From DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIOSTATISTICS
Karolinska Institutet, Stockholm, Sweden

EPIDEMIOLOGY AND ETIOLOGY OF PANCREATIC CANCER

Jiaqi Huang

Stockholm 2015

All previously published papers were reproduced with permission from the publisher.
Cover figure: Designed and drawn by Dr. Na Guan; post-processed by Dr. Qi Chen.
Published by Karolinska Institutet.
Printed by Eprint AB 2015

ISBN 978-91-7676-103-8

Institutionen för Medicinsk Epidemiologi och Biostatistik

EPIDEMIOLOGY AND ETIOLOGY OF PANCREATIC CANCER

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Samuelsson hall, Scheele laboratory, Tomtebodavägen 6, Solna Campus

Torsdag den 10 December 2015, kl 9:00

av

Jiaqi Huang

Principal Supervisor:
Professor Weimin Ye
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):
Professor Johannes-Matthias Löhr
Karolinska University Hospital
Department of Clinical Science, Intervention and Technology
Division of Surgery

Docent Magnus Nilsson
Karolinska University Hospital
Department of Clinical Science, Intervention and Technology
Division of Surgery

Opponent:
Professor Gloria Petersen
Mayo Clinic
Department of Health Sciences Research
Division of Epidemiology

Examination Board:
Professor Scott Montgomery
Örebro University
Department of Medicine and Health
Division of Clinical Epidemiology and Biostatistics

Professor John Öhrvik
Karolinska Institutet
Department of Medicine

Docent Claes Jönsson
University of Gothenburg
Department of Surgery
Sahlgrenska Academy

To my beloved family

ABSTRACT

Pancreatic cancer is one of the most devastating malignancies with an extremely high fatality, resulting that its mortality rate almost equals to incidence rate. Although primary prevention is of upmost importance, the underlying etiology of this cancer remains largely unknown. Pancreatic cancer is a heterogenetic disease, and the accumulated genetic alterations play an important role in pancreatic pathogenesis. Recent advances in next-generation sequencing have enabled comprehensive cancer genomic studies. However, clinical pancreatic cancer samples are characterized as having low tumor cellularity, as a result of an abundance of stroma in the tumor microenvironment, and this presents a big challenge for direct genomic sequencing for clinical pancreatic cancer samples. In this thesis, we aimed to enrich our knowledge of the etiology of pancreatic cancer with regard to several infectious agents and poor oral hygiene. Of note, we took the challenge to directly sequence clinical pancreatic cancer samples with a broad range of tumor cellularities, and attempted to depict its variant profile.

In **Study I**, we retrieved all hepatitis C virus (HCV) and hepatitis B virus (HBV) infection notifications in Sweden from records in a national surveillance database at the Swedish Institutet for Infectious Disease Control (SMI) from 1990 to 2006, and followed them for pancreatic cancer occurrence by the end of 2008. The pancreatic cancer risk in the exposed population was compared with that in a matched reference population. Hazard ratios (HRs) were derived from Cox proportional hazards regression models. The main finding in this study is that the subjects with HCV infection had a 60% increased risk after adjustment for potential confounders. Therefore, the finding implied that HCV infection may be associated with a higher pancreatic cancer risk but further studies are warranted to confirm the observed association. The point estimate in this study also suggested an excessive risk among subjects with HBV infection, however, without statistical significance due to a lack of study power.

In **Study II**, we took advantage of the population-based prevalence study of oral mucosal lesions conducted in Uppsala County in central Sweden during 1973-74. The study population was followed through linkages with the Swedish population and health registers. A total of 19 924 participants were included in the final analysis, with 126 pancreatic cancer ascertained during an average of 28.7 years of follow-up. Among all tested indicators of poor oral hygiene, we found that fewer teeth at baseline appeared to increase pancreatic cancer risk, although the relative risk estimates were not statistically significant. Among the subjects with more than 10 teeth, subjects with unacceptable dental plaque had a doubled risk of pancreatic cancer compared with those without dental plaque after controlling for potential confounding factors. Subjects with *Candida*-related or denture-related oral mucosal lesions, or tongue lesions, compared with those without any of the three lesions, showed a 70%, 30% and 80% increased pancreatic cancer risk, respectively.

In **Study III**, we carried out a nested case-control study within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, including 448 pancreatic cancer cases

and their individually matched control subjects. We measured serum antibodies against *Helicobacter pylori* (*H. pylori*) and pepsinogens I and II (markers for presence of chronic corpus atrophic gastritis) by enzyme-linked immunosorbent assays. Conditional logistic regression models were used to estimate odds ratios (ORs). Overall, our results demonstrated that pancreatic cancer risk was neither associated with *H. pylori* seropositivity nor CagA seropositivity. On the other hand, our findings showed that presence of chronic corpus atrophic gastritis was non-significantly associated with an increased pancreatic cancer risk. Although based on small numbers, the association was particularly prominent among individuals seronegative for both *H. pylori* and CagA (OR=5.66; 95% confidence interval: 1.59, 20.19; *p* value for interaction <0.01).

In **Study IV**, we conducted a case-only study that sourced from a population-based case-control study of pancreatic cancer in Stockholm, Sweden. This study included patients with pancreatic ductal adenocarcinoma (PDAC) who underwent resection surgery between 2007 and 2012 (n=73). Patients were followed from diagnosis until death or the end of the study. We used an Anchored Multiplex Polymerase chain reaction (AMP)-based method for profiling variants in a panel of 65 selected genes. Our findings suggested that the AMP-based next-generation sequencing method can detect variants with allelic frequencies as low as 1% given sufficient sequencing depth. *KRAS* G12 mutations were completely confirmed by Sanger sequencing for high-allele-frequency samples (>5%), and also fully confirmed by allele-specific PCR and digital PCR for low-allele-frequency samples (1%-5%). The results demonstrated that *KRAS* mutant subtype G12V is related to a worse prognosis in PDAC patients, and transversion variants are more common among smokers.

In conclusion, we found that HCV, as an infectious agent, may be associated with a higher pancreatic cancer risk. Our findings also support the hypothesis that poor oral hygiene plays a key role in the development of pancreatic cancer. On the other hand, we observed a null association between *H. pylori* infection and pancreatic cancer risk in the western European populations, but a suggested positive association between chronic corpus atrophic gastritis and pancreatic cancer risk based on a small sample size. Further studies are warranted to verify whether severe gastric atrophy contributes to pancreatic carcinogenesis. AMP-based next generation sequencing is a sensitive and accurate method for profiling tumor variants in PDAC. Future studies with larger sample sizes are needed to explore the role of tumor variants in PDAC prognosis and the impact of environmental risk factors on tumor mutational profile.

LIST OF SCIENTIFIC PAPERS

I. Huang J, Magnusson M, Törner A, Ye W, Duberg AS. Risk of pancreatic cancer among individuals with hepatitis C or hepatitis B virus infection: a nationwide study in Sweden. Br J Cancer. 2013 Nov 26;109(11):2917-23.

II. Huang J, Roosaar A, Wilson K, Axéll T, Ye W. A prospective cohort study on poor oral hygiene and pancreatic cancer risk. Int J Cancer. 2015 Aug 1. doi: 10.1002/ijc.29710. [Epub ahead of print]

III. Huang J, Zagai U, Hallmans G, Nyrén O, Engstrand L, Stolzenberg-Solomon R, Duell E. J, Overvad K, Katzke V. A, Kaaks R, Jenab M, Young Park J, Murillo R, Trichopoulou A, Lagiou P, Bamia C, Bradbury K, Riboli E, Capellá G, [coauthors from EPIC working groups], Bueno-de-Mesquita H.B(as), Ye W. *Helicobacter pylori* infection and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Manuscript*

IV. Huang J, Löhr M, Nilsson M, Segersvärd R, Matsson H, Verbeke C, Heuchel R, Kere J, Iafrate AJ, Zheng Z, Ye W. Variant profiling of candidate genes in pancreatic ductal adenocarcinoma. Clin Chem. 2015 Nov;61(11):1408-16.

CONTENTS

1 Introduction 1
2 Background 3
2.1 Descriptive epidemiology of pancreatic cancer 3
2.2 Clinical aspects of pancreatic cancer 3
2.2.1 Symptoms 3
2.2.2 Diagnosis 3
2.2.3 Subtypes of pancreatic cancer 4
2.3 Risk factors 4
2.3.1 Hepatitis C virus (HCV) or hepatitis B virus (HBV) infection 5
2.3.2 Poor oral hygiene 5
2.3.3 *Helicobacter pylori* (*H. pylori*) infection 5
2.3.4 Chronic corpus atrophic gastritis 6
2.4 Current hypothesis for pathogenesis 6
2.4.1 Inflammation 6
2.4.2 *N*-nitrosamine compounds 6
2.4.3 Insulin resistance 7
2.4.4 Genetic susceptibility 7
2.5 Tumor mutations 7
3 Aims 11
4 Materials and methods 13
4.1 Study materials 13
4.1.1 Swedish data sources 13
4.1.2 European data sources 15
4.2 Identification of pancreatic cancer 15
4.3 Laboratory method 16
4.3.1 Commercially available test: Enzyme linked immunosorbent assays (ELISAs) (Study III) 16
4.3.2 Anchored Multiplex PCR (AMP) method (Study IV) 17
4.4 Statistical analyses 18
4.4.1 Standardized incidence ratio (Study I) 18
4.4.2 Cox proportional hazards regression model (Study I, II) 19
4.4.3 Logistic regression model (Study III, IV) 19
4.4.4 Survival analysis (Study IV) 20
5 Results 21
5.1 Study I 21
5.2 Study II 23
5.3 Study III 26
5.4 Study IV 29
5.4.1 Validation of the AMP method by sequencing a quantitative multiplex DNA reference standard 29
5.4.2 Frequency and distribution of genetic aberrations 30
5.4.3 Survival analysis according to *KRAS* mutation status 37
5.4.4 Association between tobacco smoking and genetic alterations 37

6 Discussion 39
6.1 Epidemiological considerations 39
6.1.1 Study design 39
6.1.2 Bias 40
6.1.3 Confounding 41
6.2 Strengths of the studies 42
6.3 Interpretations of findings 42
6.3.1 Pancreatic cancer risk assessment 42
6.3.2 Variant profiling in PDAC 45
7 Conclusion 47
8 Future perspectives 48
9 Acknowledgements 49
10 References 53

LIST OF ABBREVIATIONS

AMP	Anchored Multiplex PCR
bp	Base pair
CA19-9	Carbohydrate antigen 19-9
Cag A	Cytotoxin-associated gene A
CI	Confidence interval
CT	Computer tomography
ctDNA	Circulating tumor DNA
EIU	Enzyme immunounits
ELISA	Enzyme-linked immunosorbent assay
EPIC	The European Prospective Investigation into Cancer and Nutrition
ERCP	Endoscopic retrograde cholangiopancreatography
EUS	Endoscopic ultrasound
FNA	Fine needle aspiration
GWAS	Genome-wide association study
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HR	Hazard ratio
H. pylori	*Helicobacter pylori*
ICD-7	The 7th edition of International Classification of Diseases
IPMN	Intraductal papillary mucinous neoplasm
MRI	Magnetic resonance imaging
NGS	Next-generation sequencing
NRN	National registration number
OD	Optical density
OR	Odds ratio
PCR	The polymerase chain reaction

PET	Positron emission tomography
PDAC	Pancreatic ductal adenocarcinoma
PG	Pepsinogen
SAS	Statistical Analysis System
SES	Socioeconomic status
SIR	Standardized incidence ratio

1 INTRODUCTION

Pancreatic cancer is one of the most aggressive and devastating malignancies. Although considerable efforts have been made to prevent and cure this lethal disease, it still remains a major health burden worldwide, given its rapid progression, invasiveness and resistant to treatment. Due to its insidious onset, it is often diagnosed at an advanced stage, when curative surgery is no longer an option. Although pancreatic cancer is not a common malignancy, because of its high fatality rate, it ranks as the 4th or 5th most common cause of cancer-related death in developed countries, such as in Europe (1, 2). Established risk factors (3, 4) include advanced age, male sex, smoking, chronic pancreatitis, type 2 diabetes, central obesity and a family history of pancreatic cancer. However, it has been estimated in UK that the identified risk factors only account for approximately 40% of all pancreatic cancer cases (5). The etiology of pancreatic cancer remains unclear. Therefore further studies are urgently warranted to identify its risk factors, as well as to understand its genetic heterogeneity, given their key roles in the pancreatic carcinogenesis.

Recent advances in next-generation sequencing (NGS) have enabled genetic architectural understanding of cancers, including pancreatic cancer. Pancreatic cancer is a heterogeneous disease at the molecular level, and results from successive accumulation of genetic mutations from distinct types of pancreatic precursor lesions, such as pancreatic intraepithelial neoplasia which is the most common type of precursor lesion. However, a major feature of pancreatic cancer is the formation of abundant stroma, termed as desmoplastic reaction. It presents a great challenge for genomic sequencing for this tumor owing to its low tumor cellularity and thus has limited earlier studies due to dilution of the tumor cells of interest present in an overwhelming wild-type background. Therefore, studies are merited particularly with regards to deep targeted sequencing method that may provide higher analytical sensitivity with a broader coverage for clinical samples.

2 BACKGROUND

2.1 DESCRIPTIVE EPIDEMIOLOGY OF PANCREATIC CANCER

Pancreatic cancer is one of the most devastating malignancies. Its incidence ranges from 1 to 10 cases per 100 000 person-year worldwide. The incidence of pancreatic cancer is generally higher in the developed countries than the developing countries. Globally in 2007 (6), the highest mortality rates for pancreatic cancer were in the Baltic countries (more than 9.5/100 000 among men and 6/100 000 among women), whereas Latin America and Hong Kong had the lowest mortality rates (below 5/100 000 among men and 3/100 000 among women). The overall 5-year survival rate for pancreatic cancer patients is less than 5% (2, 7-9). Owing to its high fatality, pancreatic cancer ranks the 7th most common cause of cancer-related death throughout the world, and the 4th or the 5th in the developed countries, including Europe (2). Pancreatic cancer mortality rates were increasing in the developed countries from the 1950s to the 1980s, and leveled off or declined soon thereafter, especially in men. However, it is difficult to explain the secular trend, since the apparent increase might largely be interpreted as an improvement of diagnostic methods, rather than occurrence of more cases (10).

Pancreatic cancer is more common in the old age population, and the median age at diagnosis is 71 years. Therefore, it has been estimated that the global burden of pancreatic cancer will increase given the improvement of life expectancy in the coming decades (3). For men, the cumulative probability of developing pancreatic cancer through lifetime is 1%, and slightly lower for women, given the fact that men have a higher smoking rate than women (10).

2.2 CLINICAL ASPECTS OF PANCREATIC CANCER

2.2.1 Symptoms

The presenting symptoms of pancreatic cancer are largely determined by the location of the tumor in the pancreas, as well as the disease stage. Due to its insidious onset, pancreatic cancer often escapes from detection during its formative stage, and cannot be detected until an advanced stage, when a curative resection surgery is no longer possible. Approximately 60 to 70% of pancreatic cancers present in the head of the gland, and result in obstructive cholestasis and jaundice. Common symptoms in pancreatic cancer patients include abdominal pain, weight loss, asthenia and anorexia (11).

2.2.2 Diagnosis

It has always been a challenge for the diagnosis of pancreatic cancer, owing to its deep location. Imaging technologies including computer tomography (CT) scan, magnetic resonance imaging (MRI), positron emission tomography (PET), endoscopic retrograde cholangiopancreatography (ERCP), as well as endoscopic ultrasound (EUS) are critical diagnostic tools for pancreatic cancer. Studies evaluating the diagnostic ability of endoscopic ultrasound, computed tomography, and MRI in patients suspected of harboring pancreatic cancer have revealed a sensitivity of 94%, 69%, and 83%, respectively (12). Among the

diagnostically uncertain cases, EUS-guided fine needle aspiration (FNA) should be further employed for cytological assessment; however, its effectiveness still requires further evaluation.

As for potential serum biomarkers for diagnosis, carbohydrate antigen 19-9 (CA19-9) probably is the only biomarker with clinical application for the early detection or monitoring of recurrent disease (13-17). CA 19-9 is produced in the biliary cells, therefore the diseases that affect those cells will lead to elevation of serum CA 19-9, such as pancreatitis, cirrhosis and cholangitis (11). In addition, approximately 10% of the patients are unable to produce CA 19-9, with an undetectable level of CA 19-9 even in advanced stage of this disease (9). With the sensitivity of 80% and specificity of 73%, it is still premature to use this biomarker alone for a diagnostic purpose (18).

2.2.3 Subtypes of pancreatic cancer

Pancreatic cancer comprises exocrine and endocrine tumors (19). The exocrine pancreatic tumors are more common, accounting for approximately 95% of all pancreatic cancers, and encompass adenocarcinoma, acinar cell carcinoma, adenosquamous carcinoma, giant cell tumor, intraductal papillary mucinous neoplasm (IPMN), mucinous cystadenocarcinoma, pancreatoblastoma, serous cystadenocarcinoma, solid and pseudopapillary tumors (11). Over 85% of pancreatic cancers are pancreatic ductal adenocarcinoma (PDAC). This thesis has an emphasis on exocrine pancreatic cancer, and **Study IV** focuses on the PDAC subtype only.

2.3 RISK FACTORS

A better understanding of the etiological factors is critical to identify high-risk groups, and for the prevention and early detection of the disease. A study in UK estimated that the current identified risk factors account for less than 40% of all pancreatic cancer cases. Instead of one dominant factor, multifactorial environmental factors are involved in contributing to pancreatic cancer development. To date, established risk factors for pancreatic cancer include advanced age (1), male sex (20), tobacco smoking (21-25), chronic pancreatitis (26, 27), type 2 diabetes mellitus (21, 28-33), obesity (5, 34-38), a family history of pancreatic cancer (39-42), non-O blood group (43, 44), heavy alcohol consumption (45-48), height (tallness) (21, 49-52), as well as high waist-to-hip ratio (49, 53). On the other hand, previous studies have indicated that allergy (54, 55), high adiponectin level (56), intense occupational physical activity (57, 58), and high dietary folate intake (59-62) may have a potentially protective effect against pancreatic cancer. The evidence for associations between several other factors and pancreatic cancer risk is less established, including nonsteroidal anti-inflammatory drugs (63-66), fish (67, 68), soft drinks (69, 70), coffee or tea (70-72), glycemic index or glycemic load (73, 74), and plasma 25(OH)D level (75-77). Other potential risk factors which have been examined but still need further confirmation, including idiopathic thrombosis (78) , Scandinavian moist snuff (79), history of gastrectomy (80), history of cholecystectomy (81), red meat (67, 82), processed meat (82), elevated sugar intake (74) and psychological stress (83).

2.3.1 Hepatitis C virus (HCV) or hepatitis B virus (HBV) infection

Approximately 170 and 350 million people are infected with HCV and HBV throughout the world (84, 85). The geographical variations for the prevalence of HCV or HBV infection are relatively large, ranging from less than 0.5% in Western countries, 7% in the East Asian, and up to 25% in the African countries (86, 87). As a low endemic country, it was estimated in Sweden that 0.5% of the population were infected with HCV, and even lower for HBV infection (88). Prior studies demonstrated that HBV was able to replicate within the pancreas gland (89, 90), and HCV infection was associated with pancreatitis (91, 92). The pancreas is a potential extrahepatic organ for inhabitation of the hepatitis virus, which makes it biological plausible that the HCV or HBV infection might be associated with an increased risk of pancreatic cancer. However such associations are still unclear and further studies are warranted, for a better understanding of the etiology of pancreatic cancer.

2.3.2 Poor oral hygiene

The oral cavity connects between the gastrointestinal tract and external environment. Poor oral hygiene, through oral microbiome alterations, may lead to a series of events, including dental plaque, periodontal disease and tooth loss (93). Loss of tooth may be used as a proxy of bacterial load on the teeth (93). Dental plaque is formed from bacterial biofilm that adheres to the surface of teeth. It was estimated that hundreds of bacterial species inhabit in the subgingival plaque. Although oral mucosal lesions are multifactorial diseases, the oral mucosal pathogens are considered as one of the most important etiologic risk factors. Poor oral hygiene has been previously proposed to be positively associated with pancreatic cancer risk in several epidemiologic studies (93-97). Three large cohort studies provided supportive evidence for the link between periodontal disease and pancreatic cancer risk (94, 95, 98). However, the findings were largely inconsistent for the association between tooth number and pancreatic cancer risk (98, 99). The aforementioned studies were based mostly on self-reported oral hygiene indicators and only one study contained repeated measurement during follow-up. In two studies with direct dental examination, the pancreatic cancer cases were ascertained from death certificate records, and the results were based on rather small case numbers (94, 95). Therefore, the association between poor oral hygiene and pancreatic cancer still awaits confirmation by larger cohort studies with detailed oral examinations.

2.3.3 *Helicobacter pylori* (*H. pylori*) infection

H. pylori has been identified as a group I carcinogen by IARC (100), and it is widely considered to play a crucial role in the pathogenesis of noncardia gastric cancer (4). A subgroup of *H. pylori* strains contains a gene associated with cytotoxin expression, namely *CagA* (cytotoxin-associated gene A). Compared to *CagA* negative strain, the *CagA* positive strain was deemed to further increase gastric cancer risk, through enhanced inflammatory responses (101). The association between *H. pylori* and pancreatic cancer has been addressed in a number of previous studies. Yet, the findings were largely inconsistent across studies. In a prior meta-analysis including four European studies (102), a pooled 56% excessive

pancreatic cancer risk was suggested, whereas an updated meta-analysis containing seven studies from Western countries could not substantiate the positive association (103). However, a strain-specific association between *H. pylori* (CagA+ or CagA-) and pancreatic cancer risk was indicated in a recent meta-analysis study (104).

2.3.4 Chronic corpus atrophic gastritis

Chronic corpus atrophic gastritis is identified as a precursor lesion of the intestinal type of noncardia gastric cancer (105), and it is characterized by a series of events during chronic gastric inflammation, including gland loss, mucosa thinning and epithelia cell regeneration and replacement (106). Autoimmune pernicious anemia, *H. pylori* infection and long-term treatment with proton pump inhibitors are identified as risk factors for chronic corpus atrophic gastritis (106-108). Hypergastrinemia and low serum pepsinogen levels have been characterized in chronic corpus atrophic gastritis. Practically, measurement of serum pepsinogens has been widely accepted for identification of presence of chronic atrophic gastritis in the general population, due to its economical and noninvasive features. We hypothesized that chronic corpus atrophic gastritis may be associated with an increased pancreatic cancer risk, through a mechanism of low-acid-production in the stomach.

2.4 CURRENT HYPOTHESES FOR PATHOGENESIS

2.4.1 Inflammation

It is widely accepted that inflammation plays a key role in the development of pancreatic cancer (109-111). The link between local inflammation and pancreatic cancer comes from studies on the pancreatic carcinogenetic effects of chronic pancreatitis, which is a well-established risk factor for pancreatic cancer (112-114). Other established risk factors, including smoking and obesity, may increase pancreatic cancer risk through promoting systemic inflammation (111). In addition, the microbiome appears to be linked to the cancer-related inflammation and plays a pathogenic role in the process of inflammation. Several studies suggested that pathogenic infection is likely to convert the commensal bacterial population from a symbiotic state to a dysbiotic state, which will lead to the formation of systemic inflammation (115, 116). Inflammation is thought to play a key role in the etiology of pancreatic cancer, as it fosters a microenvironment that facilitates accumulation of genetic alterations and initiates carcinogenesis; the sustained microenvironment thus favors for pancreatic cancer development (110, 117). Further, tumor cells are developed to evade immune surveillance, including down-regulating the tumor-specific antigens or suppressing the antitumor immune cells, which results in immune evasion and escape (118, 119).

2.4.2 *N*-nitrosamine compounds

Studies in animal models during the past three decades have demonstrated that *N*-nitrosamine carcinogens are able to induce pancreatic cancer (120-123). *N*-nitrosamine compounds may reach the pancreas either through the duodenum or through the bloodstream. The ductal epithelium is able to metabolically activate the *N*-nitrosamine carcinogens (124). The

activated carcinogens may damage DNA by generating small DNA adducts or causing single-strand breaks (125). Long-term exposure of *N*-nitrosamine compounds overwhelms DNA repair capabilities, which exerts a pancreatic carcinogenesis effect. Usually, *N*-nitrosamine is at extremely low level in the human environment. High levels of *N*-nitrosamines are thought to be accumulated through tobacco products and dietary sources (120).

2.4.3 Insulin resistance

It has been hypothesized that exposure to a higher level of insulin, as well as insulin resistance, may increase the exocrine pancreatic cancer risk (126). Due to the anatomical proximity between exocrine cells and the islets of Langerhans, it has been estimated that the exocrine cells are exposed to insulin concentrations 20-fold higher than the systemic circulation, which may facilitate pancreatic cancer cell initiation, invasion and progression (127). Numerous studies in animal models and human cancer cell lines indicated that peripheral insulin resistance may exert a ductal pancreatic carcinogenesis effect (128-132) . In a recent epidemiologic study with five prospective US cohorts, the findings implied that peripheral insulin resistance, rather than hyperglycemia or pancreatic β-cell dysfunction, is independently related to an increased pancreatic cancer risk (133).

2.4.4 Genetic susceptibility

Both inherited high-penetrance mutations and low-penetrance loci increases pancreatic cancer risk. The high-penetrance mutations are characterized by familial aggregation (134, 135), and previous studies have identified several germline mutations in *BRCA2*, *CDKN2A*, *STK11*, *APC*, *BRCA1*, *PRSS1*, *SPINK*. However these mutations account only for a small fraction of all pancreatic cancer cases (8-10%). On the other hand, multiple common susceptibility variants have been identified in genome-wide association studies (GWAS), and were reported to be associated with an altered pancreatic cancer risk (136-139), including 9q34.2 (*ABO*), 13q22.1 (*KLF5*), 5p15.33 (*TERT* and *CLPTM1*), 13q12.2 (*PDX1*), 1q32.1 (*NR5A2*), 7q32.3 (*LINC-PINT*), 16q23.1(*BCAR1*), 22q12.1 (*ZNRF3*), 17q25.1 (*LINC00673*) and 2p13.3 (*ETAA1*). The identified susceptibility loci highlight the importance of the common variants in pancreatic cancer risk, and may provide deep insights in risk stratification, early detection and targeted treatment for this lethal cancer. Biological underpinnings of those common variants, as well as exploration of the interactions between these risk variants and environmental risk factors, await further follow-up studies.

2.5 TUMOR MUTATIONS

Pancreatic cancer is a heterogenetic disease, and accumulated genetic alterations play a key role in the pancreatic carcinogenesis, promoting the progression from a normal cell to precursor lesions and ultimately to invasive and metastatic carcinoma. Pancreatic cancer is characterized by a dominance of *KRAS* mutations, the majority being point mutations, and the minority either insertions or deletions. *KRAS* is the most commonly mutated oncogene in

PDAC (140, 141), and it encodes a small GTPase that regulates downstream signaling. Commonly, the hotspot mutation for *KRAS* is on codon 12, and less frequently mutations harbored on codon 13 and codon 61. In addition, three tumor suppressor genes, including *TP53*, *CDKN2A*, and *SMAD4*, are found to be recurrently mutated in PDAC. *TP53* gene encodes a crucial component of the cellular stress response, and the mutation of *TP53* usually causes diffusion of nuclear expression by TP53 protein, which can be detected by immunohistochemistry (142). *CDKN2A* gene encodes an important cell cycle regulator, and it is reported to be inactivated through several mechanisms, including intragenic mutation, homozygous deletion, as well as promoter methylation (143) . *SMAD4* gene encodes a critical component in the transforming growth factor β signaling pathway. Somatic inactivation of *SMAD4* is usually through homozygous deletion or intragenic mutation, and is found to be associated with worse prognosis and extensive metastasis (144).

Recent advancements in technology of NGS have allowed comprehensive cancer genomic studies. In 2008, the mutational landscape of pancreatic adenocarcinoma was depicted by Jones et al. (140), including a range of molecular alterations in 63 genes that affect a core set of 12 cellular signaling pathways in most tumors. Later, Biankin et al. identified 16 genes of both known and novel mutations, according to exome sequencing on 99 early-stage PDACs (141). In a study to evaluate the clonal relationships among primary and metastatic cancer on the basis of genomic sequencing of seven tumors, the authors modelled and estimated that it requires an average of 11.7 years for the evolution from the initiation of pancreatic carcinogenesis until the birth of parental clone cell, and another 6.8 years from the birth of parental clone cell to the birth of cell with metastasis ability (145). The data provided novel insights into the genetic features of pancreatic cancer progression, and defined a broad time window for early detection when the disease is still in a curative stage.

One characteristic of pancreatic cancer is the formation of abundant stroma, containing proliferating myofibroblasts. For pancreatic cancer, the stroma is not only a mechanical barrier, but also contributes to promoting the tumor initiation, progression, invasion and metastasis. This presents a major challenge for genomic sequencing in PDAC, due to the mixture of a large proportion of stromal cells with the tumor cells. The low malignant epithelial cell content largely limits the analytical performance of various sequencing methods in PDAC. Commonly, pancreatic tumor cells are transferred into an immortalized cell line in culture, yielding tumor cell population with high purity. Alternatively, the pancreatic tumor cells are transferred into xenografting mice to eliminate stromal cells. However, it is still an open question if additional genetic modifications may be produced in the creation of these immortalized cell lines or during the xenografting. Thus these samples may not be fully representative of the spectrum of heterogeneity of the original primary tumors. Further, it is also time-consuming and, importantly, not every tumor is able to generate a successful xenograft. Therefore, sensitive and accurate methods that can be directly applied to the low-tumor-cellularity samples are urgently warranted to identify the genetic aberrations in pancreatic tumors.

Genetic analysis may provide deep insights into the etiology of PDAC. It is also crucial for the development of novel therapies as it may help stratify patients with well-characterized molecular markers into subgroups which, in turn, can be used for development of personalized treatments targeting specific pathways.

3 AIMS

The overarching aim of this thesis is to enrich our understanding of the etiology and epidemiology of pancreatic cancer.

The specific aims are:

- To explore the association between hepatitis C or hepatitis B virus infection and pancreatic cancer risk.
- To determine whether poor oral hygiene is associated with a higher pancreatic cancer risk.
- To evaluate the possible association between *H. pylori* infection and pancreatic cancer risk, as well as to examine whether presence of chronic corpus atrophic gastritis is related to an increased pancreatic cancer risk.
- To explore molecular alterations of the selected candidate genes in PDAC, and to examine their associations with prognosis and environmental exposures.

4 MATERIALS AND METHODS

4.1 STUDY MATERIALS

4.1.1 Swedish data sources

4.1.1.1 The Swedish Cancer Register

The Swedish Cancer Register was launched in 1958 and is maintained by the National Board of Health and Welfare, using from the 7th edition of International Classification of Diseases (ICD-7) throughout. It is mandatory for health care providers, including physicians, pathologists and cytologists, to report newly diagnosed cancer cases from clinical, morphological and other laboratory examinations and autopsies. Staging information has been systematically included since 2004. The estimated overall completeness of the Swedish Cancer Register is high (>98%), but completeness rates may vary according to cancer sites (146, 147). Death certificate only cases are not included in the Swedish Cancer Register.

4.1.1.2 The Swedish Causes of Death Register

The Swedish Causes of Death Register is on the basis of death certificates, including information on the date of death, the underlying and contributory causes of death for all Swedish residents, with nationwide coverage since 1911. The causes of death are classified on the basis of ICD-7 (through 1968), ICD-8 (1969-1986), ICD-9 (1987-1996) and ICD-10 (since 1997).

4.1.1.3 The Swedish Inpatient Register

The Swedish Inpatient Register was initiated by the Swedish National Board of Health and Welfare in 1964-1965. Its national coverage was 60% in 1969, 85% in 1983, and 100% in 1987. The register contains information on the dates of admission and discharge and also diagnoses at discharge. The diagnoses were coded according to ICD-7 (through 1968), ICD-8 (1969-1986), ICD-9 (1987-1996) and ICD-10 (since 1997).

4.1.1.4 The Swedish HCV and HBV cohorts

In Sweden, both the diagnosing laboratory and clinician are required to report diagnosed HCV and HBV infections to the SMI since 1990 and 1969, respectively. The notifications from various sources were merged in the national surveillance database through the use of Personal Identification Number. We established HCV and HBV cohorts based on the identified HCV and HBV infection notifications in the surveillance database in SMI from 1990 to 2006 (148, 149).

The HCV notifications from the diagnosing laboratory are based on a positive HCV-RNA or anti-HCV test, while HBV notifications are based on positive HBsAg or HBV-DNA. The reports from clinicians include further information, for example, probable route of

transmission and whether it is an acute or chronic infection of HBV. This is determined by clinical information and medical history.

Exclusion criteria included multiple notifications, erroneous documentation, incomplete personal follow-up data, and those with acute HBV infection (N=2050), those with co-infection that were reported to both the HCV and HBV register (N=3556). Overall, a total of 39 442 patients with HCV infection and 11 511 patients with chronic HBV infection were enrolled.

To identify an appropriate reference population, a quintuple larger reference population was generated from the Total Population Register at Statistics Sweden. The individuals in the reference population were those who had never been diagnosed with a HCV or HBV infection, and were randomly selected from the general population, five to one matched with HCV- or HBV-infected patients on year of birth, sex and county of residence in Sweden. In **Study I**, follow-up started on the entry date, either their HCV/HBV notification date or the assigned index entry date (non-infected reference population), until the date of first primary cancer diagnosis, emigration out of Sweden, death or December 31, 2008, whichever occurred first.

4.1.1.5 The Swedish oral-mucosal-lesion cohort

A population-based prevalence study of oral diseases was conducted in Uppsala county of Sweden from 1973 to 1974. At baseline, a professional clinical examination and questionnaire were offered to 30 118 individuals who were aged 15 or more and residing in this area. Overall, a total of 20 333 (10 036 men and 10 297 women) accepted to participate in the study. During the clinical examination, information such as number of teeth, dental plaque and oral mucosal lesions, were collected (150).

A total of 121 persons were excluded during the cross-linkage due to the following reasons: individuals' national registration numbers (NRN) were incorrect (N=84); individuals had changed NRNs during follow-up period (N=14); individuals had illegible data (N=8); individuals were inadvertently examined twice (omit data from the second examination; N=15), leaving a total of 20 212 eligible individuals. After linkage to the Cancer Register, 288 individuals were further excluded given the reason that they had a cancer diagnosis before the entry date. In total, 19 924 participants were included in the study cohort for final analysis.

In **Study II**, follow-up of the cohort members started from the date of dental examination (1973-74), and ended at the date of diagnosis of a first malignancy, death, emigration out of Sweden, or December 31, 2012, whichever occurred first.

4.1.1.6 The Stockholm pancreatic cancer study

This study was based on the incident cases of pancreatic cancer in the Stockholm county between 2007 and 2012. Criteria of the eligible cases were: younger than 85 years of age,

born in Sweden and living in the county of Stockholm. Through a network that consisted of all surgical clinics and the only oncological clinic in Stockholm county, the study was able to obtain a majority of newly diagnosed pancreatic cancer cases in Stockholm county in this period. When a pancreatic cancer case was diagnosed, the individual was asked whether he/she would like to take part in the study. After informed consent was obtained, the study coordinator at Karolinska University Hospital then informed Statistics Sweden professional interviewers to interview the case patient. The questionnaire addresses items on demographics, socioeconomic factors, physical activity, height, weight history, tobacco smoking, snuff use, dental health, earlier concomitant diseases, and medication history. A structured food frequency questionnaire was also included.

Based on the Stockholm pancreatic cancer case-control study, the pancreatic cancer patients undergoing resection surgery from 2007 to 2012 were enrolled in this study. Fresh-frozen pancreatic cancer tissues from surgical resection (N=73) and matched peripheral blood samples (N=55) were obtained from Karolinska University Hospital. Through the NRNs, tumor characteristics and survival time of all participating patients were retrieved via the electronic patient record system at Karolinska University Hospital, which consists of detailed information on tumor stage, chemotherapy, date of surgery, and date of death. Data from the aforementioned interviews were available for 57 out of 73 PDAC patients.

4.1.2 European data sources

The European Prospective Investigation into Cancer and Nutrition (EPIC) is a large prospective cohort study. From 1992 to 2000, the study recruited 519 978 participants who mostly aged between 35 to 70 years from 23 centers in 10 European countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Spain, Norway, Sweden and the United Kingdom). A non-dietary questionnaire on lifestyle variables and a dietary questionnaire with detailed dietary items were collected at enrolment for all participants. Anthropometric measurements were conducted. Blood samples were also taken, including plasma, serum, red cells and buffy coat fractions, which were separated and aliquoted for long term storage (151). Follow-up of EPIC study started from participants' recruitment date to the date of cancer diagnosis, death, emigration, or December 31, 2006, whichever came first.

We conducted a nested case-control study within the EPIC cohort study. With the incidence density sampling procedure, cancer-free control subjects were randomly selected and individually matched to each case patient on study center, age (± 3 years), sex, date (± 3 months), time(± 2 h), and fasting status (<3 h, 3–6 h or >6 h) at blood collection. To the end of 2006, 448 cases and 448 controls were enrolled and eligible in this study.

4.2 IDENTIFICATION OF PANCREATIC CANCER

For **Study I and II**, Pancreatic cancer was identified from the Cancer Register (ICD-7: 157) for the main analyses, and also from the Causes of Death Register (ICD-9: 157; ICD-10: C25) for additional analyses.

For **Study III**, incident pancreatic cancer cases were identified through population cancer registers in Denmark, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom. In France, Germany and Greece, a combined approach was applied including linkage to health insurance records, cancer and pathology registers, and active follow-up of study participants and their next-of-kin. A total of 578 primary incident pancreatic cancer cases were identified using International Classification of Diseases, 10th Revision codes (ICD-10, C25.0-25.3, 25.7-25.9) by the end of 2006. Given the different etiology and pathogenesis, endocrine pancreatic tumor cases (ICD -O-3 C25.4, histologic type and morphology codes 8150, 8151, 8153, 8155, 8240 and 8246) were excluded from the study. Further, cases without blood samples were excluded, leaving a total of 448 eligible pancreatic cancer cases.

For **Study IV**, PDAC patients were ascertained within the framework of the clinical healthcare system of the Stockholm County.

4.3 LABORATORY METHOD

4.3.1 Commercially available test: Enzyme linked immunosorbent assays (ELISAs) (Study III)

4.3.1.1 *H. pylori* serostatus

We measured serum concentration of antibodies against whole *H. pylori* antigens using a commercial *H. pylori* IgG ELISA kit (Biohit, Helsinki, Finland). The enzyme immunounits (EIU) were determined as following: sample EIU = [mean optical density (OD) value of sample – mean OD value of blank]/[mean OD value of calibrator – mean OD value of blank]*100. The cutoff point was a value of 30 EIU; equal or more than this value was defined as *H. pylori* seropositive.

4.3.1.2 CagA serostatus

The serum concentration of antibodies against CagA was measured by a commercial *Helicobacter pylori* p120 (CagA) IgG ELISA kit (Ravo Diagnostika GmbH, Freiburg, Germany). The EIU were calculated based on the formula as following: sample EIU = [mean optical density (OD) value of sample – mean OD value of blank]/[mean OD value of calibrator – mean OD value of blank]*unit value of calibrator. CagA seropositive was determined by a value of 7.5 EIU or more.

4.3.1.3 Pepsinogen I and pepsinogen II levels

Serum levels of pepsinogen I and II were measured by pepsinogen I and pepsinogen II ELISA kits (Biohit, Helsinki, Finland), respectively. Calibration curves were produced to test the serum concentrations of PGI or PGII. A serum concentration of PGI < 25 µg/l or PGI/ PGII < 3 was considered as presence of chronic corpus atrophic gastritis, with sensitivity and specificity of 71% (CI 68-74%) and 98% (CI 97-99%), respectively (152).

4.3.1.4 Quality control

The laboratory staff was blinded regarding case/control status. In each plate for the tested samples, control samples were added and designed to monitor the test quality, including a positive control sample, a blank sample and a negative control sample that were provided by the producer, and two internal control samples (generated by pooling the sera donated by 5 healthy volunteers). The titer values yielded from the control samples provided by the producer to control the within-assay imprecision were all within the expected ranges. Coefficients of variation, generated from values of the internal controls to control the between-assay imprecision, were 9.7% for *H. pylori*, 10.3% for CagA, 5.0% for PGI, and 7.9% for PGII, respectively.

4.3.2 Anchored Multiplex PCR (AMP)-based method (Study IV)

4.3.2.1 Custom target design

A panel of 65 clinically related cancer genes, including a 39-gene sub-panel of genes with recurrent mutations in solid tumors (from COSMIC database), and an exploratory sub-panel of 26 genes (from the Global Pancreatic Genomic study), was targeted for sequencing. The full coding regions were selected for the tumor suppressor genes, and selected exons for oncogenes.

4.3.2.2 DNA extraction and construction of sequencing libraries

Extraction of genomic DNA was carried out using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA). DNA libraries were prepared according to the AMP-based method (153). In brief, 200 ng genomic DNA was sheared to a mean size of 500 bp on a Covaris S220 instrument (Covaris, Woburn, MA, USA). Sheared fragments were processed with end repair (End-Repair Mix, Enzymatics, MA, USA), adenylation (Klenow Exo-, Enzymatics; Taq Polymerase, Life Technologies), and ligation (T4 DNA Ligase, Enzymatics). Illumina MiSeq adapters have been incorporated with a molecular index for tracking purpose. The ligated products were further cleaned by solid phase reversible immobilization, and then two rounds of nested polymerase chain reaction (PCR) were run for targeted enrichment (Platinum Taq Polymerase, Life Technologies), producing sequencing-ready libraries. Quantitation of the libraries was performed by quantitative PCR (Kapa Biosystems, Woburn, MA, USA) and sequenced on a MiSeq (Illumina, San Diego, CA). Thirty-two samples were pooled and sequenced in each MiSeq run.

4.3.2.3 Validation of AMP-based method

<u>A quantitative multiplex DNA reference standard</u>

We sequenced a commercial quantitative multiplex DNA reference standard (Horizon Diagnostics, DNA HD701) to assess the analytical performance of the AMP-based method. The DNA reference standard contains 6 genes and 11 mutations (*BRAF*, *KIT*, *EGFR*, *NRAS*, *KRAS*, and *PIK3CA*) (52). The 11 mutations were all covered by the above-mentioned 39-

gene sub-panel. During data analysis, *EGFR* G719S had the highest allele frequency, and served as an internal normalization reference, and was used to adjust the actual dilution ratios and to measure expected allele frequencies for all mutations in the same spike-in sample.

Sanger sequencing

Sanger sequencing was performed to evaluate mutations in *KRAS* (codons 12 and 13) that were identified by the AMP-based assay. An ABI 3730XL (PE Applied Biosystems, Foster City, CA, USA) was used for sequencing the purified PCR products, and data were analyzed by Chromas Version 2.4.

Allele-specific PCR

Allele-specific PCR was applied to validate mutations in *KRAS* codon 12 with allelic fractions too low (<5%) to be confirmed by the Sanger platform. A few samples with higher mutant allelic fractions were randomly selected and served as positive controls for allele-specific PCR. The 3' terminal base of the allele-specific primer (G12D, G12R or G12V) was designed to be identical to its corresponding mutation. Exonuclease I/SAP (New England BioLabs) was applied to the PCR products to remove unincorporated primers. Agarose gel electrophoresis (1%) was performed to identify the target amplicons.

Digital PCR

Digital PCR was performed to validate low-level *KRAS* mutations at codon 12 (<5%). *KRAS* G12D, G12V, and G12R mutations were measured separately using TaqMan SNP Genotyping Assays optimized for digital PCR on a QuantStudio 3D Digital PCR System (Life Technologies).

4.4 STATISTICAL ANALYSES

In **study I**, the R language for statistical computing (R Core Team, 2011) was used for all statistical analyses. In **study II to IV**, all statistical analyses were performed using the Statistical Analysis System (SAS) software package, version 9.3 (SAS Institute, Cary, NC). All statistical tests were two-sided and statistical significant level was set at the 5% level.

4.4.1 Standardized incidence ratio (Study I)

In **Study I**, the standardized incidence ratio (SIR) was measured as the ratio of the observed number of pancreatic cancer cases in the HCV or HBV cohort to the expected case number from the general population. In each cohort, to calculate the expected numbers, the observed person-time in each 5-year age group by sex and the corresponding Swedish population incidence rates were used. 95% confidence intervals (CIs) were calculated based on the Max method (154).

Stratification analyses were performed by sex, infection duration, and year of birth (before/after 1960).

4.4.2 Cox proportional hazards regression model (Study I, II)

In **Study I and II**, the associations between exposure variables and risk of pancreatic cancer were examined by hazard ratios (HRs) and the corresponding 95% CIs derived from Cox proportional hazards regression models. For **Study I**, due to matching by birth year, sex and county of residence, these factors were inherently adjusted in the model. Potential confounding factors were further adjusted in the Cox model, including hospitalizations for chronic obstructive pulmonary disease, diabetes mellitus, chronic pancreatitis and alcohol abuse. Because acute pancreatitis might be a complication to interferon therapy for HCV infection, and pancreatitis is a risk factor for pancreatic cancer. We further performed additional analysis that excluded all HCV infected patients who were ever hospitalized for acute and/or chronic pancreatitis.

In assessing poor oral hygiene as a risk factor for pancreatic cancer (**Study III**), the models were adjusted for age (as time scale), sex and attained calendar period in 10-year-intervals (1973-1982, 1983-1992, 1993-2002 and 2003-2012). For potential confounding factors, we further adjusted for tobacco use (non-tobacco user, pure smoker, pure snus user or mixed type user), alcohol consumption (low vs moderate or high), and area of residence (rural, small municipality or town). Stratification analysis was performed by sex. We also assessed the interaction between poor oral hygiene and sex by including a multiplicative interaction term into the model. In addition, to assess any influence of reverse causation, we performed a sensitivity analysis that excluded the first 2 years of follow-up.

4.4.3 Logistic regression model (Study III, IV)

In **Study III**, conditional logistic regression models were applied to calculate the odds ratios (ORs) and 95% CIs for the association between exposure variables and pancreatic cancer. The aforementioned matching factors were inherently adjusted in the models. In addition, potential confounding factors were further adjusted in the models, including smoking status (never, former or current), diabetes mellitus status (no or yes), height and waist-to-hip ratio. To estimate stratum-specific effects by blood group, a 4-category new variable was created by combining two dichotomous variables, e.g. *H. pylori*/CagA seropositivity (*H. pylori* negative and CagA negative vs *H. pylori* positive or CagA positive) and blood group (O vs non-O); dummy variables were then generated and introduced into regression models. To examine whether the association between *H. pylori* and/or CagA serostatus and pancreatic cancer was significantly modified by ABO blood group (O vs non-O blood type), an interaction term was introduced into the regression model and *p* value for the interaction term was derived from a Wald test. Similarly, the stratum-specific effects of chronic corpus atrophic gastritis by *H. pylori*/CagA serostatus were examined, and the interaction between these two variables was tested. To allay the concern that socioeconomic status (SES) may be a potential confounding factor for the association between *H. pylori* and pancreatic cancer, we further performed a sensitivity analysis by additionally including SES in the models. To minimize the influence of reverse causation bias, sensitivity analysis was performed by excluding the first 2 years of follow-up. To eliminate the non-fasting plasma effect, we

further conducted a sensitivity analysis by excluding case-control pairs with the fasting status of less than six hours since their last meal to blood collection.

In **Study IV**, logistic regression model was used to examine the association between tobacco smoking and overall genetic transversion or transition rate, adjusted for sex and age at entry.

4.4.4 Survival analysis (Study IV)

In **Study IV**, survival analysis was conducted among patients of PDAC with stage T3 tumors without distant metastasis. Follow-up of patients started from the date of surgery and was censored at death or January 1, 2014, whichever occurred first. The Kaplan-Meier method was used to estimate survival curves stratified by *KRAS* mutation subtype, mutation status (yes/no) in TP53 and transforming growth factor β (TGFB) signaling pathway genes (*SMAD4* or *TGFBR2*). Log-rank test was applied to compare survival curves.

5 RESULTS

5.1 HCV OR HBV INFECTION AND PANCREATIC CANCER RISK (STUDY I)

Baseline characteristics of the HCV and HBV cohorts are shown in Table 1. The mean follow-up time was 9.1 and 9.4 years in the HCV and chronic HBV cohorts, with a total of 360 154 and 107 986 person-years at risk, respectively. The HCV cohort had a greater proportion of males, and individuals in this cohort were more likely from Nordic countries. However, chronic HBV-infected individuals tended to be immigrants from non-Nordic countries.

Table 1 Characteristics of the chronic hepatitis C virus (HCV) infection and the hepatitis B virus (HBV) infection cohorts, 1990-2006

	HCV cohort ***N*****=39 442 (%)**	**Chronic HBV cohort** ***N*****=11 511 (%)**
Deceased	7872 (20.0)	738 (6.4)
Emigrated	1326 (3.4)	780 (6.8)
Person-years accumulated during follow-up	360 154	107 986
Mean follow-up time, years	9.1	9.4
Male sex %	27 024 (68.5)	6117 (53.1)
Median age at HBV/HCV notification (males, females)	38 (38, 37)	31 (33, 29)
Year of birth, median (range)	1958 (1898, 2006)	1967 (1906, 2006)
Country of origin		
Nordic countries	35 033 (88.8)	1779 (15.5)
Non-Nordic European countries	2035 (5.2)	3166 (27.5)
Other	2305 (5.8)	6482 (56.3)
Suspected route of virus transmission		
Intravenous drug use	21 941 (55.6)	178 (1.5)
Blood/blood products	2423 (6.1)	179 (1.6)
Sexual contact	891 (2.3)	334 (2.9)
Nosocomial transmission	130 (0.3)	38 (0.3)
Mother-child	92 (0.2)	762 (6.6)
Other	848 (2.1)	2744 (23.8)
Unknown or missing	13 117 (33.3)	7276 (63.2)

In the HCV cohort, we identified 34 pancreatic cancer cases and 16.5 were expected during 340 819 person-years of follow-up (first 6 months of follow-up were excluded), rendering a two-fold increased risk of pancreatic cancer. The SIRs did not vary notably across sex or estimated duration of HCV infection, but a stronger association was found among patients born after 1960 (Table 2). In the Cox regression model, HCV infected individuals had a 90% excess risk of pancreatic cancer after adjustment of age, sex and county of residence, which is consistent with the result from the SIR analysis. Additional adjustment for potential confounders diminished the strength of association somewhat, but the excess risk still remained statistically significant (Table 3). In the sensitivity analysis, after excluding

individuals ever hospitalized for pancreatitis (acute/chronic), the results did not change substantially (data not shown).

Table 2 Standardized incidence ratio (SIR) for pancreatic cancer among 39 442 individuals infected with hepatitis C virus (HCV) compared with the general population

	Pancreatic cancer		
	O[a]	**E**[b]	**SIR (95%CI)**
Lag period after HCV-notification			
≤ 6 months	4	0.8	5.3 (1.4, 13.5)
> 6 months	34	16.5	2.1 (1.4, 2.9)
Sex[c]			
Men	23	11.5	2.0 (1.3, 3.0)
Women	11	5.0	2.2 (1.1, 3.9)
HCV-duration[c, d]			
≤ 20 years	10	5.4	1.9 (0.9, 3.4)
> 20 years	24	11.1	2.2 (1.4, 3.2)
Year of birth[c]			
before 1960	31	15.5	2.0 (1.4, 2.8)
from 1960	3	1.0	3.0 (0.6, 8.8)

[a]Observed number of pancreatic cancer cases.

[b]Expected number of pancreatic cancer cases.

[c]Excluding cases and observation-time during the first 6 months after HCV notification.

[d]Estimated duration of HCV-infection.

In the HBV cohort, we found a 40% increased pancreatic cancer risk in the SIR analysis, but without statistical significance (data not shown). In the Cox regression model, the point estimate was somewhat higher (HR=2.0 after adjustment for only matching factors and HR=1.8 after full adjustment), but still did not reach the statistical significant level (Table 3).

Table 3 Hazard ratios of pancreatic cancer associated with hepatitis C virus (HCV) infection[a], and with chronic hepatitis B virus (HBV) infection[a]

	Pancreatic cancer	
	Cases	**HR (95% CI)**
Total	154	
Non-HCV infection	120	1.0
HCV-infection[b]	34	1.9 (1.3, 2.7)
HCV-infection[c]	34	1.6 (1.04, 2.4)
Total	21	
Non-HBV infection	16	1.00
HBV-infection[b]	5	2.0 (0.7, 5.3)
HBV-infection[c]	5	1.8 (0.7, 5.1)

[a]First 6 months of follow-up after HCV or HBV notification were excluded.

[b]Inherently adjusted for birth year, sex and county of residence in the model.

[c]Adjusted additionally for chronic obstructive pulmonary disease, diabetes mellitus, chronic pancreatitis and alcohol-related diseases.

5.2 POOR ORAL HYGIENE AND PANCREATIC CANCER RISK (STUDY II)

In this study, participants with fewer teeth at baseline appeared to be older, female and had a shorter follow-up duration. Similarly, cohort members with more dental plaque were older, male, and had a shorter follow-up duration. They were also more likely to be a smoker or snus users and consumed more alcohol, compared to individuals who had no dental plaque.

Table 4 demonstrates HRs for pancreatic cancer associated with different indicators of poor oral hygiene. Subjects with fewer teeth at baseline, compared to subjects with more teeth at baseline, appeared to have an increased pancreatic cancer risk, but the HR estimates were not statistically significant. On the other hand, when we used presence of dental plaque as an indicator for poor oral hygiene, we found that subjects with an unacceptable dental plaque status had a doubled risk of pancreatic cancer. As for oral mucosal lesions, the HRs were 1.7 and 1.3 for individuals with Candida- related oral mucosal lesions and denture-related oral mucosal lesions, respectively, which were of borderline statistical significance. Furthermore, individuals with tongue lesions had a statistically significant 80% increased risk for pancreatic cancer. Presence of more than one type of oral mucosal lesions further elevated the risk of pancreatic cancer. The highest HR of pancreatic cancer was observed among the individuals with all the three studied oral mucosal lesions (HR=3.1, 95% CI: 1.3, 7.6) (Table 4).

Additional adjustment for confounding factors had a negligible effect on the relative risk estimates (Table 4). Analysis stratified by sex indicated no heterogeneity for the associations (data not shown). We further restricted the analysis by excluding the first 2 years of follow-up, and the results did not alter significantly (data not shown).

Table 4 Hazard ratios (HRs) for pancreatic cancer according to number of teeth, dental plaque and oral mucosal lesions in a cohort study from 1973 to 2012.[a]

Characteristics	Total number (N)	Cases (N)	Person-years	IR (per 100,000 person-years)[b]	Crude HR (95% CI)[c]	Multivariable-adjusted HR (95% CI)[d]
Number of teeth at baseline[e]						
21-32	9490	30	331 884	14.7	1.0	1.0
11-20	5918	51	159 886	22.4	1.3 (0.8, 2.1)	1.2 (0.7, 2.0)
0-10	4516	45	80 360	30.1	1.4 (0.8, 2.5)	1.3 (0.7, 2.3)
P_{trend}					0.27	0.38
Dental plaque[f]						
No	3203	10	105 270	12.1	1.0	1.0
Acceptable	9581	52	313 205	21.2	1.8 (0.9, 3.5)	1.8 (0.9, 3.6)
Unacceptable	2624	19	73 295	27.5	2.1 (1.0, 4.6)	2.1 (1.0, 4.7)
P_{trend}					0.07	0.07
Oral mucosal lesions						
Reference	14 121	68	440 910	18.6	1.0	1.0
Candida-related oral mucosal lesions (A)[g]	1158	12	24 519	38.9	1.7 (0.9, 3.2)	1.7 (0.9, 3.3)
Denture-related oral mucosal lesions (B)[h]	4185	44	86 680	28.7	1.3 (0.9, 2.0)	1.3 (0.9, 2.0)
Tongue lesions (C)[i]	2117	24	51 645	34.5	1.8 (1.1, 2.9)	1.8 (1.1, 2.9)
Presence of both A and B	648	8	10 773	74.3*	1.8 (0.8, 3.9)	1.8 (0.8, 4.0)
Presence of both B and C	708	11	12 682	63.2	2.1 (1.1, 4.2)	2.2 (1.1, 4.3)
Presence of both A and C	603	9	12 825	70.2*	2.7 (1.3, 5.4)	2.6 (1.3, 5.4)
Presence of all 3 lesions	302	6	4656	128.9*	3.1 (1.3, 7.6)	3.2 (1.3, 7.8)

[a]HR=Hazard ratio; CI=confidence interval.

[b]IR=incidence rate per 100 000 person-years, standardized to age distribution of person-years experienced by all participants using 5-year age categories. * Crude incidence rate per 100 000 person-years was presented, as age-standardized incidence rate might be misleading when the number of cases was small.

[c]Models were adjusted for age (as time scale), sex and attained calendar period in 10-year-intervals (1973-1982, 1983-1992, 1993-2002 and 2003-2012).

[d]Models were adjusted for age (as time scale), sex and attained calendar period in 10-year-intervals (1973-1982, 1983-1992, 1993-2002 and 2003-2012), tobacco use (non-tobacco user, pure smoker, pure snus user or mixed type user), alcohol consumption (low vs high), and area of residence (rural, small municipality or town).

[e]The number of teeth calculation was based on presence or absence of 6 different teeth at baseline (tooth #16, #21, #24, #36, #41, #44); presence of 0-1, 2-5, 6 of the selected teeth examined was assumed to correlate to 0-10, 11-20 or 21-32 remaining teeth.

[f]Dental plaque calculation was based on six selected teeth on average, cut off criteria: No dental plaque (No dental plaque present); acceptable (dental plaque covering not more than one third of the tooth surface); unacceptable (dental plaque covering more than one third of the tooth surface). Dental plaque analyses were restricted to individuals with more than 10 teeth.

[g]*Candida*-related oral mucosal lesions include pseudomembranous candidiasis, chronic candidosis, angular cheilitis, atrophic and nodular leukoplakia, median type of atrophy of tongue papillae and unspecified glossitis.

[h]Denture-related oral mucosal lesions include denture stomatitis (localized, generalized and papillomatous), denture hyperplasia, traumatic ulcer and flabby ridges.

[i]Tongue lesions include lingua fissurata, and plicated tongue, atrophy of tongue papillae, hairy tongue, coated tongue, median rhomboid glossitis, and unspecific glossitis.

5.3 *HELICOBACTER PYLORI* INFECTION AND PANCREATIC CANCER RISK (STUDY III)

At baseline of recruitment, the mean age of the cases and control subjects were 57.8 years. Cases, compared to control subjects, were more likely to be current smokers and appeared to have a history of diabetes. In general, the prevalence of *H. pylori* seropositivity, CagA seropositivity and serologically-defined presence of chronic corpus atrophic gastritis did not differ significantly between cases and controls.

Our results illustrated that neither *H. pylori* seropositivity nor CagA seropositivity was associated with pancreatic cancer risk. Additional adjustment for confounding factors in the models did not alter the results appreciably (Table 5). Analysis stratified by smoking did not show significant heterogeneity of associations (*p* value for interaction = 0.11, fully-adjusted model) (data not shown).

In the combined analysis of *H. pylori* and CagA serostatus, the ORs were close to unity for subjects seropositive for *H. pylori* or CagA when compared with those seronegative for both *H. pylori* and CagA, and the null association remained unchanged across subgroups with different combinations of *H. pylori* and CagA serostatus (Table 5). In the sub-analysis among individuals with complete ABO blood type information (N=278 case-control pairs), a combined analysis of *H. pylori*/CagA serostatus and ABO blood type was performed. We did not find any excessive risk of pancreatic cancer in association with *H. pylori*/CagA seropositivity, in either O or non-O blood group (Table 5).

On the other hand, the point estimates of OR indicated a slightly increased pancreatic cancer risk among individuals with presence of chronic corpus atrophic gastritis, although it did not reach a statistically significant level (Table 6). In order to address the possible modification effect of *H. pylori* infection on the association between chronic corpus atrophic gastritis and pancreatic cancer risk, stratified analysis was performed by *H. pylori* serostatus and a stronger association was observed in the stratum seronegative for both *H. pylori* and CagA (OR=5.66, p value for interaction < 0.01, fully-adjusted model) (Table 6).

Sensitivity analyses by additionally adjusting for SES or excluding case-control pairs with less than 2 years of follow-up, revealed largely unchanged results. When we restricted the analysis to case-control pairs with fasting status of more than six hours since the last meal, the results remained essentially similar (data not shown).

Table 5 Odds ratios (ORs) and their corresponding 95% confidence intervals (CIs) for pancreatic cancer risk according to *H. pylori* serology: a case-control study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort

	Control subjects	**Case patients**	**Crude model[1]**	**Adjusted model[2]**
	***N*, %**	***N*, %**	**OR (95% CI)**	**OR (95% CI)**
***H. pylori* serology**				
Negative	241 (53.9)	250 (56.0)	1.00	1.00
Positive	206 (46.1)	196 (44.0)	0.91 (0.68, 1.21)	0.96 (0.70, 1.31)
CagA serology				
Negative	306 (68.9)	302 (68.3)	1.00	1.00
Positive	138 (31.1)	140 (31.7)	1.02 (0.76, 1.38)	1.07 (0.77, 1.48)
***H. pylori* and CagA serology**				
H. pylori-, CagA-	214 (48.3)	218 (49.5)	1.00	1.00
H. pylori+ or CagA+	231 (51.9)	224 (50.7)	0.94 (0.70, 1.25)	0.99 (0.73, 1.35)
H. pylori-, CagA+	25 (5.6)	28 (6.4)	1.11 (0.61, 2.02)	1.14 (0.59, 2.22)
H. pylori+, CagA-	91 (20.5)	82 (18.6)	0.88 (0.61, 1.27)	0.94 (0.63, 1.38)
H. pylori+, CagA+	113 (25.5)	112 (25.4)	0.96 (0.68, 1.37)	1.04 (0.63, 1.38)
Sub-analysis*:* stratified by ABO blood group (*N*=278, case-control pair)				
O blood group				
H. pylori-, CagA-	59 (21.2)	49 (17.6)	1.00	1.00
H. pylori+ or CagA+	52 (18.7)	45 (16.2)	1.06 (0.57, 1.96)	1.17 (0.60, 2.30)
Non-O blood type				
H. pylori-, CagA-,	76 (27.3)	86 (30.9)	1.00	1.00
H. pylori+ or CagA+	91 (32.7)	98 (35.3)	0.94 (0.60, 1.46)	0.86 (0.54, 1.38)

[1]Crude conditional logistic regression model was inherently adjusted for the matching factors, including study center, sex, age at blood collection (±3 years), date of blood donation (±3 months), time of blood donation (±2 h), and fasting status (<3h, 3-6h or >6 after the last meal).

[2]Adjusted conditional logistic regression model was inherently controlled for the matching factors, and was further adjusted for height, waist-to-hip ratio, smoking status (never, former or current) and diabetes mellitus status (no or yes).

Table 6 Association between serologically-determined presence of chronic corpus atrophic gastritis and risk of pancreatic cancer: a case-control study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort

	Control subjects	**Cases**	**Crude model[1]**	**Adjusted model[2]**
	N, %	*N*, %	OR (95% CI)	OR (95% CI)
Chronic corpus atrophic gastritis[3]				
No	415 (94.1)	407 (92.1)	1.00	1.00
Yes	26 (5.9)	35 (7.9)	1.39 (0.81, 2.38)	1.35 (0.77, 2.37)
Stratifying by *H. pylori* and CagA serostatus				
***H. pylori*+ or CagA+**				
Chronic corpus atrophic gastritis (no)	204 (46.5)	202 (46.1)	1.00	1.00
Chronic corpus atrophic gastritis (yes)	22 (5.0)	18 (4.1)	0.88 (0.45, 1.72)	0.85 (0.42, 1.72)
***H. pylori*- and CagA-**				
Chronic corpus atrophic gastritis (no)	210 (47.8)	202 (46.1)	1.00	1.00
Chronic corpus atrophic gastritis (yes)	3 (0.7)	16 (3.7)	5.29 (1.53, 18.23)	5.66 (1.59, 20.19)

[1]Crude conditional logistic regression model was inherently adjusted for the matching factors, including study center, sex, age at blood collection (±3 years), date of blood donation (±3 months), time of blood donation (±2 h), and fasting status (<3h, 3-6h or >6 after the last meal).

[2]Adjusted conditional logistic regression model was inherently controlled for the matching factors, and was further adjusted for height, waist-to-hip ratio, smoking status (never, former or current) and diabetes mellitus status (no or yes).

[3]Presence of chronic corpus atrophic gastritis was defined as pepsinogen I <25 μl/l or pepsinogen I/II <3.

5.4 VARIANT PROFILING OF PANCREATIC CANCER (STUDY IV)

The average age of the patients at surgery was 68.2 years, and 53.4% were men.T3 was the predominant tumor stage (80.8% of cases). Among the 57 patients with available smoking information, the majority were ever-smokers (75.4%).

5.4.1 Validation of the AMP-based assay by sequencing a quantitative multiplex DNA reference standard

To evaluate the analytical performance of the AMP-based assay, nine spike-in samples of the reference DNA (at three dilution ratios, performed in triplicate) were sequenced in one MiSeq run. Except for one sample due to poor performance, a minimum of 250X de-duplicated (molecular index consolidated) sequencing coverage was obtained for more than 90% of target regions for the rest.

Expected allele frequencies with ranges from 6.30% to 18.38% were successfully detected by the AMP-based assay (31 variants). For the 49 variants with allele frequencies ranging from 1% to 5%, 44 variants were detected and five variants were rejected (one variant: 2.63% allele frequency and 336X depth; four variants ≤ 158X depth). Nine out of 19 variants with the very low allele frequency (0.16% to 0.96%) were detected. Figure 1 reveals that to detect variants with allele frequency of 2.5%, sequencing depth greater than ~1000X was needed, and for a frequency of 1%, ~1500X depth was needed.

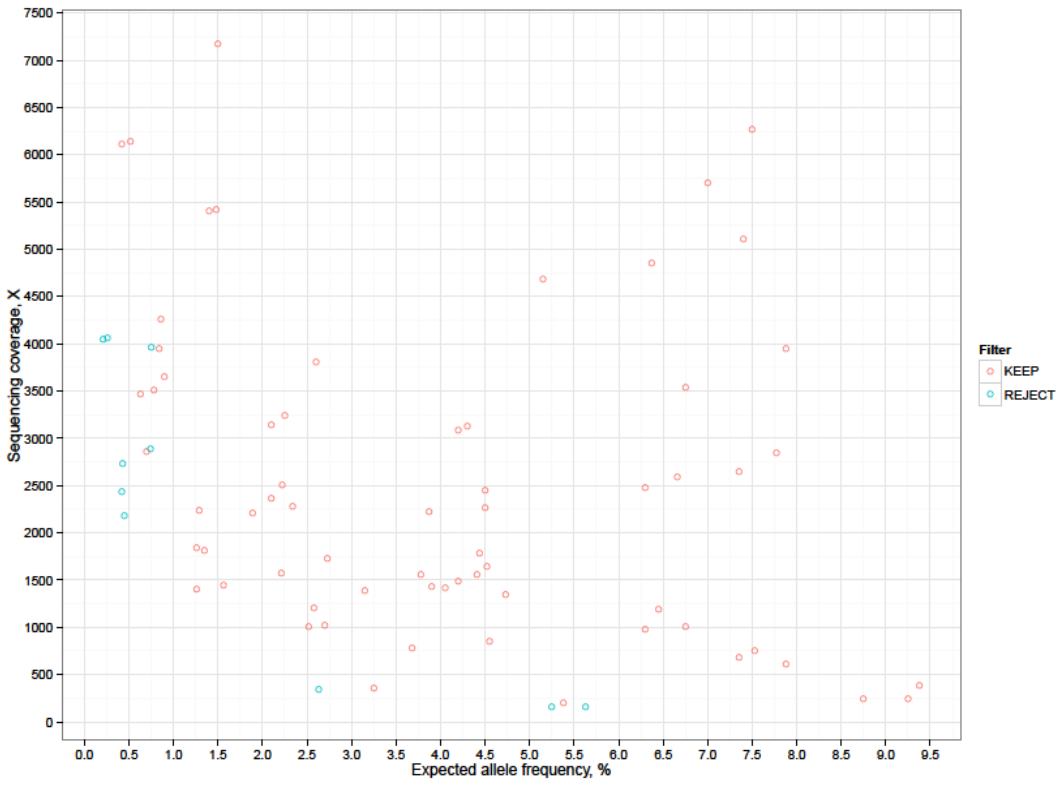

Figure 1 Detectable variant allelic frequencies according to sequencing depth, anchored multiplex PCR (AMP)-based assay.

5.4.2 Frequency and distribution of genetic aberrations

We identified a total of 95 non-synonymous variants, seven synonymous variants, and five insertions or deletions by using the AMP-based method in the 73 PDACs. Figure 2 depicts variants in 13 genes identified in at least one patient. Eighty-four percent of PDACs harbored at least one variant in the targeted regions.

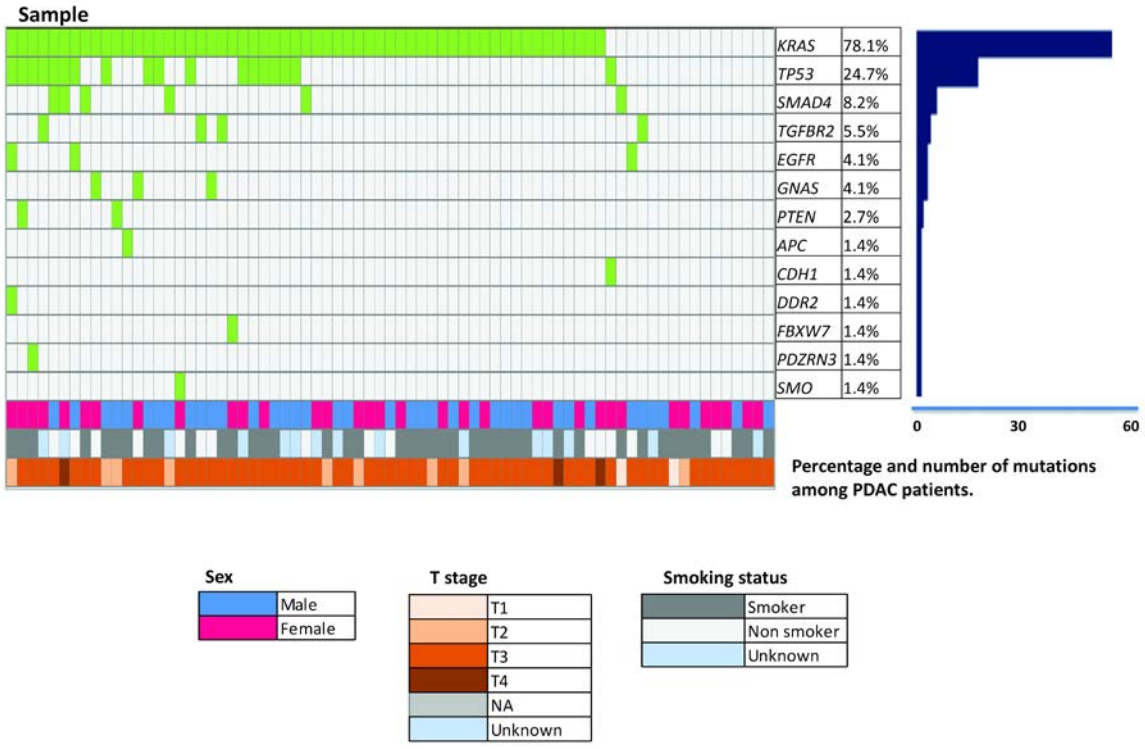

Figure 2 Concurrent and mutual exclusion of non-synonymous variants and frameshift indels observed across genes in pancreatic ductal adenocarcinoma. Tumors with and without alterations in the genes are labeled in green and grey, respectively, in the corresponding columns. Sex, T stage (according to TNM classification), and tobacco smoking status are also indicated.

The prevalence of *KRAS* was 78%, and 33% of the samples were of low-mutant-allelic-frequency (< 15%), and 8% of the samples were of very-low-mutant-allelic-frequency (<5%). The most prevalent *KRAS* mutations were G12D (GGT>GAT), G12R (GGT>CGT), and G12V (GGT>GTT). A total of 59 samples, with high mutant frequencies (> 5%) or without detectable mutations in *KRAS* on codons 12 and 13, were completely validated and confirmed by Sanger sequencing. The remaining low-mutant-allelic-frequency samples (< 5%, N=14, two samples harboring *KRAS* Q61H mutation are not shown in Table 7) detected by the AMP-based method were out of the detection limit for conventional Sanger sequencing. Mutations in these low-mutant-allelic-frequency samples were further confirmed by allele-specific PCR and digital PCR. The validation study demonstrated that the results from AMP-based assay were completely consistent with those from the Sanger sequencing (high-allelic-frequency samples), and were also fully consistent with those from

allele-specific PCR and digital PCR assays (low-allelic-frequency samples) (Table 7, Figure 3-6).

Table 7 Low and high allele frequencies of *KRAS* codon 12/13 mutations detected by anchored multiplex PCR (AMP)-based next-generation sequencing, Sanger sequencing, allele specific PCR and digital PCR assays.

Sample ID	AMP-based NGS assay	Mutant allele percentage by AMP-based NGS assay	Sanger sequencing assay	Allele specific PCR assay	Digital PCR assay	
					Relative quantification (95% confidence interval), %	Absolute quantification (95% confidence interval), copies/µL
Sample 30	p.G12D	1.14	Negative	p.G12D	2.13 (1.68 - 2.67)	5.12 (4.36 - 6.02)
Sample 65	p.G12V	1.31	Negative	p.G12V	1.17 (1.18 - 1.98)	3.98 (3.33 -- 4.75)
Sample 39	p.G12D	1.38	Negative	p.G12D	0.81 (0.59 - 1.11)	2.78 (2.23 - 3.46)
Sample 64	p.G12D	1.80	Negative	p.G12D	1.86 (1.43 - 2.42)	3.88 (3.23 - 4.66)
Sample 14	p.G12V	2.16	Negative	p.G12V	3.42 (2.68 – 4-37)	6.68 (5.70 -- 7.84)
Sample 12	p.G12V	2.52	Negative	p.G12V	2.06 (1.61 - 2.63)	4.36 (3.44 - 5.54)
Sample 26	p.G12R	2.52	Negative	p.G12R	2.49 (1.97 - 3.15)	4.65 (3.96 - 5.46)
Sample 74	p.G12V	2.62	Negative	p.G12V	5.64 (4.95 - 6.43)	16.46 (15.10 -- 17.94)
Sample 23	p.G12D	2.83	p.G12D	p.G12D	3.88 (3.22 - 4.67)	10.29 (9.08 – 11.66)
Sample 86	p.G12V	3.04	Negative	p.G12V	2.97 (2.40 – 3.67)	5.97 (5.16 - 6.90)
Sample 34	p.G12D	3.10	Negative	p.G12D	1.79 (1.43 - 2.22)	5.80 (4.99 – 6.75)
Sample 55	p.G12D	3.36	Negative	p.G12D	2.14 (1.75 - 2.61)	7.15 (6.24 -- 8.20)
Sample 43	p.G12V	7.76	Negative	p.G12V	6.37 (5.20 – 7.83)	13.14 (11.51 - 15.01)
Sample 44	p.G12R	30.34	p.G12R	p.G12R	32.28 (29.99 - 34.71)	72.60 (69.64 - 75.68)
Sample 45	p.G12V	41.45	p.G12V	p.G12V	46.80 (43.97 - 49.75)	161.73 (156.74 -- 166.87)
Sample 47	p.G12D	63.75	p.G12D	p.G12D	65.81 (62.88 - 68.80)	422.26 (414.32 - 430.35)

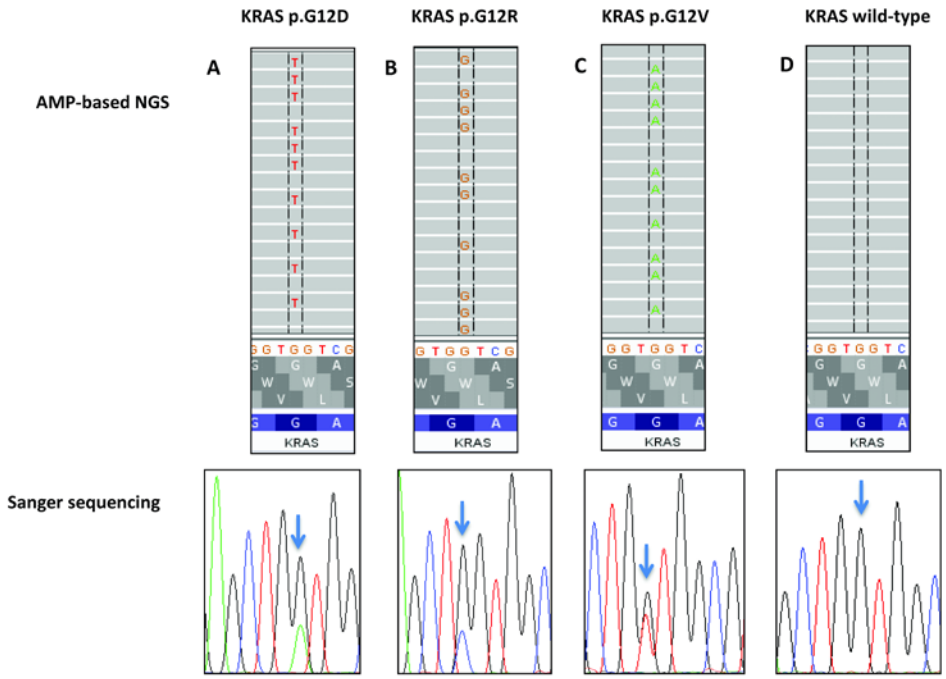

Figure 3 Identification of mutations by anchored multiplex PCR (AMP)-based next-generation sequencing, visualized using the Integrative Genomic Viewer (top panel), and confirmed by Sanger sequencing (bottom panel). (A) *KRAS* G12D mutation. (B) *KRAS* G12R mutation. (C) *KRAS* G12V mutation. (D) *KRAS* wild-type.

TP53 mutations ranked as the second most common alterations in PDAC, and were found in 25% of the samples. Of these, 6 and 7 of 18 were coexistent with *KRAS* G12D and G12R mutation, respectively. *SMAD4* mutations were identified in 8.2% of PDACs, and 5.5% of the mutations coexisted with *KRAS* G12R mutation. *GNAS* mutations were harbored in 4.1% of PDACs, and all *GNAS* mutations were concurrent with *KRAS* G12D mutation. The targeted genes with less frequent variants and the corresponding patients' clinicopathological features are shown in Figure 2. The allele frequencies of the rare variants ranged from 0.21% to 1.71%, and their filter criteria were: 1) affected in both strands of the original DNA template, 2) the total number of mutant reads from both strands was at least four.

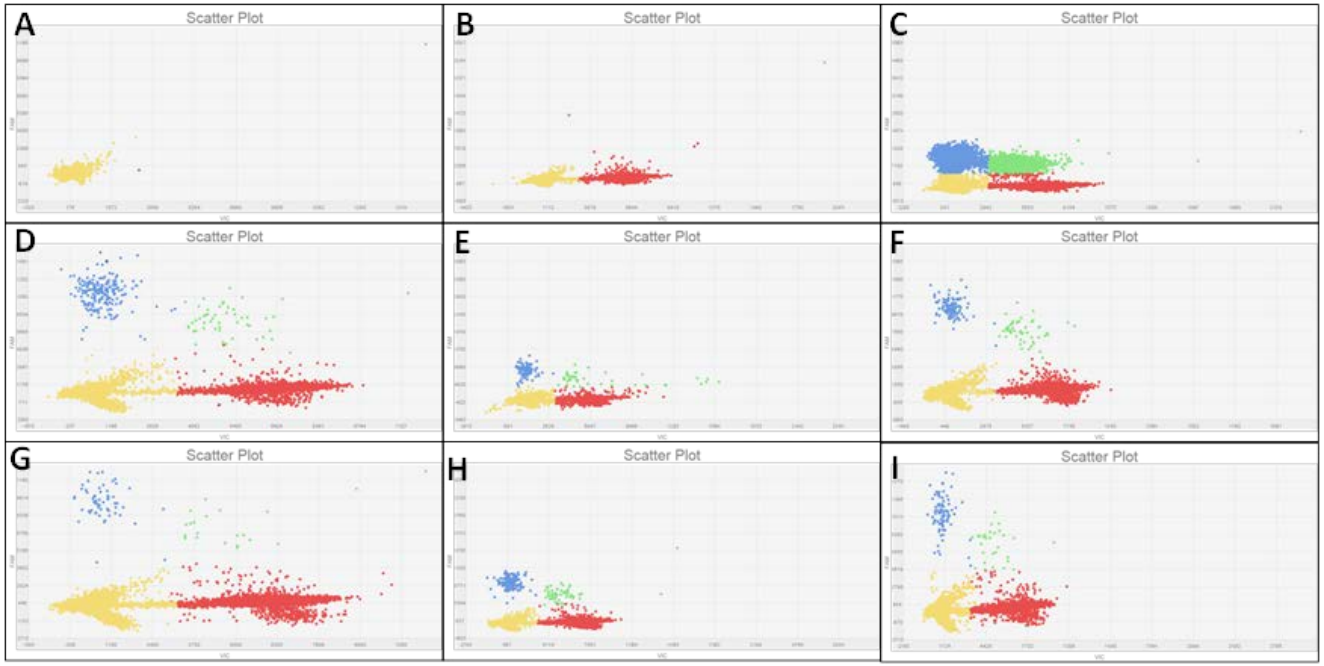

Figure 4 Validation of rare allele quantification of *KRAS* subtype mutation G12D by digital PCR assay.

The *KRAS* subtype mutation G12D was detected with a Custom Taqman SNP Genotyping Assay on the QuantStudio 3D Digital PCR system (Life Technologies). Each sample represents a mutant allelic fraction to wild-type genomic DNA background. All the samples were analyzed using QuantStudio 3D AnalysisSuite Cloud Software with the relative quantification module and absolute quantification module.

A) No DNA template control. Absolute *KRAS* G12D was 0.07 copy/µL.

B) Negative control with *KRAS* wild-type sample. Absolute *KRAS* G12D was 0.07 copy/µL.

C) Positive control with *KRAS* mutation G12D (sample 47). Absolute mutation target was 422.3 copies/µL. Mutation allele frequency was 65.8% (95% confidence interval: 62.9% - 68.8%).

Figure 4 D-I show allele frequencies of *KRAS* G12D for sample 23 (3.9%), sample 30 (2.1%), sample 34 (1.8%), sample 39 (0.81%), sample 55 (2.1%) and sample 64 (1.9%).

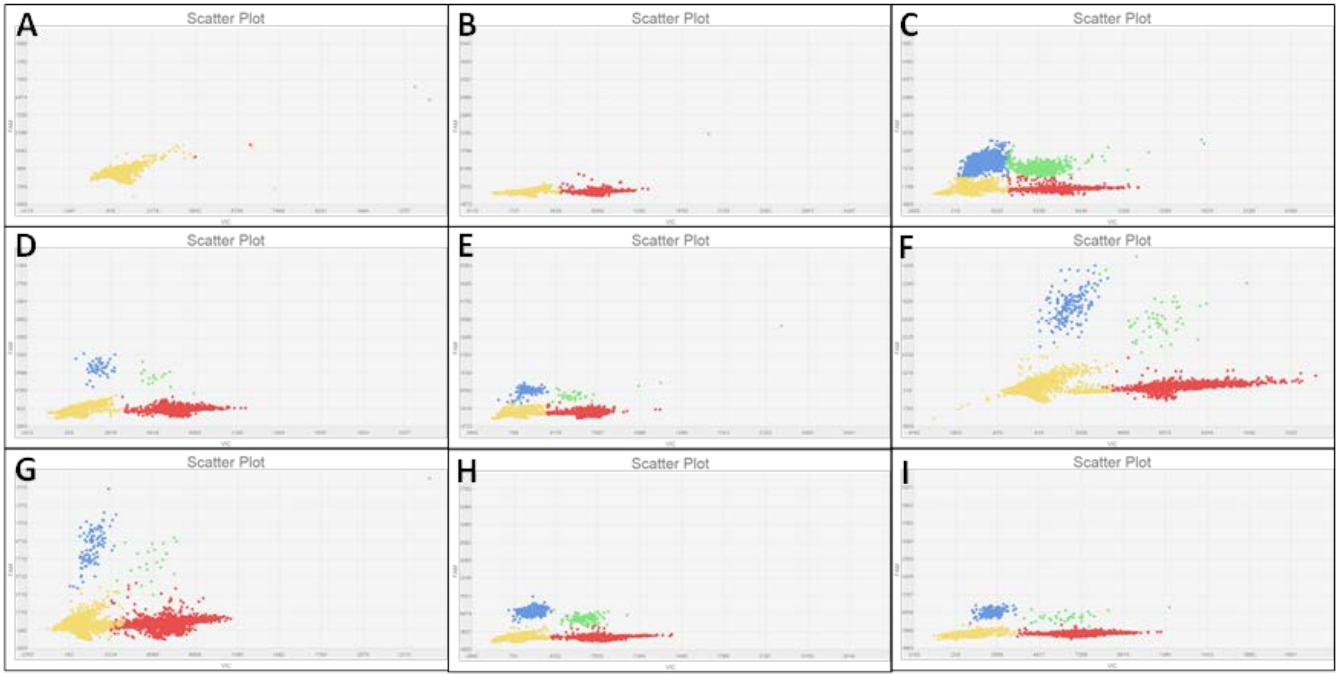

Figure 5 Validation of rare allele quantification of *KRAS* subtype mutation G12V by digital PCR assay.

The *KRAS* subtype mutation G12V was detected with a Custom Taqman SNP Genotyping Assay on the QuantStudio 3D Digital PCR system (Life Technologies). Each sample represents a mutant allelic fraction to wild-type genomic DNA background. All the samples were analyzed using QuantStudio 3D AnalysisSuite Cloud Software with the relative quantification module and absolute quantification module.

A) No DNA template control. Absolute *KRAS* G12V was 0.07 copy/µL.

B) Negative control with *KRAS* wild-type sample. Absolute *KRAS* G12V was 0.06 copy/µL.

C) Positive control with *KRAS* mutation G12V (sample 45). Absolute mutation target was 161.7 copies/µL. Mutation allele frequency was 46.8% (95% confidence interval: 44.0% - 49.8%).

Figure 5 D-I show allele frequencies of *KRAS* G12V: sample 12 (2.1%), sample 14 (3.4%), sample 43 (6.4%), sample 65 (1.2%), sample 74 (5.6%) and sample 86 (3.0%).

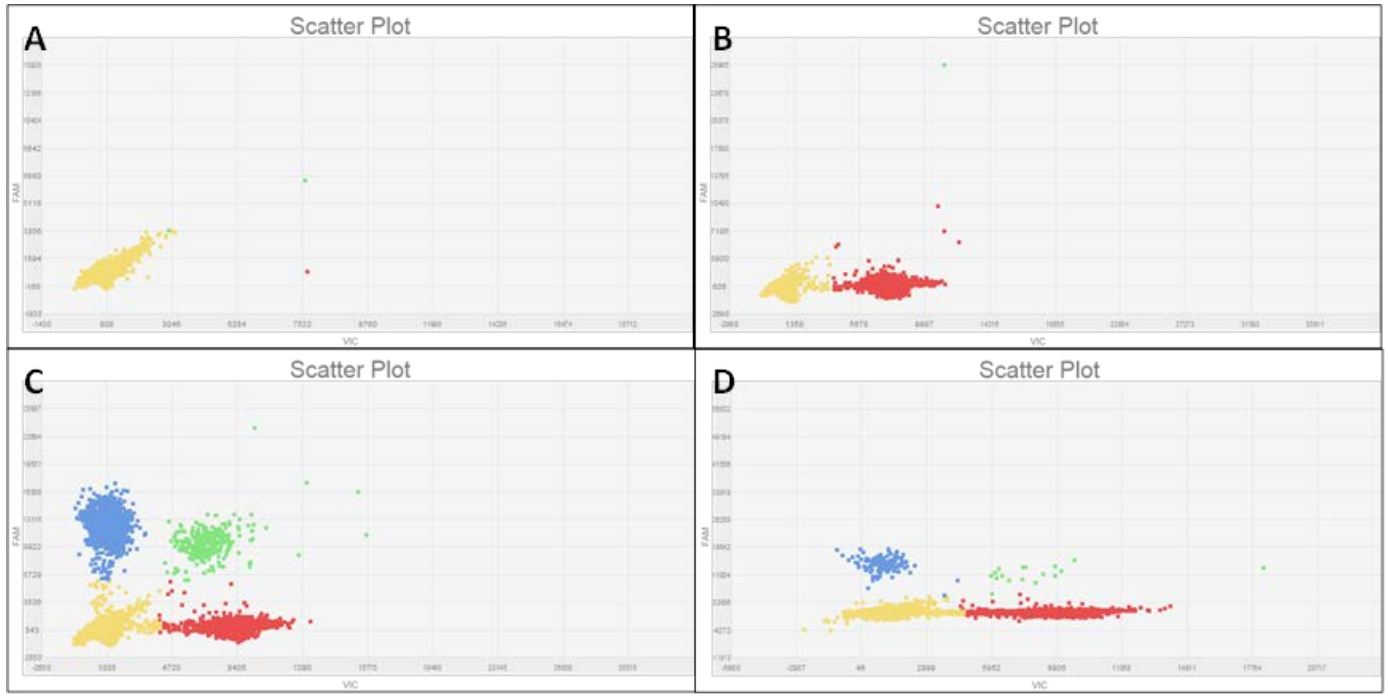

Figure 6 Validation of rare allele quantification of *KRAS* subtype mutation G12R by digital PCR assay.

The *KRAS* subtype mutation G12R was detected with a Custom Taqman SNP Genotyping Assay on the QuantStudio 3D Digital PCR system (Life Technologies). Each sample represents a mutant allelic fraction to wild-type genomic DNA background. All the samples were analyzed using QuantStudio 3D AnalysisSuite Cloud Software with the relative quantification module and absolute quantification module.

A) No DNA template control. Absolute *KRAS* G12V was 0.07copy/μL.

B) Negative control with *KRAS* wild-type sample. Absolute *KRAS* G12R was 0.06 copy/μL.

C) Positive control with *KRAS* mutation G12D (sample 44). Absolute mutation target was 72.6 copies/μL. Mutation allele frequency was 32.3% (95% confidence interval: 30.0% - 34.7%).

D) Allele frequency of *KRAS* G12R: sample 26 (2.5%).

5.4.3 Survival analysis according to *KRAS* mutation status

In the study, the median survival time of patients with wild-type *KRAS* was 26.2 months, longer than that of the patients with any *KRAS* G12 mutation (15.7 month, p =0.067, Figure 7A). The median survival time of patients with *KRAS* mutant subtype were 11.9 months for G12V, 16.5 months for G12R and 19.6 months for G12D, respectively (p =0.0197, Figure 7B). The median survival time among patients with or without mutations in *SMAD4* or *TGFBR2* was 14.1 and 20.8 months, respectively. However, the difference was not statistically significant (p =0.138, Figure 7C). Moreover, we did not find a significant difference of survival time between patients with or without *TP53* mutations (p =0.391, Figure 7D).

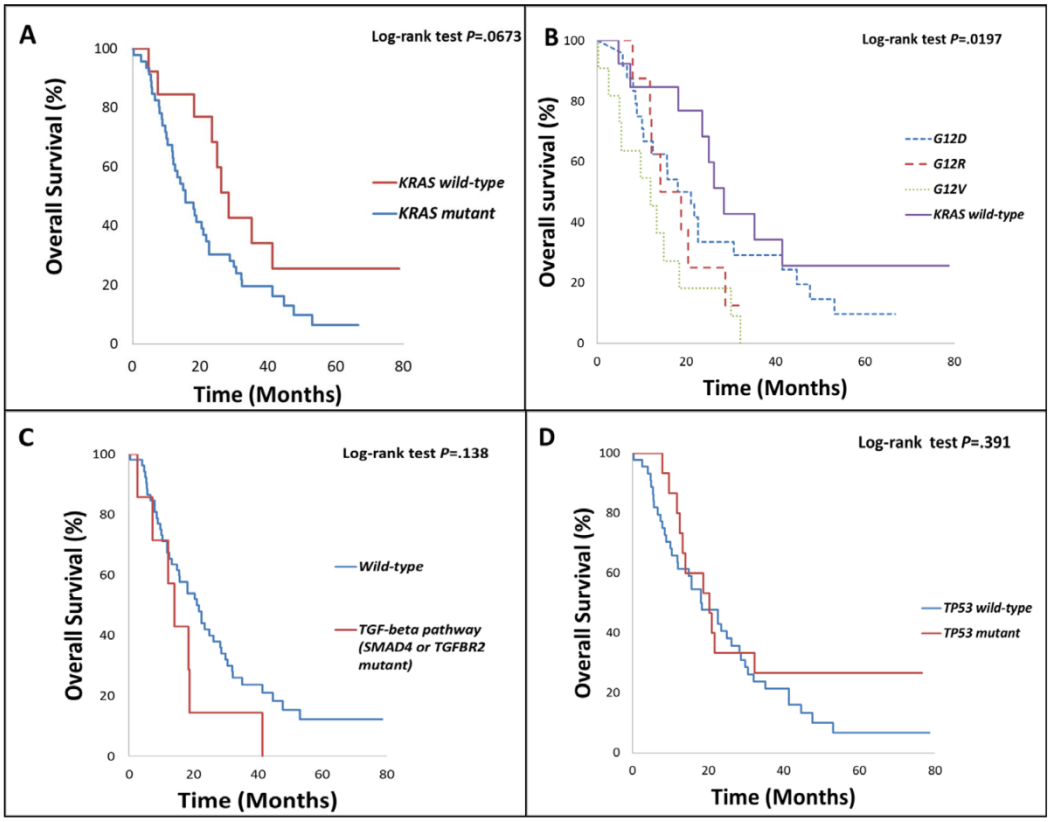

Figure 7 Overall survival analyses by the Kaplan-Meier method for pancreatic ductal adenocarcinoma patients restricted to T3 stage. (A) Overall survival by *KRAS* mutation status. (B) Overall survival by *KRAS* mutation subtype. (C) Overall survival by mutation status in TGF-β signaling pathway genes (*SMAD4* or *TGFBR2*). (D) Overall survival by *TP53* mutation status.

5.4.4 Association between tobacco smoking and genetic alterations

Tobacco smoking was associated with a greater frequency of transversion variants (OR=5.7), rather than transition variants (OR=0.7) (Table 8).

Table 8 Odds ratios (ORs) and corresponding 95% confidence intervals (CIs) for G:C>T:A transversions, G:C>C:G transversions, total transversions, and total transitions in pancreatic ductal adenocarcinoma (PDAC) in association with tobacco smoking [a]

	G:C>T:A			G:C>C:G			Transversions			Transitions[b]		
	No	**Yes**	**OR (95% CI)**	**No**	**Yes**	**OR (95% CI)**	**No**	**Yes**	**OR (95% CI)**	**Low**	**High**	**OR (95% CI)**
Never-smoker	10	3	1.0	12	2	1.0	7	6	1.0	6	7	1.0
Ever-smoker	32	11	2.3 (0.4, 11.6)	32	11	2.5 (0.4–14.9)	15	28	5.7 (1.2–27.8)	20	23	0.7 (0.2–2.9)

[a]Derived from logistic regression model, adjusted for age and sex. One case was excluded due to a lack of tumor stage information; 16 cases were excluded due to a lack of smoking information.

[b]Transitions: at 50% cutoff, <3 total transitions, low frequency; ≥3 total transitions, high frequency.

6 DISCUSSION

6.1 EPIDEMIOLOGICAL CONSIDERATIONS

6.1.1 Study design

The definition of epidemiology involves the study of the distribution and determinants of disease frequency in human populations (155). Epidemiologic research is constituted of several types of study designs, including experimental studies and observational studies. Ethics and cost limit most of the experimental studies; therefore most of epidemiologic studies are observational. In principle, there are two main types of observational studies, including cohort and case-control studies.

6.1.1.1 Cohort study

Broadly, cohort study is defined as any designated group of individuals who are followed or traced over a period of time (156), and a cohort study can be used to estimate average risks and rates (157). A prominent advantage of cohort study, compared to case-control study, is less prone to selection or recall bias. But, cohort studies are usually more expensive and time consuming. However, the Swedish health care system provides a unique opportunity that facilitates the follow-up of cohort at reasonably low cost. The national registration number contains the date of birth and 4 additional digits, and it not only enables the linkage within the medical care system, but also allows the linkage with Total Population and other demographic registers. The follow-up of **Study I and II** was through the linkage to Swedish health and population registers, including the Cancer Register, Causes of Death Register, and Emigration Register. **Study II** is a prospective cohort study with clinical examination that could collect detailed exposure data and information on potential confounders from the baseline questionnaire. The precious detailed information and virtually complete follow-up enable us to study the association between poor oral hygiene (including number of teeth, dental plaque and oral mucosal lesions) and the risk of pancreatic cancer.

In **Study I**, the exposure information, namely the HCV or HBV infection data, were identified from a nationwide register that might have minimized the possibility of exposure misclassification. To estimate relative risks, standardized incidence ratios and hazard ratios were calculated. Of note, one limitation in **Study I**, a register-based cohort, is lack of detailed information of potential confounding factors, such as smoking. To address the potential concern of confounding by smoking, the relative risks of lung cancer were also estimated.

6.1.1.2 Nested case-control study

Study III is a case-control study nested within the EPIC cohort study, with the control group selected by sampling randomly from a well-defined source population. A nested case-control study combines advantages from both cohort and case-control studies. The purpose to conduct a nested case-control study is usually that additional information besides already available data is required. However, the cost is too high to obtain this information for the

entire cohort. For example in **Study III**, it is impossible to perform the serologic measurements for half a million people in the EPIC cohort.

6.1.1.3 Case-only study

Study IV is a case-only study that includes all operated cases in the Stockholm pancreatic cancer case-control study. A case-only design can be used for multiple purposes, for instance, to evaluate gene-gene and gene-environment interaction in disease etiology, to scan for the contribution of complex genotypes, and to assess heterogeneity of outcomes (157). In **Study IV**, by using this approach, we have identified the molecular alterations in the selected cancer-related genes in PDAC in Sweden. We then explored the associations between the identified genetic alterations (transition and transversion variants) and environmental exposures (e.g. tobacco smoking). We have also examined the prognostic heterogeneity according to different genetic alterations (e.g. *KRAS* mutations, *TP53* mutations).

6.1.2 Bias

6.1.2.1 Selection bias

Selection bias is defined as 'distortions that result from procedures used to select subjects and from factors that influence study participation' (157). Selection bias can occur in a cohort study, for example due to loss to follow-up. However, in **Study I and II**, our case identification was conducted through the linkage to the Cancer Register, which has an almost 98% completeness of coverage. Information for censoring of follow-up was further derived from Causes of Death Register and Emigration Register which have also a virtually complete coverage. Selection bias can also arise in a cohort study when identification of exposed or unexposed individuals is associated with the possibility of developing the outcome of interest. In order to minimize the influence of this selection bias, sensitivity analysis was performed by excluding the first 6 months of follow-up in **Study I**, and by excluding the first 2 years of follow-up in **Study II**. In a case-control study, selection bias can occur if the controls are not representative of the population that cases originate from. The **Study III** which used a nested case-control study design suffered little from this bias, given that controls were randomly selected from the source population by a density-sampling method.

6.1.2.2 Information bias

Information bias can be introduced by measurement errors, e.g. errors in the measurement of exposure or outcome (157). Errors in the measurement are usually called classification error or misclassification. Misclassification of subjects for either exposure or outcome can be differential or non-differential (156), based on whether or not the misclassification is related with outcome or exposure. Much of discussion in misclassification has a focus on binary variable. The effects from differential misclassification can be bi-directional (overestimate or underestimate a true effect), whereas the bias from non-differential misclassification tends to dilute a true effect. In our studies, it is highly unlikely that the misclassification of the

exposure of interest was associated with the ascertainment of pancreatic cancer (study outcome).

However, non-differential misclassification might occur in our studies. First of all, there could be non-differential misclassification of the outcome in **Study I to III**, due to the difficulties in the diagnosis of pancreatic cancer. In **Study I and II**, pancreatic cancer cases were identified from the Cancer Register, in which not all cases were histopathologically verified. There is also a concern of underreporting of pancreatic cancer due to the rapid progression of this malignancy and the Swedish Cancer Register does not include death-certificate-only cases. Thus, a sensitivity analysis was performed with additional inclusion of death-certificate-only pancreatic cancer cases from the Causes of Death Register, and the results did not alter substantially. In **Study III**, a combined approach was applied to ascertain pancreatic cancer cases.

The non-differential misclassification of exposure (dichotomous variable) may bias the association towards null. In **Study I**, the HCV/HBV infection information was obtained from a parallel notification system, and it is estimated that approximately 75-80% of HCV infections are diagnosed. However, unknown infection still may exist. In addition, a resolved infection could occur among a small portion of the infected patients, either spontaneously or after treatment. This would lead to an underestimation of the pancreatic cancer risk in the HCV and HBV cohort. In **Study II**, the oral health status indicators, including tooth number, dental plaque status and presence of oral mucosal lesions, might have changed during the follow-up period. However, compared to the self-reported information from other studies, the oral hygiene information obtained from clinical dental examination was more accurate and would have reduced the non-differential misclassification of the exposures. In **Study III**, the serostatus of *H. pylori* infection might have been misclassified among cases and controls, due to unfavorable environment associated *H. pylori* clearance after long-term exposure of *H. pylori* infection and chronic gastric inflammation (158, 159). Therefore, a true positive association between *H. pylori* infection and pancreatic cancer might be biased towards null. In addition, we used serum pepsinogens I and II to determine the presence of chronic corpus atrophic gastritis, rather than using the ‘gold standard’ of the diagnosis based on histopathological examination. This could also introduce misclassification. However, there was no report that pancreatic cancer might influence pepsinogen levels. Furthermore, we determined O blood type through ABO rs505922. Although it has high linkage disequilibrium with rs8176719, it still could not fully replace rs8176719. Thus, the genotyping measurement error may arise, which would lead to non-differential misclassification of O and non-O blood type in **Study III**.

6.1.3 Confounding

Confounding, a confusion of effects (157), is a central issue for observational studies (160). Since the true association between exposure and outcome would be distorted by confounding, the confounders should be controlled, for example, in the study design stage, the analysis stage, or in a combination of the two approaches (160). One of the most prominent

limitations in the register-based studies is the lack of detailed information on the potential confounders. In **Study I**, potential confounding factors, including chronic obstructive pulmonary disease (as a proxy of smoking), diabetes mellitus, chronic pancreatitis and alcohol-related disease (as a proxy of heavy alcohol consumption) were identified from the Inpatient Register, and were further adjusted in Cox regression models. In addition, we indirectly estimated the smoking prevalence by calculating the relative risks for lung cancer. Smoking was more common in the HCV cohort than in the general population, given that we found a doubled risk of lung cancer in the HCV cohort. The relative risks for both lung cancer and pancreatic cancer were similar in either crude or fully-adjusted model, implying that smoking would not bias our finding entirely. If the finding was explained only by smoking, the HR of lung cancer in the HCV cohort should have been higher, given the fact that the relative risk of lung cancer was around 10 among smokers compared with non-smokers (161), whereas the corresponding figure was only 2 for pancreatic cancer (24). Moreover, according to Walker's finding (162), the effect of confounding is relatively small, even if both the relationships between exposure and covariate, as well as covariate and outcome, are strong. For instance, in order to completely bias an exposure effect of 1.4 by confounding, it requires the strengths of the association of confounder-exposure and confounder-outcome to be greater than 3. In **Study II**, we lacked data of BMI or waist-to-hip ratio. Besides, based on the limited information of alcohol consumption (low vs high), residual confounding might be present. However, considering their small effects on pancreatic cancer risk, these factors are unlikely to completely bias our findings.

6.2 STRENGTHS OF THE STUDIES

The shared strengths of **Study I to III** include a prospective cohort design (or based on a prospective cohort study), relatively large study population, and high completeness of follow-up. The special advantage for **Study I** is the available information from the nationwide documentation of HCV/HBV notifications. Of note for **Study II**, strengths include the professional clinical oral examination by one doctor for all participants, and the availability of information about potential confounders from baseline questionnaire. The important points of **Study III** include the access to prediagnostic blood samples and the availability of detailed information on potential confounding factors. For **Study IV**, we took advantage of the AMP-based NGS and found that this method is applicable for profiling tumor variants, even for clinical samples of low tumor cellularity.

6.3 INTERPRETATIONS OF FINDINGS

6.3.1 Pancreatic cancer risk assessment

6.3.1.1 HCV or HBV infection and pancreatic cancer risk

In **Study I**, we found a doubled risk of pancreatic cancer among HCV cohort compared with the Swedish general population. The excessive risk was consistent across strata by sex or duration of infection. Although attenuated association was observed after adjustment for

potential confounding factors, the finding still supports the hypothesis that HCV infection may be associated with an elevated pancreatic cancer risk. Moreover, we found a moderate excessive risk of pancreatic cancer among HBV-infected patients by different statistical approaches, however the size of the study cohort and the observed case numbers were too small to draw a sound conclusion.

It is still unclear about the biological mechanisms underlying the observed association between HCV and pancreatic cancer. Previous research illustrated that the pancreas is a remote location for hepatitis virus inhabitation and replication, as the pancreas and liver share common blood vessels and ducts (163). Immune response may cause chronic inflammation in the targeted organs after long-term persistent infection with HCV. Therefore, acting as a biological agent, HCV may indirectly play a role in inflammation-associated pancreatic carcinogenesis, through formation of an inflammatory microenvironment with a high concentration of growth factors and cytokines. As for the HBV infection, prior evidence suggested that HBV may replicate in the pancreas. A potential hypothesis proposes that a long period of chronic HBV infection, concurrent with damage of pancreatic epithelial cells, leads to an inflammatory response that consequently plays a role in the pancreatic cancer development.

To date, a few epidemiologic studies have examined the association between HCV/HBV and pancreatic cancer risk, whereas the results are still inconsistent. A US veteran cohort study showed a significant association between HCV and pancreatic cancer. However, the positive association was attenuated after adjusting for potential confounding factors (164). A case-control study demonstrated that the prevalence of HCV antibodies was not significantly higher among pancreatic cancer patients (1.5%) compared with controls (1%), however, a significantly higher prevalence of past exposure to HBV but not current or chronic infection, was observed among pancreatic cancer patients compared with control subjects (163). On the other hand, the REVEAL-HBV cohort study and a Chinese case-control study reported an elevated pancreatic cancer risk associated with chronic HBV infection (165, 166). In two meta-analysis studies, the authors found that chronic HCV and HBV infection increases pancreatic cancer risk, whereas past exposure to HBV may not be related to pancreatic cancer risk (167, 168).

6.3.1.2 Poor oral hygiene and pancreatic cancer risk

In **Study II**, our findings provide evidence to support the hypothesis that poor oral hygiene plays an important role in the development of pancreatic cancer. Although without statistical significance, the point estimates indicated a possible association between baseline tooth number and pancreatic cancer risk. Besides, presence of dental plaque is found to be associated with an increased pancreatic cancer risk. In addition, we found positive associations between pancreatic cancer and presence of oral mucosal lesions, including *Candida*- or denture-related oral mucosal lesions, and tongue lesions. To the best of our knowledge, this is the first prospective cohort study to explore the association between oral mucosal lesions and pancreatic cancer risk.

Increased production of carcinogenic nitrosamines and systemic inflammation are considered as potential plausible mechanisms. One poor oral hygiene indicator, i.e. a high amount of dental plaque, may be served as a reservoir for oral bacteria. In the nitrosamine-related hypothesis, individuals with poor oral hygiene may contain an abundance of nitrate-reducing bacteria, which will facilitate the formation of nitrosamines (169, 170). Nitrosamines have a carcinogenetic effect, and can induce and promote pancreatic cancer development (120, 121). On the other hand, an alternative hypothesis has been proposed. Poor oral hygiene may promote the formation of an environment in which certain oral pathogens, namely 'keystone pathogens', such as *P. gingivalis*, are able to evade host immune response and impair innate immunity. Moreover, the colonization of 'keystone pathogens' may also contribute to form a local environment that is favorable for bacterial overgrowth, which subsequently can mediate the microbial community and strengthen the conversion from a symbiotic state to a dysbiotic state (115, 116). This cascade of events may lead to a systemic inflammation state, which may play an important role in pancreatic cancer development. The inflammation state may promote the formation of a microenvironment that favors the accumulation of genetic alterations, initiates pancreatic carcinogenesis and fosters tumor development (110, 117).

6.3.1.3 *Helicobacter pylori* infection and pancreatic cancer risk

In **Study III**, a nested case-control study within a large European prospective cohort study, we observed no association between *H. pylori* infection (indicated by either *H. pylori* seropositivity or CagA seropositivity, or a combination of both) and pancreatic cancer risk. This null association remained unchanged in a combined analysis by including ABO blood type. On the other hand, based on a small sample size, our findings provided some support that chronic corpus atrophic gastritis, defined by serological pepsinogen levels, might be associated with pancreatic cancer risk, especially among the *H. pylori* seronegative individuals.

In the long period of *H. pylori* colonization in the stomach, the pathogenesis process usually goes through superficial gastritis, chronic atrophic gastritis, metaplasia and dysplasia (171). The development of multifocal atrophic gastritis causes hypo- or achlorhydria and basal hypergastrinemia due to the loss of parietal cells. The low acid environment in the stomach subsequently entails the bacterial overgrowth and promotes intragastric *N*-nitrosation catalyzation (172); through circulation in blood stream, the *N*-nitrosamines may be transported and activated in the pancreas, and may foster carcinogenesis in the pancreas. Other supporting evidence for this hypoacidity hypothesis comes from a register-based Swedish study, in which an increased pancreatic cancer risk was found in gastric ulcer patients, rather than the duodenal ulcer patients (173). Duodenal ulcer is associated with antral colonization of *H. pylori* and hyperchlorhydria, while gastric ulcer is related to corpus colonization of *H. pylori* with normo- or hypochlorhydria. In addition, long period hypo- or achlorhydria was characterized in pernicious anemia patients, and in these patients an excessive risk of pancreatic cancer has also been reported (174, 175).

A Finnish study on male smokers found that neither atrophic gastritis (defined by serologic pepsinogen level) nor histologically confirmed atrophic gastritis was associated with subsequent pancreatic cancer risk (176). However, in our study, although it was based on a small sample size, we found a positive association between chronic corpus atrophic gastritis and pancreatic cancer in the stratum seronegative for both *H. pylori* and CagA, but not in the stratum seropositive for *H. pylori* or CagA. One possible explanation is that chronic corpus atrophic gastritis might be more severe among the stratum seronegative for both *H. pylori* and CagA. Previous studies have shown that long-term advanced chronic corpus atrophic gastritis might lead to clearance of *H. pylori* colonization of the stomach mucosa, and subsequently lead to lower antibodies against the bacterium (158, 159). Nevertheless, given the small number in this stratum, we still cannot rule out the possibility that the observed positive association was due to chance. Future studies are merited to verify this association, due to the limitation of this study, and if confirmed, to further investigate the underlying mechanisms.

6.3.2 Variant profiling in PDAC

In **Study IV**, we found that the AMP-based assay can achieve analytically sensitive variant profiling in PDAC given sufficient sequencing depth. Our findings revealed a unique tumor variant profile, including dominance of *KRAS* mutations, majority in point mutations, and minority in insertions or deletions. A number of target enrichment assays, deep sequencing methods as well as variant calling algorithms have been designed and developed (177). However, it is still a challenge to identify true variants from background noises, which generally limits the analytical performance of these methods. In this study, we used the AMP-based method to enrich the targeted regions, and sequenced the targeted regions with a high depth of coverage. In a performance analytical test using a standard DNA reference, the findings elucidated that the AMP-based assay can detect tumor variants with allele frequencies as low as 1%, given sufficient sequencing depth. In addition, in the clinical PDAC samples, the detected *KRAS* mutations were completely confirmed by Sanger sequencing for high-mutant-allele-frequency samples, and also by allele-specific PCR and digital PCR for low-mutant-allele-frequency samples. The supportive evidence from these validation studies suggested that the AMP-based method is reliable for detecting tumor variants at a broad range.

In **Study IV**, we also found that patients with *KRAS* mutant G12V subtype had a worse prognosis, and the association remained largely unchanged after separate adjustment for smoking, chemotherapy treatment or *TP53* mutation status. Due to variations in study design, sample size, source population and tumor stages, the effects of *KRAS* mutations on overall survival of PDAC remain controversial (178-181). However, the aggressive biological behavior of G12V subtype was corroborated in earlier studies on other cancer types, including colorectal and lung cancers (182-184). Moreover, in vitro studies demonstrated that the GTPase activity of *G12V* is lower than G12D and wild-type *KRAS*, which can lock the mutant KRAS protein in the GTP-bound active state for a long period of time (185, 186). However, larger cohort studies are warranted to confirm these findings. Furthermore, our data

indicated that smoking is associated with transversion variants in PDAC. Previous studies indicated that *KRAS* and *TP53* mutations in pancreatic cancer patients are more common in smokers than in non-smokers (187, 188). However, some studies disproved these findings (189, 190). One study demonstrated that transition variants were 24% more common among smokers compared to non-smokers, while transversion variants were 53% more common in smokers (191). Larger studies are needed to confirm this finding and to further explore the associations between other environmental risk factors and the observed mutations in PDAC.

7 CONCLUSIONS

- HCV infection may be related to a higher pancreatic cancer risk, however, it requires larger cohort studies to verify this observed association.

- Poor oral hygiene, indicated by presence of dental plaque and oral mucosal lesions, is related to an elevated pancreatic cancer risk. This finding supports the hypothesis that poor oral hygiene has an important role in the development of pancreatic cancer.

- *H. pylori* infection is not related with pancreatic cancer risk in the western European populations. However, the suggested positive association between chronic corpus atrophic gastritis and pancreatic cancer warrants larger independent studies for verification, and, if confirmed, further studies are merited to explore the underlying mechanisms.

- The AMP-based next-generation sequencing is sensitive and accurate method for direct variant profiling of clinical pancreatic cancer samples. Using this method, we elucidated that in PDAC patients, *KRAS* mutant subtype G12V is associated with inferior survival, and that transversion variants are more common among smokers.

8 FUTURE PERSPECTIVES

Although progress has been made in recent decades in pancreatic cancer research, much remains unresolved for this malignancy, particularly as it is always diagnosed at an advanced and non-curative stage and with a dismal prognosis. In this thesis, we contributed somewhat to the further understanding of the etiology of pancreatic cancer with regards to potential infectious agents and poor oral hygiene. We have also developed a method for variant profiling of candidate genes in PDAC using AMP-based NGS.

For the potential association between infectious agents and pancreatic cancer risk, further studies with more complete information on potential confounding factors are warranted to solidify the associations. Further studies with larger sample sizes are also warranted to verify whether severe gastric atrophy contributes to pancreatic carcinogenesis. In addition, the potential pancreatic pathogenic bacteria and/or fungi from dental plaque and oral mucosal lesions need to be identified in future studies.

In **Study IV**, some targeted regions had poor depth of coverage, for instance high GC content in the *TERT* promoter and *FOXL2*. Therefore, optimization is warranted for a better coverage of the GC-rich regions; for example, future experiments should be designed to include PCR additives such as DMSO and betaine. Furthermore, small amounts of DNA materials (e.g. those obtained from EUS guided-fine needle aspiration), should be applied to the AMP-based assays to test a panel of key targeted genes simultaneously, and make the AMP-based NGS method an attractive assay to identify genetic alterations in the various pancreatic lesions, e.g. pancreatic intraepithelial neoplasia, IPMN.

Recently, detection of circulating tumor DNA (ctDNA), a ‘liquid biopsy’ technology, has emerged as a new generation of biomarker detection method. It can be used for the detection of genetic alterations that lead to initiation and progression of human cancers. It has attracted increasing attention as it is a non-invasive detection method. However, it is limited by the difficulty to identify the trace amount of cancer-specific nucleic acids among an overwhelming wild-type background. Future studies should attempt to detect mutations, such as *KRAS* mutations on codon 12, 13 and 61, in ctDNA samples of PDAC patients using AMP-based assay in a longitudinal study, and aim to expand the utilization of ctDNA for early detection, as well as for monitoring recurrence of the tumor after treatment.

9 ACKNOWLEDGEMENTS

I would like to acknowledge the scholarship from **Karolinska Institutet (KID)** that granted me and partially supported me to complete this thesis work. I also express my great gratitude to all the subjects that agree to participate in our studies, the study would not have even started without the participants.

Weimin Ye, my main supervisor, thank you for taking me as your student, and launching me into the epidemiologic journey, sharing your profound knowledge in epidemiology, biostatistics and medicine, and trusting me. Thank you for always being there whenever I need your help, your invaluable guidance, your dedication and commitment to science, your spirits in front of difficulties, and your deep insights into molecular epidemiology will impact on me for my entire research life.

Johannes-Matthias Löhr, my co-supervisor, thank you for sharing your professional and broad knowledge on pancreatic cancer during these years. I appreciate your unfailing supports, encouragement, guidance and trust. Your ever-positive attitude to science truly cultured me.

Magnus Nilsson, my co-supervisor, thank you for your generous help, encouragement and endless backing me up. I appreciate your confidence in me and your belief in my ability. Thank you for sharing your profound and broad knowledge in clinical pancreatic cancer.

A. John Iafrate, my heartfelt gratitude for all of your supports on the PDAC variant profiling paper. Thank you for taking me in your lab, and shared your profound knowledge in the cancer genomic field. Your every single input in that project is invaluable and enlightened me immensely.

Zongli Zheng, my sincere gratitude for your generous support and guidance. Your dedicated working attitude inspires and influences on me all the time.

Duberg Ann-Sofi, thank you for sharing your profound knowledge in the HCV/HBV infection diseases. Many thanks for your constant support, trusting in me, and be patient with me for my various questions. Your scientific attitude encourages me a lot!

Tony Axéll, thank you for your elegant work on the Swedish oral mucosal lesions study. I am impressed and appreciate your efforts and contributions, which make a better understanding of the impact of oral hygiene in public health.

Fang Fang, you are the rising star and a role model for all PhD students at MEB. I am so grateful for all of your generous help, invaluable advice and constant support!

Caroline Verbeke, I appreciate that you share your professional pathologic knowledge with me, and I could always get help from you whenever I ask. Thank you for your efficient and elegant work. I appreciated all of your supports and trust.

Cecilia Lundholm, you are my ultimate statistical teacher. Thank you for your great support!

My fantastic co-authors, particularly **Ann Roosaar, Ralf Segersvärd, Hans Matsson, Juha Kere, Unnur Valdimarsdóttir, Katja Fall, and EPIC writing group,** I appreciated your every single input in the papers, I am grateful for all of your encouragement and constant supports.

My dear peer doctoral fellows, **Huan Song, Dariush Nasrollahzadeh Nesheli, Zhiwei Liu, Tracy Peters, Isabella Guncha Ekheden, Nelson Ndegwa, Donal Barrett, Tingting Huang**, and close colleagues, **Ulrika Zagai, Anna Berglund, Sandra Strandberg, Amelie Plymoth, Yang Cao, Haiyun Wang, Xingdong Chen, Donghao Lv, Jianwei Zhu, Daniela Mariosa, Ruoqing Chen, Weiyao Yin, Alessandra Grotta, Lijie Ding and Katherine Wilson**, thank you for being there whenever I need your help, and thank you for supporting me all the time.

Lovely MEB members, you are making MEB a very nice working environment.

To the Huddinge group members, my fantastic colleagues, **Rainer Heuchel, Annika Wagman, Birgitta Holmgren, Ying Zhao, Xuan Li, Salvatore Nania, Jessica Norberg,** thank you for being so nice and welcoming me whenever I go there. Thank you for your always-supports!

To MGH group members, **Long P Le, Matthew Liebers, Jessica Cambry, Maristela Onozato,** for your so kindly welcoming me to the lab and taking care of me during my stay in Boston.

The TA-staff and IT/DATA-staff at MEB, I appreciated all of your help during these years.

Dan Grander, Agneta Mode, Klas Kärre and Maria Johansson. Thank you for organizing the fantastic Biomedicine Master Programme, which provided us lots of great opportunities and supported us in various aspects.

My dear Chinese friends, particularly **Chen Suo, He Gao, Bin Zhao, Heng Wang, Qinzi Yan&Shuo Liu, Qiuling Huang&Zhili Song, Xinming Wang&Celine Xu, Zheng Chang&Ci Song, Na Guan&Jianren Song, Yuning Zhang&Menghan Gao, Qin Xiao&Kai Du, Meiqiongzi Zhang&Shell, Yang Xuan, Jian Yan, Daohua Lou&Junwei Zhang, Tiansheng Shi&Qing Shen, Qi Chen, Tong Gong, Jie Song, Xu Chen, Yi Lu, Yiqing Zhan, Bojing Liu, Wei He, Haoming Yang, Xingrong Liu, Shuyang Yao, Jiangrong Wang&Xiaoyuan Ren, Sun Sun, Juan Du, Ning Wang, Qiang Zhang, Yixin Wang, Dong Yang, Fei Yang, Yuan Xue, Yunlong Yang&Ying Ye, Ning Xu, Xiaoda Wang, Jie Xu, Tianwei Gu, Qiang Lin,** thank you for being in my life during the past years, I will cherish all our colorful memories and great moments!

To my former supervisor **Xueqi Fu** in Jilin University, thank you for your scientific edification, constant support and your confidence in me.

To **my beloved parents**, thank you for your unconditional love and endless supports. You are always backing me up and making me become a stronger and better person. You are more than everything to me.

10 REFERENCES

1. Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: An update. Digestive diseases 2010;28:645-56.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: a cancer journal for clinicians 2015;65:87-108.

3. Maisonneuve P, Lowenfels AB. Risk factors for pancreatic cancer: A summary review of meta-analytical studies. International journal of epidemiology 2015;44:186-98.

4. Adami HO HD, Trichopoulos D. Textbook of cancer epidemiology – 2nd ed, chapter: Pancreatic cancer p333-348, ekbom a and trichopoulos d. Oxford University Press, New York 2008.

5. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the uk in 2010. British journal of cancer 2011;105 Suppl 2:S77-81.

6. Bosetti C, Bertuccio P, Negri E, La Vecchia C, Zeegers MP, Boffetta P. Pancreatic cancer: Overview of descriptive epidemiology. Molecular carcinogenesis 2012;51:3-13.

7. Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. Lancet 2004;363:1049-57.

8. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. CA: a cancer journal for clinicians 2008;58:71-96.

9. Hidalgo M. Pancreatic cancer. The New England journal of medicine 2010;362:1605-17.

10. Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: An overview. Nature reviews Gastroenterology & hepatology 2009;6:699-708.

11. Bond-Smith G, Banga N, Hammond TM, Imber CJ. Pancreatic adenocarcinoma. Bmj 2012;344:e2476.

12. Muller MF, Meyenberger C, Bertschinger P, Schaer R, Marincek B. Pancreatic tumors: Evaluation with endoscopic us, ct, and mr imaging. Radiology 1994;190:745-51.

13. Harsha HC, Kandasamy K, Ranganathan P, Rani S, Ramabadran S, Gollapudi S, et al. A compendium of potential biomarkers of pancreatic cancer. PLoS medicine 2009;6:e1000046.

14. Berger AC, Garcia M, Jr., Hoffman JP, Regine WF, Abrams RA, Safran H, et al. Postresection ca 19-9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: A prospective validation by rtog 9704. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2008;26:5918-22.

15. Hess V, Glimelius B, Grawe P, Dietrich D, Bodoky G, Ruhstaller T, et al. Ca 19-9 tumour-marker response to chemotherapy in patients with advanced pancreatic cancer enrolled in a randomised controlled trial. The Lancet Oncology 2008;9:132-8.

16. Ferrone CR, Finkelstein DM, Thayer SP, Muzikansky A, Fernandez-delCastillo C, Warshaw AL. Perioperative ca19-9 levels can predict stage and survival in patients

with resectable pancreatic adenocarcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2006;24:2897-902.

17. Ko AH, Hwang J, Venook AP, Abbruzzese JL, Bergsland EK, Tempero MA. Serum ca19-9 response as a surrogate for clinical outcome in patients receiving fixed-dose rate gemcitabine for advanced pancreatic cancer. British journal of cancer 2005;93:195-9.

18. Valls C, Andia E, Sanchez A, Fabregat J, Pozuelo O, Quintero JC, et al. Dual-phase helical ct of pancreatic adenocarcinoma: Assessment of resectability before surgery. AJR American journal of roentgenology 2002;178:821-6.

19. Bardeesy N, DePinho RA. Pancreatic cancer biology and genetics. Nature reviews Cancer 2002;2:897-909.

20. Wahi MM, Shah N, Schrock CE, Rosemurgy AS, 2nd, Goldin SB. Reproductive factors and risk of pancreatic cancer in women: A review of the literature. Annals of epidemiology 2009;19:103-11.

21. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2006;15:2435-40.

22. Iodice S, Gandini S, Maisonneuve P, Lowenfels AB. Tobacco and the risk of pancreatic cancer: A review and meta-analysis. Langenbeck's archives of surgery / Deutsche Gesellschaft fur Chirurgie 2008;393:535-45.

23. 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. American journal of epidemiology 2009;170:403-13.

24. 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). Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:1880-8.

25. 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. European journal of cancer 2014;50:193-203.

26. Raimondi S, Lowenfels AB, Morselli-Labate AM, Maisonneuve P, Pezzilli R. Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. Best practice & research Clinical gastroenterology 2010;24:349-58.

27. Duell EJ, Lucenteforte E, Olson SH, Bracci PM, Li D, Risch HA, et al. Pancreatitis and pancreatic cancer risk: A pooled analysis in the international pancreatic cancer case-control consortium (panc4). Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:2964-70.

28. Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. Jama 1995;273:1605-9.

29. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, Barzi F, Woodward M. Type-ii diabetes and pancreatic cancer: A meta-analysis of 36 studies. British journal of cancer 2005;92:2076-83.

30. Li D, Tang H, Hassan MM, Holly EA, Bracci PM, Silverman DT. Diabetes and risk of pancreatic cancer: A pooled analysis of three large case-control studies. Cancer causes & control : CCC 2011;22:189-97.

31. 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. European journal of cancer 2011;47:1928-37.

32. 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;24:13-25.

33. 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. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2014;25:2065-72.

34. Calle EE, Kaaks R. Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. Nature reviews Cancer 2004;4:579-91.

35. Polednak AP. Estimating the number of u.S. Incident cancers attributable to obesity and the impact on temporal trends in incidence rates for obesity-related cancers. Cancer detection and prevention 2008;32:190-9.

36. Inoue M, Sawada N, Matsuda T, Iwasaki M, Sasazuki S, Shimazu T, et al. Attributable causes of cancer in japan in 2005--systematic assessment to estimate current burden of cancer attributable to known preventable risk factors in japan. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:1362-9.

37. Dobbins M, Decorby K, Choi BC. The association between obesity and cancer risk: A meta-analysis of observational studies from 1985 to 2011. ISRN preventive medicine 2013;2013:680536.

38. Luo J, Margolis KL, Adami HO, LaCroix A, Ye W, Women's Health Initiative I. Obesity and risk of pancreatic cancer among postmenopausal women: The women's health initiative (united states). British journal of cancer 2008;99:527-31.

39. Fernandez E, La Vecchia C, Decarli A. Attributable risks for pancreatic cancer in northern italy. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 1996;5:23-7.

40. Permuth-Wey J, Egan KM. Family history is a significant risk factor for pancreatic cancer: Results from a systematic review and meta-analysis. Familial cancer 2009;8:109-17.

41. 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). International journal of cancer Journal international du cancer 2010;127:1421-8.

42. Frank C, Fallah M, Ji J, Sundquist J, Hemminki K. The population impact of familial cancer, a major cause of cancer. International journal of cancer Journal international du cancer 2014;134:1899-906.

43. 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 research 2010;70:1015-23.

44. Risch HA, Lu L, Wang J, Zhang W, Ni Q, Gao YT, Yu H. Abo blood group and risk of pancreatic cancer: A study in shanghai and meta-analysis. American journal of epidemiology 2013;177:1326-37.

45. Genkinger JM, Spiegelman D, Anderson KE, Bergkvist L, Bernstein L, van den Brandt PA, et al. Alcohol intake and pancreatic cancer risk: A pooled analysis of fourteen cohort studies. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2009;18:765-76.

46. Tramacere I, Scotti L, Jenab M, Bagnardi V, Bellocco R, Rota M, et al. Alcohol drinking and pancreatic cancer risk: A meta-analysis of the dose-risk relation. International journal of cancer Journal international du cancer 2010;126:1474-86.

47. Michaud DS, Vrieling A, Jiao L, Mendelsohn JB, Steplowski E, Lynch SM, et al. Alcohol intake and pancreatic cancer: A pooled analysis from the pancreatic cancer cohort consortium (panscan). Cancer causes & control : CCC 2010;21:1213-25.

48. Lucenteforte E, La Vecchia C, Silverman D, Petersen GM, Bracci PM, Ji BT, et al. Alcohol consumption and pancreatic cancer: A pooled analysis in the international pancreatic cancer case-control consortium (panc4). Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:374-82.

49. 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. International journal of cancer Journal international du cancer 2011;129:1708-17.

50. Aune D, Vieira AR, Chan DS, Navarro Rosenblatt DA, Vieira R, Greenwood DC, et al. Height and pancreatic cancer risk: A systematic review and meta-analysis of cohort studies. Cancer causes & control : CCC 2012;23:1213-22.

51. Emerging Risk Factors C. Adult height and the risk of cause-specific death and vascular morbidity in 1 million people: Individual participant meta-analysis. International journal of epidemiology 2012;41:1419-33.

52. Quantitative multiplex - DNA reference standards. Http://www.Horizondx.Com/quantitative-multiplex-dna-reference-standards.Html (accessed august 2015).

53. Aune D, Greenwood DC, Chan DS, Vieira R, Vieira AR, Navarro Rosenblatt DA, et al. Body mass index, abdominal fatness and pancreatic cancer risk: A systematic review and non-linear dose-response meta-analysis of prospective studies. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:843-52.

54. Gandini S, Lowenfels AB, Jaffee EM, Armstrong TD, Maisonneuve P. Allergies and the risk of pancreatic cancer: A meta-analysis with review of epidemiology and biological mechanisms. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2005;14:1908-16.

55. Olson SH, Hsu M, Satagopan JM, Maisonneuve P, Silverman DT, Lucenteforte E, et al. Allergies and risk of pancreatic cancer: A pooled analysis from the pancreatic cancer case-control consortium. American journal of epidemiology 2013;178:691-700.

56. Bao Y, Giovannucci EL, Kraft P, Stampfer MJ, Ogino S, Ma J, et al. A prospective study of plasma adiponectin and pancreatic cancer risk in five us cohorts. J Natl Cancer Inst 2013;105:95-103.

57. Bao Y, Michaud DS. Physical activity and pancreatic cancer risk: A systematic review. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2008;17:2671-82.

58. O'Rorke MA, Cantwell MM, Cardwell CR, Mulholland HG, Murray LJ. Can physical activity modulate pancreatic cancer risk? A systematic review and meta-analysis. International journal of cancer Journal international du cancer 2010;126:2957-68.

59. Larsson SC, Giovannucci E, Wolk A. Folate intake, mthfr polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: A meta-analysis. Gastroenterology 2006;131:1271-83.

60. Bao Y, Michaud DS, Spiegelman D, Albanes D, Anderson KE, Bernstein L, et al. Folate intake and risk of pancreatic cancer: Pooled analysis of prospective cohort studies. J Natl Cancer Inst 2011;103:1840-50.

61. Lin HL, An QZ, Wang QZ, Liu CX. Folate intake and pancreatic cancer risk: An overall and dose-response meta-analysis. Public health 2013;127:607-13.

62. Tio M, Andrici J, Cox MR, Eslick GD. Folate intake and the risk of upper gastrointestinal cancers: A systematic review and meta-analysis. Journal of gastroenterology and hepatology 2014;29:250-8.

63. Larsson SC, Giovannucci E, Bergkvist L, Wolk A. Aspirin and nonsteroidal anti-inflammatory drug use and risk of pancreatic cancer: A meta-analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2006;15:2561-4.

64. Capurso G, Schunemann HJ, Terrenato I, Moretti A, Koch M, Muti P, et al. Meta-analysis: The use of non-steroidal anti-inflammatory drugs and pancreatic cancer risk for different exposure categories. Alimentary pharmacology & therapeutics 2007;26:1089-99.

65. Bosetti C, Rosato V, Gallus S, Cuzick J, La Vecchia C. Aspirin and cancer risk: A quantitative review to 2011. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:1403-15.

66. Cui XJ, He Q, Zhang JM, Fan HJ, Wen ZF, Qin YR. High-dose aspirin consumption contributes to decreased risk for pancreatic cancer in a systematic review and meta-analysis. Pancreas 2014;43:135-40.

67. Paluszkiewicz P, Smolinska K, Debinska I, Turski WA. Main dietary compounds and pancreatic cancer risk. The quantitative analysis of case-control and cohort studies. Cancer epidemiology 2012;36:60-7.

68. Qin B, Xun P, He K. Fish or long-chain (n-3) pufa intake is not associated with pancreatic cancer risk in a meta-analysis and systematic review. The Journal of nutrition 2012;142:1067-73.

69. Gallus S, Turati F, Tavani A, Polesel J, Talamini R, Franceschi S, La Vecchia C. Soft drinks, sweetened beverages and risk of pancreatic cancer. Cancer causes & control : CCC 2011;22:33-9.

70. Genkinger JM, Li R, Spiegelman D, Anderson KE, Albanes D, Bergkvist L, et al. Coffee, tea, and sugar-sweetened carbonated soft drink intake and pancreatic cancer risk: A pooled analysis of 14 cohort studies. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2012;21:305-18.

71. Turati F, Galeone C, Edefonti V, Ferraroni M, Lagiou P, La Vecchia C, Tavani A. A meta-analysis of coffee consumption and pancreatic cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:311-8.

72. Chen K, Zhang Q, Peng M, Shen Y, Wan P, Xie G. Relationship between tea consumption and pancreatic cancer risk: A meta-analysis based on prospective cohort studies and case-control studies. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation 2014;23:353-60.

73. Mulholland HG, Murray LJ, Cardwell CR, Cantwell MM. Glycemic index, glycemic load, and risk of digestive tract neoplasms: A systematic review and meta-analysis. Am J Clin Nutr 2009;89:568-76.

74. Aune D, Chan DS, Vieira AR, Navarro Rosenblatt DA, Vieira R, Greenwood DC, et al. Dietary fructose, carbohydrates, glycemic indices and pancreatic cancer risk: A systematic review and meta-analysis of cohort studies. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23:2536-46.

75. Stolzenberg-Solomon RZ, Jacobs EJ, Arslan AA, Qi D, Patel AV, Helzlsouer KJ, et al. Circulating 25-hydroxyvitamin d and risk of pancreatic cancer: Cohort consortium vitamin d pooling project of rarer cancers. American journal of epidemiology 2010;172:81-93.

76. Wolpin BM, Ng K, Bao Y, Kraft P, Stampfer MJ, Michaud DS, et al. Plasma 25-hydroxyvitamin d and risk of pancreatic cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2012;21:82-91.

77. Liu SL, Zhao YP, Dai MH, You L, Wen Z, Xu JW. Vitamin d status and the risk of pancreatic cancer: A meta-analysis. Chinese medical journal 2013;126:3356-9.

78. Iodice S, Gandini S, Lohr M, Lowenfels AB, Maisonneuve P. Venous thromboembolic events and organ-specific occult cancers: A review and meta-analysis. Journal of thrombosis and haemostasis : JTH 2008;6:781-8.

79. Luo J, Ye W, Zendehdel K, Adami J, Adami HO, Boffetta P, Nyren O. Oral use of swedish moist snuff (snus) and risk for cancer of the mouth, lung, and pancreas in male construction workers: A retrospective cohort study. Lancet 2007;369:2015-20.

80. Gong Y, Zhou Q, Zhou Y, Lin Q, Zeng B, Chen R, Li Z. Gastrectomy and risk of pancreatic cancer: Systematic review and meta-analysis of observational studies. Cancer causes & control : CCC 2012;23:1279-88.

81. Lin G, Zeng Z, Wang X, Wu Z, Wang J, Wang C, et al. Cholecystectomy and risk of pancreatic cancer: A meta-analysis of observational studies. Cancer causes & control : CCC 2012;23:59-67.

82. Larsson SC, Wolk A. Red and processed meat consumption and risk of pancreatic cancer: Meta-analysis of prospective studies. British journal of cancer 2012;106:603-7.

83. Huang J, Valdimarsdottir U, Fall K, Ye W, Fang F. Pancreatic cancer risk after loss of a child: A register-based study in sweden during 1991-2009. American journal of epidemiology 2013;178:582-9.

84. Hepatitis c: Global prevalence. Wkly Epidemiol Rec 1997;72:341-4.

85. Custer B, Sullivan SD, Hazlet TK, Iloeje U, Veenstra DL, Kowdley KV. Global epidemiology of hepatitis b virus. J Clin Gastroenterol 2004;38:S158-68.

86. Dehesa-Violante M, Nunez-Nateras R. Epidemiology of hepatitis virus b and c. Archives of medical research 2007;38:606-11.

87. Lavanchy D. Hepatitis b virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. Journal of viral hepatitis 2004;11:97-107.

88. Duberg A, Janzon R, Back E, Ekdahl K, Blaxhult A. The epidemiology of hepatitis c virus infection in sweden. Euro Surveill 2008;13.

89. Shimoda T, Shikata T, Karasawa T, Tsukagoshi S, Yoshimura M, Sakurai I. Light microscopic localization of hepatitis b virus antigens in the human pancreas. Possibility of multiplication of hepatitis b virus in the human pancreas. Gastroenterology 1981;81:998-1005.

90. Yoshimura M, Sakurai I, Shimoda T, Abe K, Okano T, Shikata T. Detection of hbsag in the pancreas. Acta Pathol Jpn 1981;31:711-7.

91. Alvares-Da-Silva MR, Francisconi CF, Waechter FL. Acute hepatitis c complicated by pancreatitis: Another extrahepatic manifestation of hepatitis c virus? Journal of viral hepatitis 2000;7:84-6.

92. Torbenson M, Yeh MM, Abraham SC. Bile duct dysplasia in the setting of chronic hepatitis c and alcohol cirrhosis. Am J Surg Pathol 2007;31:1410-3.

93. Stolzenberg-Solomon RZ, Dodd KW, Blaser MJ, Virtamo J, Taylor PR, Albanes D. Tooth loss, pancreatic cancer, and helicobacter pylori. Am J Clin Nutr 2003;78:176-81.

94. Hujoel PP, Drangsholt M, Spiekerman C, Weiss NS. An exploration of the periodontitis-cancer association. Annals of epidemiology 2003;13:312-6.

95. Ahn J, Segers S, Hayes RB. Periodontal disease, porphyromonas gingivalis serum antibody levels and orodigestive cancer mortality. Carcinogenesis 2012;33:1055-8.

96. Farrell JJ, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, et al. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut 2012;61:582-8.

97. Michaud DS, Izard J, Wilhelm-Benartzi CS, You DH, Grote VA, Tjonneland A, et al. Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large european prospective cohort study. Gut 2013;62:1764-70.

98. Michaud DS, Joshipura K, Giovannucci E, Fuchs CS. A prospective study of periodontal disease and pancreatic cancer in us male health professionals. J Natl Cancer Inst 2007;99:171-5.

99. Hiraki A, Matsuo K, Suzuki T, Kawase T, Tajima K. Teeth loss and risk of cancer at 14 common sites in japanese. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2008;17:1222-7.

100. Schistosomes, liver flukes and helicobacter pylori. Iarc working group on the evaluation of carcinogenic risks to humans. Lyon, 7-14 june 1994. IARC monographs on the evaluation of carcinogenic risks to humans / World Health Organization, International Agency for Research on Cancer 1994;61:1-241.

101. Parsonnet J, Friedman GD, Orentreich N, Vogelman H. Risk for gastric cancer in people with caga positive or caga negative helicobacter pylori infection. Gut 1997;40:297-301.

102. Xiao M, Wang Y, Gao Y. Association between helicobacter pylori infection and pancreatic cancer development: A meta-analysis. PloS one 2013;8:e75559.

103. Wang Y, Zhang FC, Wang YJ. Helicobacter pylori and pancreatic cancer risk: A meta- analysis based on 2,049 cases and 2,861 controls. Asian Pacific journal of cancer prevention : APJCP 2014;15:4449-54.

104. Schulte A, Pandeya N, Fawcett J, Fritschi L, Risch HA, Webb PM, et al. Association between helicobacter pylori and pancreatic cancer risk: A meta-analysis. Cancer causes & control : CCC 2015.

105. de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, Kuipers EJ. Gastric cancer risk in patients with premalignant gastric lesions: A nationwide cohort study in the netherlands. Gastroenterology 2008;134:945-52.

106. El-Zimaity H. Gastritis and gastric atrophy. Current opinion in gastroenterology 2008;24:682-6.

107. de Vries AC, Kuipers EJ. Epidemiology of premalignant gastric lesions: Implications for the development of screening and surveillance strategies. Helicobacter 2007;12 Suppl 2:22-31.

108. Kuipers EJ. Proton pump inhibitors and gastric neoplasia. Gut 2006;55:1217-21.

109. Zambirinis CP, Pushalkar S, Saxena D, Miller G. Pancreatic cancer, inflammation, and microbiome. Cancer journal 2014;20:195-202.

110. Greer JB, Whitcomb DC. Inflammation and pancreatic cancer: An evidence-based review. Current opinion in pharmacology 2009;9:411-8.

111. Michaud DS. Role of bacterial infections in pancreatic cancer. Carcinogenesis 2013;34:2193-7.

112. 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. The New England journal of medicine 1993;328:1433-7.

113. Malka D, Hammel P, Maire F, Rufat P, Madeira I, Pessione F, et al. Risk of pancreatic adenocarcinoma in chronic pancreatitis. Gut 2002;51:849-52.

114. Talamini G, Falconi M, Bassi C, Sartori N, Salvia R, Caldiron E, et al. Incidence of cancer in the course of chronic pancreatitis. The American journal of gastroenterology 1999;94:1253-60.

115. Curtis MA. Periodontal microbiology--the lid's off the box again. Journal of dental research 2014;93:840-2.

116. Hajishengallis G, Liang S, Payne MA, Hashim A, Jotwani R, Eskan MA, et al. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. Cell host & microbe 2011;10:497-506.

117. Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420:860-7.

118. Swann JB, Smyth MJ. Immune surveillance of tumors. The Journal of clinical investigation 2007;117:1137-46.

119. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural innate and adaptive immunity to cancer. Annual review of immunology 2011;29:235-71.

120. Risch HA. Pancreatic cancer: Helicobacter pylori colonization, n-nitrosamine exposures, and abo blood group. Molecular carcinogenesis 2012;51:109-18.

121. 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;95:948-60.

122. Howatson AG, Carter DC. Pancreatic carcinogenesis: Effect of secretin in the hamster- nitrosamine model. J Natl Cancer Inst 1987;78:101-5.

123. Kokkinakis DM, Reddy MK, Norgle JR, Baskaran K. Metabolism and activation of pancreas specific nitrosamines by pancreatic ductal cells in culture. Carcinogenesis 1993;14:1705-9.

124. Anderson KE, Hammons GJ, Kadlubar FF, Potter JD, Kaderlik KR, Ilett KF, et al. Metabolic activation of aromatic amines by human pancreas. Carcinogenesis 1997;18:1085-92.

125. Hecht SS. DNA adduct formation from tobacco-specific n-nitrosamines. Mutation research 1999;424:127-42.

126. Kaaks R, Lukanova A. Energy balance and cancer: The role of insulin and insulin-like growth factor-i. The Proceedings of the Nutrition Society 2001;60:91-106.

127. Hennig R, Ding XZ, Adrian TE. On the role of the islets of langerhans in pancreatic cancer. Histology and histopathology 2004;19:999-1011.

128. Kazakoff K, Cardesa T, Liu J, Adrian TE, Bagchi D, Bagchi M, et al. Effects of voluntary physical exercise on high-fat diet-promoted pancreatic carcinogenesis in the hamster model. Nutrition and cancer 1996;26:265-79.

129. Stolzenberg-Solomon RZ, Graubard BI, Chari S, Limburg P, Taylor PR, Virtamo J, Albanes D. Insulin, glucose, insulin resistance, and pancreatic cancer in male smokers. Jama 2005;294:2872-8.

130. Pour PM, Kazakoff K, Carlson K. Inhibition of streptozotocin-induced islet cell tumors and n-nitrosobis(2-oxopropyl)amine-induced pancreatic exocrine tumors in syrian hamsters by exogenous insulin. Cancer research 1990;50:1634-9.

131. Pour PM, Stepan K. Modification of pancreatic carcinogenesis in the hamster model. Viii. Inhibitory effect of exogenous insulin. J Natl Cancer Inst 1984;72:1205-8.

132. Schneider MB, Matsuzaki H, Haorah J, Ulrich A, Standop J, Ding XZ, et al. Prevention of pancreatic cancer induction in hamsters by metformin. Gastroenterology 2001;120:1263-70.

133. Wolpin BM, Bao Y, Qian ZR, Wu C, Kraft P, Ogino S, et al. Hyperglycemia, insulin resistance, impaired pancreatic beta-cell function, and risk of pancreatic cancer. J Natl Cancer Inst 2013;105:1027-35.

134. Shi C, Hruban RH, Klein AP. Familial pancreatic cancer. Archives of pathology & laboratory medicine 2009;133:365-74.

135. Petersen GM. Familial pancreatic adenocarcinoma. Hematology/oncology clinics of North America 2015;29:641-53.

136. 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. Nature genetics 2009;41:986-90.

137. 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. Nature genetics 2010;42:224-8.

138. 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. Nature genetics 2014;46:994-1000.

139. 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. Nature genetics 2015;47:911-6.

140. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801-6.

141. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491:399-405.

142. Wood LD, Hruban RH. Genomic landscapes of pancreatic neoplasia. Journal of pathology and translational medicine 2015;49:13-22.

143. Schutte M, Hruban RH, Geradts J, Maynard R, Hilgers W, Rabindran SK, et al. Abrogation of the rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. Cancer research 1997;57:3126-30.

144. Wilentz RE, Su GH, Dai JL, Sparks AB, Argani P, Sohn TA, et al. Immunohistochemical labeling for dpc4 mirrors genetic status in pancreatic adenocarcinomas : A new marker of dpc4 inactivation. The American journal of pathology 2000;156:37-43.

145. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010;467:1114-7.

146. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the swedish cancer register: A sample survey for year 1998. Acta oncologica 2009;48:27-33.

147. Mattsson B, Wallgren A. Completeness of the swedish cancer register. Non-notified cancer cases recorded on death certificates in 1978. Acta radiologica Oncology 1984;23:305-13.

148. Duberg AS, Torner A, Davidsdottir L, Aleman S, Blaxhult A, Svensson A, et al. Cause of death in individuals with chronic hbv and/or hcv infection, a nationwide community-based register study. Journal of viral hepatitis 2008;15:538-50.

149. Hofmann JN, Torner A, Chow WH, Ye W, Purdue MP, Duberg AS. Risk of kidney cancer and chronic kidney disease in relation to hepatitis c virus infection: A nationwide register-based cohort study in sweden. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation 2011;20:326-30.

150. Axell T. A prevalence study of oral mucosal lesions in an adult swedish population. Odontol Revy Suppl 1976;36:1-103.

151. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European prospective investigation into cancer and nutrition (epic): Study populations and data collection. Public health nutrition 2002;5:1113-24.

152. Storskrubb T, Aro P, Ronkainen J, Sipponen P, Nyhlin H, Talley NJ, et al. Serum biomarkers provide an accurate method for diagnosis of atrophic gastritis in a general population: The kalixanda study. Scandinavian journal of gastroenterology 2008;43:1448-55.

153. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex pcr for targeted next-generation sequencing. Nature medicine 2014;20:1479-84.

154. Kulkarni PM, Tripathi RC, Michalek JE. Maximum (max) and mid-p confidence intervals and p values for the standardized mortality and incidence ratios. American journal of epidemiology 1998;147:83-6.

155. MacMahon B PT. Epidemiology: Principles and methods. Boston ma: Little, brown and co, 1970.

156. Rothman KJ. Epidemiology: An introduction. Oxford university press, inc., new york. 2002.

157. Rothman KJ, Greenland, S., & Lash, T.L. Modern epidemiology, 3rd edition. Philadelphia, pa: Lippincott, williams & wilkins. 2008.

158. Weck MN, Brenner H. Association of helicobacter pylori infection with chronic atrophic gastritis: Meta-analyses according to type of disease definition. International journal of cancer Journal international du cancer 2008;123:874-81.

159. Weck MN, Gao L, Brenner H. Helicobacter pylori infection and chronic atrophic gastritis: Associations according to severity of disease. Epidemiology 2009;20:569-74.

160. Aschengrau A SGR. Essentials of epidemiology in public health. London: Jones and bartlett publishers, inc, 2003.

161. Pesch B, Kendzia B, Gustavsson P, Jockel KH, Johnen G, Pohlabeln H, et al. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. International journal of cancer Journal international du cancer 2012;131:1210-9.

162. Walker AM. Observation and inference: An introduction to the methods of epidemiology. Newton, MA: Epidemiology Resources, Inc, Publishers 1991.

163. Hassan MM, Li D, El-Deeb AS, Wolff RA, Bondy ML, Davila M, Abbruzzese JL. Association between hepatitis b virus and pancreatic cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2008;26:4557-62.

164. El-Serag HB, Engels EA, Landgren O, Chiao E, Henderson L, Amaratunge HC, Giordano TP. Risk of hepatobiliary and pancreatic cancers after hepatitis c virus infection: A population-based study of u.S. Veterans. Hepatology 2009;49:116-23.

165. Iloeje UH, Yang HI, Jen CL, Su J, Wang LY, You SL, et al. Risk of pancreatic cancer in chronic hepatitis b virus infection: Data from the reveal-hbv cohort study. Liver international : official journal of the International Association for the Study of the Liver 2010;30:423-9.

166. Ben Q, Li Z, Liu C, Cai Q, Yuan Y, Wang K, et al. Hepatitis b virus status and risk of pancreatic ductal adenocarcinoma: A case-control study from china. Pancreas 2012;41:435-40.

167. Xing S, Li ZW, Tian YF, Zhang LM, Li MQ, Zhou P. Chronic hepatitis virus infection increases the risk of pancreatic cancer: A meta-analysis. Hepatobiliary & pancreatic diseases international : HBPD INT 2013;12:575-83.

168. Xu JH, Fu JJ, Wang XL, Zhu JY, Ye XH, Chen SD. Hepatitis b or c viral infection and risk of pancreatic cancer: A meta-analysis of observational studies. World journal of gastroenterology : WJG 2013;19:4234-41.

169. Nair J, Ohshima H, Nair UJ, Bartsch H. Endogenous formation of nitrosamines and oxidative DNA-damaging agents in tobacco users. Critical reviews in toxicology 1996;26:149-61.

170. Shapiro KB, Hotchkiss JH, Roe DA. Quantitative relationship between oral nitrate-reducing activity and the endogenous formation of n-nitrosoamino acids in humans. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 1991;29:751-5.

171. Correa P. Human gastric carcinogenesis: A multistep and multifactorial process--first american cancer society award lecture on cancer epidemiology and prevention. Cancer research 1992;52:6735-40.

172. Anderson KE PJ, Mack TM. Pancreatic cancer. In: Schotenfeld d, fraumeni jfj, editors. Cancer epidemiology and prevention. New York: Oxford University Press; 1996. p. 725-771.

173. Luo J, Nordenvall C, Nyren O, Adami HO, Permert J, Ye W. The risk of pancreatic cancer in patients with gastric or duodenal ulcer disease. International journal of cancer Journal international du cancer 2007;120:368-72.

174. Borch K, Kullman E, Hallhagen S, Ledin T, Ihse I. Increased incidence of pancreatic neoplasia in pernicious anemia. World journal of surgery 1988;12:866-70.

175. Hsing AW, Hansson LE, McLaughlin JK, Nyren O, Blot WJ, Ekbom A, Fraumeni JF, Jr. Pernicious anemia and subsequent cancer. A population-based cohort study. Cancer 1993;71:745-50.

176. Laiyemo AO, Kamangar F, Marcus PM, Taylor PR, Virtamo J, Albanes D, Stolzenberg-Solomon RZ. Serum pepsinogen level, atrophic gastritis and the risk of incident pancreatic cancer--a prospective cohort study. Cancer epidemiology 2009;33:368-73.

177. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, et al. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. The Journal of molecular diagnostics : JMD 2013;15:234-47.

178. Rachakonda PS, Bauer AS, Xie H, Campa D, Rizzato C, Canzian F, et al. Somatic mutations in exocrine pancreatic tumors: Association with patient survival. PloS one 2013;8:e60870.

179. Kawesha A, Ghaneh P, Andren-Sandberg A, Ograed D, Skar R, Dawiskiba S, et al. K-ras oncogene subtype mutations are associated with survival but not expression of p53, p16(ink4a), p21(waf-1), cyclin d1, erbb-2 and erbb-3 in resected pancreatic ductal adenocarcinoma. International journal of cancer Journal international du cancer 2000;89:469-74.

180. Immervoll H, Hoem D, Kugarajh K, Steine SJ, Molven A. Molecular analysis of the egfr-ras-raf pathway in pancreatic ductal adenocarcinomas: Lack of mutations in the braf and egfr genes. Virchows Archiv : an international journal of pathology 2006;448:788-96.

181. Ogura T, Yamao K, Hara K, Mizuno N, Hijioka S, Imaoka H, et al. Prognostic value of k-ras mutation status and subtypes in endoscopic ultrasound-guided fine-needle aspiration specimens from patients with unresectable pancreatic cancer. Journal of gastroenterology 2013;48:640-6.

182. Andreyev HJ, Norman AR, Cunningham D, Oates JR, Clarke PA. Kirsten ras mutations in patients with colorectal cancer: The multicenter "rascal" study. J Natl Cancer Inst 1998;90:675-84.

183. Vega F, Iniesta P, Caldes T, Sanchez A, Lopez J, Dejuan C, et al. Association of k-ras codon 12 transversions with short survival in non-small cell lung cancer. International journal of oncology 1996;9:1307-11.

184. Keohavong P, DeMichele MA, Melacrinos AC, Landreneau RJ, Weyant RJ, Siegfried JM. Detection of k-ras mutations in lung carcinomas: Relationship to prognosis. Clinical cancer research : an official journal of the American Association for Cancer Research 1996;2:411-8.

185. Al-Mulla F, Milner-White EJ, Going JJ, Birnie GD. Structural differences between valine-12 and aspartate-12 ras proteins may modify carcinoma aggression. The Journal of pathology 1999;187:433-8.

186. Krengel U, Schlichting I, Scherer A, Schumann R, Frech M, John J, et al. Three-dimensional structures of h-ras p21 mutants: Molecular basis for their inability to function as signal switch molecules. Cell 1990;62:539-48.

187. Hruban RH, van Mansfeld AD, Offerhaus GJ, van Weering DH, Allison DC, Goodman SN, et al. K-ras oncogene activation in adenocarcinoma of the human pancreas. A study of 82 carcinomas using a combination of mutant-enriched polymerase chain reaction analysis and allele-specific oligonucleotide hybridization. The American journal of pathology 1993;143:545-54.

188. Jiao L, Zhu J, Hassan MM, Evans DB, Abbruzzese JL, Li D. K-ras mutation and p16 and preproenkephalin promoter hypermethylation in plasma DNA of pancreatic cancer patients: In relation to cigarette smoking. Pancreas 2007;34:55-62.

189. Crous-Bou M, Porta M, Lopez T, Jariod M, Malats N, Alguacil J, et al. Lifetime history of tobacco consumption and k-ras mutations in exocrine pancreatic cancer. Pancreas 2007;35:135-41.

190. Schuller HM, Zhang L, Weddle DL, Castonguay A, Walker K, Miller MS. The cyclooxygenase inhibitor ibuprofen and the flap inhibitor mk886 inhibit pancreatic carcinogenesis induced in hamsters by transplacental exposure to ethanol and the tobacco carcinogen nnk. Journal of cancer research and clinical oncology 2002;128:525-32.

191. Blackford A, Parmigiani G, Kensler TW, Wolfgang C, Jones S, Zhang X, et al. Genetic mutations associated with cigarette smoking in pancreatic cancer. Cancer research 2009;69:3681-8.