# SF 424 R&R and PHS-398 Specific
# Table of Contents

| | | |
|---|---|---|
| **PI: Neklason, Deborah Wood** | Title: Genetic and Environmental Etiology of Familial Small Intestinal Carcinoid Cancer | |
| | FOA: PAR13-146 | |
| | FOA Title: NCI EXPLORATORY/DEVELOPMENTAL RESEARCH GRANT PROGRAM (NCI OMNIBUS R21) | |
| | Organization: UNIVERSITY OF UTAH | |
| *Senior/Key Personnel:* | *Organization:* | *Role Category:* |
| DEBORAH NEKLASON | University of Utah | PD/PI |

## RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved? ☒ Yes ☐ No

1.a If YES to Human Subjects

Is the Project Exempt from Federal regulations? ☐ Yes ☒ No

If NO, is the IRB review pending? ☒ Yes ☐ No

2. Are Vertebrate Animals Used? ☐ Yes ☒ No

3. Is proprietary/privileged information included in the application? ☐ Yes ☒ No

4a. Does this project have an actual or potential impact on the environment? ☐ Yes ☒ No

5. Is the research performance site designated, or eligible to be designated, as a historic place? ☐ Yes ☒ No

6. Does this project involve activities outside of the United States or partnerships with international collaborators? ☐ Yes ☒ No

**Project Summary/Abstract**

The incidence of carcinoid cancers, in particular those of the small intestine, have increased four-fold over the past 30 years. Unfortunately, the majority of these small intestinal carcinoid cancers (SICC) are diagnosed with advanced disease where prognosis is poor. Diagnosis is often missed or delayed by nearly 10 years because symptoms are ambiguous. These cancers have a major hereditary component, much of which is not explained by the rare genetic syndromes. First-degree relatives of SICC cases have an approximately 10-fold relative risk of developing SICC. Interestingly the risk to siblings is double that to parents which suggests that some aspect of a shared environment compounds the risk of developing SICCs. We hypothesize that environmental exposures and inherited genetic factors contribute to the development of small intestinal carcinoid cancers and that the interaction of these two components drives penetrance. We have the unique opportunity to use a one-of-a-kind resource to investigate environmental exposures AND inherited genetic risk factors which contribute to the development of SICCs. The resource we will use is the Utah Population Database (UPDB), a computerized integration of 7 million individuals and genealogies dating back to the 1800s. This is overlaid with statewide vital statistics, SEER cancer records, medical records, and public records. UPDB is recently expanded to include geo-spatial and environmental data to promote epidemiological research and gain insights into gene-environment interactions. This exploratory project will set the stage for future epidemiological research using this powerful resource. The aims of this proposal are as follows. 1. To identify genetic variants that explain high-risk familial SICC. We have identified 20 large multi-generation pedigrees with 3 to 7 SICC cases and statistical excess of SICC. Cases and family members will be enrolled in research, focusing on families with highest excess risk. DNA from blood and archived tumor/tissue blocks will be obtained. Whole genome sequencing will be applied to affected individuals in these large extended pedigrees to identify genetic variants predicted to be responsible for familial SICC. 2. To model environmental risk profiles of SICC using geo-coding, place and time of residence, and environmental exposure data. Using historical residential addresses from public resources, we can identify when, how long, and where each individual in our cohort of 666 SICCs resided. Follow up data extends for a median of 50 years these cases. Environmental exposure data such as source of drinking water, hazardous pollutants, or agricultural pesticide use, is overlaid on the time and space to assign exposures on an individual level and test for exposures that underlie increased risk of SICC. Finally, we will join the efforts of Aim 1 and 2 to pursue an exploratory aim to model main effects of environmental exposures with familial clustering to define potential gene-environment interactions for SICC risk. Understanding these genetic and environmental risk factors is important to allow for prevention, targeted screening of at-risk individuals, earlier diagnosis and better survival.

**Project Narrative**

Small intestinal carcinoid cancers show a strong familial risk which is influenced by unknown inherited and environmental factors. This project aims to study families with very high rates of small intestinal carcinoid cancer for genetic causes and to use state-wide environmental exposure data to identify environmental influences. Identification of individuals at risk will permit early diagnosis when the cancer is curable.

## Facilities and Resources

**Laboratory:** Laboratory space is available in Huntsman Cancer Institute (Neklason) and Genetic Epidemiology (Cannon-Albright). HCI research facility houses extensive facilities for laboratory-based research and population sciences. Dr. Neklason has been assigned individual laboratory space on the 3rd floor of the HCI, at the University of Utah. There is 300 sq. ft. of open bench space adequate for up to 3 persons to carry out routine lab work. In addition, there are shared rooms equipped for fluorescent and light microscopy, tissue culture, and storage. There are also extensive common areas such as cold rooms, dark rooms, general microscopy rooms, equipment rooms, and tissue culture rooms. Dr. Cannon-Albright has ~1000 sq. ft. of open bench space in the Division of Genetic Epidemiology fully equipped for molecular genetics work and located in Research Park, adjacent to University of Utah Medical Center.

**Office:** Dr. Neklason has a private office on the 3rd floor of HCI adjacent to the laboratories. Desk space is available in the lab for all full-time personnel. There are small conference rooms available in the office suites. Larger conference rooms and an auditorium are available in the building. The Institute provides an administrative assistant who is shared among the PI and four other investigators. The Institute has committed to support this project by providing the necessary space and administrative resources. Drs. Cannon-Albright, VanDerslice, and Curtin have fully staffed offices in their respective department in Research Park, adjacent to University of Utah Medical Center. All offices are furnished with standard office equipment, computer with secure network access, and assorted filing and cabinet storage. The groups have sufficient office space and equipment for all of personnel.

**Computer:** PC computers with Internet access and software (Microsoft Word, Excel, Powerpoint, Photoshop, Adobe Acrobat, Adobe Reader, EndNote, VectorNTI, and R) are available in the laboratory and office areas at HCI. All computers are supported by our Computer Support Group, are password protected, and adhere to University policies regarding encryption and data security. All are connected to shared high-output laser printers and graphics lab with scanners and color printers is housed within the Institute.

The computer local area network in the Genetic Epidemiology Group consists of twenty multi-processor workstations capable of extensive data analysis and 18 personal computers. We also have an HP DL380G5 Rack Mount Server and a HP DC7900 PC, externally located for data security, which contain data harvested from the UPDB database. These devices are currently used for the genetic analysis, database, graphics, and word processing needs of the Genetic Epidemiology group. All computers are inter-connected via a local Ethernet system.

The Geographic Information System (GIS) lab at the Division of Public Health has four HP workstations with i-7 or i-5 processors. These workstations are loaded with SAS (v 9.4), STATA (v 13), ArcGIS (ver 10.2), statTransfer, X-tools and other software. They are all connected to a dedicated 15TB Synology Network Attached Storage with a 5TB back-up NAS, a HP Designjet wide-format printer and a HP DesignJet plotter. A wide variety of spatial environmental data has been compiled for the US, and Utah in particular, including land use, crops, public water system water quality data, air quality data, demographic information and landscape characteristics.

**Informatics support for HCI and UPDB:** The Informatics department provides application development in support of the programs and other shared facilities of the Cancer Center. This group has developed applications such as: a clinic tracking system for the high risk cancer clinics and updated structures and tools for the Utah Population Database. The members of the informatics group have extensive experience in the field of bioinformatics, having made numerous contributions to the field over the past 15 years in both laboratory and clinical

systems, in academic and commercial environments. HCI has a permanent, dedicated computer support group that manages all computer equipment and internet resources.

**Other Resources:** The Huntsman Cancer Institute supports several shared resources, which are available for this project. The Shared Resources at HCI and the University of Utah are funded by a combination of the Cancer Center Support Grant (CCSG), HCI institutional funds, other University of Utah funding sources, and chargeback for services.

**Pedigree and Population Resource (PPR)** The PPR provides database services to users of the Utah Population Database and members of the High Risk Cancer Clinics including 1) computer access and software designed specifically to aid in analysis and management of data using sybase database management system, SAS, pedigree tools, record linking software, 2) consultation about research design and use of the database and software, and 3) collaboration on research analyses.

**Utah Population Database (UPDB)**: The UPDB is a shared resource located at the University of Utah and has been used to identify the high-risk families. The UPDB contains over seven million records and is made up of many data sets, including genealogies, cancer records from the UCR and ICR, Utah birth and death certificates, death data from the Social Security Death Index, and driver license data. There are two major data sets from which family members can be identified: 1.6 million genealogy records and 1.8 million Utah birth certificates. Both sets of records have been linked across generations. The genealogy records can encompass as many as seven generations and the birth certificates three or four generations. In pedigrees where the two sets are merged, as many as ten generations are presented. The extensive family histories available in UPDB allow the identification of familial clustering of cancer.

**Biospecimen and Molecular Pathology (BMP) Core:** The BMP, provides centralized collection, storage, and distribution of human tissue samples and blood samples to the basic science and clinical communities at the University of Utah Health Sciences Center and Huntsman Cancer Institute. The tissue resource, gives investigators the ability to evaluate tissues and blood samples from human subjects. The goal is to provide investigators with human specimens that are optimally preserved for analysis and for which relevant clinical information is available. Accordingly, the core has several -80 freezers and two liquid nitrogen storage freezers. Technical services offered by the facility include, collection and preparation of surplus surgical tissue, collection, preparation, of blood samples, tissue database management, tissue processing of Formalin Fixed Paraffin Embedded (FFPE) tissue, tissue embedding, tissue sectioning of FFPE tissue and frozen tissue, H & E staining of tissue sections, use of bright field, fluorescence, and laser capture microscopes.

**Histology:** Histology services including tissue embedding, processing, and sectioning are available through ARUP Laboratories, which is an enterprise of the University of Utah and its Department of Pathology.

**DNA Sequencing and Genotyping Core Facility:** The DNA Sequencing Core Facility is part of the Health Science Center at the University of Utah. The core provides DNA Sequencing services to the University of Utah research community and off-campus researchers. They employ the latest technologies to generate high quality data with a fast turnover and competitive prices. In support of DNA Sequencing activities they utilize state-ofthe- art DNA sequencers, genotyping platforms and lab robotics. Data from standard DNA sequencing and genotyping services are typically reported to customers within 24 hrs. Sample information can be submitted

on line and sequencing data files are also available on line for download using a simple and secure interface.

**DNA/Peptide Synthesis Core Facility:** This core offers investigators routine and specialty oligonucleotides and chemically synthesized peptides and Edman chemistry protein and peptide sequencing.

**University of Utah Center for High Performance Computing (CHPC):** CHPC is an organization of professional faculty and staff dedicated to providing access to high performance computing for research and education. The CHPC provides space, infrastructure, and systems support for large-scale computing and advanced networking systems. Computing equipment is housed in a newly renovated state-of-the-art 74,000 sq. ft. facility with 24-hour staff support, security, and high-efficiency cooling and energy supply systems. The CHPC cluster uses a Portable Batch System (PBS) job submission queue that facilitates large-scale analysis jobs. The University cluster also supports the OpenMPI message-passing interface, which allows for development and testing of scalable genome arithmetic software. All resources are HIPAA compliant. In total, the UCGD has available 2340 CPU cores and 2.5 PB of disc storage through a combination of dedicated UCGD computing infrastructure and shared resources. UCGD core faculty have complete control over these resources and will use them as needed to meet the needs of the research.

**Clinical:** The Huntsman Cancer Hospital (HCH) is a 442,000 square foot, 100 inpatient room hospital that houses state-of-the-art diagnostic, detection and treatment equipment, complete surgical facilities as well as facilities critical to cancer care including 102 outpatient exam/procedure rooms, eight operating rooms, three endoscopy suites and a large radiation oncology suite with radiographic fluoroscopic imaging (RF) capability and a cyclotron to generate isotopes for PET scanning, assisting in state-of-the-art cancer diagnostic imaging and disease detection. The University of Utah Hospitals and Clinics and Health Sciences Center (UUHSC) has been newly expanded to 589,259 gross square feet (GSF) of patient care area from its previous 442,260 (GSF). The hospital has a 425 patient bed capacity and is currently staffed for 385. There are a total of 22 on-site clinics, 17 diagnostic and therapeutic services and 5 hospital surgical services. The hospital houses the Gastroenterology endoscopy suite where 450 Endoscopic Retrograde Cholangiopancreatography procedures are performed annually.

## PHS 398 Cover Page Supplement

OMB Number: 0925-0001

### 1. Project Director / Principal Investigator (PD/PI)

Prefix:

First Name: DEBORAH

Middle Name: W

Last Name: NEKLASON

Suffix:

### 2. Human Subjects

Clinical Trial? ☒ No ☐ Yes

Agency-Defined Phase III Clinical Trial?* ☐ No ☐ Yes

### 5. Human Embryonic Stem Cells

* Does the proposed project involve human embryonic stem cells? ☒ No ☐ Yes

## PHS 398 Research Plan

OMB Number: 0925-0001

Please attach applicable sections of the research plan, below.

| | |
|---|---|
| 1. Introduction to Application (for RESUBMISSION or REVISION only) | |
| 2. Specific Aims | 2015_06_29_Sp_Aims_Final1013966022.pdf |
| 3. Research Strategy* | 2015_06_29_R21_Carcinoid_Res_Strategy_Final1013966021.pdf |
| 4. Progress Report Publication List | |
| **Human Subjects Sections** | |
| 5. Protection of Human Subjects | 2015_06_29_Human_Subjects_Final1013966020.pdf |
| 6. Inclusion of Women and Minorities | 2015_06_26_Inclusion_of_Women_and_Minorities_Final1013889163.pdf |
| 7. Inclusion of Children | 2015_06_26_Children_Final1013889160.pdf |
| **Other Research Plan Sections** | |
| 8. Vertebrate Animals | |
| 9. Select Agent Research | |
| 10. Multiple PD/PI Leadership Plan | |
| 11. Consortium/Contractual Agreements | |
| 12. Letters of Support | LOS1013889009.pdf |
| 13. Resource Sharing Plan(s) | 2015_06_26__Resource_Sharing_Plan_Final1013889158.pdf |
| **Appendix (if applicable)** | |
| 14. Appendix | |

## SPECIFIC AIMS

The incidence of small intestinal carcinoid cancers (SICC) has been steadily rising over the past 30 years, increasing four-fold from 1973-2003 (1). These cancers have a major hereditary component, much of which is not explained by the rare genetic syndromes. Nordic population studies report a strong concordance between first degree relatives with a standardized incidence ratio of 28 in siblings and 10 in parents (2). The increase in concordance between siblings versus parents suggests that some aspect of a shared environment compounds the risk of developing SICCs. We have the unique opportunity to use a one-of-a-kind resource to investigate environmental exposures AND inherited genetic risk factors which contribute to the development of SICCs. Understanding these risk factors is important to allow for prevention, targeted screening of at-risk individuals, earlier diagnosis and better survival. Unfortunately, the majority of SICCs are diagnosed with advanced disease where 5-year survival is less than 50% (1, 3). These cancers are difficult to diagnose because symptoms are ambiguous until late stages and diagnosis is missed or delayed by nearly 10 years (4).

The resource we will use is the Utah Population Database (UPDB), a computerized integration of 7 million individuals and genealogies dating back to the 1800s. This is overlaid with statewide vital statistics, SEER cancer records, medical records, and public records. UPDB is recently expanded to include geo-spatial and environmental data to promote epidemiological research and gain insights into gene-environment interactions. This exploratory project will set the stage for future epidemiological research using this powerful resource.

The aims of this proposal are as follows.

1. **To identify genetic variants that explain high-risk familial SICC.** Using UPDB, we show high relative risk in first degree relatives similar to the Nordic studies, extend this risk through third-degree relatives and have identified large pedigrees for genetic study. We have identified 666 SICC cases in UPDB and 384 have genealogy on 3 or more generations. Of these, there are 20 multi-generation pedigrees with 3 to 7 SICC cases and statistical excess of SICC. Cases and family members will be enrolled in research, focusing on families with highest excess risk. DNA from blood and archived tumor/tissue blocks will be obtained. Whole genome sequencing will be applied to affected individuals in these large extended pedigrees to identify genetic variants predicted to be responsible for familial SICC.

2. **To model environmental risk profiles of SICC using geo-coding, place and time of residence, and environmental exposure data**. Using historical residential addresses from public resources, we will identify when, how long, and where each individual in our cohort of 666 SICC resided. Environmental exposure data such as air quality, source of drinking water, hazardous pollutants, or agricultural pesticide use will be overlaid on the time and space to assign exposures on an individual level and test for exposures that underlie increased risk of SICC. Additionally, tobacco and alcohol histories will be extracted on the 94% of cases having medical records in UPDB. Data on controls (5 per case) matched for age, gender, birth cohort, and residential histories (rural vs. urban) will be obtained from UPDB.

Exploratory Aim: Using data from Aims 1 and 2, we will model main effects of environmental exposures with familial clustering to define potential gene-environment interactions for SICC risk. The environmental exposure profiles obtained through our work in Aim 2 can be applied in the context of the large families versus singleton cases. This body of work is important to

identify genetic and environmental risk factors for SICC and define individuals who would benefit from screening due to their genetic risk or environmental exposure.

Understanding the risk factors underlying SICC is important for prevention, diagnosis and treatment. This is clearly a familial but complex disease of multiple genetic and environmental influences. Using our unique database of 22 million genealogy, medical and environmental records and our experience in genetic discovery we are well positioned to define these risk factors.

## RESEARCH STRATEGY

A: SIGNIFICANCE

It is unclear why the incidence of small intestinal carcinoid cancers (SICC) has been increasing over the past 30 years. The key to reversing this trend is to better understand risk factors, identify individuals at risk, and implement early diagnosis tools. SICCs are often diagnosed at advanced stages when prognosis is dramatically diminished. Because the relative risk of SICC is so high in first-degree relatives, knowledge of genetic risk factors could identify those who would benefit from early and specific screening. Additionally, by understanding environmental exposures that compound the risk of carcinoid cancers these exposures could be reduced and/or individuals exposed to these elements offered screening for early diagnosis.

B: INNOVATION

Our experience in using the Utah Population Database (UPDB), a computerized genealogy linked to state wide cancer diagnosis data allows us to consider hypotheses and analyses not typically performed. Access to extremely large multi-generational pedigrees with multiple cases of SICC will increase our ability to identify rare variants underlying risk for this cancer. The population composing UPDB is not inbred and represents a heterogeneous population of Northern European descent typical of much of the USA (1-4). This resource is combined with the powerful new bioinformatics approaches under development by the Utah Center for Genomic Discovery specific to genomic analysis of large Utah pedigrees (letter of support from Dr. Yandell). Finally, environmental exposures can be addressed on an individual level through the recent development of Geo-Spatial Resource for Integrated Data (GRID) within UPDB and tested in the context of risk. GRID is truly innovative in a big data sense. With available historic statewide data from sources like birth records, census records, driver's license records, we are able to estimate the geographic residence of an individual at specific periods of time (spatio-temporal data). This is overlaid with historic records like industries or water sources at that time and place to assign an individual exposures that have occurred over their lifetime. Large control cohorts can easily be matched to cases. This exploratory project will set the stage for future epidemiological research using this powerful resource.

**Table 1: Examples of the 22 million records on 8 million unique individuals in UPDB**

| Sources (examples of data) | Beginning Year | Number Records |
|---|---|---|
| Birth Certificates (residence, parent demographic) | 1915 | 2,699,784 |
| Death Certificates (cause of death, tobacco) | 1904 | 829,759 |
| Cancer Records (age, location, stage, histology) | 1966 | 312,543 |
| Driver License (residence, BMI) | 1975 | 3,485,563 |
| U.S. Census of Utah (residence, occupation) | 1880 | 1,748,481 |
| Ambulatory Surgery and Inpatient Hospital (medical histories, tobacco) | 1996 | 8,526,517 |

C: APPROACH

### Background

Carcinoids are neuroendocrine tumors with the majority occurring in the GI tract (67%). Incidence of carcinoid tumors is about 2.47 (0.9 for small intestine alone) per 100,000 population per year with the rate steadily increasing 4-fold over the past 30 years (5, 6).

Carcinoid tumors of the small intestine derive from enterochromaffin cells of the neuroendocrine system in the gut and during development originate from the same stem cells as the rest of the gut epithelium (not neural crest). These cells contain a large amount of the body's store of serotonin and in response to stimuli in the lumen (chemical, mechanical, pathological), release of serotonin regulates gut motility and secretion as well as triggering nausea signals to the brain. Carcinoid tumors comprise approximately 40% of all small intestinal primary tumors and are similar in frequency to adenocarcinomas of the small intestine (7-9).

A number of rare genetic syndromes include carcinoid cancer within their tumor spectrum, including syndromes caused by mutations in *MEN1, RET, CKDN1B, PRKAR1A, NF1, TSC1, TSC2* and *VHL* (10). Additionally, a new gene has been associated with familial SICC, *IPMK*, in one of 33 families with multiple cases (11). *IPMK* and the syndromes explain a portion of familial carcinoids; however, the rarity of these syndromes coupled with the rarity of SICC within their tumor spectrum leaves much of the genetic predisposition undefined. Much of what is known about heritable contribution of small intestinal carcinoids has come from the Nordic populations. Studies have consistently shown a high relative risk of 9- to 11-fold in first-degree relatives and standardized incidence ratio in first-degree relatives of 13 (8, 12). The study from Finland reported a standardized incidence ratio of concordant carcinoid histology of SICC of 28 in siblings and 10 in parents (8). The increase in concordance between siblings versus parents suggests that some aspect of a shared environment compounds the risk of developing SICCs. Cigarette smoking and alcohol have been associated with an elevated risk of carcinoid tumors (odds ratio of 5.8 and 4.4 respectively), but occupation and socioeconomic status were not (9).

Our team's interest in SICCs arose when a SICC case presented to our oncology team at Huntsman Cancer Institute (HCI). Two additional cousins also had been diagnosed with SICC. Consequently, family members were enrolled in research, medical records reviewed, and blood samples obtained for genetic study. The three cousins grew up together on adjacent farms which begs the question if some environmental exposure contributed to the familial risk. This observation combined with the 1) rise in incidence of small intestinal carcinoids, 2) increased concordance between siblings versus parents, and 3) reduced penetrance (asymptomatic often undiagnosed), led us to the following hypothesis that we are well positioned to address. We hypothesize that environmental exposures and inherited genetic factors contribute to the development of SICCs and that the interaction of these two components drives penetrance. We are able to explore this hypothesis using a unique resource, the Utah Population Database (UPDB), developed and managed by University of Utah. A computerized Utah genealogy was created in the 1970s (13, 14) to represent the Utah pioneers and their descendants; many family records date back before the 1830s and contain data for 1.6 million individuals in 7 or more generation pedigrees. Today the Utah genealogy is part of the UPDB, a computerized integration of pedigrees, vital statistics, and medical and public records. Important to this work, the statewide Surveillance, Epidemiology and End Results (SEER) Utah Cancer Registry data from 1966 to present has been linked to the UPDB genealogy data, allowing description of the observed familial clustering of cancer in Utah. Many cancer-predisposing genes have been discovered using families identified through this resource notably, *BRCA1, BRCA2, NF1, p16* and *APC*.

We have used UPDB data resources to demonstrate familial clustering of SICCs in Utah (manuscript under review). Genealogical Index of Familiality (GIF) method was applied to test the hypothesis of excess relatedness of individuals diagnosed with carcinoid cancer in the small intestine (15, 16). Overall significant excess pairwise relatedness is found for the 384 cases ($p<0.001$) with genealogy in UPDB. In addition, when relationships closer than first cousin were ignored, significant excess relatedness was still observed ($p=0.041$). Relative risks (RR) for

SICC were estimated in first-, second-, and third-degree relatives of cases (based on population rates), and for spouses (Table 2). While all three types of first degree relatives had estimated RRs greater than 1, only the RR for siblings was significantly elevated (RR = 13.37), but sample sizes were small. No carcinoid cancers of the small intestine were observed among the 316 spouses of the 384 cases.

**Table 2. Estimated RRs for SICC in relatives of 384 cases with family history**

| Degree relative | n | Observed | Expected | p-value | RR | 95% CI |
|---|---|---|---|---|---|---|
| First | 3499 | 9 | 0.98 | 1.9e-6 | 6.43 | 3.21-11.51 |
| Sibs | 1582 | 8 | 0.60 | 2.4e-7 | 13.37 | 5.77-26.34 |
| Parents | 717 | 1 | 0.15 | 0.143 | 6.48 | 0.17-36.11 |
| Children | 1207 | 1 | 0.23 | 0.206 | 4.33 | 0.11-24.11 |
| Second | 11702 | 8 | 1.89 | 7.6e-4 | 4.24 | 1.83-8.35 |
| Third | 31002 | 12 | 5.28 | 8.2e-3 | 2.27 | 1.17-3.97 |
| Spouse | 316 | 0 | 0.12 | - | - | - |

The UPDB can also be used to look for geo-spatial clustering of cancers to explore the risks of environmental exposures as described with bladder cancer incidence in Utah associated with EPA designated Toxic Release Inventory sites (17). While the UPDB taps into an impressive array of data sets about individuals, there is a wealth of information about their neighborhoods, environmental exposures and access to services that can be exploited by locating individuals *to a place in time.* We are well positioned to use UPDB data resources to investigate environmental clustering of small intestinal carcinoid cancers in Utah. The UPDB has recently developed a Geo-Spatial Resource for Integrated Data (GRID) whereby public information about where and when individuals reside in a certain geocode is captured and assigned to the individual. This then, can be overlaid with geo-spatial environmental exposure data such as air quality or source of water to assign an exposure value on an individual basis.

**Aim 1: To identify genetic variants that explain high-risk familial SICC.**

Familial SICCs are likely to arise from multiple genetic and environmental influences and the consequential variable penetrance creates challenges to identify the genetic causes. Our approach is to focus on large highrisk pedigrees with distantly affected individuals so that the genetic component will be enriched in affected cases and the background noise (shared environment and excess of shared genetics) is reduced. In a recent SICC study, a responsible gene (*IPMK*) was identified in only one of 33 families; this one family was the only of the 33 with affected 3rd degree relatives sequenced (versus 1st or 2nd) (11). This isolated success demonstrates the power of large extended families to identify SICC risk genes. Large extended pedigrees are the strength of Utah. The most current version of the UPDB extends to 15 generations and includes over 2.3 million individuals with at least 3 generations of data. The genealogy data is linked to the Utah Cancer Registry (UCR) data including primary site, histology, stage, grade and survival. All cancer cases recorded in UCR have pathology confirmation. Dr. Cannon-Albright's group has been using this resource for over 30 years. Dr. Neklason has used this resource for the past 13 years to study genetics of gastrointestinal cancers. In addition to the team's many years of experience in the identification

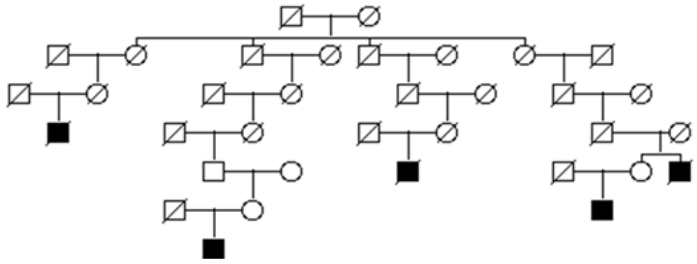

**Figure 1. Example of high-risk carcinoid cancer of the small intestine pedigree with 5 cases**

of breast, colon, melanoma, and prostate cancer predisposition genes, they have used the UPDB to identify high-risk cancer pedigrees and have collected and stored over 35,000 blood samples in thousands of high-risk pedigrees for multiple different cancers some which overlap with the SICC pedigrees.

**Identification of high-risk pedigrees**: Dr. Albright's group has created software to identify all cancer cases sharing a common ancestor in UPDB. To identify high-risk pedigrees we identify founders in the UPDB with a significant excess of SICC ($p<0.05$) among their descendants. We count all descendants for each founder (a pedigree) by birth year cohort, sex, birth state (Utah or not), and birthplace (urban or rural), and, using SICC rates from the UPDB, we calculate the expected number of cases among the descendants. Pedigrees with a significant excess of SICC cases ($p<0.05$) are considered high-risk. We identified 73 high-risk pedigrees ($p<0.05$) with 2 to 7 cases: 1 pedigree with 7 cases, 2 pedigrees with 6 cases, 2 pedigrees with 5 cases, 3 pedigrees with 4 cases, 12 pedigrees with 3 cases, and 53 pedigrees with 2 cases identified. Figure 1 shows an example pedigree with 5 SIC. The founder has almost 10,000 descendants; only 0.78 carcinoid cancers of the small intestine are expected in these descendants ($p=1.6e^{-4}$). Among the descendants there is an overall excess of cancer with 315 observed and 283.2 expected ($p=0.03$).

**Enrollment of high-risk familial SICC pedigree members:** We will recruit 30 SICC cases per year into research with priority based on 20 pedigrees with highest risk and multiple cases in the pedigree. Recruitment is done through UCR whereby permission is obtained from the subject or their next of kin, to release their name and contact information to the study. Only names of individuals who indicate an interest are provided to the research team. In the past 15 years, we have enrolled cancer patients referred from UCR with a >70% participation rate. The study team explains the study and obtains consent for participation including collection of medical records, archived tumor blocks, blood sample, and permission to contact relatives. We use the family advocate system for expansion of pedigrees and have been successful in expanding pedigrees from 77% of the contacts obtained from previous genetic studies with an average of 8 referrals per contact. Family advocacy has proven to be a highly effective method of recruitment of large pedigrees whereby distant branches can be recruited independently. When possible, unaffected family members will be referred for clinical evaluation to our colleagues here at Huntsman Cancer Institute (see letter of support from Dr. Gilcrease). Through ongoing research efforts, we have DNA from blood of 22 individuals and from FFPE blocks of an additional 16 individuals. At least 15 are in familial clusters; 2 pairs of distant cousins, 2 pairs siblings, the remainder being single cases sampled in a cluster. Exome sequencing of the cousins and siblings showed no clear mutations in known syndromes or *IPMK*, however a single individual harbors a potential splice mutation in *IPMK*. Analysis of novel variants is ongoing.

**DNA preparation and whole genome sequencing:** DNA is extracted from blood using the Qiagen AUTOPURE LS automated DNA extractor. We also obtain non-cancer DNA from formalin fixed paraffin embedded (FFPE) tissues archived from resection of the cancer. Tumor and normal cells are identified and marked on a slide by a pathologist. Cells are micro-dissected and DNA extracted. We have obtained sufficient amounts (>200 ng) of high quality double stranded non-tumor DNA from 14 of 16 SICC cases from FFPE blocks. Similarly, DNA from carcinoid tumor in the FFPE blocks will be extracted to examine for somatic alterations to compliment genetic discoveries.

We plan to perform whole genome sequencing on 10 high-risk familial carcinoid cases per year, selecting 2 or more cases from a pedigree and prioritizing pedigrees with the highest familial standardized incidence ratio (FSIR) (18). We are applying an affected-only design because non-

penetrance is suspected to be common; 21% of asymptomatic relatives of cases were found to have SICC when targeted PET imaging was used (11). Cases will be evaluated for mutations in *IPMK* by sequencing, for family histories suggestive of cancer syndromes with carcinoids in their tumor spectrum and excluded if there is clear evidence for a cause. Additional cases collected which do not undergo whole genome sequencing will be used as a confirmation cohort for findings. Genome sequencing will be contracted through a sequencing facility with Illumina X-10 sequencing. This platform consistently provides 30x mean coverage with >90% covered by 20x.

**Analysis of genome sequence, variant prioritization, and segregation studies:** Current best-practices will be used for genome alignment, alignment polishing, variant calling, variant prioritization and copy number changes. These are maintained on the Utah Genome Project wiki page and are carefully vetted and tested by the bioinformatics community (http://weatherby.genetics.utah.edu/UGP/wiki/index.php/Main_Page). Dr. Neklason, the PI of this proposal, is Program Director of Utah Genome Project and works closely with the bioinformatics teams of the University of Utah Center for Genetic Discovery led by internationally recognized bioinformaticians (please see letter of support from Mark Yandell). This multi-million dollar investment works to maintain a robust informatics analysis pipeline and computational power to support genomics-based research discovery. Genomes are annotated and scored using multiple prioritization algorithms including pVAAST which incorporates pedigree relationship information with a probabilistic search tool for identifying damaged genes based on phylogeny and amino acid substitution (19-21). To our knowledge, there are not robust programs to interpret variants in regulatory regions. We envision that these will be available in the next two years and will apply to these genomes. Modifications in promoter regions, however, can often be interpreted and tested.

Genetic variants and insertions/deletions that are shared by affected family members and suspected to drive SICC will be confirmed by Sanger sequencing and segregation in the family. The implicated genes will be screened in all available SICC cases as a confirmation cohort (approximately 90 cases).

**Aim 2: To model environmental risk profiles of SICC using geo-coding, place and time of residence, and environmental exposure data**

The concept behind this aim is that we will create a cohort of individuals with SICC and controls matched on gender, birth year, and residential histories. We will then define where and when the individuals resided through geo-coding and construct a time-varying record of spatial location. Next, we will assign specific quantities of the different environmental exposures associated with when and where the individuals resided. Personal environmental exposures (such as tobacco and alcohol use) will also be assigned to the individual using UPDB data linked with hospital records. As shown in Table 3, we have long exposure periods where we are reasonably confident the SICC reside in Utah. Follow up in Utah (births, voter's registration, driver's license, etc) extends for a median of 50 years in the 603 Utah residents diagnosed with SICC.

**Environmental Exposures:** Metrics of environmental exposure will be generated for each case and matched controls by spatially linking the residential location(s) of the individual for the relevant period in their life to actual or modeled environmental data. All residential addresses available through UPDB will be geo-coded using algorithms developed for GRID. This includes birth and death records of the individual and their parents, driver license, census records, voter registration and other public data sources. We have piloted methods to locate these individuals at birth by using historic maps and street information and will fully develop this approach as part

of this project. Figure 2 demonstrates a visual overlay of geocoded data; the residential location of carcinoid cases at time of diagnosis and agricultural land use.

**Table 3: Characteristics of SICC population data captured in UPDB**

| Characteristic | N (%) |
|---|---|
| Total | 666 |
| Male : Female | 376 : 290 (56%:43%) |
| 3-generations in UPDB? Yes: No | 385 : 281 (58%:42%) |
| Birth place Utah? Yes : No | 355 : 311 (53%:47%) |
| Tobacco use indicated? Yes : No | 145: 521 (22%:88%) |
| Medical Record in UPDB? Yes: No | 624: 42 (94%:6%) |
| Utah resident at time diagnosis | 603 (91%) |
| Median residence time in Utah* | 50 years |
| 25th percentile residence time in Utah* | 30 years |

* *evidence of residence in Utah ≥ 1 year*

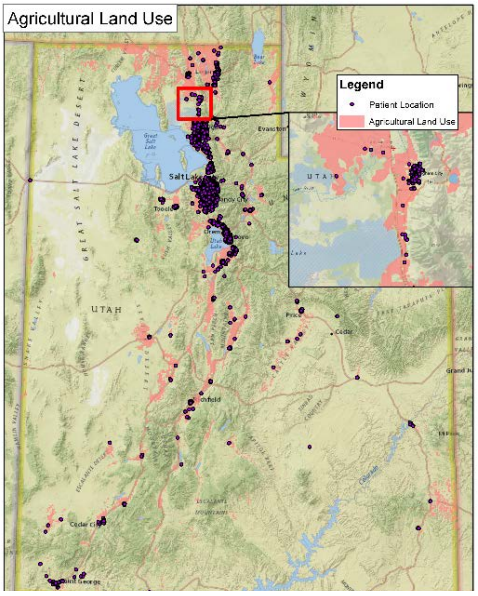


**Figure 2: Geographic location of SICC diagnosis (purple) and agrictultural land use (peach).**

There is a wide array of spatially-referenced environmental information that has been compiled during the development of GRID. Exposure to contaminants in drinking water is generated by spatial linkage to a map of water system service boundaries and database of administrative water quality (22). Measures of potential exposure to agricultural pesticide will be based on proximity to agricultural areas and specific cropping patterns (23-25). Exposure to air pollutants will be based on monitored values (available back to 1970) and predicted values based on land use regressions for areas lacking a monitor (26). Dispersion models are also used to generate estimated concentrations of hazardous air pollutants from the Toxic Release Inventory (TRI) sites.

By way of example, we used these data sources to show that exposure to traffic emissions was significantly associated with the risk of testicular cancer in Utah (highest vs. lowest quintile: OR=1.39, 95% CI: 1.04-1.86) (funded by R03 CA159357). Exposure to traffic emissions was be based on a Gaussian dispersion model from multiple points sources (every 10 m) along each major roadway, summed across all points within 500 m, and based on annual average traffic density for each road segment and regional wind patterns.

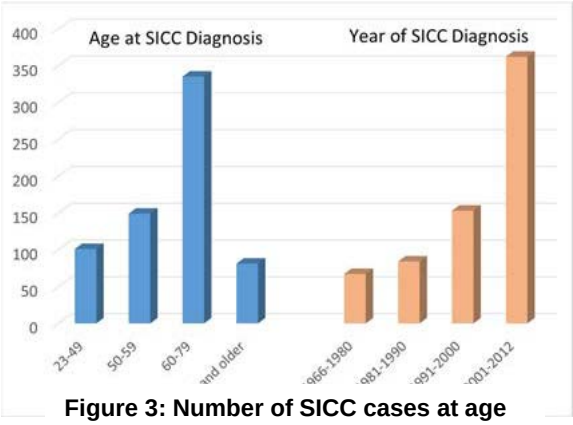


**Figure 3: Number of SICC cases at age of diagnosis and year.**

We would expect exposures earlier in life (first 30 years) to be a driving force in SICC development. Currently, the largest numbers of SICC recorded in UCR are in birth cohorts from 1920 to 1950. With an average age of diagnosis of 65, the corresponding largest years of diagnoses are from 1985 to 2015 (Figure 3). Many important early life exposures for this group would be prior to 1980. Creating measures of early life exposures will be a challenge as digitized data before 1980 is limited. Based on pilot work conducted as part of GRID, we will compile and georeferenced historical manufacturing, transportation, agricultural lands, population density and measures of community wealth. Measures of potential exposure derived from these data will be used to compare cases and controls.

**Exposure to confounding environmental risk factor (tobacco and alcohol):** The Utah population will have less exposures to confounders like tobacco and alcohol because smoking and drinking are proscriptive within the major religion in the state, Church of Latter Day Saints (aka LDS; Mormon) (9, 27). In the analysis, we are able to identify individuals who are active in the church through baptism and membership records and match controls similarly.

From medical records, we can capture the timing, amount and duration of tobacco and alcohol use. Of the 666 SICC cases, 94% have medical records in UPDB. Since 2009, tobacco-use history data (intensity, duration, date quit if applicable) and alcohol consumption is collected by University of Utah Healthcare from questionnaires given at all inpatient as well as most outpatient encounters. Data from these tobacco- and alcohol-use histories are available to supplement ICD-9 billing code indications of cigarette smoking and alcohol use in the electronic record, especially since a lack of ICD-9 diagnosed tobacco use does not mean an individual was unexposed to tobacco. We also use smoking history as an example of data readily available through UPDB record linking. We find that 22% of the 666 cases have evidence of smoking history, in contrast to 2011 statistics of 11% for Utah.

**Table 4: Tobacco exposure in SICC cases in Utah**

| | | N | % |
|---|---|---|---|
| Total cases | | 666 | 100.0% |
| Tobacco use indicated | No | 521 | 78.2% |
| Tobacco use indicated | Yes | 145 | 21.8% |
| **Tobacco indicator:** | **Description** | | |
| ICD-9 code 305.1 | Tobacco use disorder | 20 | 13.8% |
| ICD-9 code V15.82 | Personal history of tobacco use | 35 | 24.1% |
| ICD-9 305.1 and V15.83 | Tobacco use and personal history | 9 | 6.2% |
| Death certificate (DC), | Code='probably contributed' | 10 | 6.9% |
| tobacco-related code | Code='didn't contribute' | 45 | 31.0% |
| DC and ICD-9 code(s), | Code='probably contributed' | 13 | 9.0% |
| 3015.1 and/or V15.82 | Code='didn't contribute' | 13 | 9.0% |

*ICD-9 diagnoses in statewide inpatient/ambulatory records from 1996-2013. Death Certificates available from 1904-2013.*

**Exploratory Aim: Model the main effects and interaction of environmental exposures with familial clustering to define potential gene-environment interactions for SICC risk.**

Separating shared environment from shared genetics can be a major challenge when studying familial risk, especially with rare conditions such as SICC. The large multi-generation pedigrees that we propose to study offer a unique opportunity to explore genetic and environmental interactions. The environmental exposure profiles obtained through our work in Aim 1 can be applied in the context of the large families (genetic risk) versus singleton cases (absence of genetic risk) to determine if they have an additive effect. Also, because most of the cases in the large families are distantly related, they may not share the same environment. For example, the SICC cases shown in the large UPDB pedigree (Figure 1) are sufficiently far apart that they likely represent different birth cohorts and possibly different geographic communities. The aim is highly exploratory in nature because the number of SICC cases available is small, but it will

potentially lead to novel models of SICC risk that can be explored in future work. Our initial work has identified 8 sibships, 8 second-degree relatives, 12 third-degree relatives and over 150 more distantly related individuals in which environmental exposures can be estimated and related to genetic distance to model gene-environment interactions.

Risks of developing gastrointestinal carcinoids from environmental exposures will be estimated using conditional logistic regression models. The interaction between familial risk and environmental exposures will be assessed by estimating these models stratified by membership in high-risk pedigrees with excess clustering of familial disease, environmental exposures, or both. Multivariable regression-based models are a natural way to test for statistical interaction defined as departure from linear regression using categorical, continuous or categorical data, which may be time-varying. We will incorporate logistic models and Cox regression (for longitudinal exposures) to detect main effects and gene-environment interactions into our Monte Carlo-based multilocus gene-gene association framework, to allow utmost flexibility to identify susceptibility genes and exposures in association analyses in which individuals are related (28).

**Future studies beyond the scope of this grant**

Development of GRID methods and capturing the environmental data in a geo-temporal fashion can be applied to many other complex conditions. Specific to SICC we have the opportunity to develop better ways to diagnose this condition earlier. One is through PET/CT phenotyping in conjunction with HCI Radiology Department (letter from Dr. Yap). The other is to mine the 94% of SICC with medical records for comorbidities and symptoms preceding diagnosis which can provide guidelines to recognize SICC earlier.

## PROTECTION OF HUMAN SUBJECTS

IRB and the Resource for Genetics and Epidemiology (RGE) application are approved for this research study, including permission to use Utah Population Database (UPDB) and data specific to this application. The primary IRB is IRB # 2657, Genetic analysis of colorectal and other GI cancers (Cannon-Albright, PI) and additional subjects have been collected for genetic study under University of Utah IRBs # 30546, Hereditary Gastrointestinal Cancer Registry (Neklason, PI).

### ***1. RISKS TO THE SUBJECTS***

#### a. Human Subjects Involvement and Characteristics

Both of the IRB approved studies will be contributing data and subjects to the study.

IRB 2657 enrollment is based on UPDB data and histories and enrolls GI cancer cases and their family members. The aims of the study are to identify high-risk kindreds, sample individuals in those kindreds, and gather family history, DNA and questionnaire data. Participants are the subject of genetic study and project aims to analyze gene-environment interactions in cancer susceptible individuals and control individuals. The work of Aim 1 and 2 will be conducted under the scope of this IRB.

IRB 30546 is based on clinical and family referral. It includes over 1800 subjects from high risk GI cancer families. Eligibility requirements include a family history of GI cancers, clinical or genetic diagnosis of a genetic syndrome, or family members. Index cancer cases are referred to the Registry. The family advocate system is then used to obtain permission to contact additional kindred members. Permission is obtained to collect family and medical histories, medical records, tumor blocks, and a blood sample for DNA and cell lines.

#### *b. Sources of Research Material*

The types of research material obtained for these studies include biospecimens (blood for DNA and cell lines, and archival tumor blocks), medical records of colon related procedures, questionnaire data, and public records accessed by UPDB. The blood, DNA, cell lines and questionnaires are obtained for research purposes. The medical records and archival tumor blocks are existing records and specimens. All biospecimens are labeled with a sequential number that can be linked back to the research participant by study personnel only. The specimens, records, and data will be collected specifically for the aims of this research project but will be available to other related studies if the participant agrees to this in the signed consent form. In addition, participants have consented to review of their medical records pertinent to their cancer screening history or diagnosis of cancer.

#### *c. Potential Risks*

The risks of blood drawing are superficial bruising, bleeding from the site of puncture, and uneasiness associated with needles. A minor risk is associated with maintaining confidentiality of family history, medical records and genetic material.

## ***2. ADEQUACY OF PROTECTION AGAINST RISKS***

### ***a. Recruitment and Informed Consent***

The confidentiality of all subjects in UPDB is rigorously protected and overseen by the Resource for Genetic and Epidemiologic Research (RGE) at University of Utah which governs access to the UPDB data as mandated by the state government. Access to UPDB data is gained through a waiver of consent.

Subjects for genetic study are recruited after their kindred or cancer status is identified from UPDB and Utah Cancer Registry records. Utah Cancer Registry will be the third party which makes the initial contact of potential research participants (or their next of kin if deceased) and obtains permission from the subject or their next of kin, to release their name and contact information to the study. Only names of individuals who indicate an interest will be provided to the researcher. If the subject (or their next of kin) agrees to be contacted by researchers, the study is explained and subjects are given opportunity to participate in the study. The family advocate system is used to identify persons in families as described above. Verbal permission is obtained from the family advocates prior to contacting each subject. Family members are then recruited by phone contact and mail, first informing them that a study is ongoing in their family because of the possibility of increased risk for cancer. The principle investigator and study coordinator seek signed consent for the study, however the voluntary nature of the study is highly emphasized. The printed informed consent form describes the study in detail and is IRB approved.

### ***b. Protection Against Risk***

i) **Physical:** blood drawing: only trained phlebotomists are used.

ii) **Confidentiality:** All personnel working in the study have undergone nationally developed human research training for the protection of human subjects required by the IRB. All information obtained from this study will be kept confidential. All specimens are assigned a coded identifying number, and identifying information and records are kept in secured files in Huntsman Cancer Institute or Division of Genetic Epidemiology at the University of Utah. Any type of medical information or biological sample will be shared only among authorized persons for specific and necessary medical and scientific purposes. The family advocate system also allows genealogy, contact information and health status to be shared with the study. However, the study does not share information with the participants about other family members. Information obtained by investigators is only communicated to the individual to whom the information pertains. The confidentiality measures have been completely effective in our kindred and genetic studies to date. No one will be identified when we publish information from this study. No information will be given to insurance providers or employers.

iii) **Data security**. All project staff members sign confidentiality agreements. The user must have a valid account and password to logon to network and a separate login and password to access research participant data. A firewall prevents unauthorized access to the system from external sources and all computers are encrypted. All paper copies of data are stored in locked file cabinets and/or locked offices. Data are not downloaded to “laptop” computers that are taken out of the institution.

***3. POTENTIAL BENEFITS OF THE PROPOSED RESEARCH TO THE SUBJECTS AND OTHERS***

No direct benefit for cases is anticipated. There is potential for benefit to subjects if a genetic or environmental factor putting them at risk of cancer is identified. The benefit to society is also direct as the data obtained leads to appropriate prevention strategies in high-risk individuals and high-risk exposures. Additionally, as molecular mechanisms are better defined, new approaches for prevention and control will emerge.

***4. IMPORTANCE OF THE KNOWLEDGE TO BE GAINED***

The genetic, clinical and molecular characterization of inherited cancer risk is the focus of our present and long-term work. This work is important to understand risk factors that may be avoided leading to prevention efforts in families and individuals at higher risk for carcinoid cancer. In the laboratory the identification of additional susceptibility genes and the characterization of molecular expression profiles will suggest new targets for intervention, both preventive and therapeutic.

**Inclusion of Women**
An equivalent number of men and women will be examined in this investigation because small intestinal carcinoid cancers affect both genders equally. This equivalence will hold true for all aspects of the study.

**Inclusion of Minorities**
Individuals will be enrolled in research regardless of race and ethnicity. Utah Cancer Registry and the Cancer Data Registry of Idaho reflect the population of Utah and surrounding states. The population composition for Utah based on 2013 census is: 91.6% White (13.4% Hispanic; 79.7 % non-Hispanic), 1.3% Black, 1.5% Native American, 2.3% Asian, and 1.0% Pacific Islander and 2.3% two or more races. University of Utah Health Sciences Center offers translation services to the medical and research community so that non-English speaking minorities will have the same opportunities to engage in research. Huntsman Cancer Institute supports Hispanic Patient Navigators who assist families with their clinical experience in a culturally sensitive way and are available for research studies. Huntsman Cancer Institute also supports a Native American Outreach program that works with tribes throughout the Western United states.

Although the Utah population is predominantly white, the genetic and environmental discoveries made using large pedigrees in Utah Population Database have applied to all races and ethnicities across the world.

OMB Number: 0925-0002

## Planned Enrollment Report

**Study Title:** Genetic and Environmental Etiology of Familial Small Intestinal Carcinoid Cancer

**Domestic/Foreign:** Domestic

**Comments:**

| Racial Categories | Ethnic Categories | | | | Total |
|---|---|---|---|---|---|
| | Not Hispanic or Latino | | Hispanic or Latino | | |
| | Female | Male | Female | Male | |
| American Indian/Alaska Native | 5 | 5 | 0 | 0 | **10** |
| Asian | 8 | 8 | 0 | 0 | **16** |
| Native Hawaiian or Other Pacific Islander | 3 | 3 | 0 | 0 | **6** |
| Black or African American | 4 | 4 | 0 | 0 | **8** |
| White | 264 | 264 | 43 | 43 | **614** |
| More than One Race | 6 | 6 | 0 | 0 | **12** |
| **Total** | **290** | **290** | **43** | **43** | **666** |

Study 1 of 1

**Inclusion of Children**

Children will not be included in this study. Small intestinal carcinoid cancers are typically diagnosed after age 50 years, so inclusion of children is not indicated.

**Resource Sharing Plan**

Sharing of data and materials will be in accordance with relevant consents, privacy authorizations, IRB documentation, policy, regulations, and applicable privacy laws. We will adhere to the NIH policy on Sharing of Unique Research Resources and Guidelines for Recipients of NIH Grants.

Data Sharing Plan

The investigators of this proposal concur that data sharing is the foundation for many important goals of the scientific community, including encouraging open scientific inquiry, promoting new research, enabling the exploration of topics beyond those originally envisioned by the investigators, and permitting the creation of new datasets by combining data from multiple sources. The investigators agree to make their data and progress available to the general scientific community in a timely manner so that other researchers can add their expertise and create additional collaborations to accelerate the progress towards clinically useful treatments for cancer. In the case of resources from human subjects that are of a limited nature (e.g. primary tissues), sharing with outside investigators will be carefully prioritized in order to maximize these resources.

The investigators are committed to promote data sharing by publishing the data collected through this research and to make available molecular resources described in these publications. We will lodge any genome-specific and molecular data with NCBI (eg. dbGaP). We agree that data-sharing will include all data from the funded research that can be shared without compromising the rights and privacy of the subjects in our study, regardless of whether the data have been used in a publication. Data sharing will occur in a timely fashion, no later than publication of the main findings from the final dataset. Patentable data will be protected.

*Protection of Research Participants*

We are in accord with NIH data-sharing policy, in that data from human subjects derived from these studies will be shared only if the identity and privacy of the research participants can be protected and if research participant has not denied data sharing. De-identification of data will be meticulously completed with awareness that our participants' privacy must be protected in agreement with all applicable laws and regulations. Data will be stripped of identifiers that would permit linkages to individual research participants and will exclude variables that could lead to deductive disclosure of the identity of individual subjects.

The Utah Resource for Genetic and Epidemiologic Research (RGE) governs access to certain data and research resources provided to the University of Utah for use in biomedical research, including the Utah Population Database, which includes family history records, vital records, cancer registry records, driver license records, and others. These records are linked together to form multi-generational pedigrees as well as longitudinal person-level data. Access to these research resources requires review and approval by the RGE Review Committee and by an Institutional Review Board (IRB). All access is project-specific and must be renewed annually. Personnel on approved projects must sign Confidentiality Agreements prior to use of any data.

We will utilize the expertise of Ken Smith, Director of the Utah Population Database and Jahn Barlow, Director of the Resource for Genetic and Epidemiologic Research to strip identifiers from datasets for publications, presentation, and other data sharing. We are fortunate to have local professionals, of note, Dr. Jeffery Botkin, with the substantial skill set and extensive experience in data management to prevent deductive identification that is required to de-identify pedigrees and datasets. Because of the possibility of identification of pedigrees from

the unique constellation of affected individuals and their relationships to each other, we will not provide such data. We will indicate and share data for the set of case samples that are independent.

*Collaboration Approval Protocol*

Certification for access to genetic data and biomaterials will be made based on the scientific merit of the investigator(s) and the projects proposed. Review of applications will be completed after IRB and RGE approvals (from the investigator's institution and the University of Utah) are reviewed. Requests must be submitted in writing on the letterhead of the sponsoring institution at which the research will be conducted and will include 1) identifying information about the Principal Investigator and co-investigators, 2) curricula vitae, and 3) a detailed description of the research project to be conducted. The PI and an authorized representative of the recipient institution must sign the Agreement. The Agreement specifies that the investigator will only use the data and biomaterials for the specific project as described. The Agreement is not transferable to another recipient or facility. The Agreement further stipulates that neither printed nor electronic data may be copied or otherwise shared without permission. Use of the data is restricted to statistical reporting, analysis, and teaching. The agreement prohibits the user from making any efforts to identify individual cases and prohibits linking data from this archive with individually identifiable data from other datasets.

Recipients must share data that they generate within 12 months of receipt of data and/or biomaterials or upon publication of research findings, whichever comes first. Upon completion of the project, the recipient must return all biomaterials as well as clinical and genetic data received from the University of Utah or certify that the materials and data were destroyed in accordance with applicable laws and safety procedures.

Data requestors will be charged for actual costs associated with sharing of the data, including preparation and shipping the data, costs of personnel, computing time, supplies, and other directly related expenses.