN* gene, which cause BHD, and the penetrance of BHD-related phenotypes.

Methods: Exomes from 135,990 patient-participants in Geisinger's MyCode® cohort were assessed for P/LP *FLCN* variants. BHD-related phenotypes were evaluated from electronic health records. Association between *FLCN* P/LP variants and BHD-related phenotypes were assessed by Firth's logistic regression.

Results: P/LP *FLCN* variants were identified in 35 (1 in 3,234) individuals, 68.6% of whom had BHD-related phenotype(s) including: cystic lung disease (65.7%); pneumothoraces (17.1%); cutaneous manifestations (8.6%); and renal cancer (2.9%). Only four (11.4%) were clinically diagnosed with BHD.

Conclusion: In this healthcare population, the frequency of P/LP *FLCN* variants is 60 times higher than the previously reported prevalence. Although most variant-positive individuals had BHD-related phenotypes in the EHR, a minority were clinically diagnosed with BHD, likely because cutaneous manifestations, pneumothoraces, and renal cancer were observed at lower frequencies than clinical cohorts. Improved clinical recognition of cystic lung disease and education concerning its association with *FLCN* variants could prompt evaluation for BHD.

Introduction

First described in the 1970's, Birt-Hogg-Dubé syndrome (BHD; OMIM 135150) is an autosomal dominant condition associated with benign cutaneous manifestations, pulmonary cysts/blebs/bullae (hereafter referred to as “cystic lung disease”), spontaneous pneumothoraces, and renal neoplasms.¹⁻² BHD is caused by germline variants in the folliculin (*FLCN*) gene, a tumor suppressor on chromosome 17p11.2 (OMIM 607273), that encodes the folliculin protein that interacts with numerous signaling pathways.³⁻⁶ Variants resulting in premature termination of the folliculin protein are the most common variant type identified in individuals with BHD.²

BHD has an estimated prevalence of 1 in 200,000⁷, and, as of 2017, over 600 families with BHD have been described in the literature.⁸ The most common phenotypic features in individuals with BHD are benign cutaneous manifestations, including fibrofolliculomas and trichodiscomas. Such findings have been identified in 58-90% of clinically ascertained patients and typically appear after age 25.⁹ Macroscopically, trichodiscomas and fibrofolliculomas are indistinguishable from one another and are believed to be on the same morphological spectrum along with perifollicular fibromas.¹⁰ Acrochordons have also been described in patients with BHD, but these lesions are also common in the general population and their relationship to BHD is unclear.⁹

Patients with BHD also have increased risk of developing cystic lung disease and spontaneous pneumothoraces. Pulmonary cysts have been identified in 67-90% of patients with BHD.^{2, 9, 11} Cysts vary in number and are typically bilateral and located in the lower basal zones of the lungs.⁹ The literature suggests that 24-38% of patients with BHD experience a spontaneous pneumothorax with a median age of 38 and earliest age of seven years.^{2, 11}

In addition to the lung and skin findings, patients with P/LP *FLCN* variants are at an increased risk for renal neoplasia. Bilateral, multifocal renal tumors are typical and are histologically diverse with most being hybrid oncocytic tumors, chromophobe renal cell

carcinoma, and clear cell carcinoma.² It has been estimated that 12 to 34% of individuals with BHD develop renal cancers^{2, 9, 11} with a mean age of diagnosis of 50.7 years and earliest diagnosis at age 20.⁹ Additionally, individuals with BHD are at risk of developing oncocytosis, microscopic foci of oncocytic cells in the renal parenchyma, that is a precursor lesion to malignant tumors.² Other phenotypes have been reported in patients with BHD including lipomas, parathyroid adenomas, thyroid nodules, thyroid cancer, and parotid oncocytomas, but additional data are needed to determine if these are associated with BHD.²

To date, penetrance estimates of BHD-related features in individuals with P/LP *FLCN* variants have been based on families clinically ascertained due to a personal and/or family history. Population-based testing for pathogenic variants in other disease-associated genes has suggested that clinically ascertained penetrance estimates might be higher than penetrance in variant-positive individuals from broader cohorts.^{12, 13} Here, we sought to explore the prevalence of P/LP *FLCN* variants in a health care system-based population and utilize electronic health record (EHR) data to investigate the frequency of BHD-related phenotypes and prior clinical identification of the *FLCN* variant in individuals with such variants.

Materials and Methods

MyCode Participants

The Geisinger MyCode[®] Community Health Initiative (MyCode) serves as a biobank of blood and other samples from over 290,000 patient-participants who consent to health-related research.¹⁴⁻¹⁶ Participants are recruited throughout Geisinger regardless of disease or phenotype.¹⁴⁻¹⁷ The population for this study includes a subset of MyCode patient-participants, referred to as the DiscovEHR cohort¹⁴, with available exome sequencing and linked EHR data. Most patient-participants also have available SNP genotyping data. MyCode and the research outlined below are governed by the Geisinger Institutional Review Board; all participants in this study provided written informed consent.

Exome Sequencing and Intragenic CNV Calling

Exome sequencing for 135,990 adult MyCode patient-participants was performed in collaboration with the Regeneron Genetics Center, as described previously.¹⁴⁻¹⁵ Briefly, exome sequence data were aligned to human reference GRCh38 with BWA-mem¹⁸, and the resultant BAM files processed using the Picard MarkDuplicates tool (<http://broadinstitute.github.io/picard>) to flag duplicate reads. After exome data were aligned, single nucleotide variants (SNVs) and insertion/deletion variants were identified using weCall.¹⁹ Joint genotyping was performed across the entire cohort using GLnexus.²⁰ CNV calling was carried out using Copy number estimation using the Lattice-Aligned Mixture ModelS (CLAMMS) algorithm²¹ and CNV calls were validated with SNP-array CNV calls using Penn CNV.²²

***FLCN* Variant Analysis**

We reviewed exome sequencing data (n=135,990) for potential loss of function variants in the *FLCN* gene. Variants in VCF files were annotated with Ensemble Variant Effect Predictor (VEP) version 96 using RefSeq transcripts. High impact variants (nonsense, frameshift, +/- 1,2 splice site) based on the *FLCN* transcript most expressed in adult tissues (transcript variant 1, GenBank: NM_144997.7, 14 exons total with 11 coding exons, 579 amino acid residues) in genome build GRCh38.⁵ Sequence variants with an allele balance greater or equal to 0.20, genotype quality greater than or equal to 80, coverage greater than or equal to 20, and GnomAD allele frequency less than 0.001 were identified. Two *FLCN* variants that nearly met our quality thresholds were included in our analysis upon confirmation of the *FLCN* variant from a clinical genetic testing report documented in the EHR (n=1) and validation of a variant by evaluation of IGV plots (n=1). Variants were then reviewed (JMS and HS) using the American College of Medical and Genomics and Association for Molecular Pathology variant interpretation guidelines and classifications were confirmed by a third reviewer (NTS), who is board certified in laboratory genetics and genomics.²³⁻²⁵

Determination of Relatedness

To report on the number of unique families with P/LP *FLCN* variants, the frequency of BHD-related phenotypes within families, and P/LP variant prevalence in an unrelated subset of the MyCode cohort, first- and second-degree familial relationships were identified from genotype data using Pedigree Reconstruction and Identification of the Maximally Unrelated Set (PRIMUS).²⁶

Phenotype Evaluation – EHR and Imaging Review

We characterized BHD-related pulmonary, cutaneous, and renal phenotypes in variant-positive individuals via manual review of patients' EHRs. A chart review guide was developed to collect variables of interest and independent, double reviews were completed by a genetic counselor and board-certified neurodevelopmental pediatrician (JMS, SMM) between March and November 2020 (Supplemental Materials and Methods). All discrepancies between reviewers were resolved through joint review and consensus.

Cystic lung disease and renal masses might be incidentally identified on imaging studies of adjacent body parts (e.g., Computed Tomography (CT) of the abdomen includes the lung bases and lumbar spine Magnetic Resonance Imaging (MRI) includes at least part of the kidneys). Additionally, small lung cysts, blebs, and bullae may be identified, but sometimes not mentioned, in radiology reports or elsewhere in the EHR. Because of this, dedicated review of imaging studies including all or part of the kidneys and lungs was completed by a board-certified radiologist (AMD) for all patients with available imaging (Supplemental Materials and Methods). Plain X-rays were not included in this review.

Phenotype Evaluation – ICD-9/10 Diagnostic Codes

While chart review provides more accurate phenotyping, it is low throughput and often is not feasible for large sample sizes. As such, clinical manifestations of BHD-related phenotypes were also evaluated in variant positive and negative patient-participants by extracting relevant International Statistical Classification of Diseases and Related Health Problems 9 and 10 codes (ICD-9/10) from the EHR in February 2021. Selected ICD-9/10 diagnostic codes were defined

prior to manual chart review and were intended to capture BHD-related pulmonary, cutaneous, and renal phenotypes (Supplemental Table 1). While ICD-9/10 codes can provide an estimate of BHD-related phenotypes, some codes that correspond to BHD-features are non-specific and could be used for other, unrelated clinical findings (e.g., ICD-10 codes for trichodyscoma and fibrofolliculoma are broadly used for benign neoplasms of the skin), and others might not be applied even when such a feature is identified due to its indolent or benign nature (e.g., ICD-9/10 codes for pulmonary cyst). Since use of ICD-9/10 codes alone might not correctly reflect patient phenotypes,²⁷ we evaluated the ability of ICD-9/10 codes in the EHR to accurately identify BHD-related phenotypes by comparing diagnoses captured using diagnostic codes to those captured with manual review of EHR data in variant-positive individuals.

Statistical Analysis

Continuous variables are summarized as median and interquartile range and categorical variables are described as frequency and percentage. Fisher's exact and Wilcoxon rank-sum tests were performed to test significant differences in demographics between variant-positive and variant-negative individuals. Associations between *FLCN* P/LP variants and BHD phenotype were assessed by comparing the frequency of BHD-related phenotypes in variant-positive individuals with the variant-negative group using relevant ICD-9/10 diagnostic codes. The statistical significance of associations between *FLCN* P/LP variants and BHD-related phenotypes was calculated using Firth logistic regression model adjusting for age, sex and the first four principal components of ancestry. To reduce the impact of population stratification on the results, the European subset of the cohort (121,876) was used for the association analysis. The European subset used in the association analysis was identified by principal component analysis and clustering with the 1000 Genomes reference population using the same protocol applied to the UK Biobank.²⁹ Bonferroni correction was used to account for multiple testing ($\alpha = 0.05$). All analyses were conducted using R version 4.0.1.

Results

Prevalence of P/LP *FLCN* Variants

Of the 135,990 patient-participants, 35 (0.03%, 1 in 3,234) had a P/LP variant in *FLCN*. Based on genetic relatedness data,²⁶ we determined that these 35 individuals are from 28 families, representing 0.03% of the 90,563 first and second-degree unrelated subset of the cohort. Thirteen unique *FLCN* variants were identified, including five frameshift variants identified in 17 individuals, five nonsense variants identified in six patient-participants, two canonical splice site variants identified in 12 individuals, and a single intragenic deletion of exons 4-11 in one participant (Figure 1). One individual had two variants identified -- deletion of exons 4-11 and a frameshift variant in cis. Nine of the thirteen variants identified in 32 of the 35 participants were previously reported in the literature and/or submitted to the National Center for Biotechnology Information ClinVar²⁹ database. Table 1 summarizes demographics of *FLCN* variant-positive and variant-negative individuals in the MyCode cohort. Patient-participants with *FLCN* variants were 62.9% female, 97.1% self-reported their race as White, 100% self-reported non-Hispanic ethnicity, and had a median age of 62 years. Eighty-nine percent of patient-participants were alive, and the median length of EHR data was 17.43 years at the time of manual chart review. No statistically significant demographic differences were identified between the variant-positive and variant-negative individuals.

BHD-related Phenotypes – Chart Review of Patient-participants with *FLCN* Variants

The frequency of BHD-related phenotypes (i.e., cutaneous, pulmonary, and renal manifestations) was evaluated by manual chart review of each variant-positive patient-participant. We found that 68.6% (n=24/35) of variant-positive individuals had EHR-documentation of a BHD-related phenotype (Figure 2 and Table 2).

BHD-related cutaneous findings including fibrofolliculomas, perifollicular fibromas, and trichodiscomas were identified in three patient-participants (8.6%). Among all variant-positive participants, 60% (n=21/35) had documentation of an evaluation with a dermatologist and/or at least one skin biopsy. When restricting to those who were examined by a dermatologist and/or

had a skin biopsy 14.3% (n=3/21) had a BHD-related cutaneous finding. BHD-related cutaneous findings were identified in 10.7% (n=3/28) of unique families with *FLCN* variants.

Pulmonary phenotypes, including basilar cystic lung disease and/or pneumothorax, were identified in 65.7% (n=23/35) of patient-participants. Of the 23 individuals with pulmonary phenotypes, six had at least one spontaneous pneumothorax; two had exactly one and four had more than one. Participants experienced their first pneumothorax between the ages of 21 and 79 years (median 27.5 years). The individual that experienced their first pneumothorax at age 79 had a diagnosis of chronic obstructive pulmonary disease. All 23 individuals with a pulmonary phenotype had basilar cystic lung disease. Eighty-three percent (n=29/35) of variant-positive patient-participants had CT or MRI imaging that included at least a portion of the lungs (e.g., abdomen CT includes lung bases) and 37.1% (n=13/35) had adequate visualization of the entire lungs. Among those with any lung imaging and those with adequate imaging reviewed, 79.3% (n=23/29) and 100% (n=13/13) had basilar cystic lung disease, respectively. Among unique families with *FLCN* variants, 21.4% (n=6/28) had pneumothorax and 67.9% (n=19/28) had cystic lung disease in at least one family member (Figure 2).

Renal cancer was identified in one patient-participant (2.9%) with a chromophobe renal cancer diagnosed at age 71. Among the remaining 34 patient-participants without renal cancer, 79.4% (n=27/34) had imaging that included the kidneys and 67.6% (n=23/34) had adequate visualization of both kidneys in their entirety. Among those with any renal imaging and those with adequate renal imaging (e.g., bilateral renal ultrasound), 11.1% (n=3/27) and 13.0% (n=3/23) had renal masses identified, respectively. These renal masses had not been pathologically confirmed to be renal cancer.

In addition to core BHD-related phenotypes, the EHR review included evaluation of other phenotypes that have been reported in patients with BHD including thyroid cancer and parotid

oncocytomas. Bilateral parotid oncocytosis was noted in a single participant with bilateral cystic lung disease. Thyroid cancer was not noted in any individuals.

Four individuals (11.4%) from three families had a diagnosis of BHD or documentation of a germline *FLCN* variant in their EHR. Two of those individuals had EHR documentation of multiple spontaneous pneumothoraces and bilateral blebs; one had a fibrofolliculoma, bilateral blebs, and renal mass; and one had a perifollicular fibroma.

BHD-related Phenotypes – Diagnostic Code Comparison

Comparison of BHD-related phenotypes captured by ICD codes with manual chart review (Supplemental Table 2) revealed that ICD-9/10 codes for BHD-related cutaneous features lack specificity and, as such, estimate the rates of such findings in variant-positive individuals at a higher rate than observed via manual review. We found that 42.9% (n=15/35) versus 8.6% (n=3/35) had BHD-related cutaneous features based on the ICD code approach and chart review respectively. ICD codes underestimated the frequency of cystic lung disease [20.0% (n=7/35) based on ICD codes versus 65.7% (n=23/35) chart and radiology review], likely due to the benign nature of these findings and incidental identification on imaging. Diagnostic codes for pneumothorax were found to accurately identify this phenotype in variant-positive individuals [17.1% (n=6/35) ICD codes versus 17.1% (n=6/35) chart and radiology review]. All six individuals with pneumothoraces identified on manual chart review had an appropriate ICD-9/10 code in their EHR. ICD diagnostic codes for renal cancer appropriately identified the single individual with renal cancer. Diagnostic codes associated with a renal mass underestimated this phenotype [2.9% (n=1/34) ICD codes versus 8.8% (n=3/34) chart review].

The associations between *FLCN* P/LP variants and ICD-9/10 based BHD-related phenotypes were assessed by Firth's logistic regression in the unrelated, European subset of the MyCode cohort (Table 3). P/LP variants in *FLCN* were associated with a higher prevalence of pulmonary phenotypes (OR=4.16; 95% CI:11.65-9.43; p=0.026), and, more specifically, spontaneous pneumothorax (OR=15.63; 95% CI: 5.96-35.69; p=1.92x10⁻⁵) in variant positive

patient-participants. Association of *FLCN* P/LP variants with other pulmonary, cutaneous, and renal phenotypes were not statistically significant when correcting for multiple testing. We performed the association analysis in the full unrelated subset of the cohort including European and non-European minority populations, to increase sample size. The results of the full unrelated cohort were consistent with those of the European subset (Supplemental Table 3).

Discussion

Frequency of P/LP *FLCN* Variants

BHD is considered a rare disorder with an estimated prevalence of 1 in 200,000.⁷ BHD prevalence and penetrance estimates have been based on clinically ascertained families to date. This study examined exome sequencing data from a large health care population to estimate the prevalence of P/LP *FLCN* variants and frequency of BHD-related phenotypes. *FLCN* variants were identified in 28 out of 90,563 first- and second-degree unrelated individuals (1 in 3,234) suggesting that such variants are significantly more common than previously reported. This finding is consistent with the increased frequency of variants associated with other genetic syndromes identified in broader, population cohorts via genome-first approaches.³⁰⁻³²

Prevalence of BHD-related Phenotypes

In participants with P/LP *FLCN* variants from this healthcare cohort, BHD-related lung findings were the most common phenotype identified (65.7%, n=23/35). The frequency of basilar pulmonary cystic lung disease in those with CT and MRI lung imaging (79.3%, n=23/29) overlapped with the reported frequency in clinically ascertained individuals (67-90%)^{2, 9, 11} and was even higher than the reported frequency in those with adequate visualization of the lungs (100%). The frequency of pneumothorax (17.1%, n=6/35) was lower than that reported in the published literature (24-38%). Pneumothorax-associated ICD-9/10 diagnostic codes were present in the EHR of all patient-participants with pneumothoraces identified by chart review, suggesting that these codes capture the phenotype with high sensitivity. When we assessed the

association between *FLCN* P/LP variants and pneumothorax, we identified a higher prevalence of spontaneous pneumothorax diagnosis (OR=15.63 95%CI (5.96-35.69), $p=1.92 \times 10^{-5}$) in variant positive individuals. Together, these data suggest that while patient-participants with *FLCN* variants in this healthcare-based cohort may have lower rates of pneumothorax compared to the clinically ascertained population with BHD, they have higher rates than the variant-negative population. The prevalence of diagnosis codes associated with cystic lung disease was not statistically different between the variant-positive and the variant-negative patient-participants. However, this could be due to inability of the selected ICD 9/10 codes to identify all individuals with the phenotype, as a minority of variant-positive participants with this phenotype based on manual chart review had the relevant ICD codes in the EHR. Therefore, this pulmonary finding, which can be a precursor to spontaneous pneumothorax, may also occur at higher rates in individuals with *FLCN* P/LP variants. Additional research is needed to determine whether *FLCN* P/LP variants are associated with increased prevalence of basilar cystic lung disease.

Other BHD-related phenotypes, including cutaneous manifestations and renal cancer appeared to be less common in the variant-positive individuals in this cohort compared with BHD probands in the reported literature. BHD-related dermatologic findings were identified in only 8.6% ($n=3/35$) of *FLCN* variant-positive individuals compared to 58-90% of patients with BHD in the literature.⁹ When restricting the analysis to patient-participants who were examined by a dermatologist and/or had a skin biopsy that would have likely identified these benign phenotypes, only 14.3% ($n=3/21$) had BHD-related skin findings. This estimate is notably lower than the frequency of BHD-related dermatologic findings reported in the literature (14.3% vs. 58-90%).⁹ Since lack of clinical assessment does not alone account for differences in frequency, this finding is consistent with a lower penetrance of *FLCN* P/LP variants than previously reported. *FLCN* P/LP variants were not associated with an increased rate of cutaneous diagnosis, based on ICD-9/10 codes. Given the non-specific nature of cutaneous ICD-9/10

codes and their overestimation of BHD-related skin-findings in variant-positive individuals, however, additional assessment comparing the frequency of BHD-related cutaneous findings between variant-positive and variant-negative populations is needed.

Only one participant with a *FLCN* variant was identified to have renal cancer (2.9%, n=1/35). This is a lower frequency than reported in the literature (12-34%).^{2, 9,11} Although some participants are younger (e.g., five participants with *FLCN* variants are less than 40 years of age) and could go on to develop a renal cancer in their lifetime, the majority of *FLCN* variant-positive individuals in the cohort are over the age of 50.7 years (74.3%, n=26/35) which is the median age of renal cancer diagnosis reported in the literature for individuals with BHD. Due to the indolent nature of BHD-related renal cancers,² some variant-positive patient-participants might have yet to come to clinical attention. If we assume the three participants with a renal mass noted on imaging have an undiagnosed renal cancer and only consider those with adequate renal imaging as being assessed for a renal cancer, an upper limit of potential renal cancer in variant-positive individuals is 18.2% (n=4/22). *FLCN* P/LP variants were not associated with increased prevalence of renal cancer in the MyCode cohort based on EHR data. Although additional studies are needed due to small numbers, this suggests the rates of renal cancer among individuals with *FLCN* variants may not be as high as previously estimated.

These data suggest that, among individuals in a health system population who volunteered to participate in genetic research but were not selected for any disease state, the majority of individuals with P/LP *FLCN* variants have a phenotype consistent with BHD. Individual BHD-related phenotypes including pneumothorax, cutaneous findings, and renal cancer may be less common among individuals with P/LP *FLCN* variants than previously reported, a finding that is consistent with other studies examining frequency of disease-associated phenotypes in population-based cohorts.^{12,13} Although we are not aware of any prior population-based studies of *FLCN* variants, the *FLCN* gene is located within the recurrent copy

number variant region on 17p11.2 that is deleted in 90% of individuals with Smith-Magenis syndrome (SMS), hence these individuals are haploinsufficient for *FLCN*; however, there are few published reports of BHD-related phenotypes in individuals with SMS.³³⁻³⁶ Therefore, not all individuals with *FLCN* haploinsufficiency develop BHD, further suggesting that published penetrance of BHD-related phenotypes may be overestimated.

Clinical Identification of *FLCN* Variants

Even though most participants with a P/LP *FLCN* variant had features consistent with BHD documented in the EHR, the majority of variant-positive patient-participants (88.6%, n=31/35) did not have a clinical diagnosis of BHD. The majority of those with BHD-related features had basilar cystic lung disease (95.8%, n=23/24 with phenotype). Although these features are included in the diagnostic criteria for BHD,^{2,37} their presence alone did not prompt a referral to genetics or clinical genetic testing in any patient-participants with *FLCN* variants. Even in the four individuals with clinically identified *FLCN* variants or BHD diagnoses in this cohort, all had cystic lung disease but were evaluated due to other phenotypes (fibrofolliculoma, perifollicular fibroma, pneumothoraces, family history of BHD). Although cystic lung disease was identified in 23 individuals in total, nine patient-participants were only identified on radiologist review of imaging data, meaning these findings were not documented in the EHR and were not commented on in the initial radiology report. Improved clinical recognition of these benign findings and education concerning their potential association with *FLCN* variants could prompt additional evaluation for BHD. Identifying patients at risk for BHD-phenotypes could inform appropriate management. For example, in patients with BHD, a first pneumothorax might be treated more aggressively with pleurectomy/pleurodesis given the risk for recurrent pneumothoraces³⁸⁻³⁹ and identification of individuals with *FLCN* variants could enable earlier screening for renal cancer, potentially allowing detection of cancers at earlier stages.⁴⁰

Limitations

This study has several limitations. Because the MyCode population is primarily of European ancestry and from a single healthcare system, it is unclear if these findings are generalizable to more diverse patient populations. Our estimates of BHD-related phenotypes rely on data within the EHR; as such, not all patient-participants have had assessments that would identify BHD-related features. Other studies have identified differential risks for BHD-related phenotypes depending on the *FLCN* variant.³⁸ Given the small sample size and small number of recurrent variants, this analysis was not completed in this study and could be considered in future work

Conclusions

P/LP *FLCN* variants were identified in 1 in 3,234 individuals in a population-based healthcare population. Of those with such variants, 65.6% were identified to have BHD-related phenotypes; however, only 11.4% had been clinically diagnosed with BHD. This limited ascertainment is likely because classic BHD-related phenotypes, such as cutaneous manifestations, pneumothorax, and renal cancer, were observed at lower frequencies. Cystic lung disease was the most common phenotype identified in our cohort and could be used to prompt additional evaluation for BHD. This finding could inform future screening strategies for BHD in large-scale healthcare patient populations.

Data Availability

The data that support the findings of this study are available within the article and/or are available on request from the corresponding author, J.M.S.

Acknowledgements

This work was supported by the National Institutes of Mental Health of the National Institutes of Health (NIH) under award number: R01 MH074090. The authors would like to acknowledge the MyCode patient-participants that make this work possible. We would also like to acknowledge those involved in the Geisinger-Regeneron DiscovEHR Collaboration who have enabled the generation of the genomics data used in this study and Dr. H. Les Kirchner for his thoughtful review and comments that helped shape the statistical analysis.

Author Information

Funding Acquisition - C.L.M., D.H.L.; Conceptualization - J.M.S., H.S., B.M.F., S.M.M;
Supervision - B.M.F., C.L.M., M.T.O., D.H.L; Project Administration - J.M.S.; Data Curation - J.M.S., H.S., S.M.M., A.M.D.L; Investigation, Methodology - J.M.S., H.S., S.M.M., A.M.D.L, N.T.S; Resources - H.S., M.T.O.; Formal Analysis - H.S.; Visualization, Writing - original draft - J.M.S., H.S.; Writing - review & editing- J.M.S., H.S., B.M.F., C.L.M., M.T.O., D.H.L, A.M.D.L., S.M.M, N.T.S.

Ethics Declaration

The Geisinger MyCode[®] Community Health Initiative (MyCode) serves as a biobank of blood and other samples from over 290,000 patient-participants who consent to health-related research (<https://www.geisinger.org/precision-health/mycode>). MyCode and the research outlined are approved by the Geisinger Institutional Review Board. MyCode participants or their parent/legal guardian provide written consent and HIPAA authorization. As part of participation, they agree to provide samples for broad research use and permit access to data in their electronic health record for research use.¹⁴ Individual data included in this manuscript have been de-identified and presented in aggregate.

Figure 1. Pathogenic/Likely Pathogenic variants in *FLCN* identified in the MyCode cohort.

Based on build GRCh38 (NC_000017.11) and *FLCN* transcript most expressed in adult tissues (transcript variant 1, GenBank: NM_144997.7, 14 exons total with 11 coding exons, 579 amino acid residues) P=number of patient-participants with that variant and F=number of families with that variant. $\overline{\text{P}}$ denotes variants that were previously reported in the literature and/or submitted to ClinVar. ^a denotes two variants identified in the same participant.

Figure 2. Frequency of BHD-related phenotypes in the literature compared to that in individual MyCode patient-participants with P/LP *FLCN* variants and in unique MyCode

families. Exome data was used to identify first-second degree relationships and to group individuals into unique families.

References

1. Birt, A. R., Hogg, G. R., & Dubé, W. J. (1977). Hereditary multiple fibrofolliculomas with trichodiscomas and acrochordons. *Arch Dermatol*, 113(12), 1674-1677.
2. Schmidt, L. S., & Linehan, W. M. (2015). Molecular genetics and clinical features of Birt-Hogg-Dubé syndrome. *Nat Rev Urol*, 12(10), 558-569. doi:10.1038/nrurol.2015.206
3. Baba, M., Hong, S. B., Sharma, N., Warren, M. B., Nickerson, M. L., Iwamatsu, A., . . . Zbar, B. (2006). Folliculin encoded by the BHD gene interacts with a binding protein, FNIP1, and AMPK, and is involved in AMPK and mTOR signaling. *Proc Natl Acad Sci U S A*, 103(42), 15552-15557. doi:10.1073/pnas.0603781103
4. Hasumi, Y., Baba, M., Ajima, R., Hasumi, H., Valera, V. A., Klein, M. E., . . . Linehan, W. M. (2009). Homozygous loss of BHD causes early embryonic lethality and kidney tumor development with activation of mTORC1 and mTORC2. *Proceedings of the National Academy of Sciences*, 106(44), 18722-18727. doi:10.1073/pnas.0908853106
5. Nickerson, M. L., Warren, M. B., Toro, J. R., Matrosova, V., Glenn, G., Turner, M. L., . . . Schmidt, L. S. (2002). Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt-Hogg-Dubé syndrome. *Cancer Cell*, 2(2), 157-164. doi:10.1016/s1535-6108(02)00104-6
6. Schmidt, L. S., & Linehan, W. M. (2018). FLCN: The causative gene for Birt-Hogg-Dubé syndrome. *Gene*, 640, 28-42. doi:10.1016/j.gene.2017.09.044
7. Orphanet, 2008. https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=122. Accessed December 14, 2020.
8. BHD Foundation. (2017, January 2017). Published BHD Families. Retrieved from <https://www.bhdsyndrome.org/for-researchers/what-is-bhd/introduction/published-bhd-families/> Accessed December 18, 2020.

9. Jensen, D. K., Villumsen, A., Skytte, A. B., Madsen, M. G., Sommerlund, M., & Bendstrup, E. (2017). Birt-Hogg-Dubé syndrome: a case report and a review of the literature. *Eur Clin Respir J*, 4(1), 1292378. doi:10.1080/20018525.2017.1292378
10. Lopez-Garcia, D. R., Teague, D., Landis, E. T., & Sanguenza, O. P. (2014). Morphological diversity of trichodiscomas and fibrofolliculomas. *Am J Dermatopathol*, 36(9), 734-740. doi:10.1097/dad.0000000000000084
11. Gupta, N., Sunwoo, B. Y., & Kotloff, R. M. (2016). Birt-Hogg-Dube Syndrome. *Clinics in Chest Medicine*, 37(3), 475-486. doi:10.1016/j.ccm.2016.04.010
12. van Rooij, J., Arp, P., Broer, L., Verlouw, J., van Rooij, F., Kraaij, R., . . . Verkerk AJMH. (2020). Reduced penetrance of pathogenic ACMG variants in a deeply phenotyped cohort study and evaluation of ClinVar classification over time. *Genet Med* 22, 1812–1820. doi: 10.1038/s41436-020-0900-8
13. Tuke, M.A., Ruth, K.S., Wood, A.R., Beaumont, R.N., Tyrrell, J., Jones, S.E., . . . Murray A. (2019). Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genet Med.*, 21(4):877-886. doi: 10.1038/s41436-018-0271-6.
14. Carey, D. J., Fetterolf, S. N., Davis, F. D., Faucett, W. A., Kirchner, H. L., Mirshahi, U., . . . Ledbetter, D. H. (2016). The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med*, 18(9), 906-913. doi:10.1038/gim.2015.187
15. Dewey, F. E., Murray, M. F., Overton, J. D., Habegger, L., Leader, J. B., Fetterolf, S. N., . . . Carey, D. J. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science*, 354(6319). doi:10.1126/science.aaf6814
16. MyCode ScoreCard. (December 1, 2021). Retrieved from <https://www.geisinger.org/-/media/OneGeisinger/pdfs/ghs/research/mycode/mycode-scorecard.pdf?la=en>. Accessed December 11, 2021.

17. Buchanan, A. H., Lester Kirchner, H., Schwartz, M. L. B., Kelly, M. A., Schmidlen, T., Jones, L. K., . . . Sturm, A. C. (2020). Clinical outcomes of a genomic screening program for actionable genetic conditions. *Genetics in Medicine*, 22(11), 1874-1882.
doi:10.1038/s41436-020-0876-4
18. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. Retrieved from <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>
19. Genomics PLC, E. (2018, October 18, 2018). weCall. 2.0. Retrieved from <https://github.com/Genomicsplc/wecall>
20. Lin, M. F., Rodeh, O., Penn, J., Bai, X., Reid, J. G., Krasheninina, O., & Salerno, W. J. (2018). GLnexus: joint variant calling for large cohort sequencing. *bioRxiv*, 343970.
doi:10.1101/343970
21. Packer, J. S., Maxwell, E. K., O'Dushlaine, C., Lopez, A. E., Dewey, F. E., Chernomorsky, R., . . . Reid, J. G. (2016). CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. *Bioinformatics*, 32(1), 133-135. doi:10.1093/bioinformatics/btv547
22. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, SF, Hakonarson, H, Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*, 17(11):1665-74. doi: 10.1101/gr.6861907
23. Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., . . . Group, C. S. V. I. W. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation*, 39(11), 1517-1524.
doi:10.1002/humu.23626
24. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . on behalf of the, A. L. Q. A. C. (2015). Standards and guidelines for the interpretation of sequence

- variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405-423. doi:10.1038/gim.2015.30
25. Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., . . . Martin, C. L. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*, 22(2), 245-257. doi:10.1038/s41436-019-0686-8
 26. Staples, J., Qiao, D., Cho, M. H., Silverman, E. K., Nickerson, D. A., & Below, J. E. (2014). PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet*, 95(5), 553-564. doi:10.1016/j.ajhg.2014.10.005
 27. Quan, H., Li, B., Saunders, L.D., Parsons, G.A., Nilsson, C.I., Alibhai, A., Ghali, W.A.; IMECCHI Investigators. (2008). Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*, 43(4):1424-1441. doi:10.1111/j.1475-6773.2007.00822.x
 28. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), 203-209. doi:10.1038/s41586-018-0579-z
 29. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, 46(D1):D1062-D1067. doi: 10.1093/nar/gkx1153
 30. Manickam, K., Buchanan, A.H., Schwartz, M.L.B., Hallquist, M.L.G., Williams, J.L., . . . Murray, M.F. (2018). Exome Sequencing-Based Screening for BRCA1/2 Expected

Pathogenic Variants Among Adult Biobank Participants. *JAMA Netw Open*, 1(5):e182140. doi: 10.1001/jamanetworkopen.2018.2140

31. Cho, B. P. H., Nannoni, S., Harshfield, E., Tozer, D., Graf, S., Bell, S., & Markus, H. (2020). NOTCH3 variants are more common than expected in the general population and associated with stroke and vascular dementia: an analysis of 200,000 participants. *Journal of Neurology, Neurosurgery and Psychiatry*, doi:10.17863/CAM.64535
32. Dershem, R., Gorvin, C., Metpally, R.P.R., Krishnamaurthy, S., Smelser, D.T., Hannan, F.M., . . . Breitwieser, G.E.; Regeneron Genetics Center. (2020). Familial Hypocalciuric Hypercalcemia Type 1 and Autosomal-Dominant Hypocalcemia Type 1: Prevalence in a Large Healthcare Population. *Am J Hum Genet.*, 106(6):734-747. doi: 10.1016/j.ajhg.2020.04.006.
33. Dardour, L., Verleyen, P., Lesage, K., Holvoet, M., & Devriendt, K. (2016). Bilateral renal tumors in an adult man with Smith-Magenis syndrome: The role of the FLCN gene. *European Journal of Medical Genetics*, 59(10), 499–501. doi:10.1016/j.ejmg.2016.09.005
34. Truong, H.T., Dudding, T., Blanchard, C.L., & Elsea, S.H. (2010). Frameshift mutation hotspot identified in Smith-Magenis syndrome: case report and review of literature. *BMC Medical Genetics*, 11,142. doi:10.1186/1471-2350-11-142
35. Finucane, B., Savatt, J.M., Shimelis, H., Girirajan, S., Myers, S.M.(2021). Birt-Hogg-Dubé symptoms in Smith-Magenis syndrome include pediatric-onset pneumothorax. *Am J Med Genet Part A.*; 1– 3. doi:10.1002/ajmg.a.62159
36. Smith, A.C.M., Fleming, L.R., Piskorski, A.M., Amin, A., Phorphutkul, C., de la Monte, S. . . Linehan, W.M. (2014, October 18-22). Deletion of 17p11.2 encompasses FLCN with increased risk of Birt-Hogg-Dubé in Smith Magenis Syndrome: Recommendation for Cancer Screening. [Conference presentation abstract]. Annual conference of the

American Society of Human Genetics, San Diego, CA, United States.

<https://www.ashg.org/wpcontent/>

[uploads/2019/10/2014-poster-abstracts.pdf](#)

37. Menko, F.H., van Steensel, M.A., Giraud, S., Friis-Hansen, L., Richard, S., Ungari, S., . . . Maher, E.R.; European BHD Consortium. (2009). Birt-Hogg-Dubé syndrome: diagnosis and management. *Lancet Oncol*, 10(12):1199-206. doi: 10.1016/S1470-2045(09)70188-3
38. Sattler, E.C., Syunyaeva, Z., Mansmann, U., Steinlein, O.K. (2020). Genetic Risk Factors for Spontaneous Pneumothorax in Birt-Hogg-Dubé Syndrome. *Chest*, 157(5):1199-1206. doi: 10.1016/j.chest.2019.12.019. Epub 2020 Jan 17. PMID: 31958439.
39. Gupta, N., Koprass, E. J., Henske, E. P., James, L. E., El-Chemaly, S., Veeraraghavan, S., . . . McCormack, F. X. (2017). Spontaneous Pneumothoraces in Patients with Birt-Hogg-Dubé Syndrome. *Annals of the American Thoracic Society*, 14(5), 706–713. doi:10.1513/AnnalsATS.201611-886OC
40. Johannesma, P.C., Reinhard, R., Kon, Y., Sriram, J.D., Smit, H.J., van Moorselaar, R.J., . . . Postmus, P.E.; Amsterdam BHD working group. (2015) Prevalence of Birt-Hogg-Dubé syndrome in patients with apparently primary spontaneous pneumothorax. *Eur Respir J*, 45(4):1191-4. doi: 10.1183/09031936.00196914.

BHD-Associated Cutaneous Phenotypes		
Published Frequency of Cutaneous Phenotypes ⁹	Frequency in DiscovEHR participants with P/LP <i>FLCN</i> Variant	Frequency in unique DiscovEHR families with P/LP <i>FLCN</i> Variant
58-90%	8.6%	10.7%

BHD-Associated Pulmonary Phenotypes		
Published Frequency of Pulmonary Phenotypes ^{2, 9, 11}	Frequency in DiscovEHR participants with P/LP <i>FLCN</i> Variant	Frequency in unique DiscovEHR families with P/LP <i>FLCN</i> Variant
Pneumothorax 24-38% Cystic Lung Disease 67-90%	17.1% 65.7%	21.4% 67.9%

BHD-Associated Renal Cancer Phenotypes		
Published Frequency of Renal Cancer ^{2, 9, 11}	Frequency in DiscovEHR participants with P/LP <i>FLCN</i> Variant	Frequency in unique DiscovEHR families with P/LP <i>FLCN</i> Variant
Renal Cancer 12-34%	2.9%	3.5%

Table 1. Demographic Details in *FLCN* Variant Positive Patient-Participants and the Overall MyCode Cohort

Demographics	<i>FLCN</i> P/LP Variant Patient-participants (n=35)	Remaining MyCode Cohort (n=135,955)	P-value
Median Age (IQR)	62 years (54-76)	61 years (45-72)	0.20
Sex			
Female	22 (62.9%)	82,958 (61.0%)	0.86
Race			
White	34 (97.1%)	132,059 (97.1%)	0.99
Black or African American	--	2,689 (2.0%)	--
Asian	--	487 (0.4%)	--
American Indian or Alaska Native	1 (2.9%)	161 (0.1%)	--
Native Hawaiian or Pacific Islander	--	231 (0.1%)	--
Other	--	3 (0.0%)	--
Unknown	--	325 (0.2%)	--
Ethnicity			
Not Hispanic or Latino	35 (100%)	131,090 (96.4%)	0.99
Hispanic or Latino	--	2,978 (2.2%)	--
Unknown	--	1,887 (1.4%)	--
Alive During Study Period	31 (88.6%)	125,455 (92.3%)	0.34
Smoking Status			
Current Smoker	4 (11.4%)	23,094 (17.0%)	0.50
Median Length of EHR (Range)	17.43 years (0.05-22.61)	14.68 years (0.01-22.58)	--
Number with <2 years	2 (5.7%)	5,301 (3.9%)	0.40

Table 2. EHR Review for BHD-related Phenotypic Features in Variant Positive Patient-

Participants

Phenotypic Feature	Frequency in Participants with P/LP <i>FLCN</i> Variant (n=35)
BHD Diagnosis and/or Clinical Identification of <i>FLCN</i> Variant	4/35 (11.4%)
Any Personal History of BHD Cutaneous, Pulmonary, and/or Renal Phenotypes	24/35 (68.6%)
Any Cutaneous Phenotype Dermatologic Assessment (Dermatologist Exam and/or Skin Biopsy) Fibrofolliculoma, Perifollicular Fibroma, Trichodiscoma	3/35 (8.6%) 21/35 (60.0%) 3/21 (14.3%)
Any Pulmonary Phenotype Basilar Cystic Lung Disease Single Spontaneous Pneumothorax Multiple Spontaneous Pneumothoraces Lung Imaging (Adequate Visualization and Any CT/MRI Imaging) Basilar Cystic Lung Disease	23/35 (65.7%) 23/35 (65.7%) 2/35 (5.7%) 4/35 (11.4%) 13/35 (37.1%) and 29/35 (82.9%) 13/13 (100%) and 23/29 (79.3%)
Any Renal Phenotype Renal Cancer Renal Mass ^a Renal Imaging (Adequate Visualization and Any Imaging)^a Renal Mass ^a	4/35 (11.4%) 1/35 (2.9%) 3/34 (8.8%) 23/34 (67.6%) and 27/34 (79.4%) 3/23 (13.0%) and 3/27 (11.1%)
Documented Family History of BHD-related Phenotypic Feature Family History of Renal Cancer Family History of Pneumothorax Family History of Fibrofolliculomas/Papules Documented Family History of Clinical BHD or <i>FLCN</i> Variant	7/35 (20.0%) 2/35 (5.7%) 3/35 (8.6%) 2/35 (5.7%) 2/35 (5.7%)

^a excludes individual with renal cancer diagnosed

Table 3. Associations Between *FLCN* P/LP Variants and Diagnosis of BHD-related Phenotypes^a in the First- and Second-Degree Unrelated European Subset of the MyCode Cohort

Phenotype	Patient-Participants with P/LP <i>FLCN</i> Variant N=27	Remaining MyCode Cohort N=79496	Odds Ratio (95% CI)	p-value ^b
Any Pulmonary Phenotype	7 (25.9%)	6,422 (8.1%)	4.16 (1.65-9.43)	0.026*
Spontaneous Pneumothorax	6 (22.2%)	1,523 (1.9%)	15.63 (5.96-35.69)	1.92x10⁻⁵*
Pulmonary Cyst	5 (18.5%)	5,245 (18.5%)	3.40 (1.17-8.32)	0.185
Cutaneous Phenotype	10 (37.0%)	16,243 (20.4%)	2.19 (0.97-4.70)	0.405
Any Renal Phenotype	1 (3.7%)	2,200 (2.8%)	1.93 (0.21-7.43)	>0.99
Renal Mass	0 (0%)	1,693 (2.1%)	0.81 (0.01-5.81)	>0.99
Renal Cancer	1 (3.7%)	852 (1.1%)	4.94 (0.55-19.37)	0.891

^aAbility of ICD-9/10 codes in the EHR to accurately identify BHD-related phenotypes varied between phenotypes.

^bBonferroni corrected for multiple testing (7-tests).