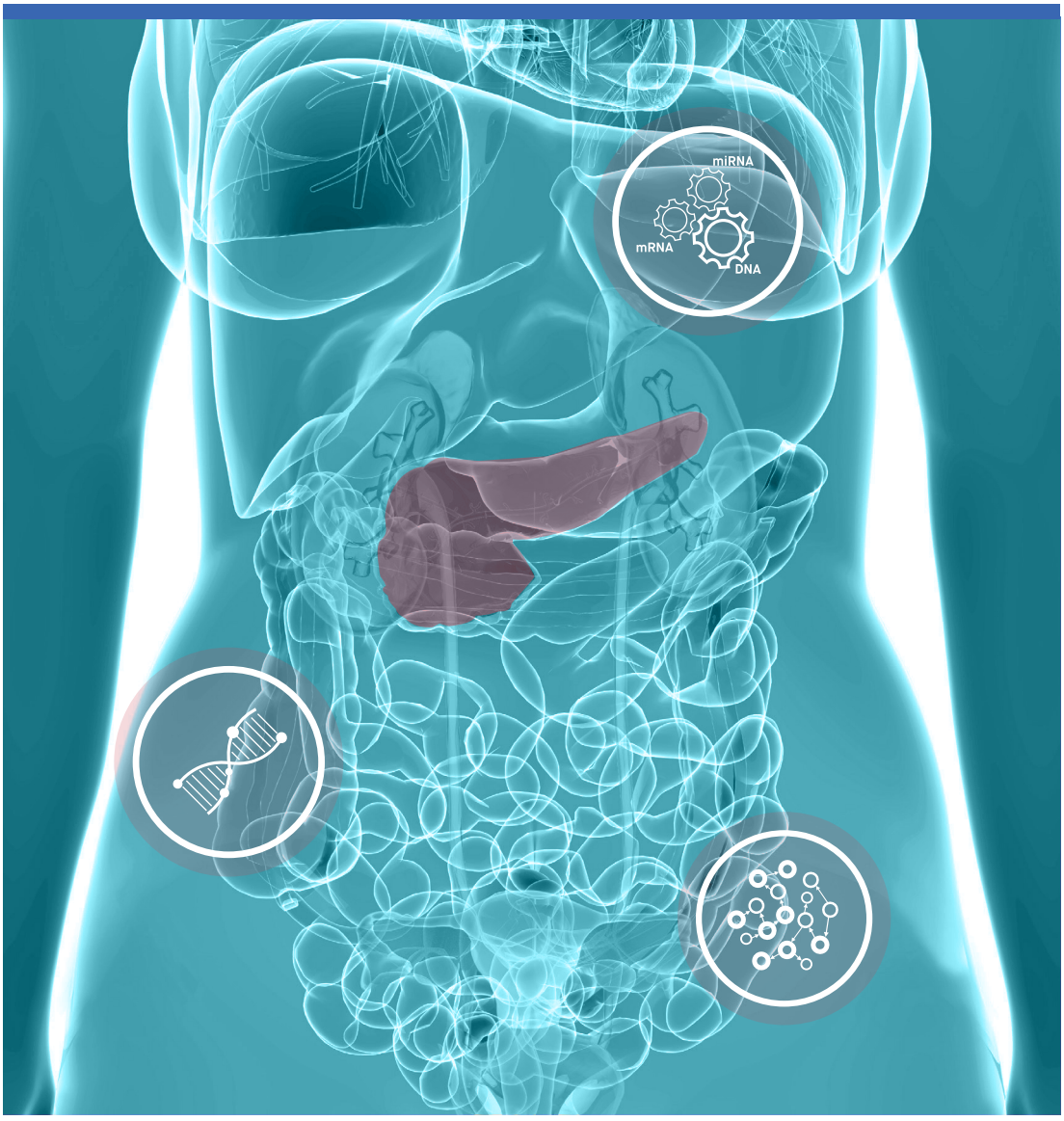
University College of Southeast Norway
Faculty of Arts and Sciences

Doctoral dissertation no. 2
2016

Vandana Sandhu

A systems biology approach to integrated molecular analysis in pancreatic and periampullary adenocarcinoma



HSN

Vandana Sandhu

A systems biology approach to integrated molecular analysis in pancreatic and periampullary adenocarcinoma

A PhD-dissertation in
Ecology

Faculty of Arts and Sciences
University College of Southeast Norway
Kongsberg, 2016

Doctoral dissertations at The University College of Southeast Norway no. 2

ISSN: 2464-2770 (print)
ISSN: 2464-2843 (online)
ISBN: 978-82-7206-407-4 (print)
ISBN: 978-82-7206-408-1 (online)

Printed: University College of Southeast Norway

Acknowledgments

Though only my denomination appears on the cover of this dissertation, a great many people have contributed to its engenderment. I owe my gratitude to all those people who have made this dissertation possible and without whom this was not possible at all. They have made my PhD experience memorable. I will cherish this experience forever.

My deepest gratitude is to my main supervisor, Professor Elin H. Kure. I have been astoundingly fortunate to have an advisor who gave me the liberty to explore on my own. Her patience and support availed me overcome many crisis situations and culminated this dissertation. Coming from a bioinformatics background, she motivated me to do translational research and gave me support and immense knowledge of the field. I could not have imagined having a better advisor and mentor for my PhD study.

Tone Jøran Oredalen and Dr. Mona Sæbø, thank you for giving me an opportunity to do my PhD at University College of Southeast Norway and for making my PhD ride a comfortable one. I also want to thank Professor Anne-Lise Børresen-Dale for allowing me to work in the Department of Cancer Genetics, Oslo University Hospital, Radiumhopitalet, and giving me excellent infrastructure and invaluable support.

My co-advisors, Dr. Urmila Kulkarni-Kale and Dr. Sangeeta Sawant, I am deeply grateful to you for encouraging me to do a PhD and for all the discussions regardless of long distance that helped me sort out the technical details of my work.

Professor Ole Christian Lingjærde insightful comments and constructive criticisms at different stages of my research were thought instigating. I am thankful to you for all the positive criticism, feedback and support.

Dr. Silje Nord, I am grateful to you for holding me to a high research standard and enforcing stringent validations for each research result, and thus edifying me how to do research.

I also want to thank Professor Caroline Verbeke, Dr. Knut Jørgen Labori and Julian Hamfjord for carefully reading and commenting on revisions of my thesis and for all the discussions regarding surgery, pathology and important considerations in the clinical

setting. I would also like to acknowledge the surgeons Trond Buanes and Knut Jørgen for providing the biobanked tissues.

I am also indebted to the members of the Elin's group with whom I have interacted during the course of my graduate studies. Particularly, I would like to acknowledge Martina Skrede, Astrid Dalsgaard, Julian Hamfjord, Inger Marie Bowitz Lothe for the many valuable discussions that helped me understand my research area better. I appreciate the efforts Martina and Astrid have taken to teach and show me the lab techniques. Special thanks to Inger Marie for collecting the tissues and performing microscopic and macroscopic pathology work.

Vilde Haakensen and Inger Marie, I am thankful to you for being such nice office mates.

I am most grateful to the collaborators for lending me their expertise and intuition to my scientific and technical problems: Dr. Peter Van Loo, Dr. David Wedge and Stefan Dentro at Welcome Trust Sanger Institute in Cambridge and Dr. Urmila Kulkarni-Kale, Dr. Sangeeta Sawant, Dr. Mohan Kale and Gaurav Dube at Savitribai Phule Pune University, India.

I am thankful to Daniel Nebdal for all the technical help. I am also grateful to Mary Anne Aune and Helga Veronica Tinnesand at University College of Southeast Norway and Mona Hagen and Gry Aarum Geitvik at Department of Cancer Genetics for helping me with administrative work.

Many friends have helped me stay sane through these three years. Their fortification and care availed me overcome setbacks and stay fixated on my graduate study. I greatly value their friendship and deeply appreciate their belief in me. I am grateful to them for helping me adjust to a new culture and a new country.

Most importantly, none of this would have been possible without the love and patience of my family. They have been a constant source of love, concern, support and strength all these three years, and I dedicate this thesis to them. I would relish expressing my heart-felt gratitude to my family. I am also grateful to my partner who supported me through this venture and providing a "writing space" and understanding me during both frustrating and joyful phases of my PhD.

Finally, I appreciate the financial support from University College of Southeast Norway and Oslo University Hospital that funded the research discussed in this dissertation. The financial support from the Norwegian Radium Hospital Research Foundation, The South-Eastern Regional Health Authority and Hole's Foundation is also greatly appreciated.

Summary

Pancreatic and periampullary adenocarcinoma (PA) is a highly lethal disease, for which the mortality rate closely parallels the incidence rate. Most of the PA patients remain asymptomatic until the disease reaches an advanced stage with an overall survival of <5%. More than 85% of patients are unressectable due to locally advanced or metastatic disease at time of diagnosis.

There are no early diagnostic or good prognostic markers of pancreatic cancer. Most PA arise from microscopic non-invasive epithelial proliferations within the pancreatic ducts, referred to as pancreatic intraepithelial neoplasia and less frequently from other benign lesions like Intraductal papillary mucinous neoplasm and mucinous cystic neoplasm. Surgical resection is regarded as the only potentially curative treatment. The studies of PA have suggested structural variations and mutations in *KRAS, SMAD4, CDKN2A* and *TP53* as the major driver events. However, the overall prevalence of genetic events is highly variable between PA from different patients.

The overall goal of this thesis is to gain insight into the molecular mechanisms of PA using a system biology approach. The miRNA and mRNA expression profiling of 85 PA tissue samples stratified the patients based on the morphological subtypes pancreatobiliary and intestinal. The study identified deregulated pathways and potential prognostic markers between the two morphological subtypes. Copy number aberration analysis of 60 PA samples also identified differences between the morphological subtypes. By integrating transcriptomic and genomic data we identified driver genes and pathways in PA. The integration of multiple *–omics* data takes into account cross talk at multiple levels in complex networks, and elucidate the variations that single *–omics* data cannot explain. Further, the transcriptome and genome profiles of three xenograft cell lines were compared to the original tumor they were derived from. The xenografts cell lines and original tumors had similar morphology, degree of differentiation, and genomic and transcriptomic profiles. The similarities suggest that the xenograft cell line can be use as *in vitro* models for studying the disease.

The main characteristic of PA is the abundant desmoplastic stroma, which is involved in the aggressiveness of the disease. The role of miRNAs in 20 pancreatobiliary PA, the more

aggressive subtype was deciphered. miRNAs were found to have a role in facilitating tumor stroma interactions by regulating the pathways involved in tumor stroma interaction.

This research identifies important driver genes, miRNAs and commonly aberrated genomic loci in PA of prognostic relevance. The identified intercalated network of information from the analyzed *–omics* data contribute to the current knowledge of the molecular biology of pancreatic and periampullary adenocarcinoma.

Aims of the study

The overall aim of this thesis is to gain insight into the molecular mechanisms of periampullary adenocarcinomas using a systems biology approach. Periampullary adenocarcinomas were analyzed at multiple molecular levels guided by the specific aims as follows:

1. Molecular profiling of miRNAs and mRNAs of periampullary adenocarcinomas to identify prognostic markers associated with morphological subtypes and anatomical sites.
2. To investigate miRNA and mRNA profiles and pathways that facilitate tumor-stroma interactions.
3. To investigate DNA copy number profiles, putative driver genes and associated pathways in periampullary adenocarcinomas.
4. To compare mRNA profiles of primary xenograft cell lines to the adenocarcinomas they were derived from.

List of papers

Paper I

Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas

Sandhu V., Bowitz Lothe I.M., Labori K.J., Lingjaerde O.C., Buanes T., Dalsgaard A.M., Skrede M.L., Hamfjord J., Haaland T., Eide T.J., Borresen-Dale A.L., Ikdahl T., Kure E.H.

Molecular Oncology, 2015, 9, 758-771.

Paper II

Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma

Sandhu V., Bowitz Lothe I.M., Labori K.J., Skrede M.L., Hamfjord J., Dalsgaard A.M., Buanes T., Dube G., Kale M.M., Sawant S., Kulkarni-Kale U., Børresen-Dale A.L., Lingjærde O.C., Kure E.H.

Molecular Oncology, November 2015 *in press.*

Paper III

The genomic landscape of pancreatic and periampullary adenocarcinoma

Sandhu V., Wedge D.C., Bowitz Lothe I.M., Labori K.J., Dentro S., Buanes T., Skrede M.L., Dalsgaard A.M., Lingjærde O.C., Børresen-Dale A.L., Ikdahl T., Van Loo P., Nord S., Kure E.H. (Manuscript)

Paper IV

Generation and characterization of novel pancreatic adenocarcinoma xenograft models and corresponding primary cell lines

Wennerstrom A.B., Bowitz Lothe I.M., Sandhu V., Kure E.H., Myklebost O., Munthe E.

PLoS One, 2014, 9, e103873.

Table of contents

1. Introduction

Pancreatic cancer is the fourth most common cause of cancer-related deaths in the Western countries and is projected to be the second leading cause of cancer death by 2030 [7, 52]. Annually, about 700 pancreatic ductal adenocarcinomas (PDAC) are diagnosed in Norway and the overall five-year survival rate is only 5% (Cancer Registry of Norway) [4]. The high mortality rate for pancreatic cancer patients is due to metastatic disease at time of diagnosis, lack of early symptoms of disease, resistance to therapy and a complex tumor microenvironment that hampers drug delivery to the cancer cells [1]. Almost 85% of patients diagnosed with pancreatic cancer are inoperable due to either locally advanced disease or distant spread at time of diagnosis [62]. Adenocarcinomas originating from the pancreatic head, the distal common bile duct, duodenum and the ampulla are collectively referred to as periampullary adenocarcinomas (PAs). They are treated by the same surgical procedure, and their diagnostic distinction may be difficult clinically, radiologically and morphologically.

1.1. Pancreatic cancer epidemiology

The incidence rate and mortality rate for pancreatic cancer is almost equal and the relative five-year survival rate has not improved much in the last 35 years (Figure 1). In Norway the incidence rate is slightly higher in men than in women (Cancer Registry of Norway, male to female ratio of 1.2), while in the US the burden is almost equal according to the American Cancer Society [4, 5]. The incidence rate of pancreatic cancer increases with age, and the risk is highest at ages above 60. Fewer cases are detected below the age of 40 in both males and females [4] (Figure 2).

The risk factors for developing pancreatic cancer can be inherited or non-inherited. The inherited risk factors include a family history of hereditary pancreatitis, cystic fibrosis, hereditary breast and ovarian cancers (*BRCA1* and *BRCA2* mutations), familial pancreatic cancer, hereditary nonpolyposis colorectal cancer with *MLH1* mutation, familial atypical multiple mole melanoma syndrome and Peutz-Jeghers syndrome [1, 8, 9]. Non-inherited risk factors are smoking, obesity, chronic pancreatitis, and some association was linked to diabetes and consumption of meat and elevated intake of alcohol [1, 2, 8, 9].

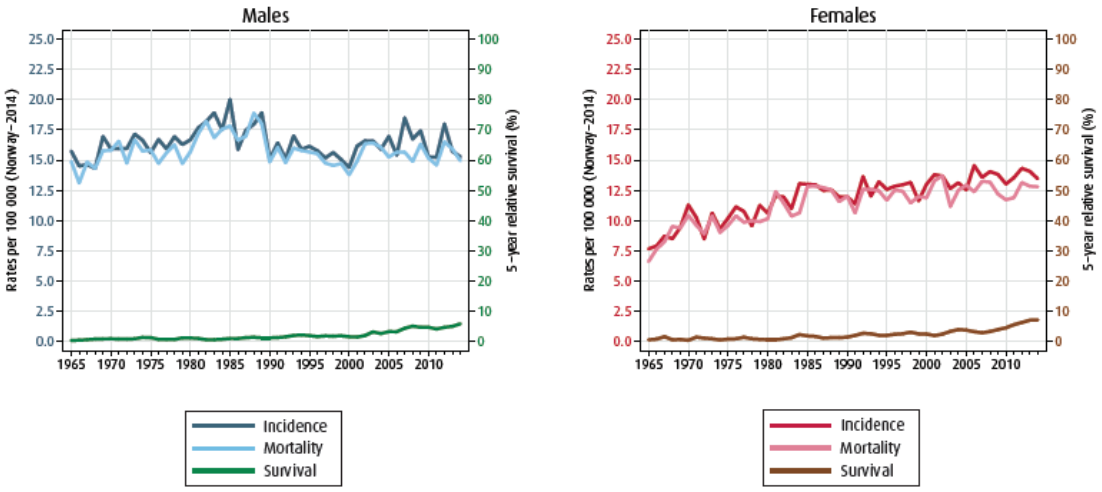


Figure 1: Relative survival, incidence rates and mortality rate of pancreatic cancer patients. Reprinted with permission from Cancer Registry of Norway [4].

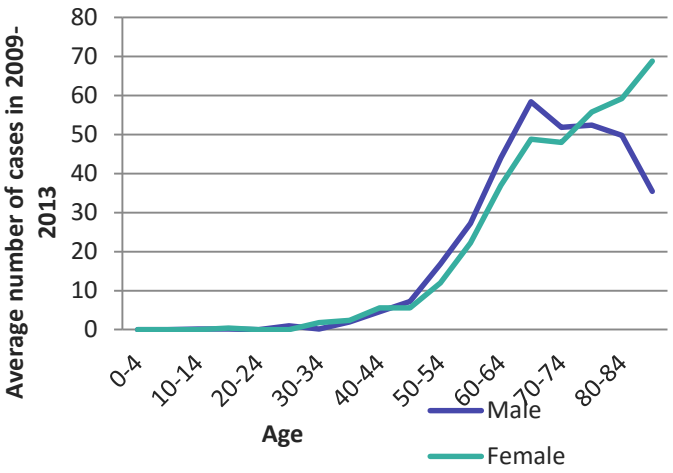


Figure 2: Average number of new cases of pancreatic cancer by age from 2009-2013 http://www.kreftregisteret.no/.

1.2. Anatomy of the pancreas

Pancreas is a hammer-shaped organ, measuring 12.5-15 cm in length and weighs 60-100 g [22]. Its gross anatomy may be divided into four subsections; head, neck, body and tail. The head of the pancreas lodges within the curve of the duodenum. The head is connected to the main portion of the pancreatic body by a slightly narrow portion called the neck, anterior to the portal vein. The pancreatic tail extends to the spleen (Figure 3) [6]. The pancreatic ducts form an intricate system of which the most peripheral ramifications drain acinar clusters and join to form intra- and interlobular ducts, which

eventually collect in the main pancreatic duct. The latter runs longitudinally through the entire pancreas to empty into the duodenum at the ampulla of Vater (Figure 3) [6].

The pancreas is a secretory structure with an internal hormonal function (endocrine) and external digestive function (exocrine). The exocrine pancreas constitutes 80% of the tissue mass of the organ and is composed of acinar cells and the draining ducts, which are organized in discrete parenchymal lobules. These cells produce and secrete inactive enzyme precursors (zymogens), which upon arrival in the duodenum are activated such that they can deploy their digestive function. The endocrine pancreas is composed of four specialized endocrine cell types, the alpha cells secreting glucagon, the beta cells secreting insulin, the delta cells secreting somatostatin and human pancreatic polypeptide cells. These four specialized endocrine cell types co-localize in small well-circumscribed clusters, which are called the islets of Langerhans. The endocrine pancreas regulates metabolism and glucose homeostasis through the secretion of hormones into the bloodstream [22].

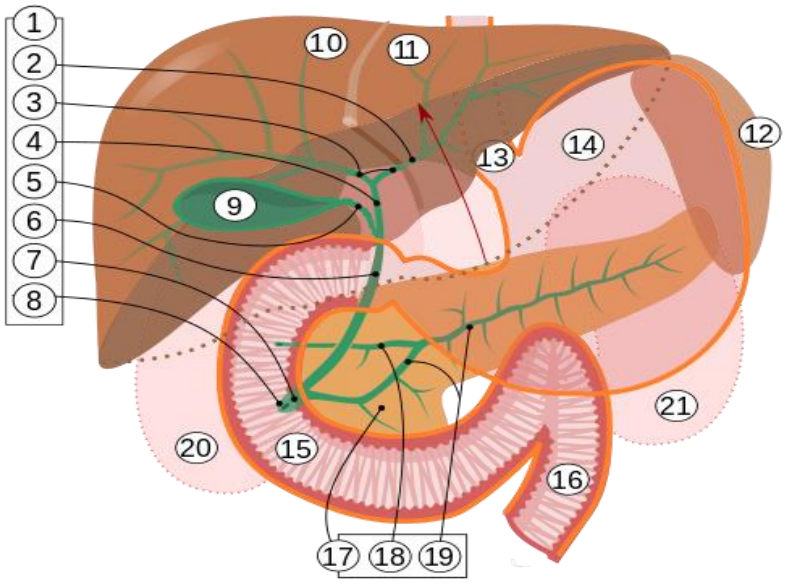


Figure 3: Anatomy of the pancreas 1=bile ducts, 2: Intrahepatic bile ducts, 3: hepatic ducts, 4: common hepatic ducts, 5: cystic ducts, 6: common bile ducts, 7: ampulla of Vater, 8: major duodenal papilla, 9: gallbladder, 10-11: right and left lobes of liver, 12: spleen, 13: esophagus, 14: stomach, 15: duodenum, 16: jejunum, 17: pancreas, 18: accessory pancreatic duct, 19: pancreatic duct, 20-21: right and left kidneys.
Source: https://commons.wikimedia.org/wiki/File:Biliary_system_multilingual.svg

1.3. Clinicopathological classification of pancreatic tumors

1.3.1. TNM staging

The extent of disease at diagnosis is recorded according to the TNM staging system, as it have been defined by The International Union Against Cancer (UICC) and The American Joint Committee on Cancer (AJCC) [10, 14]. The staging system aims at patient stratification for the purpose of treatment and outcome prediction. The TNM staging system is based on three main descriptors:

- T: Describes the size and/or extent of the primary tumor
- N: Describes the spread to regional lymph nodes
- M: Describe spread to extraregional lymph nodes or distant sites

Table 1: TNM classification of pancreatic cancer

Stage	**T**	**N**	**M**
Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1, T2, T3	N1	M0
Stage III	T4	Any N	M0
Stage IV	Any T	Any N	M1

The clinical cancer stage groups are defined by the various combinations of the T, N, and M stages. Primary Tumor (T) is broadly categorized into classes of T0 to T4, Tis and TX. They are defined as TX: primary tumor cannot be assessed; T0: No evidence of primary tumor; Tis: Carcinoma *in situ*; T1: Tumor is limited to the pancreas and < 2 cm or less in greatest dimension; T2: Tumor is limited to the pancreas and > 2 cm in greatest dimension; T3: Tumor extends beyond the pancreas, but without involvement of the celiac axis or the superior mesenteric artery; T4: Tumor involves the celiac axis or the superior mesenteric artery. Tumor stage T4 reflects surgical irresectability. The regional lymph node status is defined as N0 if there are no regional lymph node metastases, N1 if there are lymph node metastases, and NX if lymph node metastases cannot be assessed. Absence or presence of extraregional or distant metastasis is recorded as M0 or M1, respectively [10].

1.3.2. Histological grade

The histological grade of pancreatic tumors is a measure of the degree of differentiation determined by the microscopic appearance of cancer tissue. The degree of differentiation refers to the resemblance between the tumor and normal tissue of origin. Tumor differentiation is categorized as G1 to G4. The higher the grade, the more undifferentiated and aggressive the tumor is. According to the grading system recommended by UICC and AJCC, pancreatic adenocarcinomas are graded based on the percentage of glandular formation: G1 grade (well differentiated) when > 95% of the

tumor is composed of glands, G2 grade (moderately differentiated) when 50-95% of the tumor is glandular, G3 (poorly differentiated) when < 50% of the tumor is glandular, G4 for an undifferentiated tumor, and GX if the tumor grade cannot be accessed [10, 14]. The histological grade of differentiation is an independent prognostic marker of pancreatic cancer [11-12].

1.3.3. Clinical prognostic factors

The presence or absence of residual disease (R) also regarded synonymously as resection margin status. Residual tumors are categorized as R0-R2 where R0 is clear margins, R1 as residual cancer cell seen microscopically and R2 as residual cancer cells visible to the naked eye [10, 14, 22]. Vascular invasion is diagnosed clinically as histological involvement of vessels including invasion of adventitia of major branches of vessels [14]. Vascular invasion can also be categorized as V1 or V2 depending on whether one can see the vascular invasion by naked-eye inspection or only in the microscope [22]. The residual disease status and vascular invasion are prognostic factors of pancreatic cancer [13, 61].

1.3.4. Histopathological classification of pancreatic tumors

The WHO classification of tumors of the pancreas [14] classifies exocrine pancreatic neoplasias as benign, premalignant or malignant. They may be cystic or solid in appearance:

- *Benign:* The benign tumors include acinar cell cystadenoma and serous cystadenoma.
- *Premalignant lesions:* The premalignant lesions include pancreatic intraepithelial neoplasia, grade 3 (PanIN-3), intraductal papillary mucinous neoplasm (IPMN) with low, intermediate or high-grade dysplasia, intraductal tubulopapillary neoplasm and mucinous cystic neoplasm (MCN) with low, intermediate or high-grade dysplasia.
- *Malignant:*
    - ✓ *Ductal adenocarcinomas:* The ductal adenocarcinomas are composed of well to moderately developed glandular and duct like structures, which infiltrate the pancreatic parenchyma and grow in haphazard pattern and

usually demonstrate luminal and intracellular production of mucin and abundant desmoplastic stromal response. Of the ductal adenocarcinomas 60-70% are localized in the head of the pancreas, 5-15% in body and 10-15% in the tail of the pancreas.

- ✓ *Ductal adenocarcinoma variants and mixed neoplasms of the pancreas:* They are ductal adenocarcinomas with mixed neoplasms and histological variants of ductal adenocarcinomas. They are rare tumors and have distinct clinical and prognostic significance. The tumors in these category include adenosquamous carcinoma (both ductal and squamous differentiation), colloid carcinoma (ductal epithelial neoplasm with stromal mucin and neoplastic cells in association with IPMNs), medullary carcinomas (poor differentiation, syncytial growth pattern with spindle cells), hepatoid carcinoma (hepatocellular differentiation), undifferentiated carcinomas (significant component of the neoplasm does not show a definitive direction of differentiation), signet ring cell carcinoma and carcinomas with mixed differentiation (> 30% cell types are of different types).

1.3.5. Anatomical subtypes of adenocarcinoma

Pancreatic ductal adenocarcinoma (PDAC) can arise in head, neck, body or tail, but is more common in the head of the pancreas. Adenocarcinomas located in pancreatic head, which constitutes the majority of adenocarcinomas, can be divided based on site of origin [14]. The four sites of origin are the pancreatic ducts, the distal bile duct, the ampulla and the duodenum, which are collectively referred to as periampullary adenocarcinoma (PA)[15].

1.3.6. Morphological subtypes

PAs are subtyped based on morphology as pancreatobiliary or intestinal [17, 18]. The intestinal subtype resembles colonic cancer in that the cancer cells form fairly large glands and have oval-shaped nuclei, while in the pancreatobiliary subtype the tumor

glands are smaller and branched, and the cancer cells have round nuclei [16, 17]. A more favorable prognosis has been associated with the intestinal subtype of PA [17].

1.4. Symptoms and diagnosis

The most commonly reported symptoms of PAs are loss of appetite, jaundice and abdominal pain. The symptoms like jaundice, pale colored stools, dark urine and itching arise due to bile duct obstructions and secondary hyperbilirubinemia [3]. Unfortunately, symptoms of pancreatic cancer arise in late stages of the disease and the patients usually have no symptoms until the cancer has spread to other organs.

The diagnostic work-up of pancreatic cancer patients is based on the following investigations:

1.4.1. Imaging

Imaging tests are performed to visualize the primary tumor and assess its relationship to surrounding structures, e.g. large blood vessels. Imaging also allows detection of regional or distant cancer spread, and hence staging of the disease. Furthermore, imaging may allow for assessment of the effect of (neoadjuvant) treatment, and the identification of disease recurrence. Computer tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and ultrasound can be used in the diagnostic work-up of pancreatic cancer [23, 24]. Most commonly CT is used when evaluating resectability.

1.4.2. Serum biomarkers

The only Food and Drug Administration (FDA) approved biomarker for pancreatic cancer is Carbohydrate antigen 19-9 (CA 19-9). It is measured in serum for monitoring of disease progression or recurrence [19, 20]. However, it cannot be used as a screening marker for early diagnosis and for the distinction between chronic pancreatitis and pancreatic cancer, because of its low sensitivity and specificity [20]. CA19-9 levels are associated with several cancers (including cancers of the pancreas, bile system, liver, gastrointestinal tractus) and benign lesions such as cholangitis, cirrhosis and other cholestatic diseases.

1.4.3. Biopsy/cytology

A small sample of the tumor can be obtained via a percutaneous or an endoscopic route. Optical guidance during puncturing can be based on ultrasound, CT imaging, endoscopic ultrasound (EUS) or endoscopic retrograde cholangiopancreatography (ERCP) [24]. The tumor sample can be either a small tissue piece (biopsy), which can be submitted to histological examination, or a small volume of aspirated tumor cells that can be used for cytological investigation (fine needle aspiration, FNA) [24].

1.5. Treatment

To date, surgery is considered the only potentially curative treatment, which can be complemented with neo-adjuvant and/or adjuvant chemotherapy. For patients with unresectable disease, palliative chemotherapy may be indicated. The standard operation for tumors of the pancreatic head is a pancreatoduodenectomy (PD), whereas tumors of the body or tail can be resected using a distal pancreatectomy [28]. PD (also known as Kausch–Whipple procedure) involves removal of the head of the pancreas, duodenum, distal portion of the stomach, common bile duct and gallbladder [25, 26, 27]. For resection of tumors of the body and the tail of the pancreas, distal pancreatectomy is performed, which includes resection of the pancreas to the left of superior mesenteric vein, dissection of peripancreatic lymph nodes and removal of the spleen [27, 33]. Total pancreatectomy is indicated only for patients with multilocular or large tumors of pancreas and consists of a combination of PD and distal pancreatectomy [27]. Extended lymphadenectomy (lymph nodes removal), venous or arterial resections can be performed in combination with PD and pancreatectomy in case of large or so-called borderline resectable tumors with the aim to improve the limited survival that non-surgical treatment offers for these patients [27, 189-191]. Palliative surgery is performed for patients with pancreatic head cancer, who suffer from tumor-related biliary and/or intestinal obstruction and are found to have metastatic or unresectable disease during a planned resection [34]. The surgical procedure consists of constructing a bypass between the bile duct (hepaticojejunostomy) and/or the stomach (gastroenterostomy) and a small

bowel loop. Endoscopic insertion of a plastic or metal stent through the obstructed bile duct or duodenum is an alternative approach that helps to drain the bile into the intestine [34]. Cancer cachexia is a challenging aspect in all stages of the disease and requires nutritional support, diabetic control or pancreatic enzyme supplementations.

The standard treatment for patients with resectable disease is primary surgery followed by adjuvant chemotherapy. Various drugs are used, such as 5-fluoro-uracil (5-FU) or gemcitabine, or a combination of drugs. 5-FU and gemcitabine give a limited survival advantage compared to treatment with surgery only [37, 38]. Gemcitabine has shown survival benefit over no chemotherapy [39]. Both gemcitabine and 5-FU show similar results in pancreatic cancer patients with no difference in disease-free or overall survival between the two treatments [40]. Targeted therapy with the use of erlotinib (EGFR tyrosine kinase inhibitor) plus gemcitabine is also used and has shown a marginal improved survival compared to gemcitabine alone [63]. In advanced pancreatic cancer patients, the co-administration of gemcitabine and nab-paclitaxel demonstarted better median overall survival of 8.5 months compared with 6.7 months with gemcitabine alone [43, 44]. A similar survival benefit has been shown for FOLFIRINOX (folinic acid or leucovorin, fluorouracil, irinotecan and oxaliplatin) compared to gemcitabine alone, but the treatment is associated with increased toxicity and should be reserved for patients with a good performance status [210]. The surgery-first strategy followed by adjuvant chemotherapy is the most universally accepted evidence-based approach to resectable pancreatic cancer. However, some centers advocate neoadjuvant treatment strategies with chemotherapy or radiotherapy given before surgery [58]. The role of chemoradiation as a component of adjuvant therapy is still controversial but chemoradiation allow a small percentage of patients with locally advanced disease to undergo potentially curative resection [41, 42]. Recent studies have shown promising results for the use of FOLFIRINOX as neoadjuvant therapy for patients with borderline resectable pancreatic adenocarcinoma [35, 36].

1.6. The microenvironment of pancreatic and periampullary adenocarcinoma

The microenvironment of pancreatic adenocarcinoma consists largely of fibroblasts, pancreatic stellate cells, extracellular matrix (ECM) proteins (e.g. collagen I/III, fibronectins), endothelial cells, pericytes, immune cells, and nerve fibers [51, 53]. The stroma of malignant adenocarcinomas represents up to 90% of the volume [60]. The major cellular constituents of desmoplastic stroma in pancreatic cancer are pancreatic stellate cells and fibroblasts [53]. Pancreatic stellate cells play an important role in pancreatitis and pancreatic cancer by transforming to an activated state during which they acquire characteristics of myofibroblasts and express α-smooth muscle actin (αSMA) [55]. The presence of desmoplastic stroma has been associated with worse clinical outcomes and is believed to contribute to the aggressive nature of this tumor by fostering tumor growth, metastatic spread and drug resistance [53, 54]. Co-culturing experiments of pancreatic stellate cells with pancreatic cell lines suggest that the former increase the invasiveness properties of the cancer cells [56]. A number of pathways and growth factors are associated with stromal activation in pancreatic cancer like transforming growth factor (TGF–β), hepatocyte growth factor (HGF), fibroblast growth factor (FGF), epidermal growth factor (EGF) and epithelial-mesenchymal transition (EMT) [53].

1.7. Molecular biology of pancreatic cancer

1.7.1. Pancreatic cancer development

Pancreatic cancer develops through the progression of precursor lesions to invasive adenocarcinoma of pancreas. The most common neoplastic precursor lesion is PanIN. Pancreatic cancer can also evolve from IPMNs or MCN [45, 46]. IPMN is common, but only a small proportion of patients with IPMN progress to invasive carcinoma. There are many morphological alterations with advancing PanIN stages as compared to normal ducts. Cell proliferation rate and dysplastic growth rate increase in later PanIN stages. PanINs acquire many somatic genetic alterations that are similar to those found in

invasive pancreatic cancer, and the prevalence of these genetic alterations rises with advancing PanIN stages [45, 47, 48]. Specifically, *CDKN2A* and *KRAS* mutations, loss of heterozygosity at chromosomes 9p, 17p and 18q harboring *CDKN2A*, *TP53* and *SMAD4*, respectively are common events [49, 50]. *KRAS* mutations have been identified as one of the earliest events found in about 30% of early lesions, and the mutation rate increases as disease progresses and reaches nearly 100%. The loss of function of the tumor suppressors *CDKN2A*, *INK4A* and *ARF* are observed in advanced lesions. *TP53* mutations, loss of *SMAD4* and *BRCA2* arise in later stages of PanINs. The loss of telomerase activity or telomerase shortening is observed in early PanIN stages. The telomerase shortening causes inactivation of the DNA damage response checkpoint, which in turn leads to genomic instability. Reactivation of telomerase activity occurs in advanced stages of pancreatic adenocarcinoma, which facilitates growth of immortal cancer cells (Figure 4) [48, 212].

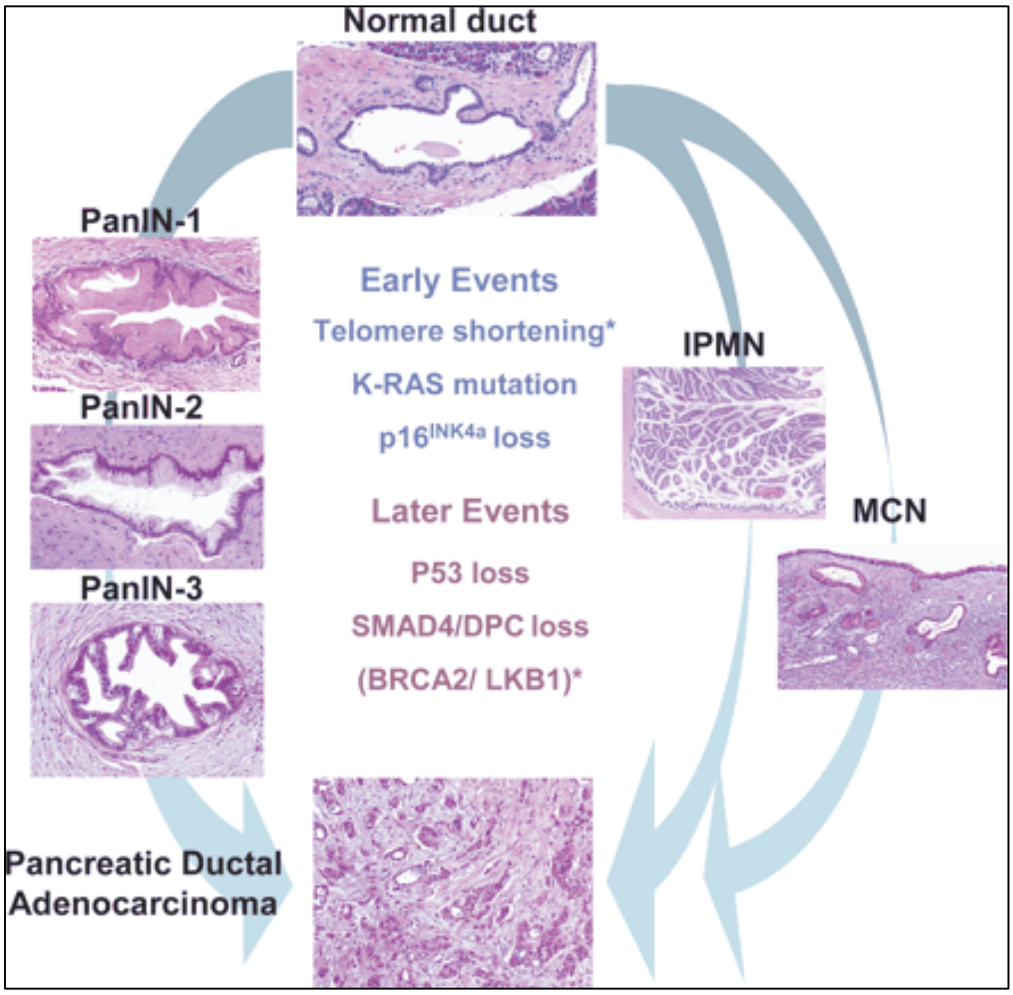


Figure 4: Genetic and morphological changes from PanINs, IPMN and MCN to invasive adenocarcinoma. Asterisks indicate events that are not common to all the precursors *BRCA2* loss and telomerase shortening is common in PanINs while *LKB1* loss is common in IPMN and PDACs. Adapted from *Hezel et al., 2006* [212].

1.7.2. Central dogma of molecular biology

The "Central dogma of molecular biology" defines the flow of information from DNA to mRNA (transcription) to protein (translation) [87]. This unidirectional flow of information can be used as a basis to study complex interactions in biological processes. Based on the outcome of the Human Genome project there are an estimated 25,000 protein coding genes in the human genome and many additional genes produce non-coding RNAs like transfer RNA, ribosomal RNA and miRNA, the latter are known to regulate gene expression [100, 101]. Studying multiple *–omics* (meaning complete knowledge) data

such as transcriptomic, genomics and proteomics by using a systems biology approach is a step toward deciphering molecular interactions.

1.7.3. miRNAs

miRNAs are small, evolutionarily conserved, single stranded and non-coding RNA molecules [67]. They are hairpin like structures and consist of 21-25 nucleotides, and can regulate gene expression at the post-transcriptional level [68, 69]. More than 2000 entries of human miRNAs are identified and available at http://www.mirbase.org/ (Release 21) [70]. It is predicted that miRNAs account for 1-5% of the human genome and regulate 30% of the protein coding genes [64, 71, 72]. They play a crucial role in regulation of gene expression and in controlling diverse cellular and metabolic pathways [64, 68]. They have also been shown to have a role in regulation of tumor-stroma interactions [192-193].

The miRNA biogenesis is initiated in the nucleus and subsequently developed to a mature miRNA in the cytoplasm. miRNA processing in the nucleus is a two-step process. The first step is the transcription of pri-miRNA by RNA polymerases II/III. The second step is cleavage of the pri-miRNA to pre-miRNA by the microprocessor complex Drosha-DGCR8 (Pasha). The pre-miRNA, which is a hairpin like structure is transported from the nucleus to the cytoplasm by Exportin-5-Ran-GTP. In the cytoplasm, RNAase DICER and the double stranded RNA-binding protein TRBP complex cleave the pre-miRNA to a mature miRNA. Further, the functional miRNA strand is loaded together with the Argonaute (Ago2) protein and forms RNA induced silencing complex (RISC) while the passenger strand is degraded. The mature miRNA silence the target mRNA through mRNA cleavage, translational repression or deadenylation [64] (Figure 5).

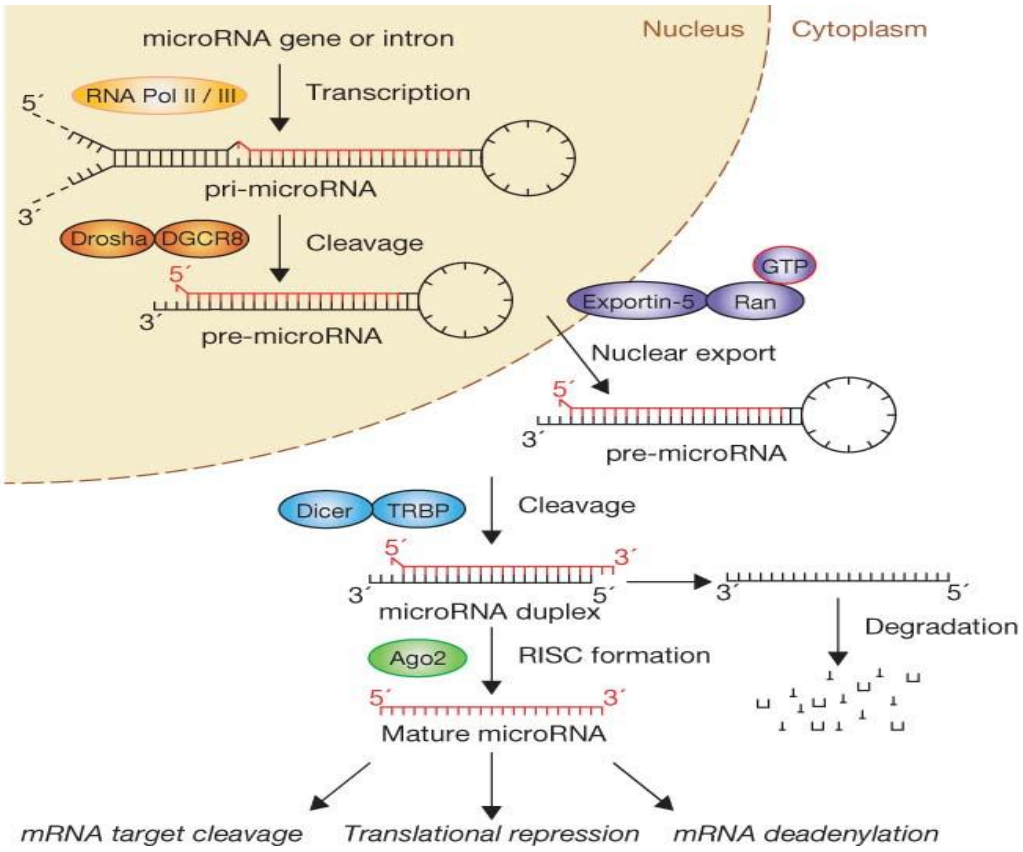


Figure 5: The miRNA biogenesis mechanism. RISC: RNA induced silencing complex, Pol II/III: RNA polymerases II and III, pre-microRNA: precursor miRNA, pri-miRNA: primary miRNA. Reprinted by permission from Macmillan Publishers Ltd: [Nature Cell Biology] [64], copyright (2009).

miRNAs bind to target sites by incorporating nearly perfect complementarity in plants, while in humans stringent requirements are required for binding. In humans, a miRNA binds to a target gene by perfect base pairing in only 2-8 regions called the 'seed' at 5'-end, and there should also be a reasonable complementarity to the miRNA 3'-end to stabilize the interaction [79]. miRNAs exhibit two properties called multiplicity and cooperativity. Multiplicity means that one miRNA can target more than one gene while cooperativity means one gene can be targeted by more than one miRNA [78]. There are many target site prediction algorithms available for predicting miRNA target genes that are based on sequence complementarity between miRNA and mRNA, thermodynamics, machine learning based approaches and evolutionary conservation [81]. Despite the large number of miRNA target site prediction tools the sensitivity and specificity of the tools are low [81]. These tools predict many targets that may depend on cell type, binding of additional cofactors, but it is also context dependent [197, 198]. Only a small subset of

miRNA-mRNA interactions are tumor related [198], therefore identifying functionally relevant target genes and pathways associated with dysregulated miRNA expression profiling is challenging. Publically available databases (miRTarbase and miRecords) are curated based on literature mining of already published experimentally validated targets and are useful resource for validating the interactions [82, 83].

miRNAs are categorized into miRNA families based on mature miRNA, sequence and structure of pre-miRNAs [84]. The members of miRNA families share many common target genes and co-target critical tumor suppressor genes [85, 86]. miRNAs can act as oncomiRs and miR-suppressors in various cancers. For instance, miRNA-15a, miRNA-16, miRNA-18 and miRNA-20a target oncogenes like *MYC* and tumor-suppressors like *RB1* in lymphomas and lung cancers [74, 195]. Studies have also shown that miRNAs can act as potential predictive and prognostic markers in cancers [76, 77].

1.7.4. Single nucleotide polymorphism and copy number aberrations

A single nucleotide polymorphism (SNP) is a variation in a single nucleotide in a specific position in a DNA sequence, and occurs at >1% within a population [88]. SNPs are known to be associated with phenotypic variation either through direct causal effects or by linkage disequilibrium [93]. SNP data analysis allows simultaneous measurement of allele specific copy number at many different SNP loci in the genome [99]. It allows thorough analysis of the genome accounting for non-aberrant cell infiltration, tumor aneuploidy and getting information of tumor subclones [99]. In cancer genomics the focus is on identifying altered genomic regions in genes and pathways by using high throughput technologies. The main goal is to identify genes that play a crucial role in cancer development, diagnosis and its treatment.

Variation in the human genome consists of two major types: (1) single nucleotide variation, in the form of DNA base-pair substitutions and short indels, and (2) structural variation affecting many base pairs, including inversions, translocations, insertions, deletions, and duplications [88]. The HapMap and ENCODE projects were initiated to describe the common patterns of human genetic variations, and these projects have facilitated the validation of millions of SNPs [89, 90, 91]. Copy number variants (CNV) are defined as DNA segments that are 1 kb or larger in size and are present at variable copy

number in comparison to a reference genome [92]. It has become apparent that they are quite common in the human genome. They can have dramatic phenotypic consequences as a result of altering gene dosage, disrupting coding sequences, or perturbing long-range gene regulation [92, 93]. CNVs between unrelated individuals differ by only 0.5% [95], while the copy number aberrations (CNA) in the cancer genome is substantial (i.e. hallmarks of cancer) [96]. CNA affect a large fraction of the cancer genome and may lead to copy number gains, amplifications and deletions [94]. Homozygous deletions are genes where both alleles are lost, while hemizygous deletions are genes where only one copy is lost resulting in loss of heterozygosity. Gains are defined as one or more copies of an allele at a genomic locus. Tumor suppressor genes are often deleted or mutated, destroying their tumor suppresive role while oncogenes are often amplified or overexpressed [98]. CNAs have critical roles in activating oncogenes and in inactivating tumor suppressors [97]. Evidence has been presented that increased copy number can be positively or negatively correlated with gene expression levels. For example, deletion of 10q23.31 is associated with downregulation of PTEN in endometrial cancer [211].

1.7.5. Dysregulated cancer pathways

The hallmarks of cancer as defined by Hanahan and Weinberg in 2011 includes ten essential biological processes, which are sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, deregulating cellular energetics, avoiding immune destruction, genome instability and destruction, tumor promoting inflammation and activating invasion and metastasis [29, 92]. Cancers accumulate genetic changes in a stepwise manner and specific molecular pathways, tumor suppressors and oncogenes are responsible for deregulating the normal machinery of the cell. The comprehensive genetic analysis of 24 PDACs performed by Jones et al. uncovered 1562 mutations in 1327 genes, assembeled into 12 core-signaling pathways (see Figure 6) [179]. The mutation spectrum in tumors is broad and heterogenous, where the genetic alteration corresponding to each pathway varies between the tumors, but the core pathways were mutated in 68% to 100% of the PDAC tumors [176, 179].

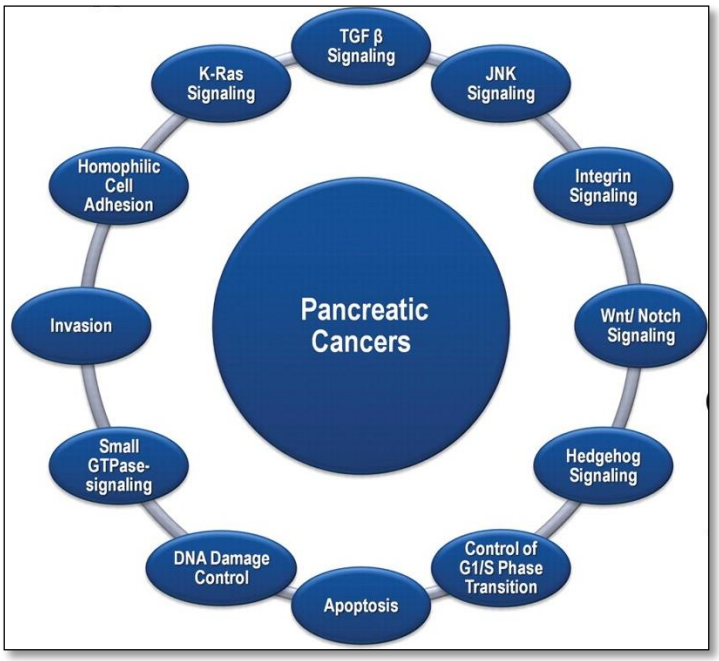


Figure 6: The figure shows the pathways commonly dysregulated in pancreatic cancer. The figure is adapted from Jones et al, 2012 (reprinted by permission from The American Association for the advancement of Science: Science [179]).

There is a large diversity of genetic changes occurring at varying frequency in pancreatic tumors leading to dysregulation of signaling pathways and disturbing the normal functions of cells. The change in functioning of biosynthetic and metabolic pathways confers growth advantages to cancerous cell and for its malignant phenotypes. The genetic profiles of pancreatic tumors are very different from normal cells, and miRNA and mRNA expression profiling allow measurement of biological activities, processes and changes in tumor cells. Expression profiling can be used to explore the underlying tumor biology with the goal of improving response to therapy and prognosis.

The important pathways known in various cancers for inhibiting apoptosis are inactivation of p53 and the PI3K pathway [196]. The genomic stability of the cells are regulated by various pathways like ATM and ATR in response to DNA damaging agents, which in turn activates p53 signaling that controls cell cycle and DNA repair [77]. RB1 and p53 are tumor suppressors that act as gatekeepers for cell growth and division and keep control of the cell cycle. The TGF-β pathway keeps check on cell cycle arrest by regulating the SMADs, which in turn regulates transcription of various cell cycle molecules [207]. Proliferative signaling like RAS and PI3K activates downstream signaling like MAPK, which

in turn modulates the expression of various genes involved in cell growth and survival [200]. The signaling pathways are intertwined and communicate with each other and regulate the functioning of the cell. These pathways are often dysregulated in various cancers including pancreatic cancer [205, 206]. Recently, Waddell et al. sequenced 100 PDAC genomes and showed that pancreatic cancer has a large number of structural variations [177]. Another recent publication by Childs et al. showed that common variations at 2p13.3, 3q29, 7p13 and 17q25.1 are associated with susceptibility to pancreatic cancer [208]. Besides these, a number of studies have been published documenting genomic alterations in pancreatic cancers using whole genome expression profiling for miRNAs and mRNAs; whole genome sequencing (WGS) and genotyping and various other methods [177-179, 186-188]. Despite of all the studies, we have scarce knowledge about the interactions between multiple mRNA, miRNAs and pathways. Thus, simultaneous analysis of different *-omics* features, which does not act in isolation but crosstalk at multiple levels in complex networks can elucidate the variations that single *-omics* data cannot explain.

A unique characteristic of pancreatic tumors is the presence of abundant desmoplastic stroma, which facilitates tumor migration, angiogenesis and expansion [57]. Losses of function in genes encoding DNA-repair enzymes promote inactivation of tumor suppressors and encode proteins that inhibit progression [104]. mRNA expression analysis allows identification of tumor suppressors and oncogenes that are often differentially regulated in the cancer genome [102, 103]. The pathways and gene expressions that are tissues specific and initiate tumor growth and development are not as well defined as for some other cancers like the Wnt-catenin pathway in colorectal cancer and BRCA1 gene mutation in breast and ovarian cancers [143, 144]. Also, the dysregulation of miRNAs is associated with cancer development and progression and can affect multiple pathways [74, 75]. There are many studies showing differentially expressed miRNAs in solid tumors of breast, pancreas, lung, colon, prostate and stomach [73, 74]. More than 50% of miRNAs are localized in fragile chromosomal sites and in the vicinity of areas of gene amplification and deletion implicating direct involvement of miRNAs in tumoriogenesis [75, 194]. A recent study on The Cancer Genome Atlas (TCGA) pan-cancer data sets showed that oncogenic miRNA families like miR-17, miR-19, miR-

130, miR-93, miR-18, miR-455 and miR-210 co-target tumor suppressor genes and pathways like PI3K, TGF-β and p53 [86]. This suggests an oncogenic role of miRNAs and the ability of miRNAs to simultaneously affect many genes of the same pathway, which makes them unique candidates for targeted therapy [76].

2. Materials and methods

2.1. Ethical considerations

The study was approved by the Regional Ethics Committee for Medical and Health Research Ethics South-East Norway (265-08412c), the Norwegian Data Inspectorate (08/01409-2/MHN) and the Norwegian Directorate of Health (08/7927). Informed written consent was obtained from all patients. The animal work conducted was performed according to protocols approved by the National Animal Research Authority in compliance with the European Convention of the Protection of Vertebrates Used for Scientific Purposes (approval ID 3275 and 3530; http://www.fdu.no/). All surgeries were performed with curative intent.

Table 2: The cohorts analyzed in this thesis; OUH: Oslo University Hospital; TCGA: The Cancer Genome Atlas, PA: periampullary adenocarcinoma, PDAC: Pancreatic ductal adenocarcinoma (http://cancergenome.nih.gov/tcga).

Paper	Cohort	*-omics* data	Number of tumor samples	Brief description
I	OUH	miRNA mRNA	• 85 • 12 • 06	• miRNA and mRNA expression profiling of PA • mRNA expression profiling of normal tissue • miRNA expression profiling of normal tissue
II	OUH	miRNA mRNA	• 20 • 8 • 12	• miRNA expression profiling of paired PDAC tumor/stroma • miRNA expression profiling of normal FFPE tissue • mRNA expression profiling of PDAC
III	OUH TCGA	SNP mRNA	• 60 • 52 • 127 • 120	• SNP6 profiling of PA from the OUH cohort • mRNA expression profiling of PA from the OUH cohort • SNP6 profiling of PDAC from the TCGA cohort • mRNA expression analysis of PDAC from the TCGA cohort
IV	OUH	mRNA	• 3 • 3 • 2	• mRNA expression profiling of xenograft cell lines • mRNA expression profiling of corresponding PDAC • mRNA expression profiling of normal tissue

2.2. The Oslo University Hospital cohort (papers I, II, III and IV)

Fresh frozen tumor tissues, adjacent normal tissue and blood samples were collected from patients treated at Oslo University Hospital from 2008 to 2011. All the patients underwent pancreatoduodenectomy with curative intent. A total of 85 samples were included in paper I. The samples included in papers II, III and IV were subsets of the cohort included in paper I. Of the 85 patients, 41 were males (48%) and 44 (52%) females with a median age of 67 years (range 34-84 years). All the samples were classified based on site of origin, histological subgroups and tumor stage in accordance with pTNM classification of malignant tumors [14]. Four of the samples included in the cohort had M1 metastases, while the rest of the tumors had M0 status. The majority of the samples in the cohort had either moderate (n=54) or poor differentiation (n=30) with the exception of one sample, which was well differentiated. The majority of the patients had localized tumor of stage II with few patients with stages III and IV. The KRAS mutations for codons 12 and 13 were found in 75% of the PAs. Based on pTNM classification the samples were classified as either pancreatobiliary or intestinal histological subtype and located in one of four anatomical sites of origin in the pancreatic head (the pancreatic ducts, the bile duct, the ampulla or the duodenum).

The miRNA and mRNA expression analyses on fresh frozen PA tissues (n=85) were used in paper I. In paper II, miRNA expression profiles were compared between 20 formalin fixed paraffin embedded (FFPE) matched paired carcinoma and stroma samples (n=20). In addition, the mRNA expression data from fresh frozen samples were correlated to the miRNA expression profiles in stroma and carcinoma tissues. In paper III whole genome SNP analysis was performed on PAs (n=60) and analyzed together with corresponding mRNA expression data. In paper IV xenografts and corresponding in vitro xenograft cell lines (n=3) were derived from pancreatic ductal adenocarcinomas. The mRNA expression profile of each xenografts cell line was compared with its primary tumors.

2.3. The Cancer Genome Atlas cohort (paper III)

SNP data analysis was performed on pancreatic cancer samples (n=127) from The Cancer Genome Atlas (TCGA) cohort. The majority of the samples were PDACs (n=111) with the

remaining (n=16) being variant subtypes of PDACs. In the TCGA cohort 75 (59%) were males and 52 (41%) were females with a median age of 65 years (range 35-88 years). The majority of the samples (57%) were of stage II.

2.4. Total RNA extraction

Total RNA was isolated from whole sections of fresh frozen tumor and adjacent normal tissues from patients with pancreatitis according to the manufacturer's instructions (mirVana miRNA Isolation Kit Ambion/Life Technologies) for papers I, II, III and IV. The RNA concentration was measured using a Nanodrop ND-1000 spectrophotometer, and the quality assessed on a Bioanalyzer 2100 (Agilent). RNA was extracted from formalin fixed paraffin embedded (FFPE) tissue cylinders using the miRNeasy FFPE kit (Qiagen) as described by the manufacturer (paper II). For the FFPE tissue, areas dominated with tumor cells and desmoplastic stromal cells were marked and cylinders were drawn. Separate slides were made and examined to ensure that there was no mixing of tumor and stromal cells. The RIN values of the total RNA were above 5.0 for 80% of the samples.

2.5. DNA extraction

DNA was extracted from tumor tissue using the Maxwell Tissue DNA kit on the Maxwell 16 Instrument (Promega) (paper III). Briefly, 5 x 20 µm sections were homogenized in 300 µl lysis buffer and added to the cartridge. The method is based on purification using paramagnetic particles as a mobile solid phase for capturing, washing and elution of genomic DNA. Elution volume used was 200 µl. DNA was extracted from 6 ml EDTA blood using the QiAamp DNA Blood BioRobot MDxKit on the BioRobot MDx (Qiagen). The method is based on lysis of the sample using protease, followed by binding of the genomic DNA to a silica-based membrane and washing and elution in 200 µl buffer AE. DNA from normal tissue was extracted using column-based technology (Qiagen, and Aros Applied Biotechnology AS, Denmark).

2.6. The principle of microarray technology

Hybridization microarray is based on the principle of the hybridization of mRNA/miRNA to immobilized oligonucleotide probes representing the sequences of interest. The

samples are fluorescently labeled and hybridized on microarray slides. The hybridization and washing conditions are optimized to obtain a maximum specificity and sensitivity. The arrays are scanned after several washing steps to remove unspecific hybridization and the fluorescence is quantified [105]. The expression level of each probe is calculated after pre-processing of the data, which includes background correction, data transformation and normalization of the raw data. Further, downstream analysis is done to identify important biological processes (Figure 7).

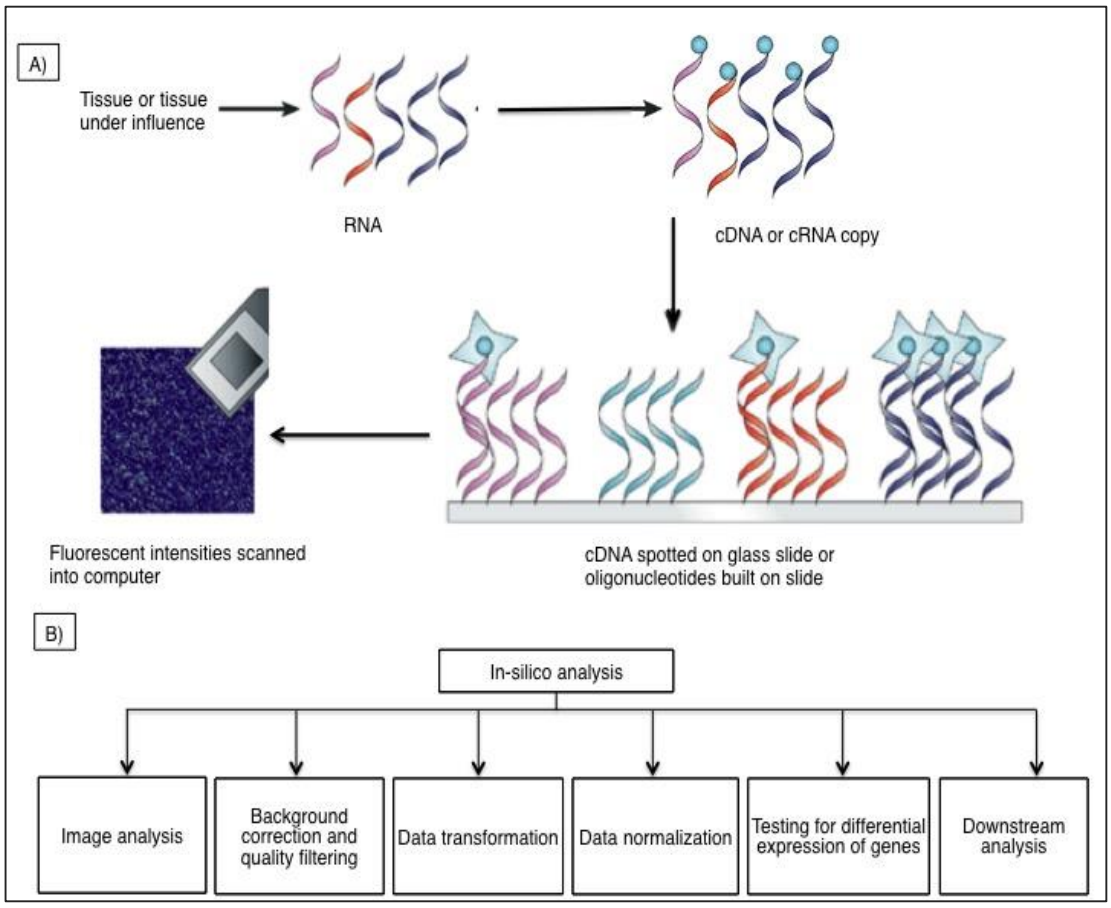


Figure 7: a) Schematic representation of microarray analysis of the samples. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Drug Discovery] [105], copyright (2002). b) Shows the steps of in-silico data processing.

2.7. mRNA expression profiling

The mRNA expression profiling was done using total RNA from fresh frozen tissue with 100 ng total RNA as starting point. The total RNA was converted to cDNA, amplified, labeled with Cy-3 and hybridized to SurePrint G3 Human GE 8x60K microarrays containing 42,545 probes. Hybridization signals were extracted using Feature Extraction software

10.7.3.1 (Agilent Technologies, Santa Clara, CA, USA). The mRNA microarray data was background corrected and quantile normalized. Filtering of controls and low expressed probes was performed by calculating the 95th percentile of the negative control probes on each array, and for probes that were at least 10% brighter than the negative controls on at least 50% of the arrays. The mRNA expression data for fresh frozen PA samples are accessible in Gene Expression Omnibus (GEO) database with accession number GSE60979. The mRNA expression data for the three xenograft cell lines generated from PA samples is submitted to GEO with accession number GSE58561.

2.8. miRNA expression profiling

miRNA is incorporated in lipoprotein complexes and thereby protected against degradation by RNases and chemicals. miRNA isolated from FFPE tissue is normally of sufficient quality to be analyzed by different microarray and sequencing methods. The miRNA expression profiling using 150 ng total RNA from fresh frozen PA tissues was used in paper I. The RNA quality was determined using a Bioanalyzer 2100 (Agilent) and the input was based on quantification using Nanodrop ND-1000. Total RNA was dephosphorylated and labeled by Cyanine-3-pCp before hybridization to SurePrint G3 Human v16 miRNA 8x60K microarrays containing probes for 1368 miRNAs. The arrays were washed and scanned on an Agilent DNA Microarray Scanner and hybridization signals were extracted using Feature Extraction software 10.7.3.1 (Agilent Technologies, Santa Clara, CA, USA). The miRNA microarray data was normalized using Robust Multiarray Average approach (RMA) and undetected probes were flagged and filtered if not detected in 50% of the array replicates. The miRNA data was deposited in GEO with accession number GSE60978.

The miRNA expression profiling published in paper II was performed by Exiqon Services, Denmark on FFPE tissue from paired tumor/stroma and normal samples. The microarray system used by Exiqon is based on the miRCURY LNA^{TM} technology. The primers and probes contain Locked Nucleic Acid (LNA) that are RNA analogs constraining their conformation and thereby making them more stable and increase the sensitivity and specificity of their hybridization to microarrays. The paired tumor/stroma RNA samples and reference RNA sample were labeled with Hy3 and Hy5 fluorescent labels,

respectively. The samples were mixed pairwise and hybridized to the miRCURY LNA™ microRNA Array 7th (Exiqon, Denmark) which captures 2042 probes, targeting miRNAs for human tissues. The miRCURY LNATM microRNA array slides were scanned using the Agilent G2565BA Microarray Scanner System (Agilent Technologies, Inc., USA), and the image analysis was carried out using ImaGene® 9 (miRCURY LNA™ microRNA Array Analysis Software, Exiqon, Denmark). The miRNA expression data was background corrected and quantile normalized. The miRNA expression data is accessible through GEO accession number GSE71533.

2.9. Genotyping using SNP arrays

Copy number alterations of periampullary adenocarcinomas in Paper III were analyzed using genome-wide human SNP 6.0 array (Affymetrix, Santa Clara CA, USA). The SNP 6.0 array has 1.8 million genetic markers, including 906,600 SNPs and 946,000 copy number probes. Briefly, total genomic DNA (500 ng) was fractioned with restriction enzymes Nsp I and Sty I, followed by ligation of adapters to the fragments. All the adaptors are ligated to fragments irrespective of size to the cohesive 4 bp overhangs. In the next step, adaptor ligated DNA fragments were amplified using a generic primer that recognize the adaptor sequence, and fragments are amplified from 200 to 1,100 bp size range. PCR amplification products for each restriction enzyme digest were purified using polystyrene beads. The amplified DNA was then fragmented, labeled and hybridized to the microarray slides. After hybridization, washing was done to remove the background signals. The fluorescence signals were recorded on the GeneChip scanner 3000 7G and allele frequencies were determined based on the fluorescence intensity.

2.10. *KRAS* mutational analysis

Tumor DNA was screened for the presence of seven *KRAS* mutations in codons 12 and 13 (G12R, G12S, G12C, G12D, G12A, G12V, G13D) using the Wobble enhanced amplification refractory mutation system (WE-ARMS) reported in paper I. The WE-ARMS method is an in-house method developed using TaqMan MGB® probes from Life Technologies. It is based on the requirement of Taq DNA polymerases for a perfect match between the 3'-end of a PCR primer and its target gene sequence to perform primer-extension. To

achieve adequate discrimination between the different KRAS mutations a mismatch was introduced in the third position from the 3'-end (wobble) of the primer, thereby increasing the specificity [133, 134].

2.11. Statistics and bioinformatics

2.11.1. Preprocessing of microarray data

Normalization of raw microarray data is essential to reduce the effect of technical variation in the data, in order to look at biological variations [106]. First, the raw data were visualized as scatter and box plots to find the distribution of the samples and outliers. Probe signals were background corrected and multiple probes representing a single gene in mRNA expression profiling data were reduced to a single value by averaging the expression of the probes. The mRNA expression data in papers I and IV were quantile normalized and low expressed, undetected and control probes were removed from the analysis. The miRNA data in papers I and II were normalized using Robust Multiarray Average approach (RMA) and quantile normalization, respectively. In quantile normalization the probe intensities are adjusted to obtain identical intensity distribution in all arrays. This method is widely adopted for normalization and is motivated by quantile-quantile plots (qqplot). The qqplot compares the distribution of samples and if the quantiles of the samples have similar distribution they are represented as straight diagonal lines. This suggests that by transforming the quantiles of arrays, the dataset can have similar distribution across the arrays [109]. In the RMA normalization method, the signals are first background corrected using the normexp convolution model by fitting a normal and exponential convolution model to the vector of observed intensities. The normal convolution represents background intensities and exponential convolution represents signal intensities. After background correction, these signals are log2 transformed and finally the estimate of the miRNA gene signal is obtained by quantile normalizing the data [107, 108].

2.11.2. Identifying differentially expressed mRNAs and miRNAs

In microarray data, identification of differentially expressed mRNAs and miRNAs between the test and the reference sample set can be carried out based on the distribution of samples and sample size. Parametric tests are usually performed if the data is normally distributed, while non-parametric tests are used if the sample size is small and data do not follow a normal distribution [111]. However, parametric tests can perform well with skewed and non-normal distributions as well [110].

2.11.3. Parametric tests

The moderated t-test (modT-test) was performed to find the differentially expressed mRNA/miRNA in papers I and II. The test statistics for the modT-test is d/(s + s0) where d is a difference between two group means, s is a pooled standard deviation estimate, and s0 is a small constant. The small constant is added to avoid divisions by very small variance estimates, when performing a large number of tests on high throughput data chances are high that a few of the variance estimates will be very small [112, 113]. In paper I, the modT-test was used to identify the differentially expressed mRNAs and miRNAs between different morphological and anatomical subtypes of PAs and between PAs and normal samples. In paper II, the modT-test was used to find differentially expressed miRNAs between carcinoma versus normal, carcinoma versus stroma, and stroma versus normal tissues. A one-way analysis of variance (ANOVA) test was used in paper II to test the null hypothesis that carcinoma, stroma and normal samples are drawn from populations with the same mean value. If the P-value was less than 0.05, the null hypothesis was rejected.

2.11.4. Non-parametric tests

The Mann-Whitney U test was performed to identify the differentially expressed miRNAs between carcinoma and stroma samples (paper II). The Mann-Whitney U test is a non-parametric test that sorts and ranks the outcome variables in ascending order, and test whether the medians in two groups are different based on the ranks of the observations in the combined group [115]. The non-parametric Kruskal-Wallis ANOVA was also carried out to test whether the mean ranks of carcinoma, stroma and normal samples were the same and originate from the same distribution. The null hypothesis of the Kruskal-Wallis

ANOVA is that the mean ranks of the groups are the same and originate from the same distribution [114].

2.11.5. Multiple testing corrections

In microarray and other high throughput data analyses, several thousands of features (genes/miRNAs/SNPs) are simultaneously tested across different conditions, and each feature is considered independent from each other [117]. In such analyses, the probability of getting false positive results increase, and the features that are found to be statistically different between conditions could be by mere chance only and might not be true positives [117]. Multiple testing corrections adjust P-values to correct for the increased probability of false positives when multiple tests are performed in parallel [116]. The false discovery rate (FDR) adjustment or Benjamin and Hochberg's correction was used in all the papers where multiple statistical tests were applied. In FDR correction, the P-values are first sorted and ranked. Then each value is multiplied by the number of tests (N) and divided by its rank to give adjusted P-values [116].

2.11.6. Permutation tests

Permutation tests or randomization tests such as the Hypergeometric test, the Fisher's exact test and the Chi square test were used in papers I, II and III to check the significance of the results. The Fisher's exact test and the Hypergeometric test were performed in paper I to determine if the clustering of samples into different groups was statistically significant. The Fisher's exact test was applied in paper II to study the clustering of tumor and stroma samples in the clusters, and to study if the genes common between two comparisons (i.e. stroma versus normal and tumor versus normal) were statistically significant. The Chi square test was carried out in paper III to test significant difference in frequency of aberration in two different groups under comparison. The Fisher's exact test is a non-parametric statistical test used in the analysis of contingency tables. It works well on small datasets to test whether the proportion of one variable are different depending on the value of other variables [118]. A Chi squared test is a parametric statistical test in which the sampling distribution of the test statistic is Chi squared. It is used to find the significant association or goodness of fit between two variables. The Chi square is the

sum of the squared difference between the observed (o) and the expected (e) data (or the deviation, d), divided by the expected data in all possible categories [119, 120]. The Hypergeometric test is a probability based statistical test and use the hypergeometric distribution to calculate the significance of an association [121].

2.11.7. Clustering

Hierarchical clustering is one of the most popular unsupervised methods for gene expression analysis. Typically, the purpose of clustering analysis in cancer research is to identify distinct gene expression patterns in order to identify new biological classes with clinical implications [122]. The meaning of "unsupervised" in clustering analysis is that the classes are unknown *a priori* and need to be discovered from the data. The data from gene expression microarrays can be represented as an expression matrix where the rows represent the expression of mRNAs/miRNAs and the columns represent the samples. In hierarchical clustering, genes with similar expression patterns are grouped together and are connected by a series of branches called dendrogram. In order to apply hierarchical clustering, the user must first define a similarity measure (i.e. how to calculate the distance between two entities to be clustered) and a linkage method (i.e. how to calculate the distance between two clusters). Common similarity measures are Pearson's and Spearman's correlation, as well as 1-d where d is the Euclidean distance. Common linkage methods are complete linkage (the maximum distance between a member of the first cluster and a member of the second cluster), single linkage (the minimum distance between a member of the first cluster and a member of the second cluster) and average linkage (the average distance between members of the two clusters). The objective of the clustering algorithm is to compute a dendrogram where the node represents the similarity between the two objects and the branch reflects the degree of similarity between the two. A node is created when joining the two genes with the highest similarity score, and a gene expression profile is computed for the node by averaging observations for the joined elements. The similarity matrix (consisting of all the similarity values between entities) is updated with this new node replacing the two joined elements, and the process is repeated n-1 times until only a single element remains [122]. Hierarchical clustering was used in papers I-IV.

2.11.8. Sparse Principal Component Analysis

Sparse Principal Component Analysis (SPCA) is an extension of the classical Principal Component Analysis (PCA) method introduced by Jolliffe et al. in 1986. PCA reduces the dimensionality of large data sets by finding linear combination of all the variables (expression data) and transform them into new variables called principal components (PCs), which can explain the variance in the samples [123]. The drawback of PCA is that a linear combination of the entire input variable (expression of all the probes) is difficult to interpret in case of high dimensional data [124]. SPCA overcomes this problem by applying sparsity constraint on the input data. SPCA is more concise than PCA and gives a more interpretable set of PCs with sparse loadings for large genomic data by applying Penalized Matrix Decomposition (PMD) in which lasso (elastic) penalty is applied [125]. It thus reduces the dimensions of large data sets and therefore represents a useful way of exploring the naturally arising sample classes based on expression profiles. SPCA has apparent advantages compared to PCA, such as better interpretability, and is generally thought to be computationally much more extensive. SPCA was used to analyze carcinoma, stroma and normal sample distributions in paper II [125].

2.11.9. Correlation analysis

Correlation analysis is used to quantify the association between two continuous variables. The correlation coefficient is a measure of linear association between two variables and it quantifies the direction and strength of the association. The value of the correlation coefficient ranges between -1 and 1 and it quantifies the direction and strength of the linear association. The value of 1 indicates that two variables have positive association, meaning high level of one variable are associated with high level of the other variable. A value of -1 indicates that the variables have negative association meaning, that the high level of one variable are associated with low levels of other variable. A value of 0 indicates no association. Correlation analysis was carried out for miRNA and mRNA data in paper II and for mRNA and copy number data in paper III. The Pearson's correlation coefficient was calculated in both papers, which is the covariance of the two variables divided by the product of their standard deviations.

2.11.10. miRNA target prediction

miRNA target site predictions were carried out using web-based tools (miRanda and Targetscan) in paper II. The miRanda tool is based on identifying potential binding sites by searching for high complimentarity regions on the 3'-UTRs. The scoring matrix used by the algorithm is built so that complimentary bases at the 5'-end of the miRNAs are rewarded more than those at the 3'-end. Therefore, the tool ensures the perfect or almost perfect match at the seed region of the miRNAs. Thermodynamic and cross species conservation features were used in the algorithm for predicting the miRNA targets in miRanda [126, 127]. Targetscan predicts the biological targets of miRNAs by searching for the presence of 8-mer, 7-mer and 6-mer sites of miRNAs based on the criteria that miRNAs complementarity is limited to six nucleotide seeds, followed by an additional 3'-match of an adenosine anchor at position 1. Therefore, in the Targetscan tool perfect complimentarity is not the criteria but conservation of target site and weighted score are used for prediction of target site [128, 129].

2.11.11. Databases of experimentally validated miRNA targets

miRWalk was used for finding the experimentally validated miRNA target sites linked with mRNA and deregulated pathways in PA (paper I). The validated target module in miRWalk archives experimentally validated miRNA interaction information associated with genes, pathways, organs, disease, cell lines, OMIM (Online Mendelian Inheritance in Man) disorders and literature of miRNAs [130]. The miRtarbase4 is a curated database of the miRNA target interactions collected by text mining of pertinent literature. The miRtarbase4 database was used in paper II for finding the experimentally validated miRNA target sites that are anti-correlated with the mRNA expression. The miRNA targets reported in these databases are validated by experiments like reporter assays, western blot, microarray and next generation sequencing experiments [131, 132].

2.11.12. Copy number analysis

Copy number aberration profiles from the OUH and the TCGA cohorts were analyzed (paper III). Segmental copy number information was derived for each sample using the Battenberg pipeline (https://github.com/cancerit/cgpBattenberg/) [140]. Briefly, the

Battenberg pipeline phases heterozygous SNPs using the 1000 genomes genotypes as a reference panel using impute2 tool [140]. In the phasing step, the reference panel is combined with the samples collected from a genetically similar population. The reference samples are genotyped at a subset of these sites and predict unobserved genotypes in the study sample [140, 141]. This step corrects phasing errors in regions with copy number changes through segmentation. Segmentation of copy number data and assessment of breakpoints were done using the allele specific piecewise constant fitting (ASPCF) algorithm [142]. After segmentation of the resulting B-allele frequency (BAF) values, t-tests were performed on the BAFs of each copy number segment to determine if they correspond to the value resulting from a fully clonal copy number change. If not, the copy number segment was represented as a mixture of two different copy number states, with the fraction of cells bearing each copy number state estimated from the average BAF of the heterozygous SNPs in that segment. The Battenberg pipeline has an advantage over the in-house algorithm ASCAT [99], as it is more sensitive particularly for samples with low cellularity like pancreatic cancer. The Battenberg pipeline models allele-specific copy number as a function of the SNP data, the tumor ploidy and the aberrant cell fraction and subsequently select the solution that is closest to non negative integer copies at all assayed loci in the genome. Its been used in paper III to estimate tumor cell fraction, tumor ploidy and copy numbers profiles for PAs.

The genome instability index (GII) was calculated for both the cohorts. It is measured as the fraction of aberrant probes throughout the genome above or below the ploidy. The GII score was used for defining the percentage of genomes aberrated in the two cohorts and it was correlated with the ploidy of the samples. The frequency plots were generated based on aberration scores for all the samples per segment. For each sample an aberration score was calculated per segment. The aberration score is 1 if the total copy number per segment is larger than the average ploidy corresponding to copy number gain. It is -1 if smaller than the average ploidy corresponding to deletion, and zero otherwise. The driver genes on chromosomal loci frequently gained and deleted in PAs were found by mapping the most frequently occurring chromosomal aberrations (threshold < 25%) on known oncogenes and tumor suppressors [151-153].

2.11.13. Pathway analysis

Pathway analysis was performed to map the list of genes enriched in known pathways and the outcome is reported in papers I, II and III. It was analysed using the GSEA tool from the Broad Institute to identify deregulated pathways (paper I). GSEA is a statistical tool for microarray data analysis based on a modified Kolmogorov–Smirnov statistics. It ranks all the genes based on the expression ratio between the normal and the tumor samples and determines whether the high-ranking genes are enriched with genes of specified pathways [135]. The gene sets were selected from four different sources: BioCarta, KEGG, Pathway Interaction Database and Reactome from the MSigDb database version v4.0.

The pathway analysis in papers II and III were performed using the web based Gene set analysis toolkit (WebGestalt) http://bioinfo.vanderbilt.edu/webgestalt/ [136, 137]. WebGestalt uses the Hypergeometric test for enrichment evaluation analysis at $P < 0.05$ after Benjamin and Hochberg's correction. The minimum number of genes required for a pathway to be considered significant can be preset. The WebGestalt analysis results give pathway enriched genes, the number of genes enriched and raw P-value (rawP) from the Hypergeometric test, the Benjamin and Hochberg's corrected P-value (adjP), the number of reference genes in the category (C score), the number of genes in the gene set and in the category (O score), the expected number in the category (E score) and the ratio of enrichment (R score). In paper II, the Gene set enrichment analysis was carried out to identify genes significantly associated with differentially expressed miRNAs in carcinoma and stroma. In paper III the Gene set enrichment analysis was used to identify biological pathways in PAs in the two cohorts.

2.11.14. Survival analysis

Survival analysis is a method to model time until the occurrence of event of interest such as death or relapse of disease. The survival function gives the probability of surviving up to time (t), and the hazard function gives the probability that the event will occur, per unit time, given that the individual has survived up to the specified time. Survival analysis is based on three variables, 1) the dependent variable is the time until the occurrence of a well-defined event. 2) The censored observations for which the event of interest has

not occurred at the time of data analysis, and 3) predictor variables whose effect on the time of occurrence need to be assessed [139]. The Kaplan-Meier estimator was used to estimate and graph survival probabilities over time, and the Log-rank test was performed to compare survival between two or more groups over time (papers I and III).

To identify potential prognostic markers (i.e. miRNAs and mRNAs) for pancreatobiliary and intestinal adenocarcinomas in paper I, Cox proportional hazard (Cox PH) analysis was performed. It allows testing for differences in survival times of two or more groups of interest, while allowing adjustment of covariates of interest. The Cox PH model provides useful and easy to interpret information regarding the relation between the hazard function and predictors [138]. The results were confirmed by the Log-rank test ($P < 0.05$). The expression for each sample was designated as high if the expression was higher than the 75^{th} percentile expression value within each histopathological subgroup, otherwise as low. Overall survival (OS) time was calculated from date of surgery to time of death. OS data were obtained from the National Population Registry in Norway. Recurrence free survival (RFS) time was calculated from date of surgery to date of recurrence of disease. Recurrence was defined as radiological evidence of intra-abdominal soft tissue around the surgical site or distant metastases.

The outline of the workflow of the thesis is represented graphically in Figure 8.

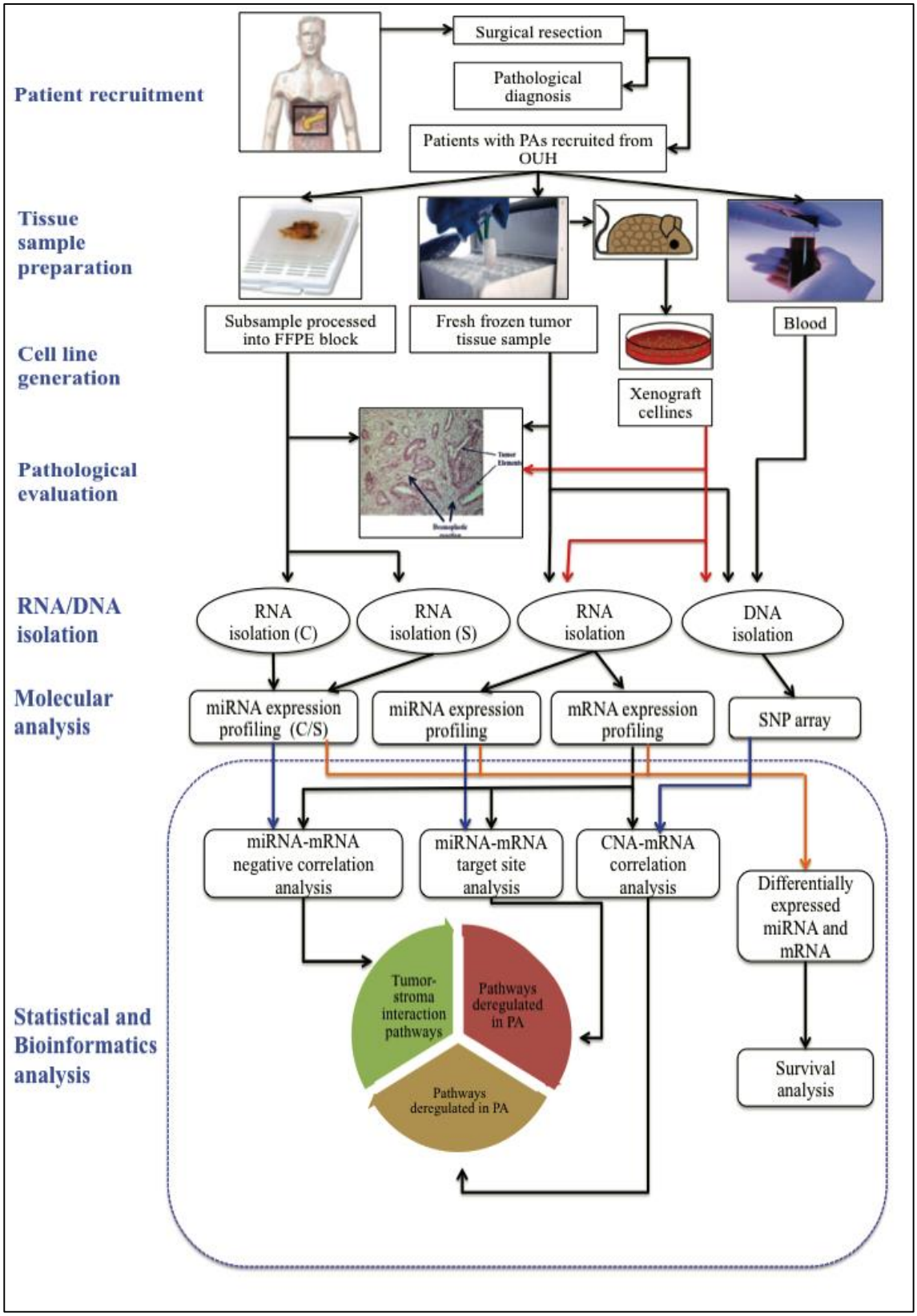


Figure 8: The figure shows the workflow of the analysis performed in the thesis.

3. Results in brief

3.1. Paper I

Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas

Sandhu V., Bowitz Lothe I.M., Labori K.J., Lingjærde O.C., Buanes T., Dalsgaard A.M., Skrede M.L., Hamfjord J., Haaland T., Eide T.J., Børresen-Dale A.L., Ikdahl T., Kure E.H.

Molecular Oncology 9, 758-771, 2015.

This paper reports the study undertaken to characterize 85 periampullary adenocarcinomas (PAs) by mRNA and miRNA expression profiling. PAs originating in the pancreatic head can be anatomically divided into four locations; the pancreatic ducts, the bile duct, the ampulla and the duodenum, which fall into two histological subtypes that is pancreatobiliary or intestinal. The molecular basis for the differences between the PAs originating at different sites and with different morphologies was studied. The prognosis of the PAs depends on both anatomical site of origin and histological subtype. The molecular profiles (miRNAs and mRNAs) of PAs was compared using unsupervised hierarchical clustering, and checked if the tumors cluster based on site of origin or by morphological subtypes. We found that adenocarcinomas originating from the pancreatic ducts, the bile duct and the ampulla with pancreatobiliary morphology clustered together, while adenocarcinomas from the ampulla and the duodenum with intestinal morphology made a separate cluster at both the miRNA and the mRNA level. This showed that samples were more distinct at the level of morphology than at site of origin. The lists of mRNAs and miRNAs differentially expressed between ampulla of intestinal and pancreatobiliary subtypes (the only site having both the morphologies) also supported the clusterings of PAs based on morphology. Further, the miRNAs and mRNAs differentially expressed between the two histopathological subtypes were determined. Ten miRNA families; six (miR-17, miR-196, miR-192, miR-194, miR-19 and miR-548) were downregulated while four (miR-154, miR-99, miR-329 and miR-199) were upregulated. The mRNAs and miRNAs associated with improved overall and recurrence free survival for the two histopathological subtypes were identified. Specifically, for the

pancreatobiliary subtype the genes *ATM*, *PTEN*, *RB1* and miRNAs miR-592 and miR-497, and for the intestinal subtype *PDK1*, *PIK3R2*, *G6PC* genes and miRNAs miR-127-3p, miR-377* with prognostic relevance were identified. The pathways differentially enriched in the two histopathological subtypes were identified and the mRNAs/miRNAs with prognostic relevance associated with pathways were determined.

3.2. Paper II

Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma

Sandhu V., Bowitz Lothe I.M., Labori K.J., Skrede M.L., Hamfjord J., Dalsgaard A.M., Buanes T., Dube G., Kale M.M., Sawant S., Kulkarni-Kale U., Børresen-Dale A.L., Lingjærde O.C., Kure E.H.

Molecular Oncology, November 2015, in press.

The aim of this paper was to investigate the role of miRNAs in stroma of PAs and the pathways involved in facilitating tumor-stroma interactions. The differences in miRNA expressions between carcinoma cells and its associated desmoplastic stroma in pancreatobiliary subtype of periampullary adenocarcinomas (PAs) were studied. The pancreatobiliary subtype is more frequent and more aggressive than the intestinal subtype. The stroma of these tumors is associated with aggressiveness of disease and promotion of cancer cell growth, invasion, metastatic progression, de-differentiation and resistance to therapy. The miRNA expression profiling of 20 FFPE paired carcinoma and stroma tissues and eight adjacent normal tissues was performed. The miRNA expression profiles between carcinoma, stromal and adjacent normal tissues were compared. We found 43 miRNAs differentially expressed between carcinoma and stromal tissues of which 26 were upregulated and 17 were downregulated in stroma as compared to carcinoma cells. Three miRNA families (miR-17, miR-15 and miR-155) were differentially expressed between carcinoma and stromal tissues at adjusted $P < 0.05$. The levels of expression of seven miRNAs followed a trend where the expression in stroma was higher than in the carcinoma and the normal tissues. The levels of expression of the seven miRNAs miR-17, miR-20a, miR-20b, miR-223, miR-10b, miR-2964a and miR-342 were higher, while the expression of miR-519e was lower in stroma as compared to the carcinoma and the normal tissues. The differentially expressed mRNAs between tumor and normal samples and differentially expressed miRNAs between carcinoma and stromal cells were used for studying the negative correlation, that is associations where upregulation of mRNAs leads to downregulation of miRNAs and vice versa. The miRNAs anti-correlated and potential mRNA targets were found using miRNA target site

prediction tools. Gene enrichment analysis was used to identify the pathways that facilitate tumor-stroma interactions involved in pancreatic cancer. The pathways regulating tumor-stroma interactions was found such as, ECM interaction remodelling, EMT, focal adhesion pathway, TGF-β, MAPK signaling, axon guidance and endocytosis pathway. The study adds new knowledge about miRNAs involvment in tumor-stroma interaction and associated pathways.

3.3. Paper III

The genomic landscape of pancreatic and periampullary adenocarcinomas

Sandhu V., Wedge D.C., Bowitz Lothe I.M, Labori K.J., Dentro S., Buanes T., Skrede M.L., Dalsgaard A.M., Lingjærde O.C., Børresen-Dale A.L., Ikdahl T., Van Loo P., Nord S., Kure E.H.

Manuscript

The copy number aberration patterns in periampullary adenocarcinomas (PAs) from Oslo University Hospital (OUH) and The Cancer Genome Atlas (TCGA) cohorts were analyzed. A total of 60 and 127 samples from the OUH and TCGA cohorts were genotyped by using Affymetrix SNP 6.0 arrays. Chromosomal loci 1q, 3q, 7p, 8q, 13q, 18p, 19q were frequently gained while chromosomal loci 1p, 3p, 6, 8p, 9p, 17p, 18q, 22q were frequently lost in PAs in both cohorts. The copy number losses were higher (75% of samples) than copy number gains (< 30% of samples). The most frequent events were deletions of 17p and 18q in both cohorts with a co-occurrence in 60% of the samples in both cohorts at $P < 0.01$. The genes located in 17p and 18q are co-involved in deregulated pathways like the cell cycle, apoptosis, p53 and wnt signaling.

The mRNA expression data was used to study the correlation between copy number gains and losses of chromosomes with up- or downregulation of genes *in-cis*. Chromosomal amplifications of 978 and 3566 genes and chromosomal deletions of 1060 and 4953 genes were associated with gene expression in the OUH and TCGA cohorts, respectively. Specifically, we observed 19q12 amplification associated with overexpression of *CCNE1* in both the OUH and TCGA cohort. The driver genes on frequently gained and deleted chromosomal loci were also found to be significantly associated with gene expression data in both cohorts. The genes located in frequently gained and deleted chromosomal loci were used for pathway analysis and similar pathways were found enriched in both cohorts.

The chromosomal aberration differences in PAs based on the site of origin and morphology were compared. The PAs are more distinct at the level of morphology and found to have subtype specific aberrations. The chromosomal aberrations specific to the intestinal subtype were gains of 13q and 3q and deletion of 5q. Subtyping using the PAM50 gene signature suggested clustering of samples into two morphological subtypes with the pancreatobiliary subtype further divided into basal and classical. The basal subtype is highly proliferative, with poor degree of differentiation and worst prognosis as compared to the classical and intestinal subtypes.

The chromosomal loci with focal gains and deletions were tested for prognostic relevance. Gains of 18p11.22 and 19q13.2 and overexpression of genes in these loci were associated with both worse overall survival and relapse free survival.

3.4. Paper IV

Generation and Characterization of Novel Pancreatic Adenocarcinoma Xenograft Models and Corresponding Primary Cell Lines

Wennerstrom, A.B., Lothe, I.M., **Sandhu, V.,** Kure, E.H., Myklebost, O., Munthe, E.

PLoS One 9, e103873.

The xenografts were generated from pancreatic ductal adenocarcinomas; seven with pancreatobiliary and one with intestinal subtype. The morphology and differentiation grade were similar in xenografts as compared to their corresponding original tumors. Extracted cells from four of the xenografts successfully generated novel xenograft cell lines. The transcriptome profiles of the primary tumors and corresponding xenograft cell lines were compared by unsupervised hierarchical clustering analysis. The xenograft cell lines clustered with the corresponding tumor they were derived from.

The xenografts and original tumors were tested for selected biomarkers (CD36, EpCAM, CK7, vimentin, E-cadherin, Snail/Slug). The immunohistochemistry markers CK7 and CD36 showed strong enrichment of human specific antibodies in xenografts. The colony forming and tumourigenic capacity of the cell lines were tested. Injecting single cell suspensions from the established *in vitro* cultures formed palpable tumors, which showed the tumourigenic capacities of xenograft cell lines. All cell lines could form colonies except the one with intestinal morphology. The stem cell markers were characterized in all the four cell lines, and all of the cell lines had heterogeneous morphologies and expressed known pancreatic stem cell markers (CD24, CD44, CD326). The generated cell lines can be useful preclinical model for studying pancreatic cancer.

4. Discussion

The discussion is divided into two sections. In the first section methodological considerations focusing on sampling of tissues, challenges with *–omics* analysis and statistical and bioinformatical considerations will be discussed. In the second section the focus is on clinical and biological implications. A more detailed discussion of the results is presented in the separate papers.

4.1. Methodological considerations

4.1.1. Pre-analytic quantitative characterization

The high level of endogenous RNases in pancreatic tissue makes the biobanking challenging. In order to maintain high quality of the biological specimens; the fresh tissue samples from patients who receive surgery were collected by the pathologist, and microscopic and macroscopic work followed shortely thereafter. A piece of tumor tissue was snap frozen in liquid nitrogen and stored at -80°C for further processing. Later, the pathologist reevaluated the fresh frozen samples to determine tissue heterogeneity, percentage of tumor cells, stromal cells, normal cells, and content of necrotic and inflammatory cells, to ensure that the samples selected for molecular analysis were of good quality. RNA was extracted from fresh frozen tissue and the RNA Integrity Number (RIN) was measured for all the samples to ensure that RNA was intact and not degraded. The quality of RNA suggested that this procedure optimized the quality of the tissues for data analysis.

Low cellularity, abundant desmoplastic stroma and cellular heterogeneity between and within pancreatic tumors complicate molecular analysis [60, 156]. Thus, determining the non-aberrant cell fraction in each sample is important in order to estimate signals from tumorigenic and non-tumorigenic cells. The approaches used to overcome some of the issues are discussed in the next section.

4.1.2. Approaches for studying periampullary adenocarcinomas

Fresh frozen tumor samples were used for RNA and DNA analyses, respectively (papers I and III). FFPE tissues were used for miRNA expression profiling of carcinoma and stromal cells of PA samples (paper II). miRNAs are highly stable while mRNA easily degrade in FFPE tissues [65-66]. mRNA expression data from fresh frozen tissue (i.e., with both carcinoma and stromal cells) was therefore used for the integrative miRNA-mRNA analysis (paper II). The distinct sampling of carcinoma and stromal cells was possible only in FFPE tissues, and measures were taken to avoid the contamination between carcinoma and stromal cells.

The DNA was extracted from fresh frozen tissues (paper III). The in-house algorithm ASCAT was initially used for analyzing CNAs, but it did not perform well due to too low cellularity of samples since it needs > 20% aberrations to make the CNA calls. The Battenberg pipeline is advantageous over the ASCAT algorithm since it is more sensitive for tumors with low cellularity and also identifies sub-clonal cells in tumors [140]. Thus, we enriched the tumor cell content and utilized the Battenberg pipeline.

The xenografts cell lines were used in papers III and IV. The mRNA expression between cell line and its corresponding tumor was compared, which showed that xenografts cell lines have gene expression similar to the corresponding tumor (paper IV). However, the DNA analysis of one of the xenografts cell lines showed changes in average ploidy as compared to the tumor (paper III). This change in CNA profile can be due to phenotypic and genotypic drift as the cell lines are grown in artificial medium. Subpopulation of clones in cell lines may grow more rapidly than other cells and can have clonal selection which may not represent the original tumor cell population [158]. Virtanen's approach of filtering was used for clustering the mRNA data of fresh frozen samples and cell lines [159]. The differentially expressed genes between cell lines and tumors were filtered, as these genes could result from contaminating normal and stromal cells in tumor tissues or proliferation associated mRNAs [159]. This can explain the observed differences in RNA expression and DNA copy number data.

4.1.3. Challenges in miRNA and mRNA expression profiling

There are many challenges when analyzing miRNA and mRNA expression profiling data. Determining the quality of the array data is important such as high variance of probes, low signal levels due to poor hybridization or a high level of RNA degradation. The quality control (QC) plots were generated after the arrays were run to detect arrays with poor quality, and the samples with poor QC plots were discarded from the analysis. The common and powerful source of variation in expression profiling is the batch effect, and this technical source of variation can occur within the samples based on the sampling date, labeling date, date of hybridization etc. Therefore subgroup of measurements could be affected by laboratory condition or technical errors that may lead to incorrect conclusions [59]. To circumvent batch effect errors, PCA plots were generated and checked for any clustering of samples due to non-biological variations (papers I and II). The samples were hybridized on the same day and in the same batch to reduce technical variations in the data (paper II).

After running the arrays, the pre-processing of expression data is a very important step to remove any biases caused by technological errors [146]. The pre-processing step involves data exploration, background adjustment, normalization, summarization and quality assessment [145]. The miRNA and mRNA expression data used in papers I, II and IV were background corrected to remove non-specific hybridization of probes and further normalized. Normalization of data is necessary in order to remove systematic biases in the measured expression levels and biases from the use of different fluorescent dyes. The mRNA expression data was quantile normalized to ensure that intensities have a comparable distribution across the arrays and across the channels. The miRNA expression data was RMA normalized which background corrected and quantile normalized the data. The expression data was summarized by averaging the probe intensities for multiple probes, since the expression arrays have multiple probes for one mRNA. Thus, summarization of data makes downstream analysis easier.

4.1.4. Challenges in SNP array analysis

The copy number data is composed of 1.8 million genetic markers which makes the analysis of data cumbersome and computationally intensive. In order to run the samples efficently, samples were run in batches on multiple computers and analysed. Besides being computationally intensive, other inherent challenges in SNP array data analysis are: 1) distinguishing cancer specific lesions or somatically acquired aberrations from inherited copy number variations 2) finding small somatic deletions and 3) estimating the frequency of LOH [148]. The inclusion of a paired germline DNA from blood helped in separating CNAs from CNVs. The two main objectives of genomic analysis that need to be addressed are 1) to identify the driver events that contribute to proliferation and carcinogenesis and 2) passenger events that are acquired during cancer progression and without any advantegous properties for cancerous cells [149, 150]. The potential driver genes in paper III were determined by mapping the frequently mutated genes to known oncogenes [153], tumor suppressors [152] and genes frequently mutated in human cancers [151]. The frequency of aberrations was used as an important factor for determining the potential driver genes. The frequently recurring aberrations in different patients are more likely to be important in cancer progression, initiation and carcinogenesis. Additionally, the direction of changes elucidates the function of genes like oncogenes, which are prone to be overexpressed and tumor suppressors, which are prone to be underexpressed. The correlation analysis between genomic copy number and transcriptomics data also contributed in selecting the subset of genes with strong association and clinical relevance.

The Battenberg pipeline estimated the purity of the sample by decomposing tumor and normal cells and at the same time accounted for the aberrant ploidy of the tumor cells [99]. It is difficult to get CN calls from tumors with low cellularity and high heterogeneity since the CNA calling tools need a certain number of aberrations to make exact CN calls (> 20% in ASCAT). Challenges in CNA analysis are aneuploidy, non-aberrant cell admixture and intratumor heterogeneity. The Battenberg pipeline overcomes these limitation by phasing the SNP genotypes with genetically similar populations, and further segmenting to find the switch points, which were then used to re-orient segments and also facilitate

the identification of clonal and subclonal aberrations. It gives higher sensitivity for determining CNA calls for low-cellularity samples like pancreatic tumors.

4.1.5. Integration of high throughput data

The complexity of interpretation of large-scale data sets and integrating multiple data is challenging but important for deriving fundamental and applied biological information [154]. Integrating multiple forms of –omics data in the framework not only gives multiple layers of information but also captures better functional activity [155]. The correlation between CNA and mRNA expression was calculated, and up- or downregulation of 2038 genes (10%) were associated with CNA in the OUH cohort while, 8519 genes (40%) with the TCGA cohort (paper III). The observed variations (10% in OUH and 40% in TCGA) in gene expression in the two cohorts can be explained by CNA, while the difference in total variations could be due to differences in sample size (OUH (n=60) and TCGA (n=127)). The integration of mRNA and CNA data not only allowed the selection of genes, which can increase the accuracy of risk stratification based on outcome but also improved the characterization of genes involved in tumor pathways [199].

The integration of miRNA and mRNA data from PAs identified important interactions involved in deregulated pathways (papers I and II). Different statistical models were used for integration of miRNA and mRNA data and for associating the interactions with biological pathways. The integration analysis deciphered the miRNAs in stroma which are involved in tumor progression and pathways that may facilitate tumor-stroma interactions. This was done in paper II by selecting the mRNAs that are negatively associated with stromal miRNAs and finding the pathways enriched using this mRNA list. In paper I, the integration of miRNA and mRNA with deregulated pathways allowed to identify mRNAs specific to deregulated pathways in morphological subtypes.

An important question in genome analysis is how the deregulation of genes, miRNAs, pathways and genomic aberrations can affect the survival of patients. Several statistical association tests like Kaplan-Meier and Cox PH model can be applied to study survival with respect to expression of genes, miRNAs and CNA. The Cox PH model allows modeling the survival time with continuous variables like age, miRNA/mRNA expression in patients

etc. Thus, it was used to find the effect of amplification/deletion of genomic regions, expression of genes and miRNAs on OS/RFS without setting a predefined threshold for covariates in order to define groups (like group with high or low expression of miR-17). It also allowed adjustment of covariates of interest.

4.1.6. Statistics and bioinformatics considerations

One of the main interests lies in the interpretation of quantitative profiles of high throughput data generated by mRNA, DNA and miRNA profiling. For the quantification and exploration of such a huge dataset, different tools and algorithms are needed to interpret and to gain comprehensive understanding of the data. The classical analytical methods like univariate and multivariate methods are used to assess group wise differences in samples. The univariate analyses methods like t-tests, ANOVA and Wilcoxon rank tests were used to find miRNAs and mRNAs differentially expressed between two groups. Whereas, multivariate methods like PCA captures the variances within a dataset by reducing the dimensionality of data, cluster the samples and gives a better visualization of the data [21]. In general, unsupervised hierarchical clustering is also a multivariate analysis approach, which is beneficial in grouping samples based on intrinsic similarities in their measurements. However, there are issues with respect to clustering approaches like robustness, difficult interpretability and thresholds used for selecting miRNAs and mRNAs [21, 147]. In order to carefully interpreting the data and avoiding false conclusions, permutation tests like Fisher's exact test, Hypergeometric test and bootstrapping with different features and parameters were used. Also, the selection of mRNAs and miRNAs for clustering analyses was performed based on the most variable mRNAs and miRNAs across the samples. These features were selected based on the standardized thresholds used in previous publications [159, 204].

Pathway analysis is used to gain insight into the underlying biology of complex network of differentially expressed genes. There are many underlying challenges with respect to specificity and sensitivity of tools used for pathway analysis. The GSEA tool was used in paper I to determine whether the *a priori* defined sets of genes were enriched in a particular pathway based on statistical tests [161]. Multiple pathway databases like

KEGG, PANTHER, REACTOME were tested alone and in combinations when using different settings for tuning the parameters. The consensus results between different analyses were used for reporting the deregulated pathways, as the GSEA results vary with the number of pathways used in the analysis. In papers II and III, the web-based tool WebGestalt was used for pathway analysis. The gene lists were selected based on correlation between miRNA (paper II) and genomic location (paper III) of interest. The pathway analysis performed on different datasets (papers I, II and III) identified many common deregulated pathways in PAs.

The samples size is a concern in studies related to pancreatic cancer. Although more than 700 new cases are diagnosed yearly in Norway, only 15% of the patients undergo surgery resulting in studies with relative small sample sizes. This is commonly observed in published studies on pancreatic cancer. In order to avoid false negative findings due to small sample size, conclusions should be drawn with caution and univariate and multivariate analyses were carried out at justifiable level of significance and power [160]. In addition, multiple testing corrections and permutation tests were performed. The biological relevance of significant features was determined in order to avoid reporting false conclusions.

4.2. Biological considerations

4.2.1. Identification of putative markers from integration of –*omics* data

Pinpointing specific genetic changes in cells with oncogenic property is crucial. A systems biology approach to molecular data analysis improved and increased the ability to detect the genes, miRNAs and chromosomal locations that are deregulated in the cancer cells. Gene dysregulation in cancer cells can evolve via different mechanisms, and single level genome data cannot fully explain genomic variations observed in the cancer genome [31]. Many genomic variations are present within and between the different molecular levels in a biological system. Thus integration of multiple -*omics* data (miRNA, mRNA, SNP) from the same patients adds power in identifying driver and passenger events [30]. This cataloging of the genetic alteration increases the ability to identify interactions

associated with cancer. This information could be important in order to improve diagnosis and treatment decisions for the patients. The comprehensive analysis using integrative genome and transcriptome data resulted in identifying potential driver events in pancreatic tumors. These driver events were identified by finding functional hotspots with the following properties: 1) frequently aberrated 2) extreme expression levels (high or low) 3) functional advantageous to the cancer cells 4) genes with many differentially expressed neighbouring genes [80]. We identified *CCNE1*, *ERBB2*, *MAPK4* and *DCC* as some of the important putative driver genes in pancreatic cancer by integrative analysis of CNA and mRNA expression data analysis. They are oncogenes and tumor suppressors that are frequently deregulated in PAs. Various neighbouring genes like *POP4*, *UQCRFS1*, *C19orf12* along with *CCNE1* at the 19q12 locus were differentially expressed and associated with gain of 19q12. *CCNE1* is a known basal like marker for breast cancer [32]. This gene was also associated with a highly proliferative "basal like pancreatobiliary subtype", suggesting the importance of *CCNE1* in oncogenesis of pancreatic cancer. The most frequent events in both the cohorts were deletions of chromosomes 6q, 9p, 17p and 18q. Numerous tumor suppressor genes were mapped on these loci and were associated with downregulation at the mRNA level.

Commonly mutated genes (*KRAS*, *CDKN2A*, *SMAD4* and *TP53*) and deregulated miRNAs (miR-21, miR-10b and miR-34a/b) that characterize PDACs were also differentially expressed in our cohort [173-175, 177, 179]. The integrated analysis: "miRNA-mRNA target analysis", "mRNA-CNA association analysis" and "miRNA-mRNA negative association analysis" implemented in papers I, II and III identified many miRNA-mRNA-CNA-pathway networks. For instance, upregulation of the genes *CCNB1* and *CEP55* were associated with downregulation of miR-16, which is linked to proliferative signaling (paper II). Various miRNA families (miR-17, miR-15, miR-515) were upregulated in stroma, and the association analysis of these miRNAs with mRNA data suggested that they were linked to pathways associated with tumor-stroma interaction (paper II). Further, the co-occuring events like deletions of 17p13 and 18q21 indicated that genes (*TP53*, *SMAD2*, *BCL2*, *PIK3R5*, *YWHAE*, *SMAD4* and *SERPINB5*) located in these regions were involved in the same signaling pathways like Wnt signaling, p53 signaling, cell cycle

and apoptosis (paper III). Such regions could be potential targets for therapeutics, but due to interconnectedness and rewiring of signaling pathways, therapeutic targeting is challenging since cells often acquire resistance through the acquisition of compensatory mechanism [201].

4.2.2. Molecular classification of periampullary adenocarcinomas

The molecular differences between subtypes of PAs are poorly understood and PAs from other site of origin than the pancreatic ducts are less studied. Since they have histopathological and prognostic differences [17, 18], we stratified the tumors based on site of origin and morphology. The molecular analyses of the miRNA, mRNA and SNP data suggested that PAs are more distinct at the level of morphology than at the site of origin. The morphological subtypes have differences in pathway regulations, like Wnt and interferon signaling, which were upregulated in pancreatobiliary, while insulin recycling was downregulated in the intestinal subtype (paper I). This is an interesting observation since the clinical symptoms of the two subtypes are very similar, however their diagnostic distinction may be difficult clinically, radiologically and morphologically. At the genomic level, the two morphological subtypes had similar CNAs, while amplification of chromosome 13 and deletion of 5q were specific to the intestinal subtype (paper III). The observed differences between the subtypes showed that they are molecularly distinct and may follow different clinical courses.

Various systematic approaches based on global gene expression signature and CNAs have been proposed for classification of PDACs. Collisson et al. classified PDACs, the most common and aggressive form of PAs into different subgroups using 64 genes, and Gutierrez et al. and Waddell et al. clustered samples based on genomic aberrations [177, 203, 204]. The recent 50-gene signature published by Moffitt et al. classified the PDACs into two subgroups (classical and basal) with different prognosis [202]. By using Collisson classifiers for subtyping the PDACs in our cohort the results could not be validated. This was also the case for Moffitt et al. [202]. We also subtyped the PAs using the PAM50 gene signature [209], which classified the samples based on morphology (pancreatobiliary and

intestinal) and further subtyped the pancreatobiliary samples into basal and classical, similar to the classes defined by Moffitt et al. We observed that both the PAM50 and the Moffitt et al. gene signatures classify tumors based on degree of differentiation. The patients with "basal like pancreatobiliary tumors" with poor degree of differentiation had the worst prognosis.

4.2.3. Clinical implication of molecular classification of periampullary adenocarcinomas

In the management of pancreatic cancer patients various clinical factors are studied and metastatic status, stage and grade of disease, nodal status, tumor size and location have been linked with prognosis [162-164]. In recent years, there has been a focus on identifying miRNAs and mRNAs as prognostic and predictive markers. The heterogeneity of pancreatic tumors poses specific challenges, thus it may be important to stratify patients according to mutation spectrum of tumor to find better treatment options. Targeted therapies represent the new era of drug treatment, and knowledge on perturbed pathways are important for designing/utilizing relevant drugs, designing good clinical trials and improving survival for these patients.

Various studies have shown that the miRNome exhibit unique miRNA signatures in blood, tissues, stool or bile [165]. A meta-analysis of 1525 PDAC patients correlating expression of miRNA showed that high levels of miR-21, miR-10b, miR-155, miR-203, miR-196a/b, miR-222 and low levels of miR-34a/b were associated with worst OS [165]. The combination of measuring plasma miRNAs miR-16, miR-196a with CA-19-9 was shown to be more effective in diagnosing patients with PDAC than only measuring CA-19-9 [170]. The upregulation of miR-10b, miR-155 and miR-106 were identified as predictive markers to identify PDACs in bile and plasma [171]. We identified upregulation of miR-106b and miR-10b in stroma as compared to carcinoma cells, which dictates that stroma could be the source of these miRNAs in plasma.

A number of studies have identified the utility of studying miRNA, mRNA and CNA as prognostic and predictive markers [157, 171-175, 180]. Molecular features of the tumor affect the prognosis of PA patients, thus we identified potential prognostic markers within the morphological subtypes at the miRNA and mRNA level (paper I). The identified potential prognostic miRNAs were also found to have prognostic relevance in other cancers like miR-592 in colorectal cancer, miR-377* in hepatocellular carcinomas, miR-497 in cervical cancer and miR-127 in lymphomas [166-169]. The genes (*PTEN*, *RB1*, *ATM*, *PIK3R2*, *G6PC* and *PDPK1*) were identified as potential subtype specific prognostic biomarkers. Additionally, the amplifications of 19q13 and 18p11.22 were associated with more "aggressive cancer" and poor OS. The association between aggressiveness of pancreatic cancer and amplification of 19q13 has also been reported previously [180]. The genes *SERTAD5* and *ERCC1* in locus 19q13, and COLEC12 and RAB12 in locus 18p11.22 were associated with worst OS and RFS. These genes have also been identified as potential prognostic biomarkers in various other cancers [181-185]. The identified miRNAs, mRNAs and CNAs may act as potentially valuable biomarkers for the clinical setting if validated in larger cohorts.

5. Conclusions and future perspective

In this thesis, multidimensional high throughput data of miRNA, mRNA and SNPs was analyzed to get a comprehensive view of the molecular changes in periampullary adenocarcinomas. These tumors are highly heterogeneous and have abundant desmoplastic stroma, which makes the analysis complicated and at the same time suggests the need of in-depth understanding of molecular changes.

The molecular characterization of PAs supported the hypothesis that they are more distinct at the level of morphology than at the site of origin. The unsupervised hierarchial clustering using miRNA and mRNA data showed that PAs cluster based on morphology at the molecular level. Furthermore, different identified pathways were deregulated in subtypes of PAs. Additionally, the role of miRNAs and mRNAs as potential prognostic markers was identified in subtypes of PAs (paper I).

The role of miRNAs in carcinoma and stromal cells were explored further to characterize miRNAs with respect to stromal component of pancreatobiliary type of PAs. The miRNA expression profiles of carcinoma cells, stromal cells and normal tissues were compared. The negatively associated miRNAs (differentially expressed in stroma) and mRNAs (differentially expressed in tumor + stroma) were mapped to deregulated pathways. The pathways regulating the tumor-stroma interactions were found (paper II).

A broad and comprehensive analysis of copy number alteration events in PAs was performed. The copy number alterations in both the OUH and TCGA cohorts showed that copy number losses were more frequent than gains. Putative driver genes and the deregulated pathways were identified in both cohorts. The amplifications of genomic locations 19q13 and 18p11 were associated with worst survival. Many genes located in these regions were shown to have prognostic importance. The subtyping of PAs based on PAM50 identified two “new” classes of pancreatobiliary subtype (basal and classical). Patients with the basal like subtype had the worst prognosis (paper III). The comparison of each xenografts cell line with its original tumor showed similar gene expression profiles and CNA patterns (paper III and IV).

The long-term goal is the validation and application of biological markers that can aid clinical decision-making. For most patients the disease reoccurs shortly after surgery, and a substantial number of patients acquire resistant to chemotherapy. Studying the molecular events and modeling the risk of recurrence for these patients may help to improve diagnostics and identify patients with poor outcome. The next step forward could be integration of proteome and methylation data to the existing layer of –omics data, which may help in identifying molecular markers to stratify the aggressiveness of disease, risk of recurrence and response to chemotherapy. Studying these multiple molecular interactions involved in various cellular and biological pathways need better approaches, algorithms and analysis pipelines for deciphering the intercalated network of information. It would be interesting to verify and use tools where multiple layers of data could be integrated to identify deregulated pathways of significance to clinical outcome.

6. Abbreviations

Ago2: Argonaute 2

AJCC: The American Joint Committee on Cancer

ANOVA: Analysis of variance

ASCAT: Allele specific copy number analysis tool

αSMA: Alpha smooth muscle actin

BAF: B allele frequency

CGH: Comparative genomic hybridization

CNA: Copy number aberration

CNV: Copy number variation

CT: Computer tomography

DNA: Deoxyribonucleic acid

eQTL: expression Quantitative trait loci

ECM: Extracellular matrix

EMT: Epithelial mesenchymal transition

FDA: Food and Drug Administration

FDR: False discovery rate

FFPE: Formalin fixed paraffin embedded

FGF: Fibroblast growth factor

FU: Fluoro-uracil

GEO: Gene expression omnibus

GII: Genome instability index

GSEA: Gene set enrichment analysis

HGF: Hepatocyte growth factor

HGP: Human Genome Project

HPP: Human pancreatic polypeptide

IPMN: Intraductal papillary mucinous neoplasm

LOH: Loss of heterozygosity

MAPK: Mitogen-activated protein kinase

MCN: Mucinous cystic Neoplasms

miRNA: microRNA

Modt: Moderated t test

MRI: Magnetic resonance imaging

mRNA: messenger RNA

OS: Overall survival

OUH: Oslo University Hospital

PA: Periampullary adenocarcinoma

PanIN: Pancreatic intraepithelial neoplasia

PCA: Principal component analysis

PCF: Piecewise constant fitting

PD: Pancreatoduodenectomy

PDAC: Pancreatic ductal adenocarcinoma

PET: Position emission tomography

PI3K: Phosphatidylinositol 3-kinase

Pre-miRNA: Precursor microRNA

Pri-miRNA: Primary microRNA

qRT-PCR: Quantitative real time polymerase chain reaction

RFS: Relapse free survival

RIN: RNA integrity number

RISC: RNA induced silencing complex

RMA: Robust multiarray analysis

SNP: Single nucleotide polymorphism

SPCA: Sparse principal component analysis

TCGA: The Cancer Genome Atlas

TGF: Transforming growth factor

TS: Thymidylate synthase

We-ARMS: Wobble enhanced amplification refractory mutation system

WGS: Whole genome sequencing

WHO: World Health Organization

7. References

1. Decker, G.A., Batheja, M.J., Collins, J.M., Silva, A.C., Mekeel, K.L., Moss, A.A., Nguyen, C.C., Lake, D.F., Miller, L.J., 2010. Risk factors for pancreatic adenocarcinoma and prospects for screening. Gastroenterology & hepatology 6, 246-254.

2. Zheng, W., McLaughlin, J.K., Gridley, G., Bjelke, E., Schuman, L.M., Silverman, D.T., Wacholder, S., Co-Chien, H.T., Blot, W.J., Fraumeni, J.F., Jr., 1993. A cohort study of smoking, alcohol consumption, and dietary factors for pancreatic cancer (United States). Cancer causes & control : CCC 4, 477-482.

3. Holly, E.A., Chaliha, I., Bracci, P.M., Gautam, M., 2004. Signs and symptoms of pancreatic cancer: a population-based case-control study in the San Francisco Bay area. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2, 510-517.

4. Cancer Registry of Norway. Cancer in Norway 2014-Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway, 2015.

5. Siegel, R., Ma, J., Zou, Z., Jemal, A., 2014. Cancer statistics, 2014. CA: a cancer journal for clinicians 64, 9-29.

6. Gray H. Lewis WH, ed. Gray's Anatomy of the Human Body. 20th ed. New York, NY: Bartleby.com; 2000.

7. American Cancer Society. Cancer Facts & Figures 2014. Atlanta: American Cancer Society; 2014.

8. Silverman, D.T., Schiffman, M., Everhart, J., Goldstein, A., Lillemoe, K.D., Swanson, G.M., Schwartz, A.G., Brown, L.M., Greenberg, R.S., Schoenberg, J.B., Pottern, L.M., Hoover, R.N., Fraumeni, J.F., Jr., 1999. Diabetes mellitus, other medical conditions and

familial history of cancer as risk factors for pancreatic cancer. British journal of cancer 80, 1830-1837.

9. Tersmette, A.C., Petersen, G.M., Offerhaus, G.J., Falatko, F.C., Brune, K.A., Goggins, M., Rozenblum, E., Wilentz, R.E., Yeo, C.J., Cameron, J.L., Kern, S.E., Hruban, R.H., 2001. Increased risk of incident pancreatic cancer among first-degree relatives of patients with familial pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 7, 738-744.

10. Edge, S.B., Compton, C.C., 2010. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of surgical oncology 17, 1471-1474.

11. Wasif, N., Ko, C.Y., Farrell, J., Wainberg, Z., Hines, O.J., Reber, H., Tomlinson, J.S., 2010. Impact of tumor grade on prognosis in pancreatic cancer: should we include grade in AJCC staging? Annals of surgical oncology 17, 2312-2320.

12. Geer, R.J., Brennan, M.F., 1993. Prognostic indicators for survival after resection of pancreatic adenocarcinoma. American journal of surgery 165, 68-72; discussion 72-63.

13. Allema, J.H., Reinders, M.E., van Gulik, T.M., Koelemay, M.J., Van Leeuwen, D.J., de Wit, L.T., Gouma, D.J., Obertop, H., 1995. Prognostic factors for survival after pancreaticoduodenectomy for patients with carcinoma of the pancreatic head region. Cancer 75, 2069-2076.

14. Sobin LH, Gospodarowicz MK,Wittekind C, TNM Classification of Malignant Tumors. 7th ed. New York, NY: Wiley-Blackwell; 2009.

15. Yeo, C.J., Sohn, T.A., Cameron, J.L., Hruban, R.H., Lillemoe, K.D., Pitt, H.A., 1998. Periampullary adenocarcinoma: analysis of 5-year survivors. Annals of surgery 227, 821-831.

16. Kimura, W., Futakawa, N., Yamagata, S., Wada, Y., Kuroda, A., Muto, T., Esaki, Y., 1994. Different clinicopathologic findings in two histologic types of carcinoma of papilla of Vater. Japanese journal of cancer research : Gann 85, 161-166.

17. Westgaard, A., Tafjord, S., Farstad, I.N., Cvancarova, M., Eide, T.J., Mathisen, O., Clausen, O.P., Gladhaug, I.P., 2008. Pancreatobiliary versus intestinal histologic type of differentiation is an independent prognostic factor in resected periampullary adenocarcinoma. BMC cancer 8, 170.

18. Westgaard, A., Pomianowska, E., Clausen, O.P., Gladhaug, I.P., 2013. Intestinal-type and pancreatobiliary-type adenocarcinomas: how does ampullary carcinoma differ from other periampullary malignancies? Annals of surgical oncology 20, 430-439.

19. Glenn, J., Steinberg, W.M., Kurtzman, S.H., Steinberg, S.M., Sindelar, W.F., 1988. Evaluation of the utility of a radioimmunoassay for serum CA 19-9 levels in patients before and after treatment of carcinoma of the pancreas. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 6, 462-468.

20. Goonetilleke, K.S., Siriwardena, A.K., 2007. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology 33, 266-270.

21. Bartel, J., Krumsiek, J., Theis, F.J., 2013. Statistical methods for the analysis of high-throughput metabolomics data. Computational and structural biotechnology journal 4, e201301009.

22. Campbell F, Verbeke CS. (2013) Pathology of the Pancreas: A Practical Approach. London, UK: Springer, chapter 9.

23. Miura, F., Takada, T., Amano, H., Yoshida, M., Furui, S., Takeshita, K., 2006. Diagnosis of pancreatic cancer. HPB : the official journal of the International Hepato Pancreato Biliary Association 8, 337-342.

24. Carpelan-Holmstrom, M., Louhimo, J., Stenman, U.H., Alfthan, H., Haglund, C., 2002. CEA, CA 19-9 and CA 72-4 improve the diagnostic accuracy in gastrointestinal cancers. Anticancer research 22, 2311-2316.

25. Whipple, A.O., Parsons, W.B., Mullins, C.R., 1935. Treatment of Carcinoma of the Ampulla of Vater. Annals of surgery 102, 763-779.

26. Kausch W. Das Karzinoma der Papilla duodeni and seine radikale Entfernung. Betir Klin Chir 1912;78:439–86

27. Bachmann, J., Michalski, C.W., Martignoni, M.E., Buchler, M.W., Friess, H., 2006. Pancreatic resection for pancreatic cancer. HPB : the official journal of the International Hepato Pancreato Biliary Association 8, 346-351.

28. Traverso, L.W., Longmire, W.P., Jr., 1978. Preservation of the pylorus in pancreaticoduodenectomy. Surgery, gynecology & obstetrics 146, 959-962.

29. Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646-674.

30. Huang, N., Shah, P.K., Li, C., 2012. Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. Briefings in bioinformatics 13, 305-316.

31. Burrell, R.A., McGranahan, N., Bartek, J., Swanton, C., 2013. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 501, 338-345.

32. Prat, A., Adamo, B., Fan, C., Peg, V., Vidal, M., Galvan, P., Vivancos, A., Nuciforo, P., Palmer, H.G., Dawood, S., Rodon, J., Ramon y Cajal, S., Del Campo, J.M., Felip, E., Tabernero, J., Cortes, J., 2013. Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. Scientific reports 3, 3544.

33. Lillemoe, K.D., Kaushal, S., Cameron, J.L., Sohn, T.A., Pitt, H.A., Yeo, C.J., 1999. Distal pancreatectomy: indications and outcomes in 235 patients. Annals of surgery 229, 693-698; discussion 698-700.

34. Cascinu, S., Jelic, S., Group, E.G.W., 2009. Pancreatic cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 20 Suppl 4, 37-40.

35. Christians, K.K., Tsai, S., Mahmoud, A., Ritch, P., Thomas, J.P., Wiebe, L., Kelly, T., Erickson, B., Wang, H., Evans, D.B., George, B., 2014. Neoadjuvant FOLFIRINOX for borderline resectable pancreas cancer: a new treatment paradigm? The oncologist 19, 266-274.

36. Ferrone, C.R., Marchegiani, G., Hong, T.S., Ryan, D.P., Deshpande, V., McDonnell, E.I., Sabbatino, F., Santos, D.D., Allen, J.N., Blaszkowsky, L.S., Clark, J.W., Faris, J.E., Goyal, L., Kwak, E.L., Murphy, J.E., Ting, D.T., Wo, J.Y., Zhu, A.X., Warshaw, A.L., Lillemoe, K.D., Fernandez-del Castillo, C., 2015. Radiological and surgical implications of neoadjuvant treatment with FOLFIRINOX for locally advanced and borderline resectable pancreatic cancer. Annals of surgery 261, 12-17.

37. Kalser, M.H., Ellenberg, S.S., 1985. Pancreatic cancer. Adjuvant combined radiation and chemotherapy following curative resection. Archives of surgery 120, 899-903.

38. Neoptolemos, J.P., Stocken, D.D., Friess, H., Bassi, C., Dunn, J.A., Hickey, H., Beger, H., Fernandez-Cruz, L., Dervenis, C., Lacaine, F., Falconi, M., Pederzoli, P., Pap, A.,

Spooner, D., Kerr, D.J., Buchler, M.W., European Study Group for Pancreatic, C., 2004. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. The New England journal of medicine 350, 1200-1210.

39. Oettle, H., Post, S., Neuhaus, P., Gellert, K., Langrehr, J., Ridwelski, K., Schramm, H., Fahlke, J., Zuelke, C., Burkart, C., Gutberlet, K., Kettner, E., Schmalenberg, H., Weigang-Koehler, K., Bechstein, W.O., Niedergethmann, M., Schmidt-Wolf, I., Roll, L., Doerken, B., Riess, H., 2007. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial. Jama 297, 267-277.

40. Neoptolemos, J.P., Stocken, D.D., Bassi, C., Ghaneh, P., Cunningham, D., Goldstein, D., Padbury, R., Moore, M.J., Gallinger, S., Mariette, C., Wente, M.N., Izbicki, J.R., Friess, H., Lerch, M.M., Dervenis, C., Olah, A., Butturini, G., Doi, R., Lind, P.A., Smith, D., Valle, J.W., Palmer, D.H., Buckels, J.A., Thompson, J., McKay, C.J., Rawcliffe, C.L., Buchler, M.W., European Study Group for Pancreatic, C., 2010. Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: a randomized controlled trial. Jama 304, 1073-1081.

41. White, R.R., Reddy, S., Tyler, D.S., 2005. The role of chemoradiation therapy in locally advanced pancreatic cancer. HPB : the official journal of the International Hepato Pancreato Biliary Association 7, 109-113.

42. O'Reilly, E.M., 2011. Adjuvant therapy for pancreas adenocarcinoma: where are we going? Expert review of anticancer therapy 11, 173-177.

43. Von Hoff, D.D., Ervin, T., Arena, F.P., Chiorean, E.G., Infante, J., Moore, M., Seay, T., Tjulandin, S.A., Ma, W.W., Saleh, M.N., Harris, M., Reni, M., Dowden, S., Laheru, D., Bahary, N., Ramanathan, R.K., Tabernero, J., Hidalgo, M., Goldstein, D., Van Cutsem, E., Wei, X., Iglesias, J., Renschler, M.F., 2013. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. The New England journal of medicine 369, 1691-1703.

44. Abbruzzese, J.L., Hess, K.R., 2014. New option for the initial management of metastatic pancreatic cancer? Journal of clinical oncology: official journal of the American Society of Clinical Oncology 32, 2405-2407.

45. Vincent, A., Herman, J., Schulick, R., Hruban, R.H., Goggins, M., 2011. Pancreatic cancer. Lancet 378, 607-620.

46. Aho, U., Zhao, X., Lohr, M., Andersson, R., 2007. Molecular mechanisms of pancreatic cancer and potential targets of treatment. Scandinavian journal of gastroenterology 42, 279-296.

47. Hruban, R.H., Maitra, A., Goggins, M., 2008. Update on pancreatic intraepithelial neoplasia. International journal of clinical and experimental pathology 1, 306-316.

48. Bardeesy, N., DePinho, R.A., 2002. Pancreatic cancer biology and genetics. Nature reviews. Cancer 2, 897-909.

49. Yamano, M., Fujii, H., Takagaki, T., Kadowaki, N., Watanabe, H., Shirai, T., 2000. Genetic progression and divergence in pancreatic carcinoma. The American journal of pathology 156, 2123-2133.

50. Moskaluk, C.A., Hruban, R.H., Kern, S.E., 1997. p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. Cancer research 57, 2140-2143.

51. Feig, C., Gopinathan, A., Neesse, A., Chan, D.S., Cook, N., Tuveson, D.A., 2012. The pancreas cancer microenvironment. Clinical cancer research : an official journal of the American Association for Cancer Research 18, 4266-4276.

52. Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., Matrisian, L.M., 2014. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research 74, 2913-2921.

53. Rasheed, Z.A., Matsui, W., Maitra, A., 2012. Pathology of pancreatic stroma in PDAC, in: Grippo, P.J., Munshi, H.G. (Eds.), Pancreatic Cancer and Tumor Microenvironment, Trivandrum (India).

54. Maehara, N., Matsumoto, K., Kuba, K., Mizumoto, K., Tanaka, M., Nakamura, T., 2001. NK4, a four-kringle antagonist of HGF, inhibits spreading and invasion of human pancreatic cancer cells. British journal of cancer 84, 864-873.

55. Omary, M.B., Lugea, A., Lowe, A.W., Pandol, S.J., 2007. The pancreatic stellate cell: a star on the rise in pancreatic diseases. The Journal of clinical investigation 117, 50-59.

56. Sato, N., Maehara, N., Goggins, M., 2004. Gene expression profiling of tumor-stromal interactions between pancreatic cancer cells and stromal fibroblasts. Cancer research 64, 6950-6956.

57. Rhim, A.D., Oberstein, P.E., Thomas, D.H., Mirek, E.T., Palermo, C.F., Sastra, S.A., Dekleva, E.N., Saunders, T., Becerra, C.P., Tattersall, I.W., Westphalen, C.B., Kitajewski, J., Fernandez-Barrena, M.G., Fernandez-Zapico, M.E., Iacobuzio-Donahue, C., Olive, K.P., Stanger, B.Z., 2014. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer cell 25, 735-747.

58. Tzeng, C.W., Fleming, J.B., Lee, J.E., Xiao, L., Pisters, P.W., Vauthey, J.N., Abdalla, E.K., Wolff, R.A., Varadhachary, G.R., Fogelman, D.R., Crane, C.H., Balachandran, A., Katz, M.H., 2012. Defined clinical classifications are associated with outcome of patients with anatomically resectable pancreatic adenocarcinoma treated with neoadjuvant therapy. Annals of surgical oncology 19, 2045-2053.

59. Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., Irizarry, R.A., 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature reviews. Genetics 11, 733-739.

60. Luo, G., Long, J., Zhang, B., Liu, C., Xu, J., Ni, Q., Yu, X., 2012. Stroma and pancreatic ductal adenocarcinoma: an interaction loop. Biochimica et biophysica acta 1826, 170-178.

61. Verbeke, C.S., Gladhaug, I.P., 2012. Resection margin involvement and tumour origin in pancreatic head cancer. The British journal of surgery 99, 1036-1049.

62. Gillen, S., Schuster, T., Meyer Zum Buschenfelde, C., Friess, H., Kleeff, J., 2010. Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages. PLoS medicine 7, e1000267.

63. Moore, M.J., Goldstein, D., Hamm, J., Figer, A., Hecht, J.R., Gallinger, S., Au, H.J., Murawa, P., Walde, D., Wolff, R.A., Campos, D., Lim, R., Ding, K., Clark, G., Voskoglou-Nomikos, T., Ptasynski, M., Parulekar, W., National Cancer Institute of Canada Clinical Trials, G., 2007. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 25, 1960-1966.

64. Winter, J., Jung, S., Keller, S., Gregory, R.I., Diederichs, S., 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nature cell biology 11, 228-234.

65. Doleshal, M., Magotra, A.A., Choudhury, B., Cannon, B.D., Labourier, E., Szafranska, A.E., 2008. Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. The Journal of molecular diagnostics : JMD 10, 203-211.

66. Xi, Y., Nakajima, G., Gavin, E., Morris, C.G., Kudo, K., Hayashi, K., Ju, J., 2007. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. Rna 13, 1668-1674.

67. He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nature reviews. Genetics 5, 522-531.

68. Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297.

69. Lai, E.C., 2003. microRNAs: runts of the genome assert themselves. Current biology : CB 13, R925-936.

70. Kozomara, A., Griffiths-Jones, S., 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research 42, D68-73.

71. Rajewsky, N., 2006. microRNA target predictions in animals. Nature genetics 38 Suppl, S8-13.

72. Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H., Cuppen, E., 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21-24.

73. Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., Prueitt, R.L., Yanaihara, N., Lanza, G., Scarpa, A., Vecchione, A., Negrini, M., Harris, C.C., Croce, C.M., 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America 103, 2257-2261.

74. Garzon, R., Fabbri, M., Cimmino, A., Calin, G.A., Croce, C.M., 2006. MicroRNA expression and function in cancer. Trends in molecular medicine 12, 580-587.

75. Calin, G.A., Croce, C.M., 2006. MicroRNA-cancer connection: the beginning of a new tale. Cancer research 66, 7390-7394.

76. White, N.M., Fatoohi, E., Metias, M., Jung, K., Stephan, C., Yousef, G.M., 2011. Metastamirs: a stepping stone towards improved cancer management. Nature reviews. Clinical oncology 8, 75-84.

77. Reinhardt, H.C., Aslanian, A.S., Lees, J.A., Yaffe, M.B., 2007. p53-deficient cells rely on ATM- and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage. Cancer cell 11, 175-189.

78. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human MicroRNA targets. PLoS biology 2, e363.

79. Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature reviews. Genetics 9, 102-114.

80. Van Loo, P., Nilsen, G., Nordgard, S.H., Vollan, H.K., Borresen-Dale, A.L., Kristensen, V.N., Lingjaerde, O.C., 2012. Analyzing cancer samples with SNP arrays. Methods in molecular biology 802, 57-72, Chapter 4

81. Peterson, S.M., Thompson, J.A., Ufkin, M.L., Sathyanarayana, P., Liaw, L., Congdon, C.B., 2014. Common features of microRNA target prediction tools. Frontiers in genetics 5, 23.

82. Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., Chien, C.H., Wu, M.C., Huang, C.Y., Tsou, A.P., Huang, H.D.,

2011. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic acids research 39, D163-169.

83. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., Li, T., 2009. miRecords: an integrated resource for microRNA-target interactions. Nucleic acids research 37, D105-110.

84. Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., Backofen, R., 2007. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS computational biology 3, e65.

85. Jansson, M.D., Lund, A.H., 2012. MicroRNA and cancer. Molecular oncology 6, 590-610.

86. Hamilton, M.P., Rajapakshe, K., Hartig, S.M., Reva, B., McLellan, M.D., Kandoth, C., Ding, L., Zack, T.I., Gunaratne, P.H., Wheeler, D.A., Coarfa, C., McGuire, S.E., 2013. Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif. Nature communications 4, 2730.

87. Crick, F., 1970. Central dogma of molecular biology. Nature 227, 561-563.

88. Wiszniewska, J., Bi, W., Shaw, C., Stankiewicz, P., Kang, S.H., Pursley, A.N., Lalani, S., Hixson, P., Gambin, T., Tsai, C.H., Bock, H.G., Descartes, M., Probst, F.J., Scaglia, F., Beaudet, A.L., Lupski, J.R., Eng, C., Cheung, S.W., Bacino, C., Patel, A., 2014. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. European journal of human genetics : EJHG 22, 79-87.

89. International HapMap, C., 2003. The International HapMap Project. Nature 426, 789-796.

90. International HapMap, C., 2005. A haplotype map of the human genome. Nature 437, 1299-1320.

91. Consortium, E.P., 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306, 636-640.

92. Freeman JL, Perry GH, Feuk L, Redon R, et al, 2006. Copy number variation: new insights in genome diversity. Genome Res; 16:949–961.

93. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S.W., Tavare, S., Deloukas, P., Hurles, M.E., Dermitzakis, E.T., 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315, 848-853.

94. Stratton, M.R., Campbell, P.J., Futreal, P.A., 2009. The cancer genome. Nature 458, 719-724.

95. Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061-1073.

96. Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. Cell 100, 57-70.

97. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., Meyerson, M., 2010. The

landscape of somatic copy-number alteration across human cancers. Nature 463, 899-905.

98. Chin, L., Hahn, W.C., Getz, G., Meyerson, M., 2011. Making sense of cancer genomic data. Genes & development 25, 534-555.

99. Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., Perou, C.M., Borresen-Dale, A.L., Kristensen, V.N., 2010. Allele-specific copy number analysis of tumors. Proceedings of the National Academy of Sciences of the United States of America 107, 16910-16915.

100. International Human Genome Sequencing, C., 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931-945.

101. Pennisi, E., 2012. Genomics. ENCODE project writes eulogy for junk DNA. Science 337, 1159, 1161.

102. Wheeler, D.A., Wang, L., 2013. From human genome to cancer genome: the first decade. Genome research 23, 1054-1062.

103. Zadran, S., Remacle, F., Levine, R.D., 2013. miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. Proceedings of the National Academy of Sciences of the United States of America 110, 19160-19165.

104. Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell, 2000, Molecular Cell biology, 4th edition, New York, W.H. Freeman publications.

105. Butte, A., 2002. The use and analysis of microarray data. Nature reviews. Drug discovery 1, 951-960.

106. Simon, R., 2009. Analysis of DNA microarray expression data. Best practice & research. Clinical haematology 22, 271-282.

107. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249-264.

108. Lopez-Romero, P., 2011. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library. BMC genomics 12, 64.

109. Bolstad, B. M., Irizarry R. A., Astrand, M, and Speed, T. P. (2003) A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. Bioinformatics 19(2) ,pp 185-193.

110. Altman, D.G., Bland, J.M., 1995. Statistics notes: the normal distribution. Bmj 310, 298.

111. Kitchen, C.M., 2009. Nonparametric vs parametric tests of location in biomedical research. American journal of ophthalmology 147, 571-572.

112. Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology 3, Article3.

113. Smyth, G.K., Michaud, J., Scott, H.S., 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067-2075.

114. Kruskal W. and Wallis W.A, 1952 "Use of ranks in one-criterion variance analysis". Journal of the American Statistical Association 47 (260): 583–621. doi:10.1080/01621459.1952.10 483441.

115. Mann, Henry B.; Whitney, Donald R. (1947). "On a test of whether one of two random variables is stochastically Larger than the other". Annals of Mathematical Statistics 18 (1): 50–60. doi:10.1214/aoms/1177730491. MR 22058. Zbl 0041.26103

116. Y. Benjamini, Y. Hochberg Controlling the false discovery rate: a practical and powerful approach to multiple testing J. Roy. Stat. Soc. B., 57 (1995), pp. 289–300

117. Noble, W.S., 2009. How does multiple testing correction work? Nature biotechnology 27, 1135-1137.

118. Fisher, R. A., 1922. "On the interpretation of χ2 from contingency tables, and the calculation of P". Journal of the Royal Statistical Society 85(1): 87–94. doi:10.2307/2340521. JSTOR 2340521

119. Yates, F, 1934. "Contingency table involving small numbers and the χ2 test". Supplement to the Journal of the Royal Statistical Society 1(2): 217–235.JSTOR 2983604

120. Plackett RL, 1983, Karl Pearson and the Chi-squared test. Int Stat Rev;51:59-72

121. Johnson, N. L., Kotz, S., and Kemp, A. W. (1992) Univariate Discrete Distributions, Second Edition. New York: Wiley.

122. Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America 95, 14863-14868.

123. Jolliffe IT (1986) Principal Component Analysis. Springer-Verlag, pp. 487. DOI: 10.1007/b98835. ISBN 978-0-387-95442-4

124. Witten, D.M., Tibshirani, R., Hastie, T., 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10, 515-534.

125. Witten, D.M., Tibshirani, R.J., 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. Statistical applications in genetics and molecular biology 8, Article28.

126. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human MicroRNA targets. PLoS biology 2, e363.

127. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S., 2003. MicroRNA targets in Drosophila. Genome biology 5, R1.

128. Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15-20.

129. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of mammalian microRNA targets. Cell 115, 787-798.

130. H. Dweep, C. Sticht, P. Pandey, N. Gretz, 2011, miRWalk-database: prediction of possible miRNA binding sites by “walking” the genes of three genomes,J. Biomed. Inform., 44, pp. 839–847

131. Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., Chien, C.H., Wu, M.C., Huang, C.Y., Tsou, A.P., Huang, H.D., 2011. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic acids research 39, D163-169.

132. Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y., Jian, T.Y., Lin, F.M., Chang, T.H., Weng, S.L., Liao, K.W., Liao, I.E., Liu, C.C., Huang, H.D., 2014. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic acids research 42, D78-85.

133. Molinari F, Frattini M, 2012, KRAS mutational test for metastatic colorectal cancer patients: not just a technical problem, Expert Rev Mol Diagn. 2012 Mar;12(2):123-6. doi: 10.1586/erm.11.94.

134. Hamfjord J, Stangeland AM, Skrede ML, Tveit KM, Ikdahl T, Kure EH, 2011, Wobble-enhanced ARMS method for detection of KRAS and BRAF mutations, Diagn Mol Pathol. 2011 Sep;20(3):158-65. doi: 10.1097/PDM.0b013e31820b49e2.

135. A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, 2005,Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. United States America, 102 (2005), pp. 15545–15550.

136. Zhang, B., Kirov, S.A., Snoddy, J.R. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res, 33(Web Server issue), W741-748.

137. Wang, J., Duncan, D., Shi, Z., Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res, 41 (Web Server issue), W77-83.

138. Hosmer D. W., Lemeshow S., May S, 2008, Applied Survival Analysis: Regression Modeling of Time to Event Data (Wiley Series in Probability and Statistics) 2nd edition. Wiley-Interscience.

139. Kleinbaum DG, 1996, Survival analysis — a self-learning text. New York: Springer-Verlag.

140. Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S.L., Hinton, J., Menzies, A., Stebbings, L.A., Leroy, C., Jia, M., Rance, R., Mudie, L.J., Gamble, S.J., Stephens, P.J., McLaren, S., Tarpey, P.S., Papaemmanuil, E., Davies, H.R., Varela, I., McBride, D.J., Bignell, G.R., Leung, K., Butler, A.P., Teague, J.W., Martin, S., Jonsson, G., Mariani, O., Boyault, S., Miron, P., Fatima, A., Langerod, A., Aparicio, S.A., Tutt, A., Sieuwerts, A.M., Borg, A., Thomas, G., Salomon, A.V., Richardson, A.L., Borresen-Dale, A.L., Futreal, P.A., Stratton, M.R., Campbell, P.J., Breast Cancer Working Group of the International Cancer Genome, C., 2012. The life history of 21 breast cancers. Cell 149, 994-1007.

141. Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 5, e1000529.

142. Nilsen, G., Liestol, K., Van Loo, P., Moen Vollan, H.K., Eide, M.B., Rueda, O.M., Chin, S.F., Russell, R., Baumbusch, L.O., Caldas, C., Borresen-Dale, A.L., Lingjaerde, O.C., 2012. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC genomics 13, 591.

143. Fodde, R., Smits, R., Clevers, H., 2001. APC, signal transduction and genetic instability in colorectal cancer. Nature reviews. Cancer 1, 55-67.

144. Petrucelli, N., Daly, M.B., Feldman, G.L., 1993. BRCA1 and BRCA2 Hereditary Breast and Ovarian Cancer, in: Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J.H., Bird, T.D., Fong, C.T., Smith, R.J.H., Stephens, K. (Eds.), GeneReviews(R), Seattle (WA).

145. Durinck, S., 2008. Pre-processing of microarray data and analysis of differential expression. Methods in molecular biology 452, 89-110.

146. Quackenbush, J., 2002. Microarray data normalization and transformation. Nature genetics 32 Suppl, 496-501.

147. Rahnenfuhrer, J., 2005. Clustering algorithms and other exploratory methods for microarray data analysis. Methods of information in medicine 44, 444-448.

148. Heinrichs, S., Li, C., Look, A.T., 2010. SNP array analysis in hematologic malignancies: avoiding false discoveries. Blood 115, 4157-4161.

149. Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., Du, J., Kau, T., Thomas, R.K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R.M., Demichelis, F., Hatton, C., Rubin, M.A., Garraway, L.A., Nelson, S.F., Liau, L., Mischel, P.S., Cloughesy, T.F., Meyerson, M., Golub, T.A., Lander, E.S., Mellinghoff, I.K., Sellers, W.R., 2007. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proceedings of the National Academy of Sciences of the United States of America 104, 20007-20012.

150. Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., Sougnez, C., Gabriel, S.B., Hernandez, B., Shen, H., Laird, P.W., Getz, G., Meyerson, M., Beroukhim, R., 2013. Pan-cancer patterns of somatic copy number alteration. Nature genetics 45, 1134-1140.

151. Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., Cooper, C.S., 2010. A census of amplified and overexpressed human cancer genes. Nature reviews. Cancer 10, 59-64.

152. Min Zhao, Jingchun Sun, Zhongming Zhao (2013) TSGene: a web resource for tumor suppressor genes. Nucleic Acids Research, 41: D970-D976

153. P. Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman & Michael R. Stratton (2004). A census of human cancer genes, Nature Reviews Cancer 4, 177-183 | doi:10.1038/nrc1299

154. Joyce, A.R., Palsson, B.O., 2006. The model organism as a system: integrating 'omics' data sets. Nature reviews. Molecular cell biology 7, 198-210.

155. Stransky, B., Barrera, J., Ohno-Machado, L., De Souza, S.J., 2007. Modeling cancer: integration of "omics" information in dynamic systems. Journal of bioinformatics and computational biology 5, 977-986.

156. Vilardell, F., Iacobuzio-Donahue, C.A., 2010. Cancer gene profiling in pancreatic cancer. Methods in molecular biology 576, 279-292.

157. Iacobuzio-Donahue, C. A., R. Ashfaq, A. Maitra, N. V. Adsay, G. L. Shen-Ong, K. Berg, M. A. Hollingsworth, J. L. Cameron, C. J. Yeo, S. E. Kern, M. Goggins, and R. H. Hruban. (2003). Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. Cancer Res 63, 8614–22

158. Burdall, S.E., Hanby, A.M., Lansdown, M.R., Speirs, V., 2003. Breast cancer cell lines: friend or foe? Breast cancer research : BCR 5, 89-95.

159. Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, et al. (2002) Integrated classification of lung tumors and cell lines by expression profiling. Proc Natl Acad Sci U S A 99: 12357–12362. doi: 10.1073/pnas.192240599

160. Biau, D.J., Kerneis, S., Porcher, R., 2008. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clinical orthopaedics and related research 466, 2282-2288.

161. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545-15550.

162. Kuhlmann KFD, de Castro SMM, Wesseling JG, ten Kate FJW, Offerhaus GJA, Busch ORC, van Gulik TM, Obertop H, Gouma DJ (2004) Surgical treatment of pancreatic adenocarcinoma: actual survival and prognostic factors in 343 patients. Eur J Cancer 40: 549–558

163. Ridwelski K, Meyer F, Ebert M, Malfertheiner P, Lippert H (2001) Prognostic parameters determining survival in pancreatic carcinoma and, in particular, after palliative treatment. Digest Dis 19(1): 85–92

164. Stocken, D.D., Hassan, A.B., Altman, D.G., Billingham, L.J., Bramhall, S.R., Johnson, P.J., Freemantle, N., 2008. Modelling prognostic factors in advanced pancreatic cancer. British journal of cancer 99, 883-893.

165. Frampton, A.E., Krell, J., Jamieson, N.B., Gall, T.M., Giovannetti, E., Funel, N., Mato Prado, M., Krell, D., Habib, N.A., Castellano, L., Jiao, L.R., Stebbing, J., 2015. microRNAs with prognostic significance in pancreatic ductal adenocarcinoma: A meta-analysis. European journal of cancer 51, 1389-1404.

166. Mosakhani, N., Lahti, L., Borze, I., Karjalainen-Lindsberg, M.L., Sundstrom, J., Ristamaki, R., Osterlund, P., Knuutila, S., Sarhadi, V.K., 2012. MicroRNA profiling predicts survival in anti-EGFR treated chemorefractory metastatic colorectal cancer patients with wild-type KRAS and BRAF. Cancer genetics 205, 545-551.

167. Luo, M., Shen, D., Zhou, X., Chen, X., Wang, W., 2013. MicroRNA-497 is a potential prognostic marker in human cervical cancer and functions as a tumor suppressor by targeting the insulin-like growth factor 1 receptor. Surgery 153, 836-847.

168. Jiang, J., Gusev, Y., Aderca, I., Mettler, T.A., Nagorney, D.M., Brackett, D.J., Roberts, L.R., Schmittgen, T.D., 2008. Association of MicroRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. Clinical cancer research : an official journal of the American Association for Cancer Research 14, 419-427.

169. Goswami, R.S., Atenafu, E.G., Xuan, Y., Waldron, L., Reis, P.P., Sun, T., Datti, A., Xu, W., Kuruvilla, J., Good, D.J., Lai, R., Church, A.J., Lam, W.S., Baetz, T., Lebrun, D.P., Sehn, L.H., Farinha, P., Jurisica, I., Bailey, D.J., Gascoyne, R.D., Crump, M., Kamel-Reid, S., 2013. MicroRNA signature obtained from the comparison of aggressive with indolent non-Hodgkin lymphomas: potential prognostic value in mantle-cell lymphoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 31, 2903-2911.

170. Liu, J., Gao, J., Du, Y., Li, Z., Ren, Y., Gu, J., Wang, X., Gong, Y., Wang, W., Kong, X., 2012. Combination of plasma microRNAs with serum CA19-9 for early detection of pancreatic cancer. International journal of cancer. Journal international du cancer 131, 683-691.

171. Cote, G.A., Gore, A.J., McElyea, S.D., Heathers, L.E., Xu, H., Sherman, S., Korc, M., 2014. A pilot study to develop a diagnostic test for pancreatic ductal adenocarcinoma based on differential expression of select miRNA in plasma and bile. The American journal of gastroenterology 109, 1942-1952.

172. Greither, T., Grochola, L.F., Udelnow, A., Lautenschlager, C., Wurl, P., Taubert, H., 2010. Elevated expression of microRNAs 155, 203, 210 and 222 in pancreatic tumors is

associated with poorer survival. International journal of cancer. Journal international du cancer 126, 73-80.

173. Giovannetti, E., Funel, N., Peters, G.J., Del Chiaro, M., Erozenci, L.A., Vasile, E., Leon, L.G., Pollina, L.E., Groen, A., Falcone, A., Danesi, R., Campani, D., Verheul, H.M., Boggi, U., 2010. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer research 70, 4528-4538.

174. Preis, M., Gardner, T.B., Gordon, S.R., Pipas, J.M., Mackenzie, T.A., Klein, E.E., Longnecker, D.S., Gutmann, E.J., Sempere, L.F., Korc, M., 2011. MicroRNA-10b expression correlates with response to neoadjuvant therapy and survival in pancreatic ductal adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 17, 5812-5821.

175. Jamieson, N.B., Morran, D.C., Morton, J.P., Ali, A., Dickson, E.J., Carter, C.R., Sansom, O.J., Evans, T.R., McKay, C.J., Oien, K.A., 2012. MicroRNA molecular profiles associated with diagnosis, clinicopathologic criteria, and overall survival in patients with resectable pancreatic ductal adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 18, 534-545.

176. Yachida, S., Iacobuzio-Donahue, C.A., 2013. Evolution and dynamics of pancreatic cancer progression. Oncogene 32, 5253-5260.

177. Waddell, N., Pajic, M., Patch, A.M., Chang, D.K., Kassahn, K.S., Bailey, P., Johns, A.L., Miller, D., Nones, K., Quek, K., Quinn, M.C., Robertson, A.J., Fadlullah, M.Z., Bruxner, T.J., Christ, A.N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., Nourbakhsh, E., Wani, S., Wilson, P.J., Markham, E., Cloonan, N., Anderson, M.J., Fink, J.L., Holmes, O., Kazakoff, S.H., Leonard, C., Newell, F., Poudel, B., Song, S., Taylor, D., Waddell, N., Wood, S., Xu, Q., Wu, J., Pinese, M., Cowley, M.J., Lee, H.C., Jones, M.D., Nagrial, A.M., Humphris, J., Chantrill, L.A., Chin, V., Steinmann, A.M., Mawson, A., Humphrey, E.S., Colvin, E.K., Chou,

A., Scarlett, C.J., Pinho, A.V., Giry-Laterriere, M., Rooman, I., Samra, J.S., Kench, J.G., Pettitt, J.A., Merrett, N.D., Toon, C., Epari, K., Nguyen, N.Q., Barbour, A., Zeps, N., Jamieson, N.B., Graham, J.S., Niclou, S.P., Bjerkvig, R., Grutzmann, R., Aust, D., Hruban, R.H., Maitra, A., Iacobuzio-Donahue, C.A., Wolfgang, C.L., Morgan, R.A., Lawlor, R.T., Corbo, V., Bassi, C., Falconi, M., Zamboni, G., Tortora, G., Tempero, M.A., Australian Pancreatic Cancer Genome, I., Gill, A.J., Eshleman, J.R., Pilarsky, C., Scarpa, A., Musgrove, E.A., Pearson, J.V., Biankin, A.V., Grimmond, S.M., 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518, 495-501.

178. Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J., Chang, D.K., Cowley, M.J., Gardiner, B.B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C.J., Gill, A.J., Pinho, A.V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J.L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pinese, M., Mead, R.S., Humphris, J.L., Kaplan, W., Jones, M.D., Colvin, E.K., Nagrial, A.M., Humphrey, E.S., Chou, A., Chin, V.T., Chantrill, L.A., Mawson, A., Samra, J.S., Kench, J.G., Lovell, J.A., Daly, R.J., Merrett, N.D., Toon, C., Epari, K., Nguyen, N.Q., Barbour, A., Zeps, N., Australian Pancreatic Cancer Genome, I., Kakkar, N., Zhao, F., Wu, Y.Q., Wang, M., Muzny, D.M., Fisher, W.E., Brunicardi, F.C., Hodges, S.E., Reid, J.G., Drummond, J., Chang, K., Han, Y., Lewis, L.R., Dinh, H., Buhay, C.J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R.E., Yung, C.K., Serra, S., Onetto, N., Mukhopadhyay, D., Tsao, M.S., Shaw, P.A., Petersen, G.M., Gallinger, S., Hruban, R.H., Maitra, A., Iacobuzio-Donahue, C.A., Schulick, R.D., Wolfgang, C.L., Morgan, R.A., Lawlor, R.T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M.A., Mann, K.M., Jenkins, N.A., Perez-Mancera, P.A., Adams, D.J., Largaespada, D.A., Wessels, L.F., Rust, A.G., Stein, L.D., Tuveson, D.A., Copeland, N.G., Musgrove, E.A., Scarpa, A., Eshleman, J.R., Hudson, T.J., Sutherland, R.L., Wheeler, D.A., Pearson, J.V., McPherson, J.D., Gibbs, R.A., Grimmond, S.M., 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399-405.

179. Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.M., Fu, B., Lin, M.T., Calhoun, E.S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D.R., Hidalgo, M., Leach, S.D., Klein, A.P., Jaffee, E.M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J.R., Kern, S.E., Hruban, R.H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W., 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801-1806.

180. Kuuselo, R., Simon, R., Karhu, R., Tennstedt, P., Marx, A.H., Izbicki, J.R., Yekebas, E., Sauter, G., Kallioniemi, A., 2010. 19q13 amplification is associated with high grade and stage in pancreatic cancer. Genes, chromosomes & cancer 49, 569-575.

181. Darwish, H., Cho, J.M., Loignon, M., Alaoui-Jamali, M.A., 2007. Overexpression of SERTAD3, a putative oncogene located within the 19q13 amplicon, induces E2F activity and promotes tumor growth. Oncogene 26, 4319-4328.

182. Bauman, J.E., Austin, M.C., Schmidt, R., Kurland, B.F., Vaezi, A., Hayes, D.N., Mendez, E., Parvathaneni, U., Chai, X., Sampath, S., Martins, R.G., 2013. ERCC1 is a prognostic biomarker in locally advanced head and neck cancer: results from a randomised, phase II trial. British journal of cancer 109, 2096-2105.

183. Yoshida, T., Kobayashi, T., Itoda, M., Muto, T., Miyaguchi, K., Mogushi, K., Shoji, S., Shimokawa, K., Iida, S., Uetake, H., Ishikawa, T., Sugihara, K., Mizushima, H., Tanaka, H., 2010. Clinical omics analysis of colorectal cancer incorporating copy number aberrations and gene expression data. Cancer informatics 9, 147-161.

184. Espinal-Enriquez, J., Munoz-Montero, S., Imaz-Rosshandler, I., Huerta-Verde, A., Mejia, C., Hernandez-Lemus, E., 2015. Genome-wide expression analysis suggests a crucial role of dysregulation of matrix metalloproteinases pathway in undifferentiated thyroid carcinoma. BMC genomics 16, 207.

185. Tiseo, M., Bordi, P., Bortesi, B., Boni, L., Boni, C., Baldini, E., Grossi, F., Recchia, F., Zanelli, F., Fontanini, G., Naldi, N., Campanini, N., Azzoni, C., Bordi, C., Ardizzoni, A., Bio, F.t.g., 2013. ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. British journal of cancer 108, 1695-1703.

186. Donahue, T.R., Tran, L.M., Hill, R., Li, Y., Kovochich, A., Calvopina, J.H., Patel, S.G., Wu, N., Hindoyan, A., Farrell, J.J., Li, X., Dawson, D.W., Wu, H., 2012. Integrative survival-based molecular profiling of human pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 18, 1352-1363.

187. Harada, T., Chelala, C., Bhakta, V., Chaplin, T., Caulee, K., Baril, P., Young, B.D., Lemoine, N.R., 2008. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. Oncogene 27, 1951-1960.

188. Harada, T., Chelala, C., Crnogorac-Jurcevic, T., Lemoine, N.R., 2009. Genome-wide analysis of pancreatic cancer using microarray-based techniques. Pancreatology : official journal of the International Association of Pancreatology 9, 13-24.

189. Tol, J.A., Gouma, D.J., Bassi, C., Dervenis, C., Montorsi, M., Adham, M., Andren-Sandberg, A., Asbun, H.J., Bockhorn, M., Buchler, M.W., Conlon, K.C., Fernandez-Cruz, L., Fingerhut, A., Friess, H., Hartwig, W., Izbicki, J.R., Lillemoe, K.D., Milicevic, M.N., Neoptolemos, J.P., Shrikhande, S.V., Vollmer, C.M., Yeo, C.J., Charnley, R.M., International Study Group on Pancreatic, S., 2014. Definition of a standard lymphadenectomy in surgery for pancreatic ductal adenocarcinoma: a consensus statement by the International Study Group on Pancreatic Surgery (ISGPS). Surgery 156, 591-600.

190. Mollberg, N., Rahbari, N.N., Koch, M., Hartwig, W., Hoeger, Y., Buchler, M.W., Weitz, J., 2011. Arterial resection during pancreatectomy for pancreatic cancer: a systematic review and meta-analysis. Annals of surgery 254, 882-893.

191. Bockhorn, M., Uzunoglu, F.G., Adham, M., Imrie, C., Milicevic, M., Sandberg, A.A., Asbun, H.J., Bassi, C., Buchler, M., Charnley, R.M., Conlon, K., Cruz, L.F., Dervenis, C., Fingerhutt, A., Friess, H., Gouma, D.J., Hartwig, W., Lillemoe, K.D., Montorsi, M., Neoptolemos, J.P., Shrikhande, S.V., Takaori, K., Traverso, W., Vashist, Y.K., Vollmer, C., Yeo, C.J., Izbicki, J.R., International Study Group of Pancreatic, S., 2014. Borderline resectable pancreatic cancer: a consensus statement by the International Study Group of Pancreatic Surgery (ISGPS). Surgery 155, 977-988.

192. Zhang, Y., Yang, P., Wang, X.F., 2014. Microenvironmental regulation of cancer metastasis by miRNAs. Trends in cell biology 24, 153-160.

193. Chou, J., Shahi, P., Werb, Z., 2013. microRNA-mediated regulation of the tumor microenvironment. Cell cycle 12, 3262-3271

194. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., Croce, C.M., 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proceedings of the National Academy of Sciences of the United States of America 101, 2999-3004.

195. Kent, O.A., Mendell, J.T., 2006. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. Oncogene 25, 6188-6196.

196. Franke, T.F., Hornik, C.P., Segev, L., Shostak, G.A., Sugimoto, C., 2003. PI3K/Akt and apoptosis: size matters. Oncogene 22, 8983-8998.

197. Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.

198. Jacobsen, A., Silber, J., Harinath, G., Huse, J.T., Schultz, N., Sander, C., 2013. Analysis of microRNA-target interactions across diverse cancer types. Nature structural & molecular biology 20, 1325-1332.

199. Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., Ljung, B.M., Esserman, L., Albertson, D.G., Waldman, F.M., Gray, J.W., 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer cell 10, 529-541.

200. Castellano, E., Downward, J., 2011. RAS Interaction with PI3K: More Than Just Another Effector Pathway. Genes & cancer 2, 261-274.

201. Locasale, J.W., 2012. Metabolic rewiring drives resistance to targeted cancer therapy. Molecular systems biology 8, 597.

202. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., Smyla, J.K., Anderson, J.M., Kim, H.J., Bentrem, D.J., Talamonti, M.S., Iacobuzio-Donahue, C.A., Hollingsworth, M.A., Yeh, J.J., 2015. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nature genetics 47, 1168-1178.

203. Gutierrez, M.L., Munoz-Bellvis, L., Abad Mdel, M., Bengoechea, O., Gonzalez-Gonzalez, M., Orfao, A., Sayagues, J.M., 2011. Association between genetic subgroups of pancreatic ductal adenocarcinoma defined by high density 500 K SNP-arrays and tumor histopathology. PloS one 6, e22315.

204. Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., Feiler, H.S., Ko, A.H., Olshen, A.B., Danenberg, K.L., Tempero, M.A., Spellman, P.T., Hanahan, D., Gray, J.W., 2011. Subtypes of pancreatic

ductal adenocarcinoma and their differing responses to therapy. Nature medicine 17, 500-503.

205. Sellers, W.R., Kaelin, W.G., Jr., 1997. Role of the retinoblastoma protein in the pathogenesis of human cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 15, 3301-3312.

206. Levine, A.J., 1997. p53, the cellular gatekeeper for growth and division. Cell 88, 323-331.

207. Buenemann, C.L., Willy, C., Buchmann, A., Schmiechen, A., Schwarz, M., 2001. Transforming growth factor-beta1-induced Smad signaling, cell-cycle arrest and apoptosis in hepatoma cells. Carcinogenesis 22, 447-452.

208. Childs, E.J., Mocci, E., Campa, D., Bracci, P.M., Gallinger, S., Goggins, M., Li, D., Neale, R.E., Olson, S.H., Scelo, G., Amundadottir, L.T., Bamlet, W.R., Bijlsma, M.F., Blackford, A., Borges, M., Brennan, P., Brenner, H., Bueno-de-Mesquita, H.B., Canzian, F., Capurso, G., Cavestro, G.M., Chaffee, K.G., Chanock, S.J., Cleary, S.P., Cotterchio, M., Foretova, L., Fuchs, C., Funel, N., Gazouli, M., Hassan, M., Herman, J.M., Holcatova, I., Holly, E.A., Hoover, R.N., Hung, R.J., Janout, V., Key, T.J., Kupcinskas, J., Kurtz, R.C., Landi, S., Lu, L., Malecka-Panas, E., Mambrini, A., Mohelnikova-Duchonova, B., Neoptolemos, J.P., Oberg, A.L., Orlow, I., Pasquali, C., Pezzilli, R., Rizzato, C., Saldia, A., Scarpa, A., Stolzenberg-Solomon, R.Z., Strobel, O., Tavano, F., Vashist, Y.K., Vodicka, P., Wolpin, B.M., Yu, H., Petersen, G.M., Risch, H.A., Klein, A.P., 2015. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nature genetics 47, 911-916.

209. Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk

predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 27, 1160-1167.

210. Conroy, T., Desseigne, F., Ychou, M., Bouche, O., Guimbaud, R., Becouarn, Y., Adenis, A., Raoul, J.L., Gourgou-Bourgade, S., de la Fouchardiere, C., Bennouna, J., Bachet, J.B., Khemissa-Akouz, F., Pere-Verge, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., Ducreux, M., Groupe Tumeurs Digestives of, U., Intergroup, P., 2011. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. The New England journal of medicine 364, 1817-1825.

211. Peterson, L.M., Kipp, B.R., Halling, K.C., Kerr, S.E., Smith, D.I., Distad, T.J., Clayton, A.C., Medeiros, F., 2012. Molecular characterization of endometrial cancer: a correlative study assessing microsatellite instability, MLH1 hypermethylation, DNA mismatch repair protein expression, and PTEN, PIK3CA, KRAS, and BRAF mutation analysis. International journal of gynecological pathology: official journal of the International Society of Gynecological Pathologists 31, 195-205.

212. Hezel, A.F., Kimmelman, A.C., Stanger, B.Z., Bardeesy, N., Depinho, R.A., 2006. Genetics and biology of pancreatic ductal adenocarcinoma. Genes & development 20, 1218-1249.

Paper I

Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas

Sandhu V., Bowitz Lothe I.M., Labori K.J., Lingjaerde O.C., Buanes T., Dalsgaard A.M., Skrede M.L., Hamfjord J., Haaland T., Eide T.J., Borresen-Dale A.L., Ikdahl T., Kure E.H.

Molecular Oncology, 2015, 9, 758-771.

MOLECULAR ONCOLOGY 9 (2015) 758–771

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas

V. Sandhu[a,f,1], I.M. Bowitz Lothe[a,b,1], K.J. Labori[c], O.C. Lingjærde[d], T. Buanes[c,g], A.M. Dalsgaard[a], M.L. Skrede[a], J. Hamfjord[a], T. Haaland[b], T.J. Eide[b,g], A.-L. Børresen-Dale[a,g], T. Ikdahl[e,2], E.H. Kure[a,f,*,2]

[a]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[b]Department of Pathology, Oslo University Hospital, Oslo, Norway
[c]Department of Hepato-Pancreato-Biliary Surgery, Oslo University Hospital, Oslo, Norway
[d]Department of Informatics, University of Oslo, Oslo, Norway
[e]Department of Oncology, Oslo University Hospital, Oslo, Norway
[f]Department of Environmental and Health Studies, Faculty of Arts and Sciences, Telemark University College, Telemark, Norway
[g]Faculty of Medicine, University of Oslo, Oslo, Norway

ARTICLE INFO

Article history:
Received 24 June 2014
Received in revised form
2 December 2014
Accepted 8 December 2014
Available online 19 December 2014

Keywords:
Periampullary adenocarcinoma
Pancreatic cancer
Molecular expression profiling
Prognostic markers
Signaling pathways
Histological subtypes

ABSTRACT

Periampullary adenocarcinomas include four anatomical sites of origin (the pancreatic duct, bile duct, ampulla and duodenum) and most of them fall into two histological subgroups (pancreatobiliary and intestinal). Determining the exact origin of the tumor is sometimes difficult, due to overlapping histopathological characteristics. The prognosis depends on the histological subtype, as well as on the anatomical site of origin, the former being the more important. The molecular basis for these differences in prognosis is poorly understood. Whole-genome analyses were used to investigate the association between molecular tumor profiles, pathogenesis and prognosis. A total of 85 periampullary adenocarcinomas were characterized by mRNA and miRNA expressions profiling. Molecular profiles of the tumors from the different anatomical sites of origin as well as of the different histological subtypes were compared. Differentially expressed mRNAs and miRNAs between the two histopathological subtypes were linked to specific molecular pathways. Six miRNA families were downregulated and four were upregulated in the pancreatobiliary type as compared to the intestinal type ($P < 0.05$). miRNAs and mRNAs associated with improved overall and recurrence free survival for the two histopathological subtypes were identified. For the pancreatobiliary type the genes *ATM*, *PTEN*, *RB1* and the miRNAs miR-592 and miR-497, and for the intestinal type the genes *PDPK1*, *PIK3R2*, *G6PC* and the miRNAs miR-127-3p, miR-377* were linked to enriched pathways and identified as prognostic markers. The molecular signatures identified may in the future guide the clinicians

* *Corresponding author*. Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.
E-mail address: Elin.Kure@rr-research.no (E.H. Kure).
[1] The authors have contributed equally.
[2] Shared senior authorship.
http://dx.doi.org/10.1016/j.molonc.2014.12.002

in the therapeutic decision making to an individualized treatment, if confirmed in other larger datasets.

1. Introduction

Periampullary adenocarcinomas (PAs) can originate from the pancreatic ducts, the distal common bile duct, the ampulla of Vater and the duodenum. Adenocarcinomas of all the four anatomical sites are classified according to the histological appearance and most of them fall into either pancreatobiliary (P) or intestinal (I) type (Albores-Saavedra et al., 2007; Westgaard et al., 2008). The adenocarcinomas with origin in the ampulla, at the intersection of the pancreas, the common bile duct and the duodenum, are those that typically can exhibit either an intestinal or pancreatobiliary appearance or a mixture of the two. Due to periampullary growth and overlapping histological phenotypes, accurate histological classification and anatomical base of the tumor may be difficult to ascertain. The different subgroups have similar clinical presentation, and are resected by pancreatoduodenectomy with curative intent. According to current national guidelines, only patients with pancreatic and duodenal adenocarcinomas are treated with adjuvant chemotherapy. Despite advances in surgery, radiotherapy and chemotherapy, the prognosis is poor. Pancreatic ductal adenocarcinoma (PDAC) is the most common and the most aggressive type, having a 5-year overall survival of 20% following treatment with curative intent (Neoptolomos et al., 2010, Oettle et al., 2013, Winter et al., 2012). Adjuvant chemotherapy on PAs has proven to have an additional effect of survival compared to surgery alone (Neoptolemos et al., 2012, 2010, 2004, Oettle et al., 2007). Interestingly, there is evidence that histological subtype may be of even stronger prognostic impact than the anatomical site of origin (Westgaard et al., 2008). The patients suffering from adenocarcinoma of intestinal type do better than patients with adenocarcinoma of pancreatobiliary type (Albores-Saavedra et al., 2007, Westgaard et al., 2008, 2013). Patients with tumors of the same histological subtype, but from different anatomical site of origin, have similar overall survival (Westgaard et al., 2013). Due to the differences in pathogenesis and survival, accurate histological classification of the periampullary adenocarcinomas is crucial for an adequate prognostic estimation and an individualized therapeutic decision. Molecular classification of the tumor may contribute to individualized treatment of the patients.

Based upon the analyses of small series of PDACs, molecular subtypes for this anatomical site have been proposed (Collisson et al., 2011; Donahue et al., 2012, Overman et al., 2013). PAs originating from other sites than the pancreatic ducts have not yet been profiled by both miRNAs and mRNAs expression although an integrative platform for miRNA, mRNA and proteins has recently been published (Thomas et al., 2014). The aim of the present study was to gain molecular insight into the pathogenesis of PAs by identifying miRNAs and mRNAs expression profiles of tumors from the different anatomical sites of origin, as well as of the two main histological subtypes and to relate their significance to survival. Pathway-based strategies were applied to identify dysregulated pathways specific for the various molecular subtypes identified of PAs. Identification of dysregulated mRNAs and miRNAs in PAs associated with better overall and recurrence free survival may help understand the pathogenesis of the disease and identify potential drug targets.

2. Materials and methods

2.1. *Patients and specimens*

A total of 85 patients were included in the present investigation, 41 (48.2%) males and 44 (51.8%) females, with median age of 67 years (range 34–84 years). Fresh frozen tumor and adjacent normal tissue were collected from patients with adenocarcinomas admitted to Oslo University Hospital (2008–2011) for pancreatoduodenectomy with curative intent. The macroscopic and microscopic pathology work followed a standardized protocol. The diagnoses were independently verified according to the WHO Classification of Tumors of the Digestive System (Bosman et al., 2010), by two experienced pathologists. Only ductal adenocarcinomas of either pancreatobiliary or intestinal type, as first described by Kimura and colleagues (Kimura et al., 1994), were included in the project. The adenocarcinomas were classified according to site of origin, histological subgroups and tumor stage, in accordance with the pTNM Classification of Malignant Tumors (Sobin et al., 2009). Additionally, the resection margins, perineural and vascular tumor invasion were recorded (Table 1). Tumors ($n = 4$) that were of mixed pancreatobiliary and intestinal types were classified according to the WHO classification as the dominant subtype. Molecular analysis was performed on frozen tissue samples, and the diagnoses were set separately on this material, to ensure that the morphology matched the original diagnoses on the formalin fixed paraffin embedded (FFPE) material. Out of a total of 114 patients with adenocarcinomas, 85 patients with tumors in the periampullary region met the criteria for inclusion in this project, 68 pancreatobiliary tumors and 17 intestinal tumors (Figure 1). The remaining 29 tumors were excluded from the study due to variant type adenocarcinomas, metastatic adenocarcinomas, or low percentage of estimated tumor cells of the prepared specimen ($< 10\%$).

The characteristics of the 85 PAs are presented in Table 1. A total of 52 PDACs were used in the analyses, of which 49

Table 1 – Pathological characteristics of tumors with reference to site of origin and histopathology. Number of mRNA and miRNA samples from PDAC (P) (n = 49, n = 52), bile duct (P) (n = 8, n = 8), ampulla (P) (n = 8, n = 8), ampulla (I) (n = 7, n = 7), Duodenum (I) (n = 9, n = 9), normal (n = 12, n = 6).

Periampullary adenocarcinomas total number = 85		PDAC (P)	Bile duct (P)	Ampulla (P)	Ampulla (I)	Duodenum (I)
		Total (P) number, n = 68			Total (I) number, n = 17	
		Number (%)	Number (%)	Number (%)	Number (%)	Number (%)
Histopathology of PAs	Pancreato-biliary	52 (76.5)	8 (11.8)	8 (11.8)	0	0
	Intestinal	0	0	0	7 (41.2)	10 (58.8)
Differentiation grade of tumors	Poor	20 (38.4)	4 (50)	4 (50)	0	2 (20)
	Moderate	31 (59.6)	4 (50)	4 (50)	7 (100)	8 (80)
	Well	1 (1.9)	0	0	0	0
pT	T1	4 (7.7)	0	0	0	1 (10)
	T2	5 (9.6)	0	0	6 (85.7)	2 (20)
	T3	43 (82.7)	8 (100)	5 (62.5)	1 (14.3)	3 (30)
	T4	0	0	3 (37.5)	0	4 (40)
N	N0	12 (23)	4 (50)	5 (62.5)	4 (57.1)	4 (40)
	N1	40 (76.9)	4 (50)	3 (37.5)	3 (42.8)	5 (50)
	N2	–	–	–	–	1 (10)
M	M0	50 (96.1)	8 (100)	8 (100)	7 (100)	8 (80)
	M1	2 (3.8)	0	0	0	2 (20)
R	R0	22 (42.3)	6 (80)	7 (90)	7 (100)	9 (90)
	R1	30 (57.6)	2 (20)	1 (10)	0	1 (10)
Vessel infiltration	Yes	26 (50)	4 (50)	5 (62.5)	2 (28.5)	7 (70)
	No	26 (50)	4 (50)	3 (37.5)	5 (71.4)	3 (30)
Perineural infiltration	Yes	51 (98)	7 (87.5)	7 (87.5)	1 (14.28)	3 (30)
	No	1 (2)	1 (12.5)	1 (12.5)	6 (85.7)	7 (70)
***KRAS* mutations in codons 12 and 13**	Mutated	45/52 (87)	6/8 (75)	6/8 (75)	2/7 (29)	6/10 (60)
	Wild type	7/52 (13)	2/8 (25)	2/8 (25)	5/7 (71)	4/10 (40)

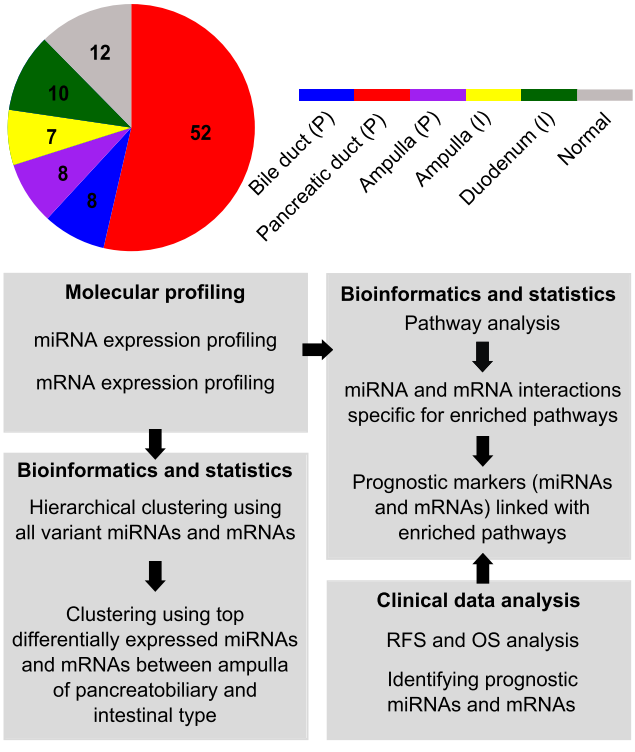


Figure 1 – The analysis pipeline. The coloring of the circle corresponds to the five different subgroups and normal tissue samples.

samples were analyzed for mRNAs and 51 for miRNAs, with 48 samples analyzed for expression of both mRNAs and miRNAs. Ten adenocarcinomas from the duodenum were examined; nine samples were profiled for miRNAs and nine for mRNAs, of which eight samples were in common for miRNAs and mRNAs. The sample distribution is shown in Figure 1. A total of 65 and 67 pancreatobiliary samples were used in the analysis of mRNAs and miRNAs, respectively. Of the intestinal samples, 16 were analyzed for both mRNAs and miRNAs. The study was approved by the Regional Ethics Committee, and all patients included had given their signed informed consent.

2.2. *Specimen preparation and RNA purification*

The tumor samples available were sectioned and stained with HE in preset intervals. The tumor cells content was estimated as a fraction of the whole HE section (10 × 20 mm) by a trained pathologist. We selected the section with the highest estimated tumor percentage. Total RNA was isolated from whole sections of fresh frozen tumor (n = 85) and adjacent normal tissues from patients with pancreatitis (n = 12) according to the manufacturer's instructions (mirVana miRNA Isolation Kit Ambion/Life Technologies). The RNA concentration was measured using a Nanodrop ND-1000 spectrophotometer, and the quality assessed on a Bioanalyzer 2100 (Agilent). The RIN-values ranged from 2.0 to 9.9 with 16 of 85 samples below 5.0.

2.3. mRNA expression analysis, pre-processing and normalization

A total of 100 ng total RNA was converted to cDNA, amplified and labeled with Cy-3 using LowInput QuickAmp Labeling Kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. After cleaning of the samples on RNeasy Mini Columns (Qiagen) the efficiency of the labeling reaction labeled cRNA was measured using Nanodrop ND-1000. 600 ng Cy-3 labeled cRNA was hybridized to SurePrint G3 Human GE 8x60K microarrays containing probes for 42,545 individual mRNAs for 17 h at 65 °C using Hi-RPM Gene Expression Hybridization Kit (Agilent Technologies, Santa Clara, CA, USA). The arrays were washed according to the manufacturer's instructions, using Gene Expression Wash Buffer Kit (Agilent Technologies, Santa Clara, CA, USA), and scanned on an Agilent DNA Microarray Scanner. Hybridization signals were extracted using Feature Extraction 10.7.3.1 (Agilent Technologies, Santa Clara, CA, USA). The mRNA microarray data was background corrected and quantile normalized. Filtering of controls and low expressed probes was performed by calculating the 95% percentile of the negative control probes on each array, and probes that were at least 10% brighter than the negative controls on at least 50% of the arrays were included in the further analysis.

2.4. miRNA expression analysis, pre-processing and normalization

Some of the samples were vacuum centrifuged to achieve the required concentration of 50 ng/μl. 200 ng total RNA was dephosphorylated and labeled by Cyanine-3-pCp using miRNA Complete Labeling and Hybridization kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The labeled total RNA was cleaned using Micro Bio-Spin 6 columns to remove DMSO contamination and unincorporated Cyanine-3-pCp, vacuum centrifuged until dryness and dissolved in RNase-free H_2O. Samples were hybridized to SurePrint G3 Human v16 miRNA 8x60K microarrays containing probes for 1368 miRNAs for 20 h at 55 °C. The arrays were washed according to the manufacturer's instructions, using Gene Expression Wash Buffer Kit (Agilent Technologies, Santa Clara, CA, USA), and scanned on an Agilent DNA Microarray Scanner. Hybridization signals were extracted using Feature Extraction 10.7.3.1 (Agilent Technologies, Santa Clara, CA, USA). The miRNA microarray data was normalized using Robust Multiarray Average approach and undetected probes were flagged and filtered if not detected in 50% of the array replicates using AgimicroRNA package Version 2.10.0 in R (Lopez-Romero, 2011). The data is accessible through GEO accession number GSE60980.

2.5. KRAS mutational analysis

Tumor DNA was screened for the presence of seven *KRAS* mutations in codons 12 and 13; (*KRAS* g.34G>C (p.G12R), g.34G>A (p.G12S), g.34G>T (p.G12C), g.35G>A (p.G12D), g.35G>C (p.G12A), g.35G>T (p.G12V), g.38G>A (p.G13D)) as previously described (Hamfjord et al., 2011).

2.6. Bioinformatical and statistical analysis

2.6.1. Clustering and differential expression analysis of the five subgroups

Hierarchical clustering was performed on the PAs (n = 85) from the different sites of origin. All the intrinsically variable mRNAs (s.d. > 0.8) and miRNAs (s.d. > 0.5) were used in the analysis. Thus, 7516 mRNAs and 246 miRNAs from the microarray data sets were used for cluster analysis. Complete linkage together with Pearson's correlation coefficient and Spearman's rank correlation for genes and samples were used, respectively (Berry et al., 2010). The grouping of samples was based on molecular profiles independent of morphology or any clinical features.

Hierarchical clustering based on the histological subtypes (i.e. pancreatobiliary (n = 68) and intestinal (n = 17)) were performed separately using the intrinsically variable mRNAs (s.d. > 0.8) and miRNAs (s.d. > 0.5) using Pearson's correlation coefficient and Spearman's rank correlation for genes and samples, respectively.

The moderated t-test (modT) (Smyth et al., 2004, 2005) was carried out to identify differentially expressed mRNAs between ampullary adenocarcinomas of pancreatobiliary (ampulla (P)) and intestinal type (ampulla (I)) at adjusted $P < 0.05$ (Benjamini and Hochberg, 1995). Using the same procedure, the differentially expressed miRNAs between ampulla (P) and ampulla (I) were identified. The analysis was carried out using the Bioconductor package limma in R (Gentleman et al., 2004). The differentially expressed miRNAs and mRNAs were used for hierarchical clustering of all the PAs using complete linkage and Pearson's correlation coefficient and Spearman's rank correlation for genes and samples, respectively.

2.6.2. Survival analysis

Survival analysis was performed using the Kaplan–Meier estimator as implemented in the KMsurv package (Therneau and Grambsch, 2000) and the log-rank test. Overall survival (OS) time was calculated from date of surgery to time of death. OS data were obtained from the National Population Registry in Norway. Four patients with distant metastasis (M1) at time of resection, and one patient that died from cardiac arrest were not included in the survival analysis. Recurrence free survival (RFS) time was calculated from date of surgery to date of recurrence of disease. Recurrence was defined as radiological evidence of intra-abdominal soft tissue around the surgical site or of distant metastasis.

To identify potential prognostic markers (i.e. miRNAs and mRNAs) for pancreatobiliary and intestinal adenocarcinomas, Cox regression analysis was performed. The results were confirmed by the log-rank test ($P < 0.05$). The Kaplan–Meier survival curves for the identified prognostic markers were plotted. The expression for each sample was designated as high if the expression was higher than the 75th percentile expression value within each histopathological subgroup, otherwise as low.

2.6.3. Pathway analysis

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was used to identify deregulated pathways in the PAs.

GSEA is a statistical tool for microarray data analysis based on a modified Kolmogorov–Smirnov statistics. It ranks all the genes based on the expression ratio between the normal and the tumor samples and determines whether the high-ranking genes are enriched with genes in a specified pathway. GSEA was used to identify pathways enriched in the five subgroups of the PAs (i.e. from each site of origin including its histopathological subgroup) compared to the normal tissue samples. All gene sets were selected from four different sources: BioCarta, KEGG, Pathway Interaction Database and Reactome as documented in MSigDb version v4.0.

Enrichment score, nominal P value and FDR for each pathway enriched in the five subgroups using GSEA were calculated. Z-scores obtained from the P values and FDR scores were plotted using JavaTree view (Saldanha, 2004).

2.6.4. *Pathway based interaction of miRNAs and mRNAs*

All the validated interactions curated from miRwalk database (Dweep et al., 2011) of differentially expressed miRNAs between the pancreatobiliary and intestinal adenocarcinomas, and statistically enriched mRNAs from GSEA analysis for each pathway were extracted. The prognostic miRNA and mRNA markers were linked to the enriched pathways.

3. Results

An overview of the analyses carried out on the PAs using molecular profiling data of miRNAs and mRNAs, clinical data and *KRAS* mutation status is summarized in Figure 1.

3.1. *mRNA expression data analysis of all PAs*

Unsupervised hierarchical clustering was performed using all the intrinsically variable mRNAs (i.e. 7516 mRNA) having the s.d. $>$ 0.8 for all the 85 PAs. The heatmap (Figure 2) shows three main clusters, where 11 of the 16 profiled tumors with intestinal subtype cluster together. In addition to the main cluster, consisting of pancreatobiliary (35/65) samples, the normal samples form a cluster together with seven samples from the PDACs, three from the bile duct and two with intestinal subtype from the ampulla. Three of the pancreatobiliary samples from the ampulla cluster with the intestinal samples. No difference in survival was observed between the three main clusters of the heatmap (data not shown). The robustness of clustering of samples based on morphological subtype was tested by varying the threshold, inclusion and exclusion of samples, and by varying the clustering parameters.

To identify clustering patterns independent of histopathology, separate hierarchical clustering was carried out for all samples using the most intrinsically variant genes for intestinal ((n = 16) (s.d $>$ 0.8) i.e. 4422 mRNAs), and pancreatobiliary ((n = 65) (s.d $>$ 0.8) i.e. 4249 mRNAs) subtypes, respectively. In both figures Figures S.1a and S.1b two main clusters are observed. In S.1a the pancreatobiliary samples cluster together in one major cluster and the normal samples cluster together with a few of the pancreatobiliary samples with low tumor content. In S.1b the intestinal samples form one major cluster, and the normal samples cluster together in a separate cluster with a few intestinal samples of low tumor content. Interestingly, none of the histological subtypes form clusters based on site of origin. Despite the small sample size in each subgroup except for the PDACs, a high significance is associated with molecular profile clusters and morphological subtypes at P value $<$ 0.001 from both hypergeometric distribution test and Fischer's exact test, indicating that clustering is not by chance but has a biological significance. The relative association between molecular profiles and site of origin was studied, and we could not find any strong association between molecular profiles and site of origin.

The moderated t-test was carried out between the two histological subtypes from the ampullary site of origin, (ampulla (P) n = 8 and ampulla (I) n = 7) to identify differentially expressed genes with P $<$ 0.05 (Supplementary Table S.1a). The 360 mRNAs, differentially expressed between ampulla (P) and ampulla (I) were used to cluster all the PAs (Figure 3). The intestinal samples from ampulla (5/7) and duodenum (4/9) clustered together and separately from the pancreatobiliary samples.

The moderated t-test was carried out between all the pancreatobiliary and the intestinal samples to identify differentially expressed genes with FDR adjusted P $<$ 0.05 (Supplementary Table S.2a). In total 374 and 173 genes were upregulated in the pancreatobiliary and the intestinal type, respectively. Genes that are differentially expressed (upregulated) in the pancreatobiliary type are transcriptionally regulated by the *WNT3A*, *TGFB1*, *HDAC*, *BDNF* and *ERBB4*. In the intestinal type the differentially expressed (upregulated) genes are transcriptionally regulated by *CDKN2A*, *RB1*, *PPARA*. These genes were identified using functional analysis performed in the Ingenuity Pathway Analysis tool (Ingenuity Systems) by using Fisher's exact test (P $<$ 0.000001) (Supplementary Tables S.3a and S.3b). A list of the genes that were identified both in the comparison between pancreatobiliary versus intestinal types and in the comparison between the ampulla (P) versus ampulla (I) types is provided in Supplementary Table S.4.

3.2. *miRNA expression data analysis of all PAs*

Hierarchical clustering was performed using intrinsically variable miRNAs having s.d. $>$ 0.5 for all the PAs (i.e. 256 miRNAs). The heatmap shows miRNAs distributing the samples into three major clusters (Figure 4). The first cluster includes the majority of the pancreatobiliary samples, the second cluster with more than half of the intestinal samples (9/16) including three ampulla (P) samples. The third cluster contains the normal samples and a few samples with low tumor content. The same distribution was observed for the mRNA clustering of three ampulla (P) samples. At both the miRNA and mRNA level clear clusters of pancreatobiliary and intestinal samples were observed.

To identify clustering patterns independent of histopathological subtype, separate hierarchical clustering was performed of pancreatobiliary and intestinal samples. The most variant miRNAs i.e. 508 for the intestinal and 498 for the pancreatobiliary type were used, respectively (Supplementary Figures S.2a and S.2b). In Figure S.2a the pancreatobiliary subtype from the ampulla (5/8 samples) cluster together, while PDAC and bile duct samples are intermixed. The normal

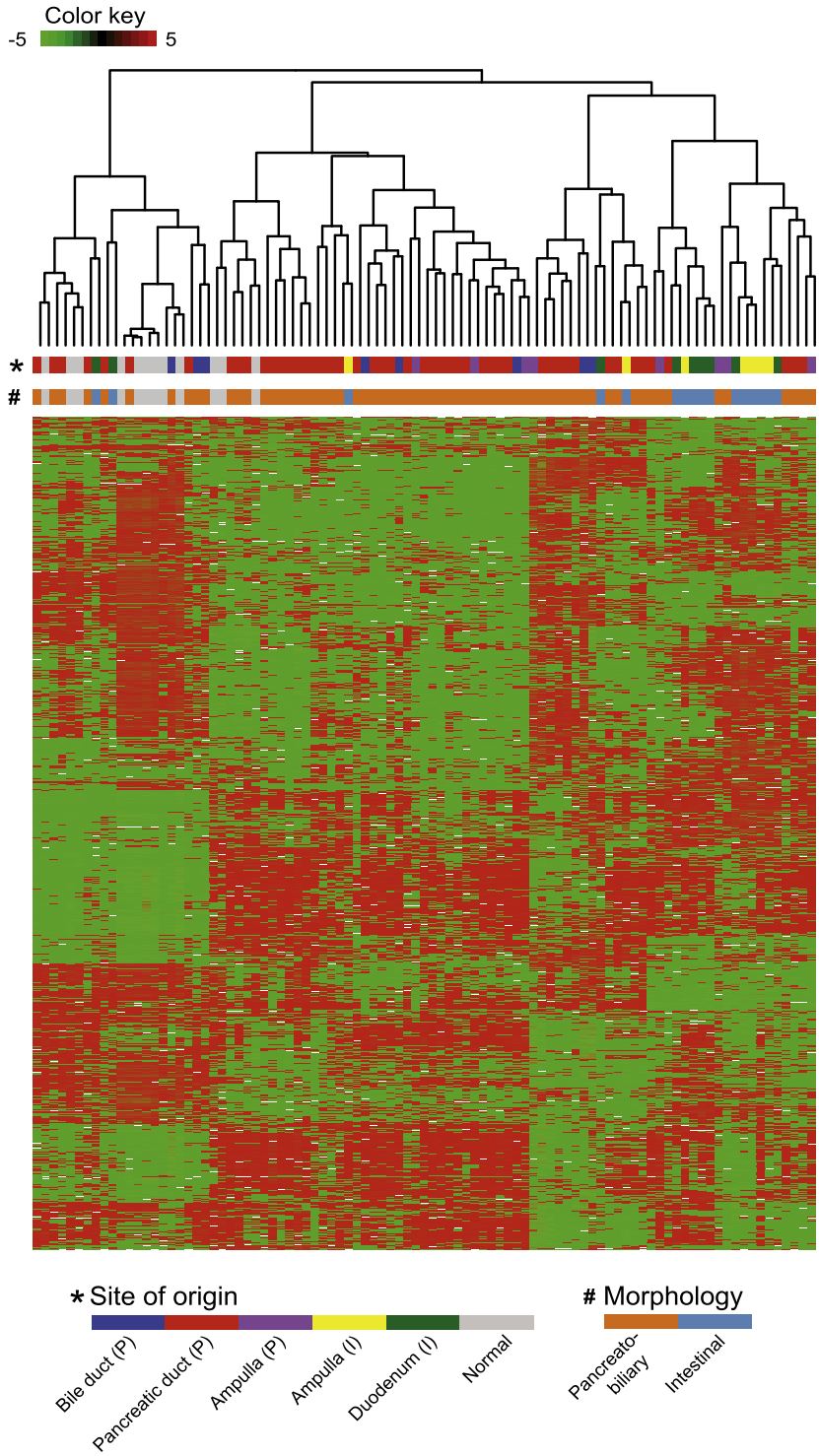


Figure 2 – Hierarchical clustering of the 85 PAs, using the 7516 most variable mRNAs across the dataset.

samples form a separate cluster. In Figure S.2b, two major clusters are observed, one with the normal samples and one with the intestinal samples. The samples from ampulla and duodenum do not cluster with respect to site of origin.

The moderated t-test was carried out between ampulla (P) and ampulla (I) to identify differentially expressed miRNAs with $P < 0.05$, (Supplementary Table S.1b). A total of 35 miRNAs differentially expressed between ampulla (P) and ampulla (I) were used for clustering all the PAs (Figure 5). The pancreatobiliary samples from the PDAC, bile duct and ampulla do not cluster with respect to site of origin. Seven of the intestinal samples form a separate cluster, while three samples cluster with the normal samples. This is most likely due to the lower tumor content in these samples.

The moderated t-test was carried out to identify differentially expressed miRNAs between the pancreatobiliary and

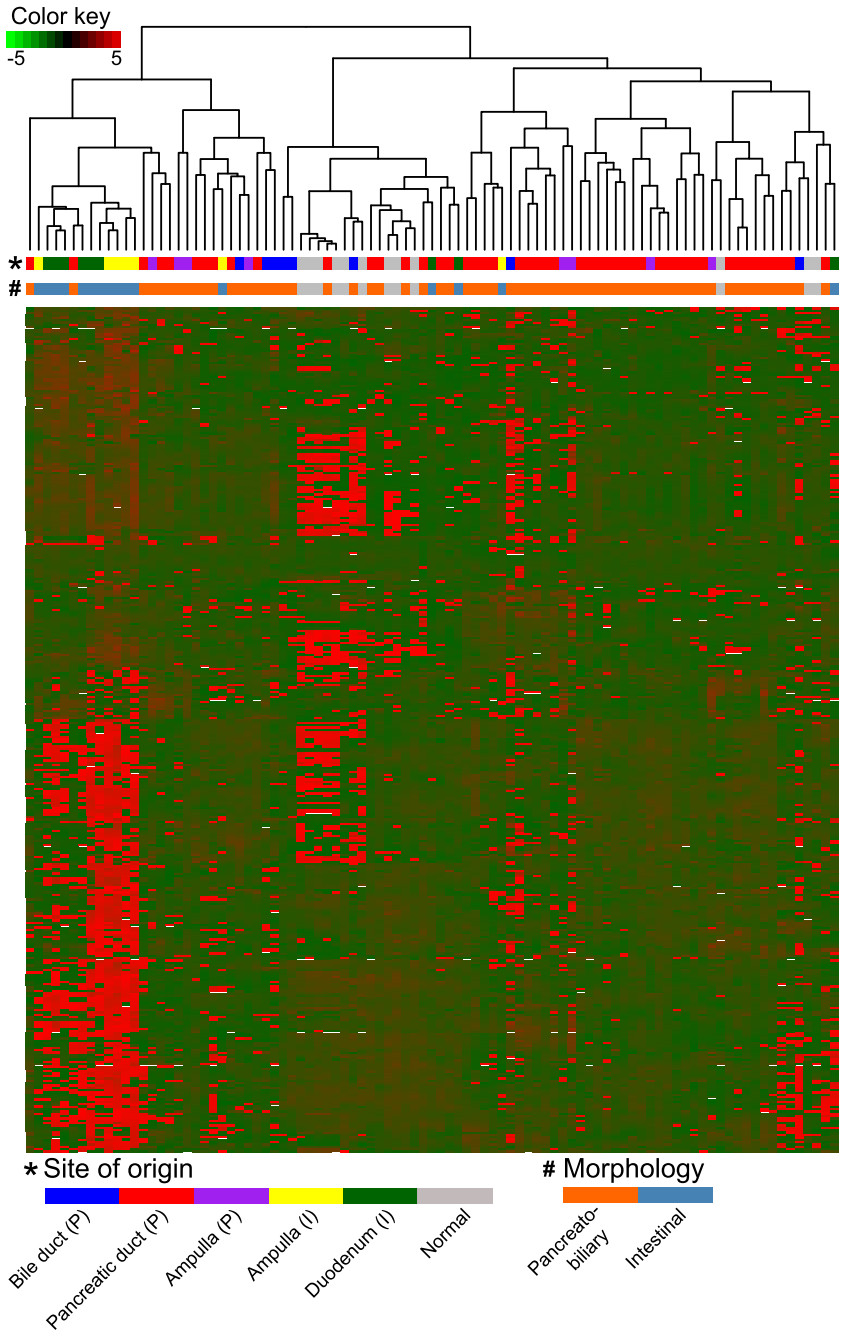


Figure 3 – Hierarchical clustering of the PA samples based on differentially expressed mRNAs (n = 360) between eight samples from the ampulla (P) and seven samples from the ampulla (I).

intestinal samples with $P < 0.05$ (Supplementary Table S.2b). Six miRNA families are downregulated in the samples with pancreatobiliary histopathology compared to the intestinal type (miR-17, miR-196, miR-192, miR-194, miR-19 and miR-548), while four miRNA families are upregulated in the pancreatobiliary type compared to the intestinal one (miR-154, miR-99, miR-329 and miR-199); see Supplementary Table S.5. Both in the pancreatobiliary versus intestinal samples, and the ampulla (P) versus ampulla (I) comparison, the miR-18a, miR-18b, miR-196a, miR-196b, miR-203, miR-20a, miR-378* and miR-625 are all differentially expressed.

3.3. *Survival analysis*

Median RFS for patients with PDAC (P) is 10 months, 16.5 months for patients with distal bile duct adenocarcinoma (P), 20.5 months for patient with ampullary adenocarcinoma (P), 21 months for patients with duodenal adenocarcinoma (I) and 39 months for patients with ampullary adenocarcinoma (I) (Figure 6). A similar trend was observed for OS (Supplementary Figure S.3) with P value = 0.0015. For patients with the intestinal subtype the median RFS was 28 months and the median OS was 31 months, while for patients with the

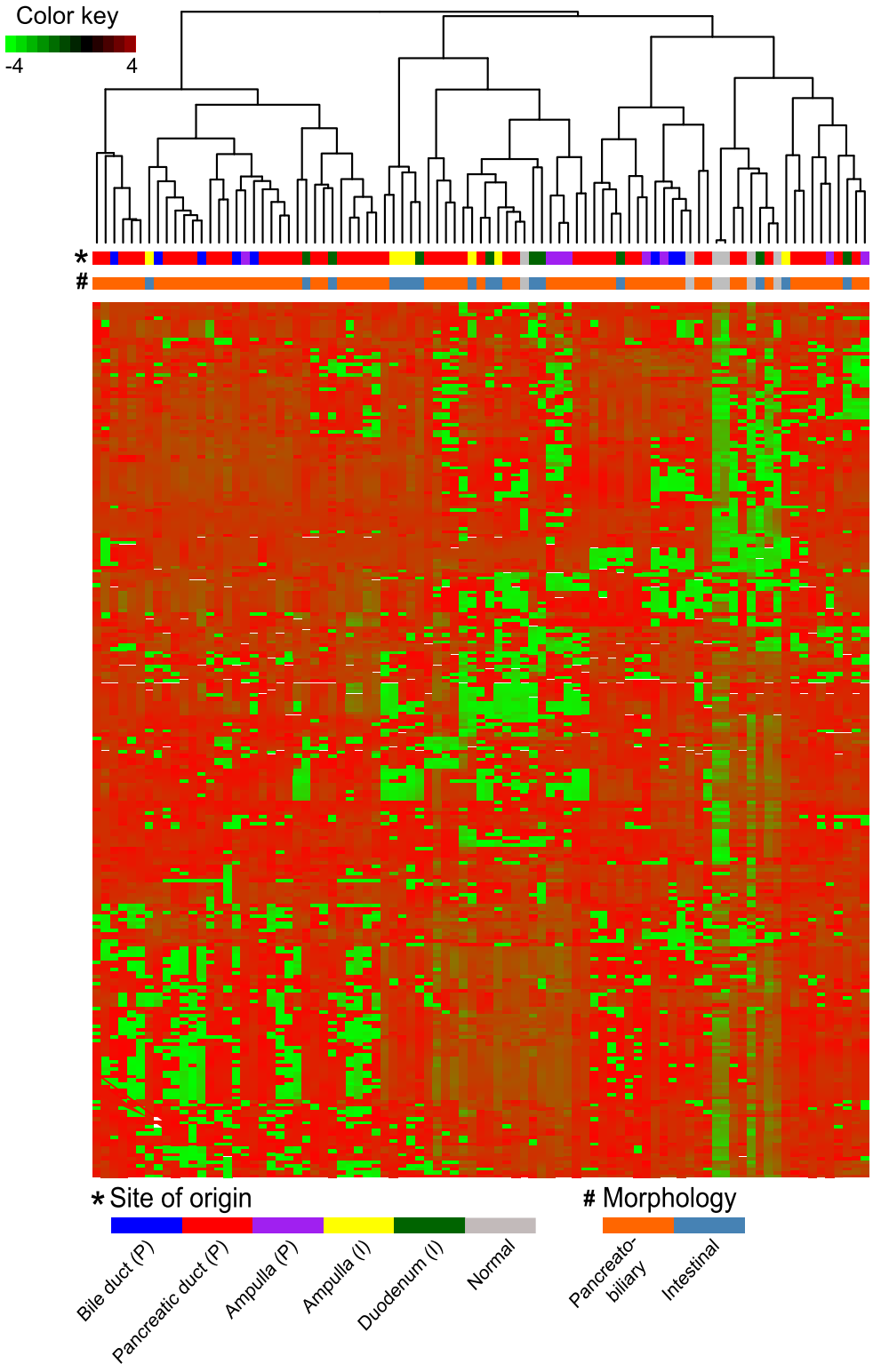


Figure 4 – Hierarchical clustering of the PAs (n = 85) based on intrinsically variable miRNAs (n = 256).

pancreatobiliary subtype the median RFS was 12.5 months and the median OS was 20 months with P value = 0.01.

None of the three major clusters observed in the heatmap of Figures 2 and 3 had any significant differences in survival.

3.4. *Pathway analysis*

Gene set enrichment analysis (GSEA) was used to identify the significantly enriched pathways among the differentially expressed genes (mRNAs) from each of the five subgroups

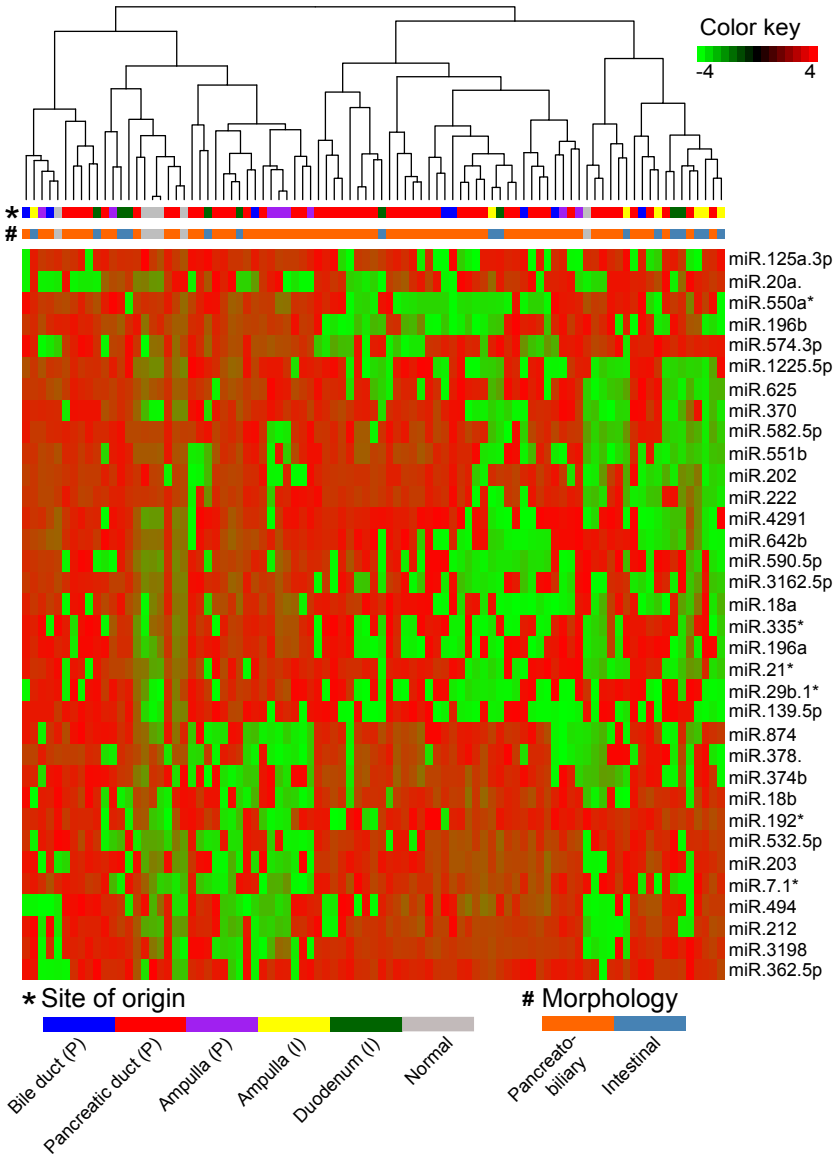


Figure 5 – Hierarchical clustering of the PAs (n = 85) based on differentially expressed miRNAs (n = 35) between the ampulla (P) and the ampulla (I).

(Figure 7) (Subramanian et al., 2005). Normal tissues were used as a reference for comparisons with each subgroup. A total of 15 pathways were significantly enriched in at least one of the five subgroups at $P < 0.05$ and with a false discovery rate FDR < 0.25. The Wnt signaling and interferon gamma signaling pathways were enriched only in the PAs with pancreatobiliary histopathology. The glucose metabolism, glycolysis, phosphatidylinositol signaling and DNA repair signaling pathways are all enriched only in the PAs of the intestinal type. Both the insulin receptor recycling and insulin receptor signaling cascades are downregulated in the intestinal PAs and unaffected in the pancreatobiliary type. Pathways like ATM, ATR, PPAR signaling, p73, RB1, p53 and E2F, are affected in all subgroups, but at different relative levels. Supplementary Table S.6 gives the enrichment score, nominal P-value and FDR value for each pathway enriched in the five subgroups using GSEA. The Z-scores obtained from the P values and FDR score were plotted using the JavaTree view (Saldanha, 2004).

3.5. *Cox regression to identify prognostic markers in patients with adenocarcinomas of pancreatobiliary and intestinal type*

The Cox regression survival analysis linked better OS or RFS with upregulation or downregulation of 302 and 274 mRNAs in patients with pancreatobiliary and intestinal histopathology, respectively (Supplementary Tables S.7a and S.7b). From these genes, upregulation of four tumor suppressor genes (*PTEN*, *RB1*, *ATM and KANK1*) are linked to better OS or RFS. The downregulation of three oncogenes *PAX5*, *PTTG2* and *JAK3* are linked with better OS or RFS in patients with adenocarcinomas of pancreatobiliary type. For patients with adenocarcinoma of intestinal type, downregulation of four

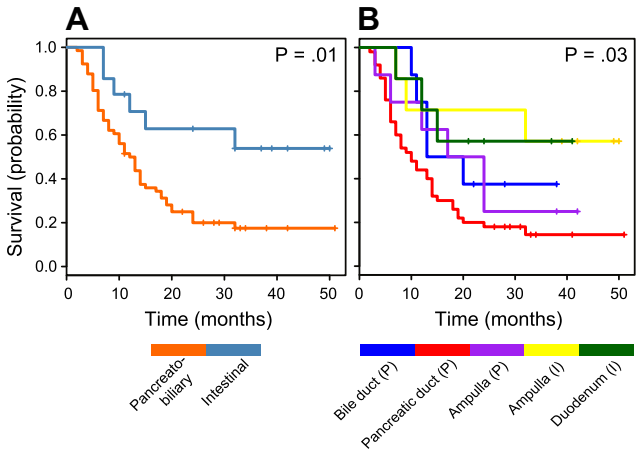


Figure 6 – A. Kaplan–Meier curve demonstrating recurrence free survival (RFS) for patients with intestinal (I) versus pancreatobiliary (P) subtype of adenocarcinomas. B. Kaplan–Meier curve demonstrating RFS for patients in the five subgroups of PAs.

oncogenes (*ELL*, *PDGFB*, *PIM1* and *AR*) and upregulation of a tumor suppressor gene (*PYCARD*) are linked to better OS or RFS.

Downregulation and upregulation of miR-196b and miR-592 respectively, are linked with both better RFS and OS at $P < 0.05$ for Cox regression test analysis with a significant log-rank $P < 0.05$ for pancreatobiliary samples (See Materials and Methods). Upregulation of miR-95 and downregulation of miR-497 are linked to better RFS and OS for pancreatobiliary samples, respectively (Supplementary Table T.8a).

Downregulation of nine miRNAs (miR-193b*, miR-493*, miR-450a, miR-365, miR-654-3p, miR-424, miR-382, miR-127-3p and miR-485-3p) is linked to better RFS, and downregulation of two miRNAs (miR-455-3p and miR-335) is linked to better OS for intestinal samples (Supplementary Table T.8b).

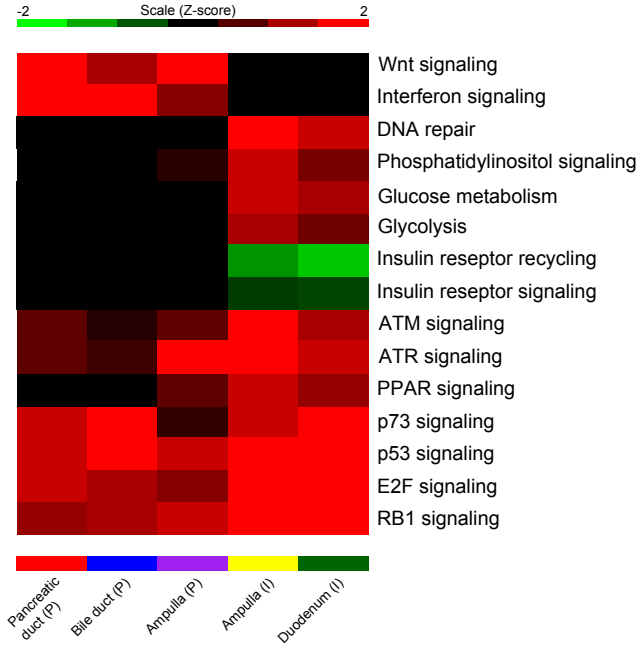


Figure 7 – The heatmap based on mRNA expressions contains enrichment P values transformed to Z-scores for the five subtypes of PAs. Red signifies maximal expression (upregulation), green signifies minimal expression (downregulation) and black shows no affect in the pathway regulation.

3.6. *Pathway based interaction of mRNA-miRNA*

Validated interactions linked with specific pathways (i.e. listed in Figure 7) from the miRwalk database (Dweep et al., 2011) were identified between differentially expressed miRNAs in the pancreatobiliary versus the intestinal sample comparison, and the genes involved in the enriched pathways were identified. The miRNA families deregulated in each of the pathways, and miRNA and mRNA expressions linked to better OS or RFS specific for each pathway are shown in Table 2. All pathway specific interactions between miRNAs and mRNAs are documented in Supplementary Table S9.

3.7. KRAS *mutation analysis*

We found a significantly higher frequency of *KRAS* mutations (84%) in the pancreatobiliary type compared to the intestinal type (47%). The *KRAS* mutational frequencies at the different sites of origin were; PDAC (P) 87%, bile duct (P) 75%, duodenum (I) 60%, ampulla (P) 75% and ampulla (I) 29% as presented in Table 1. The mutational status of *KRAS* was not significantly related to RFS and OS for any of the subgroups.

4. Discussion

The present study is the first multilevel whole-genome analysis of PAs, with results that provide new insights into current molecular knowledge of these tumors. A limitation of our study is small sample size for each group except PDAC. The imbalance in sample size between the tumor subtypes and tumor origin is a weakness. However the hierarchical clustering, both at the mRNA and miRNA level, identified clusters based on histological subtype and is supported by statistical tests. Separated hierarchical clustering did not identify site of origin specific gene expression patterns, but was distinct in the two main histological types. No significant correlation was observed between molecular subtypes and site of origin. A larger sample size in each group is required for validating the correlation.

The result of unsupervised hierarchical clustering showed three main clusters, one cluster for the intestinal type, one for the pancreatobiliary type and one for the normal samples. However, some of the tumors from both subtypes clustered together with the normal samples. Our attempt to lower the sample selection bias may have been too liberal when including the samples with tumor cells as low as 10%. This can explain why some of the tumors from both subtypes clustered together with the normal samples. This problem with low tumor content is greatest in the pancreatobiliary type carcinomas, which characteristically grow in a highly dispersed fashion. Another well-known problem that may explain the

Table 2 – The number of interactions between miRNAs and mRNAs, miRNA families and the miRNAs and mRNAs (i.e., genes) linked to better RFS and OS with reference to enriched pathways.

Pathways	Interactions (miRNA-mRNA)	Pancreatobiliary prognostic markers (miRNA and mRNA)	Intestinal prognostic markers (miRNA and mRNA)	miRNA family
Insulin regulation and recycling	175 (40–45)	*CALM1*, *PCK2*, *SLC2A4*, *PDPK1*, *PPARGC1A*, *PIK3R1*, *PHKB*, hsa-miR-196b, hsa-miR-497	*FBP1*, *G6PC*, *PCK2*, *PIK3R2*, *PDPK1*, hsa-miR-424, hsa-miR-382, hsa-miR-127-3p, hsa-miR-193b*, hsa-miR-365, hsa-miR-455-3p, hsa-miR-335	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-329
Wnt signaling	163 (40–38)	*LRP6*, *CXXC4*, *PPARD*, *CCND2*, hsa-miR-196b, hsa-miR-497, has-miR-592	*WNT5A*, *CAMK2G*, *VANGL1*, *NFATC4*, hsa-miR-455-3p, hsa-miR-335, hsa-miR-193b*, hsa-miR-365, hsa-miR-424, hsa-miR-127-3p	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154
E2F	162 (43–27)	*RB1*, *ATM*,*E2F7*	*E2B2*, *CEBPA*, hsa-miR-193b*, hsa-miR-365, hsa-miR-654-3p, hsa-miR-424, hsa-miR-127-3p, hsa-miR-377*, hsa-miR-455-3p, hsa-miR-335	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154
PPAR	150 (44–38)	*PIK3R1*, *RB1*, *RXRA*, *PCK2*, *PDPK1*, *PPARD*, *PTGS2*		miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154, miR-329
p73	117 (43–23)	*BBC3*, *WT1*, *GATA1*, *RB1*, *ITCH*	hsa-miR-193b*, hsa-miR-654-3p, hsa-miR-424, hsa-miR-127-3p, hsa-miR-335	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154, miR-329
ATM	107 (43–12)	*ATM*, hsa-miR-196b, hsa-miR-497	hsa-miR-365, hsa-miR-654-3p, hsa-miR-424, hsa-miR-127-3p, hsa-miR-377*, hsa-miR-455-3p, hsa-miR-335	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154
p53	76 (46–17)	*RB1*, *PTEN*, *BBC3*, hsa-miR-196b, hsa-miR-497, has-miR-592	*GADD45B*, hsa-miR-485-3p, hsa-miR-193b*, hsa-miR-365, hsa-miR-654-3p, hsa-miR-424, hsa-miR-127-3p, hsa-miR-377*, hsa-miR-455-3p, hsa-miR-335	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-154, miR-329
Interferon/cytokine	66 (29–14)	hsa-miR-196b, hsa-miR-497	hsa-miR-377*, hsa-miR-455-3p, hsa-miR-424, hsa-miR-365	miR-192, miR-194, miR-196, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99
RB1	44 (25–7)	*RB1*, hsa-miR-497	hsa-miR-193b*, hsa-miR-365, hsa-miR-654-3p, hsa-miR-424, hsa-miR-127-3p, hsa-miR-377*, hsa-miR-455-3p, hsa-miR-335	miR-192, miR-194, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99
Phosphatidyl-inosiol signaling	42 (27–7)	*PIK3R1*, *CALM1*, *ITPR1*, has-miR-497	*PIK3R2*, *INPPL1*, hsa-miR-335, hsa-miR-424	miR-192, miR-17, miR-18, miR-19, miR-194, miR-199, miR-99, miR-329
DNA repair	27 (16–9)	*FANCG*, *ATM*	hsa-miR-335, hsa-miR-424	miR-192, miR-17, miR-18, miR-19, miR-154
ATR	19 (13–10)	–	–	miR-192, miR-17, miR-18, miR-19, miR-196, miR-99
Glycolysis and Gluconeogenesis	6 (4–5)		*FBP1*, *G6PC*, *ALDH9A1*	–

unexpected clustering is the heterogeneity of the tumors. Four samples had mixed morphology and the fresh tumor tissue analyzed differed from the main part of the tumor on which the diagnosis was made. As a control, the diagnosis was set on both the FFPE tissue and a frozen section of the fresh material. Three of the pancreatobiliary samples from the ampulla clustered with the intestinal samples. The three tumors had mixed morphology, and were partly poorly differentiated. The unexpected clustering according to the diagnosis of the main part of the tumor, clustered as expected according to the type of the fresh sample analyzed.

The molecular findings emphasize the importance of stratifying patients with adenocarcinomas of pancreatobiliary versus intestinal type. The two histological subtypes are targeted by specific and distinct sets of miRNA families and mRNAs. Six miRNA families are downregulated and four are upregulated in the pancreatobiliary type as compared to the intestinal type adenocarcinomas ($P < 0.05$). Pancreatobiliary and intestinal type adenocarcinomas do exhibit very different clinical characteristics as shown in Table 1, thus the survival differences can be attributed to both tumor characteristics and molecular differences.

Various important pathways related to PDACs have been previously documented (Jones et al., 2008). The results from the pathway analysis suggest that different pathogenic mechanisms are involved in the two histological subtypes. Wnt signaling, which is enriched only in the pancreatobiliary type, is commonly associated with pancreatic cancer development and has been linked to *KRAS* mutations (Zhang et al., 2013). The frequency of *KRAS* mutations was significantly higher in the pancreatobiliary type PAs as compared to the intestinal type. We observed some variation in the frequency between tumors from the different sites of origin, with the highest frequency of *KRAS* mutations in the PDACs and the lowest in the ampullary adenocarcinomas of the intestinal type.

The Insulin receptor recycling and the insulin regulation pathways, which crosstalk with the glucose metabolism and phosphatidylinositol signaling pathways, play an important role in proliferation and apoptosis (Subramani et al., 2014). These pathways are all depleted in the intestinal PAs and are unaffected in pancreatobiliary type, and this finding fits with unfavorable prognosis of pancreatobiliary type over intestinal type. Deregulation of these pathways has been reported to be associated with colon and breast cancer (Chang et al., 2013 and Giovannucci, 2001).

The tumor suppressor pathways (p53 and p73), the transcription factor pathway (E2F1 and RB1), PPAR-γ and the cytokine pathways (IFN1), enriched in our study may all be potential targets for new chemotherapeutic treatment alone or in interplay between two or more signaling pathways (Sherr and McCormick, 2002).

The PPAR-γ pathway, known to inhibit tumor growth by differentiation and apoptosis, was upregulated in the intestinal PAs. Upregulation of the PPAR pathway has been associated with shorter OS, a finding contradicting a study in pancreatic cell lines given "PPAR-γ ligand troglitazone" which resulted in reduced growth (Kristiansen et al., 2006 and Yoshizawa et al., 2002). To clarify the role of PPAR in pancreatic cancer, further studies are needed.

Several miRNAs and mRNAs associated with improved OS and RFS for the two histopathological subtypes are reported. For the pancreatobiliary type the genes *ATM*, *PTEN*, *RB1* and the miRNAs miR-592 and miR-497, and for the intestinal type, the genes *PDPK1*, *PIK3R2*, *G6PC* and the miRNAs miR-127-3p, miR-377* are linked to enriched pathways and identified as prognostic markers. The miR-17 and miR-19 families are associated with the RB1, E2F1, TGF-beta, P53, PI3 kinase and hedgehog signaling pathways, where the miR-19 family is an important oncogenic component of the miR-17 cluster (Conkrite et al., 2011, Murphy et al., 2013, Hamilton et al., 2013 and Olive et al., 2009). The miR-99, miR-192, miR-194 and miR-215 families are all linked to IGF and p53 signaling (Chen et al., 2012, Murphy et al., 2013 and Pichiorri et al., 2010). TGF-beta and Wnt catenin signaling are associated with the miR-154 family (Milosevic et al., 2012). miR-196 is an important factor for *HOX* gene activation and function where *HOX* gene expression is controlled by Wnt signaling (Chen et al., 2011 and Maloof et al., 1999). The miR-192, miR-194 and miR-199 families are also associated with better OS and RFS in tumors with intestinal histology in our study. Further, the two pathways ATM and ATR signaling pathways, known to crosstalk, are enriched, but to a larger extent in the intestinal PAs. ATM signaling controls DNA double strand breaks, while ATR has an important role in DNA damage control (Cimprich and Cortez, 2008; Zou et al., 2007). Cytokines (IFN1), which acts upon cell differentiation, growth, and the immune system were upregulated in all our samples, but to a higher extent in tumors of pancreatobiliary type. It has recently been shown that by combining IFN-1 and PPAR inhibitors the antitumor effect is improved by targeting different signaling pathways with antitumor effect and tumor resistance mechanisms (Dicitore et al., 2014).

The identified miRNAs and mRNAs with prognostic significance, specific for the histopathological subtypes (P and I) enhance our understanding of the molecular basis of the PAs. Our study emphasizes the fact that PAs are distinct at the level of histopathology rather than at the level of site of origin.

There are important clinical implications of the findings of our study. Our data support the notion that prospective clinical trials on adjuvant treatment for PAs may benefit by stratifying patients based on histopathology. The histological subtypes should be taken into account when considering curative intent surgery for metastatic PAs (De Jong et al., 2010). The identified markers at the molecular level could open up for validation of predictive and prognostic biomarkers guiding in the therapeutic decision making to an individualized treatment, especially whether the patients should undergo upfront surgery or neoadjuvant treatment.

Acknowledgments

Supported by grants from South-Eastern Regional Health Authority, Holes Foundation and Oslo University Hospital. We thank all the patients who participated in the study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2014.12.002.

REFERENCES

Albores-Saavedra, J., Simpson, K., Dancer, Y.J., Hruban, R., 2007. Intestinal type adenocarcinoma: a previously unrecognized histologic variant of ductal carcinoma of the pancreas. Ann. Diagn. Pathol. 11, 3–9.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B. 57, 289–300.

Berry, M.P., Graham, C.M., McNab, F.W., Xu, Z., Bloch, S.A., Oni, T., Wilkinson, K.A., Banchereau, R., Skinner, J., Wilkinson, R.J., Quinn, C., Blankenship, D., Dhawan, R., Cush, J.J., Mejias, A., Ramilo, O., Kon, O.M., Pascual, V., Banchereau, J., Chaussabel, D., O'Garra, A., 2010. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466, 973–977.

Bosman, F.T., Carneiro, F., Hruban, R.H., Theise, N.D. (Eds.), 2010. WHO Classification of Tumors of the Digestive System, fourth Ed. International Agency for Research on Cancer.

Chang, W.W., Lin, R.J., Yu, J., Chang, W.Y., Fu, C.H., Lai, A.C., Yu, J.C., Yu, A.L., 2013. The expression and significance of insulin-like growth factor-1 receptor and its pathway on breast cancer stem/progenitors. Breast Cancer Res. BCR. 15, R39.

Chen, C., Zhang, Y., Zhang, L., Weakley, S.M., Yao, Q., 2011. MicroRNA-196: critical roles and clinical applications in development and cancer. J. Cell. Mol. Med. 15, 14–23.

Chen, Z., Jin, Y., Yu, D., Wang, A., Mahjabeen, I., Wang, C., Liu, X., Zhou, X., 2012. Down-regulation of the microRNA-99 family members in head and neck squamous cell carcinoma. Oral Oncol. 48, 686–691.

Cimprich, K.A., Cortez, D., 2008. ATR: an essential regulator of genome integrity. Nat. Rev. Mol. Cel. Biol. 9, 616–627.

Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., Feiler, H.S., Ko, A.H., Olshen, A.B., Danenberg, K.L., Tempero, M.A., Spellman, P.T., Hanahan, D., Gray, J.W., 2011. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat. Med. 17, 500–503.

Conkrite, K., Sundby, M., Mukai, S., Thomson, J.M., Mu, D., Hammond, S.M., MacPherson, D., 2011. miR-17~92 cooperates with RB pathway mutations to promote retinoblastoma. Genes Dev. 25, 1734–1745.

De Jong, M.C., Tsai, S., Cameron, J.L., Wolfgang, C.L., Hirose, K., van Vledder, M.G., Eckhauser, F., Herman, J.M., Edil, B.H., Choti, M.A., Schulick, R.D., Pawlik, T.M., 2010. Safety and efficacy of curative intent surgery for peri-ampullary liver metastasis. J. Surg. Oncol. 102, 256–263.

Dicitore, A., Caraglia, M., Gaudenzi, G., Manfredi, G., Amato, B., Mari, D., Persani, L., Arra, C., Vitale, G., 2014. Type I interferon-mediated pathway interacts with peroxisome proliferator activated receptor-gamma (PPAR-gamma): at the cross-road of pancreatic cancer cell proliferation. Biochim. Biophys. Acta 1845, 42–52.

Donahue, T.R., Tran, L.M., Hill, R., Li, Y., Kovochich, A., Calvopina, J.H., Patel, S.G., Wu, N., Hindoyan, A., Farrell, J.J., Li, X., Dawson, D.W., Wu, H., 2012. Integrative survival-based molecular profiling of human pancreatic cancer. Clin. Cancer Res. official J. Am. Assoc. Cancer Res. 18, 1352–1363.

Dweep, H., Sticht, C., Pandey, P., Gretz, N., 2011. miRWalk-database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J. Biomed. Inform. 44, 839–847.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Giovannucci, E., 2001. Insulin, insulin-like growth factors and colon cancer: a review of the evidence. J. Nutr. 131, 3109S–3120S.

Hamfjord, J., Stangeland, A.M., Skrede, M.L., Tveit, K.M., Ikdahl, T., Kure, E.H., 2011. Wobble-enhanced ARMS method for detection of KRAS and BRAF mutations. Diagn. Mol. Pathol. Am. J. Surg. Pathol. part B 20, 158–165.

Hamilton, M.P., Rajapakshe, K., Hartig, S.M., Reva, B., McLellan, M.D., Kandoth, C., Ding, L., Zack, T.I., Gunaratne, P.H., Wheeler, D.A., Coarfa, C., McGuire, S.E., 2013. Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif. Nat. Commun. 4, 2730.

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.M., Fu, B., Lin, M.T., Calhoun, E.S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D.R., Hidalgo, M., Leach, S.D., Klein, A.P., Jaffee, E.M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J.R., Kern, S.E., Hruban, R.H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W., 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801–1806.

Kimura, W., Futakawa, N., Yamagata, S., Wada, Y., Kuroda, A., Muto, T., Esaki, Y., 1994. Different clinicopathologic findings in two histologic types of carcinoma of papilla of Vater. Jpn. J. Cancer Res. Gann 85, 161–166.

Kristiansen, G., Jacob, J., Buckendahl, A.C., Grutzmann, R., Alldinger, I., Sipos, B., Kloppel, G., Bahra, M., Langrehr, J.M., Neuhaus, P., Dietel, M., Pilarsky, C., 2006. Peroxisome proliferator-activated receptor gamma is highly expressed in pancreatic cancer and is associated with shorter overall survival times. Clin. Cancer Res. official J. Am. Assoc. Cancer Res. 12, 6444–6451.

Lopez-Romero, P., 2011. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library. BMC genomics 12, 64.

Maloof, J.N., Whangbo, J., Harris, J.M., Jongeward, G.D., Kenyon, C., 1999. A Wnt signaling pathway controls hox gene expression and neuroblast migration in *C. elegans*. Development 126, 37–49.

Milosevic, J., Pandit, K., Magister, M., Rabinovich, E., Ellwanger, D.C., Yu, G., Vuga, L.J., Weksler, B., Benos, P.V., Gibson, K.F., McMillan, M., Kahn, M., Kaminski, N., 2012. Profibrotic role of miR-154 in pulmonary fibrosis. Am. J. Respir. Cel. Mol. Biol. 47, 879–887.

MSigDb: http://www.broadinstitute.org/gsea/msigdb/index.jsp (version v4.0 revised on May 31, 2013).

Murphy, B.L., Obad, S., Bihannic, L., Ayrault, O., Zindy, F., Kauppinen, S., Roussel, M.F., 2013. Silencing of the miR-17~92 cluster family inhibits medulloblastoma progression. Cancer Res. 73, 7068–7078.

Neoptolemos, J.P., Moore, M.J., Cox, T.F., Valle, J.W., Palmer, D.H., McDonald, A.C., Carter, R., Tebbutt, N.C., Dervenis, C., Smith, D., Glimelius, B., Charnley, R.M., Lacaine, F., Scarfe, A.G., Middleton, M.R., Anthoney, A., Ghaneh, P.,

Halloran, C.M., Lerch, M.M., Olah, A., Rawcliffe, C.L., Verbeke, C.S., Campbell, F., Buchler, M.W.European Study Group for Pancreatic, C, 2012. Effect of adjuvant chemotherapy with fluorouracil plus folinic acid or gemcitabine vs observation on survival in patients with resected periampullary adenocarcinoma: the ESPAC-3 periampullary cancer randomized trial. JAMA: J. Am. Med. Assoc. 308, 147–156.

Neoptolemos, J.P., Stocken, D.D., Bassi, C., Ghaneh, P., Cunningham, D., Goldstein, D., Padbury, R., Moore, M.J., Gallinger, S., Mariette, C., Wente, M.N., Izbicki, J.R., Friess, H., Lerch, M.M., Dervenis, C., Olah, A., Butturini, G., Doi, R., Lind, P.A., Smith, D., Valle, J.W., Palmer, D.H., Buckels, J.A., Thompson, J., McKay, C.J., Rawcliffe, C.L., Buchler, M.W.European Study Group for Pancreatic, C, 2010. Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: a randomized controlled trial. JAMA: J. Am. Med. Assoc. 304, 1073–1081.

Neoptolemos, J.P., Stocken, D.D., Friess, H., Bassi, C., Dunn, J.A., Hickey, H., Beger, H., Fernandez-Cruz, L., Dervenis, C., Lacaine, F., Falconi, M., Pederzoli, P., Pap, A., Spooner, D., Kerr, D.J., Buchler, M.W.European Study Group for Pancreatic, C, 2004. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. New Engl. J. Med. 350, 1200–1210.

Oettle, H., Neuhaus, P., Hochhaus, A., Hartmann, J.T., Gellert, K., Ridwelski, K., Niedergethmann, M., Zulke, C., Fahlke, J., Arning, M.B., Sinn, M., Hinke, A., Riess, H., 2013. Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. JAMA: J. Am. Med. Assoc. 310, 1473–1481.

Oettle, H., Post, S., Neuhaus, P., Gellert, K., Langrehr, J., Ridwelski, K., Schramm, H., Fahlke, J., Zuelke, C., Burkart, C., Gutberlet, K., Kettner, E., Schmalenberg, H., Weigang-Koehler, K., Bechstein, W.O., Niedergethmann, M., Schmidt-Wolf, I., Roll, L., Doerken, B., Riess, H., 2007. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial. JAMA: J. Am. Med. Assoc. 297, 267–277.

Olive, V., Bennett, M.J., Walker, J.C., Ma, C., Jiang, I., Cordon-Cardo, C., Li, Q.J., Lowe, S.W., Hannon, G.J., He, L., 2009. miR-19 is a key oncogenic component of mir-17-92. Genes Dev. 23, 2839–2849.

Overman, M.J., Zhang, J., Kopetz, S., Davies, M., Zhi-Qin, J., Stemke-Hale, K., Rummele, P., Pilarsky, C., Grutzmann, R., Hamilton, S., Hwang, R., Abbruzzese, J.L., Varadhachary, G., Broom, B., Wang, H., 2013. Gene expression profiling of ampullary carcinomas classifies ampullary carcinomas into biliary-like and intestinal-like subtypes that are prognostic of outcome. PloS One 8, e65144.

Pichiorri, F., Suh, S.S., Rocci, A., De Luca, L., Taccioli, C., Santhanam, R., Zhou, W., Benson Jr., D.M., Hofmainster, C., Alder, H., Garofalo, M., Di Leva, G., Volinia, S., Lin, H.J., Perrotti, D., Kuehl, M., Aqeilan, R.I., Palumbo, A., Croce, C.M., 2010. Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development. Cancer cell 18, 367–381.

Saldanha, A.J., 2004. Java Treeview–extensible visualization of microarray data. Bioinformatics 20, 3246–3248.

Sherr, C.J., McCormick, F., 2002. The RB and p53 pathways in cancer. Cancer cell 2, 103–112.

Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3. Article3.

Smyth, G.K., Michaud, J., Scott, H.S., 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067–2075.

Sobin, L.H., Gospodarowicz, M.K., Wittekind, Ch., International Union against Cancer, 2009. Classification of Malignant Tumorus, seventh ed. Wiley-BlackWell, Hoboken, NJ.

Subramani, R., Lopez-Valdez, R., Arumugam, A., Nandy, S., Boopalan, T., Lakshmanaswamy, R., 2014. Targeting insulin-like growth factor 1 receptor inhibits pancreatic cancer growth and metastasis. PloS One 9, e97016.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. United States America 102, 15545–15550.

Therneau, Terry M., Grambsch, Patricia M., 2000. Modeling Survival Data: Extending the Cox Model. Springer, New York, ISBN 0-387-98784-3.

Thomas, J.K., Kim, M.S., Balakrishnan, L., Nanjappa, V., Raju, R., Marimuthu, A., Radhakrishnan, A., Muthusamy, B., Khan, A.A., Sakamuri, S., Tankala, S.G., Singal, M., Nair, B., Sirdeshmukh, R., Chatterjee, A., Prasad, T.S., Maitra, A., Gowda, H., Hruban, R.H., Pandey, A., 2014. Pancreatic Cancer Database: an integrative resource for pancreatic cancer. Cancer Biol. Ther. 15.

Westgaard, A., Pomianowska, E., Clausen, O.P., Gladhaug, I.P., 2013. Intestinal-type and pancreatobiliary-type adenocarcinomas: how does ampullary carcinoma differ from other periampullary malignancies? Ann. Surg. Oncol. 20, 430–439.

Westgaard, A., Tafjord, S., Farstad, I.N., Cvancarova, M., Eide, T.J., Mathisen, O., Clausen, O.P., Gladhaug, I.P., 2008. Pancreatobiliary versus intestinal histologic type of differentiation is an independent prognostic factor in resected periampullary adenocarcinoma. BMC cancer 8, 170.

Winter, J.M., Brennan, M.F., Tang, L.H., D'Angelica, M.I., Dematteo, R.P., Fong, Y., Klimstra, D.S., Jarnagin, W.R., Allen, P.J., 2012. Survival after resection of pancreatic adenocarcinoma: results from a single institution over three decades. Ann. Surg. Oncol. 19, 169–175.

Yoshizawa, K., Cioca, D.P., Kawa, S., Tanaka, E., Kiyosawa, K., 2002. Peroxisome proliferator-activated receptor gamma ligand troglitazone induces cell cycle arrest and apoptosis of hepatocellular carcinoma cell lines. Cancer 95, 2243–2251.

Zhang, Y., Morris, J.P.t., Yan, W., Schofield, H.K., Gurney, A., Simeone, D.M., Millar, S.E., Hoey, T., Hebrok, M., Pasca di Magliano, M., 2013. Canonical wnt signaling is required for pancreatic carcinogenesis. Cancer Res. 73, 4909–4922.

Zou, L., 2007. Single- and double-stranded DNA: building a trigger of ATR-mediated DNA damage response. Genes Dev. 21, 879–885.

Supplementary figure and tables legends

Supplementary Figure S.1a: Hierarchical clustering of the pancreatobiliary (n = 65) and the normal samples using the variant mRNAs identified within the pancreatobiliary subgroup.

Supplementary Figure S.1b: Hierarchical clustering of the intestinal (n = 16) and the normal samples using the variant mRNAs identified within the intestinal subgroup.

Supplementary figure 2a: Hierarchical clustering of the pancreatobiliary (n = 67) and the normal samples using the most variable miRNAs across pancreatobiliary samples.

Supplementary figure 2b: Hierarchical clustering of the intestinal subtype (n = 16) and the normal samples using the most variable miRNAs across intestinal samples.

Supplementary figure S.3: Kaplan–Meier curve demonstrating overall survival for patients in the five subgroups of PAs.

Supplementary Table S.1a: mRNAs differentially expressed between ampulla of pancreatobiliary and intestinal type.

Supplementary Table S.1b: miRNAs differentially expressed between ampulla of pancreatobiliary and intestinal type.

Supplementary Table S.2a: Genes down- or upregulated in pancreatobiliary type compared to intestinal type.

Supplementary Table S.2b: miRNA down- or upregulated in pancreatobiliary type compared to intestinal type.

Supplementary Table S.3a: Transcription regulators of genes upregulated in pancreatobiliary type compared to intestinal type.

Supplementary Table S.3b: Transcription regulators of genes upregulated in intestinal type compared to pancreatobiliary type.

Supplementary Table S.4: List of the genes common both in the comparison between pancreatobiliary versus intestinal types and in the comparison between the ampulla (P) versus ampulla (I) type.

Supplementary Table S.5: miRNA families downregulated and upregulated in pancreatobiliary type compared to intestinal type.

Supplementary Table S.6: Enriched pathways in different adenocarcinomas.

Supplementary Table S.7a: List of mRNAs linked with better overall or relapse free survival for pancreatobiliary type samples. Enlisted are the P values (log rank test and cox regression test) and up or down regulation of mRNA linked with worst OS or RFS.

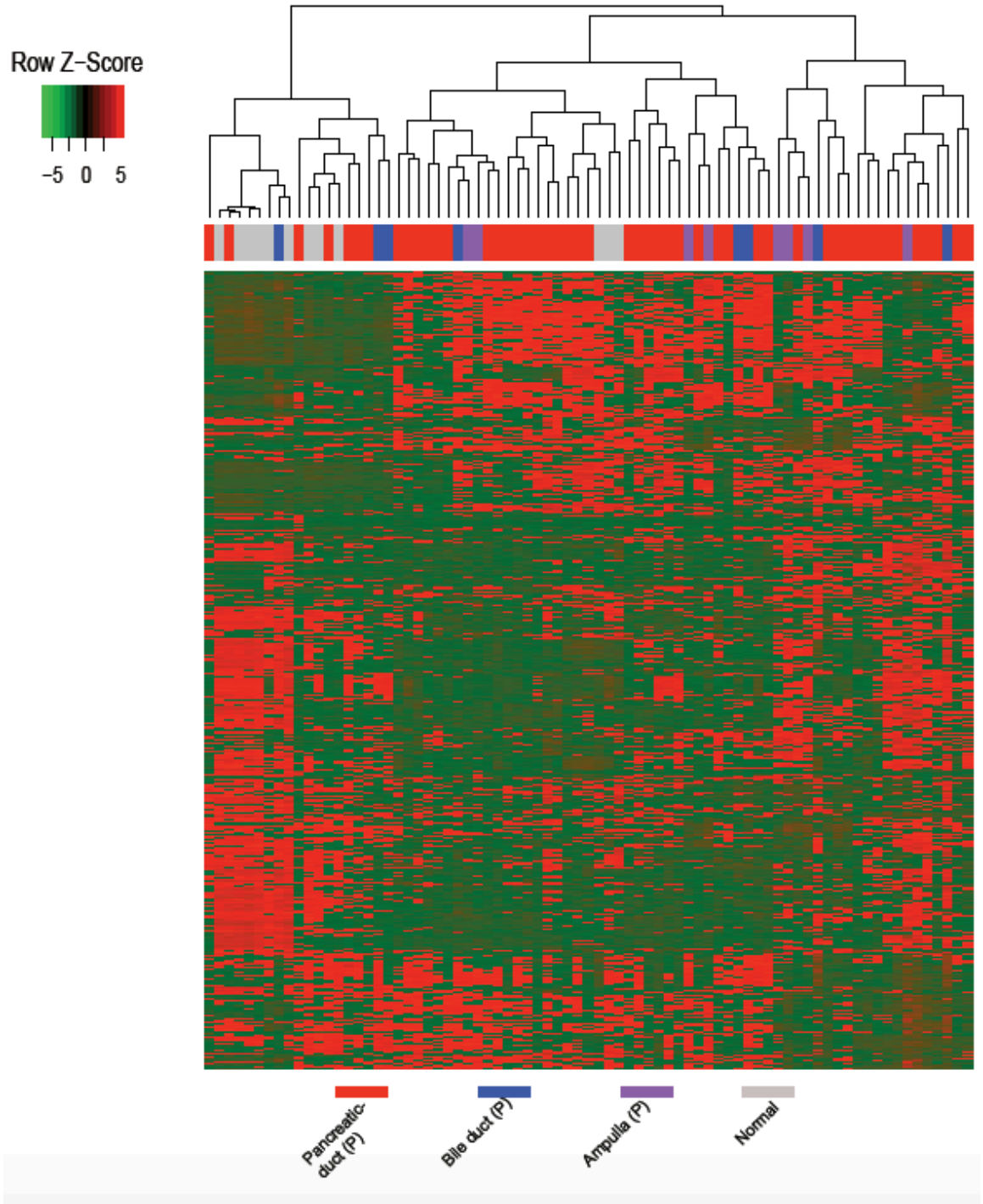
Supplementary Table S.7b: List of mRNAs linked with better overall or relapse free survival for intestinal type samples. Enlisted are the P values (log rank test and cox regression test) and up or down regulation of mRNA linked with worst OS or RFS.

Supplementary Table S.8a: List of miRNAs linked with better overall or relapse free survival for pancreatobiliary type samples. Enlisted are the P values (log rank test and cox regression test) and up or down regulation of miRNA linked with worst OS or RFS.

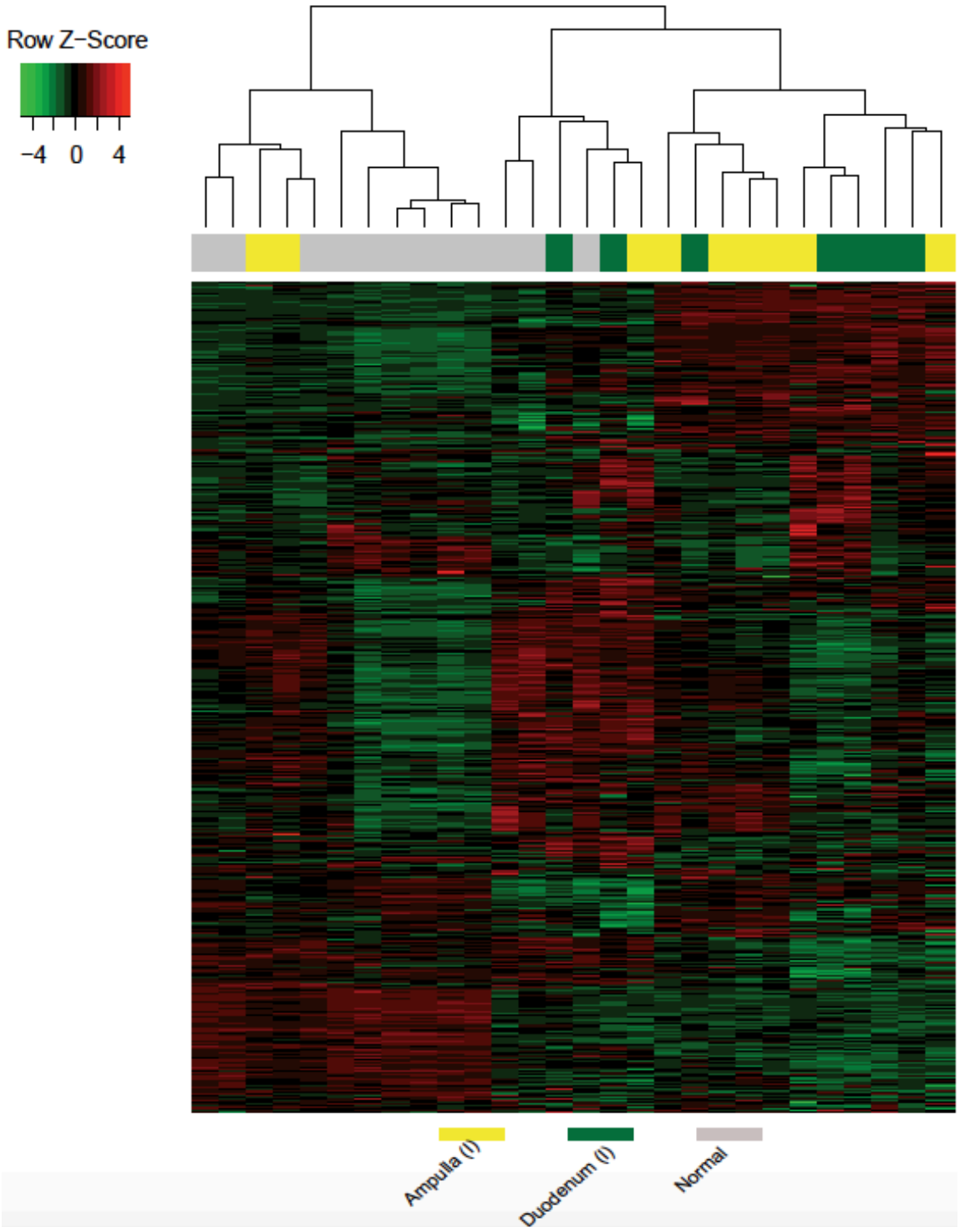
Supplementary Table S.8b: List of miRNAs linked with better overall or relapse free survival for intestinal type samples. Enlisted are the P values (log rank test and cox regression test) and up or down regulation of miRNA linked with worst OS or RFS.

Supplementary Table S9: Validated miRNA and mRNA interactions.

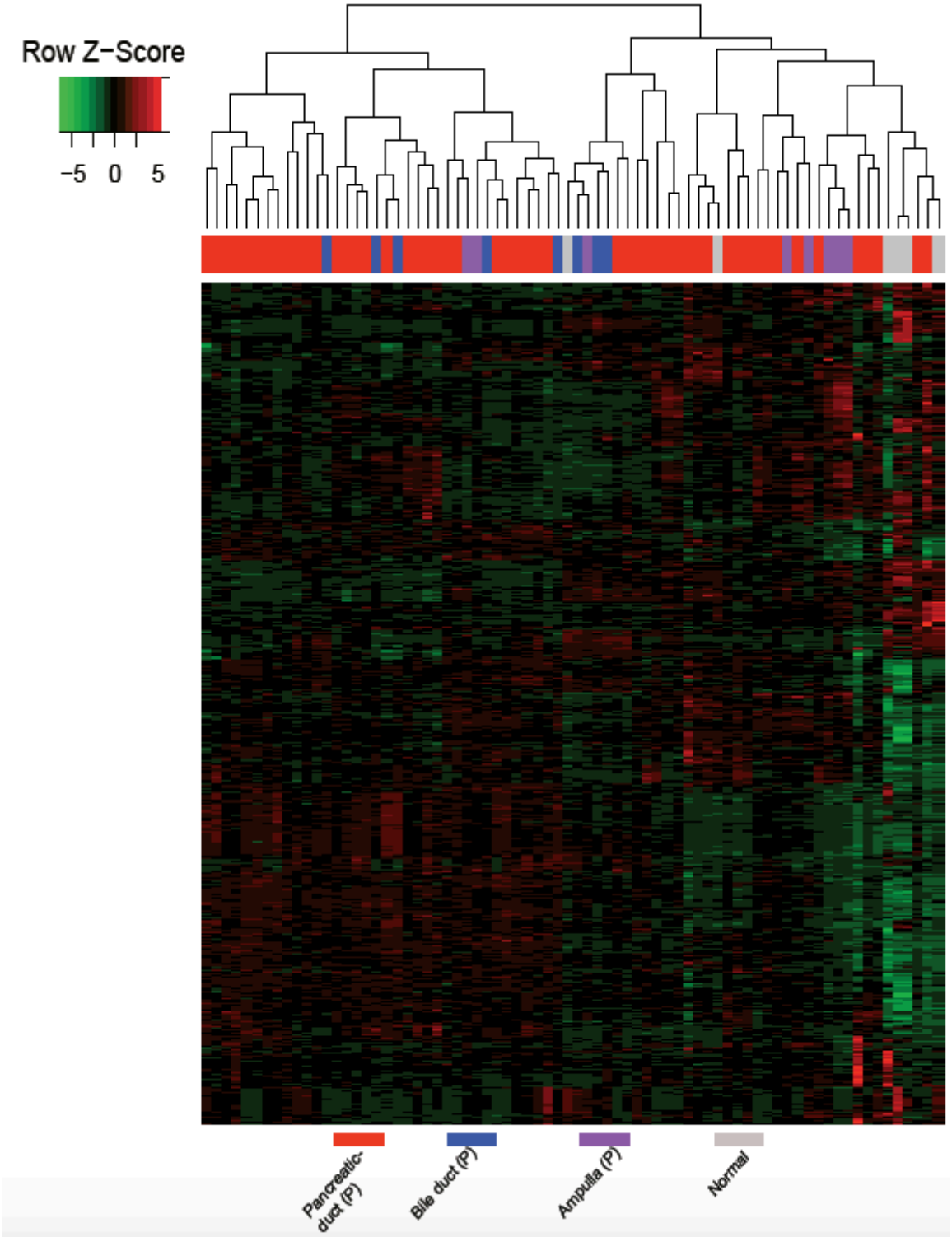
Supplementary Figure S.1a



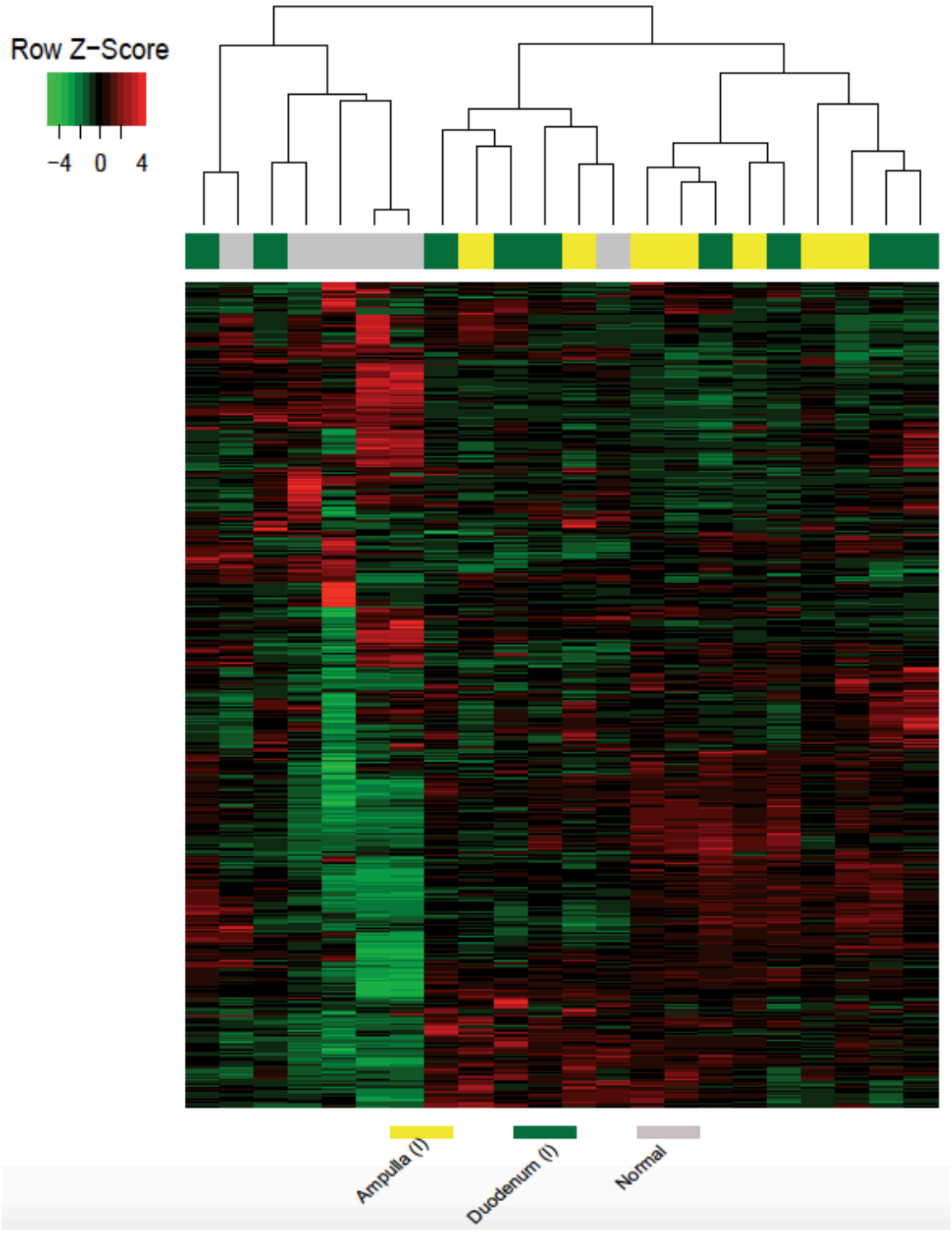
Supplementary Figure S.1b



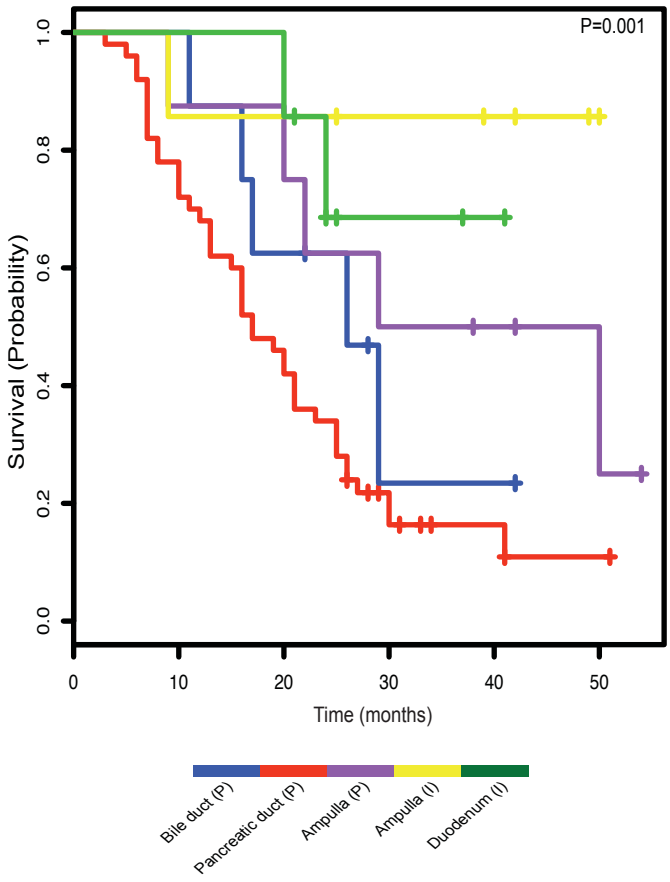
Supplementary Figure S.2a



Supplementary Figure S.2b



Supplementary Figure S.3



Supplementary tables “S.1a”, “S.2a”, “S.4”, “S.7” are available online

http://www.sciencedirect.com/science/article/pii/S1574789114002877

Supplementary Table S.1b

miRNA	log(fc)	p-value
hsa-miR-1225-5p	-0.77	0.05
hsa-miR-125a-3p	-0.94	0.05
hsa-miR-139-5p	0.46	0.03
hsa-miR-18a	0.61	0.03
hsa-miR-18b	0.41	0.02
hsa-miR-196a	1.39	0.01
hsa-miR-196b	1.36	0.03
hsa-miR-202	-0.30	0.03
hsa-miR-203	1.20	0.04
hsa-miR-20a*	0.27	0.04
hsa-miR-21*	-0.71	0.02
hsa-miR-212	-0.27	0.04
hsa-miR-222	-0.53	0.01
hsa-miR-29b-1*	0.37	0.04
hsa-miR-3162-5p	-0.68	0.04
hsa-miR-3198	-0.87	0.05
hsa-miR-335*	0.22	0.05
hsa-miR-362-5p	0.34	0.01
hsa-miR-370	-0.35	0.00
hsa-miR-374b	0.31	0.03
hsa-miR-378*	0.59	0.01
hsa-miR-4291	-0.39	0.03
hsa-miR-494	-2.14	0.05
hsa-miR-532-5p	0.34	0.05
hsa-miR-550a*	0.37	0.04
hsa-miR-551b	0.60	0.04
hsa-miR-574-3p	0.63	0.04
hsa-miR-582-5p	0.46	0.05
hsa-miR-590-5p	0.33	0.04
hsa-miR-625	0.43	0.00
hsa-miR-642b	-0.92	0.02
hsa-miR-7-1*	0.44	0.02
hsa-miR-874	-0.32	0.02

Supplementary Table S.2b

a) miRNA downregulated in pancreatobiliary type as compared to intestinal type

S.No.	miRNA	logFC	AveExpr	t	P.Value
1	hsa-miR-215	-1.31	10.09	-2.69	0.01
2	hsa-miR-192	-1.12	11.29	-2.72	0.01
3	hsa-miR-194	-1.10	10.29	-2.38	0.02
4	hsa-miR-203	-1.03	6.93	-2.86	0.01
5	hsa-miR-196a	-0.94	5.79	-3.67	0.00
6	hsa-miR-196b	-0.77	5.91	-2.66	0.01
7	hsa-miR-1	-0.73	6.98	-2.29	0.02
8	hsa-miR-133b	-0.71	6.93	-2.24	0.03
9	hsa-miR-192*	-0.68	6.15	-2.69	0.01
10	hsa-miR-146a	-0.55	8.21	-2.72	0.01
11	hsa-miR-145	-0.53	11.24	-2.00	0.05
12	hsa-miR-133a	-0.40	5.36	-2.68	0.01
13	hsa-miR-1972	-0.37	6.15	-2.07	0.04
14	hsa-miR-20a	-0.37	11.10	-2.62	0.01
15	hsa-miR-17	-0.36	9.33	-2.33	0.02
16	hsa-miR-18a	-0.36	6.71	-2.32	0.02
17	hsa-miR-28-5p	-0.29	8.34	-2.13	0.04
18	hsa-miR-378*	-0.28	5.51	-2.54	0.01
19	hsa-miR-20b	-0.27	8.28	-2.01	0.05
20	hsa-miR-19a	-0.25	10.08	-2.01	0.05
21	hsa-miR-625	-0.25	6.28	-2.35	0.02
22	hsa-miR-194*	-0.22	4.97	-2.33	0.02
23	hsa-miR-19b	-0.22	11.40	-2.15	0.03
24	hsa-miR-548c-5p	-0.21	5.37	-2.18	0.03
25	hsa-miR-4328	-0.20	5.32	-2.07	0.04
26	hsa-miR-548d-5p	-0.19	5.17	-2.06	0.04
27	hsa-miR-18b	-0.18	5.45	-2.03	0.05
28	hsa-miR-1273d	-0.14	5.11	-2.04	0.04

b) miRNA upregulated in pancreatobiliary type as compared to intestinal type

1	hsa-miR-204	0.66	6.31	3.38	0.00
2	hsa-miR-376c	0.63	8.68	2.60	0.01
3	hsa-miR-127-3p	0.59	7.91	2.73	0.01
4	hsa-miR-377	0.59	7.97	2.57	0.01
5	hsa-miR-376a	0.56	8.20	2.48	0.02
6	hsa-miR-487b	0.56	7.19	3.41	0.00

7	hsa-miR-214	0.56	10.57	2.50	0.01
8	hsa-miR-199a-5p	0.54	11.13	2.55	0.01
9	hsa-miR-381	0.52	6.94	2.89	0.00
10	hsa-miR-409-3p	0.50	6.74	2.75	0.01
11	hsa-miR-410	0.49	6.14	3.06	0.00
12	hsa-miR-218	0.47	7.63	2.24	0.03
13	hsa-miR-654-3p	0.46	6.60	2.85	0.01
14	hsa-miR-154*	0.45	6.07	2.66	0.01
15	hsa-miR-99a	0.45	10.58	2.24	0.03
16	hsa-miR-199a-3p	0.44	13.13	2.26	0.03
17	hsa-miR-136*	0.43	6.44	2.41	0.02
18	hsa-miR-100	0.43	10.66	2.12	0.04
19	hsa-miR-125b	0.42	11.94	2.32	0.02
20	hsa-miR-299-5p	0.42	6.27	2.57	0.01
21	hsa-miR-154	0.42	6.46	2.72	0.01
22	hsa-miR-493*	0.40	6.07	2.56	0.01
23	hsa-miR-337-5p	0.40	6.63	2.17	0.03
24	hsa-miR-214*	0.38	6.45	2.36	0.02
25	hsa-miR-495	0.37	6.37	2.32	0.02
26	hsa-miR-432	0.36	5.89	2.93	0.00
27	hsa-miR-379	0.36	6.05	2.56	0.01
28	hsa-miR-129*	0.36	5.46	2.37	0.02
29	hsa-miR-382	0.34	5.99	2.94	0.00
30	hsa-miR-376b	0.33	5.82	2.35	0.02
31	hsa-miR-411	0.33	5.87	2.39	0.02
32	hsa-let-7c	0.32	11.86	2.56	0.01
33	hsa-miR-299-3p	0.31	5.74	2.61	0.01
34	hsa-miR-323-3p	0.31	5.27	3.40	0.00
35	hsa-miR-130a	0.30	10.92	2.32	0.02
36	hsa-miR-485-3p	0.29	5.45	2.17	0.03
37	hsa-miR-539	0.29	5.42	3.08	0.00
38	hsa-miR-758	0.28	5.33	3.01	0.00
39	hsa-miR-369-5p	0.27	5.49	2.56	0.01
40	hsa-let-7b	0.26	14.19	2.26	0.03
41	hsa-miR-543	0.25	5.52	2.35	0.02
42	hsa-miR-409-5p	0.25	5.33	2.54	0.01
43	hsa-miR-487a	0.22	5.26	2.53	0.01
44	hsa-miR-337-3p	0.22	5.33	2.34	0.02
45	hsa-miR-376a*	0.20	5.38	2.04	0.04
46	hsa-miR-125b-2*	0.20	5.62	2.04	0.04
47	hsa-miR-1271	0.18	5.64	2.35	0.02

48	hsa-miR-485-5p	0.16	4.93	3.04	0.00
49	hsa-miR-433	0.15	5.05	2.63	0.01
50	hsa-miR-329	0.14	5.10	2.13	0.04
51	hsa-miR-1185	0.12	4.99	2.37	0.02
52	hsa-miR-654-5p	0.09	5.15	2.03	0.05
53	hsa-miR-377-	0.08	4.78	2.67	0.01

Supplementary Table S.3a

Upstream Regulator	Molecule Type	p-value of overlap	Target molecules in dataset	Mechanistic Network
Hdac	Group	1.07E-06	BCL6, GADD45B, HTRA1, IRX3, KLF9, MEOX2, NDRG4, NRP1, SFRP2, SMARCD3, SMO, SPP1, THBS4	
BDNF	Growth factor	9.50E-06	AR, DUSP1, EGR2, GAD2, ITIH3, LAMP5, LOC728392, LYPD1, NGFR, NTRK2, SFRP2, SLC16A7, SPP1, TIAM1	
GDF2	Growth factor	1.35E-05	CXCL12, GADD45B, HTRA1, NGFR, NOG, NRP1, RASL11B, SERPINF1, SPP1	
WNT3A	Cytokine	2.44E-05	CLEC11A, CPT1C, CXCL12, DCN, EGR2, IGFBP5, IRX3, LRRC17, LRRC32, NGFR, PRICKLE1, SFRP2, TGFB3, WISP2	
ESR2	ligand-dependent nuclear receptor	4.29E-05	AR, C3, CHST15, CLDN11, COMP, CXCL12, GAD2, LIMCH1, LOXL4, PTGIS, SPP1, SUSD4, THBS4	AHR, CTNNB1, ESR1, ESR2
INHA	Growth factor	7.79E-05	ADAMTSL2, BOC, COL4A4, COMP, CYP1B1, FGFR1, OBSL1, SGK1, SMARCA1	
FGF2	Growth factor	9.38E-05	AR, BCL6, CXCL12, DCN, FGFR1, FOXO1, IGFBP5, MEOX2, MGP, NGFR, NOG, NOV, NR3C1, PDGFC, PDGFRB, SPP1, ST3GAL1	CEBPA, CEBPB, FGF2, TNF
SBDS	Other	1.24E-04	GADD45B, GSTT2/GSTT2B, MAFB, MT2A, PAK3, PCDHB2, PCDHB5, SPP1	
NEUROG1	Transcription regulator	1.27E-04	C3, CFH, FAP, GFPT2, NOG, PLK2, SERPINF1	
GLIS2	Transcription regulator	1.85E-04	C3, INS, LTBP2, MGP	
FANCC	Other	2.89E-04	CXCL12, CYP7B1, FGFR1, MMP7, NR3C1, WWTR1	
SFTPA	Transporter	2.91E-04	ANKRD35, CRLF1, CYP1B1, GFPT2, MAP1B, SFRP2, ST3GAL1	

1				
CHUK	Kinase	3.86E-04	C3, CRLF1, CTSF, CYP7B1, DCN, GADD45B, MGP, NOV, SGK1, SPP1, VEGFB	AR, CEBPA, CEBPB, CHUK, CTNNB1, HTT, TAF4, TCF, TNF
ERBB2	Kinase	4.34E-04	BEX2, CADM1, CLEC11A, CXCL12, ELL2, FOXO1, GSPT2, HTRA1, IGFBP5, IRX3, LTBP2, MAP1B, NDRG4, NNMT, NPTX2, NRP1, PPP2R2B, PRKCDBP, SCG5, SRPX, ST3GAL6, THBS4, WISP2	
ERBB4	Kinase	5.52E-04	CADM1, CXCL12, DUSP1, IGFBP5, PCDHA11, PTGIS, WISP2	
APP	Other	5.86E-04	C1R, C3, CRLF1, DBN1, DCN, EVL, FAP, FGFR1, ICAM2, IGFBP5, LDHB, MAP1B, MAP2, MAP6, NGFR, NPTX1, NRP1, PAK3, PDGFRB, PJA1, RUNX1T1, SERPINF1, STMN2, UCHL1, VEGFB	APP, CEBPA, CEBPB, TNF
Cg	Complex	6.28E-04	AR, BCAT1, CLDN11, DUSP1, EFEMP1, FJX1, HSD11B1, LIMCH1, NRP1, PKIA, PLXND1, RORA, SMARCA1, SPP1, TRO	
TNF	Cytokine	6.51E-04	AR, B4GALNT1, BCL6, C10orf10, C3, CDO1, CLDN11, CLEC11A, CRLF1, CTSF, CXCL12, CYP1B1, CYP7B1, DCN, DUSP1, EGR2, FGFR1, GADD45B, GFPT2, HSD11B1, ICAM2, IGFBP5, INS, LTBP2, MEOX2, MGP, MMP7, NGFR, NNMT, NOV, NR3C1, NRP1, OSMR, PLK2, SERPINF1, SGK1, SPP1, ST3GAL6, TGFB3, TPST1, TRPC1, WISP2, ZNF365	CEBPA, CEBPB, CHUK, CTNNB1, HTT, TGFB1, TNF
TGFB1	Growth factor	6.87E-04	ANXA8L2 (includes others), BCL6, CADM1, COMP, CXCL12, DCN, DUSP1, EGR2, F13A1, FOXC1, FOXO1, GADD45B, GFPT2, GPRC5B, HTRA1, ICAM2, IGFBP5, ITGBL1, ITIH3, KLF9, LTBP2, MGP, MMP7, MSN, NBEA, NDRG4, NNMT, NOG, NOV, NRP1, PAK3, PDGFRB, PLK2, RASL11B, RBMS1, RORA, SALL2, SCG5, SGK1, SPP1, TGFB3, TMEM17, TRPC1, WISP2, ZNF365	AR, CEBPA, CEBPB, CHUK, CTNNB1, HTT, TCF, TGFB1, TNF
HTT	Transcription regulator	6.88E-04	BCL6, C1R, DCN, EGR2, EVL, FGFR1, IGFBP5, KLF9, LDHB, LTBP2, MAP2, MGP, NGFR, NPTX1, NTRK2, PER1, PRKCDBP, SCN4B, SERPINF1, SERPING1, SGK1, SPP1, TIAM1, UCHL1	HTT, TAF4

Supplementary Table S.3b

Upstream Regulator	Molecule Type	p-value of overlap	Target molecules in dataset	Mechanistic Network
HDAC1	Transcription regulator	1.17E-06	AXIN2, CDC25A, CDC25C, CEBPA, E2F2, FABP1, FABP2, LGR5, NOS2, RECQL4	EP300, HDAC1, HDAC2, PPARG
RB1	Transcription regulator	2.02E-06	CDC25A, CDC25C, CDX1, CDX2, CEBPA, EZH2, FANCA, MYB, MYBL2, PPARGC1B, RECQL4	CDK2, E2F2, E2F3, E2F4, E2f, IL6, RB1, RBL2
EP400	Other	9.31E-06	CDC25A, CDCA3, E2F2, MYBL2, PSRC1	
SREBF1	Transcription regulator	2.37E-05	ACSS2, AGMAT, CEBPA, CYP4F2, ETHE1, FASN, HNF4A, NOS2	SCAP, SREBF1, SREBF2
ISL2	Transcription regulator	5.06E-05	EPHB1, ZIC2	
APC	Enzyme	5.46E-05	AXIN2, CDX2, DPP4, FABP2, MEP1A, NOS2	
PPARA	Ligand-dependent nuclear receptor	7.97E-05	ACSS2, ADTRP, CEBPA, ETHE1, FABP1, FABP2, FASN, HNF4A, MKI67, NOS2, TXN	EP300, NR1H4, PPARA, PPARG, RXRA, SREBF1, SREBF2
CDKN2A	Transcription regulator	1.04E-04	CDC25A, CDC25C, E2F2, EZH2, FANCA, MKI67, MYBL2, RECQL4	CDKN2A, E2F4, E2f, EP300, HDAC1, HDAC2, IL6, RB1, RBL2
E2F3	Transcription regulator	1.35E-04	CDC25A, E2F2, EZH2, MYB, MYBL2	CDK2, E2F2, E2F3, E2F4, E2f, RB1
FBXL5	Enzyme	1.51E-04	SLC11A2, TFRC	
E2F4	Transcription regulator	1.64E-04	CDC25A, CDC25C, CKS2, E2F2, HIST1H3A (includes others), MKI67, MYB, MYBL2	

NR1H4	Ligand-dependent nuclear receptor	2.13E-04	CEBPA, FABP1, FABP2, FASN, HNF4A, PPARGC1B	NR1H4, PPARA, PPARG, RXRA
CSF2	Cytokine	2.53E-04	CDCA3, CDCA8, FANCA, MKI67, MYB, NOS2, NOX1, PLAGL2, RECQL4, SLC11A2, TNFSF15	
AMER1	Other	4.99E-04	AXIN2, LGR5	
ACLY	Enzyme	4.99E-04	ACSS2, FASN	
SLC9A1	Ion channel	4.99E-04	NOS2, SLC12A2	
ABCB7	Transporter	4.99E-04	TFRC, TXN	
RBL2	Other	5.43E-04	CDC25A, CDC25C, CDX1, CDX2, MYBL2	CDK2, E2F2, E2F3, E2F4, E2f, RB1, RBL2
EP300	Transcription regulator	7.27E-04	AXIN2, CDC25A, CEBPA, EPHB2, FASN, MKI67, NOS2	EP300, PPARA, PPARG, RXRA
CSF1R	Kinase	7.30E-04	LGR5, MYB, NOS2	

Supplementary Table S.5

a) Down regulated in the pancreatobiliary type as compared to the intestinal type samples

miRNA families	Members
miR-17 family	miR-17, miR-18a, miR-20a, miR-20b, miR-18b
miR-192	miR-192, miR-215
miR-196	miR-196a, miR-196b
miR-194	miR-194, miR-194*
miR-19	miR-19a, miR-19b
miR-548	(miR-548c.5p, miR-548d.5p

b) Up regulated in the pancreatobiliary type as compared to the intestinal type

miRNA families	Members
miR-154	miR-154, miR-323-3p, miR-369, miR-377, miR-377*, miR-381, miR-382, miR-409-3p, miR-409-5p, miR-410, miR-487b, miR-539, miR-1185
miR-199	miR-199a.3p, miR-199a.5p
miR-99	miR-99a, miR-100
miR-329	miR-495, miR-543, miR-329

Supplementary Table S.6

	Pancreatic duct(P)			Bile duct(P)			Ampulla(P)			Ampulla(I)			Duodenum(I)		
Name of the pathway	**NES**	**P**	**FDR q-val**	**NES**	**P**	**FDR q-val**	**NES**	**P**	**FDR q-val**	**NES**	**P**	**FDR q-val**	**NES**	**P**	**FDR q-val**
Wnt signaling	1.81	0.00	0.01	1.69	0.01	0.03	1.79	0.01	0.02	1.24	0.19	0.20	0.74	0.85	0.88
Interferon signaling	1.57	0.00	0.05	1.78	0.00	0.02	1.38	0.05	0.16	1.22	0.11	0.21	1.25	0.11	0.18
DNA repair	1.39	0.04	0.14	1.18	0.17	0.29	1.38	0.05	0.16	1.97	0.00	0.00	1.68	0.00	0.02
PI3K	1.13	0.28	0.36	1.02	0.43	0.53	1.43	0.05	0.14	1.73	0.00	0.01	1.51	0.02	0.05
Glucose metabolism	0.85	0.74	0.79	0.67	0.96	0.98	1.21	0.20	0.27	1.67	0.00	0.02	1.59	0.01	0.03
Glycolysis	1.11	0.30	0.39	1.18	0.23	0.29	1.28	0.18	0.24	1.70	0.01	0.02	1.54	0.03	0.04
Insulin recycling	-0.93	0.57	0.58	-1.34	0.15	0.15	-1.44	0.08	0.10	-1.66	0.01	0.02	-1.83	0.00	0.01
Insulin signaling	-1.17	0.17	0.31	-1.21	0.16	0.19	-1.18	0.21	0.30	-1.33	0.06	0.10	-1.35	0.05	0.10
ATM	1.46	0.05	0.09	1.36	0.10	0.16	1.40	0.09	0.14	1.99	0.00	0.00	1.77	0.01	0.01
ATR	1.55	0.02	0.06	1.47	0.06	0.11	1.68	0.01	0.04	1.99	0.00	0.00	1.83	0.00	0.00
PPAR	1.03	0.43	0.49	1.25	0.12	0.25	1.56	0.02	0.06	1.74	0.00	0.01	1.56	0.01	0.04
P73	1.58	0.01	0.05	1.75	0.00	0.02	1.42	0.05	0.13	1.59	0.00	0.03	1.82	0.00	0.00
P53	1.76	0.00	0.01	2.00	0.00	0.00	1.59	0.00	0.05	1.89	0.00	0.00	1.84	0.00	0.00
E2F	1.58	0.00	0.05	1.71	0.01	0.03	1.56	0.02	0.06	2.18	0.00	0.00	2.16	0.00	0.00
RB1	1.62	0.02	0.05	1.60	0.01	0.06	1.74	0.00	0.02	2.09	0.00	0.00	1.88	0.00	0.00

Supplementary Table S.8a

a) miRNAs linked with worst overall survival

miRNA	P value (Cox regression analysis)	P value (log rank test)	Regulation
hsa-miR-95	0.03	0.03	down
hsa-miR-592	0.04	0.03	down
hsa-miR-497	0.04	0.04	up
hsa-miR-196b	0.04	0.03	up

b) miRNAs linked with worst relapse free survival

miRNA	P value (Cox regression analysis)	P value (log rank test)	Regulation
hsa-miR-196b	0.04	0.03	up
hsa-miR-592	0.04	0.02	down
hsa-miR-95	0.03	0.01	down

Supplementary Table S.8b

a) miRNAs linked with worst overall survival

miRNA	P value (Cox regression analysis)	P value (log rank test)	Regulation
hsa-miR-193b*	0.02	0.02	up
hsa-miR-493*	0.02	0.00	up
hsa-miR-450a	0.03	0.02	up
hsa-miR-365	0.03	0.00	up
hsa-miR-654-3p	0.03	0.00	up
hsa-miR-424	0.03	0.00	up
hsa-miR-382	0.04	0.00	up
hsa-miR-127-3p	0.04	0.00	up
hsa-miR-485-3p	0.04	0.02	up

b) miRNAs linked with worst relapse free survival

miRNA	P value (Cox regression analysis)	P value (log rank test)	Regulation
hsa-miR-455-3p	0.04	0.04	up
hsa-miR-335	0.04	0.02	up

Paper II

Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma

Sandhu V., Bowitz Lothe I.M., Labori K.J., Skrede M.L., Hamfjord J., Dalsgaard A.M., Buanes T., Dube G., Kale M.M., Sawant S., Kulkarni-Kale U., Børresen-Dale A.L., Lingjærde O.C., Kure E.H.

Molecular Oncology, November 2015, *in press.*

MOLECULAR ONCOLOGY XXX (2015) 1–14

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma

V. Sandhu[a,h], I.M. Bowitz Lothe[a,b], K.J. Labori[c], M.L. Skrede[a], J. Hamfjord[a], A.M. Dalsgaard[a], T. Buanes[c,d], G. Dube[e], M.M. Kale[f], S. Sawant[e], U. Kulkarni-Kale[e], A.-L. Børresen-Dale[a,d], O.C. Lingjærde[g,i], E.H. Kure[a,h,*]

[a]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[b]Department of Pathology, Oslo University Hospital, Oslo, Norway
[c]Department of Hepato-Pancreato-Biliary Surgery, Oslo University Hospital, Oslo, Norway
[d]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[e]Bioinformatics Centre, Savitribai Phule Pune University (Formerly University of Pune), Pune, India
[f]Department of Statistics, Savitribai Phule Pune University (Formerly University of Pune), Pune, India
[g]K.G. Jebsen Centre for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[h]Department for Environmental Health and Science, Telemark University College, Bø in Telemark, Norway
[i]Department of Computer Science, University of Oslo, Oslo, Norway

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form
22 September 2015
Accepted 8 October 2015
Available online ■

Keywords:
Periampullary adenocarcinoma
Tumor-stroma interaction
miRNA/mRNA expression profiling
Stromal reaction
Pathway analysis
Tumor microenvironment
Bioinformatics and statistical analyses

ABSTRACT

Periampullary adenocarcinomas can be of two histological subtypes, intestinal or pancreatobiliary. The latter is more frequent and aggressive, and characterized by a prominent desmoplastic stroma, which is tightly related to the biology of the cancer, including its poor response to chemotherapy. Whereas miRNAs are known to regulate various cellular processes and interactions between cells, their exact role in periampullary carcinoma remains to be characterized, especially with respect to the prominent stromal component of pancreatobiliary type cancers. The present study aimed at elucidating this role by miRNA expression profiling of the carcinomatous and stromal component in twenty periampullary adenocarcinomas of pancreatobiliary type. miRNA expression profiles were compared between carcinoma cells, stromal cells and normal tissue samples. A total of 43 miRNAs were found to be differentially expressed between carcinoma and stroma of which 11 belong to three miRNA families (miR-17, miR-15 and miR-515). The levels of expression of miRNAs miR-17, miR-20a, miR-20b, miR-223, miR-10b, miR-2964a and miR-342 were observed to be higher and miR-519e to be lower in the stromal component compared to the carcinomatous and normal components. They follow a trend where expression in stroma is highest followed by carcinoma and then normal tissue. Pathway analysis revealed that pathways regulating tumor–stroma interactions such as ECM interaction remodeling, epithelial–mesenchymal transition, focal adhesion pathway, TGF-beta, MAPK signaling, axon guidance and endocytosis were differently regulated. The miRNA-

* *Corresponding author*. Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.
E-mail address: Elin.Kure@rr-research.no (E.H. Kure).

http://dx.doi.org/10.1016/j.molonc.2015.10.011

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

mRNA mediated interactions between carcinoma and stromal cells add new knowledge regarding tumor-stroma interactions.

1. Introduction

Adenocarcinomas originating from the pancreatic duct, the distal common bile ducts and the ampulla are collectively referred to as periampullary adenocarcinomas (PAs), all located in the pancreatic head. They are treated by the same surgical treatment and their diagnostic distinction may be difficult clinically, radiologically and morphologically. Recently, two histological subtypes of PAs were described: intestinal and pancreatobiliary (Westgaard et al., 2008). As implicated by the names, the latter has the typical histomorphology of pancreatic or distal bile duct cancer, and shares with these also the poor outcome. According to the Norwegian Cancer Registry (2014), the 5-year survival rate for pancreatic ductal adenocarcinomas (PDAC), which is the most common of PAs, lies at approximately 5% and has not improved during the last 10 years. Despite a large number of studies that aimed at improving survival for PDAC, treatment strategies have remained largely unsuccessful (Oettle, 2014). The high levels of heterogeneity between individual tumors as well as the extensive stromal component of the cancer (up to 80% of tumor mass) are considered main obstacles to effective treatment (Oettle, 2014). The stroma also plays an active role in disease progression and resistance to chemotherapy (Feig et al., 2012) and represents a relevant target for cancer treatment. It constitutes the microenvironment of the cancer cells and consists of fibroblasts, stellate cells, extracellular matrix, vasculature, peripheral nerves and immune cells (De Wever and Mareel, 2003). Studies have shown that miRNAs play an important role in regulating the tumor microenvironment (Zhang et al., 2014, Chou et al., 2013). It is regulated by different mechanisms like epithelial mesenchyme transition (EMT), extracellular matrix (ECM) remodeling, recruitment of metastasis-promoting stromal cells, immune escape, hypoxia, angiogenesis and modulation of key effectors such as cancer-associated fibroblasts (CAF), cell adhesion molecules (CAM) and matrix metalloproteinases (Zhang et al., 2014, Chou et al., 2013).

Studies have compared miRNA expression in pancreatic cancer with normal tissue and pancreatitis (Lee et al., 2007b, Szafranska et al., 2007). Recently, we performed an integrated expression analysis of mRNAs and miRNAs, comparing PAs with normal tissues (Sandhu et al., 2015). Lately, research focus has been shifted toward the interaction between cancer cells and its stromal components; more specifically to understand early stromal activity and the effect of stellate cells in carcinogenesis (Erkan et al., 2012). Studies on the role of miRNA and tumor microenvironment in different cancers have revealed that miRNAs are not only regulators of various cellular processes but also affect the interaction between carcinoma and stromal cells (Zhang et al., 2014). To the best of our knowledge, none of the published studies have used expression profiling of miRNAs of both the carcinomatous and stromal components from the same PAs. However, mRNA expression profiling of cancer and stromal cells has previously been compared in pancreatic cancer cell lines and stromal fibroblast induced by co-culture (Sato et al., 2004). Stromal cells have been associated with the promotion of cancer cell growth, invasion, metastatic progression, de-differentiation, and resistance to therapy (Erkan et al., 2012). Furthermore, recent data from mouse model studies uncovered a potentially host-protective role of the stroma in PDAC. Indeed, stromal depletion resulted in more aggressive, undifferentiated PDAC with enhanced EMT, profiles of infiltrating immune cells were altered, increased proliferation of cancer cells, attenuated angiogenesis and increased response to immune checkpoint blockade (Rhim et al., 2014, Ozdemir et al., 2014; Gore and Korc, 2014).

In the present study, we identified differentially expressed miRNAs in stromal versus carcinoma cells in formalin fixed paraffin embedded (FFPE) tissues from patients with PA ($n = 20$). In addition, carcinoma and stromal tissue profiles were compared with those from normal pancreatic tissues ($n = 8$) that were contained in the same specimens. Pathway analysis was carried out using mRNAs that were anti-correlated and were predicted targets of miRNAs. The aim was to identify the role of these miRNAs and the pathways that may be crucial in tumor–stroma interactions.

2. Materials and methods

2.1. Patients and specimens

A total of 20 patients with PA (10 PDACs, 9 bile duct adenocarcinomas and 1 ampullary adenocarcinoma, all of pancreatobiliary type) treated at Oslo University Hospital, were included in the present study. The median age of the patients was 69 years (range 38–79 years), seven were males and 13 females. The macroscopic and microscopic pathology examination followed a standardized protocol. The clinicopathologic characteristics of the patients are presented in Table 1. All PAs were stage T3, except for one PDAC that was limited to the pancreas (T1) and one of the bile duct adenocarcinomas that invaded adjacent structures (T4). One of the PDACs had metastasized to liver (M1 metastasis status). The present study was approved by the Regional Ethics Committee, and written informed consent was obtained from each patient.

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

Table 1 – Clinicopathologic characteristics of study series.

		Number (n = 20)	Percentage	Median
Gender	Female	13	65	–
	Male	7	35	–
Differentiation	Well	0	0	–
	Moderate	11	55	–
	Poor	8	40	–
	Undifferentiated	1	5	–
pT	T1	1	5	–
	T2	0	0	–
	T3	18	90	–
	T4	1	5	–
N	N0	6	30	–
	N1	14	70	–
M	M0	19	95	–
	M1	1	5	–
R	R0	8	40	–
	R1	12	60	–
Vessel infiltration	No	9	45	–
	Yes	11	55	–
Perineural infiltration	No	1	5	–
	Yes	19	95	–
KRAS codons 12/13	Mutated	15	75	–
	Wild type	5	25	–
Age (in years)		–	–	69.00
Disease free survival (DFS) (in months)		–	–	17.34
Overall survival (OS) (in months)		–	–	21.25

2.2. Sampling of carcinoma and stromal tissues

Hematoxylin and eosin (HE) stained sections from FFPE tissue blocks were examined by a pathologist, who marked areas with carcinoma, stroma and normal cells. Core tissue biopsies (diameter 1.0 mm; height 1–2 mm) were sampled from the selected areas of carcinoma and stroma, using Beecher Instruments' Manual Tissue Arrayer (MTA-1) from Estigen. The stroma located in close vicinity of the tumor was labeled. Using microdissected FFPE tissues enabled us to get very clean samples of stromal cells. However, the pancreatic cancer samples are heterogeneous and grow in a highly dispersed fashion. The tumor samples were sampled from regions with the higher tumor burden, preferentially >50–75% carcinoma cells. They may contain some lower percentages of stromal cells and inflammatory cells but most of the samples were predominated by carcinoma cells (Supplementary Table S1). The selected regions were examined and marked by trained pathologists. In each case new HE sections were made from these regions and the pathologist ensured that the proper tissues had been sampled from the marked areas. Two carcinoma and two stromal samples were discarded from the analysis since the contents of carcinoma or stromal cells were <10%. Paired set of carcinoma/stroma samples could be taken from 16 patients (n = 32 samples), while either only stroma samples (n = 2) or carcinoma samples (n = 2) were taken from two additional patients. Four of the 36 carcinoma samples were borderline with the content of carcinoma cells ranging from 30 to 50%. The normal samples (n = 8) were mainly composed of pancreatic acinar parenchyma cells with normal epithelium and some normal stroma. Areas of fibrosis were avoided when collecting the samples.

2.3. RNA extraction

RNA was extracted from the FFPE tissue samples by using the miRNeasy FFPE kit (Qiagen) as described by the manufacturer.

2.4. miRNA array profiling of carcinoma and stromal tissues

The tissue core biopsies of carcinoma (n = 18), stroma (n = 18) and normal (n = 8) tissues were subjected to miRNA array profiling by Exiqon Services, Denmark. The quality of the total RNA was verified by an Agilent 2100 Bioanalyzer profile. The quality control (QC) report of the samples is available in Supplementary Figure S1. A total of 700 ng RNA from both the sample and the universal reference was labeled with Hy3™ and Hy5™ fluorescent labels, respectively, using the miRCURY LNA™ microRNA Hi-Power Labeling Kit, Hy3™/Hy5™ (Exiqon, Denmark) as described by the manufacturer. The Hy3™-labeled samples and a Hy5™-labeled reference RNA sample were mixed pair-wise and hybridized to the miRCURY LNA™ microRNA Array 7th (Exiqon, Denmark), which captures probes targeting all miRNAs for human tissues (n = 2042) that are registered in the miRBASE 18.0 (Griffiths-Jones et al., 2006). Hybridization was performed according to the miRCURY LNA™ microRNA Array Instruction manual using a Tecan HS4800™ hybridization station (Tecan, Austria). Following hybridization, the microarray slides were scanned and stored in an ozone-free environment (ozone level below 2.0 ppb) in order to prevent potential bleaching of the fluorescent dyes. The miRCURY LNA™ microRNA Array slides were scanned using the Agilent G2565BA Microarray Scanner System (Agilent Technologies, Inc., USA), and the image analysis was carried out using ImaGene® 9 (miRCURY LNA™ microRNA Array Analysis Software, Exiqon, Denmark). The miRNA expression data is accessible through GEO accession number GSE71533.

2.5. miRNA expression analysis, pre-processing and normalization

The quantified signals from FFPE carcinoma, stromal and normal tissues were background corrected using the Normexp background correction method with offset value 10 (Ritchie et al., 2007). The samples were normalized using the quantile normalization method, which is found to produce the best between-slide normalization to minimize the intensity-dependent differences between samples. The threshold of detection was calculated for each individual microarray slide as 1.2 times the 25th percentile of the overall signal intensity of the slide. The miRNAs with intensities above threshold observed in less than 20% of the samples were removed from the final data set that was used for the expression analysis. For the present data set, a total of 1405 probes were discarded by this filtering procedure. A total of 677 miRNAs were used for further expression analysis.

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

2.6. mRNA expression data

The mRNA expression data for fresh frozen PAs (n = 14) and normal samples (n = 12), available in GEO with accession number GSE60979 were used in this study (Sandhu et al., 2015). The mRNA data was background corrected and quantile normalized. The moderated t-test (Smyth, 2004; Smyth et al., 2005) was used for identifying differentially expressed mRNAs between tumor and normal samples at adjusted P value < 0.05 (Benjamini and Hochberg, 1995).

2.7. Bioinformatics and statistical analyses

2.7.1. Unsupervised analysis

The unsupervised analysis was done using hierarchical clustering and sparse principal component analysis.

2.7.1.1. *Hierarchical clustering*. Unsupervised hierarchical clustering of 18 carcinoma samples, 18 stromal tissues and eight normal samples was carried out using the most variant miRNAs (n = 183) having standard deviation (s.d) > 0.8. The samples were clustered using the complete linkage method. Pearson's correlation coefficient and Spearman's rank correlation method were used as a distance measure for miRNAs and samples, respectively. The heatmap is plotted using the gplot library in R.

2.7.1.2. *Sparse principal component analysis*. Sparse Principal Component Analysis (SPCA) is an extension to the traditional principal component analysis (Jolliffe, 1986) that is more concise and gives a more interpretable set of principal components for large genomic data by applying Penalized Matrix Decomposition (PMD) (Witten and Tibshirani, 2009). It thus reduces the dimensions of large data sets and therefore represents a useful way of exploring the naturally arising sample classes based on expression profile. SPCA of all samples was carried out using the 183 most variable miRNAs. Firstly, cross validation was performed on first Sparse Principal Component (SPC) using the SPC.cv function in R. This helps in selecting the tuning parameter for SPCA, which involves applying PMD to a data matrix with lasso penalty on the columns and no penalty on the rows. The tuning parameter controls the sum of absolute values of the elements of SPCA. After tuning the parameters the SPCA was performed using the PMA library in R with parameters (sumabsv = 4 and K = 1) tuned using the SPC cross validation function (Witten et al., 2009, Witten and Tibshirani, 2009). The SPCA plot was plotted using ggplot2 library in R.

2.7.2. Identification of differentially expressed miRNAs in carcinoma and stroma

The Moderated t-tests (Smyth, 2004; Smyth et al., 2005) were carried out to identify miRNAs that are differentially expressed, between carcinoma and stromal tissues; carcinoma and normal tissues; stromal and normal tissues at Benjamini & Hochberg (BH) adjusted P < 0.05 (Benjamini and Hochberg, 1995) using the Limma module of the Bioconductor package in R. The results were confirmed using the non-parametric Mann–Whitney test (Mann et al., 1947) at BH-adjusted P < 0.05. Analysis of Variance (One-way ANOVA) was carried out to identify relative miRNA expression patterns in carcinoma, stromal and normal tissues with BH-adjusted P < 0.05 using ANOVA package in R. The results were confirmed using non-parametric Kruskal–Wallis ANOVA (Kruskal and Wallis, 1952) at BH-adjusted P < 0.05.

2.7.3. Correlations between miRNA and mRNA levels and target site prediction

Firstly, all miRNAs differentially expressed in carcinomas and stromal samples (n = 43) were compared to all mRNAs differentially expressed in tumor and normal samples in order to identify negative correlations. A Pearson correlation test was applied for this purpose, with significance threshold P < 0.05 and 95% CI for both stromal and carcinoma miRNA expression versus tumor sample (mRNA). Secondly, the miRNA-mRNA pairs found in the first step were further predicted by miRanda (John et al., 2004, Enright et al., 2003) in microRNA.org "http://www.microrna.org/microrna/" August 2010 release (Betel et al., 2008) and Targetscan release 6.2 (Lewis et al., 2003, 2005).

All the miRNA-mRNA pairs in this two-step curation were reported and further checked against experimentally validated interactions in the miRTarbase4 database (release 4.5, Nov, 2013) (Hsu et al., 2011, 2014).

2.7.4. Gene set enrichment analysis

The Gene set enrichment analysis was carried out using the WebGestalt tool (Zhang et al., 2005, Wang et al., 2013). WebGestalt uses an hypergeometric test for enrichment evaluation analysis at P < 0.05 after BH correction. The minimum number of genes required for a pathway to be considered significant is set to 10. The WebGestalt analysis gives pathway enriched genes, number of genes enriched, raw P value (rawP) from the Hypergeometric test and Benjamin and Hochberg corrected P value (adjP). The genes significantly associated with differentially expressed miRNAs in carcinoma and stroma after two-step curation in the previous step was used for the enrichment analysis.

2.7.5. Plotting the miRNA-mRNA-pathways interactions

The candidate miRNA target interactions in the enriched pathways were plotted using the tool Gephi version beta 0.8.2 (Bastian et al., 2009).

3. Results

An overview of the analysis pipeline carried out using the miRNA carcinoma/stroma and the mRNA expression data is summarized in Figure 1.

3.1. Unsupervised clustering analysis of expression data

Unsupervised hierarchical clustering of miRNAs (n = 183) showing intrinsically variable expression revealed separation of carcinoma, stromal and normal tissues (Figure 2a). The carcinoma and stromal samples clustered separately at P = 9.071e-09 (Fisher's exact test). SPCA was carried out to identify clusters based on the explained variance. The unsupervised analyses of the 183 most variable miRNAs revealed separation of the samples into carcinoma, stroma and normal

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

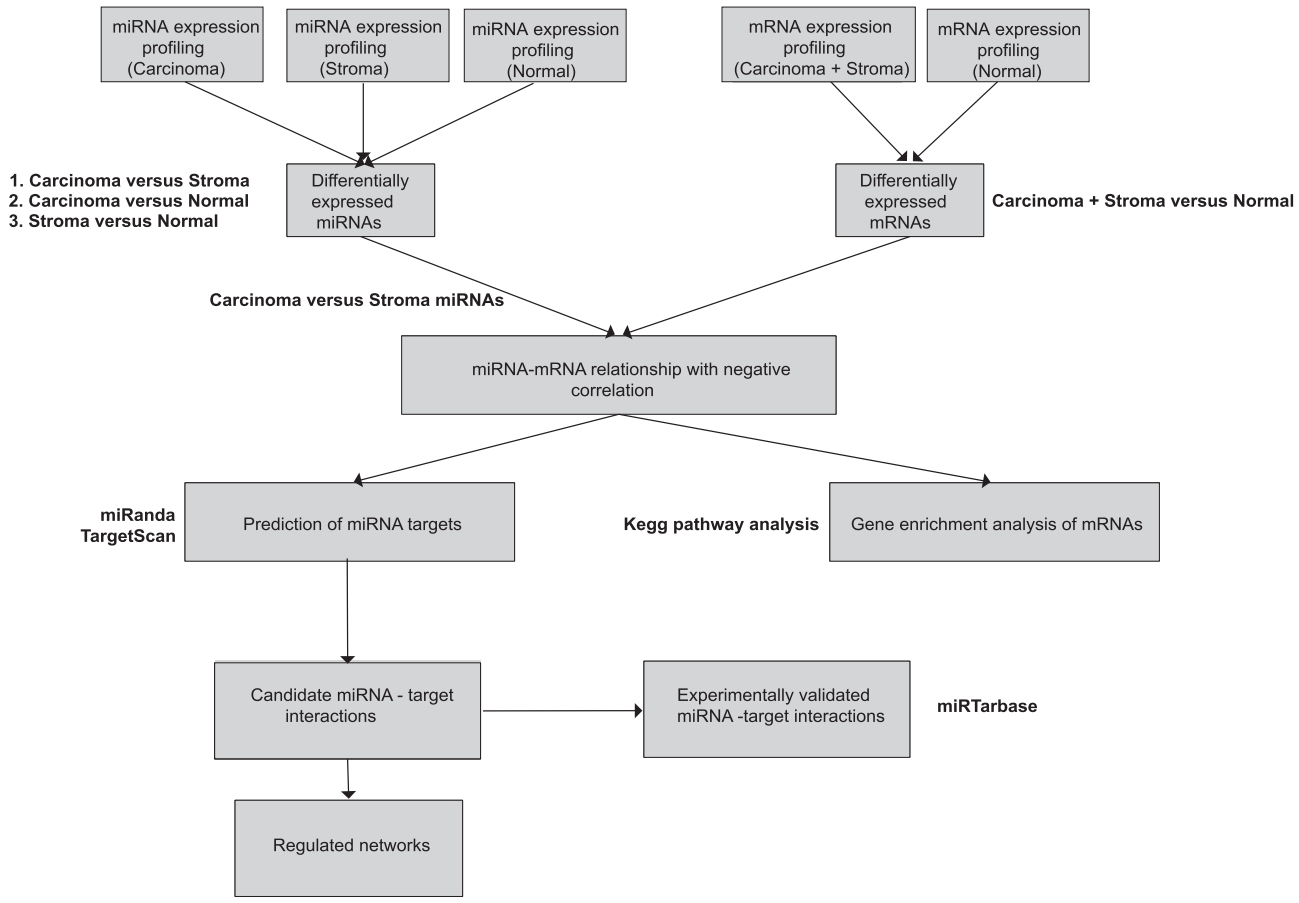


Figure 1 – Analysis pipeline of the PA samples using miRNA and mRNA expression profiling.

(Figure 2b). The normal samples clustered separately from the carcinoma and the stromal samples. An overlap between the carcinoma and stromal samples was observed in the heatmap for four of the samples, whereas, SPCA clustered carcinoma, stroma and normal samples separately.

3.2. *miRNA expression analysis*

The analysis using the Moderated t-test (modT) found that the numbers of differentially expressed miRNAs between carcinoma versus stromal tissues, stromal versus normal tissues and carcinoma versus normal tissues were 43, 287 and 331, respectively at $P < 0.05$ and adjusted for FDR correction < 5% using BH correction. The Venn diagram (Figure 3) shows the number of differentially expressed miRNAs that are common and unique in the different comparisons. There is a high number ($n = 228$) of differentially expressed miRNAs common to both carcinoma and stromal tissues as compared to the normal tissue ($P = 2.2e\text{-}44$).

The miRNAs differentially expressed in stromal tissue versus carcinoma at BH-adjusted $P < 0.05$ are enlisted in Table 2. A total of 26 and 17 miRNAs were upregulated and downregulated in stromal tissue, respectively. Some of the differentially expressed miRNAs relate to known miRNA families; miR-17, miR-15 and miR-515. The miR-17 (miR-17, miR-20a, miR-20b, miR-106a) and miR-15 (miR-15a, miR-15b, miR-16, miR-195) families were upregulated in stroma compared to carcinoma. miR-515 (miR-518a, miR-519d, miR-519e) family members were downregulated in stroma compared to carcinoma. These miRNA families were curated from mirbase database (release 21, June 2014) (www.mirbase.org), (Griffiths-Jones et al., 2006).

The miRNAs that were identified by the modT test and the Mann–Whitney test in stromal versus carcinoma tissues at BH-adjusted $P < 0.05$ are enlisted in Table 2. The miRNAs that were identified to be differentially expressed between carcinoma versus normal tissues and stromal versus normal tissues using the Moderated t-test are enlisted in the Supplementary Tables (S2a and S2b, respectively).

ANOVA test was performed on all the miRNAs to identify relative miRNA expression patterns in the three groups i.e. carcinoma, stromal and normal tissues. The test identified eight miRNAs that were upregulated in stromal as compared to carcinoma and normal tissue samples at BH-adjusted $P < 0.05$. The relative miRNA expression of six out of these eight miRNAs follows a decreasing trend, from stromal to carcinoma to normal tissues (Figure 4), while only one miRNA (miR-519e-3p) shows an increasing trend from stromal to carcinoma to normal tissues. miRNA-2964a-5p exhibits a unique pattern, as it is highly overexpressed in both stroma and carcinoma as compared to normal tissue. It is noted that the boxplots show moderate deviation from symmetry for expression of miRNAs in stroma and carcinoma as compared to normal tissue. Compared to carcinoma, the stroma shows a higher

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

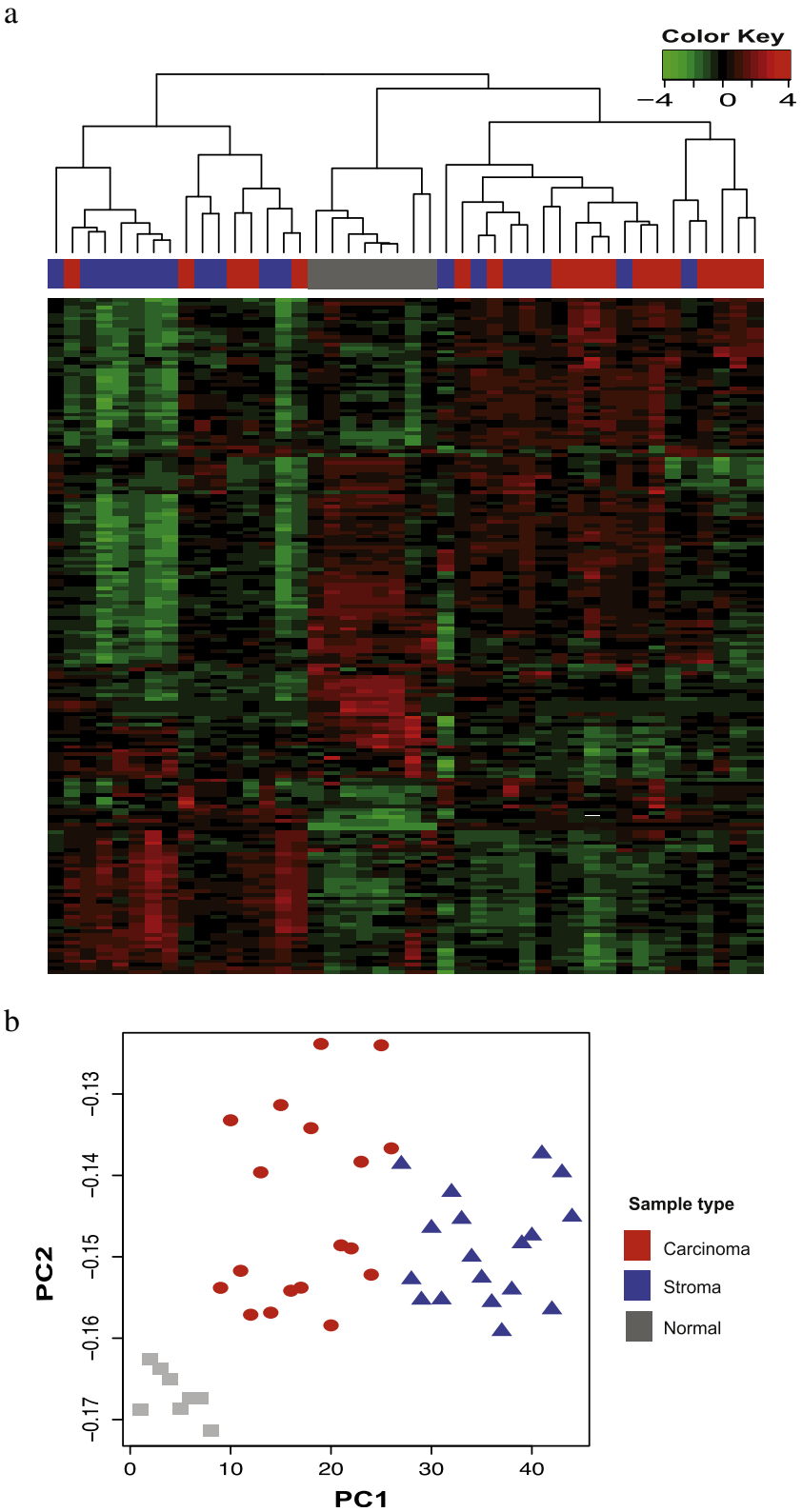


Figure 2 – 2a) The figure shows the heatmap of all the samples with the column representing the samples and the row representing the miRNAs. The samples were clustered using the complete linkage method. Pearson's correlation coefficient and Spearman's rank correlation were used as a distance measure for miRNAs and samples, respectively. The heatmap shows the separation of carcinoma, stromal and normal samples based on expression levels of the most intrinsically variable miRNAs (n = 183). The carcinoma and stromal tissues clustered separately at P = 9.071e-09 for Fisher's exact test. 2b) The figure shows the scatterplot of the first two principal components identified by the SPCA performed on all the samples and on 183 miRNAs with s.d > 0.8. The color and shape of the points refer to type of the tissues, and the x-axis represents the first principal component and the y-axis represents the second principal component of the samples.

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

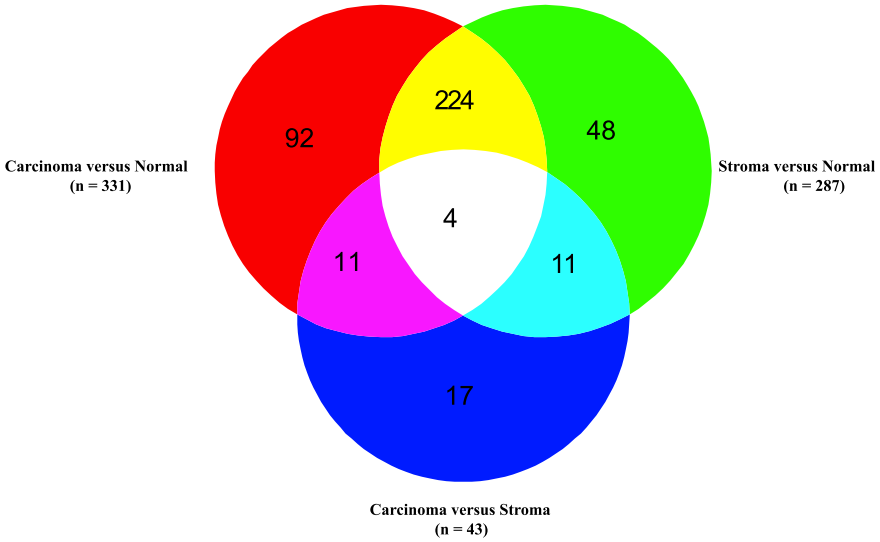


Figure 3 – The Venn diagram shows the numbers of differentially expressed miRNAs with respect to comparisons between carcinoma, stroma and normal tissues.

degree of dispersion towards the upper limit. The results are consistent with the Kruskal–Wallis rank sum test at BH-adjusted $P < 0.05$ (Supplementary Table S2c). The boxplots of 35/43 differentially expressed miRNAs between carcinoma versus stromal tissues are available in Supplementary Figure S2.

3.3. *Anti-correlation analysis and miRNA target prediction*

The total of 8670 mRNAs were found differentially expressed between tumor and normal samples (Supplementary Table S3). These differentially expressed mRNAs were used for anti-correlation analysis against 43 miRNAs differentially expressed between carcinoma and stroma. The two-step analyses for candidate miRNA-mRNA target networks (see Materials and method section 2.7.3) found 2219 carcinoma miRNA-mRNA and 1363 stroma miRNA-mRNA interactions. The total of 104 and 163 interactions were found to be experimentally validated between the carcinoma miRNA-mRNA and stroma miRNA-mRNA, respectively (Supplementary Tables S4a and S4b).

3.4. *Pathway analysis*

The list of mRNAs significantly associated with differentially expressed miRNAs between carcinoma and stroma samples were used for pathway analysis using the WebGestalt tool. Ten pathways were found to be enriched in the analysis (Table 3).

3.5. *Proliferation associated miRNAs and mRNAs*

We identified five of 11 proliferation genes included in the PAM50 gene expression signature (Parker et al., 2009) to be differentially expressed between tumors and normal tissues. The proliferation genes were anti-correlated to the carcinoma miRNA expressions and predicted by the miRNA target prediction softwares (Figure 5). Interactions between miR-20a *and CCNB1*, miR-140 and *UBE2C*, miR-16 and *UBE2C*, miR-17 and *UBE2C*, miR-20a and *UBE2C*, miR-16 and *MKI67* and miR-16 and *CEP55* were found to be experimentally validated interactions.

3.6. *The miRNA-mRNA pathway network*

The miRNA-mRNA interactions for the deregulated pathways were plotted for the stromal (Figure 6a) and carcinoma (Figure 6b) samples. The network showed closed interactions between the miR-17 and the miR-15 family members. They share many common targets. The miRNA-mRNA pairs with significant anti-correlation at $P < 0.05$ for all deregulated pathways (Table 3) are reported in Supplementary Tables S5a–j for the carcinoma and the stromal samples.

All the scripts and codes used in this project are available in Supplementary file S1.

4. Discussion

The present study is to our knowledge the first to analyze miRNA expression profiling of carcinoma, stromal and normal tissues from patients with PAs. The aim of the study was to identify pathways and miRNAs that may be crucial in interactions between carcinoma and associated stroma through the analysis of differentially expressed miRNAs and mRNAs.

The most differentially expressed miRNAs, miR-144 and miR-451a (Table 2) are located on chromosome 17 at less than 10 kb in distance from each other and are positional clusters. The positional clusters of miRNAs are known to have sequence similarity in the seed region and functional similarity with respect to the target genes and corresponding pathways (Becker et al., 2012). The miR-17 and miR-15 families were upregulated in stroma compared to carcinoma and the miR-515 family was downregulated in stroma compared to carcinoma samples. The members of the miR-17 family

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

Table 2 – (a–b): List of differentially expressed miRNAs in stroma versus carcinoma.

2a. miRNAs upregulated in stroma compared to carcinoma samples at FDR adjusted $P < 0.05$.

No.	miRNA	Log FC	Average expression	Calculated value of t statistic	P value	FDR
1	miR-144-3p	−2.019	7.119	−6.135	0.000	0.000
2	miR-451a	−2.613	10.244	−5.730	0.000	0.000
3	miR-126-3p	−1.298	9.379	−5.037	0.000	0.002
4	miR-140-3p	−1.089	7.194	−4.683	0.000	0.004
5	miR-126-5p	−0.873	7.619	−4.320	0.000	0.010
6	miR-223-3p	−1.268	8.682	−4.304	0.000	0.010
7	miR-16-5p	−1.230	9.828	−4.218	0.000	0.011
8	miR-10b-5p	−0.661	7.655	−4.188	0.000	0.011
9	miR-142-5p	−1.232	7.286	−4.126	0.000	0.012
10	miR-342-3p	−0.922	7.075	−3.853	0.000	0.021
11	miR-20b-5p	−0.821	7.091	−3.743	0.001	0.025
12	miR-191-5p	−0.853	7.782	−3.665	0.001	0.027
13	miR-195-5p	−1.102	8.682	−3.635	0.001	0.027
14	miR-130a-3p	−0.941	8.265	−3.613	0.001	0.027
15	miR-17-5p	−0.966	7.419	−3.593	0.001	0.027
16	miR-15b-5p	−0.675	7.501	−3.586	0.001	0.027
17	miR-15a-5p	−0.736	8.357	−3.566	0.001	0.027
18	miR-139-5p	−0.420	6.284	−3.557	0.001	0.027
19	miR-106a-5p	−0.752	6.685	−3.454	0.001	0.033
20	miR-20a-5p	−0.829	8.089	−3.434	0.001	0.034
21	miR-4301	−1.137	10.638	−3.327	0.002	0.041
22	miR-140-5p	−0.694	6.967	−3.320	0.002	0.041
23	miR-142-3p	−1.130	9.648	−3.316	0.002	0.041
24	miR-19b-3p	−0.763	7.800	−3.201	0.003	0.049
25	miR-2964a-5p	−0.492	10.990	−3.178	0.003	0.049
26	miR-378a-3p	−0.364	7.698	−3.133	0.003	0.049

2b. miRNAs downregulated in stroma compared to carcinoma samples at FDR adjusted $P < 0.05$

No.	miRNA	Log FC	Average expression	Calculated value of t statistic	P value	FDR
1	miR-519e-3p	0.515	7.052	3.987	0.000	0.017
2	miR-519d	1.107	8.433	3.919	0.000	0.019
3	miR-615-3p	0.513	7.887	3.819	0.000	0.021
4	miR-4633-5p	0.742	10.231	3.697	0.001	0.027
5	miR-551b-5p	0.539	6.155	3.635	0.001	0.027
6	miR-4506	0.630	6.826	3.488	0.001	0.031
7	miR-4436b-5p	0.602	9.532	3.355	0.002	0.041
8	miR-518a-5p/527	0.843	6.789	3.303	0.002	0.041
9	miR-636	0.415	6.376	3.260	0.002	0.045
10	miR-5010-3p	0.341	6.983	3.207	0.002	0.049
11	miR-596	0.542	6.779	3.183	0.003	0.049
12	miR-3676-5p	0.492	10.425	3.170	0.003	0.049
13	miR-4288	0.563	9.374	3.164	0.003	0.049
14	miR-5571-5p	0.934	7.865	3.146	0.003	0.049
15	miR-4742-3p	0.385	8.269	3.136	0.003	0.049
16	miR-302c-5p	0.508	7.561	3.127	0.003	0.049
17	miR-1298	0.553	7.741	3.123	0.003	0.049

(miR-20a, miR-20b, miR-19b and miR-106a) and the miR-15 family (miR-15, miR-16a and miR-195) share many common target genes and are closely interacting (Figure 6a and 6b). The members of these families have been associated with key regulatory roles of carcinogenesis. Members of both the miR-17 and the miR-15 family targets proliferation genes (Figure 5). The miR-15 family members, miR-15a and miR-16, are also known to be associated with regulation of cancer-associated fibroblasts as well as with tumor growth in prostate cancer (Bonci et al., 2008) and angiogenesis in multiple myeloma (Sun et al., 2013).

The miRNAs that were found differentially expressed between carcinoma and stroma play significant roles in various cancers. Briefly, miR-191 and miR-130a are associated with cancer progression in breast and colorectal cancer, respectively (Nagpal et al., 2013; Liu et al., 2013). miR-223 is involved in migration and invasion by targeting the *Mef2c* gene in breast cancer (Yang et al., 2011). It is also linked to increased migration, invasion and decreased cell adhesion in gastric cancer (Li et al., 2011). Both miR-126 and miR-378 have roles in angiogenesis and in repressing the recruitment of mesenchymal stem cells in various cancers (Chou et al., 2013; Lee et al., 2007a). miR-17, miR-19b, miR-451a, and miR-139 all have roles in the regulation of cancer-associated fibroblasts by targeting various downstream cytoskeleton regulatory proteins, cell–cell adhesion molecules and cell-matrix molecules

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

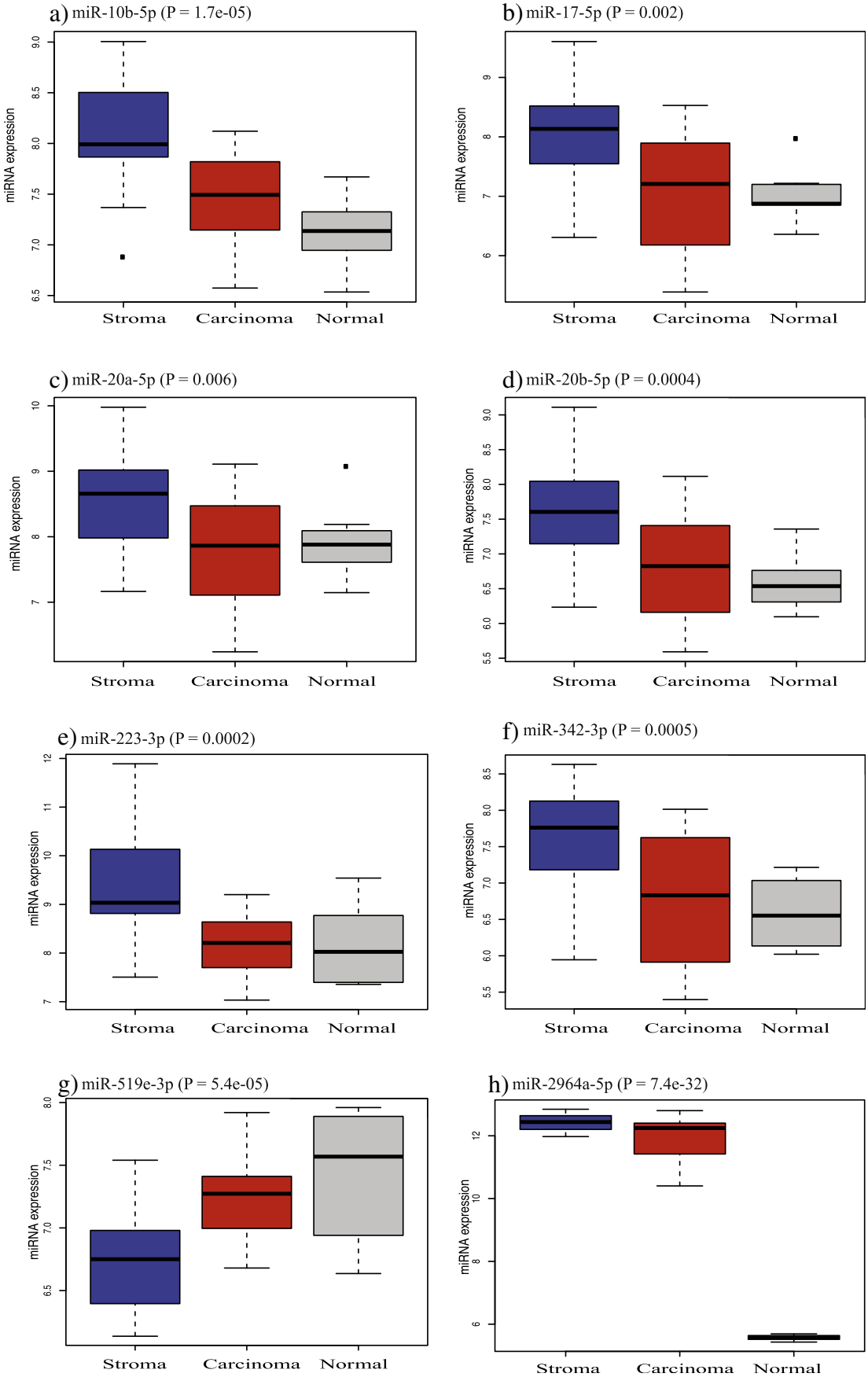
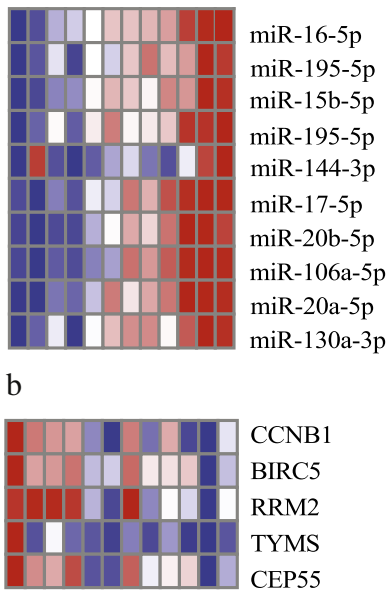


Figure 4 – The boxplots show the relative expression of the eight most differentially expressed miRNAs in stroma (blue), carcinoma (red) and normal tissues (grey) at BH-adjusted P < 0.05.

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

Table 3 – Pathway analysis using mRNAs anti-correlated to miRNAs differentially expressed between carcinoma versus stromal expression and predicted as miRNA target sites by miRanda and Targetscan tools.

No.	Pathways	P value	Adjusted P value	Predicted miRNA-mRNA interactions (in carcinoma)	Predicted miRNA-mRNA interactions (in stroma)
1	Metabolic pathways	1.42e-15	2.58e-13	83	31
2	Endocytosis	1.21e-12	1.10e-10	14	40
3	Pathways in cancer	2.52e-11	1.53e-09	48	79
4	Axon guidance	3.02e-10	1.38e-08	18	34
5	MAPK signaling	7.87e-09	2.86e-07	43	42
6	Cell cycle	2.98e-08	9.04e-07	5	52
7	Focal adhesion	6.42e-08	1.67e-06	31	40
8	ECM receptor remodelling	1.24e-07	2.82e-06	17	13
9	TGF beta signaling	1.12e-07	2.30e-06	11	15
10	p53 signaling pathway	2.24e-06	4.08e-05	5	31



c

miRNA	mRNA	Corelation	P value
hsa-miR-16-5p	CCNB1	-0.71	0.00
hsa-miR-195-5p	CCNB1	-0.64	0.01
hsa-miR-15b-5p	CCNB1	-0.69	0.01
hsa-miR-15a-5p	CCNB1	-0.77	0.00
hsa-miR-144-3p	BIRC5	-0.54	0.03
hsa-miR-20b-5p	RRM2	-0.64	0.01
hsa-miR-17-5p	RRM2	-0.58	0.02
hsa-miR-106a-5p	RRM2	-0.52	0.04
hsa-miR-20a-5p	RRM2	-0.75	0.00
hsa-miR-16-5p	TYMS	-0.70	0.01
hsa-miR-195-5p	TYMS	-0.60	0.02
hsa-miR-130a-3p	TYMS	-0.63	0.01
hsa-miR-15b-5p	TYMS	-0.66	0.01
hsa-miR-15a-5p	TYMS	-0.66	0.01
hsa-miR-16-5p	CEP55	-0.65	0.01
hsa-miR-195-5p	CEP55	-0.66	0.01
hsa-miR-130a-3p	CEP55	-0.68	0.01
hsa-miR-15b-5p	CEP55	-0.73	0.00
hsa-miR-15a-5p	CEP55	-0.74	0.00

Figure 5 – The figure shows the heatmap of the miRNAs and mRNAs associated with proliferation of the cells. 5a) The heatmap shows expression of the miRNA anti-correlated to the proliferation genes in the carcinoma and stromal samples 5b) The heatmap shows expression of proliferation genes in the periampullary adenocarcinoma samples. 5c) The miRNA-mRNA pairs anti-correlated at adjusted P value < 0.05.

(Valastyan and Weinberg, 2011). We identified miR-17 and miR-19 in focal adhesion and in the ECM pathways. miR-195 has a function in proliferation, migration and cell adhesion to the extracellular matrix in endocrine cancers (Jain et al., 2014). In pancreatic cancer, miR-106a suppresses the translation of *CDKN1A, PTEN, TGFBR2* and *KLF11*, activates proliferation-signaling pathways, and induces EMT (Jung et al., 2012). miR-144 induces EMT by downregulating the *Notch-1* gene through an miR-144 regulated mechanism (Sureban et al., 2011). In breast cancer, miR-10b induces EMT by targeting TGF-β1 (Han et al., 2014). Upregulation of the miR-515 family members, miR-515-3p, miR-519d and miR-518e, is associated with metastatic colonization of melanoma cells (Mueller et al., 2009). The miR-10b, miR-155 and miR-106a were recently identified in bile and plasma of pancreatic cancer patients and have been reported as prognostic markers (Cote et al., 2014). We found miR-10b and miR-106b (miR-106a and miR-106b are both members of the miR-17 family) to be upregulated in the stroma, which may be the source of these miRNAs in plasma.

The pathways that are dysregulated in PAs include various pathways (Table 3). Pathways like MAPK, TGF-beta, p53 and cell cycle signaling have previously been reported in an integrated whole genome analysis of mRNAs and miRNAs in PA samples (Sandhu et al., 2015). Other pathways dysregulated in the present study include the ECM interaction, focal adhesion molecule signaling, endocytosis and axon guidance. The integrated analysis of differentially expressed miRNA (candidate miRNAs) between carcinoma and stromal cells, and differentially expressed mRNA between carcinoma and normal samples identified pathways that may be important in regulating tumor–stroma interactions and that facilitate cell proliferation. The predicted miRNA-mRNA interactions in carcinoma and stromal samples identified a number of

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

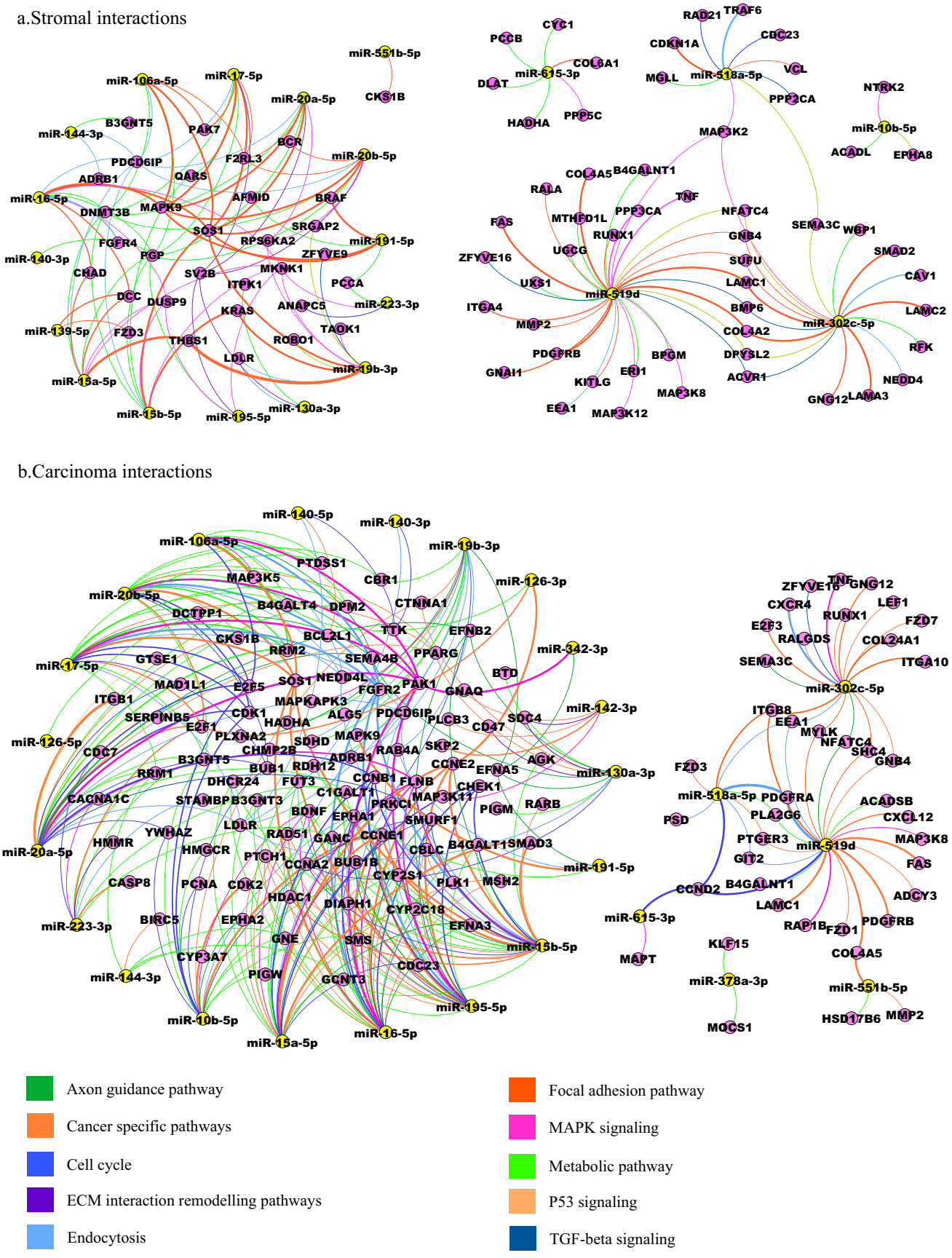


Figure 6 – The figure shows the interaction between the miRNAs and mRNAs in the deregulated pathways in the stroma and carcinoma samples. 6a) The miRNA-mRNA-pathway network shows the miRNA and the targeted mRNAs in the deregulated pathways in the stromal samples. 6b) The miRNA-mRNA-pathway network shows the miRNA and the targeted mRNAs in the deregulated pathways in the carcinoma samples.

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

genes from different pathways. Known oncogenes and tumor suppressor genes like *KRAS, FAS, DCC, SMAD2, MAPK9* and *CDKN1A* were anti-correlated to miRNA expressions in the stroma and were predicted as potential targets. While, in the carcinoma, genes like *CCNE1, FAS, E2F1, FGFR* were anti-correlated to miRNA expressions in the carcinoma cells and were predicted as potential targets.

EMT is an important mechanism of regulating tumor stroma interaction (Zhang et al., 2014; Chou et al., 2013). Various signaling pathways like MAPK, TGF-beta, PI3K-AKT, JNK can cooperate to induce full EMT progression (Lamouille et al., 2014). We found TGF-beta signaling and MAPK signaling pathway enriched in our samples. We identified the genes *SMURF1, SMAD2/3* and *BMP6* as targets of candidate miRNAs in the TGF-beta pathways. SMURFs have a key role in EMT, cell migration, fibrosis and cancer (Izzi and Attisano, 2006) and *SMAD2/3* form complex with *SMAD4* and induces EMT (Xu et al., 2009). Growth factor receptors like *FGFR2, PDGFRA* were found to be associated candidate miRNA targets in the MAPK signaling pathway. These genes regulate EMT by MAPK signaling (Xu et al., 2009). In addition we found the *PAK1, MAPK9 (JNK2)* and *TRAF6* genes in the MAPK signaling pathway, which are important genes in activation of JNK mediated EMT signaling (Lamouille et al., 2014; Alcorn et al., 2008).

Besides EMT regulation by MAPK, TGF-beta and JNK, we found ECM interaction and the focal adhesion pathway to be dysregulated, which also regulate tumor-stroma interaction (Zhang et al., 2014, Chou et al., 2013). ECM interactions not only regulate cell differentiation, its protein domains bind cell adhesion molecule, transduce signals and also regulate microenvironment of cells (Hynes, 2009). The deregulated ECM dynamics can lead to deregulated cell proliferation, migration, loss of cell differentiation and survival (Hynes, 2009). Various collagens *(COL4A2, COL4A5* and *COL24A1)*, laminins *(LAMC1* and *LAMC2)* and integrins *(ITGB8, ITGB1* and *ITGA10)* were differentially expressed and found to be candidate miRNA targets in focal adhesion and ECM interaction pathways. The desmoplasia was induced in pancreatic cancer by TGF-beta signaling, collagens of ECM and the *PDGF* gene in pancreatic cancer (Lohr et al., 2001). The interaction analysis indicates that the TGF-beta signaling pathway, ECM and *PDGF* gene in the focal adhesion pathway might be regulated by candidate miRNAs.

The endocytosis pathway helps in material uptake, regulation of signal transduction and morphogenetic processes like cell adhesion and migration. The deregulated endocytic pathway in human tumors lead to disunion of cell–cell junctions like adherence and tight junctions and transform to malignant cells by delaying endocytosis mediated inactivation of growth factor receptors (Mosesson et al., 2008). We found alterations of the RAB family gene *RAB4A* and the cytoskeleton genes *LDLR, FGFR2, CAV1* associated with the endocytosis pathway as target of candidate miRNAs. The gene *CXCR4* has previously been associated with tumor-stroma interaction in pancreatic cancer (Singh et al., 2012), and *CXCR4* was identified as a candidate target gene in the endocytosis network pathway in our study as well.

The axon guidance pathway is known to be dysregulated in pancreatic cancer (Biankin et al., 2012) and it was also enriched in our samples. Recently, studies have shown the role of nerve cells in promoting tumor progression. The mechanism involves release of neurotransmitters directly into the vicinity of tumor and stroma cells to activate corresponding membrane receptors in pancreatic cancer (Kim-Fuchs et al., 2014, Jobling et al., 2015). We found that genes associated with the axon guidance pathway (*CXCR4, DCC, PAK7, ROBO1* and *KRAS*) were enriched in the interaction analysis.

A challenge in the current study was to obtain carcinoma and stroma samples with satisfactory purity. This is difficult due to the fact that pancreatobiliary PAs grow in a highly dispersed fashion. A degree of heterogeneity can therefore not be excluded and is probably reflected in the heatmap depicting hierarchical clustering (Figure 2), where four of the stromal samples clustered with their respective carcinomas.

miRNAs found to be differentially expressed between carcinoma and stromal cells may contribute to an oncogenic phenotype, and the molecular mechanisms through which they operate should be further investigated. The role of the tumor stroma in PAs is complex, while some studies indicated that it contributes to the aggressiveness of the cancer (Maehara et al., 2001), others recently suggested a protective role (Rhim et al., 2014; Ozdemir et al., 2014). In this study, miRNAs that are differentially expressed in carcinoma and stroma were identified, along with their associated pathways. The findings indicate that these miRNAs are involved in tumor–stroma interactions, which may make them attractive as potential therauptic targets.

Acknowledgments

Supported by grants from the South-Eastern Regional Health Authority, Hole's Foundation, The Radium Hospital Foundation, Oslo University Hospital and Telemark University College. We thank all the patients who participated in the study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2015.10.011.

REFERENCES

Alcorn, J.F., Guala, A.S., van der Velden, J., McElhinney, B., Irvin, C.G., Davis, R.J., et al., 2008. Jun N-terminal kinase 1 regulates epithelial-to-mesenchymal transition induced by TGF-beta1. J. Cell Sci. 121, 1036–1045.

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: Anopen source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Becker, L.E., Lu, Z., Chen, W., Xiong, W., Kong, M., Li, Y., 2012. A systematic screen reveals microRNA clusters that significantly regulate four major signaling pathways. PloS One 7, e48474.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B. 57, 289–300.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., Sander, C., 2008. The microRNA.org resource: targets and expression. Nucleic Acids Res. 36, D149–D153.

Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., et al., 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399–405.

Bonci, D., Coppola, V., Musumeci, M., Addario, A., Giuffrida, R., Memeo, L., et al., 2008. The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nat. Med. 14, 1271–1277.

Chou, J., Shahi, P., Werb, Z., 2013. microRNA-mediated regulation of the tumor microenvironment. Cell Cycle 12, 3262–3271.

Cote, G.A., Gore, A.J., McElyea, S.D., Heathers, L.E., Xu, H., Sherman, S., et al., 2014. A pilot study to develop a diagnostic test for pancreatic ductal adenocarcinoma based on differential expression of select miRNA in plasma and bile. Am. J. Gastroenterol. 109, 1942–1952.

De Wever, O., Mareel, M., 2003. Role of tissue stroma in cancer cell invasion. J. Pathol. 200, 429–447.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S., 2003. MicroRNA targets in Drosophila. Genome Biol. 5, R1.

Erkan, M., Hausmann, S., Michalski, C.W., Fingerle, A.A., Dobritz, M., Kleeff, J., et al., 2012. The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. Nat. Rev. Gastroenterol. Hepatol. 9, 454–467.

Feig, C., Gopinathan, A., Neesse, A., Chan, D.S., Cook, N., Tuveson, D.A., 2012. The pancreas cancer microenvironment. Clin. Cancer Res. Off J. Am. Assoc. Cancer Res. 18, 4266–4276.

Gore, J., Korc, M., 2014. Pancreatic cancer stroma: friend or foe? Cancer Cell 25, 711–712.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34, D140–D144.

Han, X., Yan, S., Weijie, Z., Feng, W., Liuxing, W., Mengquan, L., et al., 2014. Critical role of miR-10b in transforming growth factor-beta1-induced epithelial-mesenchymal transition in breast cancer. Cancer Gene Ther. 21, 60–67.

Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., et al., 2011. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 39, D163–D169.

Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., et al., 2014. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 42, D78–D85.

Hynes, R.O., 2009. The extracellular matrix: not just pretty fibrils. Science 326, 1216–1219.

Izzi, L., Attisano, L., 2006. Ubiquitin-dependent regulation of TGFbeta signaling in cancer. Neoplasia 8, 677–688.

Jain, M., Zhang, L., Boufraqech, M., Liu-Chittenden, Y., Bussey, K., Demeure, M.J., et al., 2014. ZNF367 inhibits cancer progression and is targeted by miR-195. PloS One 9, e101423.

Jobling, P., Pundavela, J., Oliveira, S.M., Roselli, S., Walker, M.M., Hondermarck, H., 2015. Nerve-cancer cell cross-talk: a novel promoter of tumor progression. Cancer Res. 75, 1777–1781.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human microRNA targets. PLoS Biol. 2, e363.

Jolliffe, I.T., 1986. Principal component analysis. Springer-Verlag, ISBN 978-0-387-95442-4, p. 487. http://dx.doi.org/10.1007/b98835.

Jung, C.J., Iyengar, S., Blahnik, K.R., Jiang, J.X., Tahimic, C., Torok, N.J., et al., 2012. Human ESC self-renewal promoting microRNAs induce epithelial–mesenchymal transition in hepatocytes by controlling the PTEN and TGFbeta tumor suppressor signaling pathways. Mol. Cancer Res. MCR 10, 979–991.

Kim-Fuchs, C., Le, C.P., Pimentel, M.A., Shackleford, D., Ferrari, D., Angst, E., et al., 2014. Chronic stress accelerates pancreatic cancer growth and invasion: a critical role for beta-adrenergic signaling in the pancreatic microenvironment. Brain Behav. Immun. 40, 40–47.

Kruskal, W., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47 (260), 583–621. http://dx.doi.org/10.1080/01621459.1952.10, 483441.

Lamouille, S., Xu, J., Derynck, R., 2014. Molecular mechanisms of epithelial–mesenchymal transition. Nat. Rev. Mol. Cell Biol. 15, 178–196.

Lee, D.Y., Deng, Z., Wang, C.H., Yang, B.B., 2007a. MicroRNA-378 promotes cell survival, tumor growth, and angiogenesis by targeting SuFu and Fus-1 expression. Proc. Natl. Acad. Sci. U.S.A. 104, 20350–20355.

Lee, E.J., Gusev, Y., Jiang, J., Nuovo, G.J., Lerner, M.R., Frankel, W.L., et al., 2007b. Expression profiling identifies microRNA signature in pancreatic cancer. Int. J. Cancer. J. Int. du Cancer 120, 1046–1054.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of mammalian microRNA targets. Cell 115, 787–798.

Li, X., Zhang, Y., Zhang, H., Liu, X., Gong, T., Li, M., et al., 2011. miRNA-223 promotes gastric cancer invasion and metastasis by targeting tumor suppressor EPB41L3. Mol. Cancer Res. MCR 9, 824–833.

Liu, L., Nie, J., Chen, L., Dong, G., Du, X., Wu, X., Tang, Y., Han, W., 2013. The oncogenic role of microRNA-130a/301a/454 in human colorectal cancer via targeting Smad4 expression. PLoS One 8 (2), e55532. Epub 2013 Feb 5.

Lohr, M., Schmidt, C., Ringel, J., Kluth, M., Muller, P., Nizze, H., et al., 2001. Transforming growth factor-beta1 induces desmoplasia in an experimental model of human pancreatic carcinoma. Cancer Res. 61, 550–555.

Maehara, N., Matsumoto, K., Kuba, K., Mizumoto, K., Tanaka, M., Nakamura, T., 2001. NK4, a four-kringle antagonist of HGF, inhibits spreading and invasion of human pancreatic cancer cells. Br. J. Cancer 84, 864–873.

Mann, Henry, B., Whitney, Donald, R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60. http://dx.doi.org/10.1214/aoms/1177730491. MR 22058. Zbl 0041.26103.

Mosesson, Y., Mills, G.B., Yarden, Y., 2008. Derailed endocytosis: an emerging feature of cancer. Nat. Rev. Cancer 8, 835–850.

Mueller, D.W., Rehli, M., Bosserhoff, A.K., 2009. miRNA expression profiling in melanocytes and melanoma cell lines reveals miRNAs associated with formation and progression of malignant melanoma. J. Invest. Dermatol. 129, 1740–1751.

Nagpal, N., Ahmad, H.M., Molparia, B., Kulshreshtha, R., 2013. MicroRNA-191, an estrogen-responsive microRNA, functions as an oncogenic regulator in human breast cancer. Carcinogenesis 34, 1889–1899.

Oettle, H., 2014. Progress in the knowledge and treatment of advanced pancreatic cancer: from benchside to bedside. Cancer Treat. Rev. 40, 1039–1047.

Ozdemir, B.C., Pentcheva-Hoang, T., Carstens, J.L., Zheng, X., Wu, C.C., Simpson, T.R., et al., 2014. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell 25, 719–734.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., et al., 2009. Supervised risk predictor of breast

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

cancer based on intrinsic subtypes. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 27, 1160–1167.
Rhim, A.D., Oberstein, P.E., Thomas, D.H., Mirek, E.T., Palermo, C.F., Sastra, S.A., et al., 2014. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer Cell 25, 735–747.
Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., et al., 2007. A comparison of background correction methods for two-colour microarrays. Bioinformatics 23, 2700–2707.
Sandhu, V., Bowitz Lothe, I.M., Labori, K.J., Lingjaerde, O.C., Buanes, T., Dalsgaard, A.M., et al., 2015. Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas. Mol. Oncol. 9, 758–771.
Sato, N., Maehara, N., Goggins, M., 2004. Gene expression profiling of tumor-stromal interactions between pancreatic cancer cells and stromal fibroblasts. Cancer Res. 64, 6950–6956.
Singh, A.P., Arora, S., Bhardwaj, A., Srivastava, S.K., Kadakia, M.P., Wang, B., et al., 2012. CXCL12/CXCR4 protein signaling axis induces sonic hedgehog expression in pancreatic cancer cells via extracellular regulated kinase- and Akt kinase-mediated activation of nuclear factor kappaB: implications for bidirectional tumor-stromal interactions. J. Biol. Chem. 287, 39115–39124.
Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3. Article3.
Smyth, G.K., Michaud, J., Scott, H.S., 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067–2075.
Sun, C.Y., She, X.M., Qin, Y., Chu, Z.B., Chen, L., Ai, L.S., et al., 2013. miR-15a and miR-16 affect the angiogenesis of multiple myeloma by targeting VEGF. Carcinogenesis 34, 426–435.
Sureban, S.M., May, R., Lightfoot, S.A., Hoskins, A.B., Lerner, M., Brackett, D.J., et al., 2011. DCAMKL-1 regulates epithelial–mesenchymal transition in human pancreatic cells through a miR-200a-dependent mechanism. Cancer Res. 71, 2328–2338.
Szafranska, A.E., Davison, T.S., John, J., Cannon, T., Sipos, B., Maghnouj, A., et al., 2007. MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. Oncogene 26, 4442–4452.
Valastyan, S., Weinberg, R.A., 2011. Roles for microRNAs in the regulation of cell adhesion molecules. J. Cell Sci. 124, 999–1006.
Wang, J., Duncan, D., Shi, Z., Zhang, B., 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 41, W77–W83. Web Server issue.
Westgaard, A., Tafjord, S., Farstad, I.N., Cvancarova, M., Eide, T.J., Mathisen, O., et al., 2008. Pancreatobiliary versus intestinal histologic type of differentiation is an independent prognostic factor in resected periampullary adenocarcinoma. BMC cancer 8, 170.
Witten, D.M., Tibshirani, R., Hastie, T., 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10, 515–534.
Witten, D.M., Tibshirani, R.J., 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. Stat. Appl. Genet. Mol. Biol. 8. Article28.
Xu, J., Lamouille, S., Derynck, R., 2009. TGF-beta-induced epithelial to mesenchymal transition. Cell Res. 19, 156–172.
Yang, M., Chen, J., Su, F., Yu, B., Su, F., Lin, L., et al., 2011. Microvesicles secreted by macrophages shuttle invasion-potentiating microRNAs into breast cancer cells. Mol. Cancer 10, 117.
Zhang, B., Kirov, S.A., Snoddy, J.R., 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 33, W741–W748. Web Server issue.
Zhang, Y., Yang, P., Wang, X.F., 2014. Microenvironmental regulation of cancer metastasis by miRNAs. Trends Cell Biol. 24, 153–160.

FURTHER READING

Web links

Cancer registry of Norway: http://www.kreftregisteret.no/en/.
miRbase: http://www.mirbase.org/ (release 21, June, 2014).
miRNA.org: http://www.microrna.org/microrna/home.do (release August, 2010).
miRTarBase: http://mirtarbase.mbc.nctu.edu.tw/ (Release 4.5, November 1, 2013).
TargetScan: http://www.targetscan.org/ (Release 6.2, June 2012).
WebGestalt: http://bioinfo.vanderbilt.edu/webgestalt/ (Update 24 May, 2013).

Please cite this article in press as: Sandhu, V., et al., Differential expression of miRNAs in pancreatobiliary type of periampullary adenocarcinoma and its associated stroma, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.011

Supplementary figure and tables legends

Supplementary Figure S.1: The figure shows the QC plots for all the samples used in the project.

Supplementary Figure S2: The box plots showing the relative expression of the 35 of the differentially expressed miRNAs in stroma (blue), carcinoma (red) and normal tissues (grey) at BH - adjusted $P < 0.05$.

Supplementary Table S.1: The table shows the estimated tumor percentage of samples used in the analysis.

Supplementary Table 2: The table shows the miRNAs differentially expressed in comparison between carcinoma, stroma and normal samples.
S.2a: List of differentially expressed miRNAs in stroma versus normal tissues.
S.2b: List of differentially expressed miRNAs in tumor versus normal tissues.
S.2c: List of differentially expressed miRNAs between three groups; carcinoma, stroma and normal tissues identified using the Kruskal-Wallis ANOVA.

Supplementary Table S.3: List of mRNA differentially expressed between fresh frozen tumor and stroma samples versus normal samples at BH - adjusted $P < 0.05$.

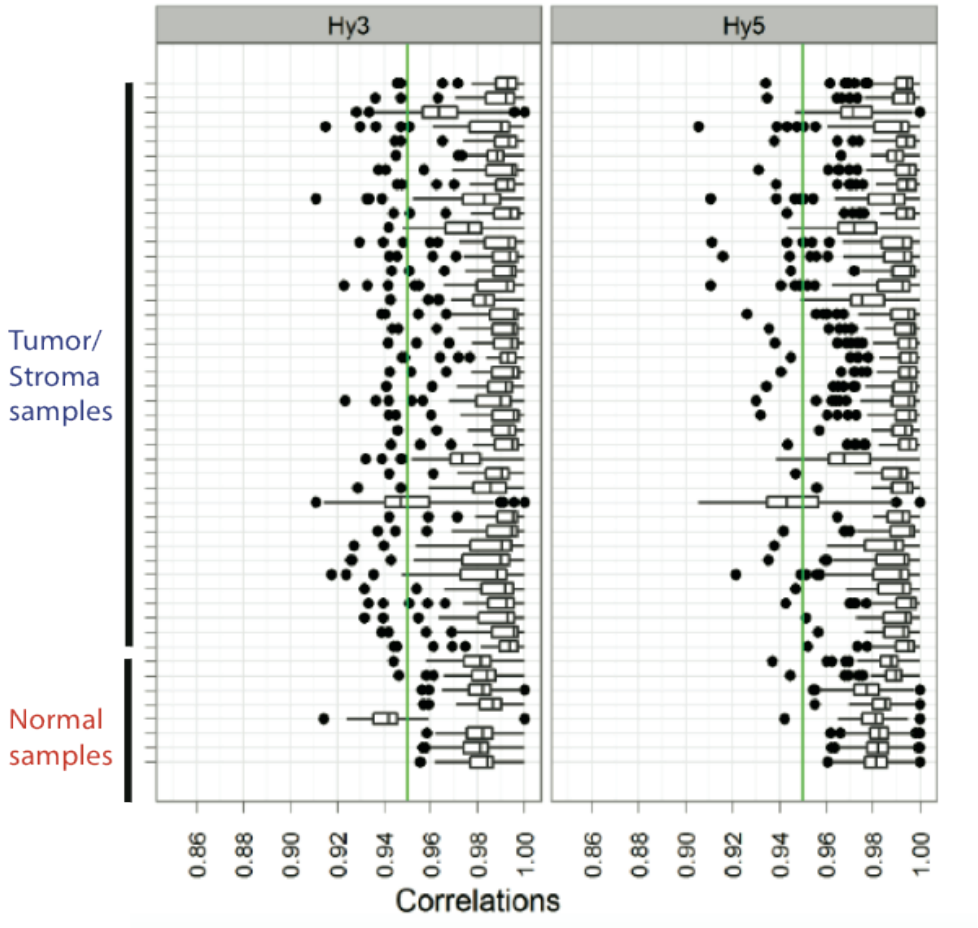
Supplementary Table S.4: The table shows the miRNA-mRNA pairs interaction in stroma and carcinoma samples. 4a: List of miRNA-mRNA pairs that are anti-correlated at $P < 0.05$ in stroma samples (Predicted and Validated values). 4b: List of miRNA mRNA pairs that are anti-correlated at $P < 0.05$ in carcinoma samples (Predicted and Validated values).

Supplementary Tables S.5a- 5j: List of miRNA-mRNA interactions in enriched pathways for the carcinoma and the stromal samples.

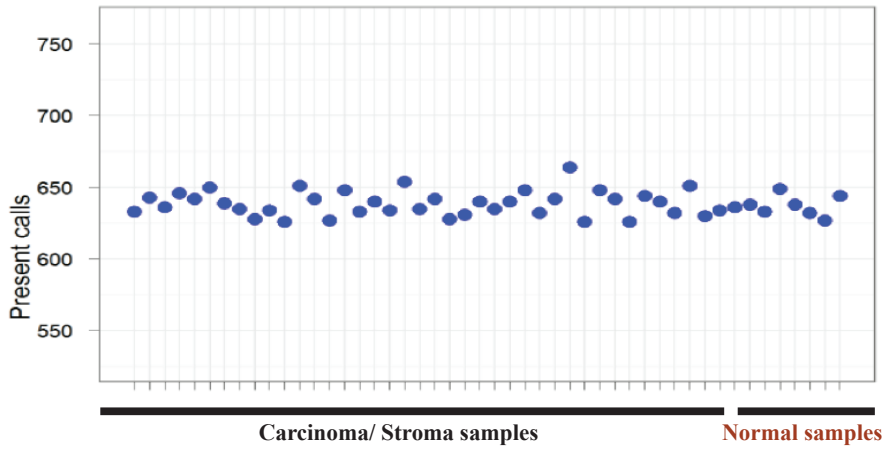
Supplementary File S1: All the scripts used in the article.

Supplementary Figure S.1

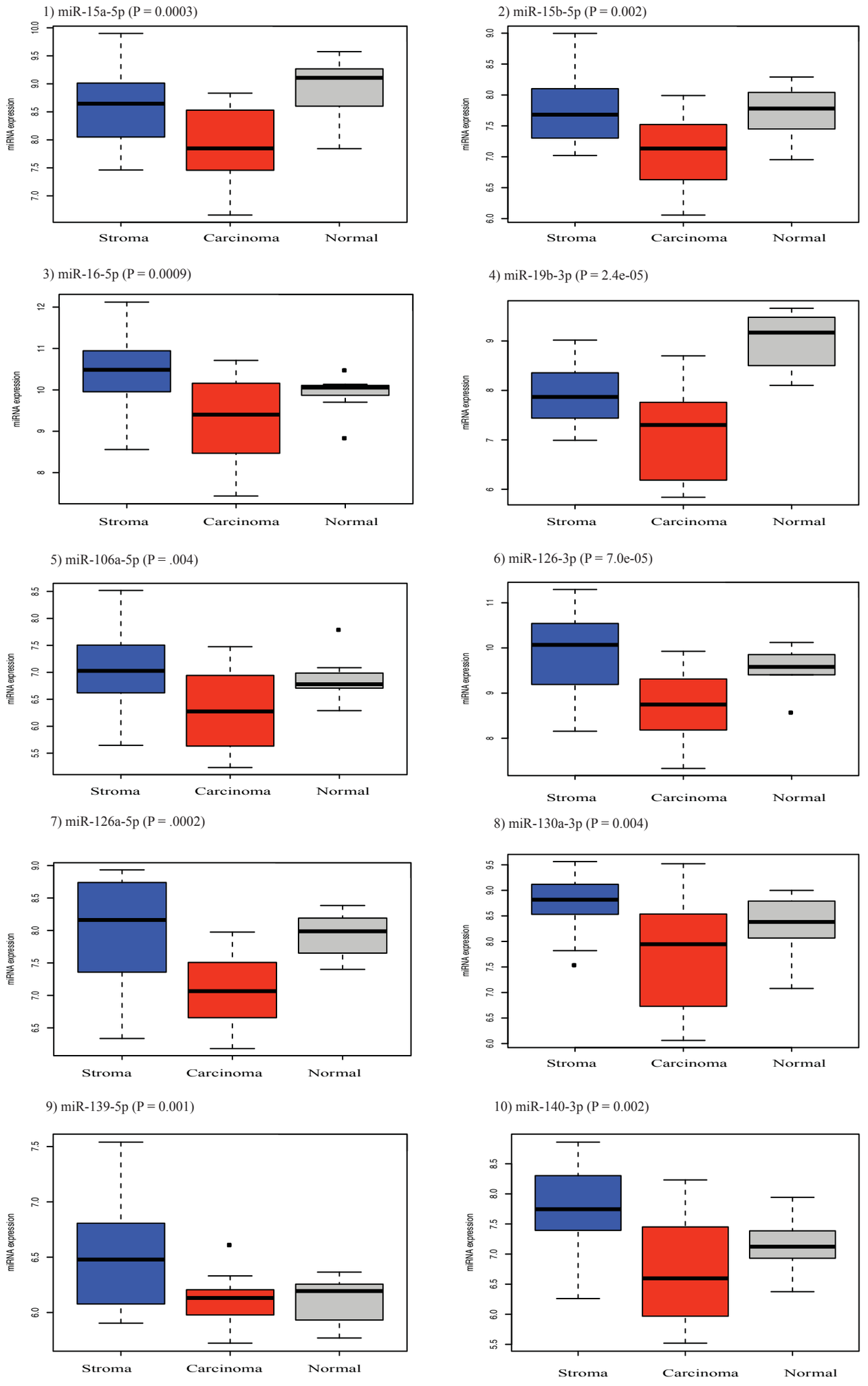
A. QC spike in box plots for carcinoma, stroma and normal samples.



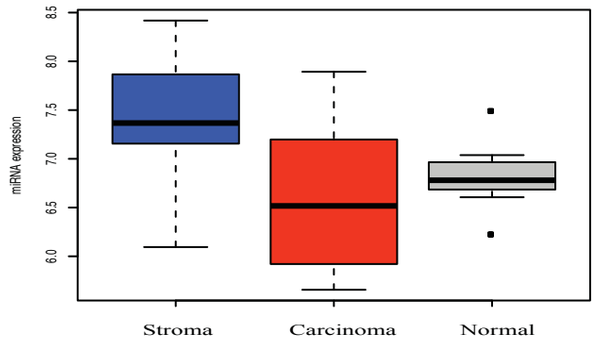
B. Present calls for carcinoma, stroma and normal samples.



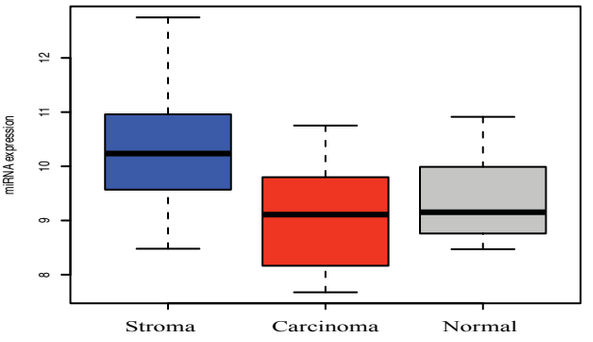
Supplementary Figure S2



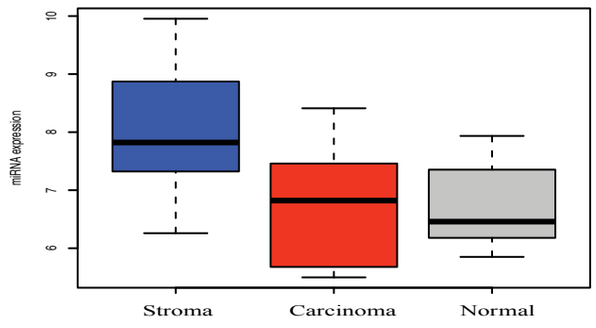
11) miR-140-5p (P = 0.007)



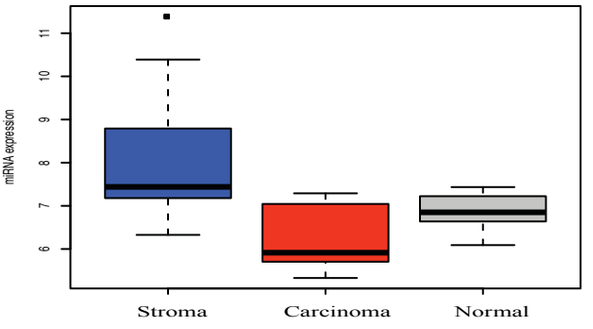
12) miR-142-3p (P = 0.008)



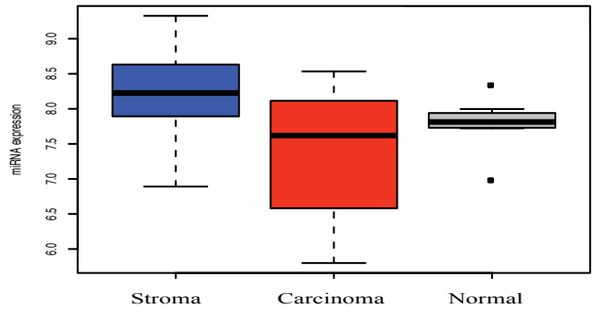
13) miR-142-5p (P = 0.0003)



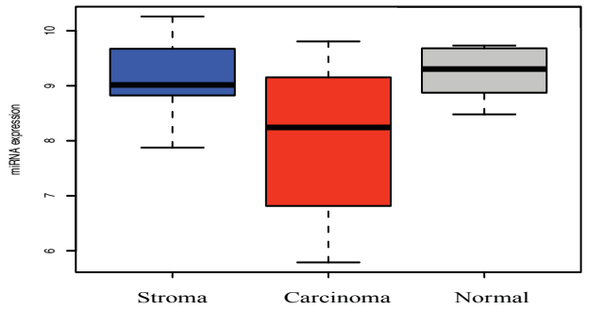
14) miR-144-3p (P = 2.7e-05)



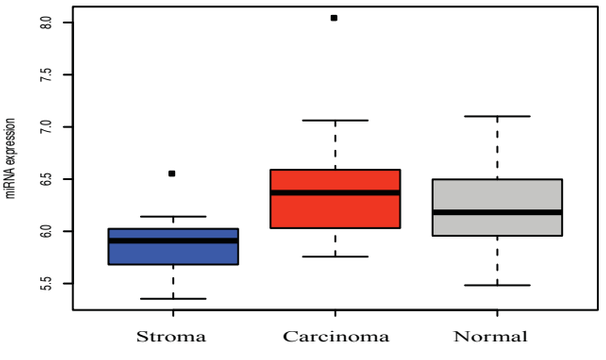
15) miR-191-5p (P = .003)



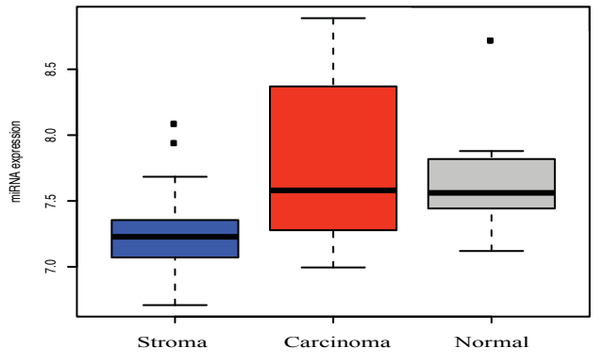
16) miR-195-5p (P = .001)



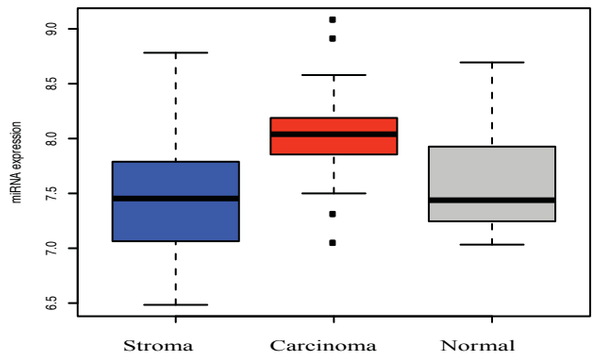
17) miR-551b-5p (P = .003)



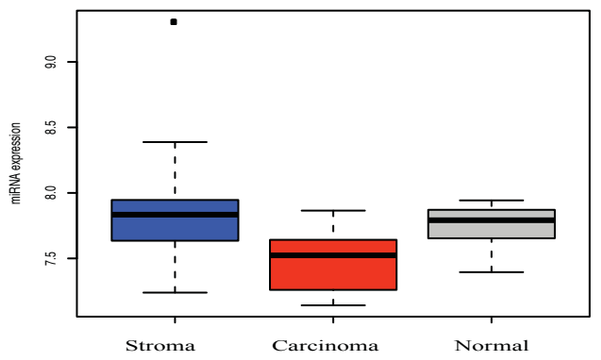
18) miR-302c-5p (P = 0.01)



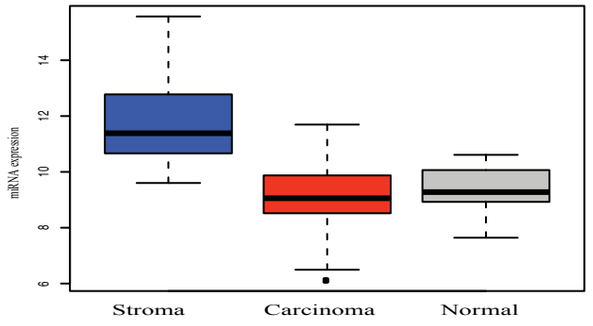
19) miR-1298 (P = 0.01)



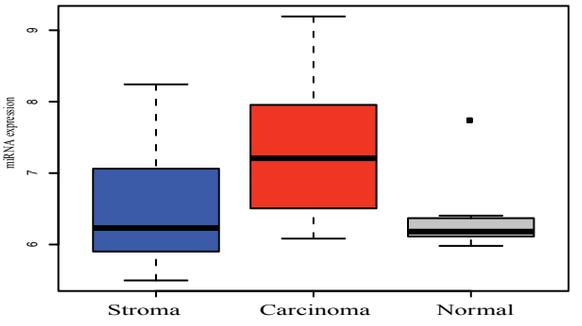
20) miR-378a-3p (P = 0.009)



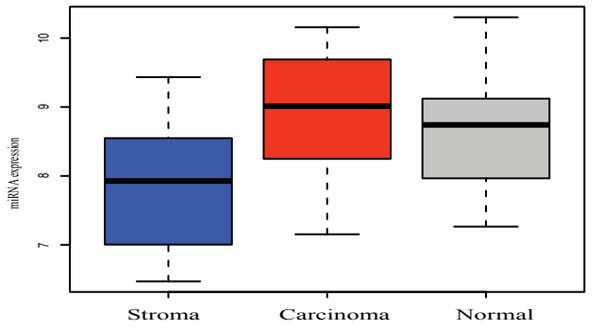
21) miR-451a (P = 3.9e-06)



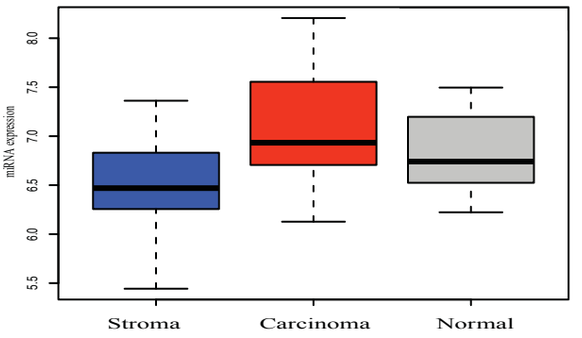
22) miR-518a-5p (P = 0.003)



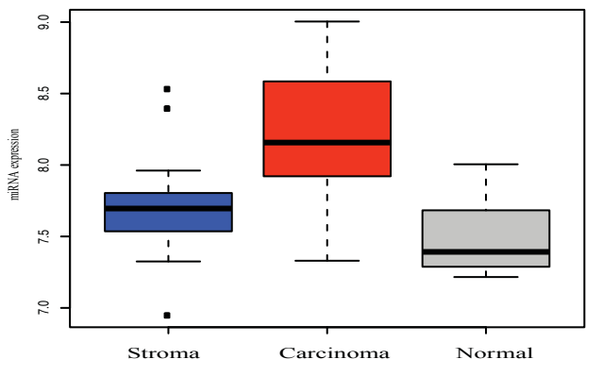
23) miR-519d (P = 0.001)



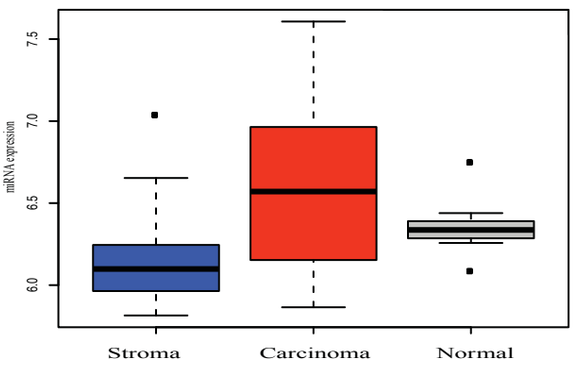
24) miR-596 (P = 0.001)



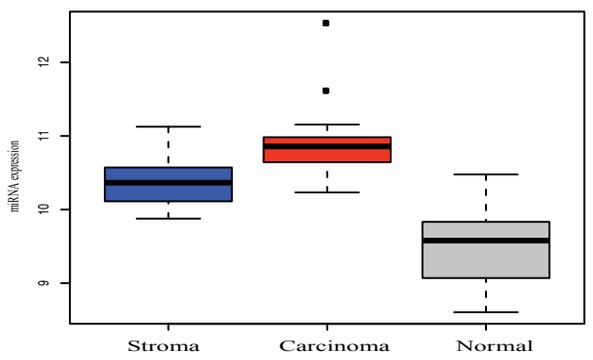
25) miR-615-3p (P = 7.9e-05)



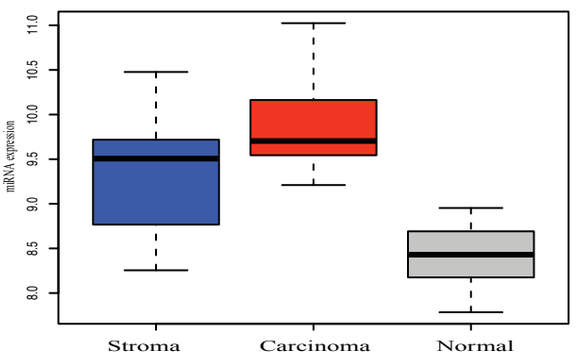
26) miR-636 (P = .008)



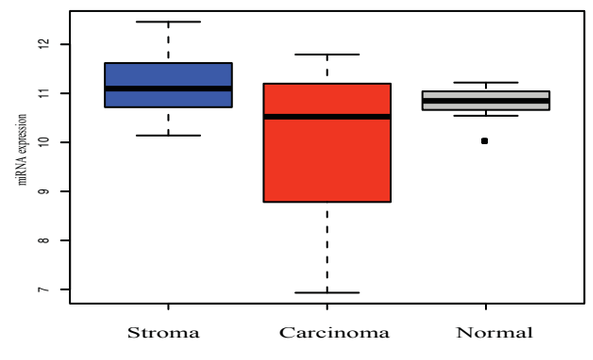
27) miR-3676-5p (P = 1.4e-07)



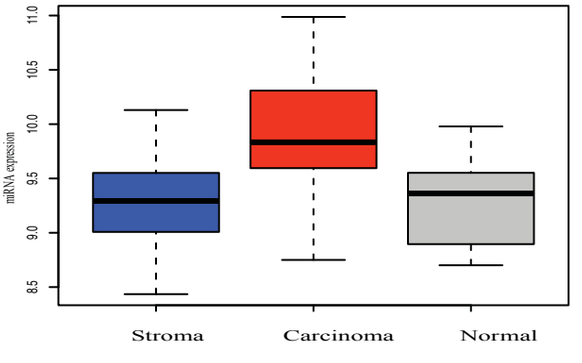
28) miR-4288 (P = 8.7e-07)



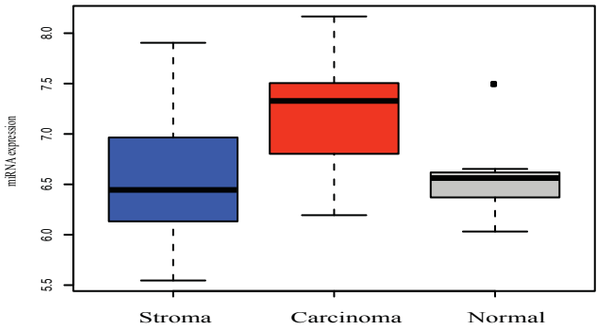
29) miR-4301 (P = 0.009)



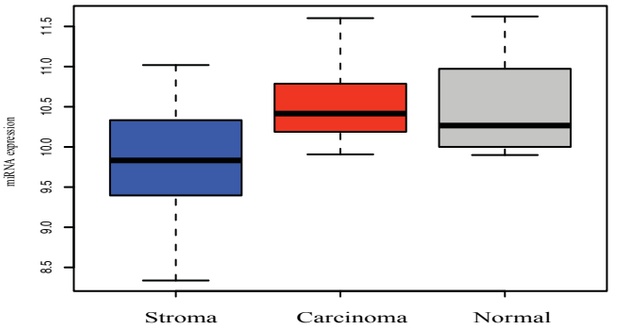
30) miR-4436-5p (P = 0.003)



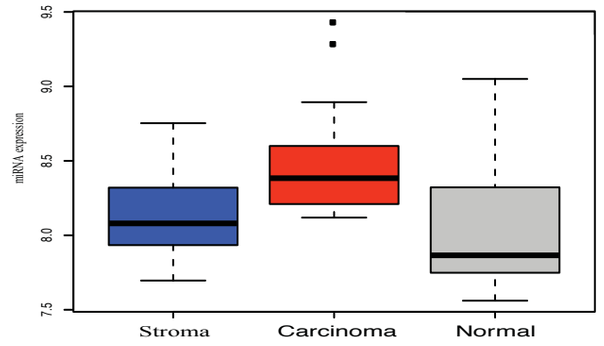
31) miR-4506 (P = 0.002)



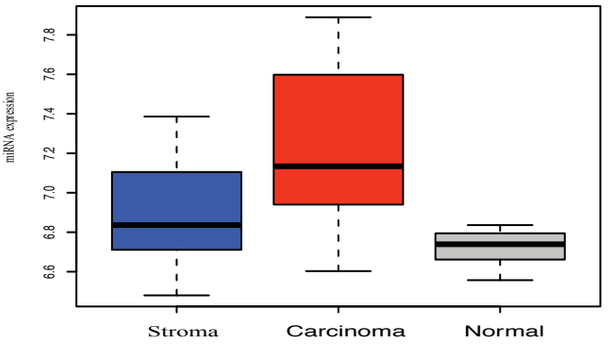
32) miR-4633-3p (P = 0.001)



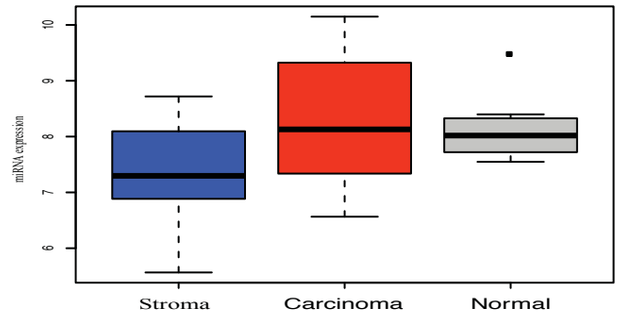
33) miR-4742-3p (P = 0.003)



34) miR-5010-3p (P = 0.0006)



35) miR-5571-5p (P = 0 .01)



Supplementary Table S.1

No.	Patient id	Tumor %
1	Patient 1	80
2	Patient 2	35
3	Patient 3	30
4	Patient 4	70
5	Patient 5	40
6	Patient 6	75
7	Patient 7	50
8	Patient 8	50
9	Patient 9	50
10	Patient 10	50
11	Patient 11	50
12	Patient 12	35
13	Patient 13	Excluded
14	Patient 14	75
15	Patient 15	80
16	Patient 16	75
17	Patient 17	70
18	Patient 18	50
19	Patient 19	90
20	Patient 20	Excluded

Supplementary Tables "S.2", "S.3", "S.4", "S.5" are available online
"http://www.sciencedirect.com/science/article/pii/S1574789115001891

Paper III

The genomic landscape of pancreatic and periampullary adenocarcinoma

Sandhu V., Wedge D.C., Bowitz Lothe I.M., Labori K.J., Dentro S., Buanes T., Skrede M.L., Dalsgaard A.M., Lingjærde O.C., Børresen-Dale A.L., Ikdahl T., Van Loo P., Nord S., Kure E.H.

Manuscript.

The genomic landscape of pancreatic and periampullary adenocarcinoma

V Sandhu[1,7], DC Wedge[8], IM Bowitz Lothe[1,2], KJ Labori[3], SC Dentro[8,11], T Buanes[3,4], ML Skrede[1], AM Dalsgaard[1], OC Lingjærde[5,6], A-L Børresen-Dale[1,4], T Ikdahl[9,10], P Van Loo[11,12], S Nord[1], EH Kure[1,7,*]

[1]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[2]Department of Pathology, Oslo University Hospital, Oslo, Norway
[3]Department of Hepato-Pancreato-Biliary Surgery, Oslo University Hospital, Oslo, Norway
[4]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[5]Department of Computer Science, University of Oslo, Oslo, Norway
[6]K.G. Jebsen Centre for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[7]Department for Environmental Health and Science, Telemark University College, Bø in Telemark, Norway
[8]Wellcome Trust Sanger Institute, Hinxton, United Kingdom
[9]Department of Oncology, Oslo University Hospital, Oslo, Norway
[10]Akershus University Hospital, Nordbyhagen, Norway
[11]The Francis Crick Institute, London, UK
[12]Department of Human Genetics, University of Leuven, Leuven, Belgium

*Corresponding author: Elin H. Kure, Elin.Kure@rr-research.no

Abstract

Despite advances in diagnostics for pancreatic and periampullary adenocarcinoma, the 5 years overall survival is still < 5%. Periampullary tumors are neoplasms that arise in the vicinity of the ampulla of Vater (PDAC, duodenum, bile duct and ampulla). In this study we have analyzed copy number aberrations using Affymetrix SNP 6.0 arrays in 60 periampullary adenocarcinomas from Oslo University Hospital. We identified genome wide copy number aberrations, putative driver genes, enriched pathways and potential prognostic markers. These results were validated in a cohort from The Cancer Genome Atlas. We observed and validated frequent gains of the chromosomal loci 1q, 3q, 7p, 8q, 13q, 18p and 19q; and loss of 1p, 3p, 6, 8p, 9p, 17p, 18q and 22q. In contrast to many other solid tumors periampullary adenocarcinomas have higher frequencies of genomic deletions than of gains. Deletions of 17p13 and 18q21/22 co-occurred in 60% of the tumors. The genes in these loci are associated with cell cycle, apoptosis, p53 and Wnt signaling. Significant correlation between copy number aberrations and gene expression *in cis* was observed in 19q12 (*CCNE1* amplicon) and 17q12 (*ERBB2* amplicon). Morphological subtypes of periampullary adenocarcinomas (i.e. pancreatobiliary or intestinal) harbor many common genomic aberrations. However, gains of 13q and 3q, and deletions of 5q are specific to the intestinal subtype. Gains of 18p11 (18p11.21-23, 18p11.31-32) and 19q13 (19q13.2, 19q13.31-32) and subsequent overexpression of the genes in these loci were associated with decreased survival, and may serve as potential prognostic markers.

Introduction

Pancreatic cancer is the fourth most common cause of cancer-related deaths in the Western countries and is projected to be second leading cause of cancer death by 2030 *(Rahib et al, 2014)*. The incidence rate and mortality rate for pancreatic cancer is almost equal with the five-year survival rate < 5%. In general, tumor evolution is either driven by mutations or by copy number aberrations (CNA) *(Ciriello et al., 2013)*. CNAs have critical roles in activating oncogenes and inactivating tumor suppressor genes, thereby targeting the various hallmarks of cancer *(Beroukhim et al., 2010*, *Hanahan et al., 2000)*. Studies have shown that alterations in pancreatic cancer range from single nucleotide polymorphisms to large-scale rearrangements, leading to copy number gains, amplifications and deletions *(Waddell et al., 2015, Jones et al., 2008)*. In the field of cancer genomics the focus is on identifying altered genomic regions and pathways by using high throughput technologies and further relate these to biological and phenotypic effects. This in-depth knowledge has already led to substantial advances in diagnostics and therapeutics in other cancers such as the *HER2* oncogene in breast cancer *(Coussens et al., 1985)*. This is the most common therapeutic target for treatment of breast cancer by monoclonal antibody trastuzumab that can bind to and inactivate the HER2 receptor *(Carter et al, 1992)*.

A number of studies have been published documenting whole genome expression profiling data with deregulated mRNAs, miRNAs and pathways in pancreatic cancer *(Sandhu et al., 2015, Iacobuzio-Donahue et.al, 2012, Jones et al., 2008)*. Previous studies on relatively smaller sample sizes of pancreatic ductal adenocarcinomas (PDAC) have documented homozygous deletions of 1p, 3p, 6p, 9p, 12q, 13q, 14q, 17p and 18q, and amplifications of 1q, 2q, 3q, 5q, 7p, 7q, 8q, 11p, 14q, 17q and 20q (*Harada et al., 2008, 2009, Gutierrez et al., 2011*). Recently a study of 75 PDACs and 25 cell lines derived from PDAC patients were analyzed using Illumina SNP arrays and whole genome SOLID sequencing *(Waddell et.al, 2015)*. The results showed that PDAC is a disease of mainly structural alterations, classified by the number and distribution of structural variation events. Another recent publication showed that the common variations at 2p13.3, 3q29, 7p13 and 17q25.1 are associated with susceptibility to pancreatic cancer

(Childs et al., 2015). Despite these studies, the knowledge about pancreatic cancer initiation, development and progression is scarce. Some of these limitations are due to small samples sizes and lack of validation in different cohorts.

In the present study a total of 187 tumors (60 patients from the Oslo University Hospital (OUH) cohort and 127 patients from The Cancer Genome Atlas (TCGA) cohort) were profiled using Affymetrix SNP 6.0 arrays. To gain insight into copy number changes in this cancer entity that is characterized by low tumor purity, we analyzed the data using the Battenberg pipeline, which is able to detect subtle changes in copy number signals through phasing both parental haplotypes *(Nik-Zainal et al., 2012, doi: 10.5281/zenodo.16107)*. We identified CNAs in tumors originating from the pancreatic ducts, the bile duct, the ampulla and the duodenum, collectively called periampullary adenocarcinomas (PAs) in the OUH cohort of 60 patients. We found frequent gains of the chromosomal loci 1q, 3q, 7p, 8q, 13q, 18p and 19q; and frequent losses of 1p, 3p, 6, 8p, 9p, 17p, 18q and 22q. The putative driver genes and pathways deregulated in the PA samples were identified in the OUH cohort and validated in the TCGA cohort. The patients with gains of 18p11 (18p11.21-23, 18p11.31-32), 19q13 (19q13.2, 19q13.31-32) and subsequent overexpression of genes within these genomic loci had poor prognosis. In the present study we identified altered genomic regions, genes and pathways using genotyping and expression data analysis to explore the tumor biology and its significance for survival. The majority of the significant observations were validated in the TCGA cohort.

Results

1. Genomic aberrations in periampullary adenocarcinomas

Genomic aberrations that occurred most frequently in the PAs were identified with high confidence by comparing the aberration patterns in the OUH and the TCGA cohorts. In the OUH cohort, deletion of chromosome 18q was the most frequent event, occurring in 46 of the 60 tumors (77%). Focal deletions of 9p21 and 9p23 were found in 70% of the tumors, loss of 17p13 and 17p12 were found in 68% and loss of 6q and 8p in > 50% of the tumors. Focal gains were observed for the following locations: 8q24.21 (32%),

18q11.2 (33%), 13q33.3 (30%), 3q25.31 (30%), 7p21.3 (28%), 19q13.2 (25%), 1q25.3 (25%) and 1q31 (25%) (Figure 1).

In the TCGA cohort, deletion of 18q was also the most frequent event, deleted in 100 of the 127 tumors (78%). Focal deletion of 9p21 and 9p23 were found in 62% of the tumors, loss of 17p12 and 17p13 in 74% and loss of 6q and 8p were found in 61% and 43% of the tumors, respectively. Focal gains of the chromosomal locations 8q24.21 (43%), 18q11.2 (28%), 13q33.3 (22%), 3q25.31 (20%), 7p21.3 (31%), 19q13.2 (27%), 1q25.3 (42%) and 1q31 (39%) were also observed (Figure 1).

The frequency plots of chromosomal aberrations in the OUH and TCGA cohorts are shown in Figure 1. The deletions and copy number gain events in the individual tumors of the OUH and TCGA cohorts are plotted as a heatmap in Supplementary figure S1. The aberration patterns of the PAs were broadly similar in the two cohorts. Approximately 30% of the tumors in both cohorts have copy number gains, while most (> 75%) of the tumors carry deletions.

2. Clinicopathological characteristics of periampullary adenocarcinomas

The clinicopathological characteristics of the OUH and TCGA cohorts were comparable, with the majority of the tumors of stage T3 and grade G2 in both cohorts (Table 1). The average genome instability index (GII) for the OUH and the TCGA cohorts were 0.33 and 0.37, respectively. The clinicopathological characteristics of only 55 samples (excluding 3 IPMNs and 2 xenograft cell lines) are presented in Table 1. The cell lines were included in the analysis since they purely contain carcinoma cells, while the tumor samples are heterogeneous and contain signals from stromal cells. CNA profiles of two of the three xenograft cell lines were compared to their original tumors. One of the cell lines had a different average ploidy than its corresponding tumor (Supplementary figure S2). There was no corresponding tumor tissue available for CNA profiling for the third xenograft cell line. The three Intraductal Papillary Mucinous Neoplasia (IPMN) samples were more normal-like with very few chromosomal aberrations, like deletions of chromosome 9p and 10q and amplification of 1q (Supplementary figure S3).

Correlation analysis between ploidy and GII showed a positive correlation for both the OUH cohort (0.48, Pearson's correlation at $P < 0.0001$) and the TCGA cohort (0.72, Pearson's correlation at $P < 0.0001$) (Supplementary Figure S4).

3. Copy number changes characterizing periampullary adenocarcinomas with different morphology

To identify CNAs between subtypes of PAs, we calculated frequencies of copy number gains and deletions and plotted them based on morphology (Figure 2A) as well as site of origin (Supplementary figure S5). The aberrations specific to morphological subgroups or sites of origin ($P < 0.05$, chi squared test) are plotted in Figures 2B and 2C. Strikingly, the frequency plots based on morphology showed multiple alterations specific to either the pancreatobiliary subtype (from PDAC, bile duct and ampulla) or the intestinal subtype (from ampulla and duodenum). The genomic aberrations specific to the intestinal subtype are loss of 4q, 5q and gains of 3q and 13 chromosomal loci.

Copy number gains at chromosomes 13q14.3/22.1/32.1/34 and deletions of loci in chromosomes 5q11.2/13.3/21.3, 18p11.22/11.23/11.31 and 18q12.3 were more evident in the intestinal than in the pancreatobiliary subtype (Figures 2A and 2B). Although the number of samples in the pancreatobiliary ($n = 41$) and the intestinal ($n = 16$) subtypes are small, the differences are highly significant (chi square test *P*-values < 0.001). In contrast, whole arm deletion of chromosome arms 6q, 8p, 9p, 17p, 18p and 18q were more common in the pancreatobiliary than in the intestinal subtype (Figure 2B). Focal deletion of 18q11 was more frequent in the ampulla of the intestinal subtype and the duodenum than in the tumors of pancreatobiliary subtype (chi square test *P*-values < 0.001) (Figure 2C). The deletions of 4q (57%) and 5q (57%) were observed in the ampulla of intestinal subtype and in the duodenum, respectively.

Due to strikingly different aberration patterns in the two morphological subtypes of PAs, clustering of samples using the PAM50 gene signature was performed *(Parker et al., 2009)*. The PAM50 gene signature clustered the samples broadly into intestinal and pancreatobiliary subtypes, where the latter clustered into basal-like and classical. Using

the gene signature defined by Moffitt *(Moffitt et al., 2015)* also clustered our samples into basal-like and classical subtypes (Supplementary figure S6A-B). The tumors in the basal-like cluster were poorly differentiated and were highly proliferative as compared to the classical subtype (P = 5.9e-05, two-way ANOVA test) (Supplementary figures S6A and S6D).

4. Driver genes in periampullary adenocarcinomas

Putative driver genes were identified in the most frequently deleted and amplified chromosomal regions by mapping the genes to known oncogene and tumor suppressor gene lists as described in Materials and Methods. We identified putative driver genes in each chromosomal locus, and the average frequencies of putative driver genes in deleted and amplified genomic loci for both cohorts are presented in Figure 3. The genes located on frequently deleted locations are, *RUNX3* and *EPHB2* on chromosome 1p; *PBRM1* and *LTF* on chromosome 3p; *MYB* and *PRDM1* on chromosome 6; *CDKN2A* and *CDKN2B* on chromosome 9; *MAP2K4* and *PIK3R5* on chromosome 17. Multiple genes including *MAPK4, SMAD2, SMAD4, DCC* and *BCL2* were found on chromosome 18, which was the most frequent deletion event in both cohorts. The genes *AKT3* on chromosome 1q; *EGFR, PIK3CG* on chromosome 7; *MYC, PTK2* on chromosome 8; *ERCC5* on chromosome 13 and *CCNE1* on chromosome 19 were amplified. Supplementary table S1 shows the amplified or deleted genes in PAs and their frequencies of aberration in both the cohorts. Putative driver genes in the regions that were exclusively aberrant in the intestinal subtype are; *KLF5, RAP2A* and *IRS2* on chromosome 13 which were amplified, and *PIK3R1, PLK2* and *PPAP2A* on chromosome 5 and *PTPRM* gene on chromosome 18p which were deleted. These genes are frequently deleted or amplified in the intestinal subtypes of PAs and are known tumor suppressors or oncogenes.

5. Integrated analysis of copy number and gene expression data

To determine genomic hotspots, we carried out correlation analysis of copy number and gene expression data. We found that amplification or deletion of chromosomal loci were associated with up- or downregulation of genes in both the cohorts. The upregulation of 974 and downregulation of 1060 genes in OUH cohort and upregulation of 3566 genes

and downregulation of 4953 genes in TCGA cohort were associated with copy number aberration *in cis*. Of these 795 of 2034 genes from OUH cohort were validated in the TCGA cohort. Expressions of *FBXL20* and *MED1* on 17q12 (*ERBB2*-amplicon) and *POP4, CCNE1, C19orf12* and *UQCRFS1* on 19q12 were highly correlated with the tumors copy number profiles *in cis* ($P < 0.001$, Pearson's correlation test) in the OUH cohort. We validated that the expression of *POP4, CCNE1, C19orf12* and *UQCRFS1* on 19q12 were associated with the tumors copy number profiles in TCGA cohort. Deletions and gains of various chromosomal loci were associated with downregulation of tumor suppressor genes like *SMAD2, PBRM1* and *TNFRSF10A,* and overexpression of oncogenes such as *JAK2* and *FAS* (Supplementary table S2), respectively in both the cohorts. Several of our reported driver genes were found significantly correlated with copy number alteration *in cis* for both cohorts. Examples include *NEK3, GRAMD3, CCNE1, PHLPP1, PINX1, MLLT3* and *CD274*. The Pearson correlation coefficient values, *P*-values from correlation test and frequencies of deletions/amplifications are reported for both cohorts (Supplementary tables S2A-2D).

6. Co-occurrence of chromosomal aberrations in periampullary adenocarcinomas

The co-occurrence of the most frequently occurring events was compared to identify the likelihood of co-involvement in dysregulation of pathways in PAs. Chromosomal deletions at locations 17p13 and 18q21/18q22 co-occurred in both the cohorts at $P < .01$ (Fisher exact test). The co-occurrence frequency of deletion of 17p13 and 18q21/22 in the tumors from the OUH and the TCGA cohorts were 60% and 62%, respectively. Deletion of 17p13 occurred in 63% (OUH cohort) and 74% (TCGA cohort), and deletions of 18q21/18q22 occurred in 70% (OUH cohort) and 79% (TCGA cohort) of the tumors. The candidate genes located on chromosome 17p13 and 18q21 are involved in cell cycle regulation (*TP53* and *YWHAE* (17p13), *SMAD2* and *SMAD4* (18q21)), p53 signaling *(TP53* (17p13) and *SERPINB5* and *PMAIP1* (18q21)), apoptosis *(TP53 and PIK3R5* (17p13) and *BCL2* (18q21)) and Wnt signaling *(TP53* (17p13) and *SMAD4* (18q21)).

7. Gene Set Enrichment Analysis

To identify pathways deregulated in PAs, gene set enrichment analysis was performed using the WebGestalt tool *(Zhang et al., 2005, Wang et al., 2013)*. The pathways significantly enriched in both cohorts (FDR < 0.05) and genes deregulated in more than 20% of the samples in each pathway are reported in **Supplementary table S3.** The top pathways deregulated in both the cohorts and also significantly enriched in both gene expression *(Sandhu et al., 2015)* and copy number data are reported in Table 2. The pathways associated with frequently co-deletions of 17p and 18q (cell cycle, apoptosis and p53 signaling) are also reported in Table 2.

8. Prognostic implications of copy number gain

To determine the prognostic implications of copy number gain, survival analysis were performed for the focally amplified regions and also for genes located in these regions. We focused on prognostic values of focal amplicon regions since an oncogenes or tumor suppressor is more likely to drive focal regions than whole arm amplification or deletion. Kaplan-Meier survival analyses showed that gain of the chromosomal region 18p11 (18p11.21-23, 18p11.31-32) were associated with decreased disease free survival (DFS) and overall survival (OS) at $P < 0.01$. The amplifications of the genes *RAB12* (Ras oncogene family member) and *COLEC12* located on 18p11.22 and 18p11.32 respectively were associated with decreased DFS (Figures 4 A-C). Gain of 19q13 (19q13.2, 19q13.31-32) and amplification of the genes *SERTAD3* and *ERCC1* located on 19q13.2 and 19q13.32 respectively, were associated with decreased OS at $P < 0.05$ (Figures 4 D-F).
The prognostic relevance of PAM50 classification into intestinal, basal-like and classical subtypes was determined by plotting a Kaplan-Meier survival curve. Patients with basal like tumor had worst survival (Supplementary figure 6C).

Discussion

The frequencies of gains and deletions were comparable between the OUH and TCGA cohorts, where ≈ 30% of the samples had copy number gains and ≈ 75% had deletions as the most frequent events. The validation using the TCGA cohort supports the findings in the OUH cohort with respect to genomic aberration patterns, the candidate driver genes

and pathways. We identified gains of 18p11 (18p11.21-23, 18p11.31-32) and 19q13 (19q13.2, 19q13.31-32) as potential prognostic markers in the OUH cohort. Due to limited clinical annotation of the TCGA data, we could not validate it in the TCGA cohort.

The correlation analysis between gene expression and copy number data identified genes that play an important role in pancreatic cancer tumor biology. Out of nine genes in the 19q12 amplicon, the expressions of four genes (*CCNE1, POP4, UQCRFS1* and *C19orf12*) were significantly correlated with gain of 19q12. Studies have identified 19q12 gain in ER-negative grade III breast cancers, and that silencing of *POP4* and *CCNE1* reduce cell viability in cancer cells harboring this amplification *(Natrajan et al., 2012)*. Amplification of *CCNE1* is typically found in basal-like breast cancers, and is associated with increased proliferation *(Cancer Genome Atlas, N., 2012)*. We found relatively higher expression of *CCNE1* in the basal-like as compared to the classical subtype ($P < 0.01$, t-test). Further, gain of 17q12 was associated with overexpression of *FBXL20* and *MED1* genes within the *ERBB2* amplicon. *FBXL20* and *MED1* have been identified as frequent centromeric border genes. These co-amplified genes are important contributor to cancer progression *(Kao and Pollack, 2006)*.

The CNA analysis using the Battenberg pipeline is advantageous for low cellularity samples like PAs. As a consequence, we have high power to detect aberrations specific to individual subtypes. One of the important findings is that the PA are more distinct at the level of morphology than at the site of origin, which is consistent with our previously published data of mRNA and miRNA expression profiling analyses of tumors from the same patients *(Sandhu et al., 2015)*. The morphology specific alterations for the intestinal subtype were deletion of 5q, gains of 13q and 3q22-26. In the duodenal tumors, all of which are of the intestinal subtype, a site of origin specific 4q deletion was observed. The driver genes (*PLK2, PIK3R1* and *PTPRM*) associated with morphological subtypes were also differentially expressed between the pancreatobiliary and intestinal subtypes in our previous study of the same cohort. Additionally, the genes *PIK3R1, NEK3, SASH1* and *RASA3* and multiple genes on chromosome 17p13.1 were also defined as potential

prognostic markers for the intestinal subtype *(Sandhu et al., 2015)*. The intestinal subtype also showed deletion of 5q, gain of chromosomal regions 3q21-26 and overexpression of *CCNL1* and *KLF5* and downregulation of *PLK2* and *PPP3CA*. Similar patterns were also identified in other tumor types *(Cancer Genome Atlas, N., 2012, Prat et al., 2013, Maire et al, 2013 and Zheng et al., 2009)*. Both mRNA and CNA profiling showed the importance of morphology specific sub typing of PAs. Further, subtyping of PAs, using the PAM50 classifier showed that subgroups of pancreatobiliary tumors are basal-like with the worst prognosis. The PAM50 gene signature classified tumors based on degree of differentiation as it has genes for proliferation, cell cycle, keratins, cell adhesion etc. The classification of PA samples based on PAM50 and "Moffitt's gene signature" are highly correlated showing the usefulness of PAM50 in subtyping. However, the copy number aberration patterns between the basal-like and classical subgroups were not significantly different in the two cohorts (data not shown).

The gene enrichment analyses based on copy number and gene expression data highlighted many common pathways. The high frequency of co-deletion of 17p and 18q is not by mere chance; but the genes from these genomic locations are associated with the cell cycle, apoptosis, p53 and Wnt signaling pathways in both cohorts. The similarities between the xenograft cell line CNA profile and its original tumor are comparable with gene expression results shown previously *(Wennerstrom et al., 2014)*.

Gains of 18p11 in PAs have been linked to poor DFS and OS and gain of 19q13 with poor DFS. The 19q13 chromosomal locus is commonly amplified in various cancers, including ovarian cancer *(Tang et al., 2002),* breast cancer *(Bellacosa et al., 1995),* pancreatic cancer *(Kuuselo et al., 2010)* and non-small cell lung cancer *(Kim et al., 2005)*. Kuuselo and colleagues have shown that gain of 19q13 is associated with higher grade, stage and outcome in pancreatic cancer *(Kuuselo et al., 2010)*. The overexpression of the genes *SERTAD3* and *ERCC1* located on 19q13.2 and 19q13.32 respectively were found to be associated with the worst overall survival in this study. *SERTAD3* is a putative oncogene that induces E2F activity and promotes tumor growth and has previously been putatively linked with DFS *(Darwish et al., 2007)*. The *ERCC1* gene is a

known prognostic biomarker in head and neck cancers *(Bauman et al., 2012)* and non-small cell lung cancer *(Tiseo et al., 2013)*.

Deletion of 18p11.22 is specific to the intestinal subtype, whereas gain of this locus occurs in 10 of the pancreatobiliary samples. 18p11.22 has also been suggested as a novel lung cancer susceptibility locus in never smokers *(Ahn et al., 2012)*. *RAB12* on chromosome 18p11.22 is a Ras oncogene family member and upregulation of this gene is associated with poor DFS. *RAB12* is also overexpressed in colorectal cancers *(Yoshida et al., 2010)*. Further, the *COLEC12* gene is known prognostic marker in anaplastic thyroid cancer *(Espinal-Enriquez et al., 2015)* and brain tumors *(Donson et al., 2009)*.

The present study has a relatively large sample size in both cohorts. The results provide new knowledge of the genomic changes characteristics of pancreatic cancer and may prove valuable to the clinical setting.

Material and Methods

DNA extraction

DNA was extracted from tumor tissue using the Maxwell Tissue DNA kit on the Maxwell 16 Instrument (Promega). Briefly 5 x 20 µm sections were homogenized in 300 µl Lysis Buffer and added to the cartridge. The method is based on purification using paramagnetic particles as a mobile solid phase for capturing, washing and elution of genomic DNA. Elution volume was 200 µl. DNA was extracted from 6 ml EDTA blood using the QiAamp DNA Blood BioRobot MDx Kit on the BioRobot MDx (Qiagen). This was done at Aros Applied Biotechnology AS, Aarhus, Denmark, and the Department of Medical Genetics, Oslo University Hospital. The method is based on lysis of the sample using protease, followed by binding of the genomic DNA to a silica-based membrane and washing and elution in 200 µl buffer AE. DNA from normal tissue was extracted at Aros Applied Biotechnology AS, Aarhus, Denmark according to their Standard Operation Procedures (SOP's) for extraction with a column based technology (Qiagen). Tissue specimens were homogenized in Qiagen Tissuelyzer homogenizer. The amount of tumor

cells in the sections used for DNA isolation were estimated on HE-stained sections cut before and after cutting of sections used for DNA isolation.

Tumor and matching normal samples

A total of 60 samples of fresh frozen tumor tissues with origin in the four different periampullary locations and corresponding normal DNA samples from EDTA blood; 28 from pancreas, 4 from bile duct, 6 from ampulla of pancreatobiliary type, 7 from ampulla of intestinal type, 9 from duodenum, 3 IPMN samples and 3 samples from xenografts cell lines generated from PDAC patients were analyzed using Affymetrix SNP 6.0 from OUH cohort. Further, for validation of the results, 127 PDAC samples from the TCGA cohort https://tcga-data.nci.nih.gov/tcga/ were analyzed.

Affymetrix SNP 6.0 arrays

The Affymetrix SNP 6.0 arrays include 1.8 million genetic markers, including 906,600 SNPs and 946,000 copy number probes. DNA digestion, labeling and hybridization were performed according to the manufacturer's instruction (Affymetrix, Santa Clara, CA, USA).

Data analysis

Copy number aberration profiles from the OUH (n = 60) and the TCGA (n = 127) cohorts were generated. Segmental copy number information was derived for each sample using the Battenberg pipeline (https://github.com/cancerit/cgpBattenberg/) as previously described *(Nik-Zainal et al., 2012*) to estimate tumor cell fraction, tumor ploidy and copy numbers. The Battenberg pipeline has an advantage over other tools, as it has greater sensitivity particularly for samples with low cellularity, frequently observed in pancreatic tumors. Briefly, the tool phases heterozygous SNPs with use of the 1000 genomes genotypes as a reference panel using Impute2 *(Howie et al., 2009)*, and corrects phasing errors in regions with copy number changes through segmentation *(Nilsen et al., 2012)*. After segmentation of the resulting b-allele frequency (BAF) values, t-tests are performed on the BAFs of each copy number segment to identify whether they correspond to the value resulting from a fully clonal copy number change. If not, the copy number segment

is represented as a mixture of two different copy number states, with the fraction of cells bearing each copy number state estimated from the average BAF of the heterozygous SNPs in that segment. Purity/ploidy fits for each sample were subjected to rigorous manual quality control, and where necessary, samples were refit to a new solution. The genome instability index (GII) was calculated for both the cohorts; it is measured as the fraction of aberrant probes throughout the genome above or below the ploidy. Correlation analysis was carried out to identify any association between GII and tumor ploidy.

Frequency plots

For each tumor, an aberration score was calculated per copy number segment. The aberration score is set to 1 if total copy number per segment is larger than the average ploidy of the tumor, i.e. corresponding to a copy number gain, and to -1 if it is smaller than the average ploidy of the tumor, corresponding to deletion. Remaining segments were scored to zero. The frequency plots were generated based on aberration score for all samples per segment. The whole genome allelic aberration frequency plots for the OUH cohort based on the four anatomical locations (pancreatic ducts, bile duct, ampulla, and duodenum), the two morphologies (pancreatobiliary and intestinal) and for validation in the TCGA cohort were plotted using ggplot2 library in R version 3.1.2. The radial plots were drawn for regions significantly different at P-value < 0.001 for Chi Square test in samples under comparison. Hierarchical clustering of the OUH and the TCGA cohort samples were done using Spearman's distance measure for cytobands and complete linkage method; where gain was given a score of 1, deletion as -1 and 0 otherwise.

mRNA expression analysis

The mRNA expression data for the OUH cohort has previously been published for the PA samples *(Sandhu et al, 2015, Wennerstrom et al., 2014)* with GEO accession numbers GSE60979 and GSE58561. The data was background corrected and quantile normalized. For the TCGA cohort, gene expression levels were assayed by RNA sequencing, RSEM (RNaseq by Expectation-Maximization) normalized per gene, as obtained from the TCGA consortium. The PAM50 gene signature was used for hierarchical clustering of PAs using Spearman's correlation as distance measure and complete linkage method. The

proliferation score was calculated as average gene expression of 11 proliferative genes namely; *CCNB1, UBE2C, BIRC5, CDC20, PTTG1, RRM2, MKI67, TYMS, CEP55, KNTC2* and *CDCA1 (Parker et al., 2009),* two way ANOVA test was done to estimate the significant difference in proliferation score between two groups.

Driver genes in amplified and deleted regions

The genes located in the amplified and deleted regions were mapped using ENSEMBL (GRCh37 genome assembly) genome annotation for SNP6 arrays.

The genes in amplified and deleted chromosomal location were identified based on the following criteria:

1. Frequency of occurrence: The genes that occur most frequently in the deleted/amplified chromosomal regions. The threshold was set to > 25%.
2. Genes that were mapped in the COSMIC cancer gene census (*Futreal et.al, 2003*) list for most frequently mutated genes in cancer, census of amplified and overexpressed genes in cancer (n = 77) *(Santarius et al., 2010)* and tumor suppressor gene list (n = 718) *(Zhao et.al, 2013).*

Correlation analysis of copy number aberrations and gene expression data

The Pearson correlation coefficient was calculated to estimate the correlation between total copy number and expression data for 52 PAs and three cell lines in OUH cohort, and 120 of 127 PDACs in TCGA cohort. Expression data for the three IPMNs, and two PAs in OUH cohort and seven PDACs in TCGA cohort were unavailable. The *P*-values are reported for the significant association between the allele frequency and the gene expression correlation test at $P < 0.05$.

Gene Set Enrichment Analysis

We performed a KEGG pathway based analysis using the Web based Gene set analysis tool-kit (WebGestalt) *(Zhang et al., 2003, Wang et al., 2013)* to identify biological pathways with enrichment of genes amplified or deleted in the OUH and the TCGA cohorts. WebGestalt uses a Hypergeometric test for enrichment evaluation analysis at P <

0.05 after Benjamin and Hochberg's correction and the minimum number of genes required for a pathway to be considered significant is set to 10. The WebGestalt analysis results gives pathway enriched genes, number of genes enriched, raw *P*-value (rawP) from the Hypergeometric test, Benjamin and Hochberg's corrected *P*-value (adjP), the number of reference gene in the category (C score), number of genes in the gene set and also in the category (O score), expected number in the category (E score) and ratio of enrichment (R score).

Survival analysis

Survival analysis was performed using the Kaplan–Meier estimator as implemented in the KMsurv package *(Therneau and Grambsch, 2000)* and the log-rank test in R version 3.1.2. Overall survival (OS) time was calculated from date of surgery to time of death. OS data were obtained from the National Population Registry in Norway. Three patients with distant metastases (M1) at time of resection, and one patient that died from cardiac arrest were not included in the survival analysis. Recurrence free survival (RFS) time was calculated from date of surgery to date of recurrence of disease. Recurrence was defined as radiological evidence of intra-abdominal soft tissue around the surgical site or of distant metastasis.

The Kaplan-Meier survival curve is plotted for focal amplified regions of PA samples and for genes located on focal amplicon regions. The expression for each sample was designated as high if the expression was higher than median expression otherwise low. The *P*-value from log rank test is reported for significant findings.

References

1. Ahn, M.J., Won, H.H., Lee, J., Lee, S.T., Sun, J.M., Park, Y.H., Ahn, J.S., Kwon, O.J., Kim, H., Shim, Y.M., et al., 2012. The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations. Human genetics 131, 365-372.

2. Bauman, J.E., Austin, M.C., Schmidt, R., Kurland, B.F., Vaezi, A., Hayes, D.N., Mendez, E., Parvathaneni, U., Chai, X., Sampath, S. et al., 2013. ERCC1 is a prognostic biomarker in locally advanced head and neck cancer: results from a randomised, phase II trial. British journal of cancer 109, 2096-2105.

3. Bellacosa A, de Feo D, Godwin AK, Bell DW, Cheng JQ, Altomare DA, Wan M, Dubeau L, Scambia G, Masciullo V. Molecular alterations of the AKT2 oncogene in ovarian and breast carcinomas. Int J Cancer. 1995; 64:280–285

4. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M. et al., 2010. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899-905.

5. Cancer Genome Atlas, N., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70.

6. Carter, P., Presta, L., Gorman, C.M., Ridgway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E., Shepard, H.M., 1992. Humanization of an anti-p185HER2 antibody for human cancer therapy. Proceedings of the National Academy of Sciences of the United States of America 89, 4285-4289.

7. Childs, E.J., Mocci, E., Campa, D., Bracci, P.M., Gallinger, S., Goggins, M., Li, D., Neale, R.E., Olson, S.H., Scelo, G et al., 2015. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nature genetics 47, 911-916.

8. Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., Sander, C., 2013. Emerging landscape of oncogenic signatures across human cancers. Nature genetics 45, 1127-1133.

9. Coussens, L., Yang-Feng, T.L., Liao, Y.C., Chen, E., Gray, A., McGrath, J., Seeburg, P.H., Libermann, T.A., Schlessinger, J., Francke, U., et al., 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. Science 230, 1132-1139.

10. Darwish, H., Cho, J.M., Loignon, M., Alaoui-Jamali, M.A., 2007. Overexpression of SERTAD3, a putative oncogene located within the 19q13 amplicon, induces E2F activity and promotes tumor growth. Oncogene 26, 4319-4328.

11. Iacobuzio-Donahue, C.A., Velculescu, V.E., Wolfgang, C.L., Hruban, R.H., 2012. Genetic basis of pancreas cancer development and progression: insights from whole-exome and whole-genome sequencing. Clinical cancer research: an official journal of the American Association for Cancer Research 18, 4257-4265.

12. Donson, A.M., Birks, D.K., Barton, V.N., Wei, Q., Kleinschmidt-Demasters, B.K., Handler, M.H., Waziri, A.E., Wang, M., Foreman, N.K., 2009. Immune gene and cell enrichment is associated with a good prognosis in ependymoma. Journal of immunology 183, 7428-7440.

13. Espinal-Enriquez, J., Munoz-Montero, S., Imaz-Rosshandler, I., Huerta-Verde, A., Mejia, C., Hernandez-Lemus, E., 2015. Genome-wide expression analysis suggests a crucial role of dysregulation of matrix metalloproteinases pathway in undifferentiated thyroid carcinoma. BMC genomics 16, 207.

14. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. et al., 2014. Ensembl 2014. Nucleic acids research 42, D749-755.

15. Gutierrez, M.L., Munoz-Bellvis, L., Abad Mdel, M., Bengoechea, O., Gonzalez-Gonzalez, M., Orfao, A., Sayagues, J.M., 2011. Association between genetic subgroups of pancreatic ductal adenocarcinoma defined by high density 500 K SNP-arrays and tumor histopathology. PloS one 6, e22315.

16. Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. Cell 100, 57-70.

17. Harada, T., Chelala, C., Bhakta, V., Chaplin, T., Caulee, K., Baril, P., Young, B.D., Lemoine, N.R., 2008. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. Oncogene 27, 1951-1960.

18. Harada, T., Chelala, C., Crnogorac-Jurcevic, T., Lemoine, N.R., 2009. Genome-wide analysis of pancreatic cancer using microarray-based techniques. Pancreatology : official journal of the International Association of Pancreatology 9, 13-24.

19. Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 5, e1000529.

20. Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A et al., 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801-1806.

21. Kao J, Pollack JR. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer. 2006; 45:761-9.

22. Kim TM, Yim SH, Lee JS, Kwon MS, Ryu JW, Kang HM, Fiegler H, Carter NP, Chung YJ. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. Clin Cancer Res. 2005;11:8235–824.

23. Kuuselo, R., Simon, R., Karhu, R., Tennstedt, P., Marx, A.H., Izbicki, J.R., Yekebas, E., Sauter, G., Kallioniemi, A., 2010. 19q13 amplification is associated with high grade and stage in pancreatic cancer. Genes, chromosomes & cancer 49, 569-575.

24. Maire, V., Nemati, F., Richardson, M., Vincent-Salomon, A., Tesson, B., Rigaill, G., Gravier, E., Marty-Prouvost, B., De Koning, L., Lang, G. et al., 2013. Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. Cancer research 73, 813-823.

25. Min Zhao, Jingchun Sun, Zhongming Zhao (2013) TSGene: a web resource for tumor suppressor genes. Nucleic Acids Research, 41: D970-D976.

26. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., Smyla, J.K., Anderson, J.M., Kim, H.J., Bentrem, D.J., Talamonti, M.S., Iacobuzio-Donahue, C.A., Hollingsworth, M.A., Yeh, J.J., 2015. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nature genetics 47, 1168-1178.

27. Natrajan, R., Mackay, A., Wilkerson, P.M., Lambros, M.B., Wetterskog, D., Arnedos, M., Shiu, K.K., Geyer, F.C., Langerod, A., Kreike, B. et al., 2012. Functional characterization of the 19q12 amplicon in grade III breast cancers. Breast cancer research : BCR 14, R53.

28. Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M. et al., 2012. The life history of 21 breast cancers. Cell 149, 994-1007.

29. Nilsen, G., Liestol, K., Van Loo, P., Moen Vollan, H.K., Eide, M.B., Rueda, O.M., Chin, S.F., Russell, R., Baumbusch, L.O., Caldas, C. et al., 2012. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC genomics 13, 591.

30. P. Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman & Michael R. Stratton (2004). A census of human cancer genes2Nature Reviews Cancer 4, 177-183 | doi:10.1038/nrc1299.

31. Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z. et al., 2009. Supervised risk predictor of breast cancer based

on intrinsic subtypes. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 27, 1160-1167.

32. Prat, A., Adamo, B., Fan, C., Peg, V., Vidal, M., Galvan, P., Vivancos, A., Nuciforo, P., Palmer, H.G., Dawood, et al., 2013. Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. Scientific reports 3, 3544.

33. Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., Matrisian, L.M., 2014. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research 74, 2913-2921.

34. Sandhu, V., Bowitz Lothe, I.M., Labori, K.J., Lingjaerde, O.C., Buanes, T., Dalsgaard, A.M., Skrede, M.L., Hamfjord, J., Haaland, T., Eide, T.J. et al., 2015. Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas. Molecular oncology 9, 758-771.

35. Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., Cooper, C.S., 2010. A census of amplified and overexpressed human cancer genes. Nature reviews. Cancer 10, 59-64.

36. Tang TC, Sham JS, Xie D, Fang Y, Huo KK, Wu QL, Guan XY. Identification of a candidate oncogene SEI-1 within a minimal amplified region at 19q13.1 in ovarian cancer cell lines. Cancer Res. 2002;62:7157–7161.

37. Tiseo, M., Bordi, P., Bortesi, B., Boni, L., Boni, C., Baldini, E., Grossi, F., Recchia, F., Zanelli, F., Fontanini, G. et al., 2013. ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. British journal of cancer 108, 1695-1703.

38. Waddell, N., Pajic, M., Patch, A.M., Chang, D.K., Kassahn, K.S., Bailey, P., Johns, A.L., Miller, D., Nones, K., Quek, K.,et al., 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518, 495-501.

39. Wang, J., Duncan, D., Shi, Z., Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res, 41 (Web Server issue), W77-83.

40. Wennerstrom, A.B., Lothe, I.M., Sandhu, V., Kure, E.H., Myklebost, O., Munthe, E., 2014. Generation and characterisation of novel pancreatic adenocarcinoma xenograft models and corresponding primary cell lines. PloS one 9, e103873.

41. Yoshida, T., Kobayashi, T., Itoda, M., Muto, T., Miyaguchi, K., Mogushi, K., Shoji, S., Shimokawa, K., Iida, S., Uetake, H. et al., 2010. Clinical omics analysis of colorectal cancer incorporating copy number aberrations and gene expression data. Cancer informatics 9, 147-161.

42. Zhang, B., Kirov, S.A., Snoddy, J.R. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res, 33(Web Server issue), W741-748.

43. Zheng, H.Q., Zhou, Z., Huang, J., Chaudhury, L., Dong, J.T., Chen, C., 2009. Kruppel-like factor 5 promotes breast cell proliferation partially through upregulating the transcription of fibroblast growth factor binding protein 1. Oncogene 28, 3702-3713.

44. Zhou, Y., Han, C., Li, D., Yu, Z., Li, F., Li, F., An, Q., Bai, H., Zhang, X., Duan, Z., et al., 2015. Cyclin-dependent kinase 11(p110) (CDK11(p110)) is crucial for human breast cancer cell proliferation and growth. Scientific reports 5, 10433.

Figures and tables legends:

Figure 1: The frequency of copy number aberrations in the OUH and the TCGA cohort. The x-axis represents the genomic position while the y-axis represents the frequencies of amplification and deletion. The red and green bars in the figure show amplifications and deletions of the chromosomes. Note the overrepresentations in deletions in both cohorts.

Figure 2: Frequency of copy number alterations in periampullary adenocarcinomas.
2A) The frequency plot of the genomic aberration pattern for the pancreatobiliary and intestinal subtypes. The x-axis represents the genomic position and is divided into 22 facets for the 22 chromosomes. The y-axis represents the frequency of chromosomal gains and losses for the two subtypes based on morphology.
2B) The radial plot shows the relative percentage of chromosomal copy number gains and losses based on morphology. The cytobands are marked in red and green for gains and losses, respectively. The color bars refer to the morphology of the samples.
2C) The radial plot shows the percentage of copy number gains and losses based on site of origin. The cytobands are marked in red and green for gains and losses, respectively. The color bars refer to the site of origin of the sample.

Figure 3: The figure shows the average frequencies of amplifications and deletions of genes in both the OUH and TCGA cohorts. The x-axis shows the frequencies of amplifications and deletions of genes, and the color bars on the y-axis represents the chromosomes. The cytobands and genes are marked in red and green for amplifications and deletions, respectively.

Figure 4: Kaplan Meier OS and DFS for patients carrying PA tumors in the OUH cohort 4 A) Kaplan Meier DFS curves according to 18p11.22 chromosomal locus status. 4 B-C) Kaplan Meier DFS curves based on *RAB12* and *COLEC12* gene expression status. The genes are located in loci 18p11.22 and 18p11.32, respectively. 4 D) Kaplan Meier OS curves according to 19q13.2 chromosomal status. 4 E-F) Kaplan Meier OS curves based on *SERTAD3* and *ERCC1* gene expression status. The genes are located on 19q13.2 and 19q13.32, respectively. In Figure 4 D, tumors carrying deletions and tumors without

19q13 alterations (normal samples) were combined in the survival analysis due to limited number of tumors with deletion events (n = 4).

Table 1: Clinical features of the PAs from the OUH cohort (n = 55) and from the TCGA cohort (n = 127). Note: For comparative reasons, the overall and disease free survival in patients from the OUH cohort were only calculated for PDAC since the TCGA cohort is primarily composed of PDAC tumors.

Table 2: Pathway analysis using frequently amplified and deleted genes in the OUH and the TCGA cohorts. The table shows the enriched pathways, number of genes enriched in the pathway, *P*-value and FDR corrected *P*-value using the Benjamini and Hochberg correction.

Figure 1

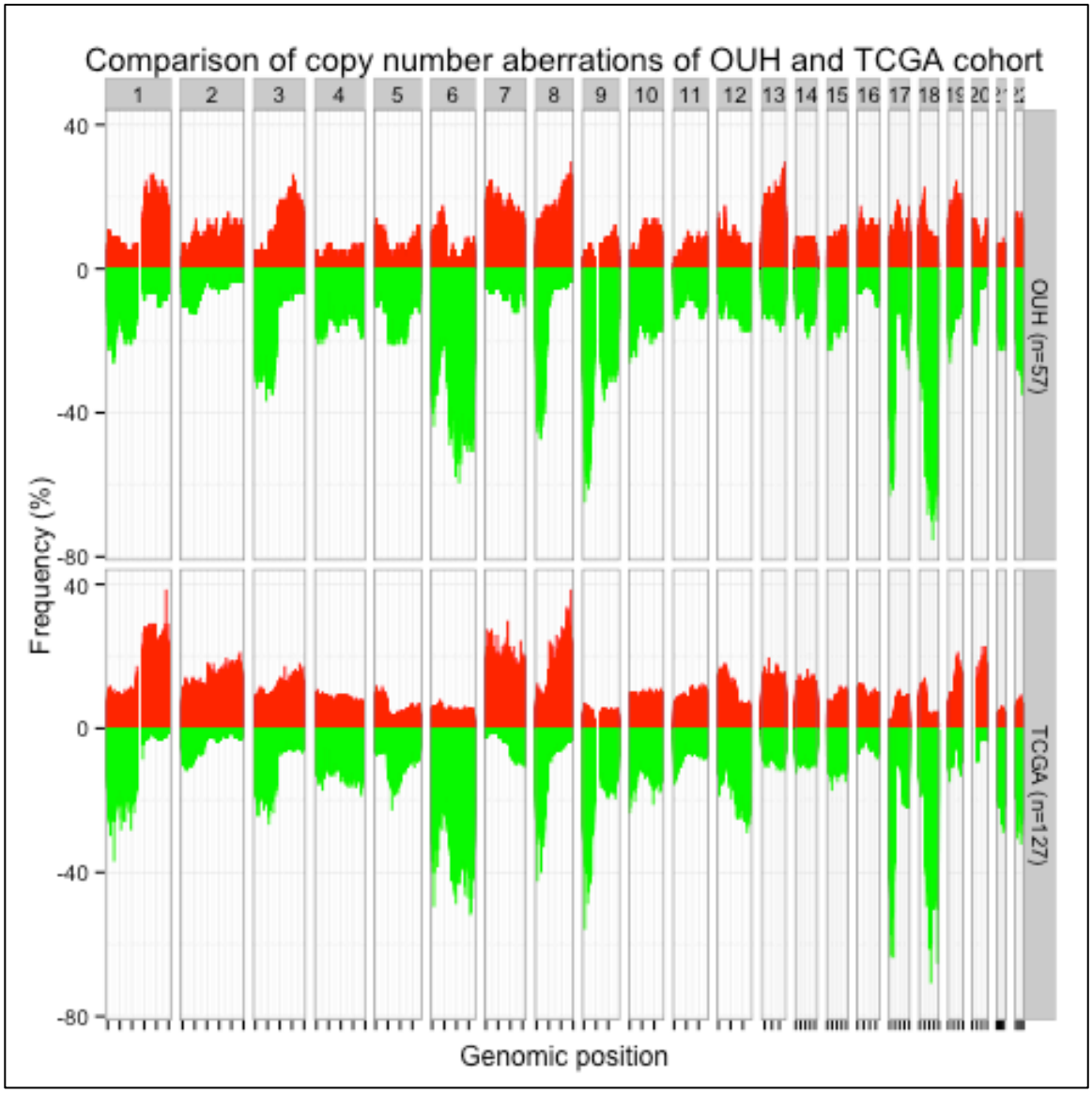
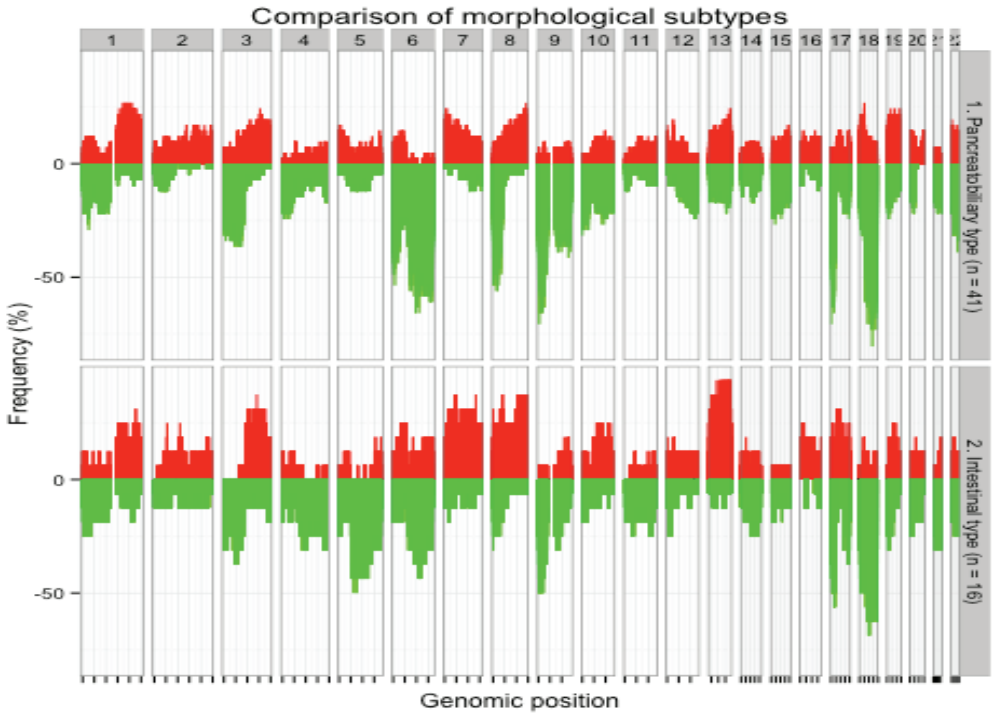
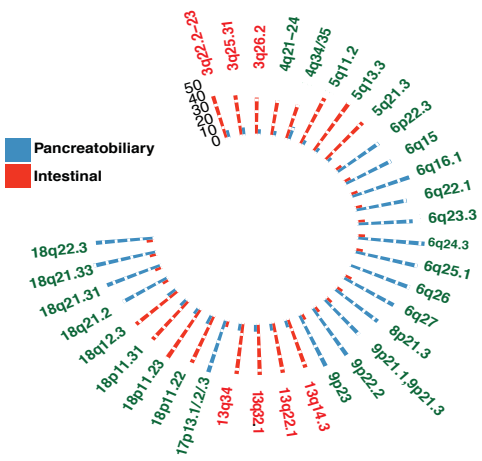

Comparison of copy number aberrations of OUH and TCGA cohort
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
40
0
-40
-80
Frequency (%)
OUH (n=57)
TCGA (n=127)
Genomic position

Figure 2

2A.



2B. Morphology based



2C. Site of origin based

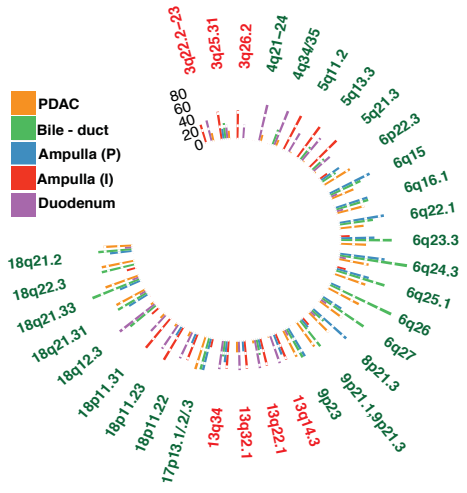


Figure 3

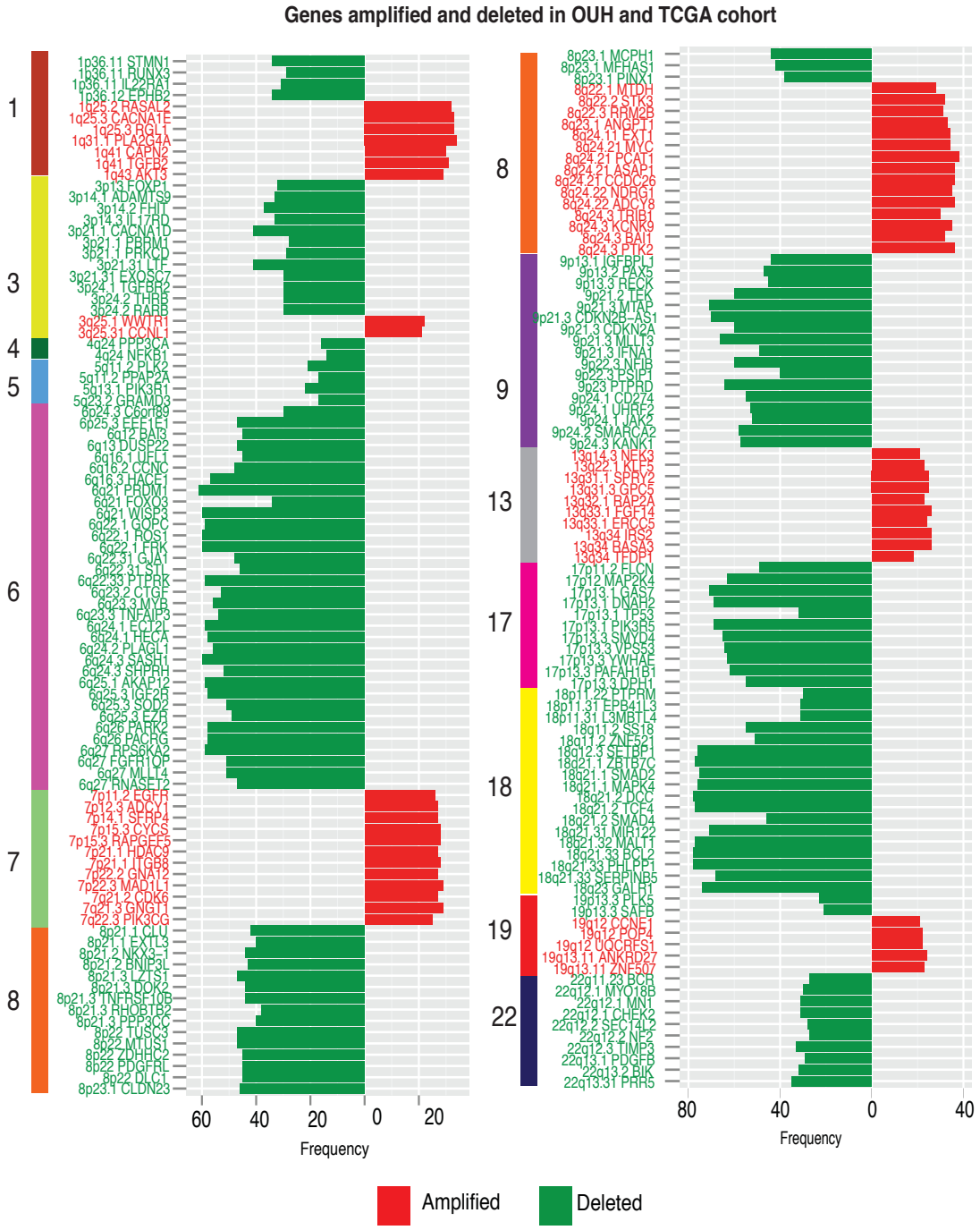

Genes amplified and deleted in OUH and TCGA cohort
Frequency
Frequency
Amplified
Deleted

Figure 4

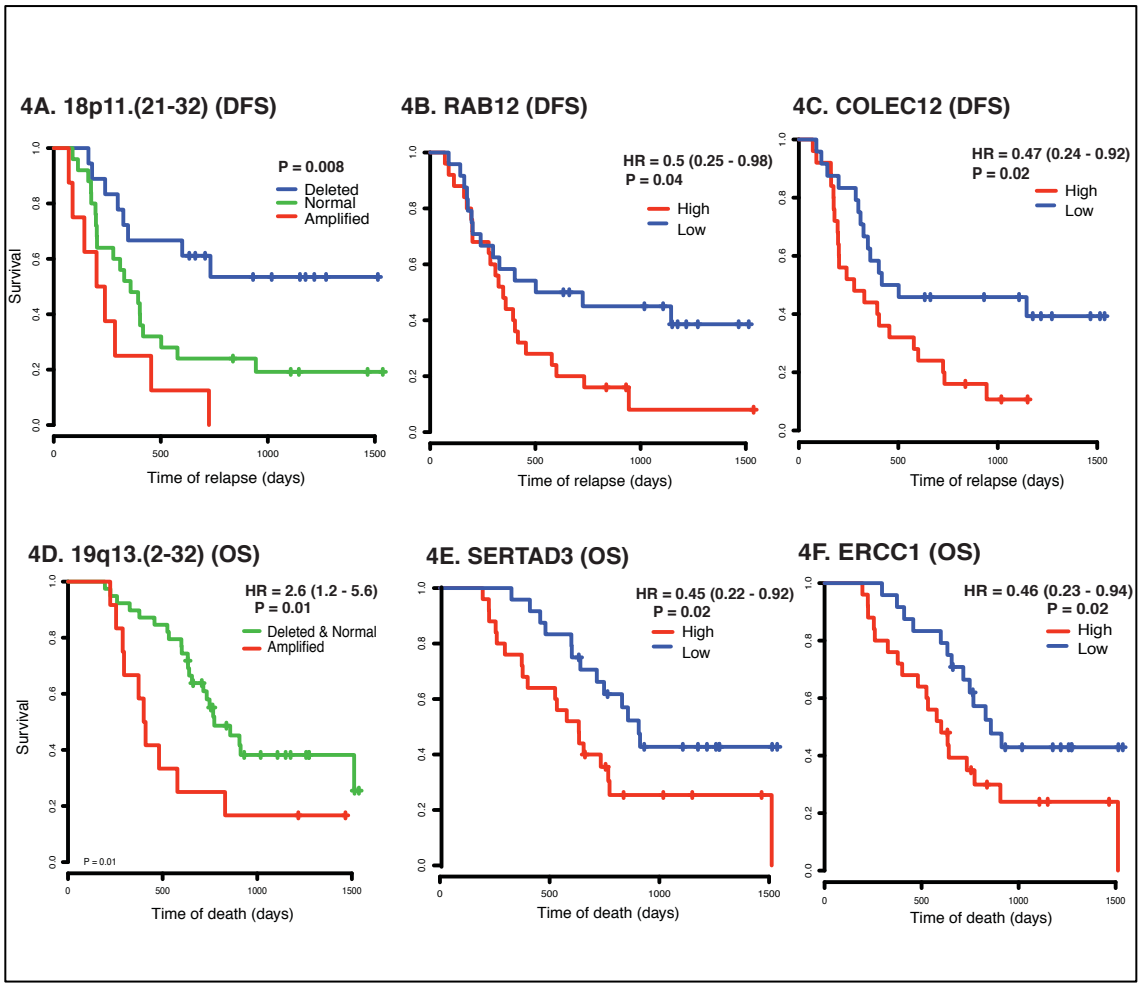

4A. 18p11.(21-32) (DFS)
P = 0.008
Deleted
Normal
Amplified
Survival
Time of relapse (days)
4B. RAB12 (DFS)
HR = 0.5 (0.25 - 0.98)
P = 0.04
High
Low
Time of relapse (days)
4C. COLEC12 (DFS)
HR = 0.47 (0.24 - 0.92)
P = 0.02
High
Low
Time of relapse (days)
4D. 19q13.(2-32) (OS)
HR = 2.6 (1.2 - 5.6)
P = 0.01
Deleted & Normal
Amplified
P = 0.01
Survival
Time of death (days)
4E. SERTAD3 (OS)
HR = 0.45 (0.22 - 0.92)
P = 0.02
High
Low
Time of death (days)
4F. ERCC1 (OS)
HR = 0.46 (0.23 - 0.94)
P = 0.02
High
Low
Time of death (days)

Table 1

Clinical Features		The OUH cohort (n=55) Frequency (percentage)	The TCGA cohort (n=127) Frequency (percentage)
Gender	Female	30 (55)	52 (41)
	Male	25 (45)	75 (59)
Type	PDAC	29 (52)	111(87)
	PDAC-other subtypes	-	16(13)
	Bile duct	4 (7)	-
	Ampulla pancreatobiliary type	6 (10)	-
	Ampulla intestinal type	7 (12)	-
	Duodenum	9 (16)	-
pT	T1	4 (7)	2(1)
	T2	9 (16)	11(9)
	T3	36 (65)	110(87)
	T4	6 (11)	3(2)
	TX	-	1(1)
N	N0	19 (35)	34(27)
	N1	35 (64)	91(72)
	N2	1 (2)	0(0)
	NX	-	2(1)
M	M0	52 (95)	59(47)
	M1	3 (5)	4(3)
	MX	-	64(50)
R	R0	37 (67)	68(54)
	R1	18 (33)	42(33)
	R2	-	4(3)
	RX	-	4(3)
	Not available	-	9(7)
Differentiation/ Grade	Well (G1)	18 (33)	14(11)
	Moderately (G2)	37 (67)	72(57)
	Poor (G3)	-	39(31)
	Undetermined (GX)	-	2(1)
Mean overall survival		578 days	247 days
Median disease free survival		259 days	-
Median age		65 years	66 years

Table 2

	OUH cohort			TCGA cohort		
Pathways	Number of genes	*P*-value	Adjusted *P*-value	Number of genes	*P*-value	Adjusted *P*-value
MAPK signaling	44	6.83E-07	5.18E-06	65	2.2E-13	3.1E-12
Jak-STAT signaling	29	3.86E-06	1.95E-05	44	6.0E-12	5.2E-11
Cell cycle	22	0.0001	0.0003	27	1.7E-05	3.8E-05
p53 signaling	13	0.0014	0.0026	20	1.8E-06	5.9E-06
Apoptosis	14	0.005	0.0081	25	1.5E-07	6.9E-07
Insulin signaling	19	0.0071	0.0104	22	0.007	0.008
TGF-beta signaling	12	0.0219	0.0269	16	0.003	0.004
Wnt signaling	18	0.0312	0.0374	34	5.8E-07	2.2E-06

Supplementary figure and tables legends

Supplementary Figure S1: The figure shows the heatmap of the OUH cohort in figure 1a and for the TCGA cohort in figure 1b. The rows represent the cytobands while the columns represent the samples. The color bars on y-axis represent the chromosomes. The red color in the heatmap represents the amplification, green color represent the deletion while black represent the normal state.

Supplementary Figure S2: The figure shows the CNA profile of the two xenografts cell lines and their original tumors. The minor allele frequency is in blue and purple is the total copy number.

Supplementary Figure S3: The figure shows CNA profile of three IPMN samples. The minor allele frequency is in blue and purple is the total copy number.

Supplementary Figure S4: The figure shows the scatter plot for the OUH cohort in figure 4A and for the TCGA cohort in figure 4B. The x-axis represent the genome instability index while the y-axis represent the ploidy of the tumors.

Supplementary Figure S5: The figure shows the frequency plot of the genomic aberrations of periampullary adenocarcinomas based on site of origin. The x-axis represents the genomic position and is divided into 22 facets for 22 chromosomes. The y-axis represents the frequency of chromosomal gain and losses for five subtypes based on site of origin. The number of samples in each subtypes are PDAC: 28, bile duct: 4, ampulla P: 6, ampulla I: 7, duodenum: 9.

Supplementary Figure S6: The figure shows the subtyping of PA samples. 6A) The heatmap shows the clustering of PA samples using PAM50 gene signature. The PA samples were clustered using Spearman distance measure and complete linkage method. 6B) The heatmap shows the clustering of PDAC samples by Moffitt et al gene signature using Spearman distance measure and complete linkage method. 6C) The Kaplan Meier

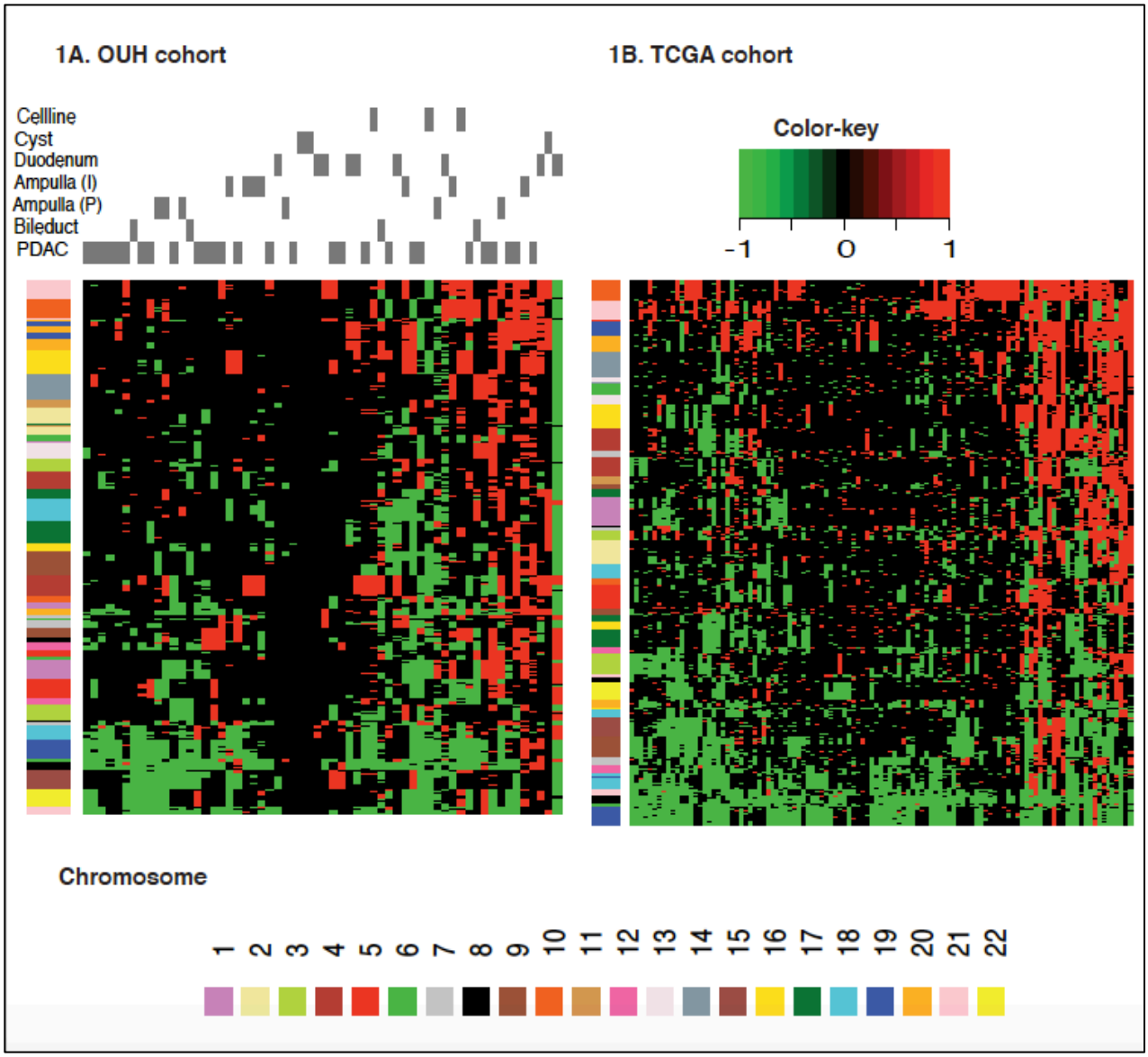
survival curve shows the OS for three subclasses found using PAM50 classifier. 6D) The boxplot shows the proliferation score for basal and classical type of PDACs.

Supplementary Table S1: Shows the genes, chromosomal location and frequency of deletion and amplification of genes in OUH (PAs (n=54), xenograft cell lines (n=3) and IPMN (n=3) and TCGA cohort (n=127).

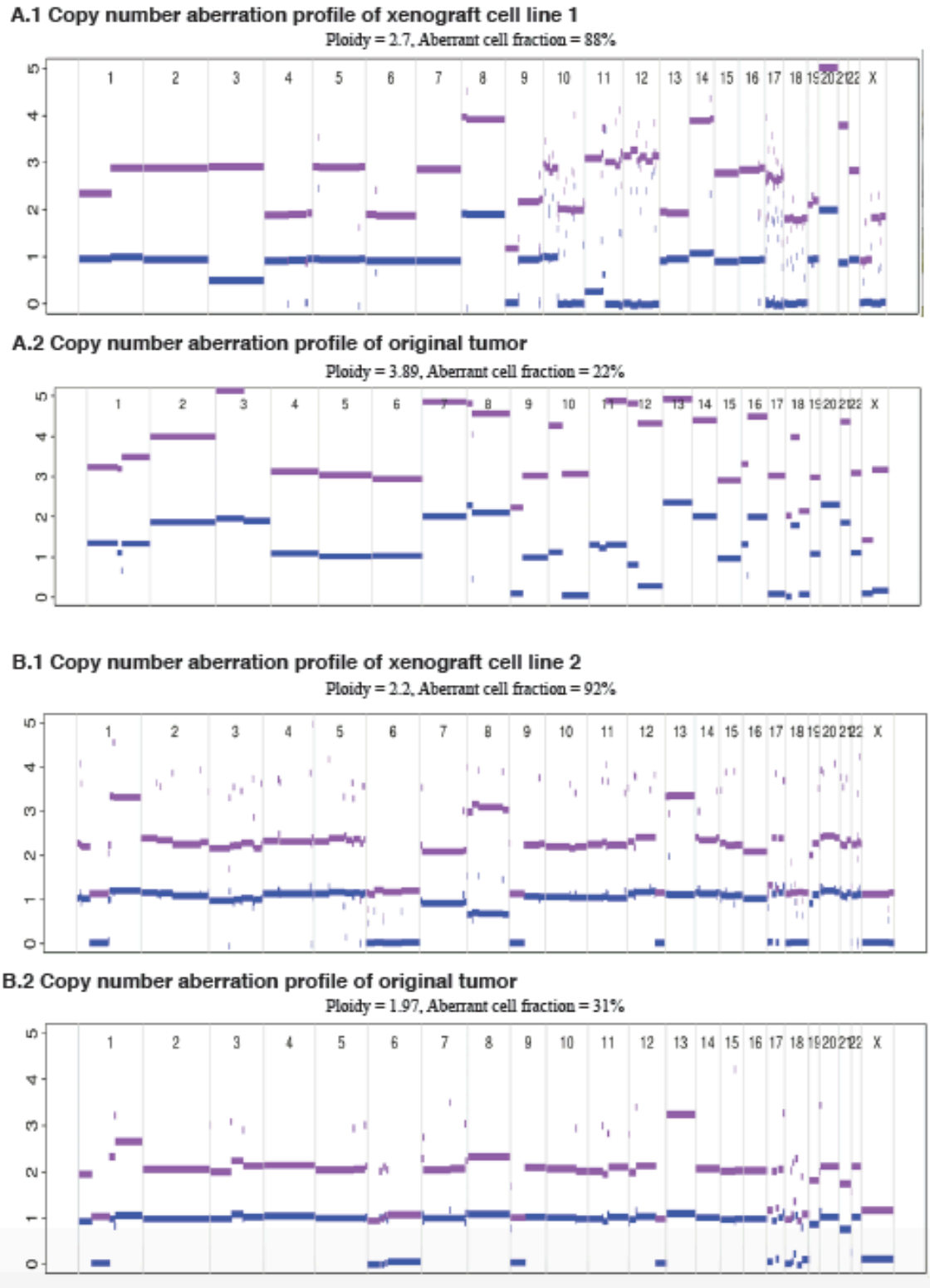
Supplementary Table S2: 2a) Table showing the corelation score (r) between allele frequency and gene expression for amplified region, p values from correlation test, frequency of amplification in PA samples in OUH cohort. 2b) Table showing the correlation score (r) between allele frequency and gene expression for deleted region, p values from correlation test, frequency of deletion in PA samples in OUH cohort. 2c) Table showing the correlation score (r) between allele frequency and gene expression for amplified region, p values from correlation test, frequency of deletion in TCGA cohort.
2d) Table showing the correlation score (r) between allele frequency and gene expression for deleted region, p values from correlation test, frequency of amplification in TCGA cohort.

Supplementary Table S3: The table shows the pathways deregulated in the OUH and TCGA cohort, #Gene = No. of genes found dysregulated in the pathway and statistical values, Statistics column list the C score= number of genes in the pathway, O score= No. of enriched genes in the pathway, E score = expected number in the pathway, R= ratio of enrichment score, rawP = raw P score from hypergeometric test and adjP= Adjusted p score from multiple testing correction.

Supplementary Figure S1



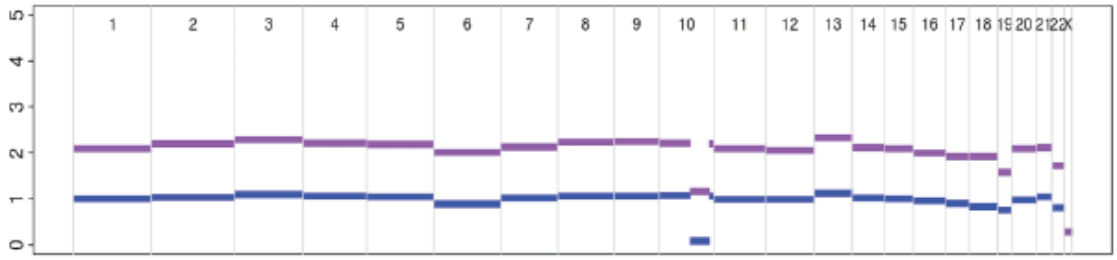
Supplementary Figure S2



Supplementary Figure S3

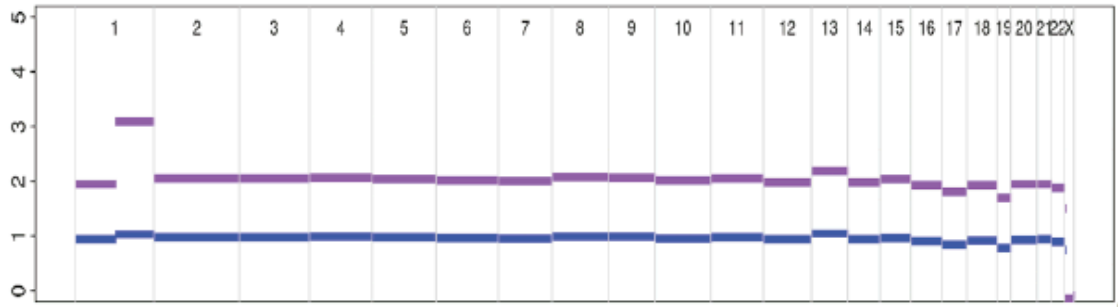
A. IPMN patient 1

Ploidy = 2.10, Aberrant cell fraction = 22%



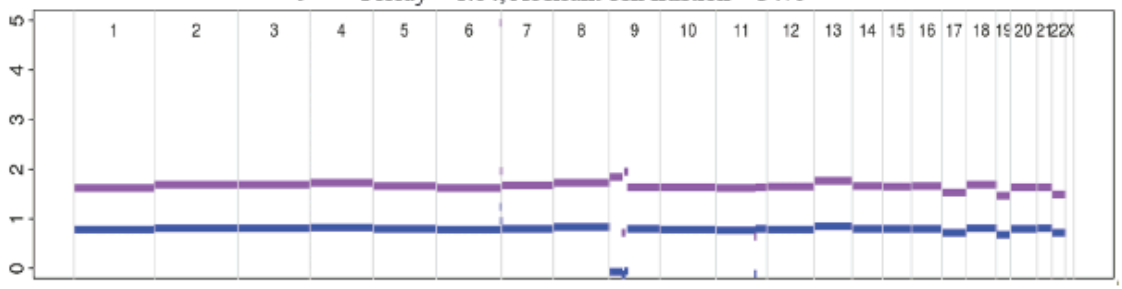
B. IPMN patient 2

Ploidy = 2.04, Aberrant cell fraction = 28%

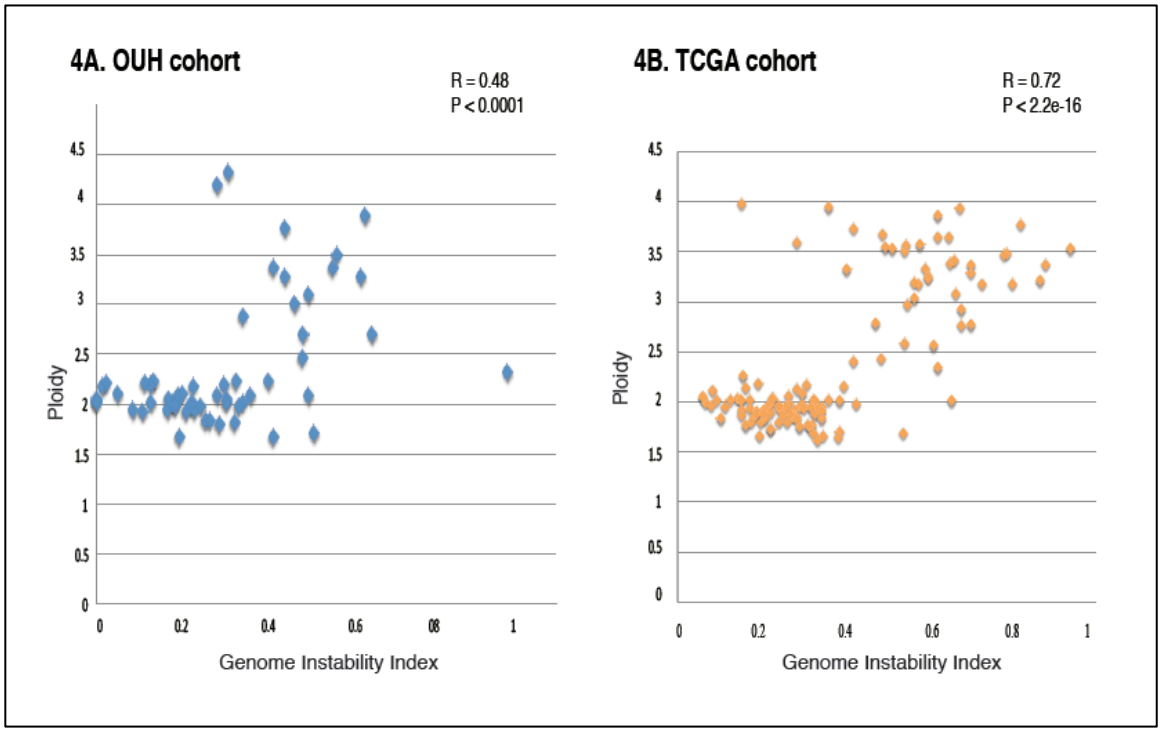


C. IPMN patient 3

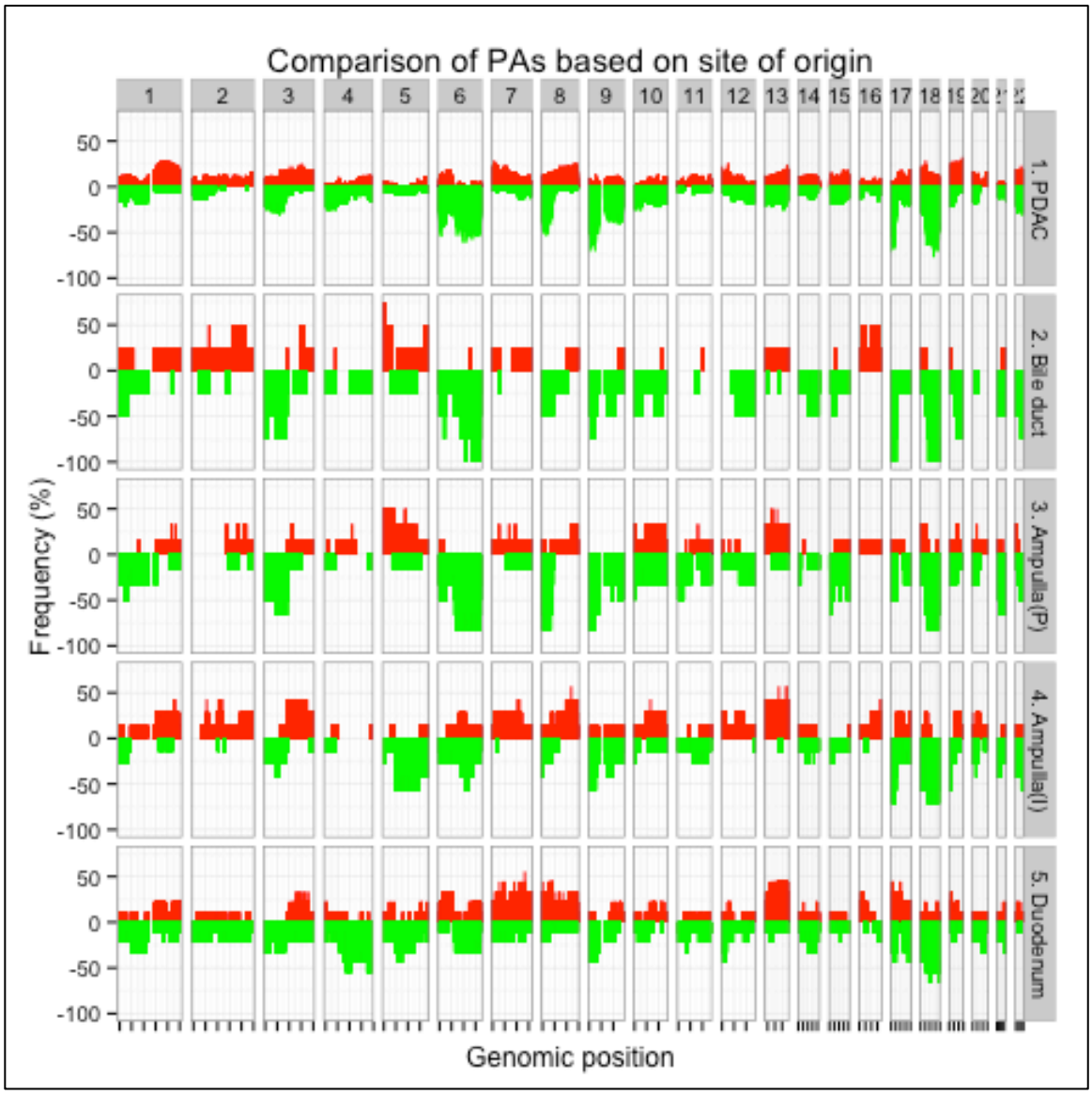
Ploidy = 1.64, Aberrant cell fraction = 34%



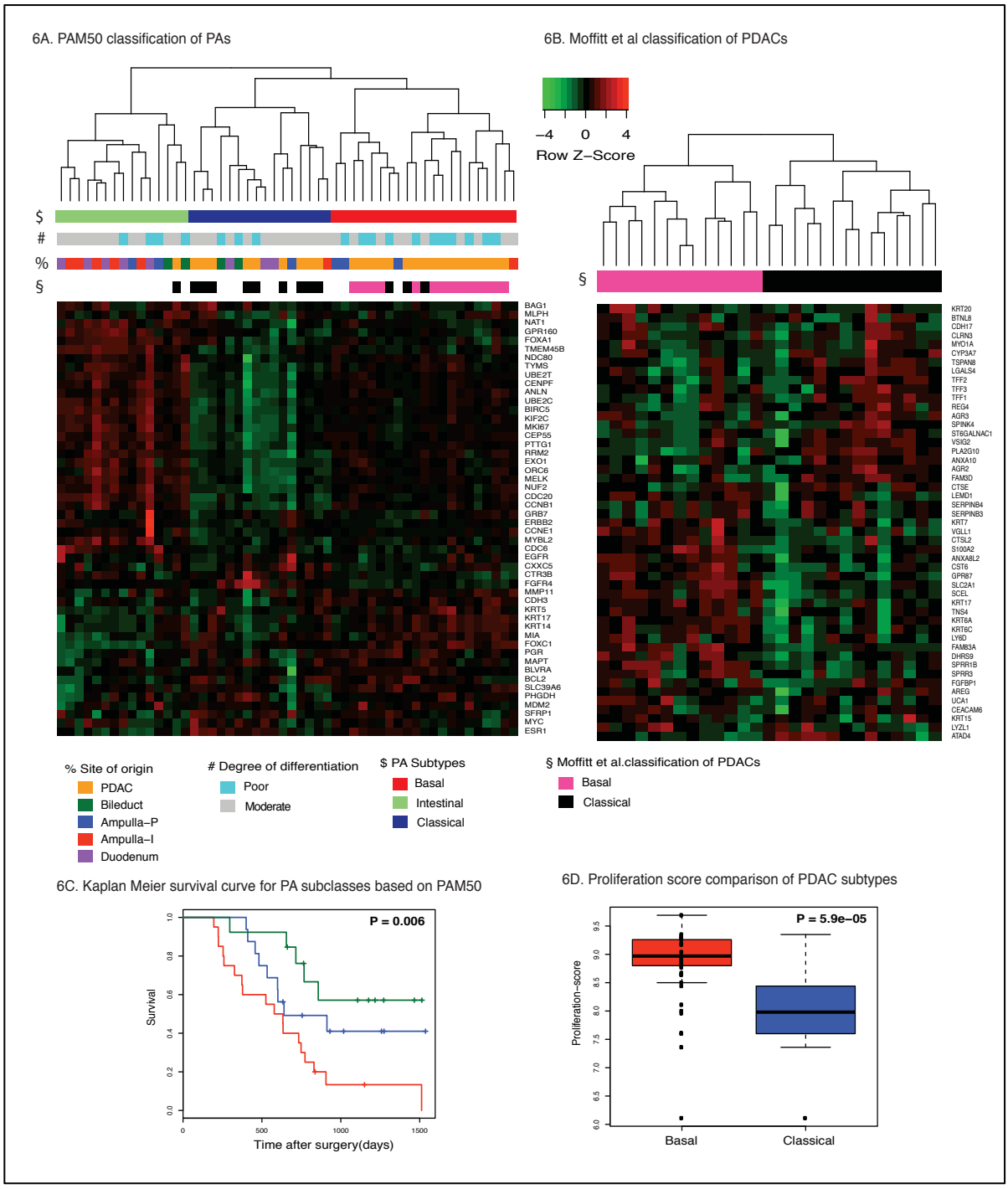
Supplementary Figure S4



Supplementary Figure S5



Supplementary Figure S6



Supplementary Table S1

1a: Table shows the frequency of genes being deleted.

Chromosome	Cytoband	OUH	TCGA
1	1p36.11	STMN1 : 28	STMN1 : 40
	1p36.11	RUNX3 : 26	RUNX3 : 31
	1p36.11	IL22RA1 : 26	ILL22RA1 : 36
	1p36.12	EPHB2 : 28	EPHB2 : 39
3	3p13	FOXP1 : 37	FOXP1 : 27
	3p14.1	ADAMTS9 : 35	ADAMTS9 : 31
	3p14.2	FHIT : 39	FHIT : 34
	3p14.3	IL17RD : 37	IL17RD : 28
	3p21.1	CACNA1D : 41	CACNA1D : 41
	3p21.1	PBRM1 : 33	PBRM1 : 23
	3p21.1	PRKCD : 32	PRKCD : 25
	3p21.31	LTF : 41	LTF : 40
	3p21.31	EXOSC7 : 34	EXOSC7 : 25
	3p24.1	TGFBR2 : 32	TGFBR2 : 28
	3p24.2	THRB : 32	THRB : 28
	3p24.2	RARB : 32	RARB : 28
4	4q24	PPP3CA : 15	PPP3CA : 17
	4q24	NFKB1 : 15	NFKB1 : 13
5	5q11.2	PLK2 : 20	PLK2 : 21
	5q11.2	PPAP2A : 20	PPAP2A : 13
	5q13.1	PIK3R1 : 20	PIK3R1 : 23
	5q23.2	GRAMD3 : 20	GRAMD3 : 13
6	6q12	BAI3 : 40	BAI3 : 50
	6q13	DUSP22 : 43	DUSP22 : 50
	6q16.1	UFL1 : 43	UFL1 : 46
	6q16.2	CCNC : 49	CCNC : 46
	6q16.3	HACE1 : 57	HACE1 : 56
	6q21	PRDM1 : 61	PRDM1 : 61
	6q21	FOXO3 : 61	FOXO3 : 6
	6q21	WISP3 : 59	WISP3 : 61
	6q22.1	GOPC : 59	GOPC : 59
	6q22.1	ROS1 : 59	ROS1 : 61
	6q22.1	FRK : 59	FRK : 61
	6q22.31	GJA1 : 46	GJA1 : 49
	6q22.31	STL : 42	STL : 50
	6q22.33	PTPRK : 56	PTPRK : 61
	6q23.2	CTGF : 51	CTGF : 55
	6q23.3	MYB : 56	MYB : 55
	6q23.3	TNFAIP3 : 54	TNFAIP3 : 54
	6q24.1	ECT2L : 57	ECT2L : 60
	6q24.1	HECA : 56	HECA : 59
	6q24.2	PLAGL1 : 54	PLAGL1 : 58
	6q24.3	SASH1 : 56	SASH1 : 63
	6q24.3	SHPRH : 52	SHPRH : 52
	6p24.3	C6orf89 : 28	C6orf89 : 31
	6p25.3	EEF1E1 : 43	EEF1E1 : 51
	6q25.1	AKAP12 : 56	AKAP12 : 62
	6q25.3	IGF2R : 54	IGF2R : 61

	6q25.3 6q25.3 6q26 6q26 6q27 6q27 6q27 6q27	SOD2 : 49 EZR : 45 PARK2 : 54 PACRG : 54 RPS6KA2 : 56 FGFR1OP : 51 MLLT4 : 51 RNASET2 : 45	SOD2 : 52 EZR : 52 PARK2 : 62 PACRG : 62 RPS6KA2 : 62 FGFR1OP : 51 MLLT4 : 51 RNASET2 : 48
8	8p21.1 8p21.1 8p21.2 8p21.2 8p21.3 8p21.3 8p21.3 8p21.3 8p21.3 8p22 8p22 8p22 8p22 8p22 8p23.1 8p23.1 8p23.1 8p23.1	CLU : 44 EXTL3 : 42 NKX3-1 : 47 BNIP3L : 46 LZTS1 : 49 DOK2 : 49 TNFRSF10B : 46 RHOBTB2 : 44 PPP3CC : 42 TUSC3 : 49 MTUS1 : 49 ZDHHC2 : 47 PDGFRL : 47 DLC1 : 47 CLDN23 : 47 MCPH1 : 44 MFHAS1 : 42 PINX1 : 40	CLU : 40 EXTL3 : 38 NKX3-1 : 41 BNIP3L : 40 LZTS1 : 44 DOK2 : 39 TNFRSF10B : 41 RHOBTB2 : 32 PPP3CC : 38 TUSC3 : 44 MTUS1 : 44 ZDHHC2 : 42 PDGFRL : 43 DLC1 : 43 CLDN23 : 44 MCPH1 : 43 MFHAS1 : 42 PINX1 : 35
9	9p13.1 9p13.2 9p13.3 9p21.2 9p21.3 9p21.3 9p21.3 9p21.3 9p21.3 9p22.3 9p22.3 9p23 9p24.1 9p24.1 9p24.1 9p24.2 9p24.3	IGFBPL1 : 47 PAX5 : 51 RECK : 49 TEK : 58 MTAP : 68 CDKN2B-AS1 : 70 **CDKN2A** : 70 MLLT3 : 67 IFNA1 : 44 NFIB : 58 PSIP1 : 40 PTPRD : 68 CD274 : 58 UHRF2 : 54 JAK2 : 53 SMARCA2 : 61 KANK1 : 58	IGFBPL1 : 40 PAX5 : 43 RECK : 40 TEK : 61 MTAP : 73 CDKN2B-AS1 : 70 CDKN2A : 50 MLLT3 : 64 IFNA1 : 54 NFIB : 61 PSIP1 : 39 PTPRD : 60 CD274 : 52 UHRF2 : 51 JAK2 : 50 SMARCA2 : 55 KANK1 : 56
17	17p11.2 17p12 17p13.1 17p13.1 17p13.1 17p13.1 17p13.3 17p13.3 17p13.3 17p13.3 17p13.3	FLCN : 49 MAP2K4 : 60 GAS7 : 68 DNAH2 : 63 TP53 : 32 **PIK3R5** : 66 SMYD4 : 61 VPS53 : 61 YWHAE : 56 PAFAH1B1 : 54 DPH1 : 53	FLCN : 48 MAP2K4 : 65 GAS7 : 73 DNAH2 : 74 TP53 : 32 PIK3R5 : 72 SMYD4 : 68 VPS53 : 67 YWHAE : 69 PAFAH1B1 : 69 DPH1 : 57

18	18p11.22 18p11.31 18p11.31 18q11.2 18q11.2 18q12.3 18q21.1 18q21.1 18q21.1 18q21.2 18q21.2 18q21.2 18q21.31 18q21.32 18q21.33 18q21.33 18q21.33 18q23	PTPRM : 35 EPB41L3 : 35 L3MBTL4 : 35 SS18 : 58 ZNF521 : 56 SETBP1 : 72 ZBTB7C : 74 **SMAD2 :** 70 **MAPK4 :** 70 **DCC :** 75 TCF4 : 74 **SMAD4 :** 46 MIR122 : 68 MALT1 : 75 **BCL2 :** 75 PHLPP1 : 75 SERPINB5 : 67 GALR1 : 72	PTPRM : 24 EPB41L3 : 26 L3MBTL4 : 26 SS18 : 52 ZNF521 : 46 SETBP1 : 79 ZBTB7C : 80 SMAD2 : 79 MAPK4 : 81 DCC : 81 TCF4 : 80 SMAD4 : 46 MIR122 : 74 MALT1 : 79 BCL2 : 80 PHLPP1 : 80 SERPINB5 : 69 GALR1 : 76
19	19p13.3 19p13.3	PLK5 : 28 SAFB : 28	PLK5 : 18 SAFB : 13
22	22q11.23 22q12.1 22q12.1 22q12.1 22q12.2 22q12.2 22q12.3 22q13.1 22q13.2 22q13.31	BCR : 25 MYO18B : 28 MN1 : 28 CHEK2 : 28 SEC14L2 : 26 NF2 : 25 TIMP3 : 26 PDGFB : 26 BIK : 30 PRR5 : 33	BCR : 28 MYO18B : 32 MN1 : 33 CHEK2 : 33 SEC14L2 : 29 NF2 : 29 TIMP3 : 39 PDGFB : 31 BIK : 33 PRR5 : 37

1b: Table shows the frequency of genes being amplified.

Chromosome	Cytoband	Oslo	TCGA
1	1q25.2 1q25.3 1q25.3 1q31.1 1q41 1q41 1q43	RASAL2 : 25 CACNA1E : 25 RGL1 : 25 PLA2G4A : 26 CAPN2 : 23 TGFB2 : 23 AKT3 : 22	RASAL2 : 39 CACNA1E : 40 RGL1 : 40 PLA2G4A : 41 CAPN2 : 36 TGFB2 : 38 AKT3 : 36
3	3q25.1 3q25.31	WWTR1 : 25 CCNL1 : 26	WWTR1 : 19 CCNL1 : 16

7	7p11.2 7p12.3 7p14.1 7p15.3 7p15.3 7p21.1 7p21.1 7q21.2 7q21.3 7p22.2 7p22.3 7q22.3	EGFR : 21 ADCY1 : 25 SFRP4 : 23 CYCS : 26 RAPGEF5 : 26 HDAC9 : 24 ITGB8 : 26 CDK6 : 23 GNGT1 : 25 GNA12 : 23 MAD1L1 : 25 PIK3CG : 21	EGFR : 31 ADCY1 : 28 SFRP4 : 30 CYCS : 29 RAPGEF5 : 30 HDAC9 : 30 ITGB8 : 29 CDK6 : 31 GNGT1 : 33 GNA12 : 31 MAD1L1 : 33 PIK3CG : 29
8	8q22.1 8q22.2 8q22.3 8q23.1 8q24.11 8q24.21 8q24.21 8q24.21 8q24.21 8q24.22 8q24.22 8q24.3 8q24.3 8q24.3 8q24.3	MTDH : 23 STK3 : 26 RRM2B : 26 ANGPT1 : 28 EXT1 : 28 MYC : 25 PCAT1 : 32 ASAP1 : 30 CCDC26 : 30 NDRG1 : 30 ADCY8 : 30 TRIB1 : 27 KCNK9 : 30 BAI1 : 28 PTK2 : 32	MTDH : 32 STK3 : 37 RRM2B : 35 ANGPT1 : 38 EXT1 : 40 MYC : 42 PCAT1 : 43 ASAP1 : 42 CCDC26 : 42 NDRG1 : 39 ADCY8 : 42 TRIB1 : 33 KCNK9 : 39 BAI1 : 35 PTK2 : 39
13	13q14.3 13q22.1 13q31.1 13q31.3 13q32.1 13q33.1 13q33.1 13q34 13q34 13q34	NEK3 : 25 KLF5 : 25 SPRY2 : 28 GPC5 : 27 RAP2A : 25 FGF14 : 30 ERCC5 : 30 IRS2 : 32 RASA3 : 30 TFDP1 : 26	NEK3 : 16 KLF5 : 20 SPRY2 : 22 GPC5 : 22 RAP2A : 21 FGF14 : 22 ERCC5 : 17 IRS2 : 20 RASA3 : 21 TFDP1 : 09
19	19q12 19q12 19q12 19q13.11 19q13.11	CCNE1 : 21 POP4 : 23 UQCRFS1 : 23 ANKRD27 : 25 ZNF507 : 22	CCNE1 : 20 POP4 : 20 UQCRFS1 : 20 ANKRD27 : 22 ZNF507 : 24

“Supplementary Table 2 is not added because of the large size”.

Supplementary Table S3

	OUH Cohort		TCGA Cohort	
PathwayName	**#Gene**	**Statistics**	**#Gene**	**Statistics**
Adherens junction	15	C=73; O=15; E=5.45; R=2.75; rawP=0.0003; adjP=0.0008	16	C=73; O=16; E=6.69; R=2.39; rawP=0.0008; adjP=0.0011
Adipocytokine signaling pathway	11	C=68; O=11; E=5.08; R=2.17; rawP=0.0116; adjP=0.0160	20	C=68; O=20; E=6.23; R=3.21; rawP=1.80e-06; adjP=5.94e-06
Amoebiasis	30	C=106; O=30; E=7.91; R=3.79; rawP=1.13e-10; adjP=2.80e-09	37	C=106; O=37; E=9.72; R=3.81; rawP=2.82e-13; adjP=3.15e-12
Apoptosis	14	C=87; O=14; E=6.50; R=2.16; rawP=0.0050; adjP=0.0081	25	C=87; O=25; E=7.97; R=3.14; rawP=1.59e-07; adjP=6.93e-07
Arachidonic acid metabolism	10	C=59; O=10; E=4.41; R=2.27; rawP=0.0114; adjP=0.0160	18	C=59; O=18; E=5.41; R=3.33; rawP=3.27e-06; adjP=8.97e-06
Arginine and proline metabolism	12	C=54; O=12; E=4.03; R=2.98; rawP=0.0005; adjP=0.0011	11	C=54; O=11; E=4.95; R=2.22; rawP=0.0089; adjP=0.0101
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	17	C=74; O=17; E=5.53; R=3.08; rawP=2.50e-05; adjP=9.89e-05	20	C=74; O=20; E=6.78; R=2.95; rawP=7.52e-06; adjP=1.73e-05
Autoimmune thyroid disease	15	C=52; O=15; E=3.88; R=3.86; rawP=3.81e-06; adjP=1.95e-05	25	C=52; O=25; E=4.77; R=5.25; rawP=4.27e-13; adjP=4.34e-12
Axon guidance	25	C=129; O=25; E=9.63; R=2.60;	33	C=129; O=33; E=11.82; R=2.79;

		rawP=9.22e-06; adjP=4.20e-05		rawP=4.19e-08; adjP=2.32e-07
Bacterial invasion of epithelial cells	10	C=70; O=10; E=5.23; R=1.91; rawP=0.0345; adjP=0.0403	14	C=70; O=14; E=6.42; R=2.18; rawP=0.0041; adjP=0.0051
Bile secretion	11	C=71; O=11; E=5.30; R=2.08; rawP=0.0158; adjP=0.0211	18	C=71; O=18; E=6.51; R=2.77; rawP=5.38e-05; adjP=9.94e-05
Calcium signaling pathway	38	C=177; O=38; E=13.22; R=2.88; rawP=2.74e-09; adjP=3.20e-08	36	C=177; O=36; E=16.22; R=2.22; rawP=4.26e-06; adjP=1.06e-05
Cell adhesion molecules (CAMs)	25	C=133; O=25; E=9.93; R=2.52; rawP=1.60e-05; adjP=6.62e-05	49	C=133; O=49; E=12.19; R=4.02; rawP=3.21e-18; adjP=1.96e-16
Cell cycle	22	C=124; O=22; E=9.26; R=2.38; rawP=0.0001; adjP=0.0003	27	C=124; O=27; E=11.37; R=2.38; rawP=1.79e-05; adjP=3.83e-05
Chagas disease (American trypanosomiasis)	17	C=104; O=17; E=7.76; R=2.19; rawP=0.0018; adjP=0.0032	22	C=104; O=22; E=9.53; R=2.31; rawP=0.0002; adjP=0.0003
Chemokine signaling pathway	31	C=189; O=31; E=14.11; R=2.20; rawP=2.88e-05; adjP=0.0001	37	C=189; O=37; E=17.32; R=2.14; rawP=7.95e-06; adjP=1.80e-05
Chronic myeloid leukemia	15	C=73; O=15; E=5.45; R=2.75; rawP=0.0003; adjP=0.0008	18	C=73; O=18; E=6.69; R=2.69; rawP=7.97e-05; adjP=0.0001
Colorectal cancer	14	C=62; O=14; E=4.63; R=3.02; rawP=0.0002; adjP=0.0006	18	C=62; O=18; E=5.68; R=3.17; rawP=7.13e-06; adjP=1.67e-05
Cytokine-cytokine	52	C=265; O=52; E=19.79;	68	C=265; O=68; E=24.29; R=2.80;

receptor interaction		R=2.63; rawP=1.23e-10; adjP=2.80e-09		rawP=3.33e-15; adjP=8.13e-14
Cytosolic DNA-sensing pathway	13	C=56; O=13; E=4.18; R=3.11; rawP=0.0002; adjP=0.0006	16	C=56; O=16; E=5.13; R=3.12; rawP=2.83e-05; adjP=5.57e-05
Dilated cardiomyopathy	16	C=90; O=16; E=6.72; R=2.38; rawP=0.0010; adjP=0.0019	23	C=90; O=23; E=8.25; R=2.79; rawP=4.55e-06; adjP=1.11e-05
ECM-receptor interaction	20	C=85; O=20; E=6.35; R=3.15; rawP=3.37e-06; adjP=1.93e-05	27	C=85; O=27; E=7.79; R=3.47; rawP=4.76e-09; adjP=3.23e-08
Endocytosis	33	C=201; O=33; E=15.01; R=2.20; rawP=1.59e-05; adjP=6.62e-05	43	C=201; O=43; E=18.42; R=2.33; rawP=1.21e-07; adjP=5.47e-07
Fc epsilon RI signaling pathway	13	C=79; O=13; E=5.90; R=2.20; rawP=0.0055; adjP=0.0086	24	C=79; O=24; E=7.24; R=3.31; rawP=8.75e-08; adjP=4.27e-07
Fc gamma R-mediated phagocytosis	16	C=94; O=16; E=7.02; R=2.28; rawP=0.0015; adjP=0.0027	24	C=94; O=24; E=8.62; R=2.79; rawP=2.90e-06; adjP=8.40e-06
Focal adhesion	41	C=200; O=41; E=14.93; R=2.75; rawP=2.81e-09; adjP=3.20e-08	54	C=200; O=54; E=18.33; R=2.95; rawP=2.42e-13; adjP=3.15e-12
Gastric acid secretion	11	C=74; O=11; E=5.53; R=1.99; rawP=0.0210; adjP=0.0262	21	C=74; O=21; E=6.78; R=3.10; rawP=1.90e-06; adjP=6.10e-06
Glycerolipid metabolism	10	C=50; O=10; E=3.73; R=2.68; rawP=0.0034; adjP=0.0056	13	C=50; O=13; E=4.58; R=2.84; rawP=0.0004; adjP=0.0006
Glycerophospholipid metabolism	11	C=80; O=11; E=5.97; R=1.84; rawP=0.0352; adjP=0.0405	22	C=80; O=22; E=7.33; R=3.00; rawP=1.96e-06; adjP=6.13e-06

Glycolysis / Gluconeogenesis	11	C=65; O=11; E=4.85; R=2.27; rawP=0.0083; adjP=0.0120	13	C=65; O=13; E=5.96; R=2.18; rawP=0.0055; adjP=0.0066
GnRH signaling pathway	18	C=101; O=18; E=7.54; R=2.39; rawP=0.0005; adjP=0.0011	28	C=101; O=28; E=9.26; R=3.02; rawP=6.88e-08; adjP=3.50e-07
Hepatitis C	23	C=134; O=23; E=10.00; R=2.30; rawP=0.0001; adjP=0.0003	36	C=134; O=36; E=12.28; R=2.93; rawP=2.52e-09; adjP=1.81e-08
Huntington's disease	24	C=183; O=24; E=13.66; R=1.76; rawP=0.0051; adjP=0.0081	32	C=183; O=32; E=16.77; R=1.91; rawP=0.0003; adjP=0.0005
Hypertrophic cardiomyopathy (HCM)	14	C=83; O=14; E=6.20; R=2.26; rawP=0.0032; adjP=0.0054	21	C=83; O=21; E=7.61; R=2.76; rawP=1.37e-05; adjP=2.98e-05
Insulin signaling pathway	19	C=138; O=19; E=10.30; R=1.84; rawP=0.0071; adjP=0.0104	22	C=138; O=22; E=12.65; R=1.74; rawP=0.0074; adjP=0.0086
Jak-STAT signaling pathway	29	C=155; O=29; E=11.57; R=2.51; rawP=3.86e-06; adjP=1.95e-05	44	C=155; O=44; E=14.21; R=3.10; rawP=6.04e-12; adjP=5.26e-11
Leishmaniasis	12	C=72; O=12; E=5.38; R=2.23; rawP=0.0068; adjP=0.0101	23	C=72; O=23; E=6.60; R=3.49; rawP=5.72e-08; adjP=3.03e-07
Leukocyte transendothelial migration	18	C=116; O=18; E=8.66; R=2.08; rawP=0.0024; adjP=0.0041	32	C=116; O=32; E=10.63; R=3.01; rawP=9.39e-09; adjP=5.73e-08
Long-term potentiation	11	C=70; O=11; E=5.23; R=2.10; rawP=0.0143; adjP=0.0194	12	C=70; O=12; E=6.42; R=1.87; rawP=0.0246; adjP=0.0266
Lysosome	16	C=121; O=16; E=9.03; R=1.77;	18	C=121; O=18; E=11.09; R=1.62;

		rawP=0.0182; adjP=0.0233		rawP=0.0275; adjP=0.0294
MAPK signaling pathway	44	C=268; O=44; E=20.01; R=2.20; rawP=6.83e-07; adjP=5.18e-06	65	C=268; O=65; E=24.56; R=2.65; rawP=2.29e-13; adjP=3.15e-12
Melanogenesis	13	C=101; O=13; E=7.54; R=1.72; rawP=0.0377; adjP=0.0429	20	C=101; O=20; E=9.26; R=2.16; rawP=0.0008; adjP=0.0011
Melanoma	13	C=71; O=13; E=5.30; R=2.45; rawP=0.0021; adjP=0.0037	20	C=71; O=20; E=6.51; R=3.07; rawP=3.77e-06; adjP=1.00e-05
Metabolic pathways	167	C=1130; O=167; E=84.37; R=1.98; rawP=1.53e-17; adjP=1.33e-15	205	C=1130; O=205; E=103.58; R=1.98; rawP=1.32e-21; adjP=1.61e-19
Natural killer cell mediated cytotoxicity	27	C=136; O=27; E=10.15; R=2.66; rawP=2.55e-06; adjP=1.66e-05	40	C=136; O=40; E=12.47; R=3.21; rawP=1.60e-11; adjP=1.30e-10
Neuroactive ligand-receptor interaction	65	C=272; O=65; E=20.31; R=3.20; rawP=2.93e-17; adjP=1.33e-15	71	C=272; O=71; E=24.93; R=2.85; rawP=3.15e-16; adjP=9.61e-15
Neurotrophin signaling pathway	23	C=127; O=23; E=9.48; R=2.43; rawP=6.29e-05; adjP=0.0002	26	C=127; O=26; E=11.64; R=2.23; rawP=7.69e-05; adjP=0.0001
Osteoclast differentiation	21	C=128; O=21; E=9.56; R=2.20; rawP=0.0005; adjP=0.0011	29	C=128; O=29; E=11.73; R=2.47; rawP=3.92e-06; adjP=1.02e-05
p53 signaling pathway	13	C=68; O=13; E=5.08; R=2.56; rawP=0.0014; adjP=0.0026	20	C=68; O=20; E=6.23; R=3.21; rawP=1.80e-06; adjP=5.94e-06

Pancreatic cancer	14	C=70; O=14; E=5.23; R=2.68; rawP=0.0006; adjP=0.0013	21	C=70; O=21; E=6.42; R=3.27; rawP=6.95e-07; adjP=2.57e-06
Pancreatic secretion	18	C=101; O=18; E=7.54; R=2.39; rawP=0.0005; adjP=0.0011	36	C=101; O=36; E=9.26; R=3.89; rawP=2.84e-13; adjP=3.15e-12
Pathways in cancer	57	C=326; O=57; E=24.34; R=2.34; rawP=1.62e-09; adjP=2.46e-08	83	C=326; O=83; E=29.88; R=2.78; rawP=5.26e-18; adjP=2.14e-16
Phagosome	23	C=153; O=23; E=11.42; R=2.01; rawP=0.0010; adjP=0.0019	36	C=153; O=36; E=14.02; R=2.57; rawP=1.05e-07; adjP=4.93e-07
Phosphatidylinositol signaling system	15	C=78; O=15; E=5.82; R=2.58; rawP=0.0006; adjP=0.0013	15	C=78; O=15; E=7.15; R=2.10; rawP=0.0044; adjP=0.0054
PPAR signaling pathway	10	C=70; O=10; E=5.23; R=1.91; rawP=0.0345; adjP=0.0403	12	C=70; O=12; E=6.42; R=1.87; rawP=0.0246; adjP=0.0266
Protein digestion and absorption	25	C=81; O=25; E=6.05; R=4.13; rawP=5.22e-10; adjP=9.50e-09	25	C=81; O=25; E=7.42; R=3.37; rawP=3.34e-08; adjP=1.94e-07
Protein processing in endoplasmic reticulum	30	C=165; O=30; E=12.32; R=2.44; rawP=4.85e-06; adjP=2.32e-05	32	C=165; O=32; E=15.12; R=2.12; rawP=3.82e-05; adjP=7.40e-05
Purine metabolism	25	C=162; O=25; E=12.10; R=2.07; rawP=0.0004; adjP=0.0010	32	C=162; O=32; E=14.85; R=2.16; rawP=2.61e-05; adjP=5.31e-05
Pyrimidine metabolism	15	C=99; O=15; E=7.39; R=2.03; rawP=0.0066; adjP=0.0100	20	C=99; O=20; E=9.07; R=2.20; rawP=0.0006; adjP=0.0009

Regulation of actin cytoskeleton	38	C=213; O=38; E=15.90; R=2.39; rawP=4.69e-07; adjP=3.88e-06	39	C=213; O=39; E=19.52; R=2.00; rawP=2.33e-05; adjP=4.84e-05
Regulation of autophagy	12	C=34; O=12; E=2.54; R=4.73; rawP=3.40e-06; adjP=1.93e-05	12	C=34; O=12; E=3.12; R=3.85; rawP=2.76e-05; adjP=5.52e-05
Rheumatoid arthritis	13	C=91; O=13; E=6.79; R=1.91; rawP=0.0175; adjP=0.0227	25	C=91; O=25; E=8.34; R=3.00; rawP=4.10e-07; adjP=1.67e-06
RIG-I-like receptor signaling pathway	16	C=71; O=16; E=5.30; R=3.02; rawP=5.51e-05; adjP=0.0002	18	C=71; O=18; E=6.51; R=2.77; rawP=5.38e-05; adjP=9.94e-05
RNA transport	23	C=151; O=23; E=11.27; R=2.04; rawP=0.0009; adjP=0.0018	23	C=151; O=23; E=13.84; R=1.66; rawP=0.0108; adjP=0.0121
Salivary secretion	16	C=89; O=16; E=6.65; R=2.41; rawP=0.0008; adjP=0.0016	20	C=89; O=20; E=8.16; R=2.45; rawP=0.0001; adjP=0.0002
Small cell lung cancer	24	C=85; O=24; E=6.35; R=3.78; rawP=8.36e-09; adjP=8.45e-08	32	C=85; O=32; E=7.79; R=4.11; rawP=1.04e-12; adjP=9.76e-12
TGF-beta signaling pathway	12	C=84; O=12; E=6.27; R=1.91; rawP=0.0219; adjP=0.0269	16	C=84; O=16; E=7.70; R=2.08; rawP=0.0037; adjP=0.0047
Tight junction	27	C=132; O=27; E=9.86; R=2.74; rawP=1.40e-06; adjP=9.80e-06	36	C=132; O=36; E=12.10; R=2.98; rawP=1.61e-09; adjP=1.23e-08
Toll-like receptor signaling pathway	19	C=102; O=19; E=7.62; R=2.49; rawP=0.0002; adjP=0.0006	25	C=102; O=25; E=9.35; R=2.67; rawP=4.04e-06; adjP=1.03e-05

Toxoplasmosis	22	C=132; O=22; E=9.86; R=2.23; rawP=0.0003; adjP=0.0008	44	C=132; O=44; E=12.10; R=3.64; rawP=1.13e-14; adjP=2.30e-13
Tryptophan metabolism	11	C=42; O=11; E=3.14; R=3.51; rawP=0.0002; adjP=0.0006	12	C=42; O=12; E=3.85; R=3.12; rawP=0.0003; adjP=0.0005
Ubiquitin mediated proteolysis	17	C=135; O=17; E=10.08; R=1.69; rawP=0.0236; adjP=0.0286	23	C=135; O=23; E=12.37; R=1.86; rawP=0.0027; adjP=0.0035
Vascular smooth muscle contraction	17	C=116; O=17; E=8.66; R=1.96; rawP=0.0056; adjP=0.0086	29	C=116; O=29; E=10.63; R=2.73; rawP=4.56e-07; adjP=1.79e-06
Viral myocarditis	20	C=70; O=20; E=5.23; R=3.83; rawP=1.15e-07; adjP=1.05e-06	30	C=70; O=30; E=6.42; R=4.68; rawP=9.24e-14; adjP=1.61e-12
Wnt signaling pathway	18	C=150; O=18; E=11.20; R=1.61; rawP=0.0312; adjP=0.0374	34	C=150; O=34; E=13.75; R=2.47; rawP=5.97e-07; adjP=2.28e-06

Paper IV

Generation and characterization of novel pancreatic adenocarcinoma xenograft models and corresponding primary cell lines

Wennerstrom A.B., Bowitz Lothe I.M., Sandhu V., Kure E.H., Myklebost O., Munthe E.

PLoS One, 2014, 9, e103873.

OPEN ACCESS Freely available online

PLOS ONE

Generation and Characterisation of Novel Pancreatic Adenocarcinoma Xenograft Models and Corresponding Primary Cell Lines

Anna B. Wennerström[1], Inger Marie Bowitz Lothe[2,3], Vandana Sandhu[2,4], Elin H. Kure[2,4], Ola Myklebost[1], Else Munthe[1]*

1 Cancer Stem Cell Innovation Centre and Department of Tumour Biology, Institute of Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **2** Department of Genetics, Institute of Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **3** Department of Pathology, Oslo University Hospital, Oslo, Norway, **4** Department for Environmental Health and Science, Telemark University College, Telemark, Norway

Abstract

Pancreatic adenocarcinoma is one of the most lethal cancer types, currently lacking efficient treatment. The heterogeneous nature of these tumours are poorly represented by the classical pancreatic cell lines, which have been through strong clonal selection *in vitro*, and are often derived from metastases. Here, we describe the establishment of novel pancreatic adenocarcinoma models, xenografts and corresponding *in vitro* cell lines, from primary pancreatic tumours. The morphology, differentiation grade and gene expression pattern of the xenografts resemble the original tumours well. The cell lines were analysed for colony forming capacity, tumourigenicity and expression of known cancer cell surface markers and cancer stem-like characteristics. These primary cell models will be valuable tools for biological and preclinical studies for this devastating disease.

Citation: Wennerström AB, Lothe IMB, Sandhu V, Kure EH, Myklebost O, et al. (2014) Generation and Characterisation of Novel Pancreatic Adenocarcinoma Xenograft Models and Corresponding Primary Cell Lines. PLoS ONE 9(8): e103873. doi:10.1371/journal.pone.0103873

Editor: Ilse Rooman, Garvan Institute of Medical Research, Australia

Received April 30, 2014; **Accepted** July 2, 2014; **Published** August 22, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper, except the microarray data. The data are available through Gene Expression Ominbus (GSE58561).

Funding: This work was financially supported by the Cancer Stem Cell Innovation Centre (CAST-SFI) and HSØ (Grant nr. 2011090). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: elsmun@rr-research.no

Introduction

The 5-year overall survival of patients with pancreatic cancer is a dismal 6.7%. Although overall mortality for all cancers has fallen from 210 to 171 pr. 100 000 US citizens from 1992 to 2010, the overall mortality for pancreatic cancers is unchanged. So, even though pancreatic cancers account for less than 10% of the diagnosed tumours, it is the 4th leading cause of cancer-related deaths in the US (www.cancer.gov). The high mortality may partly be due to the abdominal localisation where tumours may advance without early symptoms, and are diagnosed late in the disease progression, well after the acquisition of the aggressive nature of this cancer type [1]. Thus, pancreatic cancers are often metastatic and resistant towards irradiation and chemotherapy at the time of diagnosis, with a corresponding lack of efficient treatment options for the patients. The most common malignant pancreatic tumours are the pancreatic ductal adenocarcinomas (PDACs), originating from epithelial cells lining the pancreatic ducts, accounting for more than 85% of pancreatic tumours. PDACs can be of the pancreatobiliary or intestinal type, where the pancreatobiliary is the most common [2,3]. The pancreatobiliary type tumours mostly consist of glandular and duct-like structures, well to moderately developed, growing in a desmoplastic stroma. The poorly differentiated tumours form densely packed, small irregular glands as well as solid sheets and individual cells. The intestinal type adenocarcinoma form simple or cribriform glands, and are similar to the adenocarcinomas of the large intestine in growth pattern and differentiation. The degree of differentiation in pancreatic tumours has been found to be an independent prognostic marker to the same degree as tumour size and lymph node status [4]. The "cancer stem cell" hypothesis has been under intense investigation over recent years, and in many cancer types cells with stem cell characteristics are found to generate tumours much more efficiently upon injection in mice than do bulk tumour cells [5]. These so-called "cancer stem-like cells" (CSC) have the capacity for self-renewal as well as the capacity to differentiate, and have an increased resistance to cancer treatments like chemotherapy and irradiation. Altogether, these characteristics permit CSCs to generate metastases as well as treatment-resistant recurrences. Several candidate pancreatic CSC markers have been identified, including cell surface markers CD24/CD44/CD326 [6] or CD133 [7], side population positive cells [8], and cells with aldehyde dehydrogenase activity [9].

At present, there is a lack of relevant model systems to study clinically important subpopulations of tumour cells, e.g., stem-like cells, in pancreatic cancers. Few pancreatic cancer cell lines are available, and those commonly used have been grown in culture for extended periods of time. Long term *in vitro* cultivation may induce a selective pressure to adapt to the culture conditions and the cell lines thereby no longer represent the original heterogenic

tumours whereby the cells may gain mutations or altered programming *in vitro*. Furthermore, many of the classical pancreatic cell lines were generated from metastatic tumours [10]. Few primary models are available due to the limited amounts of available surgical material. In addition, there are problems generating *in vitro* cultures due to the high stromal infiltrations in pancreatic tumours, from which rapidly growing fibroblasts tend to overgrow the adenocarcinoma cells. To overcome these difficulties we generated xenografts from surplus operation material from patients with primary pancreatic tumours and thereafter established cell lines from these xenograft-passaged tumours. The original tumours and the xenografts show the same histology regarding growth pattern and differentiation. All but one tumour that generated xenografts were of the pancreatobiliary type, three moderately differentiated and three poorly differentiated. The last tumour was a moderately differentiated PDAC of intestinal type. All the generated cell lines matched the original tumours' fingerprints, had global mRNA expression pattern resembling their corresponding original tumours and were tumourigenic when injected into NOD/SCID mice. We characterised these cell lines for cell surface expression of markers known to be important for tumourigenicity and potential cancer stem cell markers during *in vitro* passaging. A schematic overview of the workflow performed in this study is found in fig. S1.

Materials and Methods

Ethics statements

The study was approved by the Regional Committee for Medical and Health Research Ethics South-East Norway (265-08412c) and the Institutional Review Board of Oslo University Hospital, and performed according to the guidelines of the Helsinki Convention. Informed written consent was obtained from all patients. Animal work was performed according to protocols approved by the National Animal Research Authority in compliance with the European Convention of the Protection of Vertebrates Used for Scientific Purposes (approval ID 3275 and 3530; http://www.fdu.no/). All surgery was performed under sevofluran anaesthesia, and all efforts were made to minimize suffering.

Mutation analysis

The *KRAS* mutations (codons 12 and 13) were determined using the Wobble-enhanced amplification refractory mutation system [11].

Pathology/Immunohistochemistry

The macro- and microscopic pathology work followed a standardised protocol. Experienced pathologists set the final diagnoses in accordance with the WHO Classification of Tumors of the Digestive System [12]. The tumours were sub-classified into pancreatobiliary and intestinal type as first described by Kimura *et al.* [13]. Furthermore, the adenocarcinomas were classified according to site of origin and tumour stage, in accordance with the pTNM Classification of Malignant Tumours [14].

The patient and xenograft tumours were subjected to semi-quantitative evaluation by immunohistochemistry (IHC). Four μm whole sections from formalin-fixed paraffin-embedded (FFPE) tissue were mounted on Menzel Super Frost Plus glass slides. For deparaffinisation and heat-induced epitope retrieval, Dako PT link was used (Dako, Glostrup, Denmark). The immunostaining was performed on the Dako EnVision+ detection system. The Dako EnVision Flex+Target Retrieval Solution at high pH was used for preheating for all the antibodies, except for Snail+Slug antibody, where low pH was used. The slides were rinsed in Dako wash buffer according to the manufacturer's instructions. Endogenous peroxidase activity was blocked for five minutes by 0.03% H_2O_2, washed twice in Dako wash buffer and then incubated with primary antibody at room temperature for 30 minutes before an additional wash step. The next step was 30-minute incubation with the appropriate HRP-labelled polymer conjugated secondary antibodies at room temperature before wash and ten minute incubation in diaminobenzidine (DAB). The last steps were rinsing twice in water, counterstaining with haematoxylin and mounting in Diatex. The protocol was optimised for each antigen. CK7 (Dako cat. 7081, 1:300), E-Cadherin (Invitrogen, CA, USA cat. 13-1700, 1:3000), Snail+Slug (AbCam, Cambridge, UK, cat ab63371, 1:100), Vimentin (Dako cat. M0851, 1:1600), CD326 (EpCAM) (Dako cat. M3525, 1:60). P53 (Santa Cruz Biotechnology, SC, USA, cat sc-126, 1:5000) and S100A4 [15].

Tumour xenografting

Fresh, surgically excised primary pancreatic adenocarcinoma material was kept in RPMI1640 (Lonza, Switzerland) with 1× Penicillin-Streptavidin (PS) (Lonza) and 0.5 μg/ml Fungizone (Gibco, USA) on ice and transported directly to the animal facility within two hours. The patient material was cut into 2×2 mm pieces and one piece implanted under the skin on each flank of minimum four locally bred female NOD/SCID IL2R γ^0 (NSG) mice. The general health of the mice was monitored daily and xenograft growth twice a week. The mice were sacrificed when tumours reached 10–15 mm in diameter, and then the xenografts were passaged and/or used to establish *in vitro* cultures. In addition, paraffin-embedded sections were made from early passages.

Isolation and propagation of human cells *in vitro*

Xenografted tumours were excised when reaching 10 mm and cells were extracted with gentleMACS dissociator (Miltenyi, Germany) according to the manufacturers protocol "*Preparation of single-cell suspensions from implanted mouse tumours, protocol 2.2.2*". In brief, the xenografts were minced to <5 mm in size and incubated on a tube rotator with 250 U Collagenase I (Worthington, USA) and 0.77 U/ml Dispase (Roche, Switzerland) for 20 minutes at 37°C. Thereafter, samples were run on program "m_impTumor_04" on the dissociator and further rotated for another 20 minutes at 37°C. DNase I (Calbiochem, Germany) was added to a final concentration of 2000 U/ml followed by a last round of the "m_impTumor_04" program. The cell suspension was washed using PEB buffer (0.5% bovine serum albumin (Sigma-Aldrich, USA) and 2 mM EDTA (Lonza)), pelleted by centrifugation at 300 g, resuspended in PBS/2% FCS and filtered through a 40 μM filter to obtain single cells.

To remove mouse cells, magnetic bead depletion was used. The freshly isolated single cell suspension was incubated with mouse anti-H-2Kd antibody (5 μg/10^6 cells, Becton Dickinson, USA) for 20 minutes at 4°C. Excess antibody was washed away using 2% FCS in PBS (Lonza) by centrifugation at 300 g and the cells were incubated with Dynabeads Pan Mouse IgG (Invitrogen) for 30 minutes at 4°C. Thereafter, the cell suspension was placed onto the magnet for two minutes. The supernatant containing enriched human cells was transferred to a clean tube, washed, pelleted by centrifugation and subsequently resuspended in Pancreatic cell culture medium with 1× pancreatic cell culture supplement (= "pancreatic medium") from Millipore, (USA).

The cells were seeded at 20 000 cells/cm^2 in pancreatic medium and propagated at 37°C, 5% CO_2. The medium was changed every second or third day and cells were passaged when

reaching sub-confluence or every 3rd week, using trypsin/EDTA (Lonza) diluted 1:5 in PBS. Remaining mouse cells were removed by several washes with PBS after four minutes incubation in trypsin/EDTA during the first passages to avoid them overgrowing the cultures. Pictures of the cultures were taken with Olympus 1×81 with software Cell^P version 3.3 (Olympus, Germany), or using the Incucyte (Essen Bioscience, UK).

Flow cytometry analyses and sorting

Subconfluent cell cultures were trypsinized, washed, and resuspended in blocking solution (1 mg/ml human IgG, (Sigma) in PBS), at 4°C for 30 minutes. Cells were pelleted by centrifugation as above and resuspended in blocking buffer with the following monoclonal antibodies, all at concentrations recommended by the manufacturer, TRA-1-85 (clone TRA-1-85, R&D, USA), CD133/1 (clone AC133) and CD133/2 (clone 293C3) both from Miltenyi. H2kD (clone SF1-1.1), CD24 (clone ML5), CD44 (clone G44-26 (C26), CD166 (clone 3A6), CD326 (clone EBA-1), CD184 (clone ID9), SSEA-4 (clone MC813-70), CD15 (clone HI89), all from Becton Dickinson. The cells were incubated at 4°C for 30–45 minutes, washed twice in FACS flow with 2% FCS and resuspended in FACS flow with 2% FCS. Propidium iodide was added to the cells immediately before analysed on a FACS Aria II with FACS Diva software 6.1.3 (Becton Dickinson) or LSR II with FACS Diva software 5.1 (Becton Dickinson).

Aldehyde dehydrogenase activity and side population assay

Aldehyde dehydrogenase activity was measured using the Aldeflour assay kit (Stem cell technologies, USA) according to manufacturer's protocol. Briefly, 1×10^6 single cells were put in Aldeflour buffer, Aldeflour reagent was added and half of the cells were immediately transferred to a tube with DEAB inhibitor. Both tubes were incubated at 37°C for 30 minutes. Thereafter, cells were centrifuged, the supernatant discarded and the cells were resolved in cold Aldeflour buffer before beeing analysed by flow as described in the antibody section.

The efflux capacity of cells was measured in a classical side population assay. Briefly, hoechst 33342 (Sigma-Aldrich) was added to 1×10^6 single cells in 2% FCS to the final concentration of 5 μg/ml and the mixture was incubated at 37°C on a tube rotator at slow speed for 90 minutes. Thereafter, the cells were centrifuged, supernatant discarded and the cells resuspended in ice-cold PBS (Lonza) with 2% FCS and analysed on flow as described above. To provide information of the different efflux pumps active, the following inhibitors was used: Verapamil (50 μM), Reserpine (10 μM) and Fumitremorgin C (10 μM), all from Sigma.

Generation of colonies in methylcellulose

Cells sorted by flow cytometry or bulk population single cells were plated in methylcellulose (Methocult from Stem Cell Technology, France) mixed with one of the following media: Stem cell medium I: 20 ng/ml EGF (PeproTech, USA), 10 ng/ml bFGF (PeproTech, USA), 1× B27 (Gibco) and 1× PS (Lonza) in 1× Defined Keratinocyte-SFM (Gibco); Stem Cell medium II: 20 ng/ml EGF, 20 ng/ml bFGF, 1× B27, 1× PS and 1× Glutamax (Gibco) in DMEM-F12 (Gibco); Pancreatic medium: Pancreatic culture media with pancreatic cell culture supplement and PS and RPMI medium: 10% FCS, 1× PS and 1× Glutamax in RPMI 1640 (Lonza) according to manufacturers' protocol. In brief, single cell suspensions were added to media with Methocult, mixed and seeded at a density of 1000 cells/well in a non-adhesive 24 well plate (Grenier Bio-One, Germany). Two weeks post plating, the colonies were stained with 150 μl 0.4 mg/ml Thiazolyl Blue Tetrazolium Bromide (MTT) (Sigma-Aldrich) at 37°C for 4 hours or overnight. Uniform colonies of minimum 50 μm were scanned using the GelCount machine and the number and sizes were quantified using GelCount software (both from Oxford Optronix, UK).

In vivo tumourigenicity

Subconfluent cell cultures were trypsinized, washed in RPMI 1640 and filtered to obtain a single cell suspension and thereafter counted in triplet using trypan blue to exclude dead cells. Cells were injected subcutaneously into the flanks of locally bred female NSG mice and tumour size was monitored twice a week.

Microarray analysis

100 ng total RNA was converted to cDNA, amplified and labelled with Cy-3 using LowInput QuickAmp Labelling Kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The global transcription profiles were evaluated using the SurePrint G3 Human GE 8x60K microarrays (Agilent Technologies, Santa Clara, CA, USA). The mRNA microarray expression profiles from three cell lines with three replicates for two cell lines, four replicates for one cell line, three fresh frozen tumour samples and two corresponding normal tissue samples were background corrected and quantile normalised. Removal of control probes, un-annotated genes and low expressed probes was done by calculating the 95% percentile of the negative control probes on each array. Probes that were at least 10% brighter than the negative controls on at least 50% of the arrays were kept. Further, 6810 differentially expressed mRNAs were removed by carrying out a moderated *t*-test for cell line expression versus tumour samples, since these genes may result from contaminating normal and stromal cells in tumour tissues or proliferation associated mRNAs. The unsupervised hierarchical clustering was carried out using Pearson and Spearman's rank correlation coefficient using 10 948 expressed mRNAs. The data are available through Gene Expression Ominbus (GSE58561).

Results and Discussion

Generation of xenografts from pancreatic ductal adenocarcinomas

Due to the limited amounts of surgically removed tissue available and the high percentage of stromal cells in these tumours, we first generated xenografts from the tumour material in NSG mice. In this way, we expanded the tissue and avoided the problem of infiltrating human stromal cells overgrowing the cancer cells. Mouse cells that infiltrated and supported tumour growth could then be removed based on the expression of mouse MHC class I during isolation of cells from the resulting xenograft tumours.

The human samples (Ppa1-9) were obtained from patients with primary pancreatic tumours without known distant metastases, that underwent pancreatoduodenectomy with curative intent. None of the patients had received neoadjuvant treatment. The histological diagnosis of the surgical specimens confirmed that eight of the nine tumours were PDACs, and seven of these were of the expected pancreatobiliary subtype. One tumour, Ppa6, was of the intestinal subtype. The last tumour, Ppa5, was however an intraductal papillary mucinous neoplasia of pancreatobiliary type and was one of the two tumours not able to generate tumours in mice (Table 1 and 2). S100A4- and p53 statuses of the PDAC

Table 1. Characterisation of patient material.

Patient Sample	Gender	Age	Diagnosis[1]	Degree of differentiation[2]	Stage[3]	Resection margin[4]	Disease-free survival (days)	Overall-survival (days)
Ppa1	F	56	PDAC (P)	Poorly	T3	1	325	525
Ppa2	M	63	PDAC (P)	Moderately	T3	1	195	633
Ppa3	M	66	PDAC (P)	Poorly	T3	1	143	481
Ppa4	F	76	PDAC (P)	Moderately	T3	0	286	410
Ppa5	M	72	IPMN (P)	High-grade dysplasia	Tis	0	365[#]	365[#]
Ppa6	F	51	PDAC (I)	Moderately	T3	1	184	1228*
Ppa7	F	70	PDAC (P)	Moderately	T3	0	1018*	1018*
Ppa8	F	59	PDAC (P)	Poorly	T3	1	162	634
Ppa9	M	58	PDAC (P)	Moderately	T3	0	985*	985*

[1]: PDAC = Pancreatic ductal adenocarcinoma, P = pancreatobiliary subtype, I = intestinal subtype.
IPMN = Intraductal papillary mucinous neoplasia.
[2]: The degree of differentiation for each tumour is according to the pTNM Classification of Malignant Tumours [14].
[3]: T3 = Tumour extends beyond pancreas, but without involvement of celiac axis or superior mesenteric artery, Tis = Carcinoma *in situ*.
[4]: 0 = resection margin free and 1 = resection margin not free.
[#]: Patient only followed up the first year after operation, due to non-malignant disease.
*: Patient is alive, the number of days given refers to the latest follow-up date.
doi:10.1371/journal.pone.0103873.t001

Table 2. Characterisation of tumour mutation status and xenograft tumours.

Patient Sample	KRAS mut	P53[1]	S100A4 cytoplasm[2]	S100A4 nuclear[2]	Xenograft take	Implanted tumours >10 mm[3]	Xenograft growth[4]	Growth *in vitro*
Ppa1	G12V	5 (1)	5 (3)	4 (3)	10/10	13	4–8	yes
Ppa2	G12V	6 (3)	5 (3)	5 (3)	5/24	10	4–9	yes
Ppa3	G12D	6 (3)	5 (3)	5 (3)	10/10	11	8–11	no
Ppa4	G12V	6 (2)	5 (3)	4 (3)	9/10	9	5–10	limited
Ppa5	WT				0/10			
Ppa6	G12R	6 (3)	0 (0)	0 (0)	3/10	28	6–7	yes
Ppa7	WT	5 (2)	4 (1)	0 (0)	7/10	12	8–11	no
Ppa8	G12D	6 (3)	4 (2)	2 (3)	5/8	9	5–8	yes
Ppa9	G12C	6 (3)	5 (1)	0 (0)	0/8			

[1]: p53 staining is scored as hot spots where 5 = 30–60% and 6 = 60–100%, and the staining intensity is given in brackets (1 = weak, 2 = medium and 3 = strong).
[2]: S100A4 is scored as % positive cells where 1 = 1–4%, 2 = 5–9%, 3 = 10–14%, 4 = 15–49%, 5≥50%, and the staining intensities is given in brackets (1 = weak, 2 = medium and 3 = strong).
[3]: Time in weeks required for the implanted patient material to reach 10 mm.
[4]:Time in weeks required for the xenograft passage to reach 10 mm for F3 and forward.
doi:10.1371/journal.pone.0103873.t002

patient tumours was analysed by immunohistochemistry, while the KRAS status was analysed by real-time PCR. All PDAC tumours had increased levels of p53, although the expression in Ppa1 was weak. All but one PDAC patient sample had KRAS mutations, and the IPMN patient had WT KRAS (Table 2). For the metastasis-associated protein S100A4, we found cells with strong cytoplasmic and/or nuclear expression in patient material Ppa1, Ppa2 and Ppa8 (Table 2). All the original tumours were STR fingerprinted at the OUS genotyping core facility at the Norwegian Radium Hospital (Table S1).

Tumours from seven of the nine individual patients gave palpable xenografts within three months after implantation (PpaX1-9 passage F1) (Table 2). Overall, the median time for patient material to establish palpable tumours above 10 mm was 15±6 weeks. There were slight differences between pieces derived from the same tumour (±4 weeks), but these variations were most pronounced between samples from each patient, ranging from 11 to 28 weeks. The variations in time until engraftment could reflect the aggressiveness or proliferation rate of the tumour cells, the fraction of tumourigenic cells within the implanted material and/or the time the tumour cells need to adapt a new environment. Interestingly, the degree of tumour differentiation did not seem to influence the growth rate. However, tumours with poor differentiation had a higher rate of engraftment for the initial operation material, from 90% for poor differentiated tumours to 40% for moderate differentiated tumours (Table 2). This is in agreement with the observation that poorly differentiated adenocarcinomas predicts a poor survival for patients with operable pancreatic cancer [16].

All the initially established xenografts could be serially passaged in mice, and the subsequent passages generally grew faster with more homogenous growth rates between pieces from the same tumour. From the third round of passaging in mouse (F3 and onward), the time before the tumours reached 10 mm stabilized (+/−2.5 weeks) with the average generation time being characteristic for each xenograft line (Table 2). The observed decrease in passaging time during the first passages is in agreement with data from Kim *et al.* [17].

Early passage xenografts were used to make paraffin embedded slides, that later were haematoxylin- and eosin- (HE) stained. These slides were evaluated by a trained pathologist and compared with the original patient sample. In all cases the morphology and differentiation grade were judged to be similar in the xenografts and their corresponding original tumours (Fig. 1).

To further compare the xenografts with the original patient material, we performed immunohistochemistry for selected biomarkers. All examined tumours and xenografts had strong staining for the human-specific antibody against CD326 (EpCAM) and the diagnostic marker cytokeratin 7 (CK7) in the majority of the cells, confirming that all xenografts consisted of human pancreatic cancer cells of epithelial origin (Table 3). Notably, sections from the implanted sites showed that some tumours were able to infiltrate and disrupt the surrounding muscles, which might explain the paralysis observed in some mice carrying PpaX2, PpaX6 or PpaX8 xenografts. Since invasive traits, stemness and tumour aggressiveness may be induced by epithelial-mesenchymal transition (EMT) [18], and expression of EMT markers has been found to be of prognostic value for pancreas cancers [19], expression of EMT-related markers was also investigated. Using a human-specific antibody against the mesenchymal marker vimentin we observed a strong, ubiquitous staining in the stroma of the original tumours in contrast to very few positive stromal cells in the xenografts, demonstrating that very few human stromal cells remain in the xenografts, even in early passages (Table 3). While

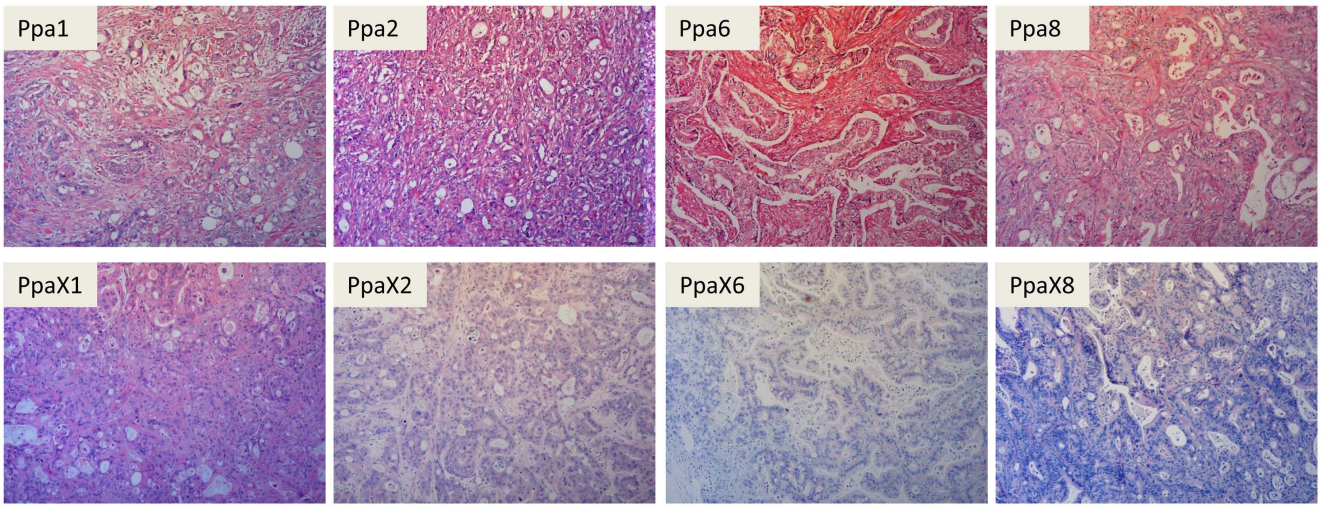


Figure 1. Morphology of the human tumours and the corresponding xenograft tumours. HE-stained sections of original patient material (upper panel) and xenograft tumours at the following passages, PpaX1 p4, PpaX2 p1, PpaX6 p7 and PpaX8 p2, (lower panel) demonstrate the overall conservation of histological features for each tumour pair.
doi:10.1371/journal.pone.0103873.g001

all original tumours and the other xenografts were negative for vimentin, PpaX2 had a strong, ubiquitous expression of vimentin in the tumour cells, which could suggest that these cells have undergone EMT. All tumours and xenografts had a strong membrane-bound expression for the epithelial marker E-Cadherin in most tumour cells, except the PpA2 original tumour that displayed a weaker staining and in fewer cells (Table 3). Interestingly, the xenografts PpaX1, PpaX2 and PpaX6 all had prevalent, strong nuclear expression of Snail and/or Slug, known inducers of EMT that represses E-Cadherin (Table 3). In contrast, all the original tumours were negative for Snail and/or Slug, and we hypothesize that Snail and/or Slug might be induced by the mouse subcutaneous microenvironment and perhaps orthotopic implantation would be more relevant for *in vivo* EMT studies. A lack of correlation between Snail- and E-Cadherin expression in pancreatic cancers has also previously been reported [20].

Establishing primary cell lines from the xenografts

When xenograft tumours that had been passaged at least three times reached 10 mm, the mice were sacrificed, tumours excised and single cells extracted using enzymatic and mechanical techniques. Before seeding the isolated cells in regular cell culture flasks, mouse cells were substantially depleted using magnetic beads coated with a mouse MHC class I antibody (H2kD, Fig. S2). Remaining mouse cells were depleted by selective trypsin/EDTA treatment during the establishment of the primary cultures, since human pancreatic cells detach after 15–20 minutes, whereas mouse cells detach fully within 1–4 minutes. Within 3–6 passages the human cells (TRA-1-85 positive) in the cultures was close to 100% and mouse cells (H2kD positive) less than 2%.

The primary cell cultures were sensitive to low seeding densities, in particular in the first passages. Therefore, extracted cells were seeded at high concentration (20 000 cells/cm^2). To optimize culture viability, maintain the relevant phenotypes and restrain growth of mouse cells, we compared various medium conditions. Initially, we compared the capacity of the cells to grow in different media, like RPMI-1640 with 10% FCS, diverse stem cell media and media specifically designed for pancreatic cells. RPMI-1640 with FCS allowed attachment and growth of the pancreatic cells, however, this medium sustained fibroblastic mouse cells and

Table 3. IHC staining of original tumours and xenograft tumours.

Tumours	CK7	EpCAM	E-Cadherin	Snail/Slug[1]	Vimentin
Ppa1	4 (3)	4 (3)	4 (2)	0 (0)	0 (0)
PpaX1	4 (3)	4 (3)	4 (3)	4 (3)	1 (3)
Ppa2	4 (3)	4 (3)	3 (2)	0 (0)	0 (0)
PpaX2	4 (3)	4 (3)	4 (3)	4 (2)	4 (3)
Ppa6	4 (3)	4 (3)	4 (3)	0 (0)	0 (0)
PpaX6	4 (3)	4 (3)	4 (3)	4 (3)	1 (1)
Ppa8	4 (3)	4 (3)	4 (3)	0 (0)	0 (0)
PpaX8	4 (3)	4 (3)	4 (3)	0 (0)	1 (1)

Sections are from the following mouse passage: PpaX1 p4, PpaX2 p1, PpaX6 p7, PpaX8 p2.
% Positive cells are given according to the following system: 1 = 1–10%, 2 = 11–40%, 3 = 41–80%, and 4≥80%, and the staining intensities are given in brackets where (1 = weak, 2 = medium and 3 = strong).
[1]: Only cells with nuclear Snail+Slug expression is counted.
doi:10.1371/journal.pone.0103873.t003

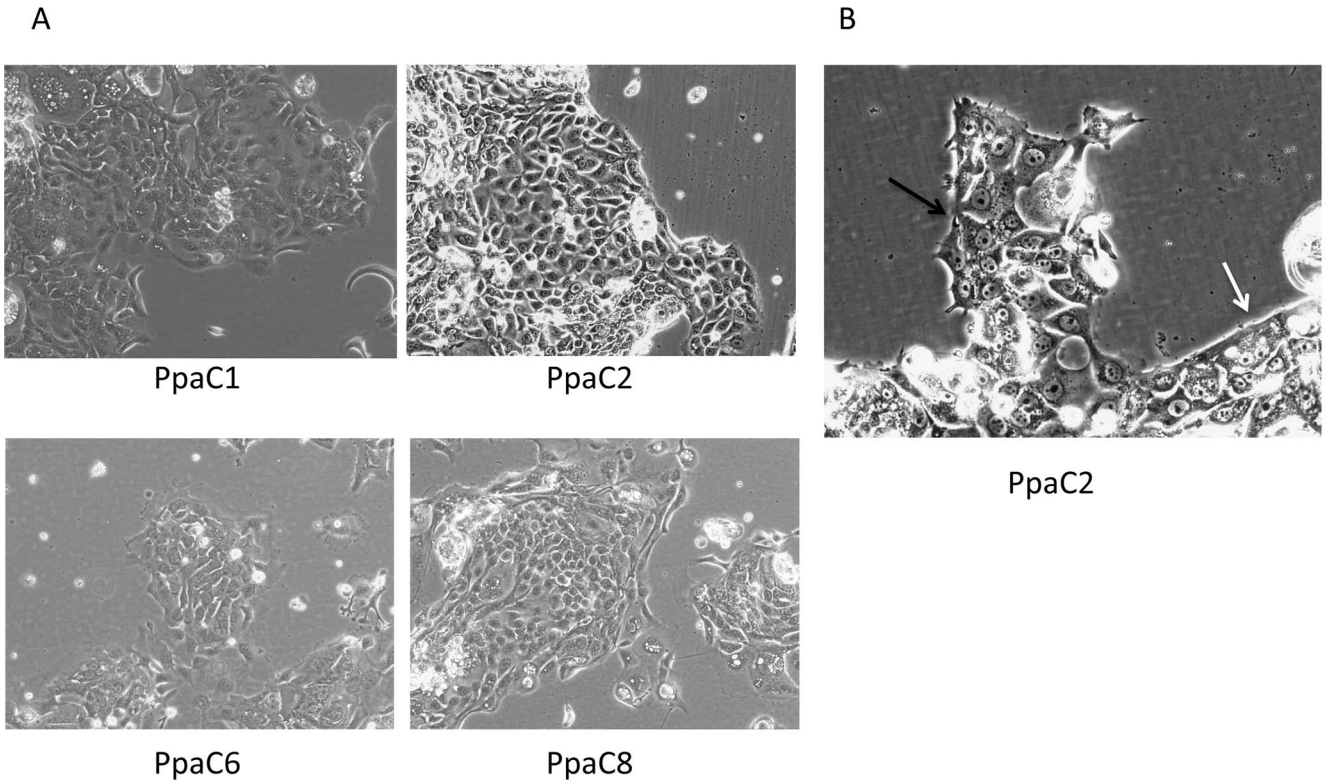


Figure 2. Global mRNA expression pattern. Heat map showing the hierarchical clustering of normal pancreatic tissue, original tumour material and the corresponding cell lines. The data set consists of 10 948 genes after filtering genes commonly regulated in fresh samples or cell lines. The RNA is isolated from the cell lines at the following passages: PpaC1 p16, PpaC6 p11 and PpaC8 p16.
doi:10.1371/journal.pone.0103873.g002

allowed them to overgrow the culture. All the general stem cell media and most of the pancreatic cell-specific media only allowed a few cells to adhere, and the cell growth was at best very poor. The serum-low Pancreatic Cell Culture Medium (= "pancreatic medium") allowed the pancreatic cells to attach and grow and at the same time limited the growth of mouse cells, and it was therefore chosen for further work.

Passaging of the cell lines was also carefully optimized. The cells tolerated neither Accutase nor TrypLE and EDTA alone did not detach the human cells within one hour. Trypsin/EDTA was tolerated when diluted 1:5 and detached the human cells within 15–20 minutes at 37°C. This was followed by one wash in PBS to remove the remaining trypsin. The resulting primary cell lines grew slowly and were initially passaged every 2–3 weeks. With time, the cells adjusted to growth *in vitro* and in some cases required passaging as often as once a week.

Extracted cells from four of the xenografts (PpaX1, PpaX2, PpaX6 and PpaX8) successfully generated primary cell lines that have been in culture for more than 10 passages *in vitro*, named PpaC1, PpaC2, PpaC6 and PpaC8, respectively. Initially, the cells grew as colonies with typical cobblestone morphology, characteristic for epithelial cells. The colonies were compact with a clear border around the colony (similar to the PpaC2 colony in Fig. 2A). After 3–4 weeks the colony borders weakened and cells grew out of the tight cobblestones in an irregular form and gained a more mesenchymal appearance. At the same time, single cells appeared that generated new colonies. In early passage cultures, the cobblestone morphology prevailed in the cultures, but over time the cultures gradually gained a mesenchymal phenotype with irregular shape growing in a disorganized pattern. This phenomenon was common for PpaC1, PpaC6 and PpaC8 (Fig. 2), however, PpaC2 kept the well-defined cobblestones for a longer time and only partly gained the mesenchymal-like growth appearance. Cultures with cobblestone morphology seemed to preferentially disseminate as cell clusters (Fig. 2B).

To test the ability of the generated cell lines to continuously grow *in vitro*, the PpaC1 culture was propagated for over 30 passages (more than 8 months) without any signs of growth decline; rather it tended to grow faster at higher passages. All four generated cell lines were fingerprinted and the fingerprints matched their corresponding original tumours (Table S1).

Transcriptome profiles of tumours and corresponding cell lines

To investigate whether the cell lines represent the parental tumours we compared the mRNA expression patterns of cells from PpaC1, PpaC6 and PpaC8 with those of the original tumours and also to non-tumourigenic pancreatic tissue from patients Ppa6 and Ppa8. To avoid differences resulting from contaminating normal and stromal cells or proliferation-associated mRNAs overexpressed in culture, we filtered the genes according to Virtanen *et al.* [21]. An unsupervised hierarchical clustering of the filtered mRNA expression patterns using Pearson and Spearman's rank correlation coefficient showed that the cell lines clustered with their corresponding tumour samples while the adjacent normal tissue samples formed a separate cluster (Fig. 3). This demonstrated that although different from the human tumours each cell line shared the overall expression pattern with their corresponding original tumour, supporting the usefulness of the cell lines as

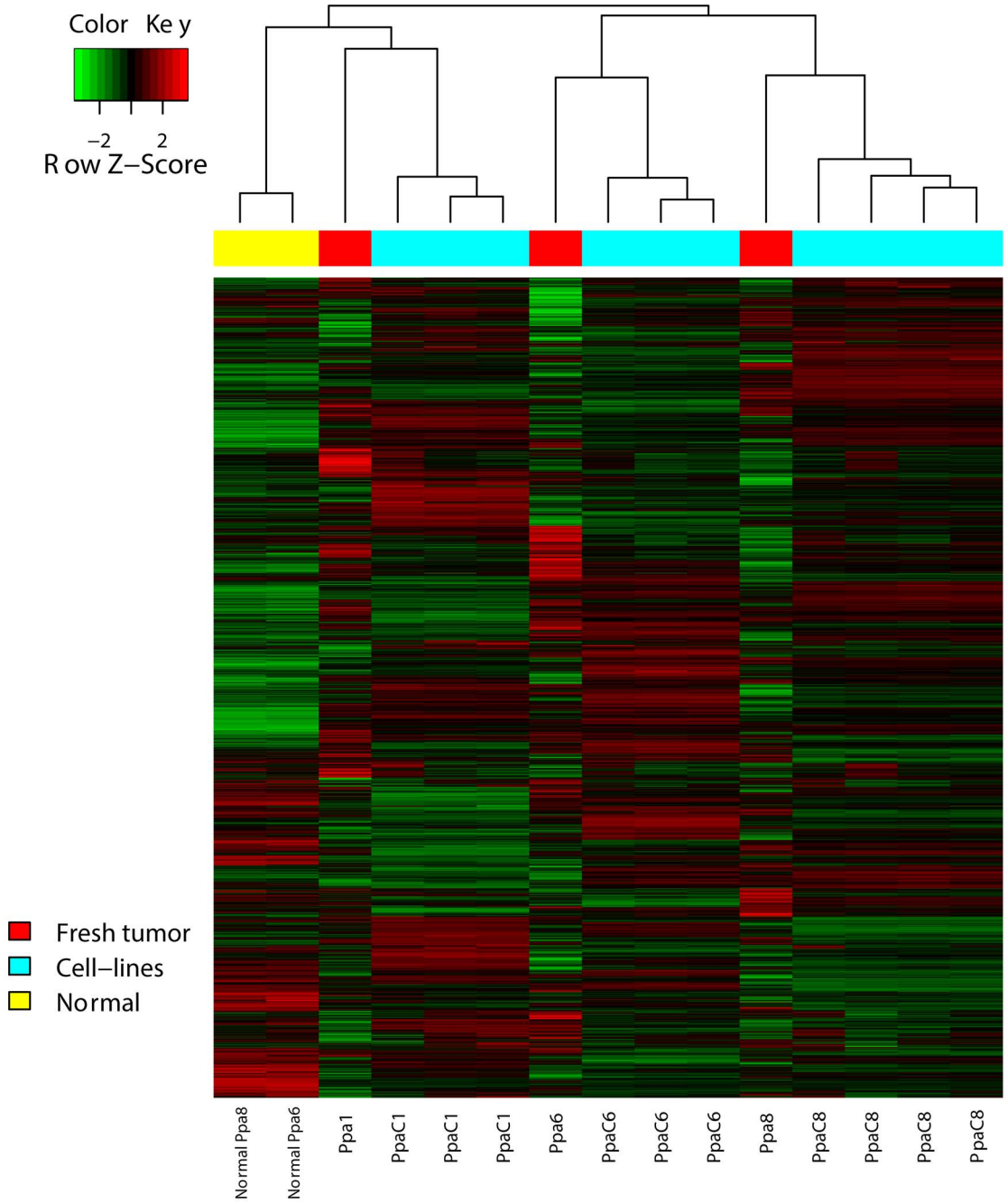


Figure 3. Morphology of the cell lines. A: Phase contrast pictures of the generated cell lines at the following passages: PpaC1 p2, PpaC2 p4, PpaC6 p8 and PpaC8 p10, at 10× magnification. B: Disseminating cell clusters in cultures with cobblestone growth pattern. The regular smooth colony border is indicated with a white arrow and the protruding group of cells with a black arrow. 20× magnification. Picture is from p4.
doi:10.1371/journal.pone.0103873.g003

models for their unique patient tumours, and also further confirm their origins from the correct patients.

Colony formation and *in vivo* tumourigenicity

The capacity to form colonies *in vitro* was investigated using anchorage-independent growth in semi-solid methylcellulose. We tested different media also for the colony assays, and despite the fact that Pancreatic medium gave the best growth under regular 2D growth conditions, the stem cell-enriching media gave the highest number of colonies (Table 4). The diameters of the colonies were however similar for all tested conditions. The number of colony-forming cells in the cultures was reproducible but differed markedly between the cell lines, from PpaC1 in which about one of three cells made colonies, to PpaC6 that did not form colonies at all (Fig. 4). The capacity to generate colonies remained stable over passages (Table 4) and it is noteworthy that the two cell lines that gave the highest number of colonies represent cancers with poor differentiation.

To confirm that the generated cell lines are tumourigenic, single cell suspensions from the established *in vitro* cultures were injected

Table 4. Colony forming capacity of PpaC1.

	in vitro passage							
Media	**p5**	**p6**	**p12**	**p18**	**p20**	**p24**	**p28**	**p31**
Stem Cell medium I	49%	37%	52%	37%	n.d.	42%	35%	n.d.
Stem Cell medium II	22%	22%	39%	41%	56%	37%	30%	41%
RPMI/10%FCS	18%	7%	24%	n.d.	n.d.	n.d.	n.d.	n.d.
Pancreatic medium	10%	n.d.	17%	n.d.	n.d.	n.d.	n.d.	n.d.

The table shows the % of cells able to generate colonies in the indicated semi-solid media in different *in vitro* passages, n.d. = not done.

doi:10.1371/journal.pone.0103873.t004

into the flank of NSG mice without the use of Matrigel. All four cell lines were able to form palpable tumours within three months from as few as 10 000 cells (Table 5). There were no signs of tumours in the 100 or 1000 cell injections for any of the cell lines after six months, which is in agreement with previous reports [6,7,22]. Importantly, despite the inability of the PpaC6 cell line to form colonies, it generated tumours in mice as efficiently as the other cell lines. It is however noteworthy that PpaC6 is derived from a tumour of the intestinal subtype while the remaining cell lines were from tumours of the pancreatobiliary subtype.

Characterization of stem cells markers

We did not perform clonal selection during generation of the cell lines in order to better conserve the chaotic and multi-mutational cellular hierarchy that exist in human tumours. Accordingly, all the established cell lines had heterogeneous morphologies, which were also reflected in the cell surface expression of known pancreatic and cancer stem cell markers (Table 6). All four cell lines had a high percentage of cells expressing the reported pancreatic cancer stem cell markers CD24, CD44 and CD326 (EpCAM) [6]. The very high number of CD326 positive cells, above 98%, confirms that the cell lines are of epithelial origin, and thus not significantly contaminated by fibroblasts, and also that even the PpaC2 cells had not gone completely through EMT. The expression of CD24 and CD44 varied both between samples and over time, but in general both markers were expressed in more than 50% of the cells in early passages. PpaC1 cells were sorted based on CD24 and CD44 expression and the colony formation capacity as well as the capacity to regenerate all subpopulations upon cultivation was examined. There was no enrichment in the colony forming capacity in the $CD24^+/CD44^+/CD326^+$ sorted cells (Fig. S3). Furthermore, when we grew the sorted cells for 2–3 passages and reanalysed the CD24 and CD44 expression, each sorted population was able to regenerate all the four subtypes (Fig. S4).

To investigate whether certain phenotypes were being selected for or induced during *in vitro* propagation, the expression of some markers was followed during *in vitro* passages for PpaC1. Over time, the percentage of $CD44^+$ cells increased and after 13 passages almost all cells had a strong cell surface expression of CD44 (Table S2). We have previously observed this change toward an almost uniform CD44+ cell population also in primary lung cancer cell lines upon passaging [23], but do not know whether this is a selection for growth *in vitro* or an aspect of *in vitro* growth.

Altogether, we could neither find any enrichment of cells with cancer stem cell characteristics based on CD24/CD44/CD326 expression, nor did we observe any hierarchical system for these markers in our cell lines. This is in agreement with our observation that the percentage of cells being able to generate spheroids in semi-solid medium was stable during passages (Table 4), although the percentages of CD44 and CD24 positive cells change over time.

Other markers for cancer stem cells in pancreatic adenocarcinoma have also been suggested, either alone or in combination with one or several of the classical CD24/CD44/CD326 markers. One of these is CD133 that has been investigated in several different cancer types, such as liposarcoma [24], glioblastoma and colorectal carcinoma [25]. In pancreatic adenocarcinoma Hermann *et al.* [7] used CD133 in combination with CD184 to identify migratory cells essential for tumour metastasis.

We investigated the expression of two epitopes of CD133, and both CD133/1 and CD133/2 positive cells could be detected in all the four cell lines with CD133/2 being more abundant.

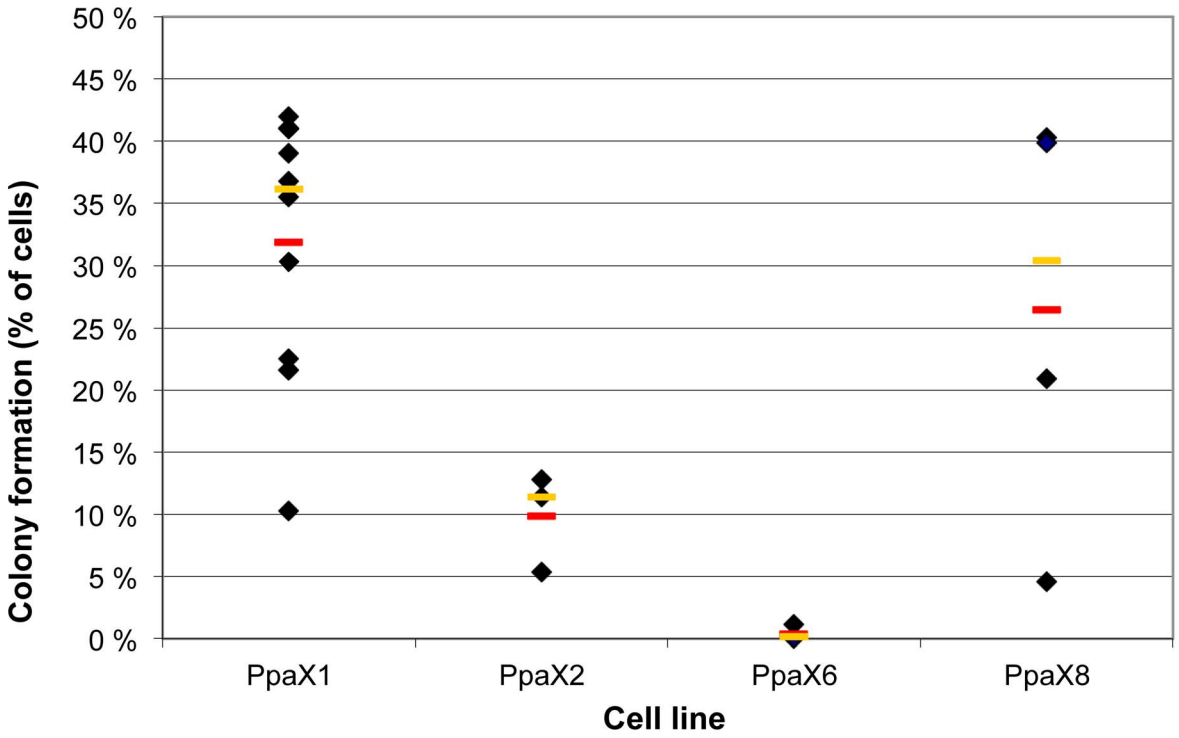


Figure 4. Colony forming capacity of the cell lines. Each diamond represent the fraction of cells able to generate colonies >50 μm in methylcellulose/stem cell medium II in one experiment (n = 3–9). The red line shows the average and the yellow line shows the median colony forming capacity for cells in the following passage span: PpaC1 p1–p31, PpaC2 p5–p13, PpaC6 p2–p7 and PpaC8 p4–p14.
doi:10.1371/journal.pone.0103873.g004

However, there was considerable variation between the cell lines; from 0.6 to 45% for CD133/1, and from 5.6 to 47% for CD133/2 (Table 6), and over time, the frequency decreased during *in vitro* passaging of PpaC1 (Table S2).

In all cell lines there was a distinct population of $CD184^+$ cells (Table 6). This is contrary to Jaiswal *et al.* [26] who only found $CD184^+$ cells (0.21%) in one of five established cell lines. As for CD133, the percentage of $CD184^+$ cells dropped with passaging, suggesting that these markers are not useful *in vitro*.

CD166 is reported to be a cancer stem cell marker in several solid cancers, including pancreatic cancer, although there are conflicting reports regarding its usefulness as a prognostic marker in pancreatic adenocarcinoma. While Khalert *et al.* [27] found that high expression of CD166 in patient samples correlated with poor survival and early tumour relapse, Tachezy *et al.* [28] found no correlation between CD166 expression and patient survival although they did find a higher expression in pancreatic cancer compared to normal tissue. Our cell lines show a high fraction of $CD166^+$ cells, in the range of 57–93% (Table 6), which is in accordance with an aggressive phenotype. Furthermore, the $CD166^+$ population was stable over time (Table S2).

Another cancer-associated protein is CD142, also known as tissue factor (TF). Although the normal function of CD142 is to initiate the coagulation cascade, it is often highly expressed in cancer tissue [29] as well as in plasma of cancer patients where it can cause thrombosis, a phenomenon known as Trousseau syndrome [30]. When investigated by flow cytometry, the pancreatic cell lines had high fractions (78–97%) of cells with membrane-bound CD142, stable over time (Tables 5 and S2). Important, also the soluble form is reported to be of importance for cancer progression [31].

Both the Side Population (SP) assay, that measures drug efflux capacity, and the Aldeflour assay, that measures the activity of aldehyde dehydrogenase, are widely used to identify or enrich for

Table 5. *In vivo* tumourigenicity.

Cells injected	PpaC1	PpaC2	PpaC6	PpaC8
250 000	4/4	5/5	-	4/4
100 000	4/4	5/5	6/6	4/4
10 000	4/4	5/5	6/6	2/4
1 000	0/4	0/5	0/6	0/4
100	0/4	-	0/6	0/4

Shown is the number of tumours growing/the number of injected sites in NSG mice for each cell lines, harvested at the following passages: PpaC1 p6, PpaC2 p12, PpaC6 p6, PpaC8 p13.
doi:10.1371/journal.pone.0103873.t005

Table 6. Cell surface expression of known cancer and stem cell markers.

Marker	PpaC1	PpaC2	PpaC6	PpaC8
TRA-1-85 +	96	98	98	98
H2kd +	1.7	0.9	1.4	0.1
CD24	62	35	97	70
CD44	64	64	86	83
CD133/1	2.0	45	7	0.6
CD133/2	7.7	47	20	5.6
CD142	78	87	97	94
CD166	57	93	87	69
CD326/EpCAM	99	94	100	97
CD184	2.1	2.2	1.1	2.4
SSEA-4	31	6.0	48	55
SSEA-1/CD15	51	48	80	53
Aldeflour	3.9	12	13	2.8
Side population	4.8	n.d.	n.d	5.8

Shown here is the average % positive cells out of single, live cells, n = at least 3, n.d = not done.
Cells were analysed in the following passage spans: PpaC1 p1–p10, PpaC2 p3–p14, PpaC6 p5–p9 and PpaC8 p4–p20.
doi:10.1371/journal.pone.0103873.t006

cells with stem cell-like properties in various cancers, including pancreatic. The PpaC1 and PpaC8 contained SP^+ subpopulations of similar size (4.8 and 5.8%, respectively) (Table 6), in both cases validated by sensitivity to pre-treatment with the membrane pump inhibitor FTC and also partially with the inhibitors verapamil and reserpine (data not shown). These high percentages of SP^+ cells are in contrast to the findings of Jaiswal *et al*. [26], who reported less than 1% SP^+ cells in all their established pancreatic cell lines. This can be due to the long time in culture, but since the side population assay is very sensitive to the metabolic conditions of the cells, this could also be due to small technical differences [32].

The cell lines originating from tumours of moderate differentiation (PpaC2 and PpaC6) had high fractions of $Aldeflour^+$ cells (12 and 13% respectively) in contrast to those originating from poorly differentiated tumours (PpaX1 and PpaX8, 2.8 and 3.9% respectively) (Table 6). These numbers are similar to those reported by Kim *et al*. [9] in untreated freshly resected patient material (12.1% and 13.7%), and in the same range as those reported by Rasheed *et al*. [22,33] for four established pancreatic cell lines (2.4–8.5%), although they reported a lower fraction in a primary pancreatic xenograft (1.7%).

Altogether, we describe a robust procedure for the generation of pancreatic adenocarcinoma cell lines, and a detailed characterisation of four new PDAC xenograft and cell lines. These cell lines are chemotherapy naïve, represent their tumours of origin well, and together with their corresponding xenografts, they represent highly relevant preclinical models for pancreatic cancer.

Supporting Information

Figure S1 Schematic figure of the overall working flow. Implanted tumours were passaged at least two times before cells were extracted to generate *in vitro* cell lines. The cell lines have now been passaged up to 32 times *in vitro*. The cell line passage number used for each cell line in each analysis is indicated in the figure legend of the relevant analysis in the main article.
(PDF)

Figure S2 Depletion of mouse cells in xenograft single cell suspensions. Cells were stained with antibodies against the human marker TRA1-85 and the mouse marker H-2Kd before and after depletion of mouse cells in the xenograft single cell suspensions. The percentages of human and mouse cells of live, single cells are indicated in the flow cytometry dot plot diagrams before and after depletion.
(PDF)

Figure S3 Colony forming capacity in isolated CD24/CD44 cell populations. CD24/CD44 subpopulations from PPaC1 cells isolated by flow cytometry assisted cell sorting were grown in methylcellulose/stem cell medium II to evaluate their colony forming abilities. The number of colonies (>50 μm)/1000 seeded cells after two weeks are shown.
(PDF)

Figure S4 Regeneration of CD24/CD44 populations from cultivated CD24/CD44 cell populations. CD24/CD44 subpopulations from PPaC1 cells isolated by flow cytometry assisted cell sorting were cultured under regular growth conditions for 2–3 passages and reanalysed for the expression of CD24 and CD44. The dot plot to the left shows the original sorting gates, and the four dot plots to the right are the cultivated isolated populations using the color-coding from the original sorted cells. In all dot plots, the Y-axis represent the CD24 expression while the X axis represent the CD44 expression.
(PDF)

Table S1 STR fingerprinting of human tumours and the generated cell lines.
(PDF)

Table S2 Cell surface marker expression during passaging of PpaC1.
(PDF)

Acknowledgments

We thank Dr. Knut Jørgen Labori and Dr. Trond Buanes for collection of tumour samples, Ellen Hellesylt for immunohistochemical staining's, Petros

Gebregziabher for animal work, Nomdo AC Westerdaal for technical assistance with flow cytometry, the genotyping core fasility at OUS for DNA fingerprinting analysis and Astrid Dalsgaard for performing the microarray analysis.

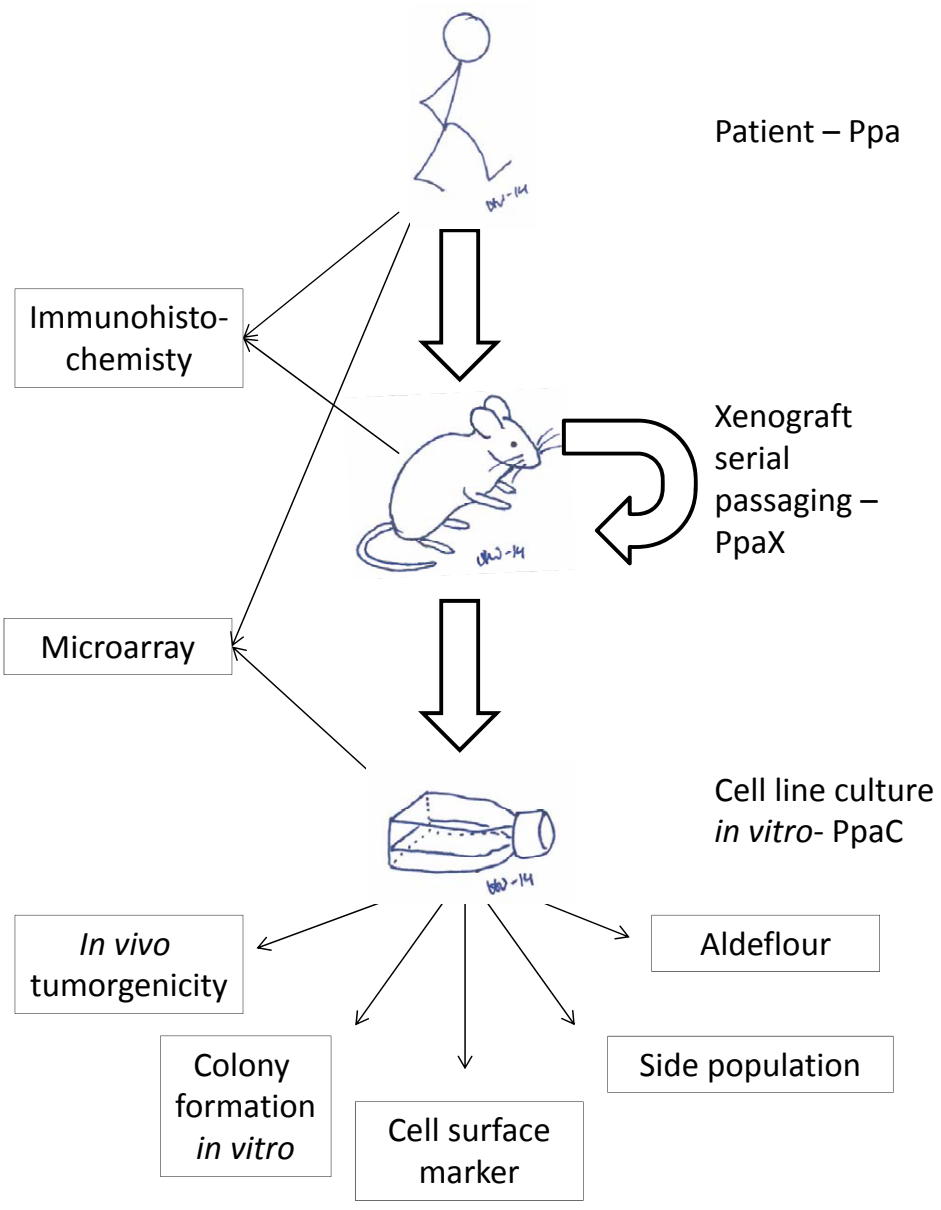
Author Contributions

Conceived and designed the experiments: ABW IMBL EHK OM EM. Performed the experiments: ABW IMBL EM. Analyzed the data: ABW IMBL VS EM. Contributed to the writing of the manuscript: ABW IMBL VS EHK OM EM.

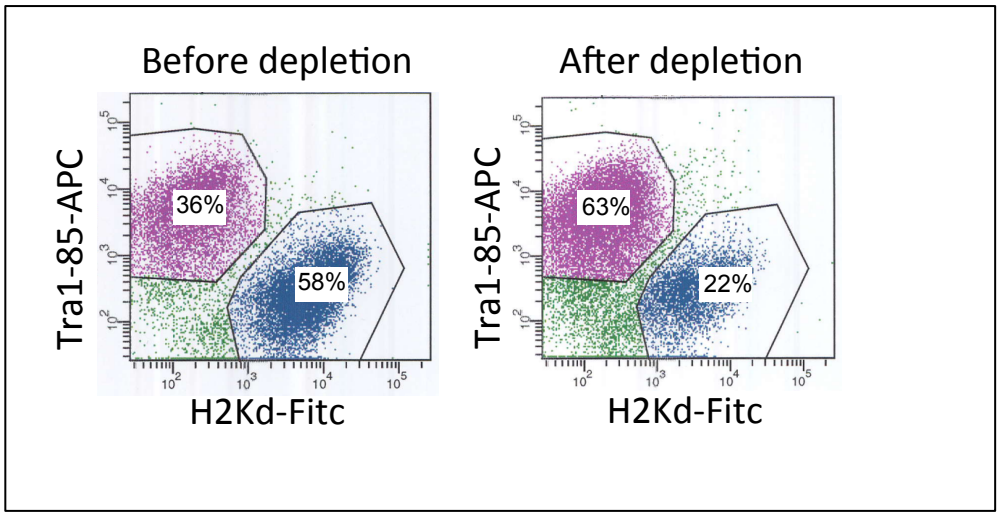
References

1. Yachida S, Jones S, Bozic I, Antal T, Leary R, et al. (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467: 1114–1117.
2. Albores-Saavedra J SK, Dancer YJ, Hruban R (2007) Intestinal type adenocarcinoma: a previously unrecognized histologic variant of ductal carcinoma of the pancreas. Annals of diagnostic pathology 11: 3–9.
3. Westgaard A TS, Farstad IN, Cvancarova M, Eide TJ, Mathisen O, et al. (2008) Pancreatobiliary versus intestinal histologic type of differentiation is an independent prognostic factor in resected periampullary adenocarcinoma. BMC Cancer 8.
4. Wasif N, Ko CY, Farrell J, Wainberg Z, Hines OJ, et al. (2010) Impact of tumor grade on prognosis in pancreatic cancer: should we include grade in AJCC staging? Ann Surg Oncol 17: 2312–2320.
5. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 100: 3983–3988.
6. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, et al. (2007) Identification of pancreatic cancer stem cells. Cancer Res 67: 1030–1037.
7. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, et al. (2007) Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell 1: 313–323.
8. Wang YH, Li F, Luo B, Wang XH, Sun HC, et al. (2009) A side population of cells from a human pancreatic carcinoma cell line harbors cancer stem cell characteristics. Neoplasma 56: 371–378.
9. Kim MP, Fleming JB, Wang H, Abbruzzese JL, Choi W, et al. (2011) ALDH activity selectively defines an enhanced tumor-initiating cell population relative to CD133 expression in human pancreatic adenocarcinoma. PLoS one 6: e20636.
10. Deer EL, Gonzalez-Hernandez J, Coursen JD, Shea JE, Ngatia J, et al. (2010) Phenotype and genotype of pancreatic cancer cell lines. Pancreas 39: 425–435.
11. Hamfjord J, Stangeland AM, Skrede ML, Tveit KM, Ikdahl T, et al. (2011) Wobble-enhanced ARMS method for detection of KRAS and BRAF mutations. Diagn Mol Pathol 20: 158–165.
12. Bosman FT, Carneiro F, Hruban RH, Theise ND (2010) WHO Classification of Tumours of the Digestive System (IARC WHO Classification of Tumours). Fourth ed: World Health Organization.
13. Kimura W, Futakawa N, Yamagata S, Wada Y, Kuroda A, et al. (1994) Different clinicopathologic findings in two histologic types of carcinoma of papilla of Vater. Jpn J Cancer Res 85: 161–166.
14. Sobin LH, Gospodarowicz MK, Wittekind C (2009) TNM Classification of Malignant Tumours. 7th ed: Wiley-Blackwell.
15. Flatmark K, Maelandsmo GM, Mikalsen SO, Nustad K, Varaas T, et al. (2004) Immunofluorometric assay for the metastasis-related protein S100A4: release of S100A4 from normal blood cells prohibits the use of S100A4 as a tumor marker in plasma and serum. Tumour Biol 25: 31–40.
16. Geer RJ, Brennan MF (1993) Prognostic indicators for survival after resection of pancreatic adenocarcinoma. Am J Surg 165: 68–72; discussion 72–63.
17. Kim MP, Evans DB, Wang H, Abbruzzese JL, Fleming JB, et al. (2009) Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. Nat Protoc 4: 1670–1680.
18. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, et al. (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 133: 704–715.
19. Yamada S, Fuchs BC, Fujii T, Shimoyama Y, Sugimoto H, et al. (2013) Epithelial-to-mesenchymal transition predicts prognosis of pancreatic cancer. Surgery 154: 946–954.
20. Hotz B, Arndt M, Dullat S, Bhargava S, Buhr HJ, et al. (2007) Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer. Clin Cancer Res 13: 4769–4776.
21. Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, et al. (2002) Integrated classification of lung tumors and cell lines by expression profiling. Proc Natl Acad Sci U S A 99: 12357–12362.
22. Rasheed ZA, Yang J, Wang Q, Kowalski J, Freed I, et al. (2010) Prognostic significance of tumorigenic cells with mesenchymal features in pancreatic adenocarcinoma. J Natl Cancer Inst 102: 340–351.
23. Wang P, Gao Q, Suo Z, Munthe E, Solberg S, et al. (2013) Identification and characterization of cells with cancer stem cell properties in human primary lung cancer cell lines. PLoS one 8: e57020.
24. Stratford EW, Castro R, Wennerstrom A, Holm R, Munthe E, et al. (2011) Liposarcoma Cells with Aldefluor and CD133 Activity have a Cancer Stem Cell Potential. Clin Sarcoma Res 1: 8.
25. Grosse-Gehling PFC, Dittfeld C, Garbe Y, Alison MR, Corbeil D, et al. (2013) CD133 as a biomarker for putative cancer stem cells in solid tumours: limitations, problems and challenges. The Journal of Pathology 229: 355–378.
26. Jaiswal KR, Xin HW, Anderson A, Wiegand G, Kim B, et al. (2012) Comparative testing of various pancreatic cancer stem cells results in a novel class of pancreatic-cancer-initiating cells. Stem Cell Res 9: 249–260.
27. Kahlert C, Weber H, Mogler C, Bergmann F, Schirmacher P, et al. (2009) Increased expression of ALCAM/CD166 in pancreatic cancer is an independent prognostic marker for poor survival and early tumour relapse. Br J Cancer 101: 457–464.
28. Tachezy M, Zander H, Marx AH, Stahl PR, Gebauer F, et al. (2012) ALCAM (CD166) expression and serum levels in pancreatic cancer. PLoS one 7: e39018.
29. Khorana AA, Ahrendt SA, Ryan CK, Francis CW, Hruban RH, et al. (2007) Tissue factor expression, angiogenesis, and thrombosis in pancreatic cancer. Clin Cancer Res 13: 2870–2875.
30. van den Berg YW OS, Reitsma PH, Versteeg HH (2012) The relationship between tissue factor and cancer progression:. Blood 119: 924–932.
31. Kasthuri RS, Taubman MB, Mackman N (2009) Role of tissue factor in cancer. J Clin Oncol 27: 4834–4838.
32. Golebiewska ABN, Bjerkvig R, Niclou SP (2011) Critical appraisal of the side population assay in stem cell and cancer stem cell research. Cell Stem Cell 8: 136–147.
33. Rasheed ZAYJ, Wang Q, Kowalski J, Freed I, Murter C, et al. (2010) Prognostic significance of tumorigenic cells with mesenchymal features in pancreatic adenocarcinoma. Journal of the National Cancer Institute 102: 340–351.

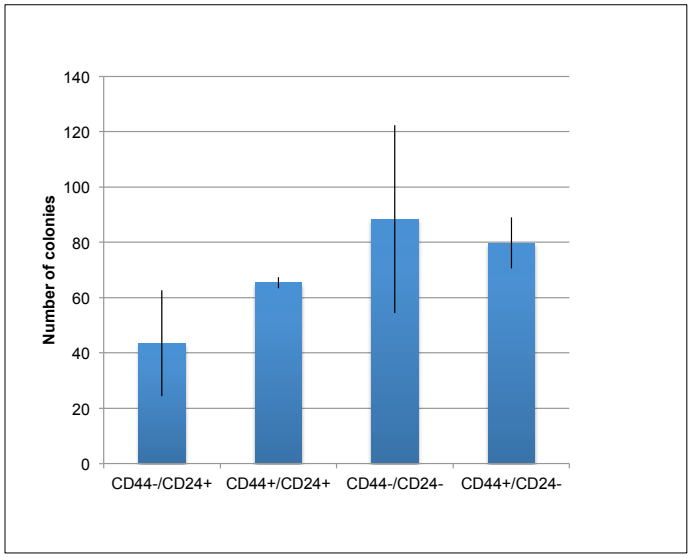
Supplementary Figure S1



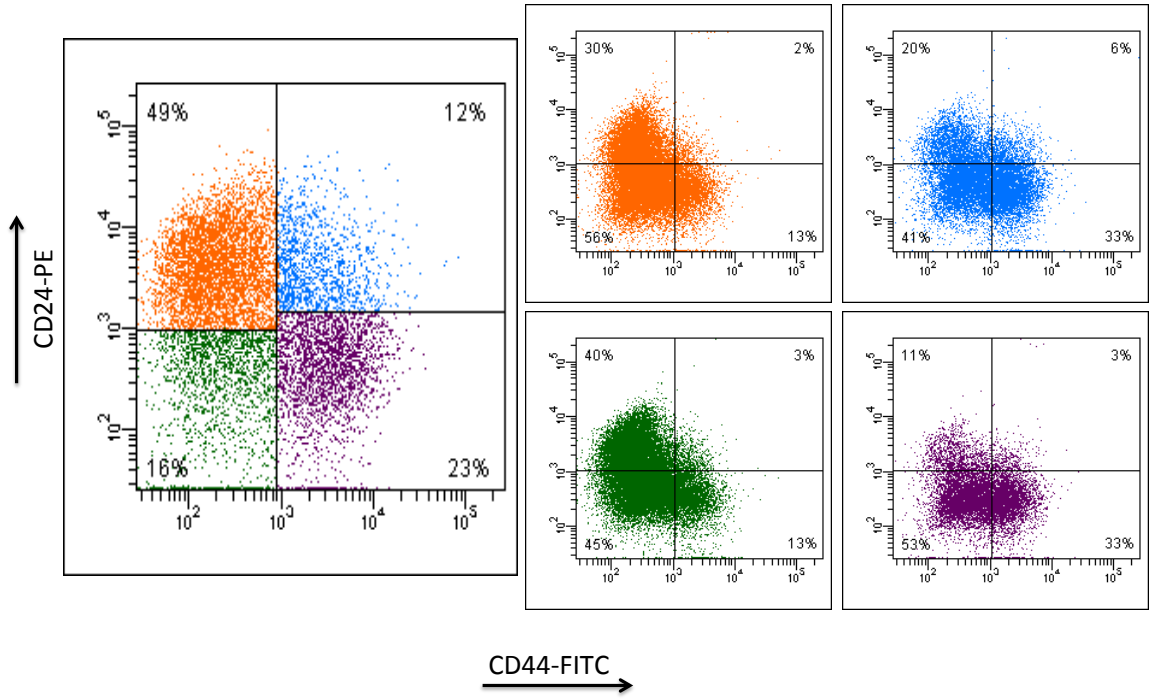
Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4



Supplementary Table S1

Sample	Amelogenin	CSF1PO	D13S317	D16S539	D18S51	D21S11	D3S1359	D5S818	D7S820	D8S1179	FGA	Penta D	Penta E	TH01	TPOX	vWA
Ppa1	x	11	10	6,10,11	14	27.2, 32	16, 17, 21	11	11	12, 14	20, 22	12, 13	7, 12	6	8, 11	16, 18
PpaC1	x	11	10	11	14	28; 32,2	16, 17	11	11	12, 14	20, 22	12,13	7, 12	6	8, 11	18
Ppa2	x, y	10, 13	11, 13	7,11,12	12, 15	27.2, 28.2	14, 17	11, 13	8, 9	11, 13	21, 23	12, 13	10, 20	6, 9.3	8, 11	14, 15
PpaC2	x, y	10, 13	11, 13	12	12	28, 29, 30	14	13	8, 9	11, 13	21,23	12, 13	10, 20	6	8, 11	12, 15
Ppa3	x, y	11, 12	11, 13	13	14, 15	29.2, 30.2	16, 17	11, 12	11	14, 15	18, 21	7, 12	11, 15	9, 9.3	8	15, 16
Ppa4	x	11, 12	8, 9	9, 13	12, 16	27.2, 29.2	16	11	8, 11	14	20, 24	9, 11	11, 14	9, 9.3	8, 11	15, 17
Ppa6	x	11, 12	9, 12	9, 12	14, 16	28.2, 29.2	16	10, 12	8, 10	13	20, 26.2	8, 13	15	4, 7	8, 9	15, 18
PpaC6	x	11, 12	9, 12	9,11,12	14	29	16	10, 12	8, 10	13	20, 26	7, 13	15	7	8, 9	15, 18
Ppa7	x	10, 11	14	6, 11	13, 19	28.2, 29.2	17, 18	11, 13	9, 10	10, 14	20, 21, 22	10, 12	7, 18	9.3	8	16, 18
Ppa8	x	11, 13	9, 12	9, 12	16, 19	28.2	14, 18	11, 13	8, 9	13	21, 23	10, 12	7, 10	4, 7	8	17
PpaC8	x	11, 13	9, 12	9, 12	16	29	14, 18	11, 13	8, 9	13	21, 23	10, 12	7, 10	7	8	17
Ppa9	x, y	11, 12	9, 11	7, 12	13, 15	28.2, 30.2	14, 18	11, 12	10, 13	12, 14	20, 22	9, 12	14, 16	6, 9.3	8, 10	18, 19

Supplementary Table S2

Supplementary Table 2. Phenotyping of PPaC1 in different passages										
	p1	p2	p4	p7	p8	p10	p13	p18	p24	p32
H2kd +	3.6		1.4		0.6	0.9		0.1		0.1
CD24			56		78	29		54	58	21
CD44	44		54		95	91	98	98	100	99
CD133/1	3.3		2.2			0.4		0.1	0.1	0.1
CD133/2	16		6.1			0.9		1.7	1.1	1.2
CD142	93							69	98	78
CD166	46							63	61	52
CD326/ EpCam			99			99		99	100	99
CD184			3.7			1.2		1.6	1.8	0.8
SSEA-4					27	11	21	9.9	30	6
CD15/ SSEA-1					60	54	94	75	91	30
Aldeflour		2.7		2.5			3.2			
Side population		5.7		3.9			2.2			

Doctoral dissertation no. 2
2016

—

A systems biology approach to integrated molecular analysis in pancreatic and periampullary adenocarcinoma

A PhD-dissertation in Ecology

—

Vandana Sandhu

—

ISSN: 2464-2770
ISBN: 978-82-7206-407-4

usn.no