HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2024 July 15.

Published in final edited form as:

Int J Cancer. 2023 July 15; 153(2): 364–372. doi:10.1002/ijc.34510.

A rare *FGF5* candidate variant (rs112475347) for predisposition to non-squamous, non-small-cell lung cancer

Lisa A Cannon-Albright[1,2,3], **Craig C Teerlink**[1,3], **Jeff Stevens**[1], **Julio C Facelli**[4,5], **Shamus R Carr**[6], **Kristina Allen-Brady**[1], **Sonam Puri**[2,8], **Joan E Bailey-Wilson**[7], **Genetic Epidemiology of Lung Cancer Consortium**[*], **Wallace Akerley**[2,8]

[1]Genetic Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah

[2]Huntsman Cancer Institute, Salt Lake City, Utah

[3]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah

[4]Department of BioMedical Informatics, University of Utah School of Medicine, Salt Lake City, Utah

[5]Clinical and Translational Science Institute, University of Utah School of Medicine, Salt Lake City, Utah

Corresponding author: LA Cannon-Albright, PhD, Genetic Epidemiology, 295 Williams Building, Research Park, Salt Lake City, UT 84108, lisa.albright@utah.edu, 801 587 9300.

[*]GELCC authorship list: Anthony M. Musolf[1], Bilal A. Moiz[1], Haiming Sun[1,2], Claudio W. Pikielny[3], Yohan Bossé[4], Diptasri Mandal[5], Mariza de Andrade[6], Colette Gaba[7], Ping Yang[8], Yafang Li[9], Ming You[10], Ramaswamy Govindan[11], Richard K. Wilson[12], Elena Y. Kupert[10], Marshall W. Anderson[10], Ann G. Schwartz[13], Susan M. Pinney[14], Christopher I. Amos[9], and Joan E. Bailey-Wilson[1]

1Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland. 2Laboratory of Medical Genetics, Harbin Medical University, Harbin, China. 3Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire. 4Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Quebec City, Quebec, Canada. 5Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, Louisiana. 6Mayo Clinic, Rochester, Minnesota. 7Department of Medicine, University of Toledo, Toledo, Ohio. 8Mayo Clinic, Scottsdale, Arizona. 9Baylor College of Medicine, Houston, Texas. 10Medical College of Wisconsin, Milwaukee, Wisconsin. 11Division of Oncology, Washington University School of Medicine, St. Louis, Missouri. 12Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio. 13Karmanos Cancer Institute, Wayne State University, Detroit, Michigan. 14Department of Environmental and Public Health Sciences, University of Cincinnati College of Medicine, Cincinnati, Ohio.

AUTHOR CONTRIBUTIONS

The work reported in the paper has been performed by the authors, unless clearly specified in the text. **LA Cannon-Albright:** conceptualization, methodology, investigation, resources, data curation, writing original draft, review and editing, supervision, project administration, funding acquisition. **CC Teerlink**: methodology, validation, formal analysis, data curation, writing original draft, review and editing. **J Stevens**: methodology, validation, formal analysis, data curation, writing original draft, review and editing. **JC Facelli**: methodology, validation, formal analysis, data curation, writing original draft, review and editing, visualization· **SR Carr**: resources, writing review and editing. **K Allen-Brady**: resources, writing review and editing. **S Puri**: resources, writing review and editing. **JE Bailey-Wilson**: validation, resources, data curation, writing original draft, review and editing, supervision, funding acquisition. **AM Musolf**: validation, resources, data curation, writing original draft, review and editing, **Genetic Epidemiology of Lung Cancer Consortium Authors***: validation, resources, data curation, writing review and editing. **W Akerley:** resources, writing review and editing, funding acquisition.

*Anthony M. Musolf, Bilal A. Moiz, Haiming Sun, Claudio W. Pikielny, Yohan Bossé, Diptasri Mandal, Mariza de Andrade, Colette Gaba, Ping Yang, Yafang Li, Ming You, Ramaswamy Govindan, Richard K. Wilson, Elena Y. Kupert, Marshall W. Anderson, Ann G. Schwartz, Susan M. Pinney, Christopher I. Amos, and Joan E. Bailey-Wilson

Ethics Statement This research was performed with approval from the University of Utah Institutional Review Board.

Conflict of Interest Author SP declares an advising/consultant relationship with AstraZeneca, G1 Therapeutics, and Jazz Pharma. The other authors have no conflict of interest to declare.

[6]Thoracic Surgery Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

[7]Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore MD

[8]Medical Oncology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah

Abstract

A unique approach with rare resources was used to identify candidate variants predisposing to familial non-squamous non-small-cell lung cancers (NSNSCLC). We analyzed sequence data from NSNSCLC-affected cousin pairs belonging to high-risk lung cancer pedigrees identified in a genealogy of Utah linked to statewide cancer records to identify rare, shared candidate predisposition variants. Variants were tested for association with lung cancer risk in UKBiobank. Evidence for linkage with lung cancer was also reviewed in families from the Genetic Epidemiology of Lung Cancer Consortium. Protein prediction modeling compared the mutation with reference. We sequenced NSNSCLC-affected cousin pairs from 8 high-risk lung cancer pedigrees and identified 66 rare candidate variants shared in the cousin pairs. One variant in the *FGF5* gene also showed significant association with lung cancer in UKBiobank. This variant was observed in 3/163 additional sampled Utah lung cancer cases, 2 of whom were related in another independent pedigree. Modeling of the predicted protein predicted a second binding site for SO_4 that may indicate binding differences. This unique study identified multiple candidate predisposition variants for NSNSCLC, including a rare variant in *FGF5* that was significantly associated with lung cancer risk and that segregated with lung cancer in the two pedigrees in which it was observed. *FGF5* is an oncogenic factor in several human cancers, and the mutation found here (W81C) changes the binding ability of heparan sulfate to FGF5, which might lead to its deregulation. These results support *FGF5* as a potential NSNSCLC predisposition gene and present additional candidate predisposition variants.

Keywords

non-squamous non-small cell lung cancer; UPDB; pedigree; predisposition; *FGF5*

INTRODUCTION

Lung cancer is the deadliest cancer in the United States. A variety of environmental risk factors for lung cancer have been identified; tobacco smoking is considered responsible for up to 90% of all lung cancers (1). Nevertheless, lung cancer does develop in non-smokers and is observed to cluster in relatives, suggesting the existence of inherited predispositions (2–5). Genome-wide association studies (GWAS) have identified common, low penetrance risk variants for lung cancer with moderate to small effect (e.g. 6) Studies of lung cancer families provide a powerful approach for identification of rare predisposition variants, and have been used to identify risk loci, including 6q23-25 and others (7–9) Whole exome sequencing of lung cancer cases in families has also revealed predisposition loci, including on chromosome arms 4q, 7p, and 12q (10).

Germline causal variants explain some familial lung cancer pedigrees. *TP53* is recognized to cause Li-Fraumeni syndrome, which causes an increased risk of lung and other cancers (11–12). Other causal germline variants have been reported in a small number of lung cancer families, including in *DICER1* (13); *SFTPA1* (14); the T790M mutation in *EGFR* (15–17); *HER2* (18); BAP1 (19); *PARK2* (20); and *RGSI7* (21). Nevertheless, causal predisposition genes/variants have not been identified for the majority of lung cancer families studied.

Here we used a powerful predisposition gene identification approach currently possible in Utah due to unique resources. These resources include linked population genealogy and cancer data which allows identification of pedigrees with a statistical excess of lung cancer, combined with a decades-old biorepository of DNAs from high-risk cancer pedigrees. In a previous analysis of evidence for an inherited contribution to lung cancer by histologic type and tobacco use (4) we showed that non-squamous, non-small cell lung cancer (NSNSCLC) in non-smokers was the only lung cancer subtype with significant evidence for a heritable contribution in the Utah resource. Based on these findings we initiated a pilot search for the genes/variants that predispose to NSNSCLC in a small study that made use of currently available resources. We hypothesized that sequence analysis of affected cousin pairs in high-risk NSNSCLC pedigrees with a statistical excess of NSNSCLC would identify new lung cancer predisposition loci.

MATERIALS AND METHODS

Identification of high-risk NSNSCLC pedigrees

The Utah Population Data Base (UPDB) represents the genealogy of the majority of the Utah population, from its founders in the mid 1800s to modern day (22). The ~3 million individuals in UPDB who have at least 2, and up to 12, generations of ancestors including a Utah founder were analyzed. The Utah genealogy data is linked to various statewide health data, including the Utah Cancer Registry (UCR). The UCR has recorded and validated all independent, primary cancers diagnosed or treated in Utah from 1966, and has been an NCI Surveillance, Epidemiology, and End-Results (SEER) Registry from 1973. Approximately 150,000 of the 3 million individuals with deep genealogy have linked SEER cancer data in UCR, including over 9,000 linked primary lung cancer cases.

The Genetic Epidemiology group at University of Utah has been studying and sampling high-risk Utah cancer pedigrees since the 1970s. The Genetic Epidemiology biorepository includes stored germline DNA, primarily extracted from whole blood, for ~35,000 individuals who have deep genealogy and are members of pedigrees studied for an excess of different cancers, including breast cancer, colorectal cancer, prostate cancer, and melanoma. As pedigrees were studied, relatives diagnosed with cancers of any site and recorded in the UCR, as well as some unaffected relatives, were screened and sampled; 163 germline DNAs were stored for individuals diagnosed with primary lung cancer, 101 of these sampled individuals were diagnosed with non-squamous non-small cell lung cancer (NSNSCLC).

Using the UPDB genealogy data, these 101 sampled NSNSCLC cases with deep genealogy were identified to be related in 50 independent clusters, or pedigrees, each pedigree including at least 2 of the sampled NSNSCLC cases. No pedigrees were completely

overlapping, but individuals could belong to more than 1 pedigree through different ancestors. Each of these "sampled" pedigrees was tested for a significant excess of lung cancer cases among the descendants as follows. All individuals in UPDB with deep genealogy were assigned to a cohort based on biological sex, 5-year birth year range, and birth state (Utah or not). Cohort-specific rates of lung cancer were estimated as the total number of lung cancer cases in each cohort divided by the total number of individuals with genealogy data in each cohort. For each pedigree the total number of lung cancer cases among the descendants was compared to the expected number of lung cancer cases based on the UPDB cohort-specific lung cancer rates. The expected number of lung cancer cases among the descendants was estimated as the sum of the cohort-specific rate of lung cancer for all descendants. An excess of observed to expected lung cancer cases ($p<0.05$) was used to identify those pedigrees classified as high-risk for lung cancer, only these pedigrees were included. Eight of the sampled high-risk pedigrees identified included a pair of sampled cousins who were each diagnosed with primary NSNSCLC; these 8 cousin pairs were selected for WES.

Whole Exome Sequence (WES) Data and Analysis/Identification of Candidate Variants

Whole exome sequencing was performed at the University of Colorado BioFrontiers Sequencing Facility on 8 pairs of affected cousins (n=14; some individuals were members of more than 1 high-risk pedigree through different ancestors). DNA libraries were prepared from 1.5 micrograms of DNA using the Agilent SureSelect Human All Exon (v7) capture kit. Samples were run on the Illumina NextSeq instrument at a depth of 50X. Reads were mapped to the human genome GRCh38 reference using BWA-mem for alignment and variants were called using Genome Analysis Toolkit version 4.1.3.0 (GATK) software following Broad Institute Best Practices Guidelines. Exome capture resulted in an average of 97.5% of target bases being covered by greater than 10x coverage across the exome with an average depth of 50X. Variants were called and filtered using standard GATK VQSR filters (QD < 2.0, QUAL < 30 MQRankSum < −12.5, ReadPosRankSum < −8.0, FS > 60.0, SOR > 3.0, max-gaussians=2 and truth-sensitivity-filter-level 99.0). Variants occurring outside the exon capture kit intended area of coverage were removed. Variants were annotated with Annovar, which contains predicted pathogenicity scores from 12 in-silico functional prediction algorithms. Rare, shared variants were identified from the 8 NSNSCLC-affected cousin pairs. All rare, shared variants were analyzed with the Qiagen Ingenuity Pathway Analysis package using the Variant Effects option. The sequencing coverage and quality statistics for each sample are summarized in Supplemental Table 1.

Risk association in UKBiobank

To further validate the rare, shared candidate variants identified in high-risk pedigrees, those variants with available data were analyzed for association with lung cancer (specified as ICD10 code C34) in a set of 2,123 Caucasian cases and 10,615 ancestrally matched controls from the UKBiobank's 488,377 total subjects genotyped on the Illumina OmniExpress SNP array (23). UKBiobank case and control subjects were matched via principal components (PCs) using ~27K independent markers that excluded several genomic regions known to adversely affect PC analysis (24). FLASHPCA2 software was used to generate eigenvectors for control selection (24). Controls were selected from among 191,466

Caucasian UKBiobank subjects who were over age 70 years of age and had no cancer diagnosis. Five controls, representing the nearest neighbors based on Euclidean distance of the first two PCs, were selected for each case. The selected UKBiobank case and control subjects were imputed to ~40M SNP markers using the Haplotype Reference Consortium's (HRC) 67K background genomes (25). For the imputation, we started with 784,256 observed SNP genotypes, and required sample genotyping >98% pre-imputation quality control using PLINK software (37) (no subjects removed). We further required a genotyping call rate <98%, HWE $p<1E^{-5}$, MAF<0.005, duplicated position in the HRC's reference genome, or site not included in the HRC's reference genome (353,578 markers were removed). The remaining 430,678 SNPs were converted to human genome Build 37 (hg19) forward strand orientation using Genotype Harmonizer software (38) and served as the basis for imputation. Imputation was performed with EAGLE v2.3 software for phasing (39) and MINIMAC3 software for imputation (40).

Selected candidate variants assayed in Utah lung cancer germline DNAs

As further validation of candidate variants and to identify additional Utah lung cancer case variant carriers, a subset of the candidate variants were selected for assay in the germline DNA for the set of 163 additional sampled Utah lung cancer cases available; 101 (62%) were diagnosed with NSNSCLC. Only a subset of variants was assayed due to the limited funding for the study.

Linkage Analysis of candidate variants in the Genetic Epidemiology of Lung Cancer Consortium (GELCC) Families

The GELCC is a consortium of investigators at multiple institutions who have been enrolling lung cancer patients with a family history of lung cancer and their informative family members into linkage and association studies for over 20 years (7–10, 30–31). Recently, the GELCC performed whole exome sequencing on 262 individuals from 28 European-American extended families with a strong family history of lung cancer, defined as four or more related persons with lung cancer (10); 60 of the individuals were diagnosed with lung cancer and an additional 81 unsequenced lung cancer cases were members of the final pedigree constructions. WES data was realigned and recalled using GATK, following their best practices including removing variants with depth (DP) <10, genotype quality (GQ) > 10, and GQ/DP > 0.5. Variants with less than 80% genotyping rates were removed, as were Mendelian inconsistencies using PLINK. Family relationships were confirmed using identity-by-descent calculations. Parametric linkage analysis was performed using TwoPointLods under an autosomal dominant model with a disease allele frequency of 1% and a penetrance of 80% for carriers and 1% for noncarriers. All individuals were classified as either affected or unknown for lung cancer status. LOD scores were calculated for each family and were then added across families for a cumulative LOD score at each variant; heterogeneity LOD (HLOD) scores were calculated at each variant (10). For the current study, LOD and HLOD scores from the candidate variants were obtained from the GELCC studies when available.

Protein prediction methods

The long isoform WT sequence of *FGF5* was obtained from UniProt (P12034) and the mutant sequence was manually obtained by changing residue # 81 from Tryptophan to Cysteine (32). Calculation of the protein structures and binding sites was performed using the RaptorX server with standard parameters (33). The variant pathogenicity characterization was obtained from the PolyPhen-2 predictor (34).

RESULTS

Sequence analysis/Identification of candidate variants.

Exome sequencing of the 14 NSNSCLC-affected cousins in the 8 high-risk pedigrees identified a total of ~13,400 rare variants (MAF<0.005) in gnomAD 3.0. Of these, ~7,000 variants were exon-specific (synonymous, non-synonymous, frameshift, splicing) coding variants. Removal of synonymous variants and poor-quality variants that failed variant quality score recalibration (VQSR) tranche filters resulted in ~3,800 candidate variants remaining. Removal of singleton variants (AC=1) resulted in 451 variants. Of these, 66 rare variants were identified as shared in at least one NSNSCLC-affected cousin pair from a high-risk lung cancer pedigree. Supplemental Table 2 details each of the 66 rare, shared candidate NSNSCLC predisposition variants identified. Five of the candidate genes identified had more than 1 different rare, shared variant observed (*CRIPAK, HRNR, MUC2, OR6K2*, and *VPS13A*). Five of the candidate variants were observed in more than 1 affected cousin pair (in genes AQP10, *GOLGA6L2*, *HS6ST1*, *NBPF1*, and *RPS15*)

Risk association in UKBiobank

Not all the candidate variants were present in UKBiobank data, due to the low frequency threshold. A total of 24 of the 66 rare, shared candidate predisposition variants identified had available genetic data present in UKBiobank. Only 1 of these 24 candidate predisposition variants showed significant association with lung cancer risk in the UKBiobank data; the variant in *FGF5* (rs112475347) was observed in 6/4246 case chromosomes and 7/21230 control chromosomes (p=0.013; OR = 4.29).

Candidate Variants in Additional Utah Lung Cancer Cases

We assayed a subset of the candidate variants in the 163 additional available germline DNA samples from Utah primary lung cancer cases of any histologic type. Variants assayed were in genes including *BLK, FGF5, FOXD4L3, ITGB5, MUC12, PI16, PIK3R3, PIK3R4, RBFOX1, SFTA3*, and *TRAF3IP3*. Two of these candidate genes were selected because they were noted to be significant in previous familial lung cancer studies (*MUC12* and *FOXD4L3*) (35), and one (*PI16*), although not significantly associated with lung cancer, was second only to *FGF5* in OR magnitude in the UKBiobank analysis (p=0.14, OR=1.82). The other variants were selected for assay due to their previously reported association with lung cancer risk, or identification with lung cancer pathways in Ingenuity. Three of the variants failed assay design at the manufacturer (in genes *ITGB5, FOXD4L3*, and *BLK*) and one failed to perform (in gene *MUC12*). For most variants assayed no variant carriers were identified among the 163 additional Utah lung cancer cases assayed (as expected for these

very rare candidate variants). Three additional lung cancer case carriers of the *FGF5* variant were the only carriers identified for any variant assayed.

Figure 1 shows the sequenced high-risk lung cancer pedigree which includes the first-cousin NSNSCLC-affected pair carrying the rare *FGF5* variant. The founders of the pedigree were born in the early 1800s in Scotland and had almost 3,000 descendants recorded in the UPDB. Among the descendants there were 6 NSNSCLC cases with only 1.9 cases expected (p=0.013); the two NSNSCLC affected cousin cases are shown with an arrow and with a "+" for variant carrier status. Of possible interest, this pedigree also had a significant excess of prostate cancers among all descendants of the founders shown (26 observed, 10.4 expected, p=3.4e-5; prostate case data not shown). Two of the other lung cancer cases in this pedigree were also diagnosed with prostate cancer (not shown), and the father of the female variant carrier shown in Figure 1 was diagnosed with prostate cancer. Decade of diagnosis for lung cancer is shown below each case.

In addition to the original sequenced pair of *FGF5* case-cousin-carriers, the *FGF5* assay of 163 additional sampled Utah lung cancer cases identified 3 additional lung cancer case carriers of the *FGF5* variant, 2 of whom were second cousins in an independent pedigree. Figure 2 shows these 2 related additional *FGF5* lung cancer case variant-carriers identified. The 2 lung cancer cases are fully shaded; the lung cancer case diagnosed in their 90s was diagnosed with a squamous histology, the other individual was diagnosed with a NSNSCLC. Of possible interest, inferred carriers among the individuals shown in Figure 2 (different ancestors of carriers, cancer status not shown) were diagnosed with colon cancer, stomach cancer, and melanoma.

Linkage Analysis in GELCC Families

As expected due to the low frequencies, most of the 66 rare candidate variants were not observed in any members of the GELCC lung cancer families. Those variants which were observed included 1 unaffected *EXTL1* variant heterozygous (het) carrier, 2 affected *TRABD2B* het carriers, 1 unaffected *OR6K2* het carrier, 1 unaffected *SYNE1* het carrier, 6 unaffected and 1 affected *ALDH1B1* het carrier, 1 unaffected *OR56A4* het carrier, 1 unaffected *PIWIL4* het carrier, 3 unaffected *NRIP2* het carriers, 1 unaffected *OR4F6* het carrier, and 1 unaffected and 1 affected *PKD1L2* het carrier. The rare *FGF5* variant observed in two Utah pedigrees was not observed in the GELCC lung cancer families. Twenty-one other *FGF5* variants were observed in the 28 GELCC lung cancer families and LOD and HLOD scores were calculated. No HLODs> 0.30 were observed for any of these *FGF5* variants.

Protein Prediction

PolyPhen-2 predicts the rare *FGF5* candidate variant observed here as a pathogenic variant, scoring it at 0.993 (sensitivity 0.70, specificity 0.97). The RaptorX WT and MUT structures show a great deal of similarity; for instance, comparison of RaptorX top models for the WT and MUT proteins (See Figure 3) show an RMS of 27.06 Å across 268 pairs, which is reduced to 0.48 Å when the comparison is pruned to the most important 131 pairs. It is also observed that the structure in the vicinity of the mutation does not show any significant

changes. Note also that the apparent change of the orientation of the signal peptide (residues 1-20) is not significant, as in both cases the signal peptide is exposed to the solvent. RaptorX predicts only one binding site for the WT of FGF5 with SO_4, this site involves residues V95, N193, K194, K199, H210, I211 and S212 with multiplicity 124. The same binding site is predicted for the mutant but with multiplicity 123, and a second binding site for SO_4 is predicted for the mutant with multiplicity 27, which may indicate a less reliable prediction (33), and involving residues N193, K199 and R205.

DISCUSSION

This unique, and small, but powerful, sequencing study for NSNSCLC was designed as a feasibility study to show the power of the high-risk pedigree approach for lung cancer predisposition gene identification. Using conveniently available stored germline DNAs from NSNSCLC cases and a population-based genealogy for Utah, NSNSCLC-affected cousin pairs who belonged to pedigrees exhibiting a significant excess of lung cancer cases were identified and sequenced. Since cousins are only expected to share 1/8 of their genetic material, and these cousins belong to validated high-risk lung cancer pedigrees, any rare variants shared in these affected cousin pairs likely represent strong candidate predisposition variants. Association with lung cancer risk in a large case/control data set from the independent UKBiobank population validated one of these candidate variants, in *FGF5*. In addition to the original variant-sharing cousin pair sequenced, an additional pair of related *FGF5*-variant-carrying Utah lung cancer cases was identified. Since many specific-site cancer predisposition genes (e.g. BRCA1, CDKN2A) have later been described to have a role in predisposition to cancers of other sites, and since *FGF5* is an oncogenic factor in several other human cancers, we note that it may be of interest that while only 2 lung cancer case carriers were identified in each of the 2 small pedigrees presented, additional inferred variant carriers (ancestors of the variant carriers shown) in each pedigree were identified to have been diagnosed with cancers of other sites.

The set of 66 candidate variants presented in Supplemental Table 2 represents a strong set of candidate predisposition variants for NSNSCLC as all are rare variants found to be shared in affected cousins belonging to high-risk lung cancer pedigrees. The specific results for the *FGF5* variant suggest it may predispose to lung cancer in the two Utah pedigrees identified. Further validation from extensions of these pedigrees, as well as in independent populations is required, but this study shows the power of this approach, even with a small number of high-risk pedigrees, and should motivate larger sequencing studies of high-risk pedigrees, many more of which are accessible using the Utah resources described.

FGF5, fibroblast growth regulator-5 is a regulator of human hair length that has been associated with cancer risk, and specifically lung cancer risk, in many studies. Overexpression of *FGF5* has been observed in multiple malignancies, including hepatocellular carcinoma (36), colorectal cancer (37), breast carcinoma (38), melanoma (39) and astrocytic brain tumors (40). Zhou (41) reported that downregulation of *FGF5* inhibits cell growth and invasion of human non-small cell lung cancer cells, and Zhao (42) has documented that increase of *FGF5* expression correlates with poor survival in lung adenocarcinoma. Finally, recognizing the importance of *FGF5* overexpression in oncogenic

processing, RNA aptamers that inhibit *FGF5* expression have been proposed as a potential therapy for cancers dependent on *FGF5* cell proliferation (43).

The protein structure modeling results do not provide an indisputable mechanism for pathogenicity of the W81C mutation reported here; the overall and local structural changes observed are modest at best. Nonetheless, the prediction of a second binding site for SO_4 in the mutant, albeit with a smaller multiplicity, may be an indication that the mutant's ability to bind to SO_4-containing compounds, like heparan sulfate, may be different. A large body of biochemical and cellular evidence points to a direct role for heparin/heparan sulfate in the formation of an active FGF/FGF receptor signaling complex (45–47). Specifically, Rapraeger et al. (44), shows that heparan sulfate binding to *FGF5* is central to its regulation, hereby we suggest that the mutation found here (W81C), may change the binding ability of heparan sulfate to *FGF5*, leading to its deregulation, which is already associated with oncogenesis (48). Further docking studies of the differential binding affinity of the mutant, when compared with the WT, will be necessary to further explore this preliminary hypothesis (49).

This study has multiple strengths, including the high-quality SEER cancer diagnosis data for all cancers, the linked genealogy data allowing identification of each individual pedigree as "high-risk" for lung cancer, and the extended relationships represented in the UPDB. The Genetic Epidemiology biorepository at University of Utah has been the source of many high-risk cancer pedigree studies and these studies have been critical to the identification of multiple cancer predisposition genes, including *BRCA1, BRCA2*, and *CDKN2A* (50–52). More recently, similar studies of small sets of high-risk pedigrees have presented multiple new cancer predisposition variants, including *GOLM1*-melanoma (53) and *ERF* –prostate/bladder cancer (54), among others. While many lung cancers are associated with tobacco use, Utah is recognized to have the lowest rates of tobacco use in the nation, which potentially served to reduce noise from this recognized lung cancer risk factor which can also cluster in relatives. Limitations of this study include the small sample size of only 8 high-risk pedigrees and the limited number of candidate variants with case/control data available in UKBiobank for validation. While the cases sequenced were NSNSCLC cases, the pedigrees were identified as high-risk for lung cancer generally, and the UKBiobank identification was of lung cancer cases only; therefore, it is possible that these findings are not specific to NSNSCLC. Finally, the population represented by the UPDB genealogy data for founders and their descendants is largely Northern European (22), and therefore the conclusions presented cannot be extrapolated beyond this population without additional studies.

This small study suggests the utility of a Utah population-wide study of all available high-risk NSNSCLC pedigrees. The two primary cancer care providers in the state have biorepositories dating back decades. Although germline DNA from blood has typically not always been stored, germline DNA potentially could be obtained from uninvolved tissues from the margins of stored formalin fixed, paraffin embedded (FFPE) tumor samples. Such a study would be powerful not only for the extended pedigrees/relationships that could be studied, but for the ability to obtain samples across several generations for individuals who have already succumbed to this deadly cancer. The identification of even a rare cancer

predisposition variant, such as reported here, if validated for pathogenicity, is valuable not only for the relatives of any carriers identified, but for the information it can provide generally on cancer predisposition gene function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This research was supported by the Utah Cancer Registry, which is funded by the National Cancer Institute's SEER Program, Contract No. HHSN261201800016I, the US Center for Disease Control and Prevention's National Program of Cancer Registries, Cooperative Agreement No. NU58DP0063200-01, with additional support from the University of Utah and Huntsman Cancer Foundation. Partial support for all datasets within the Utah Population Database is provided by the University of Utah, Huntsman Cancer Institute, and the Huntsman Cancer Institute Cancer Center Support grant, P30 CA42014 from the National Cancer Institute. LACA received partial support from the Huntsman Cancer Institute Cancer Center Support grant, P30 CA42014 from the National Cancer Institute. This project was funded in part by HCI Cancer Control and Population Sciences Program Pilot Project award number 190105, activity 30189. Some of this research was conducted using the UK Biobank Resource under Application Number 43460. JEBW and AMM were supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. The Genetic Epidemiology of Lung Cancer Consortium data collection was supported in part by Department of Health and Human Services contracts from the National Human Genome Research Institute, NIH: HHSN26820100007C, HHSN268201700012C, and HHSN75N92020C00001 and by grants U01CA76293 and U01CA243483 from the National Cancer Institute, National Institutes of Health. JCF was partially funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002538.

Data Availability Statement

The raw WES data generated in this study have been submitted to the NCBI BioProject database with the ID: PRJNA936860. Other data that support the findings of this study are available from the corresponding author upon request and with permission of UPDB.

Abbreviations:

GATK	Genome Analysis Tool Kit
GELCC	Genetic Epidemiology of Lung Cancer Consortium
GWAS	genome wide association study
HET	heterozygote
HLOD	heterogeneity Logarithm of the Odds score
HRC	Haplotype Reference Consortium
MAF	minor allele frequency
NSNSCLC	non squamous, non small cell lung cancer
PC	principal components
SEER	Surveillance, Epidemiology, and End-Results
UCR	Utah Cancer Registry

UPDB	Utah Population Data Base
WES	whole exome sequencing

REFERENCES

1. Mattson ME, Pollack ES, Cullen JW. What are the odds that smoking will kill you? American Journal of Public Health. 1987;77:425–431. [PubMed: 3826460]
2. Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. JNCI. 1963;30:289–312. [PubMed: 13985327]
3. Cannon-Albright LA, Thomas A, Goldgar DE, Gholami K, Rowe K, Jacobsen M, McWhorter WP, Skolnick MH. Familiality of cancer in Utah. Cancer Res 1994 May 1;54(9):2378–2385. [PubMed: 8162584]
4. Carr SR, Akerley W, Cannon-Albright LA. Genetic contribution to Non-squamous, Non-small cell lung cancer in non-smokers. J Thorac Oncol. 2018 Jul;13(7):938–945. [PubMed: 29626620]
5. Ooi WL, Elston RC, Chen VW, Bailey-Wilson JE, Rothschild H. "Increased familial risk for lung cancer," J.N.C.I 76:217–222 (1986).
6. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lung E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcatova I, Merletti F, Kjaerheim K, Agudo A Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P., A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit. Genes on 15q25. Nature 2008;452:633–637. [PubMed: 18385738]
7. Bailey-Wilson JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, Wiest JS, Fain P, Schwartz AG, You M, Franklin W, Klein C, Gazdar A, Rothschild H, Mandal D, Coons T, Slusser J, Lee J, Gaba C, Kupert E, Perez A, Zhou X, Zeng D, Liu Q, Zhange Q, Seminara D, Minna J, Anderson MW . A major lung cancer susceptibility locus maps to chromosome 6q23-25. AM J Hum Genet 2004;75:460–474. [PubMed: 15272417]
8. Musolf AM, Simpson CL, de Andrade M, Mandal D, Gaba C, Yang P, Li Y, You M, Kupert EY, Anderson MW, Schwartz AG, Pinney SM, Amos CI, Baile-Wilson JE. Parametric linkage analysis identifies five novel genome-wide significant loci for familial lung cancer. Hum Hered 2016;82:64–74. [PubMed: 28817824]
9. Musolf AM, Simpson CL, Moiz BA, Pikielny CW, Middlebrooks CD, Mandal D, de Andrade M, Cole MD, Gaba C, Yang P, You M, Li Y, Kupert EY, Anderson MW, Schwartz AG, Pinney Sm, Amos CI, BaileyWilson JE. Genetic Variation and Recurrent Haplotypes on Chromosome 6q23-25 Risk Locus in Familial Lung Cancer. Cancer Res. 2021 Jun 15;81(12):3162–3173. doi: 10.1158/0008-5472.CAN-20-3196. Epub 2021 Apr 14. [PubMed: 33853833]
10. Musolf AM, Moiz BA, Sun H, Pikielny CW, Bosse Y, Mandal D, de Andrade M, Gaba C, Yang P, Li Y, You M, Govindan R, Wilson RK, Kupert EY, Anderson MW, Schwartz AG, Pinney SM, Amos CI, Bailey-Wilson JE.. Whole exome sequencing of highly aggregated lung cancer families reveals linked loci for increased cancer risk on chromosome 12q, 7p and 4q. Cancer Epi Biom Prev 2020 Feb:29(2):434–442.
11. Kerrigan K, Chan J, Vagher J, Kohlmann W, Naumer A, Anson J, Low S, Schiffman J, Maese L. Lung cancer in Li-Fraumeni Syndrome. JCO Precision Oncology 2021 :5, 552–556. DOI: 10.1200/PO.20.00468 Published online March 23, 2021.
12. Ceyhan-Birsoy O, Selenica P, Chui MH, Jayakumaran G, Ptashkin R, Misyura M, Aypar U, Jairam S, Yang C, Li Y, Mehta N, Kemel Y, Salo-Mullen E, Maio A, Sheehan M, Zehir A, Carlo M, Latham A, Stadler Z, Robson M, Offit K, Ladanyi M, Walsh M, Reis-Filho JS, Mandelker Dl. Paired normal-tumor sequencing provides insights in TP53 spectrum in patients with Li-Fraumeni syndrome. J Natl Cancer Inst. 2021. doi: 10.1093/jnci/djab117.

13. De Kock L, Bah I, Wu Y, Xie M, Priest JR, Foulkes WD. Germline and somatic DICER1 mutations in a well-differentiated fetal adenocarcinoma of the lung. J Thorac Oncol 2016 11 e31–e33. [PubMed: 26886166]

14. Nathan N, Giraud V, Picard C, Nunes H, Dastot-Le Moal F, Copin B, Galeron L, De Ligniville A, Kuziner N, Reynaud-Gaubert M, Valeyre D, Couderc LJ, Chinet T, Borie R, Crestani B, Simansour M, Nau V, Tissier S, Duquesnoy P, Mansour-Hendili L, Legendre M, Kannengiesser C, Coulomb-L'Herime A, Gouya L, Amselem S, Clement A. Germline SFTPA1 mutation in familial idiopathic interstitial pneumonia and lung cancer. Hum Mol Genet 2016, 25:1457–1467. [PubMed: 26792177]

15. Bell DW, Gore I, Okimoto RA, Godin-Heymann N, Sordella R, Mulloy R, Sharma SV, Brannigan BW, Mohapatra G, Settleman J, Haber DA. Inherited susceptibility to lung cancer may be associated with the T790M durg resistance mutation in EGFR. Nat Genet 2005, 37:1315–1316. [PubMed: 16258541]

16. Lou Y, Pecot CV, Tran HT, DeVito VH, Tang XM, Heymach JV, Luthra R, Wistuba II, Zuo Z, Tsao AS. Germline mutation of T790M and dual/multiple EGFR mutations in patients with lung adenocarcinoma. Clin Lung Cancer 2016, 17, e5–e11. [PubMed: 26700910]

17. Yu HA, Arcila ME, Fleischut MH, Stadler Z, Ladanyi M, Berger MF, Robson M, Riely GJ. Germline EGFR T790M mutation found in multiple members of a familial cohort. J Thorac Oncol 2014;9:554–558. [PubMed: 24736080]

18. Yamamoto H, Higasa K, Sakaguchi M, Shien K, Soh J, Ichimura K, Furukawa M, Hashida S, Tsukuda K, Takigawa N, Matsuo K, Kiura K, Miyoshi S, Matsuda F, Toyooka S. Novel germline mutation in the transmembrane domain of HER1 in familial lung adenocarcinomas. JNCI 2014.

19. Abdel-Rahman MH, Pilarski R, Cebulla CM, Massengill JB, Christopher BN, Boru G, Hovland P, Davidorf FH. Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma and other cancers . J Med Genet 2011;48:856–859. [PubMed: 21941004]

20. Xiong D, Wang Y, Kupert E, Simpson C, Pinney SM, Gaba CR, Mandal D, Schwartz AG, Yang P, de Andrade M, Pikielny C, Byun J, Li Y, Stambolian D, Spitz MR, Liu Y, Amos CI, Bailey-Wilson JE, Anderson M, You M. A recurrent mutation in PARK2 is associated with familial lung cancer. Am J Hum Genet 2015;96:301–308. [PubMed: 25640678]

21. You M, Wang DL, Liu PY, Vikis H, James M, Lu Y, Wang Y, Wang M, Chen Q, Jia D, Liu Y, Wen W, Yang P, Sun Z, Pinney SM, Zheng W, Shu X-O, Long J, Gao Y-T, Ziang Y-B, Chow W-H, Rothman N, Petersen GM, de Andrade M, Wu Y, Cunningham JM, Wiest JS, Fain PR, Schwartz AG, Girad L, Gazdar A, Gaba C, Rothschild H, Mandal D Coons T, Lee J, Kupert E, Seminara D, Minna J, Bailey-Wilson JE, Amos CI, Anderson MW. Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene. Clin Cancer Res 2009;15:2666–2674. [PubMed: 19351763]

22. Cannon-Albright LA. (2008). Utah family-based analysis: past, present and future. Hum Hered, 65(4), 209–20. [PubMed: 18073491]

23. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sporsen T, Peakman T, Collins R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med, 12, e1001779. doi: 10.1371/journal.pmed.1001779 [PubMed: 25826379]

24. Abraham G, Qiu Y, & Inouye M (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics, 33, 2776–2778. doi: 10.1093/bioinformatics/btx299 [PubMed: 28475694]

25. The Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet, 48, 1279–1283. doi: 10.1038/ng.3643 [PubMed: 27548312]

26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 81, 559–575. doi: 10.1086/519795 [PubMed: 17701901]

27. Deelen P, Bonder MJ, van der Velde KJ, Westra H, Winder E, Hendriksen D, Franke L, Swertz MA. (2014). Genotype harmonizer: automatic strand alignment and format conversion for

genotype data integration. BMC Research Notes, 7, 901. doi: 10.1186/1756-0500-7-901 [PubMed: 25495213]

28. Loh PR, Palamara PF, & Price AL (2016). Fast and accurate long-range phasing in a UK Biobank cohort. Nat Genet, 48, 811–816. doi: 10.1038/ng.3571 [PubMed: 27270109]
29. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh P-R, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. (2016). Next-generation genotype imputation service and methods. Nat Genet, 48, 1284–1287. doi: 10.1038/ng.3656 [PubMed: 27571263]
30. Byun J, Schwartz AG, Lusk C, Wenzlaff AS, de Andrade M, Mandal D, Gaba C, Yang P, You M, Kupert EY, Anderson MW, Han Y, Li Y, Qian D, Stilp A, Laurie C, Nelson S, Zheng W, Hung RJ, Gaborieau V, Mckay J, Brennan P, Caporaso NE, Landi MT, Wu X, McLaughlin JR, Brhane Y, Bossé Y, Pinney SM, Bailey-Wilson JE, Amos CI. Genome-wide association study of familial lung cancer. Carcinogenesis. 2018 Sep 21;39(9):1135–1140. [PubMed: 29924316]
31. Amos CI, Pinney SM, Li Y, Kupert E, Lee J, de Andrade MA, Yang P, Schwartz AG, Fain PR, Gazdar A, Minna J, Wiest JS, Zeng D, Rothschild H, Mandal D, You M, Coons T, Gaba C, Bailey-Wilson JE, Anderson MW. A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. Cancer Res. 2010 Mar 15;70(6):2359–67. doi: 10.1158/0008-5472.CAN-09-3096. Epub 2010 Mar 9. [PubMed: 20215501]
32. UniProt, Consortium. (January 2015). “UniProt: a hub for protein information”. Nucleic Acids Research. 43 (Database issue): D20412. doi:10.1093/nar/gku989.
33. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nature Protocols 7, 1511–1522, 2012 [PubMed: 22814390]
34. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249 (2010). [PubMed: 20354512]
35. Kanwal M, Ding XJ, Ma ZH, Li LW, Wang P, Chen Y, Huang Y-C, Cao Y. Characterization of germline mutations in familial lung cancer from the Chinese population. Gene 2018 Jan 30;641:94–104. [PubMed: 29054765]
36. Fang F, Chang R-M, Yu L, Lei X, Xiao S, Yang H, Yang L-Y. 2015. MicroRNA-188-5p suppresses tumor cell proliferation and metastasis by directly targeting FGF5 in hepatocellular carcinoma. J. Hepatol 63, 874–885 [PubMed: 25998163]
37. Mitchell SM, Ross JP, Drew HR, Ho T, Brown GS, Saunders NFW, Duesing KR, Buckley MJ, Dunne R, Beetson I, Rand KN, McEvoy A, Thomas ML, Baker RT, Wattchow DA, Young GP, Lockett TJ, Pedersen SK, Lapointe LC, Mollow PL. 2014. A panel of genes methylated with high frequency in colorectal cancer. BMC Cancer 14, 54. [PubMed: 24485021]
38. Huang Y, Wang H, and Yang Y 2018. Expression of fibroblast growth factor 5 (FGF5) and its influence on survival of breast cancer patients. Med. Sci. Monit. 24, 3524–3530. [PubMed: 29804124]
39. Ghassemi S, Vejdovszky K, Sahin E, Ratzinger L, Schelch K, Mohr T, Peter-Vorosmarty B, Brankovic J, Lackner A, Leopoldi A, Meindl D, Pirker C, Hegedus B, Marian B, Holzmann K, Grasi-Kraupp B, Heffeter P, Berger W, Brusch M. 2017. FGF5 is expressed in melanoma and enhances malignancy in vitroandin vivo. Oncotarget 8, 87750–87762. [PubMed: 29152117]
40. Allerstorfer S, Sonvilla G, Fischer H, Spiegl-Kreinecker S, Gauglhofer C, Setinek U, Czech T, Marosi C, Buchroithner J, Pichler J, Silye R, Mohr T, Holzmann K, Grasl-Kraupp B, Marian B, Grusch M, Fischer J, Micksche M, Berger W. 2008. FGF5 as an oncogenic factor in human glioblastoma multiforme: Autocrine and paracrine activities. Oncogene 27, 4180–4190. [PubMed: 18362893]
41. Zhou Y, Yu Q, Chu Y, Zhu X, Deng J, Liu Q, Wang Q. Downregulation of fibroblast growth factor 5 inhibits cell growth and invasion of human non-small-cell lung cancer cells. J Cell Biochem 2018 Dec 5.
42. Zhao T, Qian K, Zhang Y. High expression of FGF5 is an independent prognostic factor for poor overall survival and relapse-free survival in lung adenocarcinoma. J Comput Biol 2020 Jun;27(6):948–057. [PubMed: 31553229]

43. Amano R, Namekata M, Horiuchi M, Saso M, Yanagisawa T, Tanaka Y, Ghani FI, Yamamoto M, Sakamoto T. Specific inhibition of FGF5-induced cell proliferation by RNA aptamers. Nature Scientific reports 2021 11:2976.
44. Rapraeger AC. In the clutches of proteoglycans: how does heparan sulfate regulate FGF binding? Chem Biol. 1995 Oct;2(10):645–9. doi: 10.1016/1074-5521(95)90025-x. [PubMed: 9383470]
45. Ornitz DM. FGFs, heparan sulfate and FGFRs: complex interactions essential for development. Bioessays. 2000 Feb;22(2):108–12. [PubMed: 10655030]
46. Olwin BB, Arthur K, Hannon K, Hein P, McFall A, Riley B, Szebenyi G, Zhou Z, Zuber ME, Rapraeger AC, et al. Role of FGFs in skeletal muscle and limb development. Mol Reprod Dev. 1994 Sep;39(1):90–100; [PubMed: 7999366]
47. Reuss B, von Bohlen und Halbach O. Fibroblast growth factors and their receptors in the central nervous system. Cell Tissue Res. 2003 Aug;313(2):139–57. [PubMed: 12845521]
48. Klagsbrun M, The fibroblast growth factor family: Structural and biological properties, Progress in Growth Factor Research, Volume 1, Issue 4, 1989, Pages 207–235. [PubMed: 2491263]
49. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. 2011 Jun;7(2):146–57. doi: 10.2174/157340911795677602. [PubMed: 21534921]
50. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright LA, Goldgar D, Wiseman R, Kamb A, Skolnick MH. (1994). A Strong Candidate for the Breast and Ovarian Cancer susceptibility Gene BRCA 1. Science, 266, 66–71. [PubMed: 7545954]
51. Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Mitchell JT, McArthur-Morrison J, Nguyen K, Peng Y, Samson C, Schroeder M, Snyder SC, Steele L, Stringfellow M, Stroup C, Swedlund B, Swense J, Teng D, Thomas A, Tran T, Tranchant M, Weaver-Feldhaus J, Wong AK, Shizuya H, Eyfjord JE, Cannon-Albright L, Tranchant M, Labrie F, Skolnick MH, Weber B, Kamb A, Goldgar DE. (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nat Genet, 12(3), 333–7. [PubMed: 8589730]
52. Kamb A, Shattuck-Eidens D, Eeles R, Liu Q, Gris NA, Ding W, Hussey C, Tran T, Miki Y, Weaver-Feldhaus J, McClure M, Aitken JF, Anderson DE, Bergman W, Frants R, Goldgar DE, Green A. MacLennan R, Martin NG, Meyer LJ, Youl P, Zone JJ, Skolnick MH, Cannon-Albright LA (1994). Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat Genet, 8(1), 23–6. [PubMed: 7987388]
53. Teerlink CC, Huff C, Stevens J, Yu Y, Holmen SL, Silvis MR, Trombetti K, Zhao H, Grossman D, Farnham JM, Wen J, Facelli JC, Thomas A, Babst M, Florell SR, Meyer L, Zone JJ, Leachman SA, Cannon-Albright LA.A non-synonymous variant in *GOLM1* in cutaneous malignant melanoma. JNCI, 2018 Dec 1;110(12):1380–1385. [PubMed: 29659923]
54. Cannon-Albright LA, Teerlink CC, Stevens J, Huang FW, Sipeky C, Schleutker J, Hernandez R, Facelli J, Agarwal N, Trump DL A Rare Variant in ERF (rs144812092) Predisposes to Prostate and Bladder Cancers in an Extended Pedigree. 2021 Cancers, 13(10), 2399.

NOVELTY AND IMPACT

This manuscript reports a powerful and uncommon approach analyzing validated high risk lung cancer pedigrees to identify rare candidate predisposition variants shared in affected relative pairs. This unique study identified multiple candidate predisposition variants for NSNSCLC, including a rare variant in *FGF5* that was significantly associated with lung cancer risk in an independent population, and that segregated with lung cancer in the two pedigrees in which it was observed.

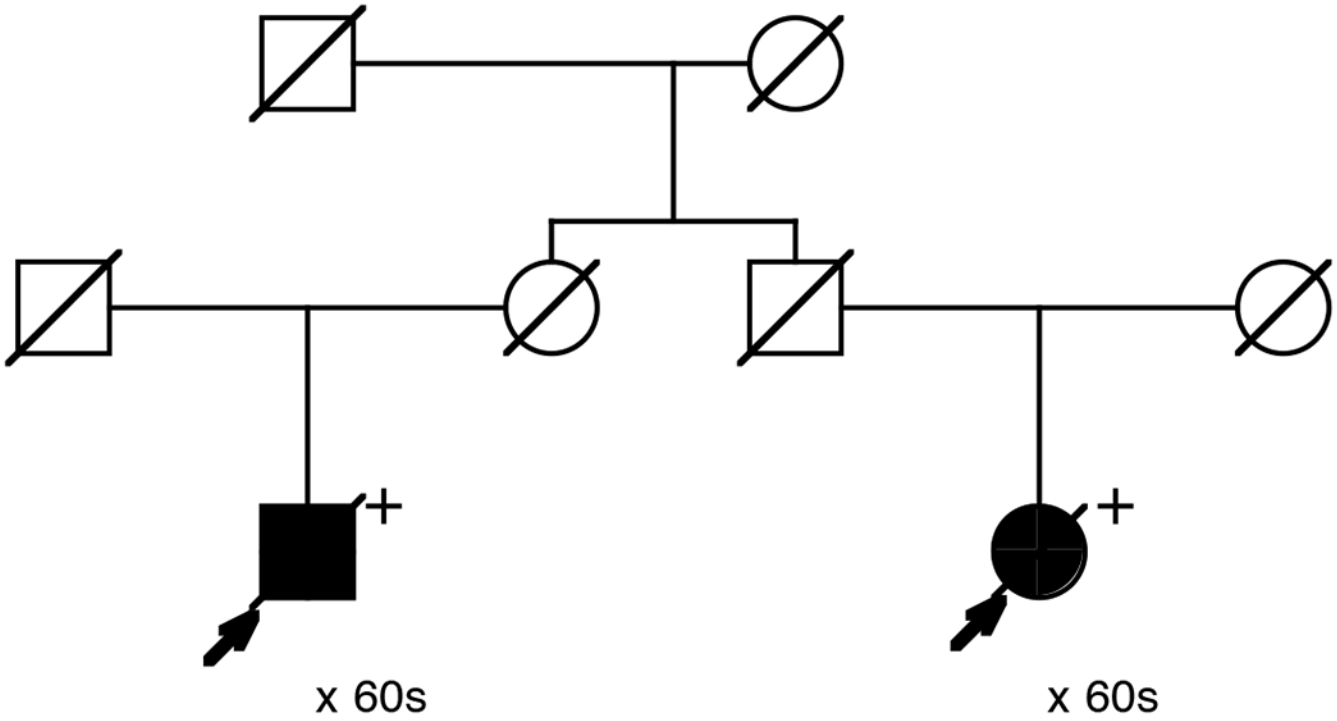

Figure 1.
The sequenced high risk NSNSCLC pedigree including a lung-cancer-affected cousin pair sharing the rare *FGF5* variant.

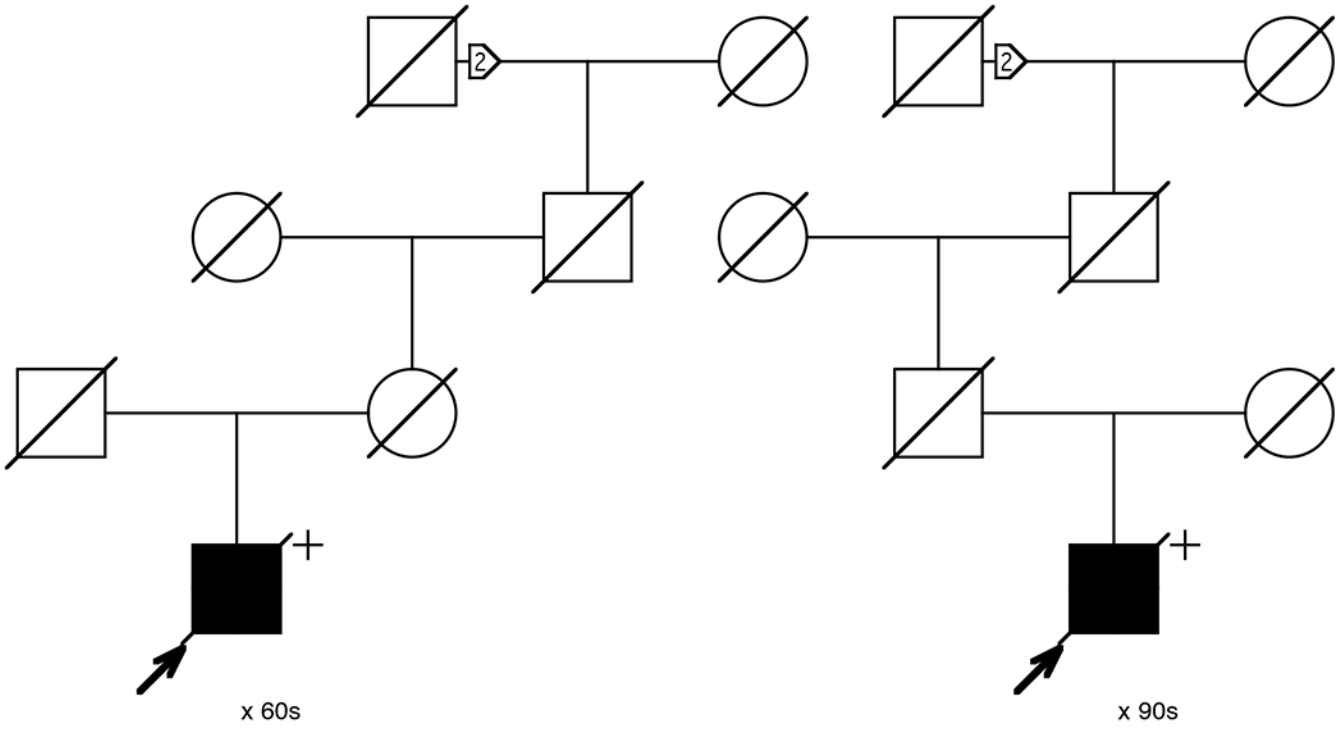

Figure 2.

Two related *FGF5* variant-carrying lung cancer cases identified by assay. The founding male had 2 marriages, shown with an arrow on the top marriage line.

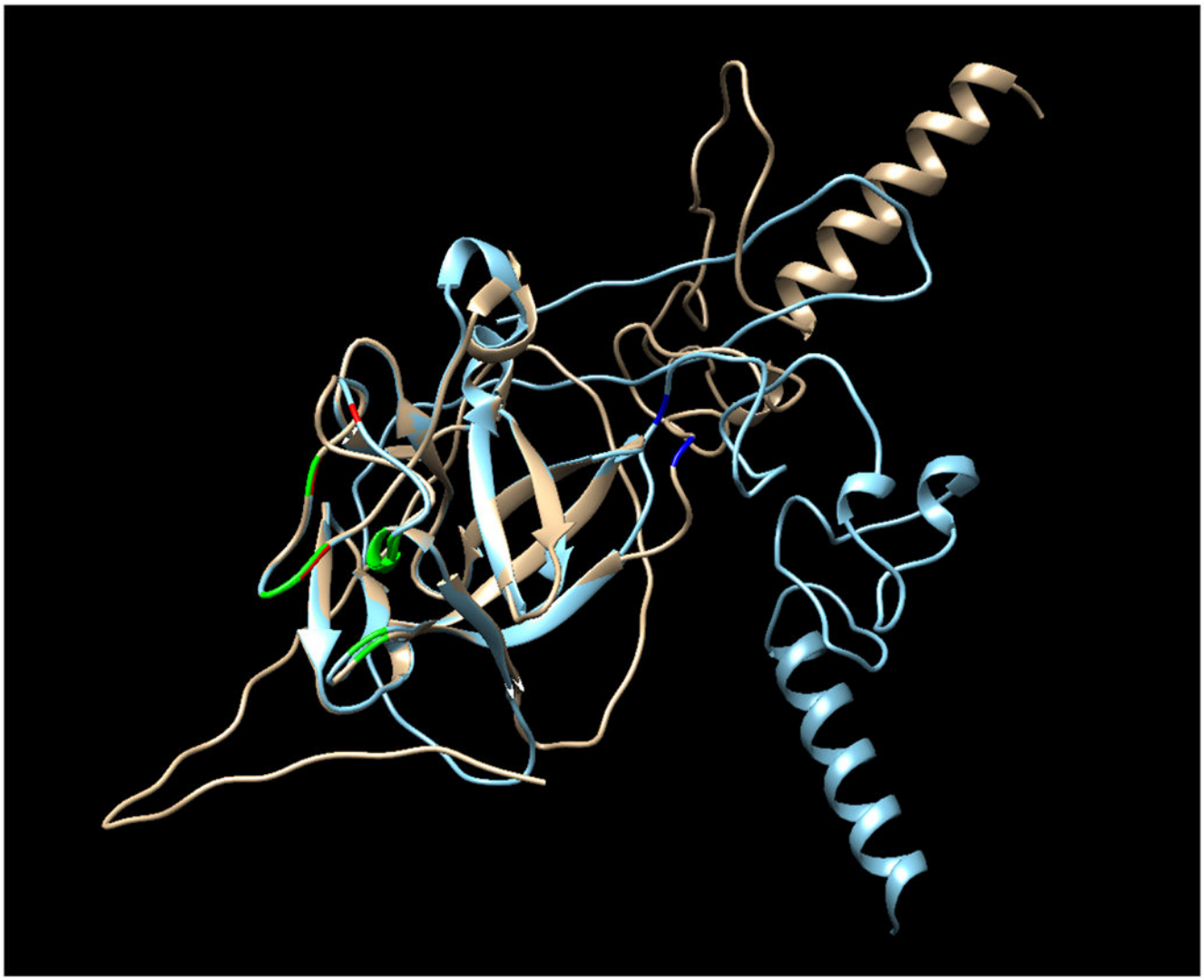

Figure 3.

Superposition of the top models predicted by RaptorX for the WT (in gold) and MUT (in light blue) structures of FGF5. The mutation site W81C is marked in dark blue for both structures, showing very little structural change. The binding sites with SO_4 predicted by RaptorX are depicted in green (V95, N193, K194, K199, H210, I211, S212) for both the WT and MUT structure; the new binding site found for the MUT structure (N193, K199 and R205) is shown in RED. The structure in the vicinity of the binding sites is not substantially changed by the mutation.