HERITABILITY AND GENETIC ARCHITECTURE

OF PANCREATIC CANCER

by
Fei Chen

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, MD
August 2018

@2018 Fei Chen
All Rights Reserved

ABSTRACT

Pancreatic cancer is currently the third leading cause of cancer deaths in the US. Identified common susceptibility loci for pancreatic cancer only account for a small fraction of phenotypic variations, highlighting the "missing heritability" problem. Array heritabilities were underestimated because of the inadequate coverage on rare variants. The goal of this study was to generate a more comprehensive and less biased estimate of heritability for pancreatic cancer and to infer the genetic architecture of pancreatic cancer through genomic partitioning of the estimated heritability.

In the Pancreatic Cancer Case-Control Consortium (PanC4) study of 3,568 cases and 3,363 controls, the heritability estimated from the imputed variants was 21.2% (s.e. = 4.8%). Genomic partitioning found more than one-third of the estimated heritability attributable to rare variants with minor allele frequency (MAF) < 0.01. Across the functional group of intronic, intergenic, coding and regulatory variants, intronic variants account for most estimated heritability. Identified GWAS loci explained 4.1% of the total phenotypic variations for pancreatic cancer. These findings suggest that the yet-to-be-identified rare and common variants are both key contributors to the "missing heritability" of pancreatic cancer, and it's important to include both coding and regulatory regions in the search of disease-causing variants for pancreatic cancer.

In the combined sample of 569 familial pancreatic cancer (FPC) cases and 754 controls from Alzheimer's Disease Neuroimaging Initiative (ADNI) study, whole-genome sequencing (WGS) variants explained 38.8% (s.e. = 15.7%) of total phenotypic variance for pancreatic cancer. This higher estimate of heritability for pancreatic cancer is likely a result of a more

completed characterization of genetic variation across the genome in the sequencing data as well as a stronger genetic background in the FPC cases.

In addition, variants in 11 established FPC genes only explained 0.4% of total phenotypic variance for pancreatic cancer in PanC4 study, which reflected the limitation of current statistical approaches in capturing the contribution of rare variants to the heritability of a trait.

Findings from this study provided valuable insights into the genetic architecture of pancreatic cancer which would help understand the disease etiology and to aid in disease screening, diagnosis, prognosis and therapy.

Advisors: Dr. Priya Duggal, PhD

Dr. Alison P. Klein, PhD

Thesis Readers: Dr. Priya Duggal, PhD

Dr. Alison P. Klein, PhD

Dr. Joan Bailey-Wilson, PhD

Dr. Peter Zandi, PhD

ACKNOWLEGEMENTS

First, I would like to express the deepest appreciation to my advisors, Dr. Alison Klein and Dr. Priya Duggal, for their continuous support and guidance throughout the whole process; for their understanding and respecting my decision of moving back to California with my families; and for their efforts and time in maintaining regular contacts and providing essential help to me during my time in California. I cannot thank them enough for the endless amount of support and direction they provided to bring me to where I am now. Without their guidance and persistent help this dissertation would not have been possible.

I would like to thank Dr. Bailey-Wilson, not only for her time and patience as a member of my dissertation committee, the departmental oral exam committee and the final oral exam committee, but also for her valuable intellectual contributions to the development of my research project.

In addition, I would like to thank all the faculties in Genetic Epidemiology, who have always been very kind and supportive to me.

Lastly, I want to thank my family and friends, my parents, my husband and my daughter in particular. I would have no emotional support system without them, and I certainly would not have survived graduate school without their ardent love and support.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEGEMENTS iv

CHAPTER ONE

INTRODUCTION

Epidemiology of Pancreatic Cancer 1
Risk Factors for Pancreatic Cancer 2
Family History 2
Cigarette Smoking 3
Chronic Pancreatitis 3
Diabetes Mellitus 4
Obesity 5
Genetic Factors 6
Family-based Heritability and Missing Heritability Problem 7
Family-based Heritability 7
Missing Heritability Problem 9
Population-based Heritability and GREML Approach 10
The Single-Component GREML Approach 10
Array-based Heritability 12
Motivations of This Study 13
Imputation and Whole-genome Sequencing Data 13
The Multi-Component GREML Approach 14
Goals of this Study 16
Supplementary Notes 18
Reference 19

CHAPTER TWO

INHERITED PANCREATIC CANCER

Family History and Risk of Pancreatic Cancer 29
Pancreatic Cancer Genes 31
BRCA1/2 31
PALB2 32
ATM 32
STK11 32
CDKN2A 33
Mismatch repair genes 33
PRSS1 33
Germline Mutations in Sporadic Cases 34
Targeted Therapy 34
Common Low-Risk Susceptibility Loci 35
PanScan I 35
PanScan II 36
PanScan III 38
PanC4 39
Imputation Analysis of PanScan I-III 41
Meta-analysis of PanScan I-III and PanC4 42
GWAS in Asian Populations 44
Current Status of Screening High-Risk Populations 45
Summary and Future Directions 47
Figures and Tables 48
Reference 52

CHAPTER THREE

HERITABILITY FOR PANCREATIC CANCER FROM IMPUTED VARIANTS

Study Populations 71
Genotyping and Imputation 75
Quality Controls on PanC4 After Imputation 75

Statistical Analysis 78
Estimates of Heritability Explained by Imputed Variants 78
Partitioning by Chromosome, MAF and Functional Annotation 81
Contributions of Identified GWAS Loci 82
Results 83
Estimates of Heritability Explained by Imputed Variants 87
Genomic Partitioning of Genetic Variation 87
The contribution of Identified GWAS Loci 94
Discussion 94
Conclusion 100
Tables 102
Supplementary Figures and Tables 105
Reference 111

CHAPTER FOUR

HERITABILITY FOR PANCREATIC CANCER FROM WGS VARIANTS

Study Populations and Genetic Data 114
Familial Pancreatic Cancer (FPC) Patients 114
Alzheimer's Disease Neuroimaging Initiative (ADNI) Subjects 115
Quality Controls on FPC-ADNI Study 116
Genotype Filtering: GQ < 20 and DP < 10 119
Variant Filtering: Segmental Duplications 120
Samples Filtering: Non-reference Discordance Rate (NRD) 120
Sample Filtering: Mismatched Sex 123
Merging of FPC WGS and ADNI WGS Data 123
Sample Filtering: Population Substructure 131
Sample Filtering: Relatedness 134
Additional Site Filtering 137
Statistical Analysis 139
Estimate of Heritability Explained by Sequenced Variants 139

Comparing Estimates of Heritability in FPC-ADNI and PanC4 Study 140
Results 140
Estimates of Heritability Explained by Sequenced Variants 144
Comparing Estimates of Heritability in FPC-ADNI and PanC4 Study 144
Discussion 145
Conclusion 148
Tables 150
Supplementary Figures and Tables 154
Reference 155

CHAPTER FIVE

CONTRIBUTION OF ESTABLISHED FPC GENES IN THE HERITABILITY OF PANCREATIC CANCER

Familial Clustering and High-penetrance Mutations for Pancreatic Cancer 157
Prevalence of High-penetrance Mutations in FPC Patients 158
Prevalence of High-penetrance Mutations in Sporadic Cases 159
Contribution of Established FPC Genes in Heritability 160
Limitation of Current Approach in Capturing Rare Variations 161
Conclusion 163
Tables 164
Reference 167

CHAPTER SIX

CONCLUSIONS, FUTURE DIRECTIONS AND PUBLIC HEALTH SIGNIFICANCE

Summary and Conclusions 172
Strengths of the Studies 175
Limitations of the Studies 177
Future Directions and Public Health Significance 178
Reference 182

LIST OF TABLES

CHAPTER TWO

Table 1 Pancreatic Cancer Predisposing Genes 48
Table 2 Genetic Susceptibility Loci for Pancreatic Cancer Identified in European Populations.. 49
Table 3 Genetic Susceptibility Loci for Pancreatic Cancer Identified in Asian Populations 51

CHAPTER THREE

Table 1 Characteristics of Analyzed PanC4 Cohort 102
Table 2 Estimates of Variance Explained by Imputed Variants from GREML-LDMS (2 MAF x 2 LD) Analysis 103
Table 3 Estimates of Variance Explained by Imputed Variants in Four Functional Groups..... 104

CHAPTER FOUR

Table 1 Characteristics of ADNI WGS Cohort at Baseline 150
Table 2 Characteristics of FPC WGS Cohort at Baseline 151
Table 3 Estimates of Heritability under Different Relatedness Cutoffs and Prevalence Values 152
Table 4 Estimates of Heritability for Pancreatic Cancer from 521,282 Variants Shared by FPC-ADNI Cohort and PanC4 Cohort..... 153

CHAPTER FIVE

Table 1 Self-reported Family History of Pancreatic Cancer in Affected Individuals 164
Table 2 Prevalence of FPC Gene Germ-line Mutations in FPC Patients..... 166

LIST OF FIGURES

CHAPTER THREE

Figure 1 Overview of quality control procedures on PanC4 imputed data. 77
Figure 2 Minor allele frequency (A) and functional annotation (B) distribution of PanC4 imputed variants. 86
Figure 3 Estimated variance explained by variants on each autosomal chromosome. 91
Figure 4 Correlations between the estimated variance explained by individual chromosome and the length of chromosomes. 92
Figure 5 The estimated proportion of phenotypic variance explained by imputed variants in six MAF categories. 93

CHAPTER FOUR

Figure 1 Overview of quality control procedures on FPC-ADNI WGS data. 118
Figure 2 Distributions of non-reference discordance rate (NRD) per sample in (A) FPC WGS cohort and (B) ADNI WGS cohort. 122
Figure 3 Comparison of FPC WGS and ADNI WGS data on chromosome 22. 127
Figure 4 Differences in site mean depth by alternative allele frequencies on chromosome 22. . 128
Figure 5 Site mean depth by alternative allele frequency groups on chromosome 22. 129
Figure 6 Differences in average site mean depth per 10-kb segment on chromosome 22. 130
Figure 7 Principal component analysis (PCA) in the merged study population. 133
Figure 8 Z_1 versus Z_0 in identity-by-descent analysis (A) and distribution of genetic similarity values (B). 136
Figure 9 Proportion of variants with missingness greater than 10%, 20%, 30%, and 40% on each chromosome. 138
Figure 10 Minor Allele Frequency (A) and Functional Annotation (B) Distribution of WGS Variants. 143

CHAPTER ONE

INTRODUCTION

Over the past 20 years, pancreatic cancer remains one of the deadliest cancers and a major public health challenge. This chapter provides the rationale of studying heritability for pancreatic cancer and how the results of the current study may guide the future strategies for gene discovery efforts. First, the key statistics of pancreatic cancer, including the prevalence, incidence, and survival rate are presented to emphasize the critical issues associated with the disease. Second, the risk factors and their impact on pancreatic cancer are reviewed. Then the limitations of the past heritability analysis on pancreatic cancer are discussed, followed by a description of the missing heritability problem in the context of pancreatic cancer. The chapter concludes with the purpose and goals of current study.

Epidemiology of Pancreatic Cancer

The pancreas is an organ located in the abdomen, behind the stomach. The pancreas has two main types of cells. The exocrine cells make up the majority of the pancreas and release enzymes for food digestion. The endocrine cells are hormone-producing cells that regulate blood sugar. These two types of cells in the pancreas form different types of tumors that present distinct characteristics and require different treatments. The vast majority of pancreatic tumors occur in the exocrine tissues. Pancreatic ductal adenocarcinoma (PDAC) is the most common type that represents > 90% of all pancreatic cancers[1].

Pancreatic cancer is a rare disease. The lifetime risk of developing cancer in the pancreas is about 1.5% by age 80 for both men and women[2]. Like all cancers, pancreatic cancer is more prevalent in aged populations. The incidence of pancreatic cancer is sharply elevated after age 55, and the median age at diagnosis is 70 years in the United States. On January 1, 2014, the

estimated prevalence of pancreatic cancer in the United States was 0.0165%[2]. Despite its low incidence, pancreatic cancer has the highest case-fatality rate of all common cancer types. It has surpassed breast cancer becoming the third-leading cause of cancer deaths in the US, responsible for an estimated 44,330 deaths in 2018. Both the incidence and mortality of pancreatic cancer have been increasing steadily over the past decade. By 2030, pancreatic cancer is predicted to become the second most common cause of cancer mortality, after lung cancer [3]. In contrast to dramatic declines in cancer-related deaths for other malignancies, the 5-year survival rate for pancreatic cancer remains low (currently 8%). The poor prognosis of pancreatic cancer is primarily attributed to its silent symptoms at early stage, the lack of reliable, accurate and non-invasive detection methods, and the lack of effective treatments for non-localized disease. More than half (52%) of patients are diagnosed at a distant stage, for which 5-year survival is less than 3%. Conventional treatments, such as chemotherapy, radiation or a combination of these modalities, have had little impact on the course of disease in late-stage cancer patients.

Risk Factors for Pancreatic Cancer

The rapid progression and short survival of patients with pancreatic cancer make research on this cancer particularly challenging. Patients are often too ill to participate in research studies, or to provide necessary information and/or bio-specimens required to answer various research questions. Studies in pancreatic cancer often have small sample sizes and hence insufficient power to draw conclusive inferences. Over the past two decades, great efforts had been made to establish longitudinal cohorts of pancreatic cancer. Large-scale collaborations of these cohorts (i.e. consortia) have provided unprecedented opportunities in revealing the causes and risk factors for pancreatic cancer.

Family History

Family history is a long-recognized risk factor for pancreatic cancer and an important predictor of disease risk. Studies have suggested that approximately 5-10% of pancreatic cancer patients report a close relative with pancreatic cancer[4,5]. Most epidemiological studies have demonstrated a 2- to 3-fold increase in risk of pancreatic cancer among individuals with affected first-degree relatives (FDRs)[6–15]. Some studies have shown even higher risk. A Swedish study reported a standardized incidence ratio (SIR) for pancreatic cancer of 1.73 (95%CI: 1.13-2.54) in offspring with at least one parent who presented with PDAC[16]. In a prospective study from the National Familial Pancreas Tumor Registry (NFPTR), the SIRs for pancreatic cancer in comparison to the SEER (Surveillance, Epidemiology, and End Results) rates were 6.4 (95% CI: 1.8-16.4) and 32.0 (95% CI: 10.2-74.7) in individuals with two and three FDRs with pancreatic cancer, respectively[17].

Cigarette Smoking

Cigarette smoking is the best-characterized risk factor for pancreatic cancer. Risk of pancreatic cancer for current smokers is about twice that for never smokers[11,18–32]. A dose-response relationship was observed with smoking intensity (i.e. number of cigarette per day) [18,19,21,22,24,33–35] and duration (i.e. number of years)[21,22,33]. A higher amount of cigarette smoking over time (i.e. pack-years) is also associated with an elevated risk of pancreatic cancer[21,24,33,36]. In several studies, smoking cessation lowered the risk of pancreatic cancer among former smokers to be comparable to never smokers after 5-10 years[19,21,22]. Given that near half of the general population being current or former smokers, it was estimated that about 11-32% of all pancreatic cancer cases could be attributable to cigarette smoking[37].

Chronic Pancreatitis

There is an increasing body of evidence suggesting chronic pancreatitis as a risk factor for pancreatic cancer. The relationship between chronic pancreatitis (CP) and pancreatic cancer was complicated. Prospective cohort studies in CP patients found up to 60% pancreatic cancer occurences within the first two years of CP diagnosis[38–42]. There was a substantially higher risk of pancreatic cancer diagnosis in the first two years, and the risk decreased with time[18,38,39,43,44]. It was suggested that a proportion of these pancreatic cancer cases may be initially misdiagnosed as CP, and hence represent a potential diagnostic delay of up to 2 years[45]. Given the rapid progression and low survival rate of pancreatic cancer, misdiagnosis would become less likely with longer periods of observation. Therefore, a lag period of two years was often considered in studies evaluating the association between CP and cancer of the pancreas. Such cohort studies had reported a 6- to 27-fold increase in risk of pancreatic cancer among CP patients compared to the general population[38–44,46–48]. In the most recent meta-analysis of 9 cohort studies and 4 case-control studies, the relative risk of pancreatic cancer in CP patients was 16.2 (95%CI: 12.6-20.7), 7.9 (95%CI: 4.3-14.7) and 3.53 (95%CI: 1.7-7.4) with a 2-year, 5-year, and >=9-year lag period, respectively[49]. Although the risk declined over time, it remained significantly elevated almost 10 years following a diagnosis of CP. In addition, increased risk of pancreatic cancer was also observed in non-smokers and non-alcoholic CP patients[25,42–44]. It was estimated that about 1-5% of CP patients would develop pancreatic cancer at 10 years after CP diagnosis[41,42,46,48,50].

Diabetes Mellitus

A bi-directional relationship has been noted between diabetes mellitus (DM) and pancreatic cancer. The risk of pancreatic cancer was correlated negatively with the duration of DM, and the highest risk was found among DM patients diagnosed within less than one year [51,52]. It was hypothesized that recent-onset diabetes could be an early manifestation of pancreatic

cancer. Early symptoms of pancreatic cancer may also favor the diagnosis of diabetes (or vice versa), thus partly explaining the excess risk in patients with a recent diagnosis of diabetes. Meta-analysis has reported an overall 2-fold increased risk of pancreatic cancer in patients with long-standing DM (duration of DM > 2 years), and a significant 50% excess risk of pancreatic cancer was still evident among patients who had diabetes for 10 or more years [51,53,54]. This association is independent of alcohol consumption, BMI, and smoking status [51,52]. Although most published studies made no distinction between type-I and type-II DM, the majority of the patients were diagnosed with type-II DM. Studies in patients with type-I DM were often small and inconclusive. In a meta-analysis of 39 pancreatic cancer studies, the risk of pancreatic cancer was similarly elevated in patients with type-I DM (RR=2.00, 95%CI: 1.37-3.01)[55]. This finding was later supported by a population-based cohort study in Taiwan involving 14,619 patients with type-I DM, where a significantly increased risk was observed in both men (SIR = 2.61, 95%CI: 1.35-4.56) and women (SIR = 4.10, 95%CI: 2.35-6.66)[56]. People with prediabetes may also have a higher risk of pancreatic cancer. There is a linear dose-response relationship between fasting blood glucose concentration and the risk of pancreatic cancer across prediabetes and diabetes, with every 0.56 mmol/L (10 mg/dL) increase in fasting blood glucose associated with a 14% increase in the risk of pancreatic cancer[57].

Obesity

Association between obesity and pancreatic cancer is often weak and less conclusive, partly due to the concurrence of obesity and diabetes. Population-based cohort studies found 1.3-2.0 fold increase in risk of pancreatic cancer among obese individuals (BMI >= 30 kg/m^2 or BMI >= 35 kg/m^2) [58–65], after adjusting for history of diabetes. Being obese is also associated with worse clinical outcome in patients with pancreatic cancer [66–69]. However, in several studies,

the weak association between obesity and pancreatic cancer was only observed in men[70–72] or women[73,74]. Higher risk of pancreatic cancer is also more likely to be observed in people who were obese in early adulthood/adolescence or in those with substantial increment in BMI [63,71,75,76]. Findings from many of these cohort studies were not conclusive due to limited power resulting from the low incidence of the disease and the competing risk of deaths among obese individuals [35,64,65,77].

Genetic Factors

Pancreatic cancer is more prevalent in families with several hereditary syndromes, including hereditary breast and ovarian cancer (HBOC), Peutz-Jeghers syndrome (PJS), familial atypical multiple mole melanoma (FAMMM), hereditary pancreatitis (HP) and Lynch syndrome (LS). These genetic syndromes are reported to be associated with a substantially higher risk of pancreatic cancer. High-penetrance germline mutations of their predisposing genes are responsible for the occurrence of the disease in 8-30% familial pancreatic cancer (FPC) cases[78–81]. These mutations have also been detected in sporadic cases with and without a family history of the disease[82–84]. Because of the high prevalence of these genes (*BRCA1/2*, *CDKN2A*, *PRSS1*, *STK11*, *PALB2*, *ATM*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*) in FPC families, they are considered the established FPC genes. An in-depth review of the risk of pancreatic cancer in these families and the predisposing genes of these genetic syndromes can be found in Chapter Two.

Genome-wide association studies (GWAS) allow for the unbiased evaluation of common genomic variants associated with pancreatic cancer. Six pancreatic cancer GWAS have been conducted by the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (PanC4) in populations of European ancestry. A total of 21 pancreatic cancer susceptibility loci located in 18 genomic regions have been discovered in European

populations. The two GWAS in Asian populations reported an additional five loci associated with pancreatic cancer. Details of these susceptibility loci can be found in Chapter Two.

Family-based Heritability and Missing Heritability Problem

A central question in the study of the inheritance of human disease is whether the observed variation of a particular trait is due to environmental factors or genetic factors – the nature versus nurture debates. Heritability is formally defined as the proportion of phenotypic variance of a trait in a given population at a given time that is attributable to variation in genetic factors. The broad-sense heritability (H^2) measures the combined effects of all loci, including possible allelic interactions within loci (dominance) and between loci (epistasis). The narrow-sense heritability (h^2) quantifies the contribution of additive genetic effects which is the sum of the average effects of parents' genes that give rise to the mean genotypic value of their progeny (Supplementary Notes). Historically, narrow-sense heritability is considered a more useful parameter to predict the selection response in the breeding of plants and livestock animals. In the study of human traits, it is generally believed that the resemblance of a trait between relatives is mostly due to the additive genetic effects[85]. Therefore, current heritability studies have been primarily focused on the narrow-sense heritability.

Heritability is a population-specific parameter since both the genetic variance and environmental variance can vary across different populations. Differences in allele frequencies, the effect sizes and the mode of gene actions lead to changes in genetic variance of a trait. The shift of environmental exposures can also change heritability. In practice, heritability of the same trait is often quite similar across populations and species[85].

Family-based Heritability

Traditionally, heritability was estimated in family-based studies, as simple functions of the regression of offspring phenotypes on parental phenotypes, the correlation of full or half-sibs, and the difference in the correlation of monozygotic (MZ) and dizygotic (DZ) twin pairs. For example, in the classic twin studies, Falconer quantifies the heritability of a trait as twice the difference between the DZ resemblance (covariance or correlation) and the MZ resemblance[86].

The rarity and fast progression of pancreatic cancer have made heritability studies of the disease particularly difficult. The often-quoted estimate of heritability for pancreatic cancer was reported by a population-based twin study. Lichtenstein and his colleagues collected data on 44,788 pairs of twins listed in the Swedish, Danish, and Finnish twin register, aiming to assess the relative importance of heritable and environmental factors in causing cancer at 28 anatomical sites [87]. The heritability for pancreatic cancer (h^2_{fam}) was estimated to be 0.36 (95% CI: 0 – 0.53). The low precision of the heritability estimate is likely due to the small number of twin pairs affected by the disease. Despite the large sample size, there were only 3 pairs of concordant MZ twins and 1 pair of concordant DZ twins in men and women combined. Similarly, the most recent twin study with 80,309 monozygotic and 123,382 dizygotic twin individuals from population-based twin registers of Denmark, Finland, Norway and Sweden failed to estimate the heritability for pancreatic cancer due to having less than five pairs of concordant twin pairs diagnosed with the disease[88]. The limited statistical power of family-based studies has prohibited the further exploration of the genetic contribution to the risk of pancreatic cancer.

A general concern with family-based heritability is the potential inflation due to either the common environmental factors shared by relatives or the non-additive genetic effects. A crucial assumption underlying the twin study design is that the common environmental variation is equal among MZ and DZ twins [89,90]. The validity of this assumption is controversial since even

monochorionic and dichorionic MZ twins show differences in the magnitude of shared environmental effects for some traits[91,92]. Violation of equal-environment assumption is particularly alarming in twin studies of behavioral phenotypes[93,94]. In addition, the design of twin studies does not allow the estimate of non-additive genetic variance which would also lead to an overestimation of the narrow-sense heritability[89,90]. Heritability estimated from other types of family-based studies depends on the appropriate partitioning of observed variation into unobserved genetic and environmental factors. Recent studies have reported an overestimation of family-based heritability from statistical models that did not account for the shared environmental factors[95]. This inflation in family-based heritability was observed in a number of well-studied quantitative traits and binary diseases, including several common cancer phenotypes (e.g. lung cancer, prostate cancer and breast cancer)[95–98]. Moreover, family-based studies require large random population-based sampling of relative pairs, regardless of their disease status. Incomplete ascertainment will lead to a biased estimate of heritability (ascertainment bias). A previous study suggested that when the rate of diagnosis among relatives of probands was higher than the rate of diagnosis in population, the estimate of heritability from family-based studies were likely inflated[99].

Missing Heritability Problem

Pancreatic cancer is a genetically highly heterogeneous disease. Despite the large sample sizes of current GWAS studies, the identified common susceptibility loci together explain < 5% of the total phenotypic variation for pancreatic cancer [100–102]. Comparing to the family-based estimate of heritability (h^2_{fam}= 36%), it appears that a large proportion of heritability is unexplained, highlighting the so-called “missing heritability” problem.

There have been extensive debates on the source of missing heritability in complex diseases or traits. It is reasonable to assume that due to the differences in genetic architecture (number, type, effect size and frequency of susceptibility variants) across traits, missing heritability may take a different form for different diseases[103]. Many explanations for missing heritability have been suggested; each has different implications for the design of the next round gene discovery efforts [104]. Although inflated family-based heritability may partially explain the missing heritability for many complex traits or diseases, such inflation is likely to be trait-specific and remains debatable for pancreatic cancer. The fact that known GWAS loci only account for a small fraction of variance explained by all common variants implies that much of the missing heritability could be due to variants whose effects are too small to reach genome-wide significance. Thus, larger sample sizes are required to increase the power of GWAS studies for identification of yet-to-be-detected common variants. Rare variants are neither directly genotyped nor well tagged by the genotyping arrays. If the missing heritability is largely due to rare variants, extending coverage on these variants through whole-exome or whole-genome sequencing (WGS) in a large sample is more important for continued progress in gene discoveries. An excess of significant GWAS associations in non-coding regions underscores the importance of including both coding and regulatory regions in the search for disease-associated variants [105,106]. Understanding the genetic architecture of the disease is the key to find missing heritability and to guide the future strategies in discovering the disease-causing genes.

Population-based Heritability and GREML Approach

The Single-Component GREML Approach

To address the missing heritability problem, alternative approaches have been developed to estimate heritability in populations of unrelated individuals. GWAS relies on a very stringent

P value (e.g. 5×10^{-8}) to reduce the occurrence of false-positive results, but it may cause many real associations to be missed. Instead of focusing on the statistical significance of individual causal variants, these approaches are designed to capture the cumulative effects of all causal variants simultaneously.

One of the most widely applied methods in heritability analysis is GREML, which stands for the genomic-relatedness-based restricted maximum-likelihood (REML) approach. This approach assumes the effect sizes of causal variants are independent of the linkage disequilibrium (LD) and are inversely proportionate to their minor allele frequency (MAF), which implies that the per-variant contribution to heritability is constant across MAF spectrum[107]. Comparing to the family-based studies, this approach utilizes the very distant relationships between unrelated individuals to estimate heritability explained by causal variants passed down from their remote common ancestors. Therefore, estimates of heritability using GREML are less biased by shared environmental factors.

The simplest form of GREML approach is the single-component GREML (GREML-SC), where the disease status as cases (=1) or controls (=0) is modeled as a function of one or more fixed effects (u) including the intercept and covariates, a random genetic effect (g), and a residual effect (e) in a mixed linear model:

$$y = u + g + e \tag{1}$$

The total phenotypic variance is the sum of the variance due to random genetic effects ($G\sigma_g^2$) and the variance due to residual effects ($I\sigma_e^2$):

$$var(y) = G\sigma_g^2 + I\sigma_e^2 \tag{2}$$

Assuming all the causal variants are known, a genetic relationship matrix (GRM) G estimated from all causal variants can be obtained to determine the genetic similarities between

pairwise individuals. Given G, the variance due to random genetic effects can be estimated using the REML approach and the proportion of phenotypic variance explained by all causal variants is defined as

$$h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2) \quad (3)$$

For binary phenotypes, the estimates of variance need to be transformed from the observed 0-1 scale to the unobserved continuous 'liability' scale using a probit transformation[108], and a population prevalence of the disease (K) need to be specified to correct for the oversampling of cases in the case-control study.

Array-based Heritability

Heritability estimated from equation (3) can be considered as the overall heritability given all causal variants were known. However, in practice, we know little about the causal variants for a trait or disease. Instead, this approach has been widely applied in GWAS to estimate the proportion of phenotypic variation explained by the additive effects of all genotyped single-nucleotide polymorphisms (SNPs), termed array heritability. In this case, a different GRM (A) is constructed based on the genetic similarities estimated from all genotyped SNPs.

$$var(y) = A\sigma_g^2 + I\sigma_e^2 \quad (4)$$

The heritability is then defined as $h^2_{array} = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$, which estimates the array heritability, not the overall heritability. Like in GWAS, how much of the phenotypic variation can be captured by genotyped SNPs depends on the LD between the causal variants and the genotyped SNPs. Because GWAS arrays target primarily SNPs with MAF > 1%, the genetic variance due to rare variants (MAF < 1%) is not well captured in array heritability estimates. As a result, the array heritability is always an underestimation of the overall heritability.

In a collection of large case-control GWAS from the National Cancer Institute's Cancer Genomic Research Laboratory, the array heritability of pancreatic cancer was estimated to be 0.098 (95% CI: 0.037 – 0.160)[102]. Using the GWAS data from the Pancreatic Cancer Cohort Consortium (PanScan), about 0.18 (95%CI: 0.06 – 0.30) of the total phenotypic variation of pancreatic cancer was explained by common genotyped single-nucleotide polymorphisms (SNPs) [100]. In the Pancreatic Cancer Case-Control Consortium (PanC4) samples of Caucasian ancestry, 0.164 (95% CI: 0.104 – 0.224) of the total phenotypic variation was due to genome-wide common variants[101]. When combining the PanC4 data with PanScan data, the array heritability was estimated to be 0.131 (95%CI: 0.099 – 0.163)[101]. The variability of array heritability in these studies likely reflects the differences in study populations (though all study populations listed above are of European ancestry) and genotyping arrays.

Motivations of This Study

The genomics era has brought new tools and resources to study the relationship between phenotype and genotype. The development of novel statistical methods offers new research avenues to estimate heritability in unpedigreed populations using different types of genetic data. The previous estimates of array heritability have direct implications on the common causal variants in the susceptibility of pancreatic cancer, but provide no information regarding the rare causal variants. In addition, recent studies have found bias in array heritability estimated using GREML-SC, due to the uneven tagging of causal variants. To obtain a more comprehensive and less biased estimate of heritability for pancreatic cancer, this study is motivated by two factors: (1) the availability of imputation and whole-genome sequencing (WGS) data, and (2) the improvements on heritability estimates using the multi-component GREML approach.

Imputation and Whole-genome Sequencing Data

Comparing to genotyping arrays, imputation and WGS data both provide a more completed characterization of genetic variation across the genome, particularly on variants at the lower end of MAF spectrum. With the extended coverage on rare and low-frequency variants, the LD between causal variants and the analyzed variants would be greatly improved using imputation and WGS data. Simulation studies have demonstrated that heritability estimated from imputed variants was able to capture almost all variation at common causal variants and a proportion of variation at rare causal variants[109]. If rare variants do contribute to the susceptibility of pancreatic cancer, we would expect to see an increase in heritability estimate using imputation data ($h^2_{imputed}$) in relative to array data. With sufficient sample sizes, WGS data in theory has the potential to capture all causal variants and hence generate the closest estimate of overall heritability, which could further approach to the family-based heritability for pancreatic cancer.

The Multi-Component GREML Approach

Despite its wide applications in GWAS studies, heritability estimated from GREML-SC is known to be biased, due to the average LD between causal variants and analyzed variants being higher or lower than the average LD between analyzed variants[110]. Even with whole-genome sequencing data, the GREML-SC estimate of heritability would be biased if the causal variants had a different MAF spectrum or LD property than the variants used in the analysis[109]. Such bias could be ameliorated by the multi-component GREML approach with additional stratification on MAF and/or LD.

The multi-component GREML approach takes a general form as,

$$y = u + \sum_t^T g_t + e \quad (5)$$

where more than one genetic components (g_t) are included in a mixed linear model. The total phenotypic variance now becomes the sum of variance attributable to each genetic component as well as the residual effects,

$$var(y) = \sum_t^T A_t \sigma_t^2 + I\sigma_e^2 \quad (6)$$

and the heritability for a trait can be calculated as

$$h^2 = \sum_t^T \sigma_t^2 / (\sum_t^T \sigma_t^2 + \sigma_e^2) \quad (7)$$

The multi-component GREML approach has two variations: GREML-MS and GREML-LDMS. In a MAF-stratified GREML (GREML-MS) analysis, the number of genetic components corresponds to the number of MAF groups defined at specific thresholds. This approach can adjust for the difference in LD caused by a difference in the MAF spectrum but is unable to correct for the region-specific LD heterogeneity across the genome. Empirical data suggested that GREML-MS was able to minimize most bias observed in GREML-SC estimates[111]. If the causal variants tend to be enriched in genomic regions with higher or lower LD than average (e.g. DNase I-hypersensitive sites), the GREML-MS estimates of heritability will be biased[109]. The LD- and MAF-stratified GREML (GREML-LDMS), which uses a sliding window approach to fit the region-specific LD heterogeneity across the genome, is demonstrated to produce more valid estimates of heritability across different simulation scenarios of causal variant distributions[109,110]. Because GREML-LDMS fits a large number of components, a much larger sample size is required to obtain acceptable standard errors (e.g. < 0.05).

By partitioning variants into narrower MAF and LD ranges in multi-component models, the LD between causal variants and analyzed variants get closer to the LD between analyzed variants, and therefore the variance estimates associated with each partitioned GRM are likely to be approximately unbiased. The multi-component GREML approach not only provides less

biased estimates of heritability, but also allow the investigation of the genetic architecture of complex traits through the estimation of heritability components from different variant sets.

Goals of this Study

Pancreatic cancer represents a major public health challenge. The limited progress in prevention, early detection, and treatment of pancreatic cancer underscore the urgent need for continued research efforts to illuminate the biology and etiology of the disease. Inherited genetic variation plays an important role in both the familial and sporadic occurrences of pancreatic cancer, and may be used to characterize populations who will benefit from early detections. Heritability (narrow-sense heritability) measures the contribution of additive genetic effects to the susceptibility of a disease. In the past, family-based heritability analysis for pancreatic cancer was limited because of the rarity of the disease. Such studies often lack the sufficient statistical power to produce meaningful estimates. In addition, heritability estimated from family-based studies may be inflated due to shared environmental factors within families. The newly developed statistical approaches allow the assessment of heritability in unrelated individuals, which teases apart the contribution of environmental factors shared among relative. However, heritability derived from GWAS studies are generally underestimated since rare variants are not well captured by genotyping arrays.

Finding the missing heritability of pancreatic cancer requires an accurate estimate of heritability and a better understanding of the genetic variation underlying the disease. The goal of this study is to comprehensively evaluate the genetic contribution to the susceptibility of pancreatic cancer using both whole-genome sequencing data and imputed data. These data provide a near-complete characterization of the genetic variation across the genome, minimizing the bias in heritability estimation due to inadequate coverage of causal variants, particularly

those at the lower end of MAF spectrum. The application of the multi-component GREML approach in this study took into account the unknown MAF distributions and LD properties of causal variants, generating a less biased estimate of heritability for pancreatic cancer. Using heritability measures, we explored the genetic architecture of pancreatic cancer through partitioning the estimated heritability by MAF, chromosomes and functional annotations. This analysis was aimed to answer critical questions about the role of rare and common variants in the susceptibility of pancreatic cancer, as well as the importance of coding and regulatory variants in pancreatic cancer risk. In addition to the reported GWAS loci for pancreatic cancer, we also quantified the contributions of established FPC genes, which further helped determine the source of missing heritability.

Supplementary Notes

Observed phenotypes (P) of a trait can be partitioned into a simple statistical model representing the contribution of the unobserved genetic (G) and environmental factors (E).

$$Phenotype(P) = Genotype(G) + Environment(E) \quad (1)$$

The variance of the observed phenotype (σ_P^2) can be expressed as a sum of unobserved underlying variance due to genetic (σ_G^2) and environmental factors (σ_E^2).

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 \quad (2)$$

The broad-sense heritability (H^2) is defined as the proportion of phenotypic variance that can be attributed to the variance of genetic factors.

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2} \quad (3)$$

The genetic variance can be further partitioned into the variance due to additive (σ_A^2), dominance (σ_D^2) and epistatic genetic effects (σ_I^2).

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 \quad (4)$$

The narrow-sense heritability (h^2) is defined as the proportion of phenotypic variance that can be attributed to the additive genetic effects.

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} \quad (3)$$

Reference

1. Cancer Facts & Figures 2017 [Internet]. [cited 2017 Sep 25]. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf

2. Howlader N, Noone A, Krapcho M, Miller D, Bishop K, Kosary C, et al. SEER Cancer Statistics Review, 1975-2014 [Internet]. National Cancer Institute, Bethesda, MD; Available from: https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission, posted to the SEER web site, April 2017.

3. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014 Jun 1;74(11):2913–21.

4. Hruban RH, Petersen GM, Ha PK, Kern SE. Genetics of pancreatic cancer. From genes to families. Surg Oncol Clin N Am. 1998 Jan;7(1):1–23.

5. Klein AP. Identifying people at a high risk of developing pancreatic cancer. Nat Rev Cancer. 2013 Jan;13(1):66.

6. Fernandez E, La Vecchia C, D'Avanzo B, Negri E, Franceschi S. Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 1994 May;3(3):209–12.

7. Silverman DT, Schiffman M, Everhart J, Goldstein A, Lillemoe KD, Swanson GM, et al. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer. 1999 Aug;80(11):1830–7.

8. Schenk M, Schwartz AG, O'Neal E, Kinnard M, Greenson JK, Fryzek JP, et al. Familial risk of pancreatic cancer. J Natl Cancer Inst. 2001 Apr 18;93(8):640–4.

9. Fehringer G, Gallinger S, Borgida A, Zhang LR, Adams L, Liu G, et al. The association of family history of cancer and medical history with pancreatic cancer risk. Pancreas. 2014 Jul;43(5):812–4.

10. Schulte A, Pandeya N, Fawcett J, Fritschi L, Klein K, Risch HA, et al. Association between family cancer history and risk of pancreatic cancer. Cancer Epidemiol. 2016 Dec;45:145–50.

11. Zheng Z, Zheng R, He Y, Sun X, Wang N, Chen T, et al. Risk Factors for Pancreatic Cancer in China: A Multicenter Case-Control Study. J Epidemiol. 2016;26(2):64–70.

12. Jacobs EJ, Rodriguez C, Newton CC, Bain EB, Patel AV, Feigelson HS, et al. Family history of various cancers and pancreatic cancer mortality in a large cohort. Cancer Causes Control CCC. 2009 Oct;20(8):1261–9.

13. Permuth-Wey J, Egan KM. Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. Fam Cancer. 2009;8(2):109–17.

14. Jacobs EJ, Chanock SJ, Fuchs CS, Lacroix A, McWilliams RR, Steplowski E, et al. Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Int J Cancer J Int Cancer. 2010 Sep 1;127(6):1421–8.

15. Austin MA, Kuo E, Van Den Eeden SK, Mandelson MT, Brentnall TA, Kamineni A, et al. Family history of diabetes and pancreatic cancer as risk factors for pancreatic cancer: the PACIFIC study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2013 Oct;22(10):1913–7.

16. Hemminki K, Li X. Familial and second primary pancreatic cancers: a nationwide epidemiologic study from Sweden. Int J Cancer. 2003 Feb 10;103(4):525–30.

17. Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJA, et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res. 2004 Apr 1;64(7):2634–8.

18. Talamini R, Polesel J, Gallus S, Dal Maso L, Zucchetto A, Negri E, et al. Tobacco smoking, alcohol consumption and pancreatic cancer risk: a case-control study in Italy. Eur J Cancer Oxf Engl 1990. 2010 Jan;46(2):370–6.

19. Vrieling A, Bueno-de-Mesquita HB, Boshuizen HC, Michaud DS, Severinsen MT, Overvad K, et al. Cigarette smoking, environmental tobacco smoke exposure and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2010 May 15;126(10):2394–403.

20. Matsubayashi H, Maeda A, Kanemoto H, Uesaka K, Yamazaki K, Hironaka S, et al. Risk factors of familial pancreatic cancer in Japan: current smoking and recent onset of diabetes. Pancreas. 2011 Aug;40(6):974–8.

21. Tranah GJ, Holly EA, Wang F, Bracci PM. Cigarette, cigar and pipe smoking, passive smoke exposure, and risk of pancreatic cancer: a population-based study in the San Francisco Bay Area. BMC Cancer. 2011 Apr 15;11:138.

22. Bosetti C, Lucenteforte E, Silverman DT, Petersen G, Bracci PM, Ji BT, et al. Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (Panc4). Ann Oncol Off J Eur Soc Med Oncol ESMO. 2012 Jul;23(7):1880–8.

23. Kuzmickiene I, Everatt R, Virviciute D, Tamosiunas A, Radisauskas R, Reklaitiene R, et al. Smoking and other risk factors for pancreatic cancer: a cohort study in men in Lithuania. Cancer Epidemiol. 2013 Apr;37(2):133–9.

24. Schulte A, Pandeya N, Tran B, Fawcett J, Fritschi L, Risch HA, et al. Cigarette smoking and pancreatic cancer risk: more to the story than just pack-years. Eur J Cancer Oxf Engl 1990. 2014 Mar;50(5):997–1003.

25. Wang Y, Duan H, Yang X, Guo J. Cigarette smoking and the risk of pancreatic cancer: a case-control study. Med Oncol Northwood Lond Engl. 2014 Oct;31(10):184.

26. Bertuccio P, La Vecchia C, Silverman DT, Petersen GM, Bracci PM, Negri E, et al. Cigar and pipe smoking, smokeless tobacco use and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (PanC4). Ann Oncol Off J Eur Soc Med Oncol ESMO. 2011 Jun;22(6):1420–6.

27. Matsuo K, Ito H, Wakai K, Nagata C, Mizoue T, Tanaka K, et al. Cigarette smoking and pancreas cancer risk: an evaluation based on a systematic review of epidemiologic evidence in the Japanese population. Jpn J Clin Oncol. 2011 Nov;41(11):1292–302.

28. Lynch SM, Vrieling A, Lubin JH, Kraft P, Mendelsohn JB, Hartge P, et al. Cigarette smoking and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. Am J Epidemiol. 2009 Aug 15;170(4):403–13.

29. Iodice S, Gandini S, Maisonneuve P, Lowenfels AB. Tobacco and the risk of pancreatic cancer: a review and meta-analysis. Langenbecks Arch Surg. 2008 Jul;393(4):535–45.

30. Antwi SO, Oberg AL, Shivappa N, Bamlet WR, Chaffee KG, Steck SE, et al. Pancreatic cancer: associations of inflammatory potential of diet, cigarette smoking and long-standing diabetes. Carcinogenesis. 2016;37(5):481–90.

31. Andersson G, Wennersten C, Borgquist S, Jirström K. Pancreatic cancer risk in relation to sex, lifestyle factors, and pre-diagnostic anthropometry in the Malmö Diet and Cancer Study. Biol Sex Differ. 2016;7:66.

32. Inoue-Choi M, Hartge P, Liao LM, Caporaso N, Freedman ND. Association between long-term low-intensity cigarette smoking and incidence of smoking-related cancer in the National Institutes of Health-AARP cohort. Int J Cancer. 2017 Sep 20;

33. Zou L, Zhong R, Shen N, Chen W, Zhu B, Ke J, et al. Non-linear dose-response relationship between cigarette smoking and pancreatic cancer risk: evidence from a meta-analysis of 42 observational studies. Eur J Cancer Oxf Engl 1990. 2014 Jan;50(1):193–203.

34. Rosato V, Polesel J, Bosetti C, Serraino D, Negri E, La Vecchia C. Population attributable risk for pancreatic cancer in Northern Italy. Pancreas. 2015 Mar;44(2):216–20.

35. Ansary-Moghaddam A, Huxley R, Barzi F, Lawes C, Ohkubo T, Fang X, et al. The effect of modifiable risk factors on pancreatic cancer mortality in populations of the Asia-Pacific region. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2006 Dec;15(12):2435–40.

36. McWilliams RR, Maisonneuve P, Bamlet WR, Petersen GM, Li D, Risch HA, et al. Risk Factors for Early-Onset and Very-Early-Onset Pancreatic Adenocarcinoma: A Pancreatic Cancer Case-Control Consortium (PanC4) Analysis. Pancreas. 2016 Feb;45(2):311–6.

37. Maisonneuve P, Lowenfels AB. Risk factors for pancreatic cancer: a summary review of meta-analytical studies. Int J Epidemiol. 2015 Feb;44(1):186–98.

38. Lowenfels AB, Maisonneuve P, Cavallini G, Ammann RW, Lankisch PG, Andersen JR, et al. Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. N Engl J Med. 1993 May 20;328(20):1433–7.

39. Karlson BM, Ekbom A, Josefsson S, McLaughlin JK, Fraumeni JF, Nyrén O. The risk of pancreatic cancer following pancreatitis: an association due to confounding? Gastroenterology. 1997 Aug;113(2):587–92.

40. Talamini G, Falconi M, Bassi C, Sartori N, Salvia R, Caldiron E, et al. Incidence of cancer in the course of chronic pancreatitis. Am J Gastroenterol. 1999 May;94(5):1253–60.

41. Wang W, Liao Z, Li G, Li Z-S, Chen J, Zhan X-B, et al. Incidence of pancreatic cancer in chinese patients with chronic pancreatitis. Pancreatol Off J Int Assoc Pancreatol IAP Al. 2011;11(1):16–23.

42. Hao L, Zeng X-P, Xin L, Wang D, Pan J, Bi Y-W, et al. Incidence of and risk factors for pancreatic cancer in chronic pancreatitis: A cohort of 1656 patients. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver. 2017 Nov;49(11):1249–56.

43. Bansal P, Sonnenberg A. Pancreatitis is a risk factor for pancreatic cancer. Gastroenterology. 1995 Jul;109(1):247–51.

44. Bang UC, Benfield T, Hyldstrup L, Bendtsen F, Beck Jensen J-E. Mortality, cancer, and comorbidities associated with chronic pancreatitis: a Danish nationwide matched-cohort study. Gastroenterology. 2014 Apr;146(4):989–94.

45. Munigala S, Kanwal F, Xian H, Agarwal B. New diagnosis of chronic pancreatitis: risk of missing an underlying pancreatic cancer. Am J Gastroenterol. 2014 Nov;109(11):1824–30.

46. Malka D, Hammel P, Maire F, Rufat P, Madeira I, Pessione F, et al. Risk of pancreatic adenocarcinoma in chronic pancreatitis. Gut. 2002 Dec;51(6):849–52.

47. Goldacre MJ, Wotton CJ, Yeates D, Seagroatt V, Collier J. Liver cirrhosis, other liver diseases, pancreatitis and subsequent cancer: record linkage study. Eur J Gastroenterol Hepatol. 2008 May;20(5):384–92.

48. Ueda J, Tanaka M, Ohtsuka T, Tokunaga S, Shimosegawa T, Research Committee of Intractable Diseases of the Pancreas. Surgery for chronic pancreatitis decreases the risk for pancreatic cancer: a multicenter retrospective analysis. Surgery. 2013 Mar;153(3):357–64.

49. Kirkegård J, Mortensen FV, Cronin-Fenton D. Chronic Pancreatitis and Pancreatic Cancer Risk: A Systematic Review and Meta-analysis. Am J Gastroenterol. 2017 Sep;112(9):1366–72.

50. Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK, Perrault J, et al. Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. J Natl Cancer Inst. 1997 Mar 19;89(6):442–6.

51. Ben Q, Xu M, Ning X, Liu J, Hong S, Huang W, et al. Diabetes mellitus and risk of pancreatic cancer: A meta-analysis of cohort studies. Eur J Cancer Oxf Engl 1990. 2011 Sep;47(13):1928–37.

52. Bosetti C, Rosato V, Li D, Silverman D, Petersen GM, Bracci PM, et al. Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium. Ann Oncol Off J Eur Soc Med Oncol. 2014 Oct;25(10):2065–72.

53. Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. JAMA. 1995 May 24;273(20):1605–9.

54. Huxley R, Ansary-Moghaddam A, Berrington de González A, Barzi F, Woodward M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. Br J Cancer. 2005 Jun 6;92(11):2076–83.

55. Stevens RJ, Roddam AW, Beral V. Pancreatic cancer in type 1 and young-onset diabetes: systematic review and meta-analysis. Br J Cancer. 2007 Feb 12;96(3):507–9.

56. Hsu P-C, Lin W-H, Kuo T-H, Lee H-M, Kuo C, Li C-Y. A Population-Based Cohort Study of All-Cause and Site-Specific Cancer Incidence Among Patients With Type 1 Diabetes Mellitus in Taiwan. J Epidemiol. 2015;25(9):567–73.

57. Liao W-C, Tu Y-K, Wu M-S, Lin J-T, Wang H-P, Chien K-L. Blood glucose concentration and risk of pancreatic cancer: systematic review and dose-response meta-analysis. BMJ. 2015 02;350:g7371.

58. Michaud DS, Giovannucci E, Willett WC, Colditz GA, Stampfer MJ, Fuchs CS. Physical activity, obesity, height, and the risk of pancreatic cancer. JAMA. 2001 Aug 22;286(8):921–9.

59. Larsson SC, Permert J, Håkansson N, Näslund I, Bergkvist L, Wolk A. Overall obesity, abdominal adiposity, diabetes and cigarette smoking in relation to the risk of pancreatic cancer in two Swedish population-based cohorts. Br J Cancer. 2005 Nov 28;93(11):1310–5.

60. Patel AV, Rodriguez C, Bernstein L, Chao A, Thun MJ, Calle EE. Obesity, recreational physical activity, and risk of pancreatic cancer in a large U.S. Cohort. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2005 Feb;14(2):459–66.

61. Stolzenberg-Solomon RZ, Adams K, Leitzmann M, Schairer C, Michaud DS, Hollenbeck A, et al. Adiposity, physical activity, and pancreatic cancer in the National Institutes of Health-AARP Diet and Health Cohort. Am J Epidemiol. 2008 Mar 1;167(5):586–97.

62. Arslan AA, Helzlsouer KJ, Kooperberg C, Shu X-O, Steplowski E, Bueno-de-Mesquita HB, et al. Anthropometric measures, body mass index, and pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Arch Intern Med. 2010 May 10;170(9):791–802.

63. Genkinger JM, Spiegelman D, Anderson KE, Bernstein L, van den Brandt PA, Calle EE, et al. A pooled analysis of 14 cohort studies of anthropometric factors and pancreatic cancer risk. Int J Cancer. 2011 Oct 1;129(7):1708–17.

64. Alsamarrai A, Das SLM, Windsor JA, Petrov MS. Factors that affect risk for pancreatic disease in the general population: a systematic review and meta-analysis of prospective cohort studies. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2014 Oct;12(10):1635-1644.e5; quiz e103.

65. Johansen D, Stocks T, Jonsson H, Lindkvist B, Björge T, Concin H, et al. Metabolic factors and the risk of pancreatic cancer: a prospective analysis of almost 580,000 men and women in the Metabolic Syndrome and Cancer Project. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2010 Sep;19(9):2307–17.

66. Bethea TN, Kitahara CM, Sonderman J, Patel AV, Harvey C, Knutsen SF, et al. A pooled analysis of body mass index and pancreatic cancer mortality in african americans. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2014 Oct;23(10):2119–25.

67. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003 Apr 24;348(17):1625–38.

68. Yuan C, Bao Y, Wu C, Kraft P, Ogino S, Ng K, et al. Prediagnostic body mass index and pancreatic cancer survival. J Clin Oncol Off J Am Soc Clin Oncol. 2013 Nov 20;31(33):4229–34.

69. Kasenda B, Bass A, Koeberle D, Pestalozzi B, Borner M, Herrmann R, et al. Survival in overweight patients with advanced pancreatic carcinoma: a multicentre cohort study. BMC Cancer. 2014 Sep 29;14:728.

70. Nöthlings U, Wilkens LR, Murphy SP, Hankin JH, Henderson BE, Kolonel LN. Body mass index and physical activity as risk factors for pancreatic cancer: the Multiethnic Cohort Study. Cancer Causes Control CCC. 2007 Mar;18(2):165–75.

71. Lin Y, Kikuchi S, Tamakoshi A, Yagyu K, Obata Y, Inaba Y, et al. Obesity, physical activity and the risk of pancreatic cancer in a large Japanese cohort. Int J Cancer. 2007 Jun 15;120(12):2665–71.

72. Rapp K, Schroeder J, Klenk J, Stoehr S, Ulmer H, Concin H, et al. Obesity and incidence of cancer: a large cohort study of over 145,000 adults in Austria. Br J Cancer. 2005 Oct 31;93(9):1062–7.

73. Inoue M, Noda M, Kurahashi N, Iwasaki M, Sasazuki S, Iso H, et al. Impact of metabolic factors on subsequent cancer risk: results from a large-scale population-based cohort study in Japan. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP. 2009 Jun;18(3):240–7.

74. Jee SH, Yun JE, Park EJ, Cho ER, Park IS, Sull JW, et al. Body mass index and cancer risk in Korean men and women. Int J Cancer. 2008 Oct 15;123(8):1892–6.

75. Genkinger JM, Kitahara CM, Bernstein L, Berrington de Gonzalez A, Brotzman M, Elena JW, et al. Central adiposity, obesity during early adulthood, and pancreatic cancer mortality in a pooled analysis of cohort studies. Ann Oncol Off J Eur Soc Med Oncol ESMO. 2015 Nov;26(11):2257–66.

76. Levi Z, Kark JD, Afek A, Derazne E, Tzur D, Furman M, et al. Measured body mass index in adolescence and the incidence of pancreatic cancer in a cohort of 720,000 Jewish men. Cancer Causes Control CCC. 2012 Feb;23(2):371–8.

77. Berrington de Gonzalez A, Sweetland S, Spencer E. A meta-analysis of obesity and the risk of pancreatic cancer. Br J Cancer. 2003 Aug 4;89(3):519–23.

78. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med Off J Am Coll Med Genet. 2015 Jul;17(7):569–77.

79. Catts ZA-K, Baig MK, Milewski B, Keywan C, Guarino M, Petrelli N. Statewide Retrospective Review of Familial Pancreatic Cancer in Delaware, and Frequency of Genetic Mutations in Pancreatic Cancer Kindreds. Ann Surg Oncol. 2016 May;23(5):1729–35.

80. Takai E, Yachida S, Shimizu K, Furuse J, Kubo E, Ohmoto A, et al. Germline mutations in Japanese familial pancreatic cancer patients. Oncotarget. 2016 Nov 8;7(45):74227–35.

81. Chaffee KG, Oberg AL, McWilliams RR, Majithia N, Allen BA, Kidd J, et al. Prevalence of germ-line mutations in cancer genes among pancreatic cancer patients with a positive family history. Genet Med Off J Am Coll Med Genet. 2017 Jul 20;

82. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious Germline Mutations in Patients With Apparently Sporadic Pancreatic Adenocarcinoma. J Clin Oncol Off J Am Soc Clin Oncol. 2017 Oct 20;35(30):3382–90.

83. Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. JAMA. 2018 Jun 19;319(23):2401–9.

84. Yurgelun MB, Chittenden AB, Morales-Oyarvide V, Rubinson DA, Dunne RF, Kozak MM, et al. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. Genet Med Off J Am Coll Med Genet. 2018 Jul 2;

85. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008 Apr;9(4):255–66.

86. Falconer D, MacKay T. Introduction to Quantitative Genetics. 4th Ed. 1996. Longmans Green, Harlow, Essex, UK;

87. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000 Jul 13;343(2):78–85.

88. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA. 2016 Jan 5;315(1):68–76.

89. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. Nat Rev Genet. 2013 Feb;14(2):139–49.

90. Vinkhuyzen AAE, Wray NR, Yang J, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet. 2013;47:75–95.

91. Reed T, Uchida IA, Norton JA, Christian JC. Comparisons of dermatoglyphic patterns in monochorionic and dichorionic monozygotic twins. Am J Hum Genet. 1978 Jul;30(4):383–91.

92. Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AHC, et al. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009 Feb;41(2):240–5.

93. Littvay L. Do heritability estimates of political phenotypes suffer from an equal environment assumption violation? Evidence from an empirical study. Twin Res Hum Genet Off J Int Soc Twin Stud. 2012 Feb;15(1):6–14.

94. Fosse R, Joseph J, Richardson K. A critical assessment of the equal-environment assumption of the twin method for schizophrenia. Front Psychiatry. 2015;6:62.

95. Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. Nat Genet. 2016;48(9):980–3.

96. Visscher PM, McEVOY B, Yang J. From Galton to GWAS: quantitative genetics of human height. Genet Res. 2010 Dec;92(5–6):371–9.

97. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. 2013 May;9(5):e1003520.

98. Liu C, Dupuis J, Larson MG, Cupples LA, Ordovas JM, Vasan RS, et al. Revisiting heritability accounting for shared environmental effects and maternal inheritance. Hum Genet. 2015 Feb;134(2):169–79.

99. Hrubec Z. The effect of diagnostic ascertainment in twins on the assessment of the genetic factor in disease etiology. Am J Hum Genet. 1973 Jan;25(1):15–28.

100. Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, et al. Most common "sporadic" cancers have a significant germline genetic component. Hum Mol Genet. 2014 Nov 15;23(22):6112–8.

101. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet. 2015 Aug;47(8):911–6.

102. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, et al. Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. J Natl Cancer Inst. 2015 Dec;107(12).

103. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet. 2002 Oct 1;11(20):2417–23.

104. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747–53.

105. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012 Sep 7;337(6099):1190–5.

106. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009 Jun 9;106(23):9362–7.

107. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010 Jul;42(7):565–9.

108. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011 Mar 11;88(3):294–305.

109. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

110. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet. 2018 Apr 26;

111. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

CHAPTER TWO

INHERITED PANCREATIC CANCER

Pancreatic cancers arise through a series of genetic events both inherited and acquired. Inherited genetic changes, both high penetrance and low penetrance, are an important component of pancreatic cancer risk, and may be used to characterize populations who will benefit from early detection. Furthermore, pancreatic cancer patients with inherited mutations may be particularly sensitive to certain targeted agents, providing an opportunity to personalized treatment. Family history of pancreatic cancer is one of the strongest risk factors for the disease, and is associated with an increased risk of cancers at other sites, including but not limited to breast, ovarian and colorectal cancer. The goal of this chapter is to discuss the importance of family history of pancreatic cancer, and the known genes that account for a portion of the familial clustering of pancreatic cancer.

Family History and Risk of Pancreatic Cancer

Family history is a long-recognized risk factor for pancreatic cancer and an important predictor of disease risk. Studies have suggested that approximately 5-10% of pancreatic cancer patients report a close relative with pancreatic cancer[1,21,2]. Most epidemiological studies have demonstrated a 2- to 3-fold increase in risk of pancreatic cancer among individual with affected first-degree relatives (FDRs)[3–12]. However, some studies have shown even higher risk. A Swedish study reported a standardized incidence ratio (SIR) for pancreatic cancer of 1.73 (95%CI: 1.13-2.54) in offspring with at least one parent affected with pancreatic ductal adenocarcinoma (PDAC)[13]. In a prospective study from the National Familial Pancreas Tumour Registry (NFPTR), the SIRs for pancreatic cancer in comparison to the SEER (Surveillance,

Epidemiology, and End Results) rates were 6.4 (95% CI: 1.8-16.4) and 32.0 (95% CI: 10.2-74.7) in individuals with two and three FDRs with pancreatic cancer[14].

Familial clustering was considered the first evidence supporting the genetic predisposition to pancreatic cancer. Reports of multiple siblings in one generation and individuals in three consecutive generations affected by pancreatic cancer are strong evidence of a hereditary form of the disease following Mendelian inheritance patterns[15–19]. This was later supported by segregation analyses, which favored a major gene model that was predicted to follow an autosomal dominant pattern of a rare allele[20]. As demonstrated in the observational epidemiologic studies, individuals with a family history of pancreatic cancer are at an increased risk of developing the disease themselves. In addition, a population-based twin study in Europe has estimated the heritability for pancreatic cancer to be 36% (95% CI: 0-53%)[21]. As more evidence of the genetic basis of pancreatic cancer has emerged, an operational definition of familial pancreatic cancer (FPC) was proposed to facilitate investigations of the inherited components of the disease. FPC is defined as kindreds with at least a pair of FDRs diagnosed with PDAC. Comparing to the general population, FPC kindred members have a 7- to 9-fold increased risk of pancreatic cancer[14,22]. Risk is even higher among members of FPC kindreds with a young-onset case (< 50 years; SIR = 9.31, 95% CI: 3.42-20.28) than those without (SIR = 6.34, 95% CI: 4.02-9.51)[22].

Both prospective and retrospective studies have found increased risks of other cancers in relatives of pancreatic cancer patients, particularly breast cancer[23,24], melanoma[24], ovarian[25], and colorectal cancer[24,26]. Study from NFPTR reported an increased risk of dying from cancer of the breast (weighted standardized mortality ratio (wSMR) = 1.66, 95%CI: 1.15-2.34), ovarian (wSMR=2.05, 95%CI: 1.10-3.49), bile duct (wSMR = 2.89, 95%CI: 1.04-6.39) and bladder

(wSMR = 1.90, 95%CI: 1.00-3.30) in FDRs of FPC probands[27]. Elevated mortality of colon cancer (wSMR = 2.31, 95%CI: 1.30-3.81) and prostate cancer (wSMR = 2.31, 95%CI: 1.14-4.20) were also observed among the relatives of young-onset (< 50 years old) pancreatic cancer probands[27]. These findings suggest a shared genetic etiology between pancreatic cancer and several other cancers, and the potential benefits of surveillance at-risk relatives for cancers with established screening guidelines.

Pancreatic Cancer Genes

Pancreatic cancer is more prevalent in families with several hereditary syndromes, for which the predisposing genes have been identified, including *BRCA1* and *BRCA2* associated with hereditary breast and ovarian cancer (HBOC), *STK11* associated with Peutz-Jeghers syndrome (PJS), *CDKN2A/p16* associated with familial atypical multiple mole melanoma (FAMMM), mismatch repair (MMR) genes associated with Lynch syndrome, and *PRSS1* associated with hereditary pancreatitis (HP). These genetic syndromes are reported to be associated with a substantially higher risk of pancreatic cancer. The recent discovery of germline mutations in *PALB2* and *ATM* genes in FPC kindreds has extended the list of established high- and moderate-risk pancreatic cancer genes (Table 1).

BRCA1/2

BRCA1 and *BRCA2* gene are well-known high-penetrant predisposing genes for hereditary breast and/or ovarian cancer (HBOC). These genes are involved in the DNA damage response and DNA double-strand breaks repair. Pancreatic cancer is the third most common cancer associated with *BRCA1*/2 mutations, though the penetrance at age 70 is much lower[28–31]. The prevalence of *BRCA2* mutations in pancreatic cancer patients is 1.4-8.2% for patients unselected for family history[32–36], about 6-10% among FPC patients[37–42], and up to 17.2% in

families with 3 or more pancreatic cancers[38,41,43]. Comparing to the general population, the risk of pancreatic cancer is about 2-6 fold in *BRCA2* carriers[28,44] and 2-5 fold in *BRCA1* carriers [29,30,44]. Several studies had reported a higher risk of pancreatic cancer in *BRCA2* carriers than in *BRCA1* carriers[28,31,44–46]. *BRCA1/2* mutation carriers are at particularly high risk (SIR = 4-10) for early onset pancreatic cancer[28–30,44]. FDRs of *BRCA1/2* carriers, regardless of their carrier status, have a significantly higher risk of pancreatic cancer than the general population[44,46].

PALB2

PALB2 gene is a tumor suppressor that interacts closely with both *BRCA1* and *BRCA2* during double-strand DNA repair. Mutations of *PALB2* had previously been associated with familial breast cancer[47]. Jones *et al.* first reported the discovery of truncating mutations of *PALB2* gene in four FPC probands from the NFPTR [48]. Since then, pathogenic mutations of *PALB2* gene have been found in 0.4-4% FPC families, majority of which were families with history of both pancreatic cancer and breast/ovarian cancer[49–55].

ATM

ATM is a breast cancer susceptibility gene that coordinates the DNA double-strand breaks repair. Deleterious mutations of *ATM* gene were first reported by Roberts and his colleagues in two FPC families with at least three members affected by pancreatic cancer[56]. In the subsequent analysis, four additional *ATM* mutations were found in 166 FPC patients compared to none in 190 spouse controls[56]. To date, *ATM* mutations are found in 1-5% patients with pancreatic cancer[35,36,42,57–59].

STK11

Peutz-Jeghers syndrome (PJS) is caused by germline mutations in the *STK11* gene[60–62]. PJS patients are at very high risk of developing cancer during their lifetimes, particularly

gastrointestinal cancer and gynecological cancer. The cumulative risk of developing any gastrointestinal cancer is 38-66% at age 70[63]. Compared to the general population, PJS patients have a 76-140 fold elevated risk for pancreatic cancer[64–66]. The cumulative risk of developing pancreatic cancer at age 70 in PJS patients is 11-55%[64–67].

CDKN2A

CDKN2A is a tumor suppressor gene that is considered a major cause of familial melanoma. In melanoma-prone families of European ancestry, pancreatic cancer is the second most common type of cancers associated with *CDKN2A* mutations. Longitudinal studies in these families have found a 15- to 80-fold increased risk of pancreatic cancer in carriers of *CDKN2A* mutations comparing to the general population[68–73]. The risk of developing pancreatic cancer is also higher in FDRs of carriers than in FDRs of non-carriers (RR = 7.4 [95%CI: 2.3-18.7])[74].

Mismatch repair genes

Hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch Syndrome, accounts for 2-5% of all colorectal cancer. It is caused by inactivating mutations of DNA mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6* and *PMS2*. While several studies found no increase in risk of pancreatic cancer in Lynch Syndrome patients[75–78], others reported an approximately 7- to 10-fold elevated risk of developing pancreatic cancer in carriers of MMR gene mutations[79–82]. The relative risk of pancreatic cancer is higher at younger age[79,80]. The cumulative risk of developing pancreatic cancer at age 70 among mutation carriers was estimated to be 3.68% (95% CI: 1.45% - 5.88%) [80].

PRSS1

Hereditary pancreatitis (HP) is autosomal-dominant disorder characterized by recurrent episodes of acute pancreatitis in childhood and frequent progression to chronic pancreatitis.

Germline mutations in *PRSS1* are responsible for the majority of HP cases. Comparing to the general population, the risk of pancreatic cancer is about 69-fold higher in HP patients, and the median age of cancer onset was at least 15 years earlier [83–86]. About 20-50% HP patients would develop pancreatic cancer at age 70[83–85,87]. The risk is even higher among smokers with HP who tend to develop pancreatic cancer 20 years before non-smokers[85,88].

Germline Mutations in Sporadic Cases

Inherited genetic alterations are not restricted to patients with FDRs affected by pancreatic cancer. Mutations of pancreatic cancer genes are also found in non-FPC patients with or without family history. Three germline mutations in *BRCA2* was found in 41 (7.3%) sporadic pancreatic cancer patients without family history[32]. Of the 14 *BRCA* mutation carriers identified from 306 consecutive incident PDAC patients, only two had a family history of PDAC[33]. Salo-Mullen *et al.* reported a 7.4% prevalence of *BRCA* mutations in 27 PDAC patients of Ashkenazi Jewish ancestry without a family history of breast, ovarian or pancreatic cancer[34]. Studies have reported 0-3% sporadic or unselected pancreatic cancer patients carrying *PALB2* mutations, leading to an aggregated prevalence of 0.75%[55]. In an unselected group of 854 pancreatic cancer patients, twelve of them had germline *BRCA2* mutations, ten with *ATM*, three with *BRCA1*, and two with *PALB2*, the majority of whom had no FDR affected by the disease[36]. Larger scale studies are currently underway to evaluate the mutation prevalence in apparently sporadic pancreatic cancer and expanding genetic testing beyond the current guidelines.

Targeted Therapy

Understanding genetic predisposition of pancreatic cancer has important implication for the development and translation of targeted therapies. Tumors with mutations in *BRCA1/2*, *PALB2*, and *ATM* are highly sensitive to DNA-damaging related treatments such as crosslinking

agents[89–91] and poly(ADP-ribose) polymerase inhibitors (PARPi)[92–95]. Preclinical studies have demonstrated improved sensitivity to chemotherapeutic agents and ionizing radiation in pancreatic cancer cells treated by these agents[96–98]. The clinical benefits of using crosslinking agents and PARPi in patients with pancreatic cancer are currently being investigated. Preliminary results have shown promising efficacy of these agents, particularly in patients with *BRCA2*-associated pancreatic cancer[99–102].

Common Low-Risk Susceptibility Loci

Genome-wide association studies (GWAS) allow for the unbiased evaluation of common genomic variants associated with pancreatic cancer. To identify common susceptibility variants, six pancreatic cancer GWAS have been conducted by the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (PanC4) in populations of European ancestry, including PanScan I in 2009[103], PanScan II in 2010[104], PanScan III in 2014[105], PanC4 in 2015[106], an imputation analysis of PanScan I-III[107] and a recent meta-analysis of PanScan I-III and PanC4. A total of 21 pancreatic cancer susceptibility loci located in 18 genomic regions have been discovered in European populations (Table 2).

PanScan I

PanScan I was a two-stage GWAS including 1,896 patients with incidence pancreatic cancer and 1,939 controls in the discovery stage, as well as 2,457 cases and 2,654 controls in the replication stage. The most significant variant (rs505922) on chromosome 9q34.2 was mapped to the first intron of *ABO* blood group gene [103]. The association of ABO loci with pancreatic cancer has been robustly replicated in studies of European[105,106,108,109] and Asian populations[110–112]. These findings are consistent with the observation that individuals with blood group O had a lower risk of pancreatic cancer than those with groups A or B. About 17% to 19.5% of all

pancreatic cancers in populations of European descent was attributable to the inheritance of a non-O blood group[113,114].

PanScan II

From 3,851 pancreatic cancer cases and 3,934 controls, PanScan II identified three novel genomic regions on chromosome 13q22.1 (a large non-genic region), chromosome 1q32.1 (*NR5A2*) and chromosome 5p15.33 (*TERT-CLPTM1L*) to be significantly associated with pancreatic cancer[104]. The locus on chromosome 13q22.1 (rs9543325) was mapped to a large gene desert flanked by the *KLF5* and *KLF12* genes. Both genes encode a transcription factor involved in cell transformation, proliferation, and carcinogenesis. Several studies have reported the overexpression of the *KLF5* gene in pancreatic cancer [115–118]. The *KLF2* gene, on the other hand, was found to be downregulated in PDAC tumor tissues, and its expression may suppress the malignant transformation of PDAC cancer cells through its regulation of beta-catenin/TCF signaling[119].

Two variants on chromosome 1q32.1 are associated with pancreatic cancer independently. The first significant variant (rs3790844) identified in PanScan II is located in the first intron of the *NR5A2* gene[104]. Imputation analysis of PanScan I-III detected the second variant in the upstream of *NR5A2* (rs2816938). *NR5A2* encodes the nuclear receptor subfamily 5 group A member 2, a transcription factor that activates or inhibits transcription of specific target genes. Overexpression of NR5A2 was observed in resected PDAC tumor tissues and was associated with reduced survival time in PDAC patients[120]. Heterozygous Nr5a2 mice exhibit increased rates of pancreatic acinar to ductal metaplasia and impaired recovery after chemically induced acute pancreatitis[121,122]. Loss of Nr5a2 accelerates the development of oncogenesis

driven by Kras[121,122]. These findings suggest a tumor suppressor role of *NR5A2* that protects pancreas from KRAS driven pre-neoplastic changes.

Four independent pancreatic cancer risk loci have now been identified in the multi-cancer *TERT-CLPTM1L* region on chromosome 5p15.33. The first pancreatic cancer risk locus identified in PanScan II is located in the intron 13 of *CLPTM1L* (rs401681). PanScan III reported a second independent risk locus on chromosome 5p15.33, tagged by a synonymous variant within the second exon of *TERT* (rs2736098) [105]. A third independent risk locus located in the first intron of *TERT* gene (rs2853677) was discovered through a candidate gene analysis in 5,550 pancreatic cancer cases and 7,585 control subjects from PANDoRA (PANcreatic Disease ReseArch) consortium and PanScan [123]. Recently, imputation of PanScan I-III and replication in PANDoRA and PanC4 found a fourth risk locus for pancreatic cancer in this genomic region (rs35226131), which is located about 200bps upstream of the transcriptional start site of *TERT* [107]. The chromosome 5p15.33 region contains two plausible candidate genes: *TERT*, which encodes the catalytic subunit of telomerase reverse transcriptase and *CLPTM1L*, which encodes the cleft lip and palate-associated transmembrane 1 like protein. TERT is a component of the protein and RNA complex that maintains telomere ends. Mutations in TERT promoter region were frequent in multiple tumor types and were correlated with increased TERT expression and telomerase activation[124]. Common variants in the *TERT* region were associated with leukocyte telomere length in patients with breast and ovarian cancer[125]. A recent study had reported an association between the minor allele of rs401681 and shorter telomere length in pancreatic cancer patients, which was consistent with the observation that telomere shortening occurs as an early event in pancreatic tumorigenesis[126–128]. Overexpression of *CLPTM1L* gene are observed in lung and pancreatic cancer tissues[129–131]. CLPTM1L protects tumor cells from genotoxic

apoptosis and is required for Ras-induced oncogenic transformation[129,130,132]. It's also found that overexpression of CLPTM1 may lead to an abrogation of normal cytokinesis and promote cell proliferation in pancreatic cancer cells[131].

PanScan III

The PanScan III study population combined 1,582 newly genotyped pancreatic cancer cases and 5,203 control subjects with PanScan I cohort and sought replication in 2,576 cases and 6,662 controls from the PANDoRA consortium. Four new risk loci for pancreatic cancer was identified on chromosome 7q23.2 (*LINC-PINT*), 16q23.1 (*BCAR1*), 13q12.2 (*PDX1*) and 22q12.1 (*ZNRF3*). The signal on 7q32.3 was marked by an intronic variant (rs6971499) in *LINC-PINT*, a long intergenic p53-induced non-protein coding RNA located between *MKLN1* (Muskelin 1) and *KLF14* (Kruppel-like factor 14). Muskelin is an intracellular protein that mediates cell adhesive and cytoskeletal responses to the extracellular matrix[133]. KLF14 is a member of the Kruppel-like family of transcription factors that may act as a suppressor of KRAS-mediated cell growth through regulation of the cyclin A promoter[134]. Loss of KLF14 may also trigger centrosome amplification, aneuploidy and spontaneous tumorigenesis[135]. KLF14 has also been associated with several metabolic phenotypes including type-2 diabetes mellitus (T2DM), a known risk factor for pancreatic cancer[136–139].

A synonymous variant residing in the last exon of *BCAR1* was noted on 16q23.1 (rs7190458). Breast cancer anti-estrogen resistance 1 (BCAR1), also known as p130Cas is a member of the Cas (Crk-associated substrate) family of adaptor proteins with important regulatory roles in migration, cell cycle control and apoptosis[140]. Altered expression and activity of p130Cas is known to promote metastasis and drug resistance in multiple cancers[140,141]. In addition, two chymotrypsinogen gene, *CTRB1* and *CTRB2* are also located closely to the

detected signal. As important members of a family of serine proteases secreted by the pancreas into the gastrointestinal tract[142], these two genes are plausible target for susceptibility variants at this locus. The detected signal for pancreatic cancer is also in proximity of a susceptibility locus (rs7202877) for type-I and type-II diabetes[143,144] that was found to impair beta-cell function[145] and influences expression of CTRB1/2 in pancreas tissues[146].

The top ranked variant on chromosome 13q12.2 (rs9581943) is located in the promoter region of the *PDX1* (pancreatic and duodenal homeobox1 protein 1) gene. Pathway analysis of GWAS data identified *PDX1*, along with *NR5A2*, *HNF1A*, and *HNF4G*, as important genes for pancreatic development[147]. The protein encoded by *PDX1* is a transcriptional activator of several genes. It is essential in the early development of pancreas[148], and plays a major role in beta-cell function and glucose-dependent regulation of insulin gene expression. Heterozygous mutations in *PDX1* resulted in impaired glucose tolerance and symptoms of diabetes as seen in maturity-onset diabetes of the young type 4 (MODY4) and late-onset T2DM[149–151].

The signal on chromosome 22q2.1 (rs16986825) maps to an intron in *ZNRF3* (zinc and ring finger 3), which encodes a cell surface transmembrane E3 uniquitin protein ligase that is a negative regulator of the WNT signaling pathway[152]. Additionally, a low-penetrance breast cancer gene, *CHEK2,* is also located in proximity to the detected signal. This gene encodes a cell-cycle checkpoint kinase that cooperates with p53, BRCA1 and ATM and regulates cell division in response to DNA damage[153]. Germline mutations and variants of *CHEK2* had been implicated in susceptibility to several cancer types[154–158], including familial pancreatic cancer[41,42,159].

PanC4

PanC4 was conducted on 9,925 pancreatic cancer cases and 11,569 controls, pooling 4,164 newly genotyped cases and 3,792 controls from nine studies in the PanC4 consortium with PanSan I and II cohorts in the gene discovery stage, and analyzed an independent set of 2,497 cases and 4,611 controls from the PANDoRA consortium in the replication stage. Not only this study replicated all previously identified risk loci for pancreatic cancer in European populations, three novel associated signals were also detected on chromosome 17q25.1 (*LINC00673*), 7p13 (*SUGCT*) and 3q29 (*TP63*)[106]. Significant association was also found on chromosome 2q13.3 (*ETAA1*), a region with prior suggestive evidence in the Han Chinese[160].

The top variant on 17q25.1 (rs11655237) maps to *LINC00673* (long inter-genic non-protein coding RNA 673). This association was also observed in a Han Chinese population (rs11655237, OR = 1.26, $P = 3.95\times10^{-14}$)[161]. Through its epigenetic regulation of gene expression, *LINKC00673* may function as an oncogene in several types of cancers. Overexpression of LINC00673 promotes tumor proliferation, invasion and metastasis in non-small-cell lung cancer [162–164] and tongue squamous cell carcinoma[165], and was correlated with poor prognosis in breast cancer[166]. In contrast, expression of LINC00673 was significantly lower in PDAC cancer cells than in normal cells and tissues and overexpression of LINC00673 in the PDAC cell line substantially reduced the rate of cell proliferation. It was found that the single-nucleotide change at rs11655237 creates a miR-1231 binding site, which diminishes the effect of LINC00673 in an allele-specific manner and thus confer susceptibility to pancreatic tumorigenesis [161].

PanC4 reported a significant association on 7p13 with an intronic variant (rs17688601) of the *SUGCT* (succinyl-CoA:glutarate-CoA transferase) gene. Mutations of this gene cause a benign form of glutaric aciduria (glutaric aciduria type III), a rare metabolic abnormality

characterized by persistent isolated accumulation or excretion of glutaric acid [167]. The role of this gene in pancreatic cancer risk is unclear.

Two strongly correlated intronic variants of *TP63* (tumor protein p63) were found to be associated with pancreatic cancer in PanC4 (top variants rs9864771). Protein encoded by *TP63* (p51/p63) is a p53 homologue with pleiotropic functions including cell proliferation, survival, apoptosis, differentiation, senescence, and aging. Frequent overexpression of p63 was observed in resected PDAC tissues[168]. It was suggested that different isoforms of p63 have opposite effects. While TAp63 induces cell death and cell cycle arrest with tumor suppressor features[169], DNp63 as the predominant isoform in pancreatic cancer cell lines, promotes pancreatic cancer growth, motility and invasion[170,171]. As a tumor suppressor, p63 had reduced anti-oncogenetic effects compared with p53 in human cancer cells[172]. However, loss of p63 can cooperate with loss of p53, leading to higher tumor burden and metastasis as seen in genetic mice models[168,171]. It is hypothesized that it is the ratio of TAp63 and DNp63 that determines the biological outcome and chemo-sensitivity.

Imputation Analysis of PanScan I-III

The imputation analysis of the GWAS data in 5,107 cases and 8,845 controls from PanScan I-III had uncovered three new pancreatic cancer signals on chromosome 1q32.1 (*NR5A2*), 8q24.21 (*MYC*), and 5p15.33 (*CLPTM1L-TERT*), all of which are independent from previously reported susceptibility variants[107].

The detected variants on 8q24.21 (rs10094872) is a novel risk loci for pancreatic cancer, independent from the previously reported loci with suggestive evidence in PanScan III (rs1561927). These two variants are both located in the 2 Mb region known to contain multiple susceptibility loci that influence risk of bladder, breast, prostate, colorectal, lung, ovarian,

pancreatic, and renal cancer[173–177]. *MYC* (MYC proto-oncogene, bHLH transcription factor) is the gene located in the closest proximity to the detected variant. Oncogene MYC is a transcription factor that has been implicated in the pathogenesis of one-third of all human malignancies, and may play an important role in KRAS-driven neoplastic transformation in the pancreas[178]. MYC overexpression occurs in up to 42% of advanced PDAC[179,180]. Activation of MYC in adult mice has led to the development of ductal adenocarcinomas with metastasis to the liver[181]. Although evidence have suggested regulatory roles of the 8q24.21 risk loci in the expression of MYC, functional analysis are warranted to allow a deeper understanding of the underlying mechanism[178].

Meta-analysis of PanScan I-III and PanC4

A recent meta-analysis of 9,040 pancreatic cancer patients and 12,496 controls of European ancestry from PanScan and PanC4 has identified a significant association on chromosome 7p12 (*TNS3*). Replications of promising signals in the PANDoRA consortium yield another four genome-wide significant loci on chromosome 1p36.33 (*NOC2L*), 8q21.11 (*HNF4G*), 17q12 (*HNF1B*) and 18q21.32 (*GRP*).

The variant at 1p36.33 (rs13303010) maps to the first intron of the *NOC2L* gene, which encodes the nucleolar complex protein 2 homology, also known as novel INHAT repressor (NIR). NIR is an inhibitor of histone acetyltransferases (HAT) that can bind to p53 at promoters and inhibit p53-mediated gene transactivation by blocking histone acetylation[182]. This protein can also bind to TAp63, a tumor suppressor protein closely related to p53, resulting in the inhibition of TAp63-dependent transactivation of p21[183]. Notably, an intronic variant of the *TP63* gene had previously been identified as a susceptibility locus for pancreatic cancer[106]. Increased expression of *NOC2L* was found in pancreatic cancer cell lines as compared with

histologically normal pancreatic cancer samples[184]. This gene is also among the seventeen genes differentially methylated in breast tumor samples[185].

The locus at 7p12 was marked by an intronic variant (rs73328514) in *TNS3*. This gene encodes Tensin 3, a member of the Tensin family of intracellular proteins that is known as a negative regulator of cell migration and adhesion [186–188] and may play an important role in tumorigenesis and cancer metastasis[186,189,187].

An intronic variant of *HNF4G* on chromosome 8q21.11(rs2941471) was found to be associated with pancreatic cancer. The *HNF4G* gene encodes hepatocyte nuclear factor 4 gamma, a member of nuclear receptor (NRs) that regulate gene expression in a spatiotemporal manner. *HNF4G* is one of the orphan NRs whose endogenous ligands and biological functional are unknown. In addition to this meta-analysis, two pathway analysis of GWAS data had each identified *HNF4G* as a susceptibility factor for pancreatic cancer[147,190]. Common variants of this gene were found to be associated with breast cancer[191]. This gene is also involved in the growth and invasion of bladder cancer[192,193]. In addition, a variant of *HNF4G* correlated with rs2941471 has been significantly associated with variations in serum urate concentrations (rs2941484, r^2 = 0.56, in 1000 Genomes EUR)[194], consistent with a previous report of elevated urate levels as an early event of pancreatic cancer[195].

The signal at 17q12 was located in the fourth intron of *HNF1B*, encoded hepatocyte nuclear factor 1-beta, another member of NRs that is known to regulate the early development of pancreas and cell differentiation[196]. Mutations of this gene cause a specific type of maturity-onset diabetes of the young (MODY5), also known as renal cysts and diabetes syndrome (MIM [137920]). Common variants of this gene are associated with type-2 diabetes mellitus and related phenotypes[197–199], and hence may also contribute to the susceptibility of pancreatic cancer.

Genetic and epigenetic changes of this gene are also associated with risk of endometrial, ovarian and prostate cancer[200–204].

At 18q21.32, the detected association was marked by rs1517037, about ~10kb upstream of the *GRP* gene. This gene encodes a member of the bombesin-like family of gastrin-releasing peptides that stimulate exocrine pancreatic secretion and release various gastrointestinal peptides[205,206]. Studies have suggested a potential role of *GRP* in pancreatic tumor growth[207].

GWAS in Asian Populations

Two GWAS have been conducted in populations of Asian descent (Table 3). The Japanese pancreatic cancer study of 991 cases and 5,209 controls found suggestive associations on chromosome 6q25.3 (*FOXQ1*), 12p11.21 (*BICD1*) and 7q36.2 (*DPP6*) [208]. The second GWAS in a Chinese population of 3,584 pancreatic cancer cases and 4,868 controls (ChinaPC) identified five susceptibility loci on chromosome 21q21.3 (*BACH1*), 21q22.3 (*TFF1*), 10q26.11 (*PRLHR*), 22q13.32 (*FAM19A5*), and 5p13.1 (*DAB2*)[160]. The most significant association identified in ChinaPC was for rs372883, a variant located in the 3' untranslated region (3'UTR) of *BACH1* (BTB domain and CNC homolog 1) gene on chromosome 21q21.3. BACH1 is a transcription factor that belongs to the cap 'n' collar type of basic region leucine zipper factor family (CNC-bZip). Recent studies have demonstrated a critical role of BACH1 in cell migration and metastasis through its regulation of metastasis-related gene expression in breast, colon and prostate cancer [209–211]. The second significant association was detected on chromosome 21q22.3 (rs1547374). This region harbors the trefoil family protein 1 (*TFF1*) gene that encodes secretory proteins expression in gastrointestinal mucosa. Upregulated expression of TFF1 in precursor lesions of PDAC, including pancreatic intraepithelial neoplasia (PanIN), intraductal papillary-mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MSNs), suggests its potential

involvement at the early stage of pancreatic carcinogenesis [212–214]. Recent studies found that reduced expression of TFF1 in the invasion front of human PDAC was associated with lymph node metastasis and poor survival in patients with PDAC[215]. In pancreatic cancers, expression of TFF1 promotes tumorigenesis by suppressing oncogene-induced senescence [216] and is correlated with increase metastasis[217]. An intronic variant (rs2255280) in *DAB2* (clathrin adaptor protein) gene region on 5p13.1 was among the identified susceptibility loci in ChinaPC. Frequent loss expression of DAB2 in human malignant cancer cells suggests its potential role as a tumor suppressor[218]. Overexpression of DAB2 inhibits cell growth, migration and invasion, and was correlated with poor survival in cancer patients[219–222]. Significant association on 10q26.11 was observed for a regulatory variant of *PRLHR* (prolactin releasing hormone receptor) gene (rs12413624). Polymorphisms of this gene was associated with colorectal cancer[223]. A intronic variant of *FAM19A5* (family with sequence similarity 19 member A5) gene on 22q13.32 was also associated with an increased risk of pancreatic cancer (rs5768709). This gene encodes a TAFA protein expressed predominately in brain and may function as brain-specific chemokines or neurokines[224]. Prior to the ChinaPC study, there is no implication of *PRLHR* or *FAM19A5* in the risk of pancreatic cancer and thus their susceptibility role is currently unknown.

Current Status of Screening High-Risk Populations

Screening and early detection of pancreatic cancer offer the best chance of reducing the high mortality rates of this disease. The goal of screening asymptomatic individuals is to identify pancreatic cancer at early stage or, ideally to identify high-grade precancerous lesions that can be resected to prevent the development of cancer. Because of the low incidence of pancreatic cancer in the general population, population level screening will demand a highly specific screening assay. Selective screening of individuals at increased risk for pancreatic cancer is considered

worthwhile. The International Cancer of the Pancreas Screening (CAPS) Consortium recommends screening on FDRs of FPC patients, patients with PJS, and carriers of *CDKN2A/p16*, *BRCA2*, and carriers of MMR gene mutations with ≥1 affected FDR[225] . FDRs of FPC patients represent a group of high-risk individuals that are relatively easy to identify in clinical settings. Of all identified risk factors for pancreatic cancer, PJS confers the greatest risk for the disease, making PJS patients good candidates for pancreatic cancer screening. Among the established pancreatic cancer genes, germline *BRCA2* mutations followed by *ATM* account for the highest percentage of inherited pancreatic cancer[36,42,58,40,34]. It is recommended that *BRCA2* mutation carriers with ≥1 affected FDR and those with two or more affected family members should be considered for screening, particularly Ashkenazi Jewish individuals. In addition, given the substantially higher risk of pancreatic cancer in patients with *CDKN2A/p16* and mismatch repair (MMR) gene mutations, screening is also recommended to those mutation carriers with >1 affected FDRs.

Endoscopic ultrasonography (EUS) and/or MRI/magnetic resonance cholangiopancreatography (MRCP) are recommended by CAPS as initial screening tools. However, the CAPS Consortium could not reach consensus on ages to initiate or stop surveillance, the interval for follow-up imaging, nor on the long-term management of initial abnormal results. Screening and early detection strategies should be accompanied by effective treatment or preventive strategies if they are to produce a significant survival benefit. Given morbidity and mortality associated with pancreatic surgery, there is little consensus about when surgery is required for a pancreatic lesion in asymptomatic high-risk individuals. Multidisciplinary assessment is however recommended to make individualized decision of the necessity of surgical intervention. The lack of consensus on many aspects of pancreatic cancer

screening underscores the need for more research to fill the knowledge gap and to make evidence-based decisions.

Summary and Future Directions

Pancreatic cancer is a rare and deadly disease with the highest case-fatality rate of any major cancer. Due to the lack of effective means for prevention, diagnosis and treatment, pancreatic cancer remains a major public health challenge. Family history, cigarette smoking, chronic pancreatitis, and diabetes are well-established risk factors for pancreatic cancer. Pancreatic cancer is fundamentally a genetic disease caused by both inherited and acquired genetic mutations. Family-based heritability analysis reported 36% of pancreatic cancer was due to genetics. Familial pancreatic cancer kindreds and patients affected by certain genetic syndromes, for example HP, PJS, HBOC, FAMMM, and HNPCC, are at particularly high risk of pancreatic cancer. About 15-20% of FPC are caused by germline mutations in one of the established pancreatic cancer gene (*BRCA2, ATM, PALB2, PRSS1*, S*TK11*, *BRCA1*, *CDKN2A*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*). The genetic basis of susceptibility underlying the majority of FPC cases, however, remains unexplained. To date, GWAS of pancreatic cancer have discovered 16 low-risk susceptibility loci in European populations and 5 in Asian populations, many of which had strong biological plausibility. Together, these GWAS loci explained < 5% of pancreatic cancer. Screening, surveillance and management guidelines for genetically high-risk individuals are currently evolving. DNA-damaging related agents are promising in treating pancreatic cancer caused by mutations in *BRCA1*, *BRCA2*, *PALB2 or ATM* genes. Identification of disease-causing genes can aid in the characterization of individuals at highest genetically defined risk in which effective prevention approach can be developed.

Figures and Tables

Table 1 Pancreatic Cancer Predisposing Genes

Genes	Predisposition Syndromes	Risk of Pancreatic Cancer*
BRCA2	Hereditary breast and/or ovarian cancer	SIR = 2.20 – 5.90
BRCA1		SIR = 1.60 – 4.73
PALB2	Familial breast cancer	Increased
ATM		Increased
STK11	Peutz-Jeghers syndrome	SIR = 76.2 – 139.7
PRSS1	Hereditary Pancreatitis	SIR = 53 – 87
CDKN2A	Familial atypical multiple mole and melanoma syndrome	SIR = 14.8 – 80.8
Mismatch repair genes (*MLH1*, *MSH2*, *MHS6* and *PMS2*)	Hereditary non-polyposis colorectal cancer	No effect up to SIR = 10.68

* SIR = standardized incidence ratio

Table 2 Genetic Susceptibility Loci for Pancreatic Cancer Identified in European Populations

Chr*	Nearest Gene(s)	Top SNP†	Minor/Major Alleles	Consequence Type	Allelic OR [95% CI]‡	P-value
1p36.33	*NOC2L*	rs13303010	G/A	intron	1.26 [1.19-1.35]	8.36×10^{-14}
1q32.1	*NR5A2*	rs3790844	T/C	intron	0.77 [0.71-0.84]	2.45×10^{-10}
1q32.1	*NR5A2*	rs2816938	A/T	upstream	1.20 [1.15-1.25]	4.88×10^{-15}
2p13.3	*ETAA1*	rs1486134	G/T	downstream	1.14 [1.09-1.19]	3.36×10^{-9}
3q29	*TP63*	rs9854771	A/G	intron	0.89 [0.85-0.93]	2.35×10^{-8}
5q15.33	*TERT, CLPTM1L*	rs401681	C/T	intron	1.19 [1.11-1.27]	3.66×10^{-7}
5p15.33	*TERT, CLPTM1L*	rs2736098	T/C	synonymous	0.80 [0.76-0.85]	9.78×10^{-14}
5p15.33	*TERT, CLPTM1L*	rs35226131	T/C	upstream	0.71 [0.63-0.80]	1.70×10^{-8}
7p12	*TNS3*	rs73328514	T/A	intron	0.85 [0.80-0.90]	1.35×10^{-7}
7p13	*SUGCT*	rs17688601	A/C	intron	0.88 [0.84-0.92]	1.41×10^{-8}
7q32.3	*LINC-PINT*	rs6971499	C/T	intron	0.79 [0.74-0.84]	2.98×10^{-12}
8q21.11	*HNF4G*	rs2941471	G/A	intron	0.89 [0.85-0.93]	6.60×10^{-10}
8q24.21	*MYC*	rs10094872	T/A	intron	1.15 [1.10-1.20]	3.22×10^{-9}
9q34	*ABO*	rs505922	T/C	intron	1.20 [1.12-1.28] (het) 1.44 [1.26-1.63] (hom)	5.37×10^{-8}
13q12.2	*PDX1*	rs9581943	A/G	upstream	1.15 [1.10-1.20]	2.35×10^{-9}
13q22.1	none	rs9543325	T/C	intergenic	1.26 [1.18-1.35]	3.27×10^{-11}

16q23.1	*BCAR1, CTRB1, CTRB2*	rs7190458	A/G	synonymous	1.46 [1.30-1.65]	1.13×10^{-10}
17q12	*HNF1B*	rs4795218	A/G	intron	0.88 [0.84-0.92]	1.32×10^{-8}
17q25.1	*LINC00673*	rs11655237	T/C	non-coding transcript exon	1.26 [1.19-1.34]	1.42×10^{-14}
18q21.32	*GRP*	rs1517037	T/C	upstream	0.86 [0.80-0.91]	3.28×10^{-8}
22q12.1	*ZNRF3*	rs16986825	T/C	intron	1.18 [1.12-1.25]	1.18×10^{-8}

* Chromosomal location in NCBI genome build 37.

† SNP = single nucleotide polymorphisms

‡ OR = odds ratio; CI = confidence interval; het = heterozygotes; hom = homozygotes

Table 3 Genetic Susceptibility Loci for Pancreatic Cancer Identified in Asian Populations

Chr*	Nearest Gene(s)	Top SNP†	Minor/Major Alleles	Consequence Type	Allelic OR [95% CI]‡	P-value
5p13.1	*DAB2*	rs2255280	C/A	intron	0.81 [0.76-0.87]	4.18×10^{-10}
10q26.11	*RPLHR*	rs12413624	T/A	regulatory region	1.23 [1.16-1.31]	5.12×10^{-11}
21q21.3	*BACH1*	rs372883	C/T	3' UTR	0.79 [0.75-0.84]	2.24×10^{-13}
21q22.3	*TFF1*	rs1547374	G/A	downstream	0.79 [0.74-0.84]	3.71×10^{-13}
22q13.32	*FAM19A5*	rs5768709	G/A	intron	1.25 [1.17-1.34]	1.41×10^{-10}

* Chromosomal location in NCBI genome build 37.

† SNP = single nucleotide polymorphisms

‡ OR = odds ratio; CI = confidence interval; het = heterozygotes; hom = homozygotes

Reference

1. Hruban RH, Petersen GM, Ha PK, Kern SE. Genetics of pancreatic cancer. From genes to families. Surg Oncol Clin N Am. 1998 Jan;7(1):1–23.

2. Klein AP. Identifying people at a high risk of developing pancreatic cancer. Nat Rev Cancer. 2013 Jan;13(1):66.

3. Fernandez E, La Vecchia C, D'Avanzo B, Negri E, Franceschi S. Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 1994 May;3(3):209–12.

4. Silverman DT, Schiffman M, Everhart J, Goldstein A, Lillemoe KD, Swanson GM, et al. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer. 1999 Aug;80(11):1830–7.

5. Schenk M, Schwartz AG, O'Neal E, Kinnard M, Greenson JK, Fryzek JP, et al. Familial risk of pancreatic cancer. J Natl Cancer Inst. 2001 Apr 18;93(8):640–4.

6. Fehringer G, Gallinger S, Borgida A, Zhang LR, Adams L, Liu G, et al. The association of family history of cancer and medical history with pancreatic cancer risk. Pancreas. 2014 Jul;43(5):812–4.

7. Schulte A, Pandeya N, Fawcett J, Fritschi L, Klein K, Risch HA, et al. Association between family cancer history and risk of pancreatic cancer. Cancer Epidemiol. 2016 Dec;45:145–50.

8. Zheng Z, Zheng R, He Y, Sun X, Wang N, Chen T, et al. Risk Factors for Pancreatic Cancer in China: A Multicenter Case-Control Study. J Epidemiol. 2016;26(2):64–70.

9. Jacobs EJ, Rodriguez C, Newton CC, Bain EB, Patel AV, Feigelson HS, et al. Family history of various cancers and pancreatic cancer mortality in a large cohort. Cancer Causes Control CCC. 2009 Oct;20(8):1261–9.

10. Permuth-Wey J, Egan KM. Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. Fam Cancer. 2009;8(2):109–17.

11. Jacobs EJ, Chanock SJ, Fuchs CS, Lacroix A, McWilliams RR, Steplowski E, et al. Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Int J Cancer J Int Cancer. 2010 Sep 1;127(6):1421–8.

12. Austin MA, Kuo E, Van Den Eeden SK, Mandelson MT, Brentnall TA, Kamineni A, et al. Family history of diabetes and pancreatic cancer as risk factors for pancreatic cancer: the PACIFIC study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2013 Oct;22(10):1913–7.

13. Hemminki K, Li X. Familial and second primary pancreatic cancers: a nationwide epidemiologic study from Sweden. Int J Cancer. 2003 Feb 10;103(4):525–30.

14. Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJA, et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res. 2004 Apr 1;64(7):2634–8.

15. Dat NM, Sontag SJ. Pancreatic carcinoma in brothers. Ann Intern Med. 1982 Aug;97(2):282.

16. Ghadirian P, Simard A, Baillargeon J. Cancer of the pancreas in two brothers and one sister. Int J Pancreatol Off J Int Assoc Pancreatol. 1987 Dec;2(5–6):383–91.

17. MacDermott RP, Kramer P. Adenocarcinoma of the pancreas in four siblings. Gastroenterology. 1973 Jul;65(1):137–9.

18. Friedman JM, Fialkow PJ. Familial carcinoma of the pancreas. Clin Genet. 1976 May;9(5):463–9.

19. Ehrenthal D, Haeger L, Griffin T, Compton C. Familial pancreatic adenocarcinoma in three generations. A case report and a review of the literature. Cancer. 1987 May 1;59(9):1661–4.

20. Klein AP, Beaty TH, Bailey-Wilson JE, Brune KA, Hruban RH, Petersen GM. Evidence for a major gene influencing risk of pancreatic cancer. Genet Epidemiol. 2002 Aug;23(2):133–49.

21. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000 Jul 13;343(2):78–85.

22. Brune KA, Lau B, Palmisano E, Canto M, Goggins MG, Hruban RH, et al. Importance of age of onset in pancreatic cancer kindreds. J Natl Cancer Inst. 2010 Jan 20;102(2):119–26.

23. Bermejo JL, Hemminki K. Familial association of histology specific breast cancers with cancers at other sites. Int J Cancer. 2004 Apr 10;109(3):430–5.

24. Hiripi E, Bermejo JL, Li X, Sundquist J, Hemminki K. Familial association of pancreatic cancer with other malignancies in Swedish families. Br J Cancer. 2009 Nov;101(10):1792.

25. Cote ML, Schenk M, Schwartz AG, Vigneau FD, Kinnard M, Greenson JK, et al. Risk of other cancers in individuals with a family history of pancreas cancer. J Gastrointest Cancer. 2007;38(2–4):119–26.

26. Hemminki K, Li X. Familial liver and gall bladder cancer: a nationwide epidemiological study from Sweden. Gut. 2003 Apr;52(4):592–6.

27. Wang L, Brune KA, Visvanathan K, Laheru D, Herman J, Wolfgang C, et al. Elevated Cancer Mortality in the Relatives of Patients with Pancreatic Cancer. Cancer Epidemiol Prev Biomark. 2009 Nov 1;18(11):2829–34.

28. Thompson D, Easton DF, Breast Cancer Linkage Consortium. Cancer Incidence in BRCA1 mutation carriers. J Natl Cancer Inst. 2002 Sep 18;94(18):1358–65.

29. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst. 1999 Aug 4;91(15):1310–6.

30. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, Hoogerbrugge N, Verhoef S, Vasen HFA, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. J Med Genet. 2005 Sep;42(9):711–9.

31. Risch HA, McLaughlin JR, Cole DEC, Rosen B, Bradley L, Fan I, et al. Population BRCA1 and BRCA2 mutation frequencies and cancer penetrances: a kin-cohort study in Ontario, Canada. J Natl Cancer Inst. 2006 Dec 6;98(23):1694–706.

32. Goggins M, Schutte M, Lu J, Moskaluk CA, Weinstein CL, Petersen GM, et al. Germline BRCA2 gene mutations in patients with apparently sporadic pancreatic carcinomas. Cancer Res. 1996 Dec 1;56(23):5360–4.

33. Holter S, Borgida A, Dodd A, Grant R, Semotiuk K, Hedley D, et al. Germline BRCA Mutations in a Large Clinic-Based Cohort of Patients With Pancreatic Adenocarcinoma. J Clin Oncol Off J Am Soc Clin Oncol. 2015 Oct 1;33(28):3124–9.

34. Salo-Mullen EE, O'Reilly EM, Kelsen DP, Ashraf AM, Lowery MA, Yu KH, et al. Identification of germline genetic mutations in patients with pancreatic cancer. Cancer. 2015 Dec 15;121(24):4382–8.

35. Hu C, Hart SN, Bamlet WR, Moore RM, Nandakumar K, Eckloff BW, et al. Prevalence of Pathogenic Mutations in Cancer Predisposition Genes among Pancreatic Cancer Patients. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2016 Jan;25(1):207–11.

36. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious Germline Mutations in Patients With Apparently Sporadic Pancreatic Adenocarcinoma. J Clin Oncol Off J Am Soc Clin Oncol. 2017 Oct 20;35(30):3382–90.

37. Hahn SA, Greenhalf B, Ellis I, Sina-Frey M, Rieder H, Korte B, et al. BRCA2 germline mutations in familial pancreatic carcinoma. J Natl Cancer Inst. 2003 Feb 5;95(3):214–21.

38. Couch FJ, Johnson MR, Rabe KG, Brune K, de Andrade M, Goggins M, et al. The prevalence of BRCA2 mutations in familial pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2007 Feb;16(2):342–6.

39. Slater EP, Langer P, Fendrich V, Habbe N, Chaloupka B, Matthäi E, et al. Prevalence of BRCA2 and CDKN2a mutations in German familial pancreatic cancer families. Fam Cancer. 2010 Sep;9(3):335–43.

40. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med Off J Am Coll Med Genet. 2015 Jul;17(7):569–77.

41. Catts ZA-K, Baig MK, Milewski B, Keywan C, Guarino M, Petrelli N. Statewide Retrospective Review of Familial Pancreatic Cancer in Delaware, and Frequency of Genetic Mutations in Pancreatic Cancer Kindreds. Ann Surg Oncol. 2016 May;23(5):1729–35.

42. Chaffee KG, Oberg AL, McWilliams RR, Majithia N, Allen BA, Kidd J, et al. Prevalence of germ-line mutations in cancer genes among pancreatic cancer patients with a positive family history. Genet Med Off J Am Coll Med Genet. 2017 Jul 20;

43. Murphy KM, Brune KA, Griffin C, Sollenberger JE, Petersen GM, Bansal R, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res. 2002 Jul 1;62(13):3789–93.

44. Mocci E, Milne RL, Méndez-Villamil EY, Hopper JL, John EM, Andrulis IL, et al. Risk of pancreatic cancer in breast cancer families from the breast cancer family registry. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2013 May;22(5):803–11.

45. Mersch J, Jackson MA, Park M, Nebgen D, Peterson SK, Singletary C, et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer. 2015 Jan 15;121(2):269–75.

46. Streff H, Profato J, Ye Y, Nebgen D, Peterson SK, Singletary C, et al. Cancer Incidence in First- and Second-Degree Relatives of BRCA1 and BRCA2 Mutation Carriers. The Oncologist. 2016 Jul;21(7):869–74.

47. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 2007 Feb;39(2):165–7.

48. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science. 2009 Apr 10;324(5924):217.

49. Tischkowitz MD, Sabbaghian N, Hamel N, Borgida A, Rosner C, Taherian N, et al. Analysis of the gene coding for the BRCA2-interacting protein PALB2 in familial and sporadic pancreatic cancer. Gastroenterology. 2009 Sep;137(3):1183–6.

50. Slater EP, Langer P, Niemczyk E, Strauch K, Butler J, Habbe N, et al. PALB2 mutations in European familial pancreatic cancer families. Clin Genet. 2010 Nov;78(5):490–4.

51. Hofstatter EW, Domchek SM, Miron A, Garber J, Wang M, Componeschi K, et al. PALB2 mutations in familial breast and pancreatic cancer. Fam Cancer. 2011 Jun;10(2):225–31.

52. Peterlongo P, Catucci I, Pasquini G, Verderio P, Peissel B, Barile M, et al. PALB2 germline mutations in familial breast cancer cases with personal and family history of pancreatic cancer. Breast Cancer Res Treat. 2011 Apr;126(3):825–8.

53. Stadler ZK, Salo-Mullen E, Sabbaghian N, Simon JA, Zhang L, Olson SH, et al. Germline PALB2 mutation analysis in breast-pancreas cancer families. J Med Genet. 2011 Aug;48(8):523–5.

54. Ghiorzo P, Pensotti V, Fornarini G, Sciallero S, Battistuzzi L, Belli F, et al. Contribution of germline mutations in the BRCA and PALB2 genes to pancreatic cancer in Italy. Fam Cancer. 2012 Mar;11(1):41–7.

55. Borecka M, Zemankova P, Vocka M, Soucek P, Soukupova J, Kleiblova P, et al. Mutation analysis of the PALB2 gene in unselected pancreatic cancer patients in the Czech Republic. Cancer Genet. 2016 May;209(5):199–204.

56. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, et al. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2012 Jan;2(1):41–6.

57. Grant RC, Selander I, Connor AA, Selvarajah S, Borgida A, Briollais L, et al. Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. Gastroenterology. 2015 Mar;148(3):556–64.

58. Takai E, Yachida S, Shimizu K, Furuse J, Kubo E, Ohmoto A, et al. Germline mutations in Japanese familial pancreatic cancer patients. Oncotarget. 2016 Nov 8;7(45):74227–35.

59. Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, et al. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. Cancer Discov. 2015 Dec 9;

60. Hemminki A, Tomlinson I, Markie D, Järvinen H, Sistonen P, Björkqvist AM, et al. Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. Nat Genet. 1997 Jan;15(1):87–90.

61. Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature. 1998 Jan 8;391(6663):184–7.

62. Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 1998 Jan;18(1):38–43.

63. van Lier MGF, Wagner A, Mathus-Vliegen EMH, Kuipers EJ, Steyerberg EW, van Leerdam ME. High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol. 2010 Jun;105(6):1258–64; author reply 1265.

64. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology. 2000 Dec;119(6):1447–53.

65. Korsse SE, Harinck F, van Lier MGF, Biermann K, Offerhaus GJA, Krak N, et al. Pancreatic cancer risk in Peutz-Jeghers syndrome patients: a large cohort study and implications for surveillance. J Med Genet. 2013 Jan;50(1):59–64.

66. Resta N, Pierannunzio D, Lenato GM, Stella A, Capocaccia R, Bagnulo R, et al. Cancer risk associated with STK11/LKB1 germline mutations in Peutz-Jeghers syndrome patients: results of an Italian multicenter study. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver. 2013 Jul;45(7):606–11.

67. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJP, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res Off J Am Assoc Cancer Res. 2006 May 15;12(10):3209–15.

68. Goldstein AM, Fraser MC, Struewing JP, Hussussian CJ, Ranade K, Zametkin DP, et al. Increased risk of pancreatic cancer in melanoma-prone kindreds with p16INK4 mutations. N Engl J Med. 1995 Oct 12;333(15):970–4.

69. Ghiorzo P, Ciotti P, Mantelli M, Heouaine A, Queirolo P, Rainero ML, et al. Characterization of ligurian melanoma families and risk of occurrence of other neoplasia. Int J Cancer. 1999 Nov 12;83(4):441–8.

70. Borg A, Sandberg T, Nilsson K, Johannsson O, Klinker M, Måsbäck A, et al. High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families. J Natl Cancer Inst. 2000 Aug 2;92(15):1260–6.

71. Goldstein AM, Struewing JP, Fraser MC, Smith MW, Tucker MA. Prospective risk of cancer in CDKN2A germline mutation carriers. J Med Genet. 2004 Jun;41(6):421–4.

72. de Snoo FA, Bishop DT, Bergman W, van Leeuwen I, van der Drift C, van Nieuwpoort FA, et al. Increased risk of cancer other than melanoma in CDKN2A founder mutation (p16-Leiden)-positive melanoma families. Clin Cancer Res Off J Am Assoc Cancer Res. 2008 Nov 1;14(21):7151–7.

73. Potjer TP, Kranenburg HE, Bergman W, de Vos tot Nederveen Cappel WH, van Monsjou HS, Barge-Schaapveld DQCM, et al. Prospective risk of cancer and the influence of tobacco use in carriers of the p16-Leiden germline variant. Eur J Hum Genet EJHG. 2015 May;23(5):711–4.

74. Mukherjee B, Delancey JO, Raskin L, Everett J, Jeter J, Begg CB, et al. Risk of non-melanoma cancers in first-degree relatives of CDKN2A mutation carriers. J Natl Cancer Inst. 2012 Jun 20;104(12):953–6.

75. Watson P, Lynch HT. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer. 1993 Feb 1;71(3):677–85.

76. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer. 1999 Apr 12;81(2):214–8.

77. Pande M, Wei C, Chen J, Amos CI, Lynch PM, Lu KH, et al. Cancer spectrum in DNA mismatch repair gene mutation carriers: results from a hospital based Lynch syndrome registry. Fam Cancer. 2012 Sep;11(3):441–7.

78. Samadder NJ, Smith KR, Wong J, Thomas A, Hanson H, Boucher K, et al. Cancer Risk in Families Fulfilling the Amsterdam Criteria for Lynch Syndrome. JAMA Oncol. 2017 Aug 3;

79. Geary J, Sasieni P, Houlston R, Izatt L, Eeles R, Payne SJ, et al. Gene-related cancer spectrum in families with hereditary non-polyposis colorectal cancer (HNPCC). Fam Cancer. 2008;7(2):163–72.

80. Kastrinos F, Mukherjee B, Tayob N, Wang F, Sparr J, Raymond VM, et al. Risk of pancreatic cancer in families with Lynch syndrome. JAMA. 2009 Oct 28;302(16):1790–5.

81. Win AK, Young JP, Lindor NM, Tucker KM, Ahnen DJ, Young GP, et al. Colorectal and other cancer risks for carriers and noncarriers from families with a DNA mismatch repair gene mutation: a prospective cohort study. J Clin Oncol Off J Am Soc Clin Oncol. 2012 Mar 20;30(9):958–64.

82. Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. Gut. 2017 Jul 28;

83. Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK, Perrault J, et al. Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. J Natl Cancer Inst. 1997 Mar 19;89(6):442–6.

84. Howes N, Lerch MM, Greenhalf W, Stocken DD, Ellis I, Simon P, et al. Clinical and genetic characteristics of hereditary pancreatitis in Europe. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2004 Mar;2(3):252–61.

85. Rebours V, Boutron-Ruault M-C, Schnee M, Férec C, Maire F, Hammel P, et al. Risk of pancreatic adenocarcinoma in patients with hereditary pancreatitis: a national exhaustive series. Am J Gastroenterol. 2008 Jan;103(1):111–9.

86. Raimondi S, Lowenfels AB, Morselli-Labate AM, Maisonneuve P, Pezzilli R. Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. Best Pract Res Clin Gastroenterol. 2010 Jun;24(3):349–58.

87. Masamune A, Kikuta K, Hamada S, Nakano E, Kume K, Inui A, et al. Nationwide survey of hereditary pancreatitis in Japan. J Gastroenterol. 2017 Aug 31;

88. Lowenfels AB, Maisonneuve P, Whitcomb DC, Lerch MM, DiMagno EP. Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis. JAMA. 2001 Jul 11;286(2):169–70.

89. van der Heijden MS, Brody JR, Gallmeier E, Cunningham SC, Dezentje DA, Shen D, et al. Functional defects in the fanconi anemia pathway in pancreatic cancer cells. Am J Pathol. 2004 Aug;165(2):651–7.

90. van der Heijden MS, Brody JR, Dezentje DA, Gallmeier E, Cunningham SC, Swartz MJ, et al. In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. Clin Cancer Res Off J Am Assoc Cancer Res. 2005 Oct 15;11(20):7508–15.

91. Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, et al. Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Mol Cancer Ther. 2011 Jan;10(1):3–8.

92. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005 Apr 14;434(7035):913–7.

93. McCabe N, Lord CJ, Tutt ANJ, Martin NMB, Smith GCM, Ashworth A. BRCA2-deficient CAPAN-1 cells are extremely sensitive to the inhibition of Poly (ADP-Ribose) polymerase: an issue of potency. Cancer Biol Ther. 2005 Sep;4(9):934–6.

94. Tutt A, Robson M, Garber JE, Domchek SM, Audeh MW, Weitzel JN, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. Lancet Lond Engl. 2010 Jul 24;376(9737):235–44.

95. Perkhofer L, Schmitt A, Romero Carrasco MC, Ihle M, Hampp S, Ruess DA, et al. ATM Deficiency Generating Genomic Instability Sensitizes Pancreatic Ductal Adenocarcinoma Cells to Therapy-Induced DNA Damage. Cancer Res. 2017 Oct 15;77(20):5576–90.

96. Jacob DA, Bahra M, Langrehr JM, Boas-Knoop S, Stefaniak R, Davis J, et al. Combination therapy of poly (ADP-ribose) polymerase inhibitor 3-aminobenzamide and gemcitabine shows strong antitumor activity in pancreatic cancer cells. J Gastroenterol Hepatol. 2007 May;22(5):738–48.

97. Chen S, Wang G, Niu X, Zhao J, Tan W, Wang H, et al. Combination of AZD2281 (Olaparib) and GX15-070 (Obatoclax) results in synergistic antitumor activities in preclinical models of pancreatic cancer. Cancer Lett. 2014 Jun 28;348(1–2):20–8.

98. Porcelli L, Quatrale AE, Mantuano P, Leo MG, Silvestris N, Rolland JF, et al. Optimize radiochemotherapy in pancreatic cancer: PARP inhibitors a new therapeutic opportunity. Mol Oncol. 2013 Jun;7(3):308–22.

99. Pishvaian MJ, Wang H, Zhuang T, He AR, Hwang JJ, Hankin A, et al. A phase I/II study of ABT-888 in combination with 5-fluorouracil (5-FU) and oxaliplatin (Ox) in patients with metastatic pancreatic cancer (MPC). J Clin Oncol. 2013 Feb 1;31(4_suppl):147–147.

100. O'Reilly EM, Lowery MA, Segal MF, Smith SC, Moore MJ, Kindler HL, et al. Phase IB trial of cisplatin (C), gemcitabine (G), and veliparib (V) in patients with known or potential BRCA or PALB2-mutated pancreas adenocarcinoma (PC). J Clin Oncol. 2014 May 20;32(15_suppl):4023–4023.

101. Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmaña J, et al. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. J Clin Oncol Off J Am Soc Clin Oncol. 2015 Jan 20;33(3):244–50.

102. Olaparib in gBRCA Mutated Pancreatic Cancer Whose Disease Has Not Progressed on First Line Platinum-Based Chemotherapy - Full Text View - ClinicalTrials.gov [Internet]. [cited 2017 Dec 2]. Available from: https://clinicaltrials.gov/ct2/show/NCT02184195

103. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009 Sep;41(9):986–90.

104. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224–8.

105. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat Genet. 2014 Sep;46(9):994–1000.

106. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet. 2015 Aug;47(8):911–6.

107. Zhang M, Wang Z, Obazee O, Jia J, Childs EJ, Hoskins J, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget. 2016 Oct 11;7(41):66328–43.

108. Willis JA, Olson SH, Orlow I, Mukherjee S, McWilliams RR, Kurtz RC, et al. A replication study and genome-wide scan of single-nucleotide polymorphisms associated with

pancreatic cancer risk and overall survival. Clin Cancer Res Off J Am Assoc Cancer Res. 2012 Jul 15;18(14):3942–51.

109. Rizzato C, Campa D, Pezzilli R, Soucek P, Greenhalf W, Capurso G, et al. ABO blood groups and pancreatic cancer risk and survival: results from the PANcreatic Disease ReseArch (PANDoRA) consortium. Oncol Rep. 2013 Apr;29(4):1637–44.

110. Risch HA, Lu L, Wang J, Zhang W, Ni Q, Gao Y-T, et al. ABO Blood Group and Risk of Pancreatic Cancer: A Study in Shanghai and Meta-Analysis. Am J Epidemiol. 2013 Jun 15;177(12):1326–37.

111. Xu H-L, Cheng J-R, Zhang W, Wang J, Yu H, Ni Q-X, et al. Re-evaluation of ABO gene polymorphisms detected in a genome-wide association study and risk of pancreatic ductal adenocarcinoma in a Chinese population. Chin J Cancer. 2014 Feb;33(2):68–73.

112. Nakao M, Matsuo K, Hosono S, Ogata S, Ito H, Watanabe M, et al. ABO blood group alleles and the risk of pancreatic cancer in a Japanese population. Cancer Sci. 2011 May 1;102(5):1076–80.

113. Wolpin BM, Chan AT, Hartge P, Chanock SJ, Kraft P, Hunter DJ, et al. ABO blood group and the risk of pancreatic cancer. J Natl Cancer Inst. 2009 Mar 18;101(6):424–31.

114. Wolpin BM, Kraft P, Gross M, Helzlsouer K, Bueno-de-Mesquita HB, Steplowski E, et al. Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. Cancer Res. 2010 Feb 1;70(3):1015–23.

115. Mori A, Moser C, Lang SA, Hackl C, Gottfried E, Kreutz M, et al. Up-regulation of Krüppel-like factor 5 in pancreatic cancer is promoted by interleukin-1beta signaling and hypoxia-inducible factor-1alpha. Mol Cancer Res MCR. 2009 Aug;7(8):1390–8.

116. Diaferia GR, Balestrieri C, Prosperini E, Nicoli P, Spaggiari P, Zerbi A, et al. Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. EMBO J. 2016 Mar 15;35(6):595–617.

117. Abjalimov IR, Zinovyeva MV, Nikolaev LG, Kopantzeva MR, Kopantzev EP, Sverdlov ED. Expression of transcription factor genes in cell lines corresponding to different stages of pancreatic cancer progression. Dokl Biochem Biophys. 2017 Jul 1;475(1):267–70.

118. Kondratyeva LG, Chernov IP, Zinovyeva MV, Kopantzev EP, Sverdlov ED. Expression of master regulatory genes of embryonic development in pancreatic tumors. Dokl Biochem Biophys. 2017 Jul 1;475(1):250–2.

119. Zhang D, Dai Y, Cai Y, Suo T, Liu H, Wang Y, et al. KLF2 is downregulated in pancreatic ductal adenocarcinoma and inhibits the growth and migration of cancer cells. Tumor Biol. 2016 Mar 1;37(3):3425–31.

120. Luo Z, Li Y, Zuo M, Liu C, Yan D, Wang H, et al. Effect of NR5A2 inhibition on pancreatic cancer stem cell (CSC) properties and epithelial-mesenchymal transition (EMT) markers. Mol Carcinog. 2017 May 1;56(5):1438–48.

121. Flandez M, Cendrowski J, Cañamero M, Salas A, Pozo N del, Schoonjans K, et al. Nr5a2 heterozygosity sensitises to, and cooperates with, inflammation in KRasG12V-driven pancreatic tumourigenesis. Gut. 2014 Apr 1;63(4):647–55.

122. Figura G von, Morris JP, Wright CVE, Hebrok M. Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. Gut. 2014 Apr 1;63(4):656–64.

123. Campa D, Rizzato C, Stolzenberg-Solomon R, Pacetti P, Vodicka P, Cleary SP, et al. TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. Int J Cancer. 2015 Nov 1;137(9):2175–83.

124. Huang D-S, Wang Z, He X-J, Diplas BH, Yang R, Killela PJ, et al. Recurrent TERT promoter mutations identified in a large-scale study of multiple tumour types are associated with increased TERT expression and telomerase activation. Eur J Cancer Oxf Engl 1990. 2015 May;51(8):969–76.

125. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013 Apr;45(4):371–84, 384e1-2.

126. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, DePinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev. 2006 May 15;20(10):1218–49.

127. van Heek NT, Meeker AK, Kern SE, Yeo CJ, Lillemoe KD, Cameron JL, et al. Telomere Shortening Is Nearly Universal in Pancreatic Intraepithelial Neoplasia. Am J Pathol. 2002 Nov 1;161(5):1541–7.

128. Bao Y, Prescott J, Yuan C, Zhang M, Kraft P, Babic A, et al. Leucocyte telomere length, genetic variants at the TERT gene region and risk of pancreatic cancer. Gut. 2017 Jun 1;66(6):1116–22.

129. James MA, Wen W, Wang Y, Byers LA, Heymach JV, Coombes KR, et al. Functional Characterization of CLPTM1L as a Lung Cancer Risk Candidate Gene in the 5p15.33 Locus. PLOS ONE. 2012 Jun 4;7(6):e36116.

130. Ni Z, Tao K, Chen G, Chen Q, Tang J, Luo X, et al. CLPTM1L Is Overexpressed in Lung Cancer and Associated with Apoptosis. PLOS ONE. 2012 Dec 26;7(12):e52598.

131. Jia J, Bosley AD, Thompson A, Hoskins JW, Cheuk A, Collins I, et al. CLPTM1L Promotes Growth and Enhances Aneuploidy in Pancreatic Cancer Cells. Cancer Res. 2014 May 15;74(10):2785–95.

132. James MA, Vikis HG, Tate E, Rymaszewski AL, You M. CRR9/CLPTM1L Regulates Cell Survival Signaling and Is Required for Ras Transformation and Lung Tumorigenesis. Cancer Res. 2014 Feb 15;74(4):1116–27.

133. Adams JC, Seed B, Lawler J. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. EMBO J. 1998 Sep 1;17(17):4964–74.

134. Fernandez-Zapico ME, Lomberk GA, Tsuji S, DeMars CJ, Bardsley MR, Lin Y-H, et al. A functional family-wide screening of SP/KLF proteins identifies a subset of suppressors of KRAS-mediated cell growth. Biochem J. 2011 Apr 15;435(2):529–37.

135. Fan G, Sun L, Shan P, Zhang X, Huan J, Zhang X, et al. Loss of KLF14 triggers centrosome amplification and tumorigenesis. Nat Commun. 2015 Oct 6;6:8450.

136. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010 Jul;42(7):579–89.

137. Wang J, Zhang J, Shen J, Hu D, Yan G, Liu X, et al. Association of KCNQ1 and KLF14 polymorphisms and risk of type 2 diabetes mellitus: A global meta-analysis. Hum Immunol. 2014 Apr;75(4):342–7.

138. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010 Aug 5;466(7307):707–13.

139. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 2011 Jun;43(6):561–4.

140. Barrett A, Pellet-Many C, Zachary IC, Evans IM, Frankel P. p130Cas: a key signalling node in health and disease. Cell Signal. 2013 Apr;25(4):766–77.

141. Nikonova AS, Gaponova AV, Kudinov AE, Golemis EA. CAS proteins in health and disease: An update. IUBMB Life. 2014 Jun 1;66(6):387–95.

142. Whitcomb DC, Lowe ME. Human pancreatic digestive enzymes. Dig Dis Sci. 2007 Jan;52(1):1–17.

143. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009 Jun;41(6):703–7.

144. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012 Sep;44(9):981–90.

145. Harder MN, Ribel-Madsen R, Justesen JM, Sparsø T, Andersson EA, Grarup N, et al. Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased β-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. J Clin Endocrinol Metab. 2013 Apr;98(4):E801-806.

146. 't Hart LM, Fritsche A, Nijpels G, van Leeuwen N, Donnelly LA, Dekker JM, et al. The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. Diabetes. 2013 Sep;62(9):3275–81.

147. Li D, Duell EJ, Yu K, Risch HA, Olson SH, Kooperberg C, et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. Carcinogenesis. 2012 Jul;33(7):1384–90.

148. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Dev Camb Engl. 1996 Mar;122(3):983–95.

149. Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev. 1998 Jun 15;12(12):1763–8.

150. McKinnon CM, Docherty K. Pancreatic duodenal homeobox-1, PDX-1, a major regulator of beta cell identity and function. Diabetologia. 2001 Oct;44(10):1203–14.

151. Melloul D, Tsur A, Zangen D. Pancreatic Duodenal Homeobox (PDX-1) in health and disease. J Pediatr Endocrinol Metab JPEM. 2002 Dec;15(9):1461–72.

152. Hao H-X, Xie Y, Zhang Y, Charlat O, Oster E, Avello M, et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. Nature. 2012 Apr 29;485(7397):195–200.

153. Magni M, Ruscica V, Buscemi G, Kim J-E, Nachimuthu BT, Fontanella E, et al. Chk2 and REGγ-dependent DBC1 regulation in DNA damage induced apoptosis. Nucleic Acids Res. 2014 Dec 1;42(21):13150–60.

154. Nevanlinna H, Bartek J. The CHEK2 gene and inherited breast cancer susceptibility. Oncogene. 2006 Sep 25;25(43):5912–9.

155. Siołek M, Cybulski C, Gąsior-Perczak D, Kowalik A, Kozak-Klonowska B, Kowalska A, et al. CHEK2 mutations and the risk of papillary thyroid cancer. Int J Cancer. 2015 Aug 1;137(3):548–52.

156. Lawrenson K, Iversen ES, Tyrer J, Weber RP, Concannon P, Hazelett DJ, et al. Common variants at the CHEK2 gene locus and risk of epithelial ovarian cancer. Carcinogenesis. 2015 Nov;36(11):1341–53.

157. Havranek O, Kleiblova P, Hojny J, Lhota F, Soucek P, Trneny M, et al. Association of Germline CHEK2 Gene Variants with Risk and Prognosis of Non-Hodgkin Lymphoma. PloS One. 2015;10(10):e0140819.

158. Wang Y, Dai B, Ye D. CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis. Int J Clin Exp Med. 2015;8(9):15708–15.

159. Lener MR, Kashyap A, Kluźniak W, Cybulski C, Soluch A, Pietrzak S, et al. The Prevalence of Founder Mutations among Individuals from Families with Familial Pancreatic Cancer Syndrome. Cancer Res Treat Off J Korean Cancer Assoc. 2017 Apr;49(2):430–6.

160. Wu C, Miao X, Huang L, Che X, Jiang G, Yu D, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. Nat Genet. 2011 Dec 11;44(1):62–6.

161. Zheng J, Huang X, Tan W, Yu D, Du Z, Chang J, et al. Pancreatic cancer risk variant in *LINC00673* creates a miR-1231 binding site and interferes with PTPN11 degradation. Nat Genet. 2016 May 23;48(7):ng.3568.

162. Shi X, Ma C, Zhu Q, Yuan D, Sun M, Gu X, et al. Upregulation of long intergenic noncoding RNA 00673 promotes tumor proliferation via LSD1 interaction and repression of NCALD in non-small-cell lung cancer. Oncotarget. 2016 May 3;7(18):25558–75.

163. Ma C, Wu G, Zhu Q, Liu H, Yao Y, Yuan D, et al. Long intergenic noncoding RNA 00673 promotes non-small-cell lung cancer metastasis by binding with EZH2 and causing epigenetic silencing of HOXA5. Oncotarget. 2017 May 16;8(20):32696–705.

164. Lu W, Zhang H, Niu Y, Wu Y, Sun W, Li H, et al. Long non-coding RNA linc00673 regulated non-small cell lung cancer proliferation, migration, invasion and epithelial mesenchymal transition by sponging miR-150-5p. Mol Cancer. 2017 Jul 11;16(1):118.

165. Yu J, Liu Y, Gong Z, Zhang S, Guo C, Li X, et al. Overexpression long non-coding RNA LINC00673 is associated with poor prognosis and promotes invasion and metastasis in tongue squamous cell carcinoma. Oncotarget. 2017 Mar 7;8(10):16621–32.

166. Abdul-Rahman U, Győrffy B, Adams BD. linc00673 (ERRLR01) is a prognostic indicator of overall survival in breast cancer. Transcription. 2017 Aug 10;1–13.

167. Sherman EA, Strauss KA, Tortorelli S, Bennett MJ, Knerr I, Morton DH, et al. Genetic mapping of glutaric aciduria, type 3, to chromosome 7 and identification of mutations in c7orf10. Am J Hum Genet. 2008 Nov;83(5):604–9.

168. Ito Y, Takeda T, Wakasa K-I, Tsujimoto M, Sakon M, Matsuura N. Expression of p73 and p63 proteins in pancreatic adenocarcinoma: p73 overexpression is inversely correlated with biological aggressiveness. Int J Mol Med. 2001 Jul 1;8(1):67–71.

169. Melino G. p63 is a suppressor of tumorigenesis and metastasis interacting with mutant p53. Cell Death Differ. 2011 Sep;18(9):1487–99.

170. Danilov AV, Neupane D, Nagaraja AS, Feofanova EV, Humphries LA, DiRenzo J, et al. DeltaNp63alpha-mediated induction of epidermal growth factor receptor promotes pancreatic cancer cell growth and chemoresistance. PloS One. 2011;6(10):e26815.

171. Flores ER, Sengupta S, Miller JB, Newman JJ, Bronson R, Crowley D, et al. Tumor predisposition in mice mutant for p63 and p73: evidence for broader tumor suppressor functions for the p53 family. Cancer Cell. 2005 Apr;7(4):363–73.

172. Kunisaki R, Ikawa S, Maeda T, Nakazaki Y, Kurita R, Harata M, et al. p51/p63, a novel p53 homologue, potentiates p53 activity and is a human cancer gene therapy candidate. J Gene Med. 2006 Sep;8(9):1121–30.

173. Park SL, Chang S-C, Cai L, Cordon-Cardo C, Ding B-G, Greenland S, et al. Associations between variants of the 8q24 chromosome and nine smoking-related cancer sites. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2008 Nov;17(11):3193–202.

174. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet. 2010 Oct;42(10):874–9.

175. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Petursdottir V, Hardarson S, et al. A common variant at 8q24.21 is associated with renal cell cancer. Nat Commun. 2013;4:2776.

176. Sur I, Tuupanen S, Whitington T, Aaltonen LA, Taipale J. Lessons from functional analysis of genome-wide association studies. Cancer Res. 2013 Jul 15;73(14):4180–4.

177. Tanskanen T, van den Berg L, Välimäki N, Aavikko M, Ness-Jensen E, Hveem K, et al. Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. Int J Cancer. 2017 Sep 28;

178. Hessmann E, Schneider G, Ellenrieder V, Siveke JT. MYC in pancreatic cancer: novel mechanistic insights and their translation into therapeutic strategies. Oncogene. 2016 Mar 31;35(13):1609–18.

179. Schleger C, Verbeke C, Hildenbrand R, Zentgraf H, Bleyl U. c-MYC activation in primary and metastatic ductal adenocarcinoma of the pancreas: incidence, mechanisms, and clinical significance. Mod Pathol Off J U S Can Acad Pathol Inc. 2002 Apr;15(4):462–9.

180. Birnbaum DJ, Adélaïde J, Mamessier E, Finetti P, Lagarde A, Monges G, et al. Genome profiling of pancreatic adenocarcinoma. Genes Chromosomes Cancer. 2011 Jun;50(6):456–65.

181. Lin W, Rajbhandari N, Liu C, Sakamoto K, Zhang Q, Triplett AA, et al. Dormant cancer cells contribute to residual disease in a model of reversible pancreatic cancer. Cancer Res. 2013 Mar 15;73(6):1821–30.

182. Hublitz P, Kunowska N, Mayer UP, Müller JM, Heyne K, Yin N, et al. NIR is a novel INHAT repressor that modulates the transcriptional activity of p53. Genes Dev. 2005 Dec 1;19(23):2912–24.

183. Heyne K, Willnecker V, Schneider J, Conrad M, Raulf N, Schüle R, et al. NIR, an inhibitor of histone acetyltransferases, regulates transcription factor TAp63 and is controlled by the cell cycle. Nucleic Acids Res. 2010 Jun;38(10):3159–71.

184. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun. 2018 Feb 8;9(1):556.

185. Titus AJ, Way GP, Johnson KC, Christensen BC. Deconvolution of DNA methylation identifies differentially methylated gene regions on 1p36 across breast cancer subtypes. Sci Rep. 2017 Sep 14;7(1):11594.

186. Martuszewska D, Ljungberg B, Johansson M, Landberg G, Oslakovic C, Dahlbäck B, et al. Tensin3 is a negative regulator of cell migration and all four Tensin family members are downregulated in human kidney cancer. PloS One. 2009;4(2):e4350.

187. Cao X, Voss C, Zhao B, Kaneko T, Li SS-C. Differential regulation of the activity of deleted in liver cancer 1 (DLC1) by tensins controls cell migration and transformation. Proc Natl Acad Sci U S A. 2012 Jan 31;109(5):1455–60.

188. Blangy A. Tensins are versatile regulators of Rho GTPase signalling and cell adhesion. Biol Cell. 2017 Mar;109(3):115–26.

189. Qian X, Li G, Vass WC, Papageorge A, Walker RC, Asnaghi L, et al. The Tensin-3 protein, including its SH2 domain, is phosphorylated by Src and contributes to tumorigenesis and metastasis. Cancer Cell. 2009 Sep 8;16(3):246–58.

190. Lee YH, Gyu Song G. Genome-wide pathway analysis in pancreatic cancer. J BUON Off J Balk Union Oncol. 2015 Dec;20(6):1565–75.

191. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013 Apr;45(4):353–61, 361e1-2.

192. Okegawa T, Ushio K, Imai M, Morimoto M, Hara T. Orphan nuclear receptor HNF4G promotes bladder cancer growth and invasion through the regulation of the hyaluronan synthase 2 gene. Oncogenesis. 2013 Jul 29;2:e58.

193. Sun H, Tian J, Xian W, Xie T, Yang X. miR-34a inhibits proliferation and invasion of bladder cancer cells by targeting orphan nuclear receptor HNF4G. Dis Markers. 2015;2015:879254.

194. Köttgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet. 2013 Feb;45(2):145–54.

195. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med. 2014 Oct;20(10):1193–8.

196. Naqvi AAT, Hasan GM, Hassan MI. Investigating the role of transcription factors of pancreas development in pancreatic cancer. Pancreatol Off J Int Assoc Pancreatol IAP Al. 2018 Mar;18(2):184–90.

197. Bonnycastle LL, Willer CJ, Conneely KN, Jackson AU, Burrill CP, Watanabe RM, et al. Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. Diabetes. 2006 Sep;55(9):2534–40.

198. Gupta V, Vinay DG, Rafiq S, Kranthikumar MV, Janipalli CS, Giambartolomei C, et al. Association analysis of 31 common polymorphisms with type 2 diabetes and its related traits in Indian sib pairs. Diabetologia. 2012 Feb;55(2):349–57.

199. Goda N, Murase H, Kasezawa N, Goda T, Yamakawa-Kobayashi K. Polymorphism in microRNA-binding site in HNF1B influences the susceptibility of type 2 diabetes mellitus: a population based case-control study. BMC Med Genet. 2015 Sep 2;16:75.

200. Berndt SI, Sampson J, Yeager M, Jacobs KB, Wang Z, Hutchinson A, et al. Large-scale fine mapping of the HNF1B locus and prostate cancer risk. Hum Mol Genet. 2011 Aug 15;20(16):3322–9.

201. Kim HJ, Bae JS, Lee J, Chang IH, Kim KD, Shin HD, et al. HNF1B polymorphism associated with development of prostate cancer in Korean patients. Urology. 2011 Oct;78(4):969.e1-6.

202. Setiawan VW, Haessler J, Schumacher F, Cote ML, Deelman E, Fesinmeyer MD, et al. HNF1B and endometrial cancer risk: results from the PAGE study. PloS One. 2012;7(1):e30390.

203. Painter JN, O'Mara TA, Batra J, Cheng T, Lose FA, Dennis J, et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. Hum Mol Genet. 2015 Mar 1;24(5):1478–92.

204. Zhao Y, Liang J, Qi JG, Yang N, Wu G, Lin YL, et al. Meta-analysis of the association between the HNF1B rs4430796 (A>G) polymorphism and risk of prostate cancer based on case-control studies. Genet Mol Res GMR. 2015 Jul 3;14(3):7426–35.

205. Jensen RT, Moody T, Pert C, Rivier JE, Gardner JD. Interaction of bombesin and litorin with specific membrane receptors on pancreatic acinar cells. Proc Natl Acad Sci U S A. 1978 Dec;75(12):6139–43.

206. Knuhtsen S, Holst JJ, Jensen SL, Knigge U, Nielsen OV. Gastrin-releasing peptide: effect on exocrine secretion and release from isolated perfused porcine pancreas. Am J Physiol. 1985 Mar;248(3 Pt 1):G281-286.

207. Avis I, Jett M, Kasprzyk PG, Cuttitta F, Treston AM, Maneckjee R, et al. Effect of gastrin-releasing peptide on the pancreatic tumor cell line (Capan). Mol Carcinog. 1993;8(4):214–20.

208. Low S-K, Kuchiba A, Zembutsu H, Saito A, Takahashi A, Kubo M, et al. Genome-wide association study of pancreatic cancer in Japanese population. PloS One. 2010 Jul 29;5(7):e11824.

209. Liang Y, Wu H, Lei R, Chong RA, Wei Y, Lu X, et al. Transcriptional network analysis identifies BACH1 as a master regulator of breast cancer bone metastasis. J Biol Chem. 2012 Sep 28;287(40):33533–44.

210. Davudian S, Shajari N, Kazemi T, Mansoori B, Salehi S, Mohammadi A, et al. BACH1 silencing by siRNA inhibits migration of HT-29 colon cancer cells through reduction of metastasis-related genes. Biomed Pharmacother Biomedecine Pharmacother. 2016 Dec;84:191–8.

211. Shajari N, Davudian S, Kazemi T, Mansoori B, Salehi S, Khaze Shahgoli V, et al. Silencing of BACH1 inhibits invasion and migration of prostate cancer cells by altering metastasis-related gene expression. Artif Cells Nanomedicine Biotechnol. 2017 11;1–10.

212. Terris B, Blaveri E, Crnogorac-Jurcevic T, Jones M, Missiaglia E, Ruszniewski P, et al. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol. 2002 May;160(5):1745–54.

213. Yeh T-S, Jan Y-Y, Chiu C-T, Ho Y-B, Chen T-C, Lee K-F, et al. Characterisation of oestrogen receptor, progesterone receptor, trefoil factor 1, and epidermal growth factor and its receptor in pancreatic cystic neoplasms and pancreatic ductal adenocarcinoma. Gut. 2002 Nov;51(5):712–6.

214. Prasad NB, Biankin AV, Fukushima N, Maitra A, Dhara S, Elkahloun AG, et al. Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. Cancer Res. 2005 Mar 1;65(5):1619–26.

215. Sunagawa M, Yamaguchi J, Kokuryo T, Ebata T, Yokoyama Y, Sugawara G, et al. Trefoil factor family 1 expression in the invasion front is a poor prognostic factor associated with lymph node metastasis in pancreatic cancer. Pancreatol Off J Int Assoc Pancreatol IAP Al. 2017 Oct;17(5):782–7.

216. Radiloff DR, Wakeman TP, Feng J, Schilling S, Seto E, Wang X-F. Trefoil factor 1 acts to suppress senescence induced by oncogene activation during the cellular transformation process. Proc Natl Acad Sci U S A. 2011 Apr 19;108(16):6591–6.

217. Arumugam T, Brandt W, Ramachandran V, Moore TT, Wang H, May FE, et al. Trefoil factor 1 stimulates both pancreatic cancer and stellate cells and increases metastasis. Pancreas. 2011 Aug;40(6):815–22.

218. Zhang Z, Chen Y, Tang J, Xie X. Frequent loss expression of dab2 and promotor hypermethylation in human cancers: a meta-analysis and systematic review. Pak J Med Sci. 2014 Mar;30(2):432–7.

219. Du L, Zhao Z, Ma X, Hsiao T-H, Chen Y, Young E, et al. miR-93-directed downregulation of DAB2 defines a novel oncogenic pathway in lung cancer. Oncogene. 2014 Aug 21;33(34):4307–15.

220. Xie Y, Zhang Y, Jiang L, Zhang M, Chen Z, Liu D, et al. Disabled homolog 2 is required for migration and invasion of prostate cancer cells. Front Med. 2015 Sep;9(3):312–21.

221. Cheong S-M, Choi H, Hong BS, Gho YS, Han J-K. Dab2 is pivotal for endothelial cell migration by mediating VEGF expression in cancer cells. Exp Cell Res. 2012 Mar 10;318(5):550–7.

222. Wang Z, Tseng C-P, Pong R-C, Chen H, McConnell JD, Navone N, et al. The mechanism of growth-inhibitory effect of DOC-2/DAB2 in prostate cancer. Characterization of a novel GTPase-activating protein associated with N-terminal domain of DOC-2/DAB2. J Biol Chem. 2002 Apr 12;277(15):12622–31.

223. Su Q, Wang Y, Zhao J, Ma C, Wu T, Jin T, et al. Polymorphisms of PRLHR and HSPA12A and risk of gastric and colorectal cancer in the Chinese Han population. BMC Gastroenterol. 2015 Aug 25;15:107.

224. Park MY, Kim HS, Lee M, Park B, Lee HY, Cho EB, et al. FAM19A5, a brain-specific chemokine, inhibits RANKL-induced osteoclast formation through formyl peptide receptor 2. Sci Rep. 2017 Nov 14;7(1):15575.

225. Canto MI, Harinck F, Hruban RH, Offerhaus GJ, Poley J-W, Kamel I, et al. International Cancer of the Pancreas Screening (CAPS) Consortium summit on the management of patients with increased risk for familial pancreatic cancer. Gut. 2013 Mar;62(3):339–47.

CHAPTER THREE

HERITABILITY FOR PANCREATIC CANCER FROM IMPUTED VARIANTS

The heritability analysis was conducted in two independent case-control studies: a subset of the Pancreatic Cancer Case-Control Consortium (PanC4) study and a combined sample of familial pancreatic cancer (FPC) patients from the Familial Pancreatic Cancer Genome Sequencing Project (FPCGSP) and participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. This chapter presents the analysis conducted in the PanC4 study using variants imputed from 1000 Genomes Project reference panels, including the estimation of heritability explained by all imputed variants, the genomic partitioning of estimated heritability by chromosomes, minor allele frequency (MAF) categories and functional annotations, as well as the evaluation of identified GWAS loci contributing to the susceptibility of pancreatic cancer.

Study Populations

The Pancreatic Cancer Case Control Consortium (PanC4) was established to improve the understanding of the etiology of pancreatic cancer through joint, or pooled data analysis (http://panc4.org)[1] from established case-control studies of pancreatic cancer. A subset of pancreatic cancer case-control studies with available bio-samples participated in the PanC4 genome-wide association study (GWAS). Participating sites included nine hospital-based or population based case-control studies including the Central Europe study coordinated by the International Agency for Research on Cancer (IARC/Central Europe)[2], Johns Hopkins Hospital (Baltimore, MD)[3,4], Mayo Clinic (Rochester, MN)[5], MD Anderson Cancer Center (Houston, TX)[6], Memorial Sloan-Kettering Cancer Center (New

York, NY)[7], University of Toronto (Toronto ON, Canada)[8], Queensland (Queensland, Australia)[9], University of California San Francisco (San Francisco, CA)[10] and Yale University (New Haven, CT)[11]. Cases were defined as individuals with adenocarcinoma of the pancreas. Control subjects were selected from the general population or hospital catchment area as detailed in

Analysis	**Model#**	**GREML Approach***	**Mixed Linear Model†**	**List of Genetic Components**	**Calculation of h^2**
Estimates of heritability explained by imputed variants	1	GREML-SC	$y = u + g + \varepsilon$	g	$h^2_{imputed} = \sigma^2_g/(\sigma^2_g + \sigma^2_e)$
	2	GREML-MS	$y = u + \sum_{i=1}^{2} g_i + \varepsilon$	$g_{MAF<0.01}$ $g_{MAF\geq0.01}$	$h^2_{imputed} = \frac{\sum_{i=1}^{2}\sigma^2_{g_i}}{\sum_{i=1}^{2}\sigma^2_{g_i} + \sigma^2_e}$
	3	GREML-LDMS	$y = u + \sum_{i=1}^{4} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below}$ $g_{MAF<0.01,LD\ above}$ $g_{MAF\geq0.01,LD\ below}$ $g_{MAF\geq0.01,LD\ above}$	$h^2_{imputed} = \frac{\sum_{i=1}^{4}\sigma^2_{g_i}}{\sum_{i=1}^{4}\sigma^2_{g_i} + \sigma^2_e}$
Genomic partitioning by chromosome	4	GREML-LDMS	$y = u + \sum_{i=1}^{88} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below,chr1}$ $g_{MAF<0.01,LD\ above,chr1}$ $g_{MAF\geq0.01,LD\ below,chr1}$ $g_{MAF\geq0.01,LD\ above,chr1}$ … $g_{MAF<0.01,LD\ below,chr22}$ $g_{MAF<0.01,LD\ above,chr22}$ $g_{MAF\geq0.01,LD\ below,chr22}$ $g_{MAF\geq0.01,LD\ above,chr22}$	$h^2_{chr1} = \frac{\sum_{i=1}^{4}\sigma^2_{g_{i,chr1}}}{\sum_{i=1}^{88}\sigma^2_{g_i} + \sigma^2_e}$ … $h^2_{chr22} = \frac{\sum_{i=1}^{4}\sigma^2_{g_{i,chr22}}}{\sum_{i=1}^{88}\sigma^2_{g_i} + \sigma^2_e}$
Genomic Partitioning by MAF	5	GREML-LDMS	$y = u + \sum_{i=1}^{12} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below}$ $g_{MAF<0.01,LD\ above}$ $g_{MAF\ 0.01-0.10,LD\ below}$ $g_{MAF\ 0.01-0.10,LD\ above}$	$h^2_{MAF<0.01} = \frac{\sum_{i=1}^{2}\sigma^2_{g_{i,MAF<0.01}}}{\sum_{i=1}^{12}\sigma^2_{g_i} + \sigma^2_e}$

				$g_{MAF\ 0.10-0.20,LD\ below}$ $g_{MAF\ 0.10-0.20,LD\ above}$ $g_{MAF\ 0.20-0.30,LD\ below}$ $g_{MAF\ 0.20-0.30,LD\ above}$ $g_{MAF\ 0.30-0.40,LD\ below}$ $g_{MAF\ 0.30-0.40,LD\ above}$ $g_{MAF\ 0.40-0.50,LD\ below}$ $g_{MAF\ 0.40-0.50,LD\ above}$	$h^2_{MAF\ 0.01-0.10} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.01-0.10}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.10-0.20} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.10-0.20}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.20-0.30} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.20-0.30}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.30-0.40} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.30-0.40}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.40-0.50} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.40-0.50}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$
Genomic partitioning by functional annotation	6	GREML-LDMS	$y = u + \sum_{i=1}^{16} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below,coding}$ $g_{MAF<0.01,LD\ above,coding}$ $g_{MAF\geq0.01,LD\ below,coding}$ $g_{MAF\geq0.01,LD\ above,coding}$ $g_{MAF<0.01,LD\ below,intergenic}$ $g_{MAF<0.01,LD\ above,intergenic}$ $g_{MAF\geq0.01,LD\ below,intergenic}$ $g_{MAF\geq0.01,LD\ above,intergenic}$ $g_{MAF<0.01,LD\ below,intronic}$ $g_{MAF<0.01,LD\ above,intronic}$ $g_{MAF\geq0.01,LD\ below,intronic}$ $g_{MAF\geq0.01,LD\ above,intronic}$ $g_{MAF<0.01,LD\ below,regulatory}$ $g_{MAF<0.01,LD\ above,regulatory}$ $g_{MAF\geq0.01,LD\ below,regulatory}$ $g_{MAF\geq0.01,LD\ above,regulatory}$	$h^2_{coding} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,coding}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{intergenic} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,intergenic}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{intronic} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,intronic}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{regulatory} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,regulatory}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$

Evaluation of GWAS loci	7	GREML-LDMS	$y = u + g_{GWAS} + \sum_{i=1}^{4} g_i + \varepsilon$	g_{GWAS} $g_{MAF<0.01,LD\ below,remaining}$ $g_{MAF<0.01,LD\ above,remaining}$ $g_{MAF\geq0.01,LD\ below,remaining}$ $g_{MAF\geq0.01,LD\ above,remaining}$	$h^2_{GWAS} = \frac{\sigma^2_{GWAS}}{\sigma^2_{GWAS} + \sum_{i=1}^{4} \sigma^2_{g_{i,remaining}} + \sigma^2_e}$

* GREML-SC = single-component GREML; GREML-MS = MAF-stratified GREML; GREML-LDMS = LD- and MAF-stratified GREML.

† The disease status in observed scale (y) is a functional of some fixed effects (μ) including intercept, age, sex, and first 20 principal components, one or more genetic components (g_i), and a residual effect (ε).

Supplementary Table 2. This study was reviewed and approved by the Institutional Review Board of the Johns Hopkins University School of Medicine and of each participating institution. Informed consent was obtained from all patients participating in this study.

Genotyping and Imputation

DNA samples from 8,052 PanC4 participants, 180 study duplicates, 176 HapMap control samples, and 26 replicates from the previous pancreatic cancer GWAS PanScan 2[12], were genotyped on the IlluminaHumanOmniExpressExome-8v1 array at the Johns Hopkins Center for Inherited Disease Research (CIDR). Genotypes were called using GenomeStudio version 2011.1, Genotyping Module 1.9.4 and GenTrain version 1.0[1]. Before imputation, individuals with excess allele sharing, duplicates, and subjects with incomplete information on age were removed. Single-nucleotide polymorphisms (SNPs) with the following characteristics were excluded: positional duplicates, more than two discordant calls in study duplicates, technical failures, or missing call rate > 2%, more than one Mendelian error in HapMap control trios, Hardy-Weinberg equilibrium p-value < 10^{-6}, sex difference in allele frequency > 0.2 for autosomes/XY in samples of European ancestry, monomorphic SNPs and SNP on sex chromosomes. Overall 7,956 samples and 821,150 SNPs passed the applied quality control filters and were imputed using IMPUTE v2[13] with the 1000 Genomes (Phase 3, v3)[14] as the reference panel.

Quality Controls on PanC4 After Imputation

Additional QC filters were applied on the imputed PanC4 data to minimize the impact of poorly imputed variants in the heritability analysis (Figure 1). After imputation, there were 7,956 samples and 81,671,345 autosomal variants in the dataset. A total of 372 known non-European samples were removed. The imputation generated three probabilities for genotype AA, AB and BB at each site of a given sample. To facilitate the heritability analysis, PLINK[15] was used to

convert the probabilities to hard genotypes. For genotype calls with high certainty (i.e. one of the probabilities was significantly higher than the other two), PLINK considers the genotype with the highest probability as the hard genotype. For genotype calls with low certainty (defined as the difference between the highest two probabilities is less than 0.1), PLINK sets the genotype to missing. After being converted to hard genotype calls, about 1.9% variants with missingness greater than 5% were excluded. As recommended for heritability analysis, variants that failed HWE exact test at $p < 10^{-6}$ and those with minor allele count less than 5 (equivalent to a minor allele frequency of 0.03%) were excluded to minimize the noise caused by extremely rare variants. More than 60% of the variants were excluded due to being monomorphic after imputation. Because heritability analysis in case-control studies is particularly sensitive to hidden relatedness in the study population, a relatedness cutoff of 0.025 was applied to further exclude 653 distantly related individuals. The genetic relationship was determined using 99,138 common (MAF > 0.05) and independent (pairwise $r^2 > 0.20$) variants directly genotyped in our dataset. The final dataset contains 6,931 samples and 16,112,030 sites, including 68,512 multi-allelic sites. After splitting these multi-allelic variants into multiple records, there were 16,184,129 variants in the final dataset. Minor allele frequencies (MAF) of these variants were calculated in PLINK. Annotation of the variants was obtained from ANNOVAR[16]. The functional predictions were derived from the NCBI Reference Sequence Database.

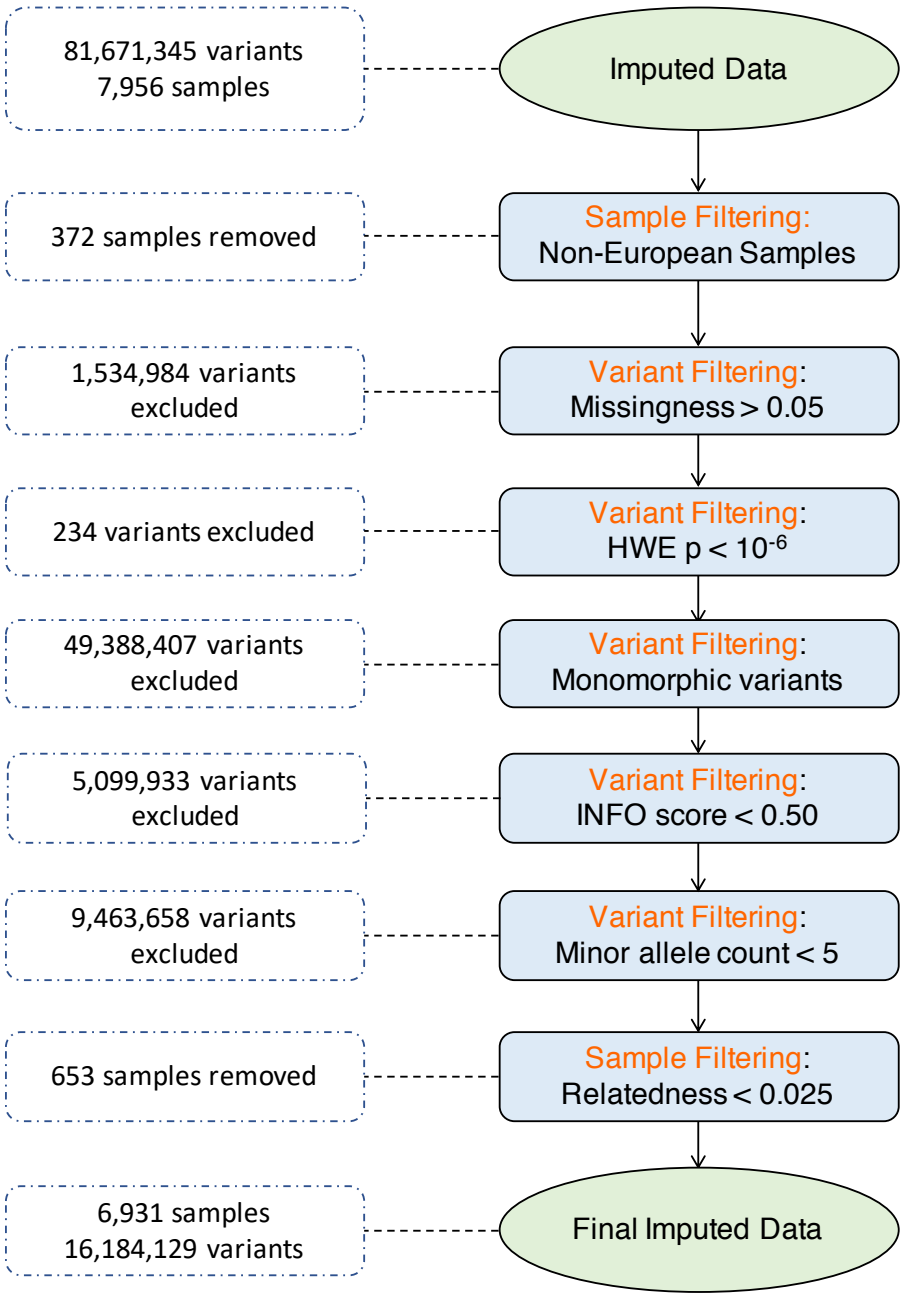


Figure 1 Overview of quality control procedures on PanC4 imputed data. Circles with a solid line represent WGS data at different quality control stage. Boxes with a solid line represent filters applied in an order directed by arrows. Boxed with dashed line detail changes in WGS data resulting from the implementation of each filter.

Statistical Analysis

This study aimed to assess the contribution of the additive genetic effects in the susceptibility of pancreatic cancer. The recently developed restricted maximum likelihood (REML) approach was used to obtain comprehensive estimates of heritability for pancreatic cancer. First, the narrow sense heritability due to additive genetic effects was estimated from all imputed variants across the genome using three forms of REML approach: GREML-SC (single-component GREML), GREML-MS (MAF-stratified GREML) and GREML-LDMS (LD- and MAF-stratified GREML)[17]. The validity of estimated heritability from these GREML approaches depends on the MAF distribution and LD properties of the causal variants, which is unknown to us. By applying all three forms of GREML approach to our dataset, we were able to evaluate the validity of the estimated heritability by comparing the results from different GREML approaches and select the model that generates a less biased estimate for subsequent analysis. Potential bias in the heritability estimation due to residual population stratification and/or cryptic relatedness were assessed to ensure the validity of the results. Next, the estimated heritability was partitioned by MAF, chromosomes, and functional annotation, to infer the genetic architecture of pancreatic cancer. The GREML-LDMS approach was applied to estimate the variance explained by each subset of variants conditioning on the rest of the genome. These joint analyses would account for the correlations between variants in different categories. Lastly, the contributions of identified GWAS loci to the disease heritability was quantified. All the models and corresponding calculation of heritability in this analysis are outlined in Supplementary Table 1.

Estimates of Heritability Explained by Imputed Variants

The heritability explained by imputed variants ($h^2_{imputed}$) was estimated using three forms of REML approach: GREML-SC, GREML-MS, and GREML-LDMS. A detailed

description of these forms of GREML approach can be found in Chapter One. In the GREML-SC analysis, the variance due to additive genetic effects of the imputed variants was estimated by fitting one single genetic relationship matrix (GRM) calculated from all autosomal variants in a mixed linear model (Model #1). The proportion of variance explained by all autosomal variants (i.e. heritability) was then defined as $\boldsymbol{h^2_{imputed}} = \boldsymbol{\sigma^2_g}/(\boldsymbol{\sigma^2_g} + \boldsymbol{\sigma^2_e})$.

In the GREML-MS analysis, all autosomal variants were stratified into two MAF groups: 0.0003 – 0.01 and 0.01 - 0.50. The predictor GRMs were calculated from variants in each MAF category and were fit jointly in a mixed linear model (Model #2). The overall heritability was then defined as $\boldsymbol{h^2_{imputed}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{2}} \boldsymbol{\sigma^2_{g_i}} /(\sum_{\boldsymbol{i=1}}^{\boldsymbol{2}} \boldsymbol{\sigma^2_{g_i}} + \boldsymbol{\sigma^2_e})$.

LD stratification was applied in the GREML-LDMS analysis to further account for the region-specific heterogeneity in LD. In brief, individual variants were first allocated to a large number of overlapping segments with an average length of 100 kb. The mean LD r^2 score for each segment was derived from the LD r^2 score of individual variants in the segment. Based on the segment LD scores, all variants were stratified into two groups (i.e. above or below the mean segment LD score). After the LD stratification, the variants in each LD group were further binned into two MAF categories (0.003-0.01 and 0.01-0.50) as described in the GREML-MS analysis. The predictor GRMs were calculated from each LD-MS group and fitted jointly in a mixed linear model (Model #3). The proportion of variance explained by all autosomal variants was calculated as $\boldsymbol{h^2_{imputed}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{4}} \boldsymbol{\sigma^2_{g_i}} /(\sum_{\boldsymbol{i=1}}^{\boldsymbol{4}} \boldsymbol{\sigma^2_{g_i}} + \boldsymbol{\sigma^2_e})$.

For a binary trait like pancreatic cancer, the estimates of variance need to be transformed from the observed 0-1 scale to the unobserved continuous 'liability' scale using a probit transformation[18]. To correct for the oversampling of cases in the case-control study, a population prevalence of the disease (K) was specified. Based on the 2009-2011 SEER report, the lifetime

risk of being diagnosed with cancer of the pancreas by age 85 was K = 0.0149 for US Whites[19]. Unless specified, the default average information algorithm was applied to run REML iterations in all heritability analysis. All models in the analysis were adjusted for age, sex and the first 20 PCs which were calculated using 99,138 common (MAF > 0.05) and independent (pairwise r^2 > 0.20) variants directly genotyped in our dataset. The incorporation of the first 20 PCs in the model is recommended to control for potential population substructure in the study sample, as the presence of population stratification would bias the estimate of heritability[17]. The variance in liability scale was reported along with its standard errors. Results from GREML-SC, GREML-MS, and GREML-LDMS were compared and the model that generates a less biased estimate of heritability was selected for the subsequent analysis.

Heritability analysis is particularly sensitive to population substructure. Any residual relatedness or population stratification in the study population could inflate the estimates of heritability. The presence of residual population stratification or relatedness can lead to correlations between chromosomes, which can be detected in a separate analysis by fitting only one chromosome at a time in a mixed linear model. In this separate analysis, the variance explained by any given chromosome ($h^2_{c_sep}$) would also capture some of the variance caused by other chromosomes. The joint analysis would protect against such inter-chromosomal correlations because the estimate of variance explained by each chromosome (($h^2_{c_joint}$) is conditional on the other autosomes in the model. To quantify the effect of remaining population structure, the difference between $h^2_{c_sep}$ and $h^2_{c_joint}$ was regressed on chromosome length (i.e. the number of variants).

$$\boldsymbol{h^2_{c_sep} - h^2_{c_joint} = b_0 + b_1 \times L_c + e} \quad \textbf{(4)}$$

The intercept of this regression (b_0) indicates inflation due to cryptic relationship (irrespective of chromosomal length L_c) and b_1 indicates inflation due to population stratification (correlates with L_c). Because the difference between $h^2_{c_sep}$ and $h^2_{c_joint}$ represents the overall effect of all the other 21 autosomes on a given chromosome, the proportion of variance attributed to the cryptic relationship and population stratification across the whole genome is approximately equal to $b_0 \times \frac{22}{21} + b_1 \times \frac{sum(L_c)}{21}$. This analysis was conducted based on results from the GREML-SC approach.

Partitioning by Chromosome, MAF and Functional Annotation

Genomic partitioning of estimated heritability was performed using GREML-LDMS approach, where all autosomal variants were stratified into 4 LD-MS groups as described previously (MAF < 0.01 vs. MAF ≥0.01 and below vs. above median LD score). Additional stratification was applied according to the type of partitioning analysis (by MAF, chromosome or functional annotation). Variance explained by each subset of variants was estimated in joint analysis conditioning on the remaining variants in the genome.

To determine the variance captured by each autosomal chromosome, the variants in 4 LD-MS groups were further allocated to 22 autosomal chromosomes, resulting in 88 LD-MS-chromosome strata. The GRMs from each LD-MS-chromosome stratum were calculated and fitted jointly in a mixed linear model (Model #4). In this analysis, the Fisher scoring approach was applied for variance estimation. The variance captured by each chromosome was aggregated from variance due to 4 LD-MS groups allocated to each chromosome. Standard errors of the summed variance were calculated from the sample variance/covariance matrix using the delta method. Linear regression was performed to assess the correlations between variance explained

by individual chromosome and the length of the chromosome. The length of a chromosome was represented by the total number of variants in the chromosome.

To improve the resolution in MAF distribution of causal variants, variants were binned into six MAF categories: 0.0003 – 0.01, 0.01 – 0.10, 0.10 – 0.20, 0.20 – 0.30, 0.30 – 0.40 and 0.40 – 0.50. Variance explained by each MAF category was estimated from a mixed linear model with 12 genetic components (Model #5). GRM calculated from each component was fitted jointly in the model conditioning on all other components. The variance captured by each MAF category was aggregated from variance due to 2 LD strata within the MAF category. Standard errors of the summed variance were calculated from the sample variance/covariance matrix using the delta method.

All imputed variants were initially annotated into six functional groups according to ANNOVAR annotations: coding, intergenic, intronic, ncRNA, UTR, and upstream/downstream. Due to the limited number of variants in some of these groups, exonic and splicing variants were grouped together as coding variants whereas variants coded as ncRNA, UTR or upstream/downstream were combined as regulatory variants. Therefore, all variants were mapped into four functional groups: coding, intergenic, intronic and regulatory. Variants in each of the four functional groups were further stratified into 2 MAF categories and 2 LD groups. In the joint analysis of all functional groups, the variance explained by each functional category was summed from variance due to 4 LD-MS strata within the functional category (Model #6). Standard errors of the summed variance were calculated from the sample variance/covariance matrix using the delta method.

Contributions of Identified GWAS Loci

The common low-risk susceptibility loci for pancreatic cancer have been reported in six genome-wide association studies (GWAS) in European populations and two GWAS in Asian populations (Chapter Two). A total of 26 GWAS loci reached the genome-wide significance were included in this analysis. The index SNP of these 26 GWAS loci were first identified in our dataset. If not available, the variants with the strongest LD (pairwise r^2) to the index SNP were included to capture the GWAS signals. Then all variants within ±250 kb of the index SNP were grouped together with the index SNP as a single genetic component. The remaining variants across the genome were stratified into 2 MAF categories and 2 LD groups as described previously. The variance explained by the GWAS loci was estimated by fitting 5 GRMs jointly in a mixed linear model (Model #7).

Results

The final sample consists of 3,568 pancreatic cancer cases and 3,363 controls (Tables

Table 4). Case and control samples were similar in sex ratio and age distributions. All subjects in this case-control cohort were aged 40 years or older.

A shows the distribution of minor allele frequencies (MAF) in the final dataset containing 16,184,129 variants. The majority of the variants are rare and low-frequency variants with a MAF < 5%. The remaining variants are evenly distributed across the MAF categories. The majority of the variants in the final dataset are intergenic (52.8%) or intronic (37.4%). About 1% of the sequenced variants are located in the coding regions (

B).

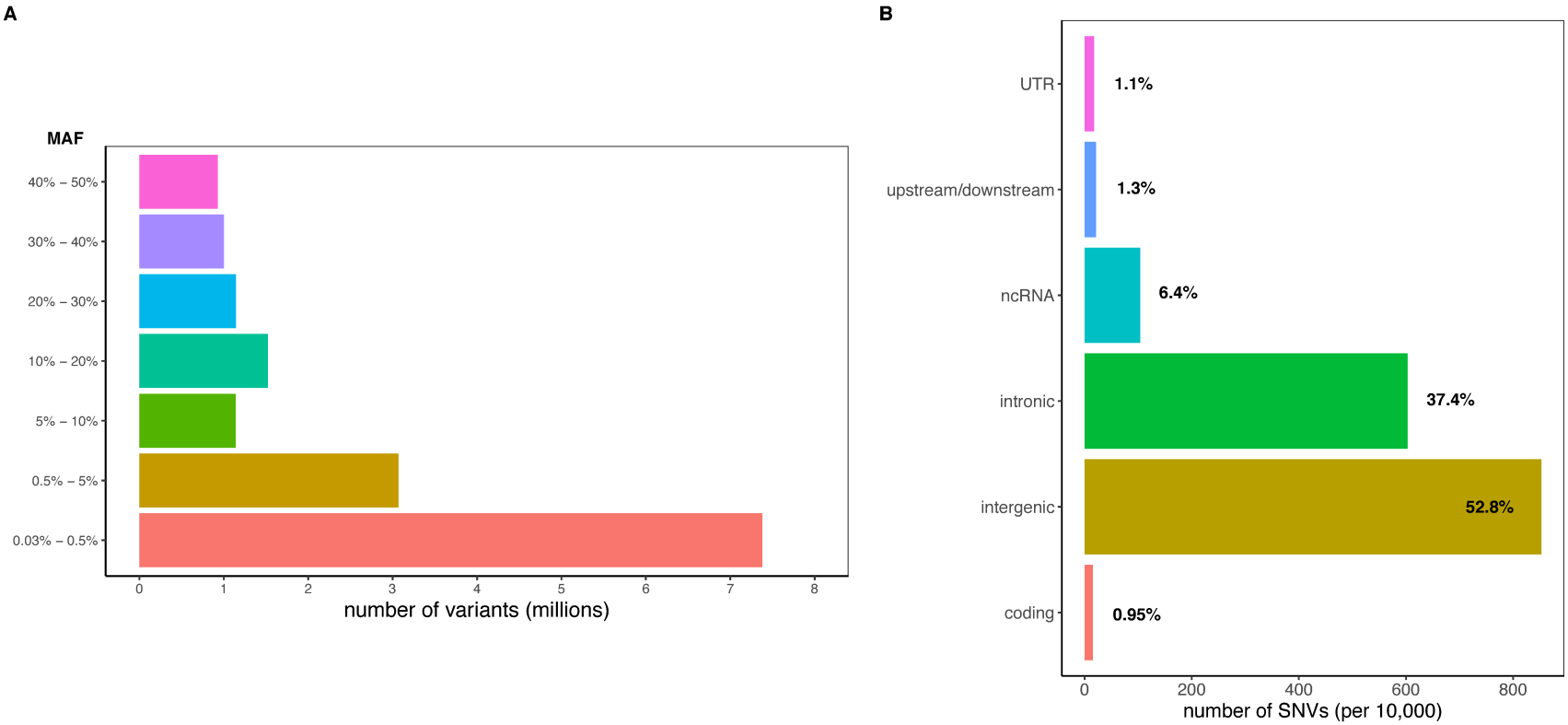


Figure 2 Minor allele frequency (A) and functional annotation (B) distribution of PanC4 imputed variants.

Estimates of Heritability Explained by Imputed Variants

The heritability for pancreatic cancer was calculated in three models applying different forms of GREML approach. The heritability for pancreatic cancer explained by all imputed variants ($h^2_{imputed}$) was estimated to be 16.0% (s.e. = 4.6%) in the single-component model (Model #1). When stratified at MAF of 0.01 (Model #2), the estimate of heritability from GREML-MS was 12.9% (s.e. = 3.9%), lower than the estimate from GREML-SC model. When further stratified by LD (Model #3), the heritability for pancreatic cancer explained by all imputed variants was estimated to be 21.2% (s.e. = 4.8%). Given the substantial differences in the estimates of heritability from these three models, the GREML-LDMS estimate was considered the less biased. Unless specified, all subsequent analysis was conducted based on the GREML-LDMS model.

By comparing the estimated variance explained by each chromosome in a separate ($h^2_{c_sep}$) analysis to a joint ($h^2_{c_joint}$) analysis, the proportion of variance attributed to the residual population stratification and/or cryptic relatedness in our dataset was estimated to be 0.31%, which is unlikely to spuriously inflate the estimate of overall heritability.

Genomic Partitioning of Genetic Variation

Genomic partitioning of the estimated heritability can provide valuable insights on the underlying genetic architecture of the disease. Partitioning by chromosome was conducted in a GREML-LDMS model (Model #4) and the estimated variance associated with each chromosome is shown in Figure 3. Chromosome 9 account for the largest proportion of genetic variation for pancreatic cancer (h^2 = 2.3%, s.e. = 1.6%), followed by chromosome 7 (h^2 = 2.1%, s.e. = 1.8%). Chromosome 8 (h^2 = 1.8%, s.e. = 1.7%), chromosome 16 (h^2 = 1.8%, s.e. = 1.4%), chromosome 5 (h^2 = 1.5%, s.e. = 1.9%), chromosome 2 (h^2 = 1.5%, s.e. = 2.0%), and chromosome 1 (h^2 =

1.5%, s.e. = 2.0%) also explained a fair amount of phenotypic variance for pancreatic cancer. On all these abovementioned chromosomes, common susceptibility loci for pancreatic cancer have been identified in GWAS studies. Regression of the variance explained by individual chromosome on the length of the chromosome found no correlations as shown in Figure 4.

The GREML-LDMS model with 2 MAF categories and 2 LD groups (Model #3) suggest that imputed variants with MAF ≤ 0.01 (48.4% of all imputed variants) explains 5.2% (s.e. = 3.6%) of the phenotypic variance of pancreatic cancer, while common variants (MAF > 0.01) account for about 16.0% (s.e. = 3.5%) of the phenotypic variance (

	Affected Individuals (N = 3,568)	**Unaffected Individuals (N = 3,363)**
Female, N (%)	1,510 (42.3%)	1,491 (44.3%)
Age, N (%)		
40-49	335 (9.4%)	358 (10.6%)
50-59	768 (21.5%)	849 (25.2%)
60-69	1275 (35.7%)	1,140 (33.9%)
70-79	900 (25.2%)	786 (23.4%)
>80	290 (8.1%)	230 (6.8%)

Table 5). When the imputed variants were stratified into 6 MAF categories and 2 LD groups (Model #5), the overall heritability for pancreatic cancer was estimated to be 19.9% (s.e. = 5.0%), which is quite similar to the result from the previous model with 2 MAF categories (Model #3). Partitioning of the estimated heritability by six MAF categories also found a substantial amount of genetic variation at rare variants for pancreatic cancer, with $h^2 = 6.9\%$ (s.e. = 3.8%) for variants with MAF ≤ 0.01, all of which is attributed to variants with lower than average LD scores (Figure 5). Variants with 0.01 < MAF ≤ 0.10 explain a comparable amount of variance for pancreatic cancer ($h^2 = 6.2\%$, s.e. = 3.1%) but mostly from variants with above average LD scores (Figure 5).

In the genomic partitioning by functional groups, intronic and intergenic variants, which contain 37.3% and 52.7% of all imputed variants, account for 12.4% (s.e. = 6.8%) and 6.0% (s.e. = 6.8%) of phenotypic variance for pancreatic cancer, respectively (

	Above mean LD		**Below mean LD**		**Row sum**	**s.e.** *	**No. Variants (%)**
	Est	**s.e.** †	**Est**	**s.e.** †			
MAF ≤ 0.01	0.0523	0.037	0	0.051	0.052	0.036	8,354,405 (51.6%)
MAF > 0.01	0.0142	0.017	0.1460	0.032	0.160	0.035	7,829,724 (48.4%)
Column sum	0.067	0.040	0.146	0.058			
Total sum	**0.212**	**0.048**					
No. Variants (%)	8,092,536 (50%)		8,091,593 (50%)				

* s.e. for row sum and column sum was calculated from variance/covariance matrix.

Table 6). Coding variants, including exonic and splicing variants, despite being the smallest functional group, explained 1.0% (s.e. = 3.9%) of phenotypic variance for pancreatic cancer. The remaining 1.8% variance (s.e. = 3.5%) was attributed to variants in regulatory regions (e.g. UTR, ncRNA and upstream/downstream).

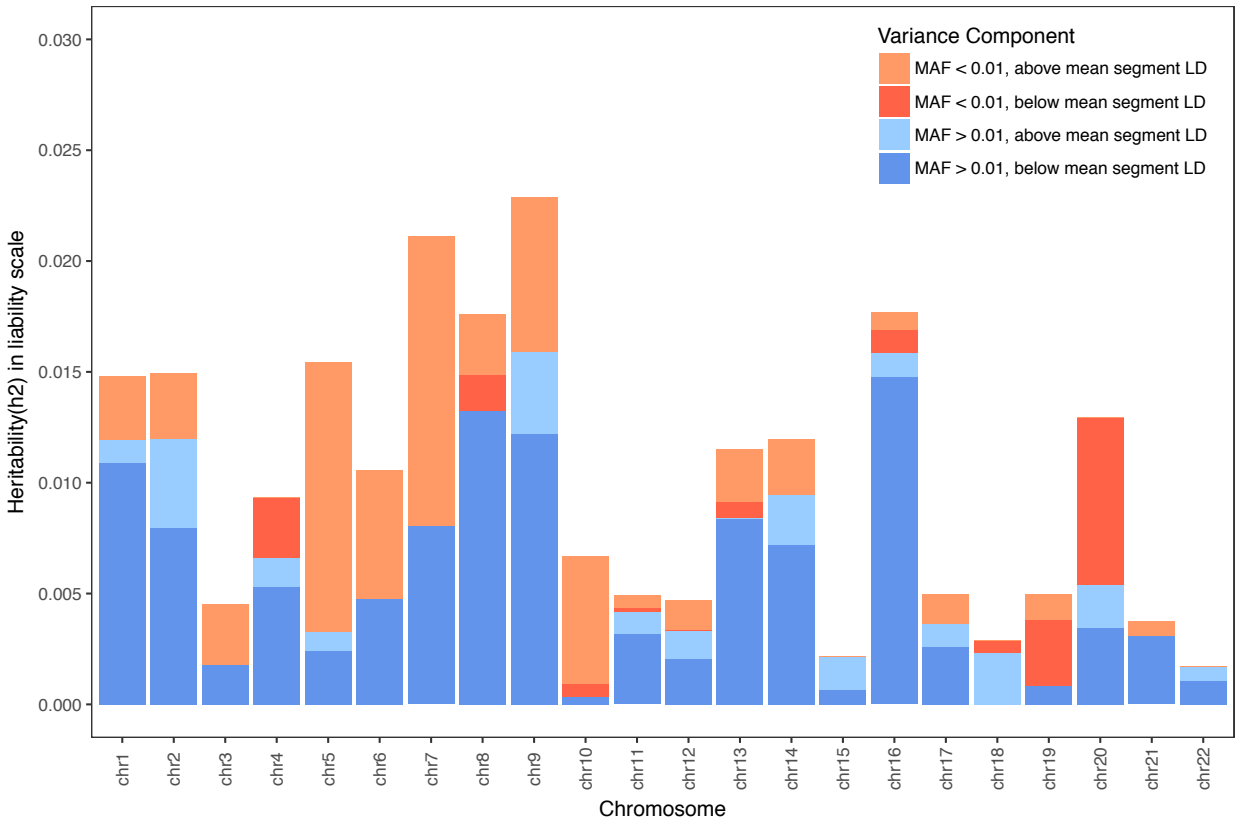


Figure 3 Estimated variance explained by variants on each autosomal chromosome. Variants on each chromosome were stratified into 2 MAF categories and 2 LD groups. The estimated variance associated with individual chromosome was aggregated from the variance explained by four MAF-LD groups.

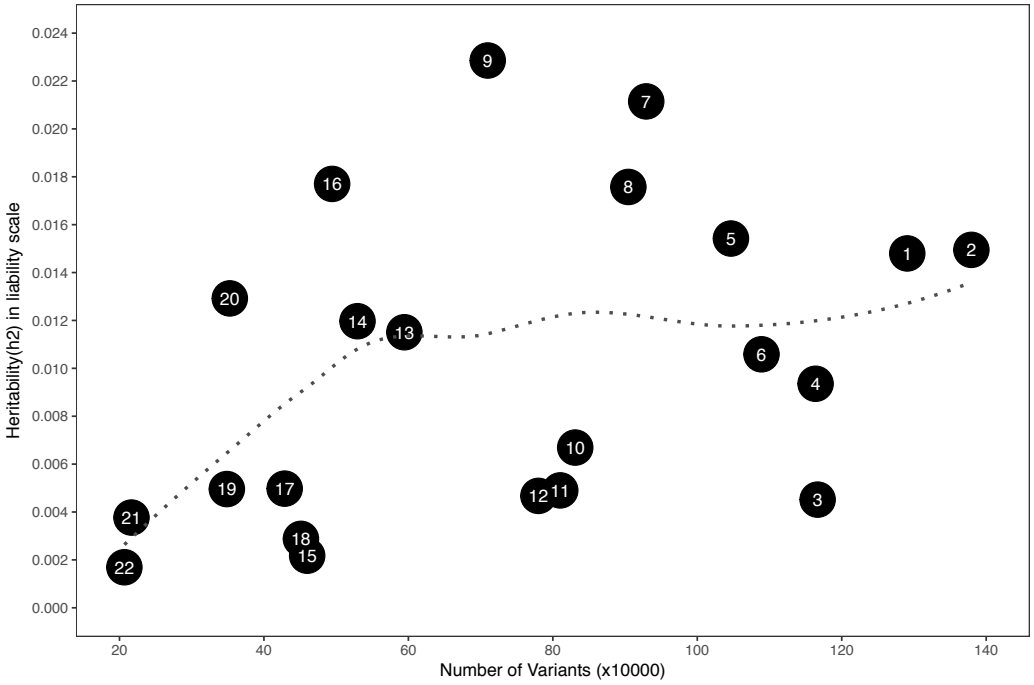


Figure 4 Correlations between the estimated variance explained by individual chromosome and the length of chromosomes. Each black circle represents a chromosome marked by the chromosome ID number. The dotted line is loess smoothed fit curve based on the data. No correlation was found between the estimated variance explained by individual chromosome and the length of chromosomes.

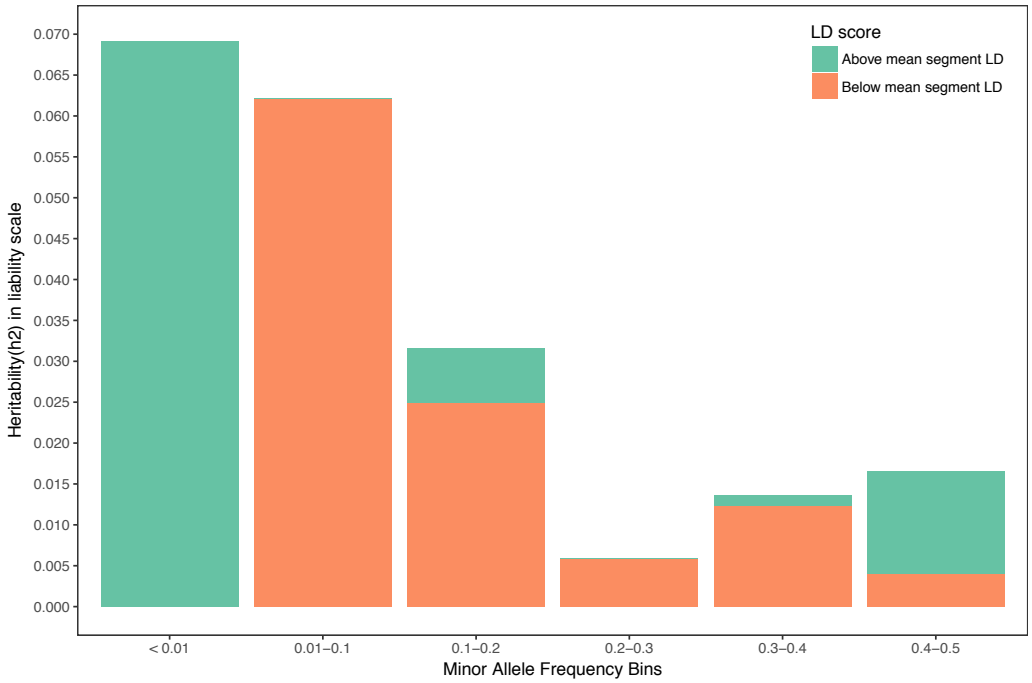


Figure 5 The estimated proportion of phenotypic variance explained by imputed variants in six MAF categories. Variants on each MAF category were further stratified into 2 LD groups (above vs. below mean segment LD score). The estimated variance attributed to each MAF category was aggregated from the variance explained by 2 LD groups.

The contribution of Identified GWAS Loci

Of the 26 common susceptibility loci reported in GWAS studies, 23 index SNPs were available in our dataset. For the three GWAS loci whose index SNP not available in our dataset, including rs2736098 on chromosome 5p13.33 (*TERT-CLPTM1L*), rs10094872 on chromosome 8q24.21 (*MYC*) and rs4795218 on chromosome 17q12 (*HNF1B*), variants in strong LD (pairwise r^2) with these index SNP were included in the analysis (Supplementary Table 3). To assess the aggregate contribution of these 26 GWAS loci to the estimated heritability for pancreatic cancer, a total of 72,225 variants located within ±250 kb of the index SNP was analyzed. Together these variants explained 4.1% (s.e. = 0.8%) of phenotypic variance for pancreatic cancer.

Discussion

Our study represents the first systematic investigation of the genetic architecture of pancreatic cancer. From the variants imputed to the 1000 Genomes Project reference panels, the heritability for pancreatic cancer explained by imputed variants ($h^2_{imputed}$) was estimated to be 21.2% (s.e. = 4.8%). This estimate is considerably lower than the estimates from traditional twin study[20](i.e. h^2_{fam} = 36%) but substantially higher than the array heritability estimates (h^2_{array}) which ranged from 9.8% to 18% [1,21,22]. Previously in the same PanC4 sample, the heritability for pancreatic cancer was estimated to be 16.4% (95% CI = 10.4 – 22.4%) from 620,357 directly genotyped SNPs using GREML-SC approach[1]. Comparing to this array heritability, the use of imputation data in our analysis allows $h^2_{imputed}$ to better capture the variance explained by rare and low-frequency causal variants. In addition, the application of GREML-LDMS approach stratifying variants by MAF and LD can minimize the differences in LD between causal variants and analyzed variants, and consequently reduce the bias associated with GREML-SC estimate.

Therefore, $h^2_{imputed}$ estimated in our analysis is an improved estimate of heritability than the previously reported array heritability.

However, heritability estimated from imputed variants is expected to miss some variance due to rare causal variants. This is because either the rare casual variants were not imputed using the chosen reference panels or they were excluded at quality control stage. Characterizing rare variants in imputation are challenging. It is likely that causal variants with low allele frequency or those unique to pancreatic cancer patients were not polymorphic in 1000 Genome Project reference populations and therefore were not imputed. It's also known that rare variants tend to have lower imputation accuracy than common variants. In our analysis, we applied a stringent INFO score cutoff at 0.50 to exclude variants with poor imputation quality, which affected 56.3% of all imputed variants. Since heritability analysis cannot account for the imputation uncertainty, probabilities of genotypes have to be converted to hard genotypes, which led to high missingness (> 5%) in about 1.9% of imputed variants. This conversion also caused some rare variants to become monomorphic in the data, making no contribution to the heritability analysis. Exclusion based on INFO score and missingness was likely to have a greater impact on rare variants than on common variants. Lastly, following the standard practice of GREML analysis, variants with minor allele count <5 were excluded from our analysis to avoid the noise and unstable estimations. This means that variants with extremely low frequency, such as mutations of established FPC genes, will never be included in the GREML based heritability analysis. As a result, their contribution to the susceptibility of pancreatic cancer can never be accurately reflected in heritability measures.

The underestimation of variance explained by rare variants may be partially explained the difference in h^2_{fam} and $h^2_{imputed}$. Whole-genome sequencing data in theory is capable of

capturing variance due to all causal variants, and hence generates an estimate of heritability closest to h^2_{fam}. In simulation studies, Yang and his colleagues showed that the proportion of variation at sequence variants captured by the 1000 Genomes Project imputation is almost independent of genetic architecture; and on average, about 97% of variation at common variants and about 68% of variation at rare variants could be captured by 1000 Genomes Project imputation[17]. In our GREML-LDMS analysis (Model #3), 5.2% (s.e. = 3.6%) and 16.0% (s.e. = 3.5%) of phenotypic variance for pancreatic cancer can be attributable to rare and common variants, respectively. Extrapolated from the proportion of variation at common and rare variants captured by 1000 Genomes Project imputation, the adjusted estimate of heritability for pancreatic cancer from $h^2 = \frac{h^2_{rare}}{0.68} + \frac{h^2_{common}}{0.97}$ would be 24.1%.

Even with the adjustment, the estimated heritability is still considerably lower than the family-based heritability (h^2_{fam}) of 36%. There are two possible explanations for the remaining missing heritability. First, the estimate of heritability from twin studies may be inflated owing to the shared environmental factors. Differences in shared environmental factors in MZ twins and DZ twins if not properly modeled may lead to an over-estimation of heritability. Although evidence of such inflation has been reported for a number of quantitative and binary traits[23], it is likely to be trait-specific and remains debatable for pancreatic cancer. Secondly, even with the adjustment, the estimated heritability for pancreatic cancer is still likely to be an underestimation since GREML estimates can never capture variance due to extremely rare mutations. Mutations of established FPC genes have previously been detected in 3-10% of sporadic pancreatic cancer patients[24–26]. Given that cases in the PanC4 study were unselected for their family history, the combined prevalence of these mutations in the PanC4 study was expected to be similar to that in published studies. In the most recent study, some of the established FPC genes were found to

have large effects on the risk of pancreatic cancer in sporadic cases, with an odds ratio ranging from 2.58 to 12.33[26]. Given the prevalence and effect sizes, variants of these established FPC genes were expected to explain a measurable fraction of phenotypic variation for pancreatic cancer. However, due to its low frequencies, these rare mutations likely were not imputed using 1000 Genome Projected reference panels or had been removed at quality controls. Without these rare pathological mutations, variants of the established FPC genes only explain a scanty amount of phenotypic variance for pancreatic cancer (0.4%) in our study. Using whole-genome sequencing data in a large enough cohort would largely lessen the underestimate of these very rare variants. However, even with WGS data, the standard practice is to exclude variants with small minor allele count (MAC, e.g. MAC < 3 or < 5), in order to minimize the noise in the analysis. As a result, the GREML estimate of heritability will always be an underestimation of the overall heritability, and the magnitude of underestimation depends on the underlying genetic architecture of the trait. For cancer phenotypes known to be affected largely by rare mutations, the bias can be substantial.

To infer the underlying genetic architecture of pancreatic cancer, the estimate of overall heritability was partitioned by MAF, chromosome, and functional annotations. Partitioning by chromosome ranked chromosome 9, 7, 8, 16, 5, 2 and 1 as top contributors to the heritability of pancreatic cancer. This is consistent with the GWAS findings as on all these chromosomes common susceptibility loci have been discovered (Chapter Two). Our results also suggest that variants in genomic regions with lower LD tend to explain a larger proportion of variance than those in regions with higher LD (14.6% vs. 6.7%, Table 1). Rare variants in lower than average LD regions appear to make little or no contribution to the estimated heritability for pancreatic cancer in our analysis (Table 5, Figure 5), which could be a result of under-representation of

such variants in the imputation data. As discussed above, since heritability estimated from imputed variants captures almost all variation at common variants but only a proportion of variation at rare variants, results from partitioning by chromosome are likely driven by common causal variants, some of which had been identified through GWAS studies.

Results from partitioning by MAF showed a non-negligible proportion of phenotypic variance explained by variants with MAF ≤ 0.01. More than 1/3 of the estimated heritability is attributable to rare variants. As suggested in simulation studies, this is likely an underestimate since not all variation at rare variants is captured by 1000 Genomes Project imputation, particularly those extremely rare variants that are known to confer risk in FPC cases. Our results demonstrate an important role of rare variants in the susceptibility of pancreatic cancer, which emphasize the necessity of WGS studies. High-coverage sequencing studies with a near-complete characterization of genetic variation would allow examination of variants with extremely low frequency or unique to specific populations. However, obtaining a sufficiently large sample to permit such sequencing analysis remains a significant challenge.

On the other hand, common variants still account for a large proportion of estimated heritability in our study. Previous GWAS studies in the PanC4 study reported that 3.0% (95% CI = 2.0 – 3.9%) of the total phenotypic variation was accounted for by established association regions[1]. In our analysis, known GWAS loci explained 4.1% of phenotypic variance for pancreatic cancer, leaving > 10% of the phenotypic variance unexplained. The increment in genetic variance due to known GWAS loci in our study is likely resulting from the inclusion of the five novel loci discovered in a recent meta-analysis of four pancreatic cancer consortium studies[27]. The large proportion of unexplained heritability highlights the importance of continuously searching for common susceptibility loci for pancreatic cancer. SNP-array based

genotyping followed by imputation will remain a more cost-effective strategy than WGS studies. However, larger sample sizes will be required to increase the power of current GWAS for identification of yet-to-be detected causal common and low-frequency variants. As imputation reference panel of large sample size (e.g. Haplotype Reference Consortium, HRC) becoming available, the imputation accuracy is expected to be greatly improved for low frequency and rare variants, hence increase the power of detecting associations on these variants[28,29]. An alternative approach to improve the imputation quality on rare variants is to use a population-specific reference panel, which would provide better enrichment of population-specific variants that are likely to be monomorphic in commonly used reference populations[30–32]. This could be particularly useful for a low-prevalence disease like pancreatic cancer, known to be largely affected by rare variants. Currently, a pancreatic cancer specific haplotype reference panel is under construction from high-coverage WGS (50x) of 638 pancreatic cancer cases of European ancestry. For future imputation and meta-analysis, there is great potential to gain power using this disease-specific reference panel, in combination with 1000 Genomes Project or HRC panels, for rare variants in particular.

Across four functional groups, intronic variants account for most phenotypic variance of pancreatic cancer (12.4%). Interestingly, 12 out of 21 GWAS loci identified in the European population are mapped to intronic variants. However, it is unclear whether these variants are of direct functional significance, as opposed to simply being in linkage disequilibrium with another functional SNP in the vicinity. The coding variants, comprising about 1% of imputed variants, account for 1% of the phenotypic variance of pancreatic cancer. This is likely an underestimate since a proportion of rare or extremely rare coding variants were not imputed or removed by quality controls. Comparing to the other three functional groups, the proportion of variants with a

MAF < 0.005 is much higher among coding variants (Supplementary Figure 1). It is possible that the poor imputation accuracy on rare and extremely rare variants has a greater impact on coding variants than variants in other three functional groups.

Conclusion

In a subset of the PanC4 sample of 3,568 pancreatic cancer cases and 3,363 controls, the heritability for pancreatic cancer is estimated to be 21.2% from imputed variants. This estimate is less biased than reported array-based heritability but remains an underestimation of overall heritability due to limited coverage on rare and extremely rare variants. It is known that rare mutations are presented in 8-30% familial cases and in 3-10% of sporadic cases. Current statistical approaches, such as GREML used in this study, is incapable of capturing variance due to these rare mutations. Genomic partitioning by chromosome ranks several autosomal chromosomes as top contributors to the estimated heritability. Since these results are largely driven by common variants, they provide very little guidance on physical locations of causal variants. Rare variants with MAF < 0.01 explain about 1/3 of the estimated heritability, suggesting an important role of these variants in the susceptibility of pancreatic cancer. Whole-genome sequencing that determines the genetic variation underlying the full spectrum of pancreatic cancer will be necessary for rare causal variant discovery. However, establishing a sufficiently large cohort for sequencing analysis is very costly. Combining existing sequencing cohorts is an interesting strategy that is worth being further investigated for its performance in identifying rare causal variants. In our study, known GWAS loci only account for 4.1% of the phenotypic variance for pancreatic cancer, significantly less than the amount of variance attributable to common variants (16.0%). Imputation and meta-analysis have proven fruitful in the search for common and low-frequency susceptibility loci for pancreatic cancer, and remain

the most cost-effective strategy in future gene discovery research for the disease. Increasing the sample size and improving the imputation accuracy would be the key to ensure progress in such studies. In addition to using a larger reference panel like Haplotype Reference Consortium, a disease-specific reference panel for pancreatic cancer would further improve the imputation on variants with extremely low frequency or unique to the disease populations.

Tables

Table 4 Characteristics of Analyzed PanC4 Cohort

	Affected Individuals (N = 3,568)	**Unaffected Individuals (N = 3,363)**
Female, N (%)	1,510 (42.3%)	1,491 (44.3%)
Age, N (%)		
40-49	335 (9.4%)	358 (10.6%)
50-59	768 (21.5%)	849 (25.2%)
60-69	1275 (35.7%)	1,140 (33.9%)
70-79	900 (25.2%)	786 (23.4%)
>80	290 (8.1%)	230 (6.8%)

Table 5 Estimates of Variance Explained by Imputed Variants from GREML-LDMS Analysis

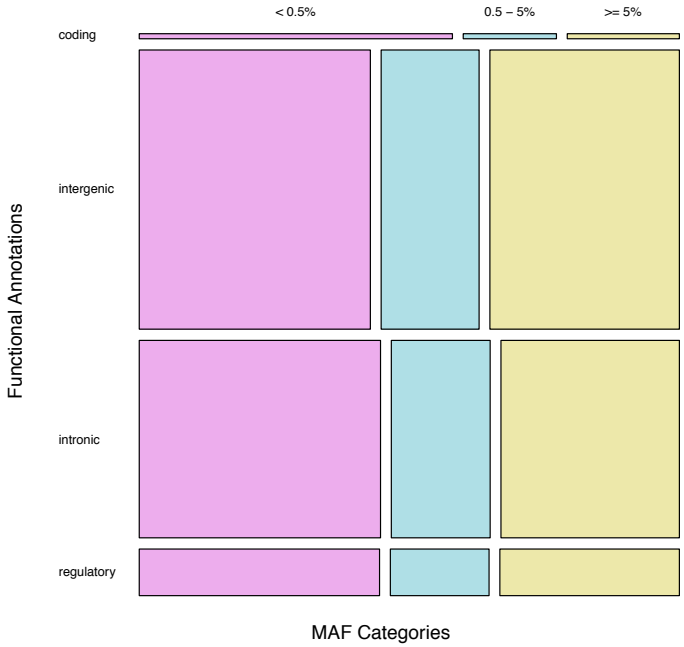
	Above mean LD		Below mean LD		Row sum	s.e. *	No. Variants (%)
	Est	s.e. †	Est	s.e. †			
MAF ≤ 0.01	0.0523	0.037	0	0.051	0.052	0.036	8,354,405 (51.6%)
MAF > 0.01	0.0142	0.017	0.1460	0.032	0.160	0.035	7,829,724 (48.4%)
Column sum	0.067	0.040	0.146	0.058			
Total sum	**0.212**	**0.048**					
No. Variants (%)	8,092,536 (50%)		8,091,593 (50%)				

* s.e. for row sum and column sum was calculated from variance/covariance matrix.

Table 6 Estimates of Variance Explained by Imputed Variants in Four Functional Groups

		Above mean LD		Below mean LD		Row sum	s.e.	No. Variants (%)	Sub-category sum	s.e.
		Est	s.e.	Est	s.e.					
Coding	MAF ≤ 0.01	0.0016	0.022	0.0036	0.029	0.005	0.036	103,133 (0.6%)	0.010	0.039
	MAF > 0.01	0.0050	0.009	0	0.014	0.005	0.017	50,753 (0.3%)		
Intergenic	MAF ≤ 0.01	0.0334	0.036	0	0.059	0.033	0.063	4,305,707 (26.6%)	0.060	0.068
	MAF > 0.01	0	0.015	0.0267	0.026	0.027	0.030	4,220,706 (26.1%)		
Intronic	MAF ≤ 0.01	0.0416	0.037	0	0.055	0.042	0.061	3,173,417 (19.6%)	0.124	0.066
	MAF > 0.01	0.0086	0.014	0.0735	0.025	0.082	0.028	2,856,226 (17.6%)		
Regulatory	MAF ≤ 0.01	0	0.024	0	0.036	0	0.043	746,866 (4.6%)	0.018	0.045
	MAF > 0.01	0.0002	0.009	0.0181	0.014	0.018	0.016	676,634 (4.2%)		

Supplementary Figures and Tables



Supplementary Figure 1 MAF distribution of variants in four functional groups. The heights of the boxes are proportional to the percentage of all variants being annotated as coding, intergenic, intronic or regulatory variants. The widths of the boxes are proportional to the percentage of variants in each functional group fallen into each MAF categories.

Supplementary Table 1 Models and Calculations of h^2 in Heritability Analysis

Analysis	Model#	GREML Approach*	Mixed Linear Model†	List of Genetic Components	Calculation of h^2
Estimates of heritability explained by imputed variants	1	GREML-SC	$y = u + g + \varepsilon$	g	$h^2_{imputed} = \sigma^2_g/(\sigma^2_g + \sigma^2_e)$
	2	GREML-MS	$y = u + \sum_{i=1}^{2} g_i + \varepsilon$	$g_{MAF<0.01}$ $g_{MAF\geq 0.01}$	$h^2_{imputed} = \frac{\sum_{i=1}^{2}\sigma^2_{g_i}}{\sum_{i=1}^{2}\sigma^2_{g_i} + \sigma^2_e}$
	3	GREML-LDMS	$y = u + \sum_{i=1}^{4} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below}$ $g_{MAF<0.01,LD\ above}$ $g_{MAF\geq 0.01,LD\ below}$ $g_{MAF\geq 0.01,LD\ above}$	$h^2_{imputed} = \frac{\sum_{i=1}^{4}\sigma^2_{g_i}}{\sum_{i=1}^{4}\sigma^2_{g_i} + \sigma^2_e}$
Genomic partitioning by chromosome	4	GREML-LDMS	$y = u + \sum_{i=1}^{88} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below,chr1}$ $g_{MAF<0.01,LD\ above,chr1}$ $g_{MAF\geq 0.01,LD\ below,chr1}$ $g_{MAF\geq 0.01,LD\ above,chr1}$ … $g_{MAF<0.01,LD\ below,chr22}$ $g_{MAF<0.01,LD\ above,chr22}$ $g_{MAF\geq 0.01,LD\ below,chr22}$ $g_{MAF\geq 0.01,LD\ above,chr22}$	$h^2_{chr1} = \frac{\sum_{i=1}^{4}\sigma^2_{g_{i,chr1}}}{\sum_{i=1}^{88}\sigma^2_{g_i} + \sigma^2_e}$ … $h^2_{chr22} = \frac{\sum_{i=1}^{4}\sigma^2_{g_{i,chr22}}}{\sum_{i=1}^{88}\sigma^2_{g_i} + \sigma^2_e}$
Genomic Partitioning by MAF	5	GREML-LDMS	$y = u + \sum_{i=1}^{12} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below}$ $g_{MAF<0.01,LD\ above}$ $g_{MAF\ 0.01-0.10,LD\ below}$ $g_{MAF\ 0.01-0.10,LD\ above}$ $g_{MAF\ 0.10-0.20,LD\ below}$ $g_{MAF\ 0.10-0.20,LD\ above}$ $g_{MAF\ 0.20-0.30,LD\ below}$ $g_{MAF\ 0.20-0.30,LD\ above}$	$h^2_{MAF<0.01} = \frac{\sum_{i=1}^{2}\sigma^2_{g_{i,MAF<0.01}}}{\sum_{i=1}^{12}\sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.01-0.10} = \frac{\sum_{i=1}^{2}\sigma^2_{g_{i,MAF\ 0.01-0.10}}}{\sum_{i=1}^{12}\sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.10-0.20} = \frac{\sum_{i=1}^{2}\sigma^2_{g_{i,MAF\ 0.10-0.20}}}{\sum_{i=1}^{12}\sigma^2_{g_i} + \sigma^2_e}$

				$g_{MAF\ 0.30-0.40,LD\ below}$ $g_{MAF\ 0.30-0.40,LD\ above}$ $g_{MAF\ 0.40-0.50,LD\ below}$ $g_{MAF\ 0.40-0.50,LD\ above}$	$h^2_{MAF\ 0.20-0.30} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.20-0.30}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.30-0.40} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.30-0.40}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{MAF\ 0.40-0.50} = \frac{\sum_{i=1}^{2} \sigma^2_{g_{i,MAF\ 0.40-0.50}}}{\sum_{i=1}^{12} \sigma^2_{g_i} + \sigma^2_e}$
Genomic partitioning by functional annotation	6	GREML-LDMS	$y = u + \sum_{i=1}^{16} g_i + \varepsilon$	$g_{MAF<0.01,LD\ below,coding}$ $g_{MAF<0.01,LD\ above,coding}$ $g_{MAF\geq0.01,LD\ below,coding}$ $g_{MAF\geq0.01,LD\ above,coding}$ $g_{MAF<0.01,LD\ below,intergenic}$ $g_{MAF<0.01,LD\ above,intergenic}$ $g_{MAF\geq0.01,LD\ below,intergenic}$ $g_{MAF\geq0.01,LD\ above,intergenic}$ $g_{MAF<0.01,LD\ below,intronic}$ $g_{MAF<0.01,LD\ above,intronic}$ $g_{MAF\geq0.01,LD\ below,intronic}$ $g_{MAF\geq0.01,LD\ above,intronic}$ $g_{MAF<0.01,LD\ below,regulatory}$ $g_{MAF<0.01,LD\ above,regulatory}$ $g_{MAF\geq0.01,LD\ below,regulatory}$ $g_{MAF\geq0.01,LD\ above,regulatory}$	$h^2_{coding} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,coding}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{intergenic} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,intergenic}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{intronic} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,intronic}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$ $h^2_{regulatory} = \frac{\sum_{i=1}^{4} \sigma^2_{g_{i,regulatory}}}{\sum_{i=1}^{16} \sigma^2_{g_i} + \sigma^2_e}$
Evaluation of GWAS loci	7	GREML-LDMS	$y = u + g_{GWAS} + \sum_{i=1}^{4} g_i + \varepsilon$	g_{GWAS} $g_{MAF<0.01,LD\ below,remaining}$ $g_{MAF<0.01,LD\ above,remaining}$ $g_{MAF\geq0.01,LD\ below,remaining}$ $g_{MAF\geq0.01,LD\ above,remaining}$	$h^2_{GWAS} = \frac{\sigma^2_{GWAS}}{\sigma^2_{GWAS} + \sum_{i=1}^{4} \sigma^2_{g_{i,remaining}} + \sigma^2_e}$

* GREML-SC = single-component GREML; GREML-MS = MAF-stratified GREML; GREML-LDMS = LD- and MAF-stratified GREML.

† The disease status in observed scale (y) is a functional of some fixed effects (μ) including intercept, age, sex, and first 20 principal components, one or more genetic components (g_i), and a residual effect (ε).

Supplementary Table 2 Description of the Studies Included in PanC4 Case-Control Consortium

Study	Cases	Controls	Accrual Years	Source of Cases	Source of Controls	Control Matching	Age of Diagnosis for Cases (SD) [†]	Age of Controls (SD) [*]	Male (%)	Cauc. (%) [†]
IARC	448	456	2006-2010	Academic hospitals	General practitioners	Age; sex; region	63.84 (11.16)	61.88 (11.88)	57	100
Johns Hopkins Hospital	315	81	2007-2011	Clinic	Spouse in-law	None (in-law)	64.29 (11.53)	63.45 (14.73)	51	93
Mayo Clinic	1104	1027	2000-2010	Clinic	Primary care patients	Age; sex; race; residence	65.92 (11.09)	63.34 (10.58)	56	93
MD Anderson	616	509	1997-2007	Hospital	Friends and spouse of non-pancreatic cancer patients	Age; race; sex	62.6 (9.69)	59.2 (10.6)	59	100
Memorial Sloan Kettering	317	139	2000-2008	Clinic	Patient spouses and patient visitors	None	64.01 (10.39)	61.68 (10.98)	64	87
Toronto	402	401	2003-2012	Population-based cancer registry	Family medicine clinic database	Age; sex; ethnicity	64.92 (11.03)	62.95 (11.73)	50	85
Queensland	559	604	2007-2011	Population-based hospitals & cancer registry	Electoral rolls	Age; sex	66.64 (11.26)	67.51 (10.91)	60	96
UCSF	253	248	2006-2010	Two UCSF clinics	Three UCSF clinics	Frequency matched by age and sex	62.52 (10.35)	60.40 (10.96)	54	82
YALE	156	366	2005-2009	Population-based hospitals & cancer registry	Enhanced RDD	Frequency matched by age and sex	67.01 (10.43)	65.15 (10.6)	59	93
Total	4170	3831					64.76	63.09	58	95

[*] SD: standard deviation

[†] Percent self-reported as Caucasians

Supplementary Table 3 Availability of the Index SNP of Pancreatic Cancer GWAS Loci in the Data

CHR*	Nearest Gene(s)	Index_SNP†	Position	Population	Availability	Alternative	Pairwise_r^2
1p36.33	*NOC2L*	rs13303010	894573	European	yes	NA	NA
1q32.1	*NR5A2*	rs3790844	200007432	European	yes	NA	NA
1q32.1	*NR5A2*	rs2816938	199985368	European	yes	NA	NA
2p13.3	*ETAA1*	rs1486134	67639769	European	yes	NA	NA
3q29	*TP63*	rs9854771	189508471	European	yes	NA	NA
5q15.33	*TERT, CLPTM1L*	rs401681	1322087	European	yes	NA	NA
5p15.33	*TERT, CLPTM1L*	rs2736098	1294086	European	no	rs548166886	1.000
5p15.33	*TERT, CLPTM1L*	rs35226131	1295373	European	yes	NA	NA
7p12	*TNS3*	rs73328514	47488569	European	yes	NA	NA
7p13	*SUGCT*	rs17688601	40866663	European	yes	NA	NA
7q32.3	*LINC-PINT*	rs6971499	130680521	European	yes	NA	NA
8q21.11	*HNF4G*	rs2941471	76470404	European	yes	NA	NA
8q24.21	*MYC*	rs10094872	128719884	European	no	rs9642880	0.598
9q34	*ABO*	rs505922	136149229	European	yes	NA	NA
13q12.2	*PDX1*	rs9581943	28493997	European	yes	NA	NA
13q22.1	none	rs9543325	73916628	European	yes	NA	NA
16q23.1	*BCAR1, CTRB1, CTRB2*	rs7190458	75263661	European	yes	NA	NA
17q12	*HNF1B*	rs4795218	36078510	European	no	rs3110645	1.000

17q25.1	*LINC00673*	rs11655237	70400166	European	yes	NA	NA
18q21.32	*GRP*	rs1517037	56878274	European	yes	NA	NA
22q12.1	*ZNRF3*	rs16986825	29300306	European	yes	NA	NA
5p13.1	*DAB2*	rs2255280	39394989	Asian	yes	NA	NA
10q26.11	*RPLHR*	rs12413624	120278944	Asian	yes	NA	NA
21q21.3	*BACH1*	rs372883	30717737	Asian	yes	NA	NA
21q22.3	*TFF1*	rs1547374	43778895	Asian	yes	NA	NA
22q13.32	*FAM19A5*	rs5768709	48929569	Asian	yes	NA	NA

* Chromosomal location in NCBI genome build 37.

† SNP = single nucleotide polymorphisms; Index SNP = variant reported to be associated with pancreatic cancer

Reference

1. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet. 2015 Aug;47(8):911–6.

2. Urayama KY, Holcatova I, Janout V, Foretova L, Fabianova E, Adamcakova Z, et al. Body mass index and body size in early adulthood and risk of pancreatic cancer in a central European multicenter case-control study. Int J Cancer. 2011 Dec 15;129(12):2875–84.

3. Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJA, et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res. 2004 Apr 1;64(7):2634–8.

4. Brune KA, Lau B, Palmisano E, Canto M, Goggins MG, Hruban RH, et al. Importance of age of onset in pancreatic cancer kindreds. J Natl Cancer Inst. 2010 Jan 20;102(2):119–26.

5. McWilliams RR, Bamlet WR, Cunningham JM, Goode EL, de Andrade M, Boardman LA, et al. Polymorphisms in DNA repair genes, smoking, and pancreatic adenocarcinoma risk. Cancer Res. 2008 Jun 15;68(12):4928–35.

6. Hassan MM, Bondy ML, Wolff RA, Abbruzzese JL, Vauthey J-N, Pisters PW, et al. Risk factors for pancreatic cancer: case-control study. Am J Gastroenterol. 2007 Dec;102(12):2696–707.

7. Olson SH, Orlow I, Simon J, Tommasi D, Roy P, Bayuga S, et al. Allergies, variants in IL-4 and IL-4R alpha genes, and risk of pancreatic cancer. Cancer Detect Prev. 2007;31(5):345–51.

8. Eppel A, Cotterchio M, Gallinger S. Allergies are associated with reduced pancreas cancer risk: A population-based case-control study in Ontario, Canada. Int J Cancer. 2007 Nov 15;121(10):2241–5.

9. Tran B, Whiteman DC, Webb PM, Fritschi L, Fawcett J, Risch HA, et al. Association between ultraviolet radiation, skin sun sensitivity and risk of pancreatic cancer. Cancer Epidemiol. 2013 Dec;37(6):886–92.

10. Duell EJ, Bracci PM, Moore JH, Burk RD, Kelsey KT, Holly EA. Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2008 Jun;17(6):1470–9.

11. Risch HA. Etiology of pancreatic cancer, with a hypothesis concerning the role of N-nitroso compounds and excess gastric acidity. J Natl Cancer Inst. 2003 Jul 2;95(13):948–60.

12. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224–8.

13. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012 Jul 22;44(8):955–9.

14. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061–73.

15. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.

16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.

17. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

18. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011 Mar 11;88(3):294–305.

19. Lifetime Risk (Percent) of Being Diagnosed with Cancer by Site and Race/Ethnicity Both Sexes, 18 SEER Areas, 2009-2011 [Internet]. [cited 2015 Nov 10]. Available from: http://seer.cancer.gov/archive/csr/1975_2011/browse_csr.php?sectionSEL=1&pageSEL=sect_01_table.15.html

20. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000 Jul 13;343(2):78–85.

21. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, et al. Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. J Natl Cancer Inst. 2015 Dec;107(12).

22. Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, et al. Most common "sporadic" cancers have a significant germline genetic component. Hum Mol Genet. 2014 Nov 15;23(22):6112–8.

23. Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. Nat Genet. 2016;48(9):980–3.

24. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious Germline Mutations in Patients With Apparently Sporadic Pancreatic Adenocarcinoma. J Clin Oncol Off J Am Soc Clin Oncol. 2017 Oct 20;35(30):3382–90.

25. Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. JAMA. 2018 Jun 19;319(23):2401–9.

26. Yurgelun MB, Chittenden AB, Morales-Oyarvide V, Rubinson DA, Dunne RF, Kozak MM, et al. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. Genet Med Off J Am Coll Med Genet. 2018 Jul 2;

27. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun. 2018 Feb 8;9(1):556.

28. Wood AR, Beaumont R, Hernandez D, Nalls M, Gibbs JR, Bandinelli S, et al. Imputation of rare variants from the new Haplotype Reference Consortium identifies associations missed by 1000 Genomes. In Baltimore, MD; 2015.

29. Iglesias AI, van der Lee SJ, Bonnemaijer PWM, Höhn R, Nag A, Gharahkhani P, et al. Haplotype reference consortium panel: Practical implications of imputations with large reference panels. Hum Mutat. 2017 Aug;38(8):1025–32.

30. Pistis G, Porcu E, Vrieze SI, Sidore C, Steri M, Danjou F, et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. Eur J Hum Genet EJHG. 2015 Jul;23(7):975–83.

31. Zhou W, Fritsche LG, Das S, Zhang H, Nielsen JB, Holmen OL, et al. Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. Genet Epidemiol. 2017 Dec;41(8):744–55.

32. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. Eur J Hum Genet EJHG. 2017;25(7):869–76.

CHAPTER FOUR

HERITABILITY FOR PANCREATIC CANCER FROM WGS VARIANTS

The heritability analysis was conducted in two independent case-control studies: a subset of the Pancreatic Cancer Case-Control Consortium (PanC4) study and a combined sample of Familial Pancreatic Cancer (FPC) patients from the Familial Pancreatic Cancer Genome Sequencing Project (FPC-GSP) and participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. This chapter presents the analysis conducted in the FPC-ADNI study, including the estimate of heritability from WGS variants in FPC-ADNI study, as well as the comparison of heritability estimates between FPC-ADNI and PanC4 study using the same set of genotyped variants.

Study Populations and Genetic Data

Familial Pancreatic Cancer (FPC) Patients

FPC-GSP is a multi-institutional study, encompassing 10 familial pancreatic cancer (FPC) registries across North American, with the goal of identifying additional pancreatic cancer susceptibility genes through whole-genome sequencing (WGS) analysis (https://www.familialpancreaticcancer.org/). Patients with FPC were enrolled into the National Familial Pancreatic Tumor Registry (NFPTR) at Johns Hopkins or one of the FPC registries participating at the following sites: Dana-Farber Cancer Institute (Boston, MA), Karmanos Cancer Institute (Detroit, MI), Mayo Clinic (Rochester, MN), McGill University Health Centre (Montreal, Quebec, Canada), Memorial Sloan Kettering Cancer Center (New York, NY), Mount Sinai Hospital (Toronto Ontario, Canada), University of Michigan (Ann Arbor, MI), University of Pennsylvania (Philadelphia, PA), and University of Pittsburgh (Pittsburgh, PA). FPC families are defined as kindreds with at least two first-degree relatives (FDRs) diagnosed with pancreatic

ductal adenocarcinoma (PDAC). Of 593 FPC kindreds, a total of 658 patients were selected for sequencing as they were either untested or tested negative for any previously known deleterious mutation in an establish FPC gene[1]. Diagnosis of pancreatic cancer was self-reported. When possible, all cancer diagnoses were pathologically confirmed.

Germline DNA samples were obtained from blood (n=454), Epstein-Barr virus transformed peripheral blood lymphocytes (lymphoblastoid cell line, LCL) (n=158), or non-tumor tissue (n=26). These samples were whole-genome sequenced using Illumina HiSeq 2000 at 30x depth of coverage and genotyped with the HumanOmni2.5-8v1 SNP array (Illumina, Inc., San Diego, CA) by Personal Genome Diagnostics. Whole genome sequencing generated 200 base pair (2×100 bp reads) per fragment in the final library. Sequence reads were analyzed using ELAND v2 (Illumina) for alignment to the human reference genome hg19/GRCh37 and CASAVA v1.7 for variant calling[2].

Alzheimer's Disease Neuroimaging Initiative (ADNI) Subjects

ADNI is an ongoing, longitudinal, multicenter study launched in 2004. The overarching goals of ADNI project were to characterize clinical, genetic, imaging and biochemical biomarkers of Alzheimer's Disease (AD) over the course of disease progression, and to establish standardized methods for imaging/biomarker collection and analysis for ultimate use in clinical trials [3]. The ADNI project has three phases, ADNI1, ADNI GO and ADNI2, with distinct research objectives. The full ADNI cohort recruited 822 individuals aged 55-90 years from 59 participating sites across the United States and Canada. The cohort consists of subjects with normal cognition, with early or late mild cognitive impairment (MCI) or with AD. Individuals with cancer were not eligible to participate in ADNI [3].

As genetic factors play an important role in Alzheimer's disease, a key aim of the ADNI is to combine genetics with imaging and clinical data to help investigate mechanisms of the disease. Of 822 ADNI participants, 818 with a sufficient amount of DNA sample were genotyped and sequenced. DNA samples of these individuals were derived from either peripheral blood (n=731) or immortalized LCL (n=87)[4]. These samples were genotyped on HumanOmni2.5 SNP array and sequenced on non-CLIA WGS platform at Illumina at 30X depth (Illumina, Inc., San Diego, CA). The WGS data were aligned and called using CASAVA on reference genome hg19/GRCh37 per Illumina's protocol in the spring of 2013. After basic quality control checks, the released dataset contains 809 ADNI samples.

Quality Controls on FPC-ADNI Study

A series of quality control (QC) procedures were implemented on WGS data to minimize the impacts of low-quality variants, mishandling of DNA samples, cryptic relatedness, population substructures and batch effects on heritability estimates. Insertion/deletion variants were not included in our analysis. **Error! Reference source not found.** presents an overview of these QC procedures. Prior to QC, the FPC WGS data consisted of 658 samples sequenced on 56,330,793 single-nucleotide variants (SNVs) with a transition to transversion ratio (Ti/Tv) of 1.47. The ADNI WGS data contained 809 samples sequenced on 59,444,512 SNVs with a Ti/Tv of 1.52. Before merging the FPC WGS and ADNI WGS data, genotypes with genotyping quality score (GQ) < 20 or read depth (DP) < 10 were set to missing. Variants in annotated super-duplicated regions were excluded. Samples were dropped if they had non-reference discordance rate (NRD) > 5% or mismatched sex. In addition, twenty FPC samples were removed due to being duplicates or with a non-FPC diagnosis. After merging the WGS data of FPC and ADNI cohort, we further removed non-European samples as well as closely related individuals (first-

and second-degree relatives). In the end, variants with the following characteristics were further excluded: monomorphic variants, missingness > 10%, Hardy-Weinberg Equilibrium exact test with $p < 10^{-}6$, and minor allele count < 3. The final dataset contains 1,323 samples on 14,055,912 single-nucleotide variants (SNVs). The Ti/Tv ratio in the final dataset was 2.07, which was considered adequate for WGS data.

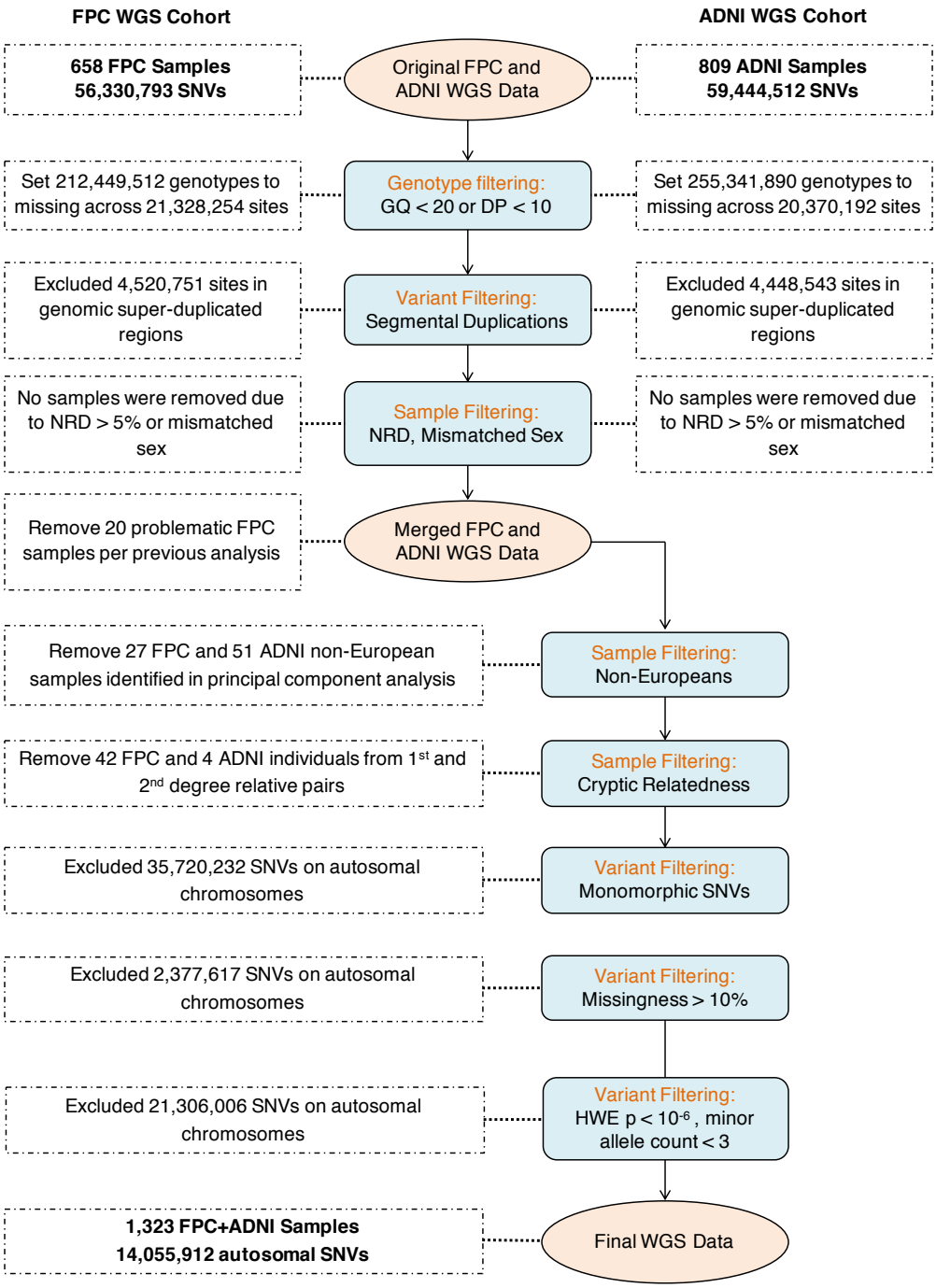


Figure 6 Overview of quality control procedures on FPC-ADNI WGS data. Circles with the solid line represent WGS data at different quality control stage. Boxes with the solid line represent filters applied in an order directed by arrows. Boxed with dashed line detail changes in WGS data resulting from the implementation of each filter. The bolded text emphasizes the number of samples and variants in WGS data at the beginning and the end of quality controls.

Genotype Filtering: GQ < 20 and DP < 10

Filtering on GQ scores was recommended for WGS analysis. CASAVA variant caller uses a Bayesian approach to update a prior genotype distribution based on the observed sequencing reads to produce a final predicted genotype distribution. This final genotype distribution is summarized as a set of quality scores (GQ). Briefly, at each site the probability of all possible genotypes was calculated based on the relative distribution of reads of different bases (i.e. A, T, C, G). Then a prior distribution based on a reference sequence was applied to adjust the genotype probabilities and to determine the GQ. Higher GQ values indicate higher confidence in the called genotypes. Filtering on GQ scores is highly recommended, but the thresholds are often arbitrary and vary by study.

Another filtering metric often used in sequencing data is read depth (DP) values. The DP values generated by CASAVA represent the total number of reads mapped to the site that passed read filtering (i.e. low-quality reads are dropped). It is not the number of reads used to call the most likely genotype, but the total number of reads used to determine the probability of all possible genotypes. No minimum DP threshold was enforced to make genotype calls. However, genotype calls with read depth greater than 3 times the chromosomal mean were removed to minimize potentially spurious calls that occur near the centromeres and other high-copy number regions. Like GQ values, DP thresholds are often arbitrary and vary by study.

GQ < 20 and DP < 10 were often applied in sequencing studies to exclude low-quality calls. To evaluate the efficiency of these two filters in our dataset, the true-variant rate and error rate were calculated in 11 pairs of duplicated FPC samples[5]. The true-variant rate is the proportion of concordant SNVs that remained in the dataset after applying the filter, whereas the error rate is the proportion of discordant SNVs that remained after applying the filter. The true-

variant rate and error rate were each averaged across 11 pairs of duplicated samples. At GQ < 20, 97.0% true variants and 2.5% errors were retained in the dataset. While at DP < 10, the true-variant rate and error rate was 96.9% and 8.7%. When both filters were applied, 96.1% true variants were retained, and the error rate was reduced to 1.3%. A higher cutoff value at GQ or DP may further lower the error rate at a high expense of removing a large number of true variant calls. Because the number of concordant genotype calls was considerably larger than the number of discordant genotype calls, the cutoff of GQ < 20 and DP < 10 would maximize the true variant calls at an acceptable error rate. Genotypes with GQ < 20 or DP < 10 were therefore set to missing, which affects about 38% and 33% variants in FPC WGS and ADNI WGS data, respectively.

Variant Filtering: Segmental Duplications

Previous studies have suggested that WGS has limited capacity to resolve the structure of duplicated regions within genomes, particularly those > 15 kb in length and show > 97% sequence identity[6]. Assembly of sequence reads from these duplicated regions is difficult and more prone to errors. Over the years a list of segmental duplications in the human genome that are > 1kb in length and > 90% identity has been constructed. The coordinates of these segmental duplications in GRCh37 reference human genome were obtained from the Segmental Duplication Database (http://humanparalogy.gs.washington.edu). SNVs located in these super-duplicated regions were excluded from our analysis, lowering the total number of SNVs by 4,520,751 (8.0%) in FPC WGS data and by 4,448,543 (7.5%) in ADNI WGS data (Figure 1).

Samples Filtering: Non-reference Discordance Rate (NRD)

A high non-reference discordance rate (NRD) is an indicator of problematic samples. For each sample in ADNI (N = 809) or FPC (N = 637) data, we calculated the NRD as the proportion

of non-reference genotypes differed between the WGS data and Omni2.5M array data. We limited this analysis to autosomal variants shared by the WGS data and Omni2.5M array data, excluding small insertion/deletions, SNVs with call rate < 0.95 and those with a MAF < 5% in Omni2.5M data. Concordant and discordant genotypes were identified using a PERL module in VCFtools (*vcf-compare*)[7]. Because sequencing errors in WGS data would affect the NRD calculation, we performed the analysis in the genome after removing the genotypes with GQ < 20 or DP < 10 and excluding variants in the super-duplicated regions.

Figure 7 showed the NRD per sample in the FPC WGS data and ADNI WGS data, respectively. The NRD rates in FPC WGS data ranged from 1.04% to 1.77% with a mean of 1.07% (Figure 7A). Sample FPC0713 with an NRD of 1.77% was distributed further away from the cluster of most samples but was considered acceptable (< 5%). The distribution of NRD rates in the ADNI WGS data was quite similar to that in the FPC WGS cohort, with a slightly higher mean of 1.11% and a slightly narrower range between 1.06% and 1.46% (Figure 7B). One sample 12AD48485 in the ADNI WGS data was separated from the main cluster but was considered acceptable as well.

All samples in FPC WGS and ADNI WGS data had NRD < 2%. Therefore, no sample in either cohort was removed due to an NRD > 5%.

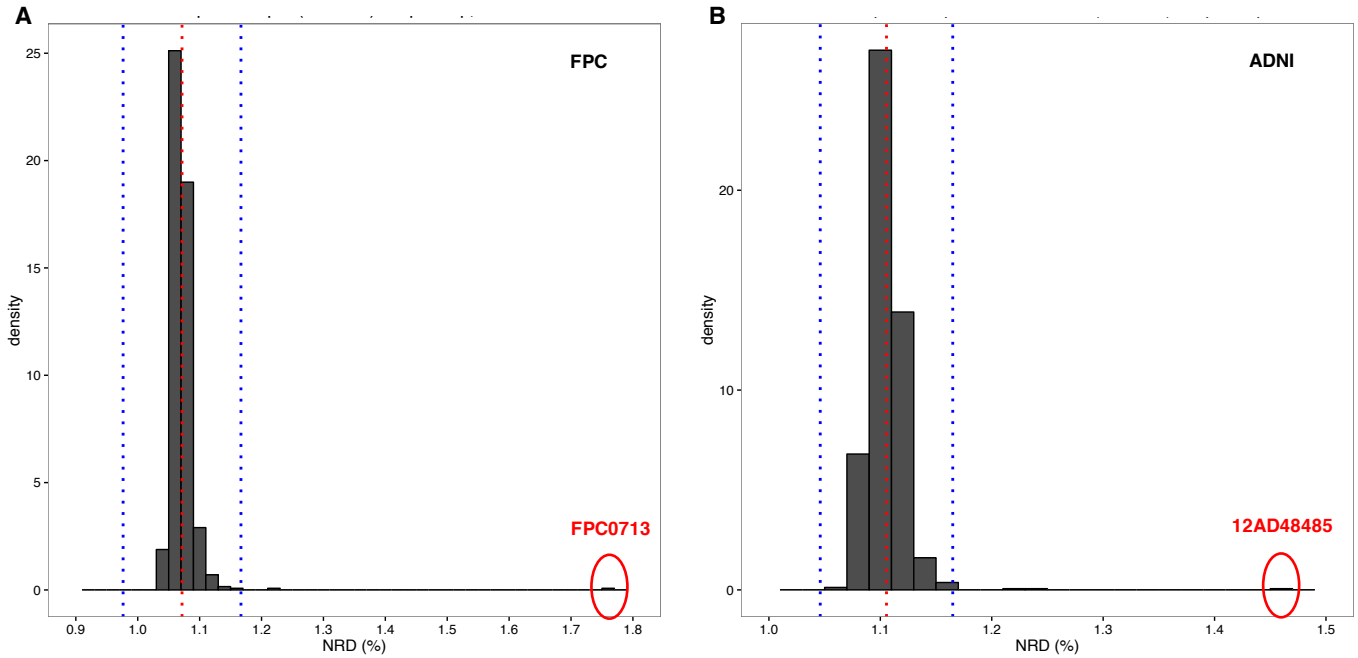


Figure 7 Distributions of non-reference discordance rate (NRD) per sample in (A) FPC WGS and (B) ADNI WGS data. The red dotted lines represent the mean NRD average across all samples in each data. The blue dotted lines represent the NRD percentage ± 3 standard deviations from the mean. Samples circled and marked with their ID in each data had an NRD significantly higher than the remaining samples.

Sample Filtering: Mismatched Sex

Mismatched sex is another indicator of problematic samples. We performed a sex check analysis using PLINK (*--check-sex ycount*)[8], calculating the X chromosome inbreeding coefficient F to determine the sex of each sample in ADNI WGS and FPC WGS data. We used bi-allelic SNVs on the sex chromosome with MAF > 5% for this analysis, excluding variants in the pseudo-autosomal regions *(--split-x*). Linkage disequilibrium (LD) based pruning was applied to these variants at a pair-wise r^2 of 0.2.

A total of 11,638 variants on X chromosome and 148 variants on Y chromosome were used to determine sex in the FPC WGS data. There were 326 females and 312 males in this cohort according to pedigree information. Based on the empirical distribution of F estimates in FPC WGS data, we set the threshold to make female calls to 0.50. This threshold flagged two samples, FPC0145 and FPC0347, since their F estimates were > 0.80. However, all female samples in FPC WGS data had zero Y genotypes, including FPC0145 and FPC0347. The count of non-missing Y genotypes in males had a wider range than in females. After considering the sex check results from Illumina, and the non-missing Y genotype counts in our analysis, we believed that the female calls on FPC0145 and FPC0347 were justified despite their high F values.

A total of 8,501 variants on X chromosome and 55 variants on Y chromosome were used to determine sex in the ADNI WGS data. There were 446 males and 363 females in the ADNI WGS data. Estimated sex given genomic data matched the reported sex in pedigrees for all samples. In conclusion, no sample was removed due to mismatched sex in either dataset.

Merging of FPC WGS and ADNI WGS Data

We merged the FPC WGS and ADNI WGS data after applying the genotype filtering at GQ < 20 or DP < 10 and the variant filtering on segmental duplications. Evaluation of NRD rate and genetically determined sex in both FPC and ADNI samples didn't remove any sample. Before merging, we evaluated the comparability of these two datasets based on two quality metrics: site mean depth and site mean GQ.

We presented the results from chromosome 22 to illustrate the details of this analysis. We first compared the FPC WGS data and ADNI WGS data to identify shared and private SNVs using BCFtools[9]. Variants were matched by positions, reference allele and at least one alternative allele. On chromosome 22, FPC WGS data contains 598,739 SNVs and this number for ADNI WGS data was 645,982 (Figure 8). A total of 335,307 SNVs were shared by these two datasets, leaving 263,432 private SNVs in FPC WGS and 310,675 private SNVs in ADNI WGS. Majority of these SNVs were bi-allelic. The genotype filtering on GQ and DP had caused some SNVs to become monomorphic for the reference allele. After excluding the monomorphic sites, 280,917 SNVs on chromosome 22 were shared between these two datasets (Figure 8).

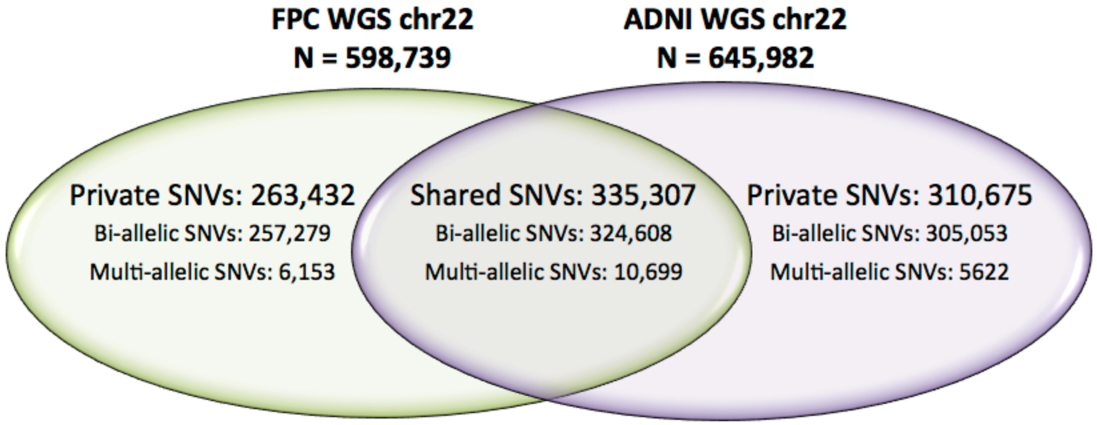
The site mean depth was defined as DP values of non-reference genotypes averaged across samples for each site. Non-reference genotypes with GQ < 20 or DP < 10 were set to missing and not included in the calculation of site mean depth. Among the 280,917 shared SNVs, the distributions of site mean depth in FPC WGS, ADNI WGS and the merged dataset were quite comparable (Figure 8). The absolute difference in site mean depth averaged across all shared SNVs was 4.83. Although the majority of the shared SNVs had an absolute difference in site mean depth ≤ 6, extremely large differences were observed in some SNVs, particularly those with low allele frequency. We plotted the difference in site mean depth of shared SNVs against the alternative allele frequency (Figure 9). The variability of difference in site mean depth was

the largest at the lower end of the alternative allele frequency (AAF). We also plotted the distributions of site mean depth by AAF groups (Figure 10). The distribution of site mean depth was more spread out in singletons and doubles (i.e. AC=1/2) than in other AAF groups. But the means and medians were very consistent across all AAF groups and were comparable in FPC WGS, ADNI WGS and the merged dataset. To evaluate the comparability of all sites on chromosome 22, not just the shared sites, we calculated the site mean depth averaged across all sites in 10-kb segments. Note that the number of SNVs located in the same 10-kb segment in FPC WGS was different from that in ADNI WGS since certain SNVs were private to the specific dataset. Then we calculated the difference in average site mean depth for each 10-kb segment between FPC WGS and ADNI WGS (Figure 11). The data suggested that the average site mean depth per 10-kb segment was consistently higher in FPC WGS than in ADNI WGS across the genomic positions on chromosome 22. But these differences were small enough to be acceptable. The same analysis demonstrated adequate comparability in all other autosomal chromosomes (data not shown).

The site mean GQ was defined as the GQ values of non-reference genotypes averaged across all samples. Genotypes excluded due to GQ < 20 or DP < 10 were not considered in the calculation. We conducted the same analysis in site mean GQ as in site mean depth, which led to the same conclusion that distributions of site mean GQ on chromosome 22 as well as other autosomal chromosomes were adequately comparable in FPC WGS, ADNI WGS and the merged dataset (data not shown).

Once the comparability of site mean depth and site mean GQ between FPC WGS and ADNI WGS was demonstrated, these two datasets were merged using BCFtools [9]. There were 20 FPC samples removed due to being duplicates or non-FPC diagnosis [10]. The merged dataset

contained 1,447 samples sequenced on 72,404,844 SNVs across all autosomal chromosomes. In the merged dataset, multi-allelic sites were split into multiple rows using BCFtools (*norm*)[9].



Distribution of Site Mean Depth (n = 280,917)

	FPC	ADNI	Abs. Difference	Merged
Min	10.0	10.0	0.0	10.0
25%	30.4	27.6	1.88	29.5
50%	37.0	33.8	3.32	35.4
75%	41.6	38.5	6.00	39.7
Max	119.0	111.0	91.0	91.2
Mean	35.5	32.7	4.83	33.9

Distribution of Site Mean GQ (n = 280,917)

	FPC	ADNI	Abs. Difference	Merged
Min	20.0	20.0	0.00	20.0
25%	119.0	110.0	6.5	102.3
50%	147.0	136.7	11.9	118.0
75%	168.0	157.8	24.3	130.5
Max	406.0	346.0	283.0	278.0
Mean	140.4	131.3	18.89	113.5

Figure 8 Comparison of FPC WGS and ADNI WGS data on chromosome 22. The green circle represents the SNVs in FPC WGS data while the purple circle represents the SNVs in ADNI WGS data. The overlap area of these two circles represents the SNVs shared by these two datasets. Distributions of site mean depth and site mean GQ were obtained from 280,917 shared SNVs that were polymorphic after merging.

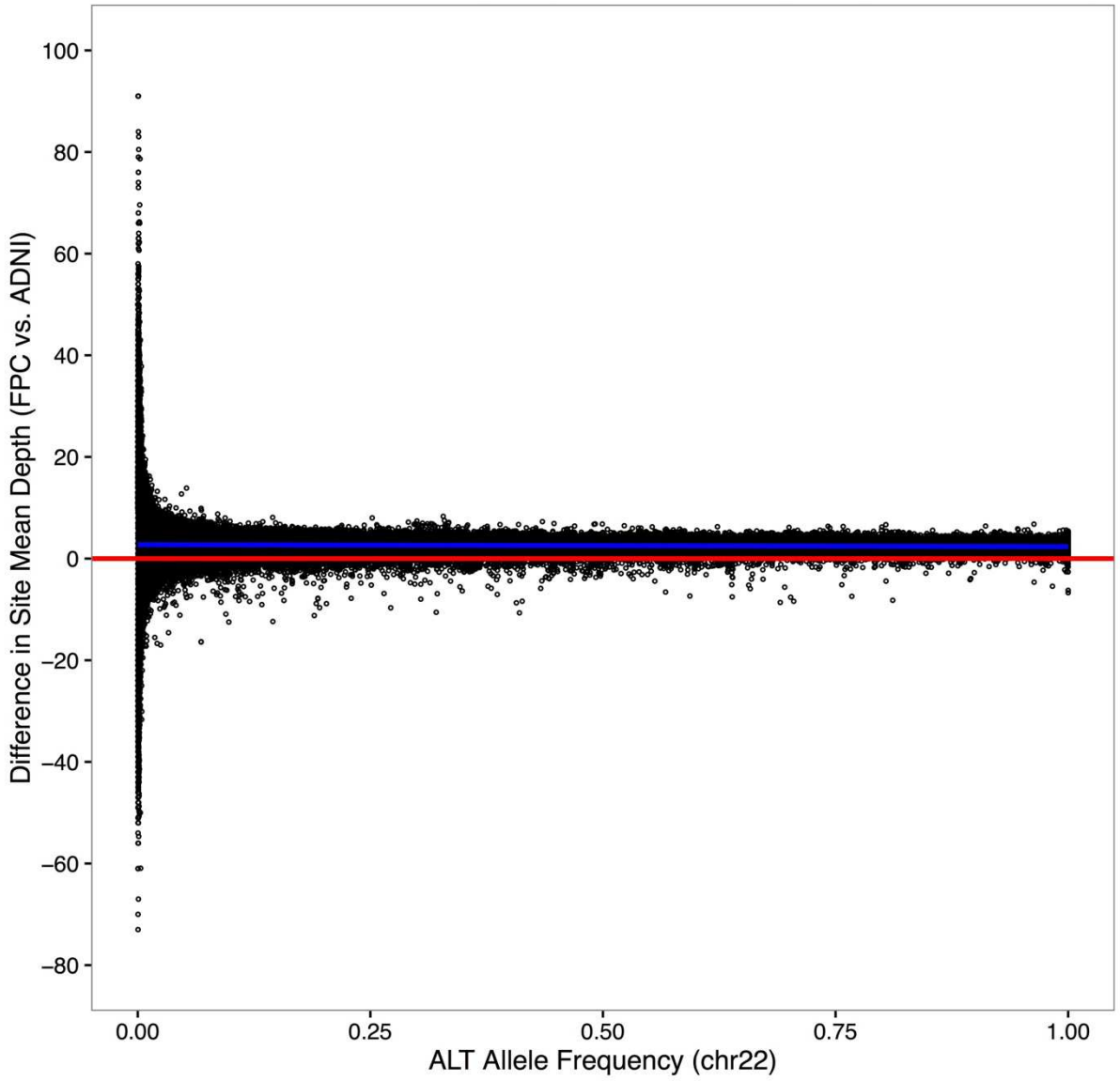


Figure 9 Differences in site mean depth by alternative allele frequencies on chromosome 22. The differences in mean depth per site were calculated among 280,917 SNVs shared FPC WGS and ADNI WGS data. The blue line indicates the difference in site mean depth (site mean depth in FPC WGS minus site mean depth in ADNI WGS) average across all sites. The red line marks the zero. The alternative allele frequency of each site was determined in the merged dataset.

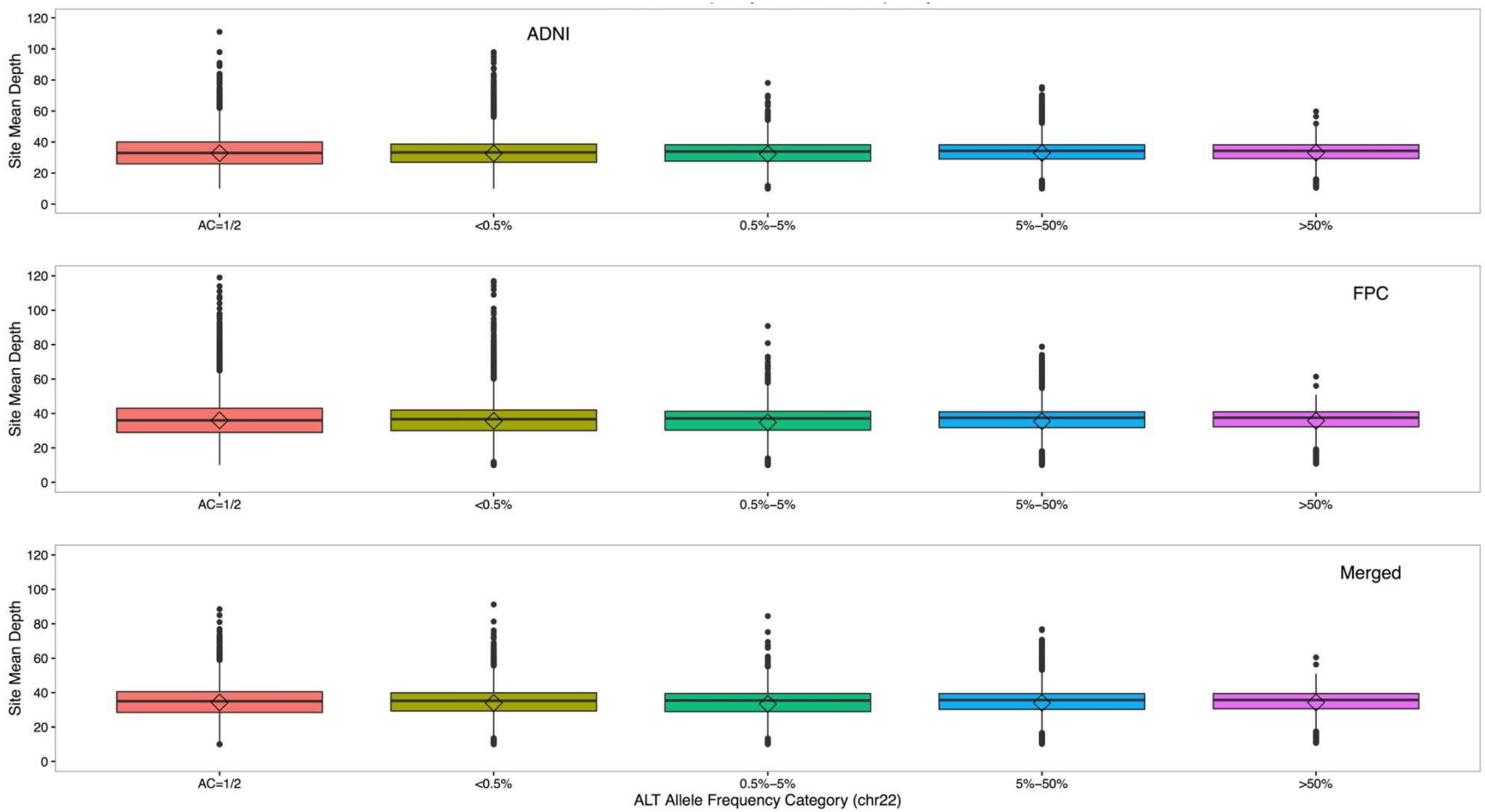


Figure 10 Site mean depth by alternative allele frequency groups on chromosome 22. The alternative allele frequencies of 280,917 SNVs shared by FPC WGS and ADNI WGS data were calculated in the merged dataset. For multi-allelic SNVs, the frequencies were summed over all alternative alleles. The diamond shape in each boxplot marks the site mean depth average across all SNVs in each alternative allele frequency group.

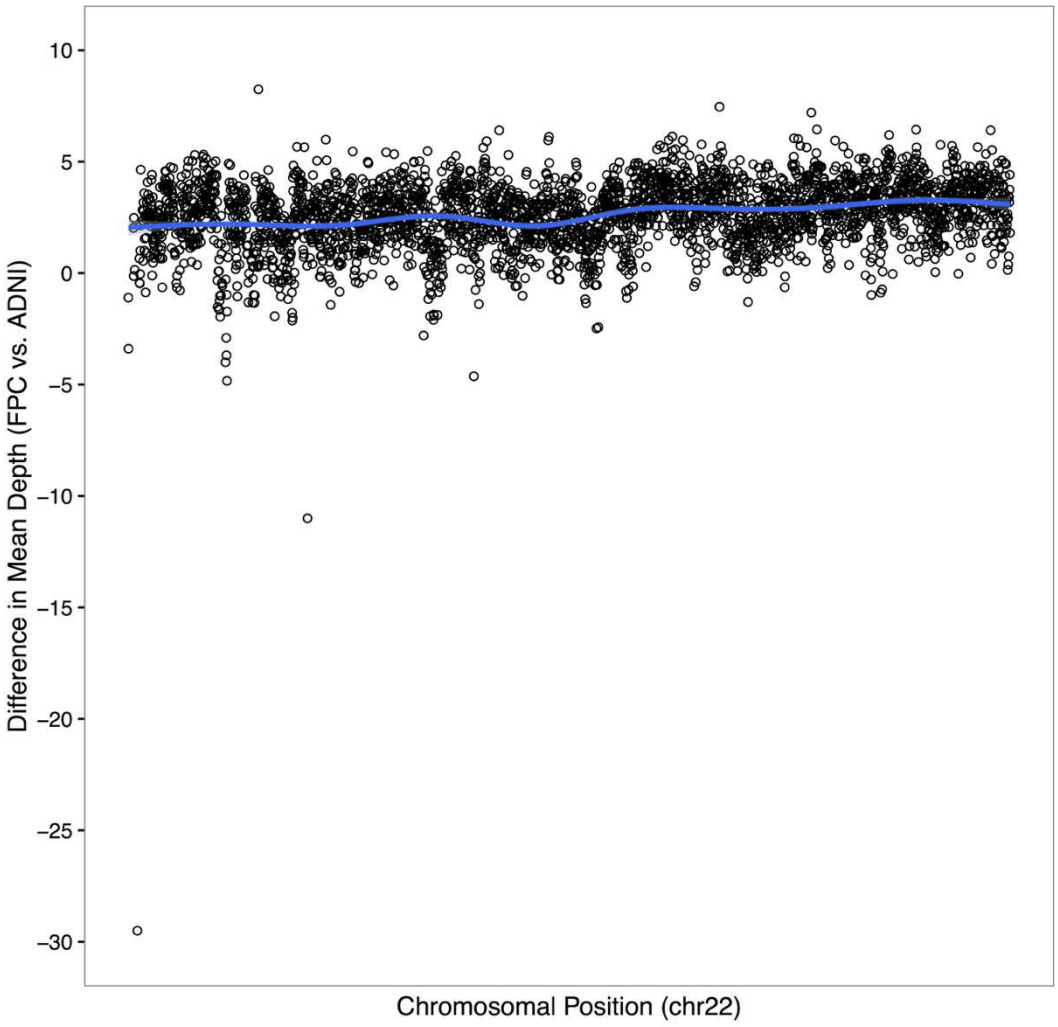


Figure 11 Differences in average site mean depth per 10-kb segment on chromosome 22. Site mean depth was averaged across all sites, shared and private, in each 10-kb segment. The differences were calculated comparing average site mean depth per 10-kb segment in ADNI WGS to that in FPC WGS. The blue line is the regressed line based on all data points.

Sample Filtering: Population Substructure

Both the FPC WGS and ADNI WGS data were self-reported to be predominately Caucasians. To identify samples of non-European ancestry, we performed principal component analysis (PCA) using reference populations from HapMap phase III. The HapMap phase III populations included two populations of European ancestry (i.e. CEU and TSI), three East Asian populations (i.e. CHD, CHB and JPT), four African populations (i.e. ASW, LWK, MKK and YRI), one South Asian population (i.e. GIH) and one Hispanic population (i.e. MEX). The genotyping data were downloaded from the International HapMap Project's FPT site (ftp://ftp.ncbi.nlm.nih.gov/hapmap/). Prior to the PCA, we excluded SNVs with any missing calls in any HapMap population, and converted the positions of the remaining SNVs in hg18/build36 to hg19/build37 using liftOver from UCSC Genome Browser[11]. Given that inferring genetic ancestry can be more challenging when family structure and cryptic relatedness is present, we only kept the founders in these HapMap populations, leaving a total of 988 samples to enter the PCA analysis. The merged FPC and ADNI WGS dataset were converted from VCF format to the binary PLINK format before the PCA analysis. During this conversion, 48,021,112 bi-allelic SNVs with at least 1 alternative allele were included.

SNVs for PCA were selected in the following steps. We first identified 336,308 common (MAF > 5%) SNVs on autosomal chromosomes shared across all HapMap populations as well as the merged dataset. Then we excluded 2,476 SNVs in high linkage disequilibrium (LD) regions and regions known to be associated with ethnicity, including the lactase regions on chromosome 2, the HLA regions on chromosome 6, and the polymorphic inversion regions on chromosome 8 and 17. LD-based pruning was conducted on the remaining SNVs at a pairwise $r^2 < 0.20$ to ensure the independence of variants. A total of 63,238 SNVs were used in PCA.

PCA was carried out using SMARTPCA in EIGENSTRAT [12]. We conducted two sets of PCA analysis on (1) the merged study population along with all HapMap phase III populations, and (2) the merged study population only. Genetic ancestry was determined based on tight clustering of study samples on the first two principal components (PCs) in the first set of analysis. In the second set of PCA analysis including only sample populations, samples with one of the first two PC values that exceeded ±1 standard deviation from the mean were considered as outliers.

The majority of the FPC and ADNI samples were clustered alongside the CEU and TSI populations, suggesting European ancestry (Figure 12A). We did, however, identify 27 FPC samples and 51 ADNI samples as outliers based on their first two PC values (Figure 12B). Among these samples, 44 were clustered closely with African populations (27 ADNI and 17 FPC) and 13 were East Asians (8 ADNI and 5 FPC). The remaining samples were inferred as Mexicans or South Asians. These 78 samples were removed from the merged dataset.

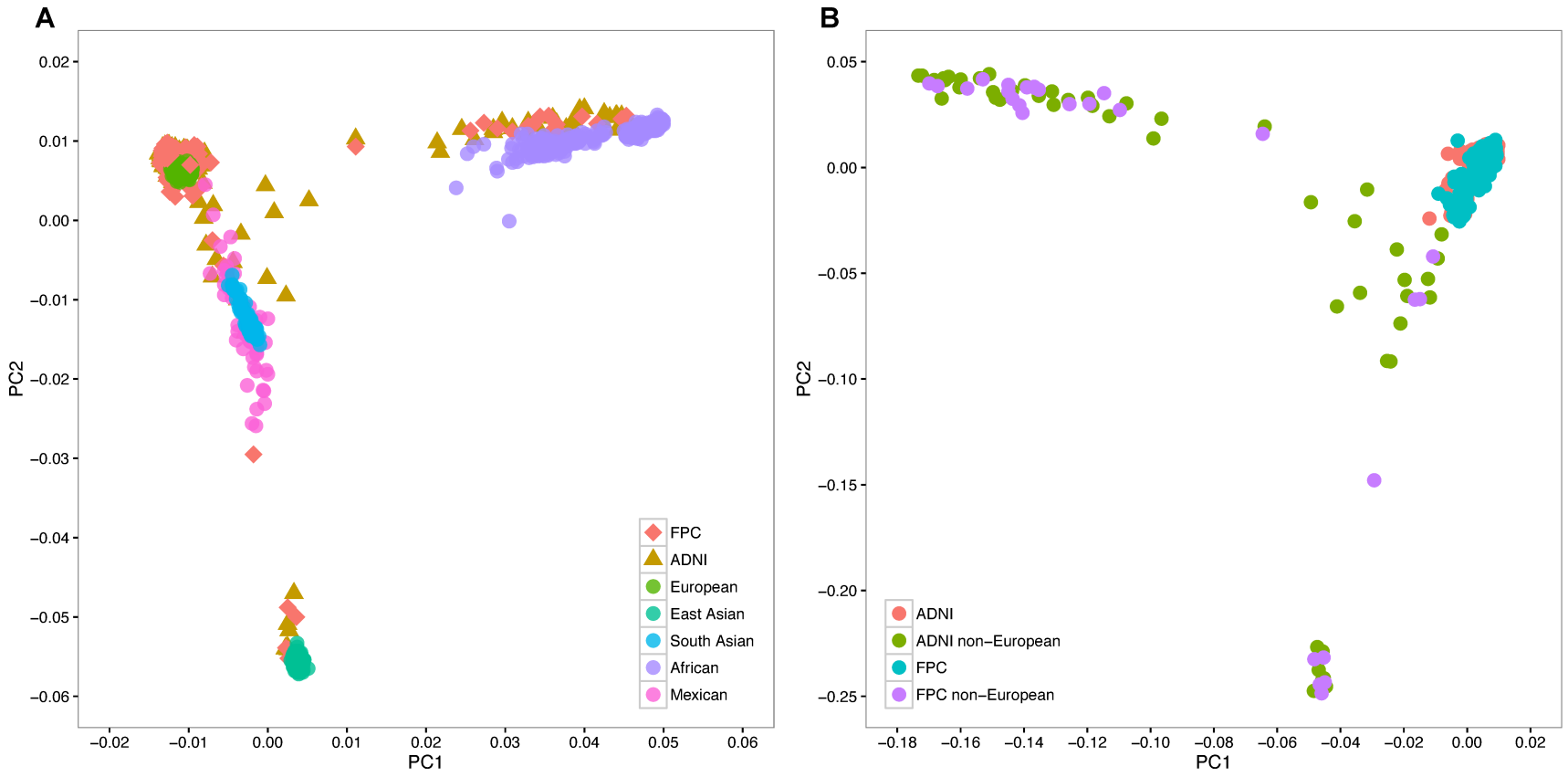


Figure 12 Principal component analysis (PCA) in the merged study population. In the first set of analysis (A), FPC samples (diamonds) and ADNI samples (triangles) were clustered alongside the HapMap phase III populations (circles). There were 44 samples clustered closely with African populations, 13 with East Asians and 21 around Mexicans or South Asians. These 78 non-European samples were marked by green and purple circles in plot (B) and were excluded from the subsequent analysis.

Sample Filtering: Relatedness

Heritability estimates in case-control studies are particularly sensitive to hidden relatedness in the study population. Including close relatives would result in the estimate of genetic variance being driven by the phenotypic correlations for these pairs, and lead to inflation possibly due to common environmental effects. To identify close relatives in the data, we first conducted an identity-by-descent (IBD) analysis separately in FPC WGS and ADNI WGS data using PLINK[8]. The genetic similarity was later assessed using GCTA[13] to determine the optimal cutoff that would eliminate the relatedness in the study population while maximizing the sample size. Prior to this analysis, 78 individuals identified as non-Europeans in the PCA analysis were excluded to minimize the bias due to population substructure, leaving 1,369 samples in the merged dataset.

First- and second-degree relatives were identified using a selected panel of SNVs in FPC WGS and ADNI WGS cohort separately. Variants were selected for IBD analysis if they were bi-allelic and independent (pairwise $r^2 < 0.10$) on autosomal chromosomes with MAF > 5%. The IBD analysis using PLINK calculated the probability of sharing 0 (Z_0), 1 (Z_1), and 2 (Z_2) alleles at each site. By comparing the estimated Z_0, Z_1, and Z_2 values to the expected, we were able to identify pairs of parent-offspring, full-siblings, half-sibling/grandparent-grandchild/avuncular, and first cousins in each cohort. In FPC WGS data, a total of 410,722 SNVs were selected for IBD analysis. We identified 3 parent-offspring pairs, 25 full-siblings, 12 first cousins and 8 pairs of other second-degree relatives (i.e. half-sibling, grandparent-grandchild or avuncular) (Figure 13A). The inferred genetic relationships of these related pairs matched the reported pedigree relationships. In ADNI WGS data, a total of 412,572 SNVs were used in the IBD analysis. Four

pairs of full-siblings were identified (Figure 13A). Pedigree information of ADNI WGS data was not available to us.

We also evaluated the relatedness using GCTA [13]. The genetic similarity for each pair of samples was calculated using variants shared by the pair. We limited this analysis to 320,379 bi-allelic SNVs with MAF > 5% on autosomal chromosomes. The relatedness value calculated from GCTA ranged from -0.025 to 0.057 (Figure 13B). We later considered two relatedness cutoffs 0.025 and 0.05 in our heritability analysis. By removing one individual from the first- and second-degree relative pairs identified in the IBD analysis, the final sample consists of 1,323 individuals.

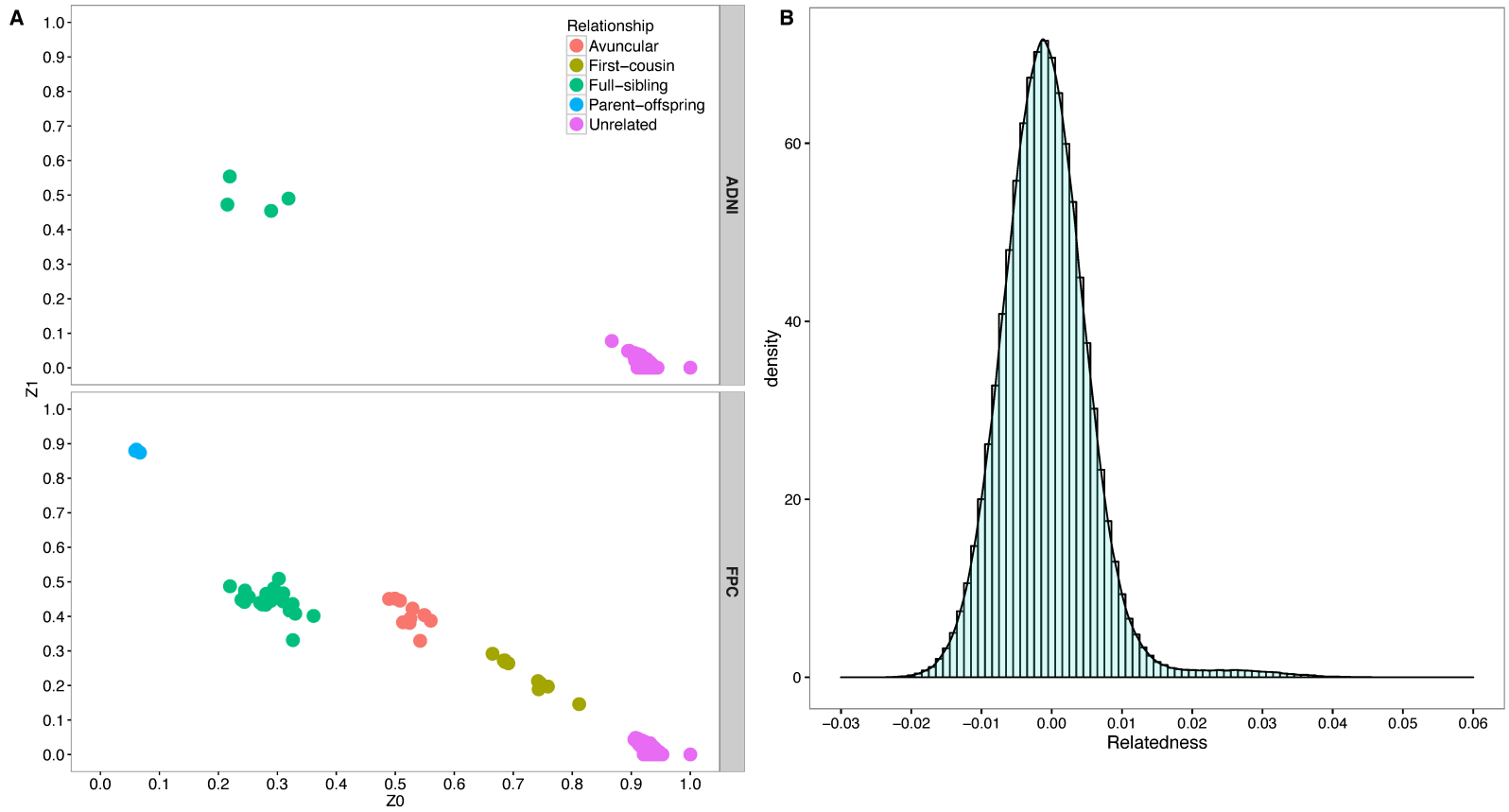
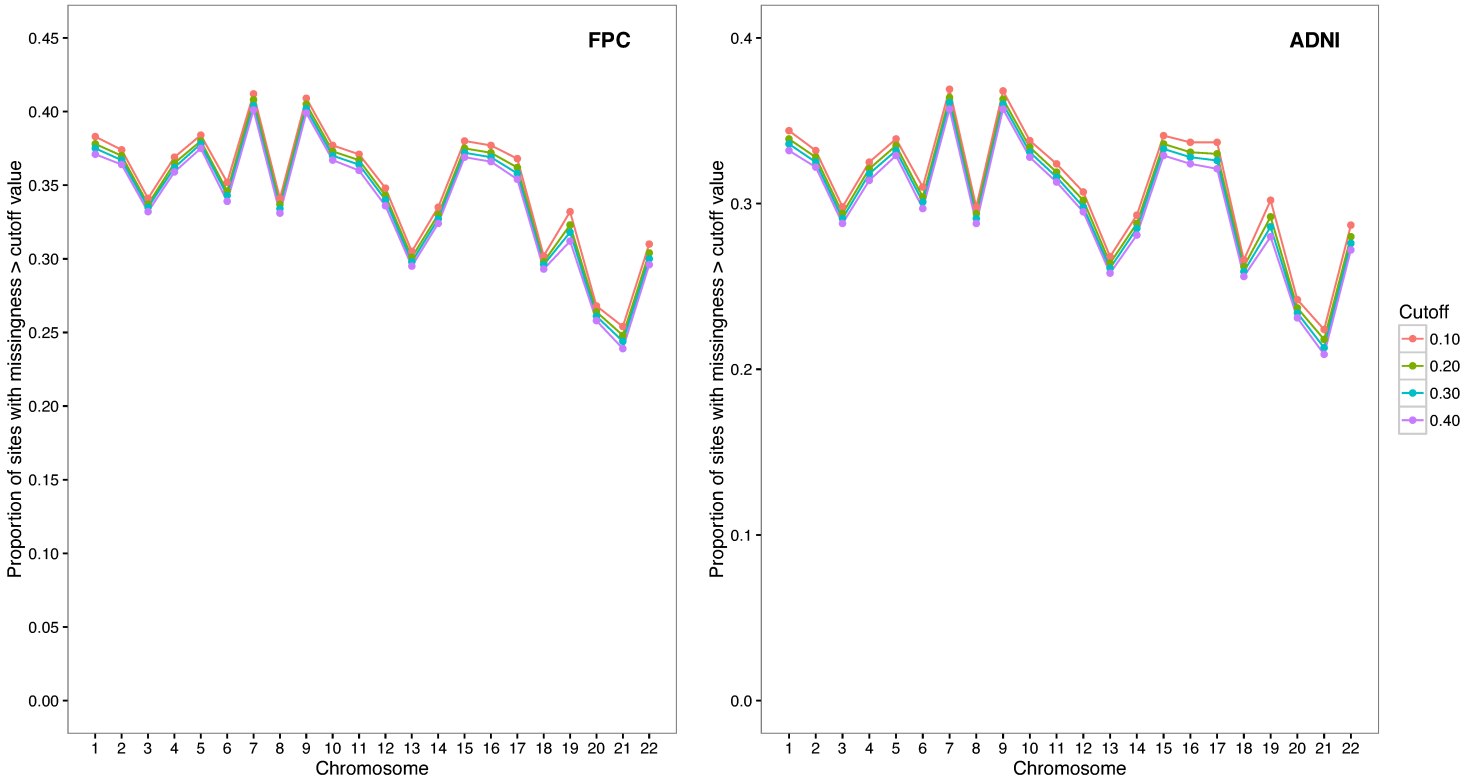


Figure 13 Z_1 versus Z_0 in identity-by-descent analysis (A) and distribution of genetic similarity values (B). Pairs of related and unrelated samples in ADNI WGS data (A above) and FPC WGS data (A below) were plotted. The genetic similarity values in (B) were estimated for samples in the merged dataset, after randomly excluding one individual from each first- and second- degree relative pairs.

Additional Site Filtering

The removal of genotypes and samples inevitably caused some variants to be homozygotes of either reference or alternative allele. About half of the SNVs (35,720,232) in the merged dataset were monomorphic and hence excluded. Missingness per site was defined as the proportion of called non-reference genotypes that were set to missing due to GQ < 20 or DP < 10. A high missingness raises concern with the poor quality of sequencing or variant calling. About 37% and 33% of variants in FPC and ADNI dataset contain missing calls, the majority of which had lost heterozygosity completely, particularly rare variants like singletons and doubletons. We assessed the proportion of variants on each chromosome with missingness greater than 10%, 20%, 30% and 40% (Figure 14). Since the differences were small across these cut-off values, a conservative value of 10% was chosen to minimize the impact of poor quality variants in the subsequent analysis, excluding an additional 2,377,617 variants. We also exclude variants that failed Hardy-Weinberg Equilibrium (HWE) exact test at $p < 10^{-6}$ (N = 51,550) and variants with minor allele count < 3 (N = 21,254,456). Variants with extremely low frequencies would cause noise and lower the precision of heritability estimates and hence were excluded from the analysis.

Figure 14 Proportion of variants with missingness greater than 10%, 20%, 30%, and 40% on each chromosome. Missingness was defined as the proportion of called genotypes that was set to missing due to GQ < 20 or DP < 10. The pattern of missingness cross chromosomes was similar in FPC and ADNI WGS dataset. A conservative cut-off value of 10% was chosen to minimize the impact of poor quality variants in the analysis.



Statistical Analysis

The analysis in this chapter is aimed to estimate the heritability for pancreatic cancer due to additive genetic effects from sequenced variants (h^2_{WGS}). The small sample size of FPC-ADNI study prohibits any stratification on MAF and/or LD. Therefore, GREML-SC approach was used to estimate the heritability. Comparisons of estimated heritability were conducted using a subset of variants shared by FPC-ADNI and PanC4 study.

Estimate of Heritability Explained by Sequenced Variants

A detailed description of the GREML approaches can be found in Chapter One. Using the GREML-SC approach, the variance due to additive genetic effects of all sequenced variants was estimated by fitting a single GRM calculated from all autosomal SNVs in a mixed linear model. The proportion of variance explained by all autosomal SNVs was defined as $\boldsymbol{h^2_{WGS}} = \boldsymbol{\sigma^2_{WGS}}/(\boldsymbol{\sigma^2_{WGS}} + \boldsymbol{\sigma^2_e})$. In this model, age, sex and first 20 principal components (PCs) were included as fix effects. The first 20 PCs were calculated using 320,379 common (MAF > 0.05) and independent (pairwise $r^2 > 0.20$) variants sequenced in the dataset. The default average information algorithm was applied to run REML iterations in all heritability analysis. A disease prevalence (K) of 0.0149 was used to correct for the ascertainment bias in the case-control cohort. Given that FPC cases constitute up to 10% of all pancreatic cancer cases and the true prevalence of FPC was unknown, the heritability was estimated with the prevalence varying from 0.1K to K. The estimates of variance were transformed from the observed 0-1 scale to the unobserved continuous 'liability' scale using a probit linear transformation[14]. The variance in liability scale was reported along with its standard errors. This analysis was conducted at relatedness cutoff of 0.05 and 0.025 to evaluate the impact of relatedness cutoffs on heritability estimates.

Comparing Estimates of Heritability in FPC-ADNI and PanC4 Study

Comparisons of estimated heritability for pancreatic cancer were performed using a subset of variants presented in both FPC-ADNI and PanC4 study. To minimize the impact of poor imputation in PanC4 data as well as the potential bias due to batch effects in FPC-ADNI data, this analysis was limited to variants that were directly genotyped in both datasets. A relatedness cutoff of 0.025 was applied to both datasets to minimize the potential bias due to residual relatedness. Heritability for pancreatic cancer was estimated with a disease prevalence (K) of 0.0149 in the PanC4 study and a disease prevalence varying from 0.1K to K in the FPC-ADNI study. Both GREML-SC and GREML-MS approach were used to estimate heritability. In the GREML-MS analysis, variants were stratified at MAF of 0.05, and the two variance components were fitted jointly in an MLM. Because of the small sample size of FPC-ADNI study, GREML-LDMS does not provide reliable estimates or convergence is not obtained. The variance in liability scale was reported along with its standard errors.

Results

The final dataset contains 1,323 samples (569 cases and 754 controls) sequenced on 14,019,172 autosomal sites, including 36,619 sites with multiple alternative alleles. After splitting the multi-allelic variants into multiple records, the final dataset consists of 14,055,912 SNVs.

A shows the distribution of MAF in the final dataset. The majority of the variants are rare and low-frequency variants with a MAF < 5%. The remaining variants are evenly distributed across the MAF categories. All variants in the final dataset were annotated by ANNOVAR[15]. The functional predictions were derived from the NCBI Reference Sequence Database. Among all variants in the final dataset, 54.2% of them were located in the intergenic regions and 37.0% were intronic. About 0.83% of the sequenced variants were coding variants (

B).

Characteristics of ADNI and FPC WGS samples at baseline are shown in Tables Table 7 and Table 8, respectively. In general, these two study populations are quite similar in terms of the age distribution and sex ratio. Individuals in both studies are primarily of European ancestry as confirmed by PCA analysis. The ADNI WGS study consists of 34.6% individuals with normal cognitive functions, 29.0% with early mild cognitive impairment and 30.4% with late cognitive impairment. The remaining 5.9% had a diagnosis of Alzheimer's' Disease (Tables

Table 7). Among the FPC cases, more than half had two relatives affected by pancreatic cancer; 30.7% had three and 13.2% had four or more relatives affected by the disease.

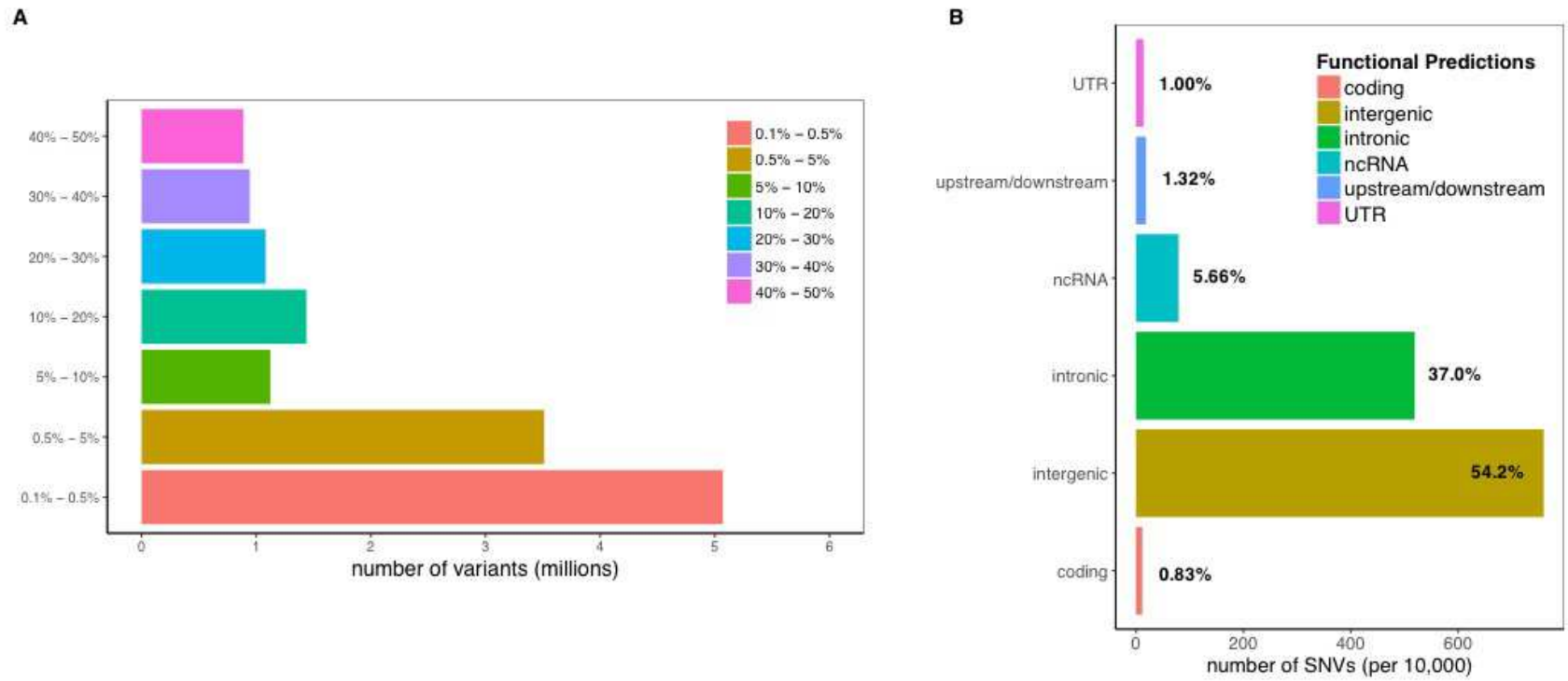


Figure 15 Minor Allele Frequency (A) and Functional Annotation (B) Distribution of WGS Variants.

Estimates of Heritability Explained by Sequenced Variants

The estimates of heritability explained by sequenced variants (h^2_{WGS}) varied by the specified population prevalence of pancreatic cancer (Table 9). Applying a relatedness cutoff further excluded distantly related individuals in our study population and hence reduced the sample size. At a relatedness of 0.025 and a sample size of 1,180 individuals, the heritability for pancreatic cancer explained by sequenced variants (h^2_{WGS}) was estimated to be 38.8% (s.e. = 15.7%), 44.0% (s.e. = 17.9%), 53.2% (s.e. = 21.6%) and 62.6% (s.e. = 25.4%) with a disease prevalence of 0.00149 (0.1K), 0.00298 (0.2K), 0.00745 (0.5K) and 0.0149 (K), respectively. Results under a relatedness cutoff of 0.05 were quite similar (Table 9).

Comparing Estimates of Heritability in FPC-ADNI and PanC4 Study

The comparison between the estimated heritability in FPC-ADNI and Panc4 study was performed using 521,282 polymorphic variants directly genotyped in both datasets. The MAF distributions of these variants are similar in FPC-ADNI and PanC4 dataset (

Supplementary Figure 2). Comparing to the full set of sequenced or imputed variants, this subset of variants has a higher proportion of common variants; over 94.4% of the 521,282 variants have a MAF > 1%. In both datasets, the relatedness cutoff of 0.025 was applied. Since one FPC case does not have array data, in total 1,179 individuals in the FPC-ADNI cohort were included in this analysis. Using the GREML-SC approach, the heritability for pancreatic cancer estimated from this subset of variants was 16.6% (s.e. = 3.3%) in the PanC4 study. In the FPC-ADNI study, the phenotypic variance for pancreatic cancer attributable to this subset of variants was 30.9% (s.e. = 11.2%) when the disease prevalence was set to 0.00149 (0.1K) and was 50.0% (s.e. = 18.1%) when the disease prevalence was 0.0149 (K). Results from the GREML-MS

analysis was similar in the PanC4 study (h^2 = 16.2%, s.e. = 3.3%) but were slightly higher in FPC-ADNI study. Specifically, when a disease prevalence of 0.00149 (0.1K) was specified, the heritability was estimated to be 32.1% (s.e. = 11.2%); and under a disease prevalence of 0.0149 (K), the estimate of heritability was 51.9% (s.e. = 18.0%) (Table 10). As expected, the estimates of heritability using this subset of variants were lower than those from the full set of imputed or sequenced variants as in Table 9 and Chapter Three.

Discussion

In this combined sample of FPC cases and ADNI participants, the heritability for pancreatic cancer from all sequenced variants was estimated to range from 38.8% (s.e. = 15.7%) to 62.6% (s.e. = 25.4%) depending on the specification of the population prevalence of pancreatic cancer (K). Misspecification of K by one factor or two could result in a 20% bias on the estimated heritability, especially when the disease is common (K >= 0.1)[14]. Unlike the PanC4 study where cases were selected regardless of the family history, affected individuals in FPC-ADNI study were all familial cases with at least two FDRs with pancreatic cancer. Since the true prevalence of FPC was unknown, four disease prevalence values, 0.0149 (K), 0.00745 (0.5K), 0.00298 (0.2K) and 0.00149 (0.1K), were evaluated to quantify the influences of K on the estimates of heritability in our study. The wide range of K specifications led to a huge difference in the estimates of heritability. Given that up to 10% of pancreatic cancer cases had a family history of the disease, the true value of heritability for pancreatic cancer is likely to be near the estimated heritability under a disease prevalence of 0.1K, where 38.8% (s.e. = 15.7%) of the phenotypic variance for pancreatic cancer was explained by all sequenced variants in the FPC-ADNI study.

However, this estimate of heritability from sequenced variants may be biased downwardly. Based on the simulation studies, estimates of heritability from the GREML-SC approach is unbiased only when the causal variants are a random subset of all sequenced variants[16]. Results from our genomic partitioning analysis in the PanC4 study (Chapter Three) suggested a distribution of causal variants on the lower end of the MAF spectrum. Moreover, the estimate of heritability in the PanC4 study has increased considerably after stratification by MAF and LD (21.2% vs. 16.0%). Given these findings in the PanC4 study, the estimate of heritability from all sequenced variants (h^2_{WGS}) using GREML-SC approach in FPC-ADNI is likely biased downwardly.

Bias in heritability estimates can also result from the exclusion of extremely rare variants. WGS data with sufficient coverage (30X) provides a near-complete characterization of genetic variations across the genome. Comparing to imputation, WGS had advantages in capturing extremely rare variants that are monomorphic in reference populations. But these extremely rare variants are often excluded from the heritability analysis to allow a reliable estimate of variance using the GREML approach. In our WGS data, more than 60% of the sequenced variants were excluded at minor allele count < 3. As a result, the genetic variance due to these variants can never be captured using current statistical analysis and the current sample size. This is particularly concerning for a cancer phenotype since rare mutations are known to play an important role in the susceptibility of most cancer types.

The estimate of heritability for pancreatic cancer from sequenced variants (h^2_{WGS})) in the FPC-ADNI study is significantly higher than the estimate from imputed variants ($h^2_{imputed}$)) in the PanC4 study (38.8% vs. 21.2%). There are three possible explanations for the difference in heritability estimates. First, the estimate of heritability in the FPC-ADNI study is inflated due to

batch effects. In the FPC-ADNI WGS data, the cases and controls were sequenced and called in different pipelines. In the presence of technical artifact, the allele frequencies between cases and controls on average will be more different than expected under the null hypothesis, and cases would appear to be more related to other cases than to controls, leading to inflated estimates of heritability. To minimize the impact of batch effects, stringent QC filters were implemented on the WGS data. In addition, the comparability of WGS data was carefully assessed before merging the FPC and ADNI datasets. Variants sequenced in the FPC and ADNI study overall showed very similar distributions of mean depth per site and mean GQ per site across the genome. The differences in these two quality metrics between the case and control samples tend to be more variable among singleton and doubleton variants, which were excluded from the heritability analysis. Although QC filters are not able to fully address the batch effects, the influence of batch effects on the estimates of heritability is likely to be small in the FPC-ADNI study. To mitigate the concern of batch effects, we also compared the estimate of heritability from 521,282 directly genotyped variants shared by the FPC-ADNI and PanC4, which also suggest a higher heritability for pancreatic cancer in the FPC-ADNI study than in the PanC4 study (32.1% vs. 16.2%). Since array data is in general less prone to technical artifact than sequencing data, the impact of batch effects on these estimates is likely to be trivial.

The second explanation for a higher estimate of heritability in the FPC-ADNI study would be the bias due to misspecification of the population prevalence of disease. Pancreatic cancer cases in FPC-ADNI study were exclusively familial cases with at least 2 FDRs affected by the disease. Currently, there is no report of the prevalence of FPC in the general population. It is estimated that up to 10% of all pancreatic cancer cases had a family history of pancreatic cancer. When limited to FPC cases, the prevalence might be even smaller. Therefore, a disease

prevalence of 0.00149, which is equivalent to 1/10 of the population prevalence of pancreatic cancer, may still be an overestimate of the true prevalence of FPC. Under a disease prevalence of 0.000745 (i.e. 1/20 of the population prevalence of pancreatic cancer), a conservative estimate of heritability from all sequenced variants in FPC-ADNI study was 34.6% (s.e. = 14.1%), which is still higher than the estimate from the PanC4 study.

Last but not least, the difference in the estimates of heritability between the FPC-ADNI and the PanC4 study may partially reflect the distinct characteristics of the study populations. As discussed above, the PanC4 study selected pancreatic cancer cases regardless of their family history, whereas all affected individuals in FPC-ADNI study were familial cases. Narrow-sense heritability is a population parameter that quantifies the relative contribution of additive genetic effects. The genetic variance depends on segregation in a population of the alleles that influence the trait, the allele frequencies, the effect sizes of the variants and the mode of gene actions[17]. Although many causal variants and genes are shared by familial and sporadic cases of pancreatic cancer, it is possible that their enrichment and Mendelian form of inheritance in high-risk families together led to a greater contribution to the susceptibility of pancreatic cancer than in general populations. On the other hand, this estimate of heritability in the FPC-ADNI study is remarkably similar to what was reported in the twin study, which was considered as the upper limit of heritability for pancreatic cancer.

Conclusion

The overall heritability for pancreatic cancer from sequenced variants was estimated to be 38.8% in the FPC-ADNI study. This estimate of heritability may be inflated due to batch effects and misspecification of disease prevalence. Though difficult to be quantified, the inflation is expected to be small. This estimate of heritability does not capture the genetic variance

attributable to extremely rare variants that are associated with the risk of pancreatic cancer. This is due to the inherent limitation of the GREML approach. The estimate of heritability from sequenced variants in the FPC-ADNI study is quite similar to the family-based heritability, but significantly higher than the estimates from imputed variants in the PanC4 study. The higher estimate of heritability in the FPC-ADNI study may suggest a greater contribution of genetic effects in familial pancreatic cancer cases than in sporadic cases. In combination with the results in the PanC4 study, the heritability for pancreatic cancer may be somewhere between 21.2% and 38.8%.

Tables

Table 7 Characteristics of ADNI WGS Samples at Baseline

Characteristics	ADNI WGS (N = 809)
Study Phase, N (%)	
ADNI1	279 (34.5%)
ADNI GO	125 (15.5%)
ADNI2	405 (50.1%)
Age at baseline (years), mean ± SD (range)	73 ± 7 (55-91)
Female gender, N (%)	363 (44.9%)
Self-report race, N (%)	
White	756 (93.4%)
Black	28 (3.5%)
Others*	25 (3.1%)
Diagnosis at baseline†, N (%)	
CN	280 (34.6%)
SMC	0 (0%)
EMCI	235 (29.0%)
LMCI	246 (30.4%)
AD	48 (5.9%)
Education (years), N (%)	
≤ 12	116 (14.3%)
>12 & ≤ 16	344 (42.5%)
> 16 & ≤ 20	349 (43.1%)

* Asian, Hawaiian/Other Pacific Islanders, American Indian/Alaskan, more than one, unknown

† CN = cognitively normal; SMC = significant memory concern; EMCI = early mild cognitive impairment; LMCI = impairment; AD = Alzheimer's Disease.

Table 8 Characteristics of FPC WGS Samples at Baseline

Characteristics	FPC WGS (N = 638)
Age group (years), N (%)	
30-49	35 (5.5%)
50-69	338 (53.0%)
70-89	251 (39.3%)
90 or above	7 (1.1%)
Unknown	7 (1.1%)
Female gender, N (%)	325 (50.9%)
No. of affected relatives, N (%)	
Two	358 (56.1%)
Three	196 (30.7%)
Four or more	84 (13.2%)

Table 9 Estimates of Heritability under Different Relatedness Cutoffs and Prevalence Values

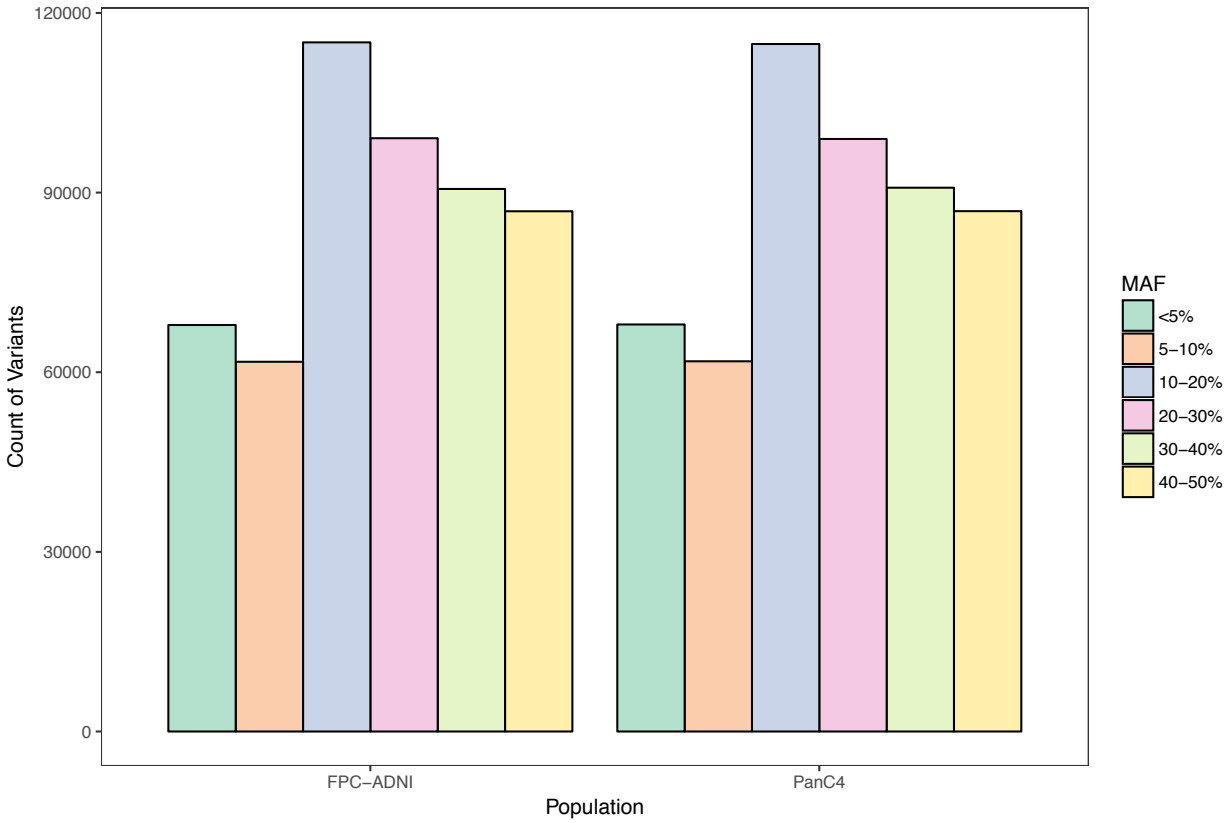
Sample Size	Prevalence	h^2	s.e.	P value
1,322 (relatedness > 0.05)	K = 0.0149	0.620	0.164	3.18×10^{-6}
	K/2 = 0.00745	0.526	0.139	3.18×10^{-6}
	K/5 = 0.00298	0.435	0.115	3.18×10^{-6}
	K/10 = 0.00149	0.384	0.102	3.18×10^{-6}
1,180 (relatedness > 0.025)	K = 0.0149	0.626	0.254	1.05×10^{-3}
	K/2 = 0.00745	0.532	0.216	1.05×10^{-3}
	K/5 = 0.00298	0.440	0.179	1.05×10^{-3}
	K/10 = 0.00149	0.388	0.157	1.05×10^{-3}

Table 10 Estimates of Heritability for Pancreatic Cancer from 521,282 Variants Shared by FPC-ADNI and PanC4 Study

Method	Cohort	Sample Size	Prevalence (K)	h^2	s.e.	P
GREML-SC	PanC4	6,931	K = 0.0149	0.166	0.033	1.70×10^{-7}
	FPC-ADNI	1,179	K = 0.0149	0.500	0.181	2.56×10^{-3}
			K/2 = 0.00745	0.424	0.154	2.56×10^{-3}
			K/5 = 0.00298	0.351	0.127	2.56×10^{-3}
			K/10 = 0.00149	0.309	0.112	2.56×10^{-3}
GREML-MS	PanC4	6,931	K = 0.0149	0.162	0.033	0.34
	FPC-ADNI	1,179	K = 0.0149	0.519	0.180	0.48
			K/2 = 0.00745	0.440	0.153	0.48
			K/5 = 0.00298	0.364	0.127	0.48
			K/10 = 0.00149	0.321	0.112	0.48

Supplementary Figures and Tables

Supplementary Figure 2 MAF Distribution of 521,282 Shared Variants in FPC-ADNI and PanC4 Study.



Reference

1. Klein AP. Genetic susceptibility to pancreatic cancer. Mol Carcinog. 2012 Jan;51(1):14–24.

2. Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, et al. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. Cancer Discov. 2015 Dec 9;

3. Weiner MW, Aisen PS, Jack CR, Jagust WJ, Trojanowski JQ, Shaw L, et al. The Alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimers Dement J Alzheimers Assoc. 2010 May;6(3):202-211.e7.

4. Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, et al. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement J Alzheimers Assoc. 2010 May;6(3):265–73.

5. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, et al. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat Biotechnol. 2011 Dec 18;30(1):61–8.

6. She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, et al. Shotgun sequence assembly and recent segmental duplications within the human genome. Nature. 2004 Oct 21;431(7011):927–30.

7. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinforma Oxf Engl. 2011 Aug 1;27(15):2156–8.

8. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.

9. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov 1;27(21):2987–93.

10. Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, et al. Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer. Cancer Discov. 2016 Feb 1;6(2):166–75.

11. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002 May 16;12(6):996–1006.

12. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.

13. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan 7;88(1):76–82.

14. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011 Mar 11;88(3):294–305.

15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.

16. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

17. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008 Apr;9(4):255–66.

CHAPTER FIVE

CONTRIBUTION OF ESTABLISHED FPC GENES IN THE HERITABILITY OF PANCREATIC CANCER

Familial Clustering and High-penetrance Mutations for Pancreatic Cancer

Inherited genetic components play an important role in the susceptibility of pancreatic cancer. It was estimated that up to 10% of pancreatic cancer patients had a family history of the disease. In recent case-control and cohort studies, 2.3 – 8.7% of pancreatic cancer patients had reported having at least one first-degree relatives (FDRs) affected by the disease, meeting the criteria of familial pancreatic cancer (FPC) (Table 11). The percentage appears to be higher in patients with early-onset pancreatic cancer (diagnosed at age ≤ 50 or < 60)[1,2]. In addition, two studies have reported a smaller proportion (< 2%) of pancreatic cancer patients having ≥ 2 FDRs with the disease [3,4]. It's important to note that obtaining an accurate family history of pancreatic cancer is very challenging as family members may be too young to have the disease when assessed, particularly for families with complex pedigree structures. Therefore, conclusive epidemiologic data on the prevalence of FPC is still lacking. Comparing to the general population, FPC kindred members have a 7- to 9-fold increased risk of pancreatic cancer[5,6]. The risk is even higher among members of FPC kindreds with a young-onset case (< 50 years; SIR = 9.31, 95% CI: 3.42-20.28) than those without (SIR = 6.34, 95% CI: 4.02-9.51)[6].

Mutations of several genes have been found to be responsible for the familial clustering of pancreatic cancer. Next-generation sequencing in FPC families had identified *ATM* and *PALB2*, previously known as familial breast cancer genes, to be associated with pancreatic cancer. Increased risk of pancreatic cancer is also associated with several inherited syndromes for which the predisposing genes have been identified. The most prominent syndromes are

hereditary breast-ovarian cancer syndrome, particularly due to germ-line mutations in *BRCA1/2*, and familial atypical multiple mole melanoma syndrome, due to mutations in *CDKN2A*. In addition, *STK11* associated with Peutz-Jeghers syndrome, mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) associated with Lynch syndrome, and *PRSS1* associated with hereditary pancreatitis are among the list of genes contributing to pancreatic cancer risk. A detailed review of these genes can be found in Chapter Two.

Prevalence of High-penetrance Mutations in FPC Patients

Given the substantial risk of pancreatic cancer associated with these genes, they may be used to characterized high-risk individuals that will benefit from early detection or targeted treatment of the disease. To determine the clinical utility of these high-penetrance mutations in identifying high-risk individuals for pancreatic cancer, efforts have been made to quantify their prevalence in FPC families (Table 12). *BRCA2* is the most frequently tested gene in these studies given its long-established connection to the risk of pancreatic cancer. Mutations in *BRCA2* also account for most FPC cases, with a prevalence ranged from 2.9% to 19.2%[7–13]. The wide range of *BRCA2* mutation prevalence in these studies may partially due to the various proportion of Ashkenazi Jewish (AJ) patients in the study population. It's known that the prevalence of three *BRCA1/2* founder mutations (*BRCA1* 185delAG, *BRCA1* 5382insC, and *BRCA2* 6174delT) is much higher in Ashkenazi Jewish patients with pancreatic cancer than in non-AJ patients[14]. Following *BRCA2*, *ATM* appears to be the second most frequently mutated genes in FPC families, with a prevalence ranged from 2.4 – 8.3%[15–17,13]. Germ-line mutations of *BRCA1, CDKN2A* and *PALB2* genes were not always detected in FPC samples[11,16–19], and their prevalence was in general lower than *BRCA2* and *ATM*[12,13,16,17,20]. A recent study from the Pancreatic Cancer Genetic Epidemiology (PACGENE) Consortium assessed the prevalence of

BRCA1, *BRAC2*, *PALB2*, and *CDKN2A* mutations in 727 unrelated probands with positive family history of pancreatic cancer and reported an overall prevalence of 8% in FPC probands, which was significantly higher than that in non-FPC probands (3.5%)[12]. A follow-up study extended the testing panel to also include *ATM*, *MSH2*, and *PMS2* gene. In 302 unrelated pancreatic cancer patients from Mayo Clinic, 11.9% of the FPC patients carried at least one deleterious mutation of these seven genes while 4.3% of the non-FPC patients were carriers[13]. The overall prevalence of FPC gene mutations was 14.5% in a Japanese study with a small sample size of 54 FPC patients[21]. A substantially higher prevalence of FPC gene mutations was reported in a US study (30.4%) but this study was small and was conducted retrospectively so that not all samples were tested for the same genes[16].

Prevalence of High-penetrance Mutations in Sporadic Cases

Germ-line mutations of the established FPC genes are not limited to FPC families or patients with a positive family history of pancreatic cancer. Mutations of these genes have also been detected in apparently sporadic pancreatic cancer cases. These patients were primarily recruited in clinical settings, unselected for their family history of pancreatic cancer. Early reports of germ-line mutations in sporadic/unselected pancreatic cancer patients have been focusing on *BRCA2*[22,23] and *CDKN2A*[24,25] genes. More recently, the investigation of the prevalence of germline mutations has been extended to include all established FPC genes in studies with increased sample sizes. Hu and his colleagues evaluated the frequency of mutations in established pancreatic cancer genes among 96 unselected pancreatic cancer patients from Mayo Clinic[26]. Review of the family history of pancreatic cancer classified 7% of the patients to be FPC cases, similar to the prevalence of FPC cases observed in the general population. Germ-line mutations were detected in 9.4% of tested samples, mostly in patients with a family history

of pancreatic cancer or other related cancers (i.e. breast cancer, colorectal cancer and endometrial cancer). In the follow-up study of 3,030 patients, deleterious mutations of established FPC genes were found in 5.5% of all tested patients, including 7.9% of patients with a family history of pancreatic cancer and 5.2% patients with no family history[27]. In a similar study, Shindo and his colleagues recruited 854 pancreatic cancer patients from Johns Hopkins Hospital, the majority of whom (87.3%) had a negative family history of pancreatic cancer[28]. Among these sporadic pancreatic cancer cases, 31 mutations in known FPC susceptibility genes were identified, equivalent to an overall prevalence of 3.5%. In a recent study of 289 resected PDAC samples, 9.7% of patients carried pathogenic or likely pathogenic germline mutations of established FPC genes. The observed higher prevalence of established FPC gene mutations in this study is likely a result of a high proportion of patients (57.1%) with a personal history of another cancer[29].

In general, pancreatic cancer patients with stronger family history tend to have a higher chance of carrying mutations of known FPC genes. Mutations of these genes have also been detected in 4-10% of cases unselected for their family history[12,13,28,27,29]. Findings from these studies indicate an essential role of these high-penetrance mutations in the susceptibility of pancreatic cancer.

Contribution of Established FPC Genes in Heritability

To quantify the contribution of the established FPC genes in the susceptibility of pancreatic cancer, we estimated the variance explained by these 11 genes in a heritability analysis using imputed data in the PanC4 study. The analyzed PanC4 study consists of 3,568 pancreatic cancer patients and 3,363 controls (more details can be found in Chapter Three). These pancreatic cancer patients were selected regardless of their family history from nine hospital- or population-based case-control studies. Thus the prevalence of these FPC gene

mutations in the PanC4 study was expected to be similar to that in the general population. A multi-component analysis using GREML-LDMS approach was conducted to evaluate the contribution of the established FPC genes. A total of 9,445 variants located within ±50 kb of gene boundaries (3' UTR to 5' UTR) of the 11 FPC genes were evaluated as a single genetic component for their aggregated contribution to the estimated heritability for pancreatic cancer. The remaining variants across the genome were stratified into 2 minor-allele frequency (MAF) categories and 2 linkage disequilibrium (LD) groups as described in Chapter Three. A disease prevalence of 0.0149 was specified to transform the estimates from observed scale to liability scale and to correct for ascertainment bias in the case-control cohort. Together these variants explained 0.4% (s.e. = 0.3%) of phenotypic variance for pancreatic cancer.

This estimate appears to be much lower than what was expected given the prevalence and effect sizes of these genes in pancreatic cancer patients. Rare mutations in established FPC genes have been shown to explain a measurable fraction (4-10%) if not more of unselected pancreatic cancer. Mutations of these genes also confer a high risk of pancreatic cancer, with the per gene odds ratio ranged from 2.58 to 12.33[27]. However, the estimate of heritability using the current statistical approach cannot sufficiently capture the genetic variance attributable to these rare variants associated with the risk of pancreatic cancer.

Limitation of Current Approach in Capturing Rare Variations

Regardless of the type of data, imputation or sequencing, germ-line mutations are difficult to be characterized due to their extremely low frequencies in the general population. In imputations they are either not imputed or imputed with poor accuracy due to being monomorphic or extremely rare in the reference population. Detecting these rare mutations are possible in sequencing but requires a large sample size, which could be extremely costly for a

rare disease like pancreatic cancer. Even if these mutations are captured with sufficient accuracy in imputation or sequencing, they are often removed at the quality control stage. Including extremely low-frequency variants in the heritability analysis is currently not recommended as they would introduce noise and lead to unstable estimates. In our analysis, all variants with minor allele count less than five in the PanC4 study were excluded. Pathological mutations of FPC genes are most likely excluded in such heritability analysis. Therefore, failed to capture the variance explained by these rare mutations is expected in any GREML analysis, which would lead to a biased estimate of heritability.

The underestimation of heritability could be more evident for cancer phenotypes known to be affected by extremely rare mutations. In simulation studies, Yang and his colleagues concluded that the GREML-LDMS estimates of heritability using whole-genome sequencing data were unbiased under four different scenarios when the causal variants were (1) a random set of all variants, (2) more common than random variants, (3) less frequent than random variants, or (4) less frequent than random variants and with lower than average LD[30]. The conclusion may hold true for complex traits like human height and BMI, in which the majority of the additive genetic variance is explained by common variants[30,31]. But the genetic architecture of many cancer types may not fall into any of the four categories. The spectrum of cancer susceptibility alleles includes mutations in genes that are highly penetrant as well as low-frequency to common variants that impart a small additional risk for cancer. Given that high-penetrance mutations will most definitively be excluded from the analysis, the variance of cancer phenotypes explained by sequencing variants would be an underestimation of their true heritability. And the degree of bias may vary by cancer types, depending on the effect size and frequency of these high-penetrance mutations.

Conclusion

Heritability is a fundamental concept of genetics and its unbiased estimation is critical for understanding the genetic architecture of complex traits as well as for designing better studies in identifying disease-causing genes. Unlike traditional family-based studies, current statistical methods such as GREML approach allow the assessment of heritability in unrelated individuals, which tease apart the contribution of environmental factors shared among relatives. Although results from these approaches are sensitive to the type of data used in the analysis (array, imputation or sequencing), they can generally be improved by stringent quality controls, larger sample sizes and/or larger imputation panels. However, even when the best practices are applied, heritability is still likely underestimated, due to the limited power of these approaches in capturing genetic variation at rare mutations. The extent of such bias is likely to be trait-dependent, and more evident in cancer phenotypes known to be greatly affected by high-penetrance germ-line mutations. This limitation should be kept in mind when interpreting estimates of heritability from these approaches.

Tables

Table 11 Self-reported Family History of Pancreatic Cancer in Affected Individuals

Study (Year)	Country	Time Period	Study Design	Number of Cases*	Type of relative	Number (%) with a family history
Fernandez et al. (1994)[32]	Italy	1983-1992	Case-Control	362	First-degree	14 (3.9%)
Silverman et al. (1999)[33]	US	1986-1989	Case-Control	459	First-degree	23 (5.0%)
Inoue et al. (2003)[34]	Japan	1988-1999	Case-Control	200	Unspecified	10 (5.0%)
Hassan et al. (2007)[35]	US	2000-2006	Case-Control	802	First-degree	45 (5.6%)
Jacob et al. (2009)[36]	US (CPS-II)	1982	Prospective Cohort	7,306	First-degree	168 (2.3%)
Jacob et al. (2010)[37]	US, Europe, China (PanScan)	Varies by study	Nested Case-Control	1,183	First-degree	76 (6.4%) 6 (0.5%) (2 or more)
Austin et al. (2013)[38]	US	2005-2009	Case-Control	538	First-degree	47 (8.7%)
Elena et al. (2013)[39]	US, Europe, China (PanScan)	Varies by study	Nested Case-Control	773	First-degree	44 (5.7%)
Fehringer et al. (2014)[40]	Canada	2003-2009	Case-Control	743	First-degree	50 (7.2%)
Mizuno et al. (2014)[41]	Japan	1993-2013	Retrospective Cohort	688	First-degree	28 (4.1%)
Rosato et al. (2015)[42]	Italy	1991-2008	Case-Control	326	Unspecified	10 (3.1%)
Mocci et al. (2015)[43]	Spain	2009-2015	Case-Control	577	First-degree	36 (6.2%)
Piciucchi et al. (2015)[1]	Italy	2006-2013	Retrospective Cohort	293	First-degree	19 (6.5%) 2 (8.0%) (age ≤ 50)
Zheng et al. (2015)[44]	China	2011-2013	Case-Control	323	Unspecified	11 (3.4%)
Lakatos et al. (2016)[45]	Hungary	2012-2014	Retrospective Cohort	354	Unspecified	13 (3.6%)
Schulte et al. (2016)[46]	Australia	2007-2011	Case-Control	592	First-degree	31 (5.2%)
Takai et al. (2016)[17]	Japan	2002-2013	Retrospective Cohort	1,197	First-degree	88 (7.3%)

McWilliams et al. (2017)[47]	US, Europe, Australia, Canada (PanC4)	Varies by study	Case-Control	1,733 (age < 60)	First-degree	157 (9.1%) (age < 60)
Huang et al. (2018)[48]	US	1993-2012	Prospective Cohort	1,455	Unspecified	49 (3.4%)
Molina-Montes et al. (2018)[4]	Europe (PanGenEU)	2009-2018	Case-Control	1,431	First-degree	104 (7.3%) 28 (1.9%) (2 or more)

* Number of cases with available information on family history of pancreatic cancer.

Table 12 Prevalence of FPC Gene Germ-line Mutations in FPC Patients

Study	Country	Sample Size*	Most Commonly Tested Genes					Overall Prevalence‡
			BRCA1	*BRCA2*	*CDKN2A*	*PALB2*	*ATM*	
Bartsch et al. (2002)[18]	Germany	18	-	-	0 (0%)	-	-	-
Murphy et al. (2002)[7]	US	29	-	5 (17.2%)	-	-	-	-
Hahn et al. (2003)[8]	Europe	26	-	5 (19.2%)	-	-	-	-
Couch et al. (2007)[9]	US	146	-	8 (5.5%)	-	-	-	-
Slater et al. (2010)[20]	Europe	81	-	-	-	3 (3.7%)	-	-
Slater et al. (2010)[10]	Germany	70	-	2 (2.9%)	0 (0%)	-	-	-
Ghiorzo et al. (2012)[11]	Italy	16	0 (0%)	0 (0%)	5 (31.3%)	0 (0%)	-	
Harinck et al. (2012)[19]	Netherlands	28	-	-	-	0 (0%)	-	-
Roberts et al. (2012)[15]	US	166	-	-	-	-	4 (2.4%)	-
Zhen et al. (2015)[12]	US	515	6 (1.2%)	19 (3.7%)	13 (2.5%)	3 (0.6%)	-	41 (8.0%)
Catts et al. (2016)[49]	US	46	4 (9.5%)	7 (16.7%)	-	0 (0%)	1 (8.3%)	14 (30.4%)
Takai et al. (2016)[17]	Japan	54	0 (0%)	3 (5.6%)	-	2 (3.7%)	2 (3.7%)	8 (14.5%)
Chaffee et al. (2017)[13]	US	185	2 (1.1%)	8 (4.3%)	4 (2.2%)	1 (0.5%)	6 (3.2%)	22 (11.9%)

* The total number of FPC probands being tested for any gene. The number of FPC probands tested for each gene may be different in the same study.

† The denominator was aggregated from the number of samples tested for each specific gene across studies.

‡ Overall prevalence of all tested genes associated with pancreatic cancer, not limited to the five listed genes.

Reference

1. Piciucchi M, Capurso G, Valente R, Larghi A, Archibugi L, Signoretti M, et al. Early onset pancreatic cancer: risk factors, presentation and outcome. Pancreatol Off J Int Assoc Pancreatol IAP Al. 2015 Apr;15(2):151–5.

2. McWilliams RR, Maisonneuve P, Bamlet WR, Petersen GM, Li D, Risch HA, et al. Risk Factors for Early-Onset and Very-Early-Onset Pancreatic Adenocarcinoma: A Pancreatic Cancer Case-Control Consortium (PanC4) Analysis. Pancreas. 2016 Feb;45(2):311–6.

3. Jacobs EJ, Chanock SJ, Fuchs CS, Lacroix A, McWilliams RR, Steplowski E, et al. Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Int J Cancer. 2010 Sep 1;127(6):1421–8.

4. Molina-Montes E, Gomez-Rubio P, Márquez M, Rava M, Löhr M, Michalski CW, et al. Risk of pancreatic cancer associated with family history of cancer and other medical conditions by accounting for smoking among relatives. Int J Epidemiol [Internet]. [cited 2018 Apr 17]; Available from: https://academic-oup-com.ezp.welch.jhmi.edu/ije/advance-article/doi/10.1093/ije/dyx269/4797025

5. Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJA, et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res. 2004 Apr 1;64(7):2634–8.

6. Brune KA, Lau B, Palmisano E, Canto M, Goggins MG, Hruban RH, et al. Importance of age of onset in pancreatic cancer kindreds. J Natl Cancer Inst. 2010 Jan 20;102(2):119–26.

7. Murphy KM, Brune KA, Griffin C, Sollenberger JE, Petersen GM, Bansal R, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res. 2002 Jul 1;62(13):3789–93.

8. Hahn SA, Greenhalf B, Ellis I, Sina-Frey M, Rieder H, Korte B, et al. BRCA2 germline mutations in familial pancreatic carcinoma. J Natl Cancer Inst. 2003 Feb 5;95(3):214–21.

9. Couch FJ, Johnson MR, Rabe KG, Brune K, de Andrade M, Goggins M, et al. The prevalence of BRCA2 mutations in familial pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2007 Feb;16(2):342–6.

10. Slater EP, Langer P, Fendrich V, Habbe N, Chaloupka B, Matthäi E, et al. Prevalence of BRCA2 and CDKN2a mutations in German familial pancreatic cancer families. Fam Cancer. 2010 Sep;9(3):335–43.

11. Ghiorzo P, Pensotti V, Fornarini G, Sciallero S, Battistuzzi L, Belli F, et al. Contribution of germline mutations in the BRCA and PALB2 genes to pancreatic cancer in Italy. Fam Cancer. 2012 Mar;11(1):41–7.

12. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med Off J Am Coll Med Genet. 2015 Jul;17(7):569–77.

13. Chaffee KG, Oberg AL, McWilliams RR, Majithia N, Allen BA, Kidd J, et al. Prevalence of germ-line mutations in cancer genes among pancreatic cancer patients with a positive family history. Genet Med Off J Am Coll Med Genet. 2017 Jul 20;

14. Lucas AL, Frado LE, Hwang C, Kumar S, Khanna LG, Levinson EJ, et al. BRCA1 and BRCA2 germline mutations are frequently demonstrated in both high-risk pancreatic cancer screening and pancreatic cancer cohorts. Cancer. 2014 Jul 1;120(13):1960–7.

15. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, et al. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2012 Jan;2(1):41–6.

16. Catts ZA-K, Baig MK, Milewski B, Keywan C, Guarino M, Petrelli N. Statewide Retrospective Review of Familial Pancreatic Cancer in Delaware, and Frequency of Genetic Mutations in Pancreatic Cancer Kindreds. Ann Surg Oncol. 2016 May 1;23(5):1729–35.

17. Takai E, Yachida S, Shimizu K, Furuse J, Kubo E, Ohmoto A, et al. Germline mutations in Japanese familial pancreatic cancer patients. Oncotarget. 2016 Nov 8;7(45):74227–35.

18. Bartsch DK, Sina-Frey M, Lang S, Wild A, Gerdes B, Barth P, et al. CDKN2A germline mutations in familial pancreatic cancer. Ann Surg. 2002 Dec;236(6):730–7.

19. Harinck F, Kluijt I, van Mil SE, Waisfisz Q, van Os TAM, Aalfs CM, et al. Routine testing for PALB2 mutations in familial pancreatic cancer families and breast cancer families with pancreatic cancer is not indicated. Eur J Hum Genet EJHG. 2012 May;20(5):577–9.

20. Slater EP, Langer P, Niemczyk E, Strauch K, Butler J, Habbe N, et al. PALB2 mutations in European familial pancreatic cancer families. Clin Genet. 2010 Nov;78(5):490–4.

21. Takai E, Totoki Y, Nakamura H, Morizane C, Nara S, Hama N, et al. Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. Sci Rep. 2015;5:18425.

22. Goggins M, Schutte M, Lu J, Moskaluk CA, Weinstein CL, Petersen GM, et al. Germline BRCA2 gene mutations in patients with apparently sporadic pancreatic carcinomas. Cancer Res. 1996 Dec 1;56(23):5360–4.

23. Özcelik H, Schmocker B, Nicola ND, Shi X-H, Langer B, Moore M, et al. Germline BRCA2 6174delT mutations in Ashkenazi Jewish pancreatic cancer patients. Nat Genet. 1997 May;16(1):17–8.

24. McWilliams RR, Wieben ED, Rabe KG, Pedersen KS, Wu Y, Sicotte H, et al. Prevalence of CDKN2A mutations in pancreatic cancer patients: implications for genetic counseling. Eur J Hum Genet EJHG. 2011 Apr;19(4):472–8.

25. Ghiorzo P, Fornarini G, Sciallero S, Battistuzzi L, Belli F, Bernard L, et al. CDKN2A is the main susceptibility gene in Italian pancreatic cancer families. J Med Genet. 2012 Mar 1;49(3):164–70.

26. Hu C, Hart SN, Bamlet WR, Moore RM, Nandakumar K, Eckloff BW, et al. Prevalence of Pathogenic Mutations in Cancer Predisposition Genes among Pancreatic Cancer Patients. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2016 Jan;25(1):207–11.

27. Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. JAMA. 2018 Jun 19;319(23):2401–9.

28. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious Germline Mutations in Patients With Apparently Sporadic Pancreatic Adenocarcinoma. J Clin Oncol Off J Am Soc Clin Oncol. 2017 Oct 20;35(30):3382–90.

29. Yurgelun MB, Chittenden AB, Morales-Oyarvide V, Rubinson DA, Dunne RF, Kozak MM, et al. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. Genet Med Off J Am Coll Med Genet. 2018 Jul 2;

30. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

31. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet. 2018 Apr 26;

32. Fernandez E, La Vecchia C, D'Avanzo B, Negri E, Franceschi S. Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 1994 May;3(3):209–12.

33. Silverman DT, Schiffman M, Everhart J, Goldstein A, Lillemoe KD, Swanson GM, et al. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer. 1999 Aug;80(11):1830–7.

34. Inoue M, Tajima K, Takezaki T, Hamajima N, Hirose K, Ito H, et al. Epidemiology of pancreatic cancer in Japan: a nested case-control study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC). Int J Epidemiol. 2003 Apr;32(2):257–62.

35. Hassan MM, Bondy ML, Wolff RA, Abbruzzese JL, Vauthey J-N, Pisters PW, et al. Risk factors for pancreatic cancer: case-control study. Am J Gastroenterol. 2007 Dec;102(12):2696–707.

36. Jacobs EJ, Rodriguez C, Newton CC, Bain EB, Patel AV, Feigelson HS, et al. Family history of various cancers and pancreatic cancer mortality in a large cohort. Cancer Causes Control CCC. 2009 Oct;20(8):1261–9.

37. Jacobs EJ, Chanock SJ, Fuchs CS, Lacroix A, McWilliams RR, Steplowski E, et al. Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Int J Cancer J Int Cancer. 2010 Sep 1;127(6):1421–8.

38. Austin MA, Kuo E, Van Den Eeden SK, Mandelson MT, Brentnall TA, Kamineni A, et al. Family history of diabetes and pancreatic cancer as risk factors for pancreatic cancer: the PACIFIC study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2013 Oct;22(10):1913–7.

39. Elena JW, Steplowski E, Yu K, Hartge P, Tobias GS, Brotzman MJ, et al. Diabetes and risk of pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. Cancer Causes Control CCC. 2013 Jan;24(1):13–25.

40. Fehringer G, Gallinger S, Borgida A, Zhang LR, Adams L, Liu G, et al. The association of family history of cancer and medical history with pancreatic cancer risk. Pancreas. 2014 Jul;43(5):812–4.

41. Mizuno S, Nakai Y, Isayama H, Kawahata S, Saito T, Takagi K, et al. Smoking, family history of cancer, and diabetes mellitus are associated with the age of onset of pancreatic cancer in Japanese patients. Pancreas. 2014 Oct;43(7):1014–7.

42. Rosato V, Polesel J, Bosetti C, Serraino D, Negri E, La Vecchia C. Population attributable risk for pancreatic cancer in Northern Italy. Pancreas. 2015 Mar;44(2):216–20.

43. Mocci E, Guillen-Ponce C, Earl J, Marquez M, Solera J, Salazar-López M-T, et al. PanGen-Fam: Spanish registry of hereditary pancreatic cancer. Eur J Cancer Oxf Engl 1990. 2015 Sep;51(14):1911–7.

44. Zheng Z, Zheng R, He Y, Sun X, Wang N, Chen T, et al. Risk Factors for Pancreatic Cancer in China: A Multicenter Case-Control Study. J Epidemiol. 2016;26(2):64–70.

45. Lakatos G, Balázs A, Kui B, Gódi S, Szücs Á, Szentesi A, et al. Pancreatic Cancer: Multicenter Prospective Data Collection and Analysis by the Hungarian Pancreatic Study Group. J Gastrointest Liver Dis JGLD. 2016 Jun;25(2):219–25.

46. Schulte A, Pandeya N, Fawcett J, Fritschi L, Klein K, Risch HA, et al. Association between family cancer history and risk of pancreatic cancer. Cancer Epidemiol. 2016 Dec;45:145–50.

47. McWilliams RR, Maisonneuve P, Bamlet WR, Petersen GM, Li D, Risch HA, et al. Risk Factors for Early-Onset and Very-Early-Onset Pancreatic Adenocarcinoma: A Pancreatic Cancer Case-Control Consortium (PanC4) Analysis. Pancreas. 2016 Feb;45(2):311–6.

48. Huang Brian Z., Le Marchand Loic, Haiman Christopher A., Monroe Kristine R., Wilkens Lynne R., Zhang Zuo-Feng, et al. Atopic allergic conditions and pancreatic cancer risk: Results from the Multiethnic Cohort Study. Int J Cancer. 2018 Jan 4;142(10):2019–27.

49. Catts ZA-K, Baig MK, Milewski B, Keywan C, Guarino M, Petrelli N. Statewide Retrospective Review of Familial Pancreatic Cancer in Delaware, and Frequency of Genetic Mutations in Pancreatic Cancer Kindreds. Ann Surg Oncol. 2016 May;23(5):1729–35.

CHAPTER SIX

CONCLUSIONS, FUTURE DIRECTIONS AND PUBLIC HEALTH SIGNIFICANCE

Summary and Conclusions

Pancreatic cancer is fundamentally a genetic disease caused by both inherited and acquired genetic alterations. Inherited genetic variation plays an important role in both the familial and non-familial (sporadic) occurrences of pancreatic cancer. Family history is a long-recognized risk factor for pancreatic cancer and an important predictor of disease risk. As previously discussed, rare mutations in 11 FPC genes confer a high lifetime risk of pancreatic cancer (Chapter Two). Large-scale genome-wide association studies (GWAS) to date have identified 26 common variants associated with a small increased risk of for pancreatic cancer in European populations and additional 5 loci in Asian populations (Chapter Two). These findings suggest that the genetic basis of pancreatic cancer consists of a wide spectrum of causal variants, ranging from extremely rare high-penetrance mutations to common low-risk susceptibility loci.

Heritability (narrow-sense heritability) is a population-based parameter that has been widely used to quantify the impact of additive genetic effect to a complex trait/disease. Prior larger scale twin studies have estimated heritability for pancreatic cancer to be 36% (95% CI: 0% - 53%)[1]. However, some studies have suggested estimates of heritability from family-based studies may be biased upwardly due to the shared environmental factors. The genomic-relatedness-based restricted maximum-likelihood (GREML) approach allows for the assessment of heritability in unrelated individuals, which could tease apart the shared environmental factors. Using this approach, the heritability for pancreatic cancer in our study was estimated to be 21.2% from imputed variants in the PanC4 study and was 38.8% from sequenced variants in the FPC-

ADNI study. These two estimates of heritability are both biased downwardly since extremely rare mutations associated with pancreatic cancer cannot be captured by the GREML approach. The higher estimate of heritability in the FPC-ADNI study could be a result of more comprehensive characterization of genetic variation in sequencing data than in imputation data, as well as the pancreatic cancer patients in the study population being exclusively FPC cases. It is possible that among FPC patients, the enrichment of high-penetrance causal variants with moderate to large effects in high-risk families together led to a greater contribution to the susceptibility of pancreatic cancer than in the general populations. Potential bias due to batch effects and/or mis-specification of disease prevalence are also possible explanations for the higher estimate of heritability in FPC-ADNI study. But their impacts are likely to be small given the stringent quality controls and the sensitivity analysis on disease prevalence. Additional analysis using the same set of variants from array data led to the same observation of a higher estimate of heritability in the FPC-ADNI study than in the PanC4 study. The estimate of heritability in the FPC-ADNI study is also very similar to the estimate from the previous twin study. These results suggest that additive genetic effects may be a greater contributor to the risk of pancreatic cancer in families with a strong genetic background, and that the true heritability for FPC could be somewhere near this estimate of heritability.

Genomic partitioning of the estimated heritability in the PanC4 study ranked chromosome 9, 7, 8, 16, 5, 2 and 1 as the top contributors to the heritability of pancreatic cancer. This is consistent with the GWAS findings as on all these chromosomes common susceptibility loci have been discovered. Across the functional group of intronic, intergenic, coding and regulatory variants, intronic variants account for most phenotypic variance of pancreatic cancer (12.4%). Interestingly, 12 out of 21 GWAS loci identified in European populations are mapped

to intronic variants. Given that the GREML approach is more powered to capture genetic variation at common variants, it was not surprising to find that these results appeared to be driven by GWAS loci. However, this does not mean that rare variants have little influence on the risk of pancreatic cancer, but rather their influence cannot be truly reflected in such analysis. For example, the contribution of coding variants is likely to be underestimated to a greater extent due to being on the lower end of the MAF spectrum than variants in other functional groups.

Results from partitioning by MAF showed that more than one-third of the estimated heritability was attributable to rare variants. Since the GREML approach can only capture a portion of variance due to rare variants, this proportion is likely to be an underestimate. Our results demonstrate an important role of rare variants in the susceptibility of pancreatic cancer, which emphasize the necessity of whole-genome sequencing (WGS) studies for the identification of variants with extremely low frequency or that are unique to specific populations. However, obtaining a sufficiently large sample to permit such sequencing analysis remain a significant challenge. On the other hand, common variants still explain a substantial amount of total phenotypic variance for pancreatic cancer (16.0%). Comparing to a previous analysis in the PanC4 study, our analysis reported a higher contribution of known GWAS loci to the susceptibility of pancreatic cancer (4.1% vs. 3.0%), likely resulting from the inclusion of the five novel loci discovered in a recent meta-analysis of four pancreatic cancer consortium studies[2]. The large proportion of still-missing heritability highlights the importance of continuing to search for common susceptibility loci for pancreatic cancer. Imputation and meta-analysis will remain a more cost-effective strategy than WGS in identifying variants with moderate to small effect sizes. However improved power must be achieved through larger sample sizes, larger or

trait-specific imputation reference panels, and a better characterization of low-frequency and rare variants.

Germ-line mutations are known to be critical in the development of pancreatic cancer. Prevalence of these high-penetrance mutations was reported to be 8-30% in FPC cases and 4-10% in sporadic cases. When evaluated in heritability analysis, variants of these 11 FPC genes together only explained 0.4% of all phenotypic variance for pancreatic cancer. These results again highlighted the inherent limitation of the GREML approach, incapable of capturing variance attributable to extremely rare mutations associated with pancreatic cancer. This limitation could yield misleading results that rare variants have a smaller influence on the susceptibility of a disease, which should be considered when interpreting results from such heritability analysis, particularly for cancer outcomes known to be affected by rare mutations.

Strengths of the Studies

Heritability for pancreatic cancer had previously been estimated from traditional family-based studies. Due to the rarity and fast progression of pancreatic cancer, these studies have limited statistical power to make conclusive inference and estimates of heritability from family-based studies may be inflated by shared environmental factors. In this study we performed a heritability analysis using the restricted maximum likelihood (REML) approach (GREML) to estimate the heritability from case-control studies of pancreatic cancer. Heritability estimated from unrelated individuals is less likely to be biased by environmental factors shared between relatives.

A number of studies have applied the GREML approach to estimate heritability for pancreatic cancer using array data. However, since variants in arrays are very selective and mostly common, these array-based heritability estimates are likely under-estimates. Our study

was conducted in two different case-control samples, each has its own strengths. The PanC4 study consists of 3,568 pancreatic cancer cases and 3,363 controls selected from nine hospital- or population-based studies, representing one of the largest case-control studies for pancreatic cancer. Samples from the PanC4 study were originally genotyped on the IlluminaHumanOmniExpressExome-8v1 array and were later imputed with 1000 Genome phase 3 populations. The imputation of array data has greatly extended its coverage to low-frequency and rare variants. In the PanC4 study, we employed three forms of GREML approach (GREML-SC, GREML-MS and GREML-LDMS) to estimate phenotypic variance explained by genetic variants. It has been demonstrated that single-component GREML (GREML-SC) estimates are biased when causal variants are not a random set of all sequence variants in terms of their MAF distribution and linkage disequilibrium (LD) structure. MAF-stratified multi-component GREML (GREML-MS) can account for the non-random MAF distribution of causal variants. Additional stratification on LD can further correct for the region-specific LD heterogeneity across the genome. The application of MAF- and LD-stratified multi-component GREML (GREML-LDMS) approach using imputation data in PanC4 study ensured our estimate of heritability be less biased than previously reported array-based estimates using the GREML-SC approach. In addition, the large sample size of the PanC4 study allows for stable estimates of variance when the number of components increases in GREML-LDMS models, and hence permits the investigation of the genetic architecture of pancreatic cancer through genomic partitioning of the estimated heritability.

The FPC-ADNI study may very likely be the only pancreatic cancer sample with WGS data. This sample was a combined sample with FPC cases from the Familial Pancreatic Cancer Genome Sequencing Project (FPC-GSP) and controls from the Alzheimer's Disease

Neuroimaging Initiative (ADNI) study, each of which is by far the largest sample of European ancestry with WGS data available for the corresponding disease. Subjects in the ADNI sample were cancer-free at baseline, and were similar in term of age distribution, sex ratio and catchment area to FPC cases, making them the ideal controls. Moreover, the cases and controls were both sequenced at Illumina using the same platform around the same period of time, allowing the merging of these WGS data to be achieved with fewer concerns about batch effects. Comparing to arrays and imputations, WGS data with sufficient coverage (30X) provides a near-complete characterization of the genetic variations across the genome, and hence minimize the bias in heritability estimation due to inadequate coverage of causal variants, particularly those at the lower end of MAF spectrum.

Limitations of the Studies

The small sample size of the FPC-ADNI study has limited our ability to evaluate the heritability for pancreatic cancer using multi-component GREML approaches. Attempts to stratify the variants into a various number of MAF categories all failed to generate a reliable estimate of variance using GCTA software. It was demonstrated that estimates of heritability from GREML-SC approach are unbiased only when the causal variants are a random subset of all sequenced variants[3]. Results from the PanC4 study suggested a distribution of causal variants for pancreatic cancer towards the lower end of the MAF spectrum. Given this, the estimate of heritability from all sequenced variants using the GREML-SC approach is likely biased downwardly.

The small size of the FPC-ADNI study also prohibits us from performing the genomic partitioning on WGS data. Our high coverage WGS data in theory is a more ideal data than imputation data to infer the genetic architecture of pancreatic cancer because of its near-complete

characterization of the genetic variation across the genome. However, any analysis involving more than one genetic component in the model failed to provide reliable estimates or convergence was not obtained.

Moreover, due to the unavailability of WGS data in an existing pancreatic cancer control population, we combined the FPC cases with subjects from ADNI study. The cases and controls were sequenced and called in slightly different pipelines. In the presence of technical artifact, the allele frequencies between cases and controls on average will be more different than expected under the null hypothesis, and cases will appear to be more related to other cases than to controls, which could lead to inflated estimates of heritability. To minimize the impact of batch effects, stringent QC filters were implemented on WGS data. In addition, the comparability of WGS data was carefully assessed before merging the FPC and ADNI dataset. Variants sequenced in the FPC and ADNI samples overall showed very similar distributions of mean depth per site and mean GQ per site across the genome. Although QC filters are not able to fully address the batch effects, the influence of batch effects on the estimates of heritability is likely to be small in the FPC-ADNI study.

Future Directions and Public Health Significance

It is estimated that pancreatic cancer will become the second leading cause of cancer death in the US by 2030. The survival rate for pancreatic cancer remains low (currently 8%) in comparison to other common cancer types, due to the lack of effective means for prevention, diagnosis and treatment. In the past few years, genetic research in pancreatic cancer has identified a number of genes and variants associated with the disease, which provide valuable insights into the complexities of its genetic architecture. Deciphering the genetic architecture is central to understanding the disease etiology and to aid in disease screening, diagnosis, prognosis

and therapy. Our study represents the first systematic investigation of the genetic architecture of pancreatic cancer.

Despite of the fruitful yields in recent GWAS and sequencing studies, the identified genes and variants for pancreatic cancer are far from explaining all heritability. Finding the missing heritability is the key step towards deciphering the genetic basis of pancreatic cancer and to help make informed decisions on the strategies and priorities of future gene discovery efforts. Our genomic partitioning analysis has found missing heritability at both rare and common variants for pancreatic cancer. Despite being underestimated, more than one-third of the estimated heritability for pancreatic cancer is attributable to rare variants, but little has been explained by known FPC genes. Since GWAS studies are not designed to detect rare associations, sequencing studies are warranted to facilitate the discovery of rare causal variants. Whole-exome sequencing can identify rare variants that directly influence gene functions but is restricted to examining the protein-coding content that comprises only 1% of the genome. In our analysis intronic variants explained the largest proportion of the estimated heritability for pancreatic cancer, emphasizing the importance of including both coding and non-coding regions in the search for disease-associated variants. WGS can measure nearly all genetic variation in the human genome and assess structural variants more accurately than WES. But owing to their cost, WGS studies have been limited by small sample sizes. Unexplained heritability at common variants was also substantial, indicating that more low-risk common susceptibility loci have yet to be identified. These variants likely have effect sizes too small to reach the genome-wide significance in current GWAS and require larger sample sizes to be detected. For low-frequency and rare variants, imputation and meta-analysis can be a more cost-effective strategy than sequencing, but the power needs to be improved with larger sample

sizes, larger imputation reference panels, or a combination with disease-specific reference panels.

Population-based heritability estimates such as those obtained in our study, provide a valuable metric for completeness of available genetic risk information, but individualized disease prevention and treatment will ultimately require identifying the variants accounting for risk in a given individual rather than on a population basis. It is our hope that results from this study provide a potentially valuable path towards the future discoveries of additional causal variants for pancreatic cancer that may permit a better characterization of individuals at the highest genetically defined risk in which effective prevention approaches can be developed and potential therapeutic targets for the disease can be explored.

Screening and early detection offer the best chance of reducing the high mortality of the disease. Selective screening of high-risk individuals is considered a cost-effective strategy. Currently, recommendation for screening is made based on individuals' family history of pancreatic cancer and their mutation status of the established FPC genes. However, based on our literature review, germ-line mutations of the established FPC genes account for about 8-30% of tested FPC cases, suggesting that pancreatic cancer is a genetically heterogeneous disease and more high-penetrance mutations are yet to be found. These high-penetrance mutations can facilitate the identification of the small proportion of the population at highest genetically defined risk for pancreatic cancer who may benefit from selective screening and early detection of the disease. Moreover, individuals with specific mutations may also benefit from targeted therapies. Tumors with mutations in *BRCA1/2*, *PALB2*, and *ATM* are highly sensitive to DNA-damaging related treatments such as crosslinking agents[4–6] and poly(ADP-ribose) polymerase inhibitors (PARPi)[7–10]. The clinical benefits of using these agents in patients with pancreatic

cancer are currently being investigated and promising results have been shown in the preliminary analysis.

Reference

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000 Jul 13;343(2):78–85.

2. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun. 2018 Feb 8;9(1):556.

3. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015 Oct;47(10):1114–20.

4. van der Heijden MS, Brody JR, Gallmeier E, Cunningham SC, Dezentje DA, Shen D, et al. Functional defects in the fanconi anemia pathway in pancreatic cancer cells. Am J Pathol. 2004 Aug;165(2):651–7.

5. van der Heijden MS, Brody JR, Dezentje DA, Gallmeier E, Cunningham SC, Swartz MJ, et al. In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. Clin Cancer Res Off J Am Assoc Cancer Res. 2005 Oct 15;11(20):7508–15.

6. Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, et al. Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Mol Cancer Ther. 2011 Jan;10(1):3–8.

7. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005 Apr 14;434(7035):913–7.

8. McCabe N, Lord CJ, Tutt ANJ, Martin NMB, Smith GCM, Ashworth A. BRCA2-deficient CAPAN-1 cells are extremely sensitive to the inhibition of Poly (ADP-Ribose) polymerase: an issue of potency. Cancer Biol Ther. 2005 Sep;4(9):934–6.

9. Tutt A, Robson M, Garber JE, Domchek SM, Audeh MW, Weitzel JN, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. Lancet Lond Engl. 2010 Jul 24;376(9737):235–44.

10. Perkhofer L, Schmitt A, Romero Carrasco MC, Ihle M, Hampp S, Ruess DA, et al. ATM Deficiency Generating Genomic Instability Sensitizes Pancreatic Ductal Adenocarcinoma Cells to Therapy-Induced DNA Damage. Cancer Res. 2017 Oct 15;77(20):5576–90.

Fei Chen
948 Summitridge Dr,
Diamond Bar, CA 91765
Cell: 917-412-1568
Email: fchen13@jhu.edu
Date of Birth: 02/25/1985 (Fuzhou, China)

EDUCATION

Ph.D., Genetic Epidemiology, 2018
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
Dissertation: "Heritability and Genetic Architecture of Pancreatic Cancer"
Advisors: Dr. Alison Klein, Dr. Priya Duggal

Sc.M., Genetic Epidemiology, 2014
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

M.A., Biomedical Informatics, 2009
Columbia University, New York, NY

B.S., Biology, 2007
Fudan University, Shanghai, China

PUBLICATIONS

Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. Summit on Translational Bioinformatics. 2010;2010:1.

Chen F, Klein AP, Klein BE, Lee KE, Truitt B, Klein R, Iyengar SK, Duggal P. Exome array analysis identifies CAV1/CAV2 as a susceptibility locus for intraocular pressure. Investigative ophthalmology & visual science. 2015 Jan 1;56(1):544-51.

Szymanski EP, Leung JM, Fowler CJ, Haney C, Hsu AP, Chen F, Duggal P, Oler AJ, McCormack R, Podack E, Drummond RA. Pulmonary nontuberculous mycobacterial infection. A multisystem, multigenic disease. American journal of respiratory and critical care medicine. 2015 Sep 1;192(5):618-28.

Chen F, Duggal P, Klein BE, Lee KE, Truitt B, Klein R, Iyengar SK, Klein AP. Variation in PTCHD2, CRISP3, NAP1L4, FSCB, and AP3B2 associated with spherical equivalent. Molecular vision. 2016;22:783.

Chen F, Szymanski EP, Olivier KN, Liu X, Tettelin H, Holland SM, Duggal P. Whole-exome sequencing identifies the 6q12-q16 linkage region and a candidate gene, TTK, for pulmonary nontuberculous mycobacterial disease. American journal of respiratory and critical care medicine. 2017 Dec 15;196(12):1599-604.

Chen F, Roberts NJ, Klein AP. Inherited pancreatic cancer. Chinese clinical oncology. 2017 Dec 29;6(6).

Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, Canzian F, Childs EJ, Hoskins JW, Jermusyk A, Zhong J. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nature communications. 2018 Feb 8;9(1):556.

Loomis SJ, Klein AP, Lee KE, Chen F, Bomotti S, Truitt B, Iyengar SK, Klein R, Klein BE, Duggal P. Exome Array Analysis of Nuclear Lens Opacity. Ophthalmic epidemiology. 2018 May 4;25(3):215-9.

RESEARCH EXPERIENCE

Genetic Epidemiology, The Sidney Kimmel Comprehensive Cancer Center,
Baltimore, MD
2014 – 2018
Dissertation research with Dr. Alison Klein
Heritability analysis for pancreatic cancer using whole-genome sequencing data and imputation data.

Genetic Epidemiology, Johns Hopkins Bloomberg School of Public Health,
Baltimore, MD
2012-2018
Research Assistant to Dr. Priya Duggal
Exome array and exome sequencing studies on ocular disorders and infectious diseases.

Biomedical Informatics, Columbia University, New York, NY
2007-2009
Research Assistant to Dr. Chunhua Weng
Secondary use of electronic health records in public health research.

PRESENTATIONS

Poster: "Exome array analysis identified CAV1/CAV2 as a susceptibility locus for intraocular pressure" Presented at the Genomic of Common Disease annual meeting, September 2014, Potomac, MD

Poster: "Intraocular Pressure and Exome Sequencing Variation in the Beaver Dam Eye Study." (Abstract/Program #728) Presented at the 65th Annual Meeting of the American Society of Human Genetics, October 2015, Baltimore, MD.

Poster: "Whole exome sequencing and linkage analysis of patients with pulmonary non-tuberculous mycobacterial infection." Presented at the International Genetic Epidemiology Society Meeting, October 2015, Baltimore, MD

Poster: ""Transcriptomic profiling of normal pancreatic duct, precursor, and pancreatic adenocarcinoma organoids." Presented at the 2017 Genetic Research Day, February 2016, Baltimore, MD

Poster: "Whole exome sequencing and linkage analysis identified *TTK* as a candidate gene for pulmonary nontuberculous mycobacterial lung disease." (Abstract/Program #1318) Presented at the 66th Annual Meeting of the American Society of Human Genetics, October 2016, Vancouver, Canada.

Poster: "Heritability of familial pancreatic cancer using whole-genome sequencing data." Presented at the 2018 Genetic Research Day, February 2018, Baltimore MD

CONFERENCES ATTENDED

2016 The American Society of Human Genetics Annual Meeting, Vancouver, Canada
2016 The Genomic of Common Disease, Baltimore, MD
2015 The American Society of Human Genetics Annual Meeting, Baltimore, MD
2015 International Genetic Epidemiology Society Meeting, Baltimore MD
2014 The Genomics of Common Disease, Potomac, MD

TEACHING EXPERIENCE

Teaching Assistant, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
Principle of Genetic Epidemiology I (1st term 2013-2014, 1st term 2014-2015, Summer Institute 2015)
Principle of Genetic Epidemiology III (3rd term 2013-2014)
Epidemiology Method 2 (2nd term 2015-2016)

HONORS AND AWARDS

The Maryland Genetics, Epidemiology, and Medicine Training Grant (MD-GEM), 2017-2018
The Maryland Genetics, Epidemiology, and Medicine Training Grant (MD-GEM), 2016-2017
The Maryland Genetics, Epidemiology, and Medicine Training Grant (MD-GEM), 2015-2016
The Maryland Genetics, Epidemiology, and Medicine Training Grant (MD-GEM), 2014-2015
The Wolfe Street Competition Award, 2015-2016