Molecular characterization of early precursors of pancreatic ductal adenocarcinoma

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Lisa Lotte Frohn
aus Duisburg

Duisburg, Oktober 2021

aus dem Institut für Pathologie
des Universitätsklinikum Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof Dr. Irene Esposito
Universitätsklinikum Düsseldorf
Institut für Pathologie

2. Prof Dr. Wiliam F. Martin
Heinrich-Heine Universität Düsseldorf
Institut für Molekulare Evolution

Tag der mündlichen Prüfung:
25.03.2022

Für meinen Papa

Table of contents

Summary 1

Zusammenfassung 3

Abbreviations 5

1 Introduction 7

1.1 Pancreatic ductal adenocarcinoma 7

1.1.1 Epidemiology, survival and risk factors 8

1.1.2 Diagnosis and therapy 10

1.2 Precursors of the PDAC 12

1.2.1 Pancreatic intraepithelial neoplasia 12

1.2.2 Atypical flat lesion 13

1.2.3 Intraductal papillary mucinous neoplasms 14

1.2.3.1 Gastric IPMN 15

1.2.3.2 Intestinal IPMN 15

1.2.3.3 Pancreatobiliary IPMN 16

1.2.4 Intraductal oncocytic papillary neoplasms 16

1.2.5 Mucinous cystic neoplasm 17

1.3 Genetic alterations involved in PDAC progression 18

1.3.1 The stepwise process from normal tissue over precursor lesions to PDAC 18

1.3.2 Chromothripsis 22

1.3.3 DNA Methylation in cancer 24

2 Aim 27

3 Materials and Methods 28

3.1 Equipment 28

3.2 Kits and enzymes 30

3.3 Chemicals/ Reagents 32

3.4 Buffer 34

3.5 Primer 35

3.6 Antibodies 36

3.7 Software and databases 37
3.8 Patient cohort 39
3.9 Histology Methods 40
3.9.1 Tissue microarray cohort 40
3.9.2 Hematoxylin and Eosin (H&E) Staining 41
3.9.3 Crystal violet staining 41
3.9.4 Immunohistochemistry 42
3.9.5 Microdissection (manual) 43
3.9.6 Microdissection with Palm Microbeam 43
3.10 Molecular biology methods 45
3.10.1 Targeted Gene Sequencing by Next-Generation-Sequencing (NGS) 45
3.10.1.1 DNA Isolation from FFPE tissue 45
3.10.1.2 Quality control of genomic DNA from FFPE 45
3.10.1.3 Amplicon-based library preparation 46
3.10.2 Low-coverage whole-genome sequencing 49
3.10.2.1 Single-cell suspension from paraffin blocks 49
3.10.2.2 DEPArray™ 50
3.10.2.3 Whole-genome amplification 51
3.10.2.4 Low-pass library preparation 54
3.10.2.5 Low-pass sequencing 55
3.10.3 Whole transcriptome sequencing by NGS 56
3.10.3.1 RNA Isolation from FFPE tissue 56
3.10.3.2 RNA quantification 57
3.10.3.3 RNASeq 57
3.10.4 Genome-wide DNA Methylation 58
3.10.4.1 Illumina Infinium MethylationEpic Array 58
3.10.5 FISH 58
3.10.6 Bioinformatic methods 59
3.10.6.1 NGS analysis 60
3.10.6.2 Analysis of copy number variations 60

3.10.6.3 Analysis pipeline RNA-Seq Data 61
3.10.6.4 Methylome 61
3.10.6.5 Pathway Analysis 61
3.10.6.6 Statistical analysis 62
4 Results 63
4.1 Mutation profile of precursor lesions 63
4.2 Detection of copy number variations 67
4.3 Proliferation rate 73
4.4 Status of the *GNAS* gene 74
4.5 Transcriptome Data 76
4.6 DNA Methylation pattern 82
4.7 Integrative analysis: gene expression and methylation 87
4.7.1 O-glycan Mucin biosynthesis 94
5 Discussion 95
5.1 Characterization of mutation patterns of precursor lesions 95
5.1.1 Impact of *GNAS* 97
5.2 The potential to progress from a precursor lesion to PDAC 99
5.2.1 Cell of origin 102
5.2.2 Gastric IPMN versus intestinal IPMN 103
5.3 Investigation of new biomarkers 105
5.3.1 Krüppel-like factor 4 105
5.3.2 Mucin 6 (*MUC6*) and Trefoil factor 1 (*TFF1*) 106
5.4 Conclusion 109
6 References 111
7 List of Tables 128
8 List of Figures 129
9 Communications 131
10 Eidesstattliche Versicherung 132

11 Attachment 133
11.1 Self-designed NGS Panel #1 133
11.2 Self-designed NGS Panel #2 139
11.3 Enrichment Analysis 145
11.4 Integrative Analysis gene list 146

Summary

Pancreatic ductal adenocarcinoma (PDAC) is an aggressive disease with high mortality rates due to late diagnosis and rapid progression. Therefore, better tools for earlier detection are urgently needed. The most common PDAC precursor lesions are pancreatic intraepithelial neoplasias (PanIN) and intraductal papillary mucinous neoplasms (IPMN). The gastric subtype of IPMN shares morphological homologies with PanIN and the two lesions are currently distinguished only according to their size. However, their outcome differs highly so that stringent criteria for proper differentiation are mandatory. This study aims at characterizing the molecular features of these two PDAC precursors for a better understanding of their biological behavior.

120 cases of resected IPMN and/or PanIN were retrospectively evaluated and reclassified according to morphologic and immunohistochemical criteria. In total, 59 different precursors were analyzed via next-generation sequencing with a customer-designed gene panel with 22 cancer-related genes. For copy number variation (CNV) detection, low-coverage whole-genome sequencing was done with formalin fixed paraffin embedded (FFPE) material from 47 cases. For transcription data, RNASeq was performed with 7 different samples (3 gastric IPMN and 4 PanIN) and 7 controls (3 acinar bulk tissue and 4 PDAC samples). For methylation status, the Illumina Infinium MethylationEPIC Array was performed in 49 samples (including 18 gastric IPMN, 8 PanIN samples, 10 intestinal IPMN samples, 7 acinar bulk tissue, 3 beta cells samples and 3 duct cells samples).

Targeted NGS confirmed the presence of the two driver mutations in the *KRAS* and *GNAS* genes in gastric and intestinal IPMN. In the copy number profile, the PanIN lesions did not display recurrent copy numbers in opposition to gastric IPMN, which had CNV in defined regions. Intestinal IPMN showed the most aberrant copy number profile, suggesting a higher transformational potential.

The methylation profile indicated a high similarity between PanIN and gastric IPMN, where only 0.36 % of the 5'-Cytosine-phosphate-Guanine-3' sites (CpGs) were differently methylated in contrast to the completely different methylation profile of the intestinal IPMN.

Based on the methylation analysis, the cell of origin of gastric IPMN and PanIN seems to be derived from the acinar compartment, whereas the cell of origin of intestinal IPMN is derived from the ductal compartment.

The transcriptome analysis was performed to compare gastric IPMN and PanIN and showed that 595 significantly different expressed genes (DEGs) were present in the two lesions (p-value < 0.01). By combining the significantly different expressed and methylated genes in an integrative analysis between PanIN and gastric IPMN, 3 different hub genes were chosen for further investigation, including the genes *MUC6*, *TFF1* and *KLF4*. The prospect marker genes were tested by immunochemistry in a cohort of 71 patients on tissue-microarray slides. The expression pattern on the protein level did not reflect the mRNA level and these markers were not suitable for detection or to distinguish the different precursors from each other. The higher expression of these genes, the copy number profile and the different activation of the downstream pathways support the concept of higher neoplastic potential of gastric IPMN compared to PanIN.

In conclusion, gastric IPMN and PanIN have a very similar morphology, genetics and methylation profile but differ in the CNV and transcriptome profile. Our data indicate a common origin with the acquisition of a higher neoplastic potential by gastric IPMN. Intestinal IPMN are distinct entities with a clear-cut neoplastic potential already in low grade lesions with completely different methylation and copy number profile.

Our study shows that the genomic based data can separate the different subtypes of PDAC precursors in harmony with the immunophenotypical profile. As long as no validated markers are found, the immunophenotypical subtyping is fundamental in identifying precursors with different risks of progression and should be included as core element in pathology reports.

Zusammenfassung

Das duktale Adenokarzinom der Bauchspeicheldrüse (PDAC) ist eine aggressive Erkrankung mit einer hohen Sterblichkeitsrate aufgrund der späten Diagnose und des schnellen Fortschreitens. Daher werden dringend bessere Verfahren zur Früherkennung benötigt. Die häufigsten PDAC-Vorläuferläsionen sind pankreatische intraepitheliale Neoplasien (PanIN) und intraduktale papillär muzinöse Neoplasien (IPMN). Der gastrale Subtyp der IPMN teilt morphologische Homologien mit PanIN, diese beiden Läsionen werden derzeit jedoch nur nach ihrer Größe unterschieden. Ihre klinische Relevanz unterscheidet sich jedoch stark, sodass strenge Kriterien für eine korrekte Differenzierung erforderlich sind. Diese Studie zielt darauf ab, die molekularen Merkmale dieser Vorläuferläsionen zu charakterisieren, um ihr biologisches Verhalten besser zu verstehen.

120 Fälle mit resezierten IPMN und/oder PanIN wurden retrospektiv ausgewertet und nach morphologischen und immunhistochemischen Kriterien reklassifiziert. Insgesamt wurden 59 verschiedene Vorläufer mittels Next-Generation-Sequencing mit einem selbst entworfenen Genpanel mit 22 krebsrelevanten Genen analysiert. Für die Detektion von Kopienzahlvariationen (CNV) wurde eine Low-coverage Gesamtgenomesequenzierung mit Formalin-fixiertem Paraffin-eingebetteten (FFPE) Gewebe von insgesamt 47 Fällen durchgeführt. Für Transkriptionsdaten wurde total RNA mit 14 verschiedenen Proben (3 gastrale IPMN und 4 PanIN und als Kontrollen 3 azinäre Gewebeproben und 4 PDAC Proben) sequenziert. Für das Methylierungsprofil wurde der Illumina Infinium MethylationEPIC Array mit insgesamt 49 Proben (darunter 18 gastrale IPMN, 8 PanIN, 10 intestinale IPMN, 7 azinäre Gewebeproben, 3 Proben mit sortierten Betazellen and 3 Proben mit sortierten Gangzellen) verwendet.

Gezielte Sequenzierung mittels NGS bestätigte zwei Treibermutationen in gastralen und intestinalen IPMNs in den Genen *KRAS* und *GNAS*. Im Kopienzahlprofil zeigen die PanIN-Läsionen keine Kopienzahlveränderungen, im Gegensatz zu den gastralen IPMN, die CNV in definierten Regionen aufweisen. Die intestinalen IPMN zeigten das am stärksten abweichende Kopienzahlprofil, was auf ein höheres Transformationspotenzial hindeutet. Das Methylierungsprofil deutet auf die Ähnlichkeit zwischen PanIN und gastralen IPMN hin, wobei nur 0,36 % der 5'-Cytosin-Phosphat-Guanin-3' (CpGs) Stellen unterschiedlich methyliert sind im Gegensatz zu dem völlig anderen Methylierungsprofil der intestinalen IPMN.

Auf Grundlage des Methylierungsprofils konnte ein Hinweis auf die Ursprungszelle der untersuchten Vorläufer-Läsionen gefunden werden. Der entstandene Abstammungsbaum zeigt, dass die gastralen IPMN und PanIN aus dem azinären Kompartiment entstehen und die intestinalen IPMN ihren Ursprung im duktalen Kompartment haben.

Die Analyse der Transkriptionsdaten zeigt 595 signifikant unterschiedlich exprimierte Gene zwischen der gastralen IPMN und PanIN (p-wert < 0.01). Durch die Kombination der signifikant exprimierten und methylierten Gene in einer integrativen Analyse zwischen PanIN und gastraler IPMN wurden 3 verschiedene Hub-Gene für weitere Untersuchungen ausgewählt, darunter die Gene *MUC6*, *TFF1* und *KLF4*. Die potenziellen Marker wurden mittels Immunhistologie in einer Kohorte von 71 Patienten auf Gewebs-Mikroarray-Schnitten getestet. Die Expressionsmuster auf Proteinebene spiegeln nicht die mRNA-Ebene wider und diese Marker sind nicht geeignet für den Nachweis oder die Unterscheidung der Vorläufer. Die höhere Expression dieser Gene, das Kopienzahlprofil und die unterschiedliche Aktivierung der nachgeschalteten Signalwege unterstützen das Konzept eines höheren neoplastischen Potenzials der gastrischen IPMN im Vergleich zu PanIN.

Zusammenfassend lässt sich sagen, dass gastrale IPMN und PanIN eine sehr ähnliche Morphologie, Genetik und ein ähnliches Methylierungsprofil aufweisen, sich aber in ihrem Kopienzahlprofil und Transkriptomprofil unterscheiden. Unsere Daten deuten auf einen gemeinsamen Ursprung mit Akquisition eines höheren neoplastischen Potentials durch die gastrale IPMN hin. Intestinale IPMN sind eigenständige Entitäten mit einem eindeutigen neoplastischen Potential, bereits in niedriggradigen Läsionen mit einem völlig anderen Methylierungs- und Kopienzahlprofil.

Unsere Studie zeigt, dass basierend auf den genomischen Untersuchungen eine Differenzierung zwischen PDAC und den Vorläuferläsionen in Übereinstimmung mit der morphologischen Einteilung gut möglich ist. Solange noch kein validierter Marker gefunden worden ist, ist die immunphänotypische Subtypisierung grundlegend für die Identifizierung von PDAC-Vorstufen mit unterschiedlichem Progressionsrisiko und sollte als Kernelement in Pathologieberichten aufgenommen werden.

Abbreviations

5-FU	5-Fluoruracil
aa	Amino acid
ACMG	American College of Medical Genetics and Genomics guidelines
ADM	Acinar-ductal-metaplasia
AFL	Atypical flat lesion
APC	Allophycocyanin
Arg	Arginine
AS	Average silhouettes
Asp	Aspartic acid
BMFZ	German: Biomedizinisches Forschungszentrum
CA 19-9	Carbohydrate antigen 19-9
cAMP	Cyclic adenosine monophosphate
CC	Colloid carcinoma
CEP	Chromosome enumeration probes
cfDNA	Cell free DNA
CGIs	CpG islands
c-NHEJ	Classical non-homologous end joining
CNV	Copy number variation
CpGs	5'-Cytosine-phosphate-Guanine-3' sites
CRM	Circumferential resection margin
CT scan	Computed tomography scan
CTCs	Circulating tumor cells
Cys	Cystine
DEG	Differentially expressed genes
DKFZ	German: deutsches Krebsforschungszentrum
DMG	Differentially methylated gene
DMP	Differentially methylated pattern
DNA	Deoxyribonucleic acid
DNMTs	DNA methyltransferase
DSBs	Double-strand breaks
EMT	Epithelial-mesenchymal transition
EUS	Endoscopic ultrasound
FAM	6-Carboxyfluorescein
FFPE	Formalin fixed paraffin embedded
FISH	Fluorescence *in situ* hybridization
FITC	Fluorescein isothiocyanate
GAPS	GTPase activation proteins
GDP	Guanosine diphosphate
GEFS	Guanine nucleotide exchange factors
Gly	Glycine
G-protein	Guanine nucleotide-binding protein
GTP	Guanosine triphosphate
H&E staining	Hematoxylin and Eosin staining

Hh	Hedgehog
His	Histidine
HML-2	Human mammary tumor virus like-2
IGR	Intergenic region
IPMN	Intraductal papillary mucinous neoplasms
IRS	Immune reactive score
Kbp	Kilo base pair
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPC	Laser Pressure Catapulting
Lys	Lysine
MAPD	Median of the absolute values of all pairwise differences
Mbp	Mega base pairs
MCN	Mucinous cystic neoplasm
MRI	Magnetic resonance imaging
MTR	Missense tolerance ratio
NGS	Next generation sequencing
PanIN	Pancreatic intraepithelial neoplasm
PCA	Principal component analysis
PCN	Pancreatic cystic neoplasms
PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PER	Pathogen enriched regions
PPI- network	Protein-protein-interaction network
qRT-PCR	Quantitative real-time polymerase chain reaction
RNA	Ribonucleic acid
RT	Room temperature
TAMRA	Tetramethylrhodamines
TMA	Tissue micro array
UDG	Uracil-DNA glycosylase
VAF	Variant allele frequency
Val	Valine
VIC	4,7,2’-Trichloro-7’-Phenyl-6-Carboxyfluorescein
WGA	Whole-genome amplification
WHO	World health organization

1 Introduction

1.1 Pancreatic ductal adenocarcinoma

The definition from the world health organization (WHO) of a pancreatic ductal adenocarcinoma (PDAC) is an invasive pancreatic epithelial neoplasm with glandular differentiation, usually demonstrating luminal and/or intracellular mucin production[1].

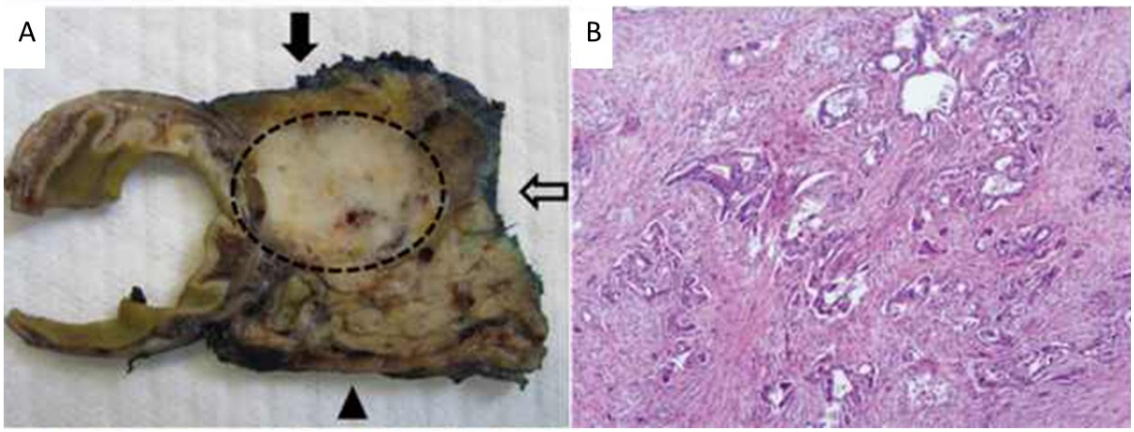


Figure 1: Macroscopy and morphology of the PDAC.
(A) Head of the pancreas with connecting gut. Macroscopically view of a PDAC; gross morphology of PDAC after axial slicing. PDAC presents as a solid ill-defined white-yellowish mass of the pancreas head (circle). (B) Typical histomorphology of PDAC. (Figure is adapted from Haeberle *et al*. 2019 [2]).

It accounts for about 90 % of the tumors in the pancreas. Most of the PDACs arise in the head of the organ (60-70 %), about 20-25 % of the tumors are located in the body or the tail[3]. The PDAC appears as a poorly-defined, white-yellowish mass, most of the time without necrosis (Figure 1 A). When the tumor is located in the head of the pancreas it typically leads to stenosis and proximal dilatation of the main pancreatic duct and/or the common bile duct. This leads to painless jaundice and obstructive chronic pancreatitis in many cases. For this reason, the size of the tumor in the head is usually smaller than in the body/tail, due to the early symptoms[4]. The typical histopathology of pancreatic ductal adenocarcinoma is a well-differentiated tumor with duct-like glandular morphology. Neoplastic glands are embedded in a desmoplastic stroma (Figure 1 B)[5].

The diagnosis of a PDAC is usually late, so most of the carcinomas are already infiltrating the surrounding structures like the ampulla of Vater, the duodenal wall, the retroperitoneal tissue, and the superior mesenteric vessels. Tumors of the body/tail region can infiltrate the stomach wall, the left side of the colon, the spleen, or the left adrenal gland. PDAC is considered a micrometastatic disease. Here a small collection of cancer cells can shed from the original tumor and spread fast to another part of the body[6].

1.1.1 Epidemiology, survival and risk factors

The PDAC accounts for 90 % of solid tumors in the pancreas. It is a very aggressive disease with a bad prognosis and will cause the second most cancer-related deaths by 2030. PDACs ranks 10th in the incidence of solid cancer in men and 9th in the incidence in women in the United States in 2019[7]. The incidence of pancreatic cancer is slightly higher in men than in women, being 5.5 per 100,000 for men and 4.0 for women[8]. PDAC is one of the most lethal malignancies with a mortality/incidence ratio of 94 %[9].

The median age at diagnosis is 71 years and PDAC rarely occurs before the age of 55. The risk also increases with advanced age, the highest incidence is reported in people over 70 years.

In 2020, the 5-year survival rate of PDAC has reached 10 % for the first time in the United States. The improvement in treatment and survival rate in comparison to other cancer types is still minimal[10]. In the last 40 years, the 5-year survival rate in cancer generally rose from 25 % to 50 % in the UK (Figure 2). In pancreatic cancer, it raised from 2 % to 10 %. The improvements, in general, are according to better and faster diagnosis and advanced treatments[11]. From the 1970s to 2010, the 1-year survival rate increased from 10 % to 22 % in men and 20 % in women. The worse improvement in pancreatic cancer is underlining the desperate need for new approaches in diagnostics and therapy of this particular type of cancer.

In the case of PDAC, the detection of early-stage tumors is most of the time incidental finding. Early detection is rare due to the fact that the progression and growth of the tumor is mostly without symptoms[8]. Symptoms mostly appear late, when the tumor is on an advanced stage. The localization of the pancreas in the retroperitoneal upper abdomen is difficult to access, which makes the diagnosis difficult. Diagnosis with imaging methods like computed tomography scan (CT scan) has limits due to the lack of pancreatic and bile duct dilatation in early stages and a lack of reliable diagnostic markers do not improve the diagnosis[12,13].

Different risk factors for pancreatic cancer exist. One of the factors is chronic pancreatitis, which can increase the risk up to 13.3- fold[14,15].

Another important risk factor is diabetes mellitus. New adult-onset of diabetes may be an early sign of pancreatic malignancy. Long-standing diabetes mellitus increased the risk for PDAC by 2- fold.

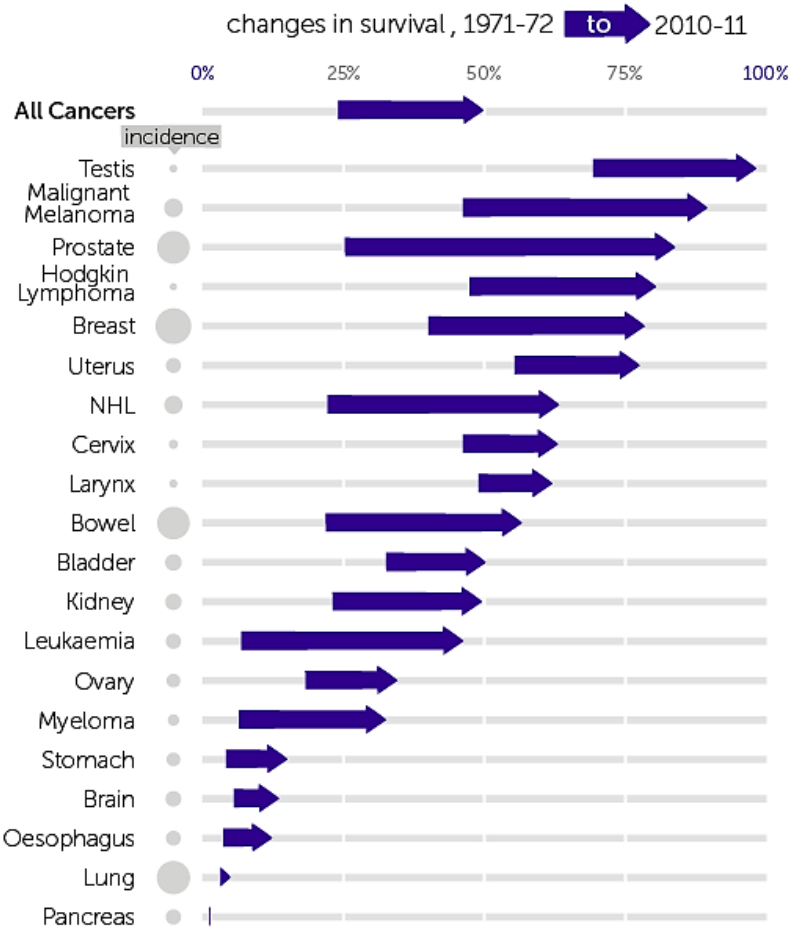


Figure 2: The 5-year survival in different cancer types between 1971 and 2011.
The figure shows the development of the 5-year survival rate of different cancer types within a time period of 40 years. The general improvement in the survival is increased from 25 % to 50 %. The survival rate of pancreatic cancer is still at a low level at about 9-10 %, with a minimal improvement over the last decades. (Figure is adapted from Cancer Research UK [10]).

Further studies are needed to understand whether diabetes can predict the onset of pancreatic cancer or be a marker for it. In a pooled study of 8,305 patients with pancreatic cancer, 30 % of the patients received a diagnosis of diabetes close to the diagnose of pancreatic cancer[16].

Additionally, 5-10 % of pancreatic cancer patients have a family history of pancreatic cancer. Individuals with affected first-degree relatives have a 9- fold increased risk of pancreatic cancer in comparison to the general population[17]. Patients with familiar history of PDAC have an increasing number of precursor lesions than those with sporadic pancreatic cancer[18]. The risk can increase up to 32- fold, when at least three or more first-degree relatives are affected by pancreatic cancer.

Besides the inherited pancreatic cancer, other familial syndromes are also related to an increase of the risk of pancreatic cancer by the inheritance of specific genetic mutations (germline mutations)[19]. Some susceptibility genes are *BRCA1* and *BRCA2* (Breast Cancer gene

1 and 2), which are known and associated with the hereditary breast-ovarian cancer syndrome. Other examples are *APC* (APC Regulator of WNT Signaling Pathway) associated with familial adenomatous polyposis, *STK11* (Serine/Threonine Kinase 11) associated with the Peutz-Jeghers syndrome, and *CDKN2A* (Cyclin Dependent Kinase Inhibitor 2A) associated with familial atypical multiple mole melanoma syndrome. In addition, *PALB2* (Partner And Localizer Of BRCA2) is associated with the Fanconi anemia DNA repair pathway and *MLH1*, *MLH2* (MutL Homolog 1 and 2), *MSH6* (MutS Homolog 6), and *PMS2* (PMS1 Homolog 2, mismatch repair system component) with hereditary nonpolyposis colorectal cancer (also called: lynch syndrome) also increase the risk to develop pancreatic cancer[20].

Aside from the genetic risk factors, various additional lifestyle risk factors for PDAC are described, like tobacco smoking, alcohol abuse, a diet high in red meat, and obesity[8,17,21–24].

1.1.2 Diagnosis and therapy

Symptoms are mostly non-specific and can include back pain, abdominal pain, digestive problems, weight loss, and new-onset diabetes mellitus. The first symptoms result from a mass effect of the growing tumor. 60-70 % of the tumors grow in the head of the pancreas, where the probability of an obstruction of the common bile duct and/or pancreatic duct is higher, which leads to earlier symptoms. This is the reason why tumors in the body or tail of the pancreas are mostly in an advanced stage by diagnosis.

Only 10-15 % of the PDACs are in a local stage for resection at the date of diagnosis. Surgery is the only curative treatment today. This small subset of resectable tumors has a 5-year survival rate of about 37 %[25]. For tumors in a regional stage, the 5-year survival rate is 12 %. Most of the PDACs (53 %) are detected in an advanced stage, when the primary tumor has already spread and metastasis appeared, mostly in the liver (80 %) or the lung (45 %) as the first site of recurrence[25–27]. The 5-year survival rate for metastatic PDAC is 3 %. This dramatic prognosis is underlining the urgent need for new approaches in the diagnosis and treatment of PDAC.

According to the guidelines from the European society for medical oncology (ESMO) 2019, different therapy options are possible, depending on the status of the tumor and the fitness of the patient[28]. The therapy options are resection, adjuvant chemotherapy, chemotherapy and/ or palliative drug tumor therapy depending on the stage of the tumor and the fitness of the patient. With only few targetable mutations, which are investigated so far, only one option

for personalized medicine exists today. Tumors with germline *BRCA1/2* mutation and a metastatic PDAC are treated efficient with olaparib[29].

One major problem in the treatment of pancreatic cancer is the challenges of chemotherapy. The chemoresistance with a standard cytostatic drug, for example with gemcitabine, occurs for different reasons. In brief, the tumor microenvironment protects the tumor from most of the chemotherapeutic agents[30]. For example, the stromal reaction, which occurs around the tumor, serve as a barrier to chemotherapeutic agents[31]. Furthermore, other common processes during pancreatic carcinogenesis, like the epithelial-mesenchymal transition (EMT), impedes the chemotherapy[32]. This is a very complex topic with many different influencing factors.

The limited therapy options and the worse prognosis make early detection very important. A couple of different approaches exist for diagnostic markers for detection of pancreatic cancer and to distinguish between benign lesions, PDAC, and chronic pancreatitis[13].

In clinical practice, the serum marker carbohydrate antigen 19-9 (CA 19-9) is the only routinely used marker for PDAC today. The marker is not recommended for screening PDAC, it is a sign of disease progression and can be used for detecting a recurrence of pancreatic cancer[33]. Only in 65 % of the resectable PDAC patient, the CA 19-9 marker is detected in an elevated level[34]. Several studies are testing various potential markers alone and in combination with CA 19- 9. A list of different marker types belonging to proteomics, metabolomics, cytokines, and noncoding RNAs are reported. All of them show a similar specificity and sensitivity as CA 19-9 or fail to distinguish between chronic pancreatitis or healthy controls[35–38].

The newest research field is liquid biopsy, where cell-free DNA (cfDNA) and/or circulating tumor cells (CTCs) are in the focus of research. For 73 % of PDAC patients, regardless of the tumor stage, CTCs are detectable. The main limitation is the lack of standardized isolation and characterization[39–41].

Another approach for early detection is the characterization of the precursor lesions, where the time span to find a premalignant lesion can be much longer (estimate 1-3 decades) as opposed to finding an early-stage tumor, which develop, on average, in 14 months to a metastatic tumor. The estimated time span from precursor lesions to an early-stage tumor is controversial discussed and variates from about 10 years up to 30 years[42–44].

1.2 Precursors of the PDAC

To improve the survival rate and the therapy options, the progression of pancreatic cancer must be fully understood. Therefore, the precursors should be the first therapeutical targets for early detection and consisted of large impact for diagnostic marker[39].

All precursors are graded according to a two-tiered grading system (low grade and high grade dysplasia) based on the degree of cytonuclear and architectural atypia[45]. Several precursors are known to date. In the following chapter, the common precursors, mainly analyzed in this study, are briefly described.

1.2.1 Pancreatic intraepithelial neoplasia

Pancreatic intraepithelial neoplasia (PanIN) is the most common precursor of PDAC, the majority (recognizable 85 %) of PDAC arise from PanIN lesions[46]. PanIN is a microscopic lesion with a diameter below 0.5 cm[47]. PanIN lesions usually occur in small ducts, mainly in the head of the pancreas. There is no dilatation of the duct, so diagnostics with imaging tools is not possible due to the size and localization[48].

PanIN lesions display a flat, mostly micropapillary architecture[1]. PanIN lesions are composed of cuboidal to columnar cells producing varying amounts of mucin. The lesions are grading according to a two-tiered grading system into a low and high grade based on the degree of cytonuclear and architectural atypia (Figure 3)[45].

The incidence of low grade PanIN increases by age. It is very common that low grade PanIN lesions are found incidentally. The lesions are found in over half of the population with an age older than 50 years[49,50]. In contrast, high grade PanIN lesions are rarely seen without invasive ductal adenocarcinoma[50]. So, high grade PanIN lesions may have clinical significance and may serve as a surrogate marker for invasive carcinoma[51]. Currently, low grade PanIN in resected specimens do not have any clinical relevance[50].

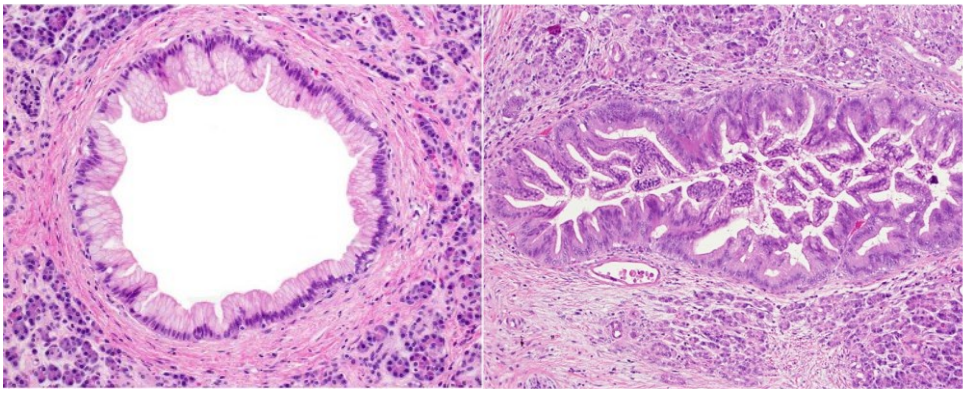


Figure 3: Morphology of a PanIN lesion and the immunohistology pattern.
PanIN low grade lesion (A) and PanIN high grade lesion (B). (Figure is adapted from Basturk *et al*. 2015 [45]).

The immunohistochemical labelling pattern of PanIN is similar to that of PDAC. None of the lesions express MUC2 (mucin 2), but most express MUC5AC (mucin 5AC) and some express MUC1 (mucin 1). In about 20 % of the low grade PanIN lesions a MUC1 expression is observable[52]. In a small cohort of 41 patients, the increasing Ki-67 proliferation index has been shown with increasing grades of dyplasia[53]. Recent studies have shown that aberrant expression of *TP53* (tumor protein p53) gene is rare and SMAD4 (SMAD Family Member 4) immunolabelling is retained in high grade PanIN lesions [54,55].

1.2.2 Atypical flat lesion

Atypical flat lesion (AFL) is a recently described microscopic (< 0.5 cm) precursor lesion, which is found in acinar-to-ductal metaplasia (ADM) areas. The lesion has been identified in a murine model of PDAC and in patients with familial pancreatic cancer[56]. It is a non-mucin-producing lesion with highly proliferative areas. AFLs show cytological atypia and are typically surrounded by a loose but highly cellular stroma[57]. AFL displays a tubular structure (Figure 4).

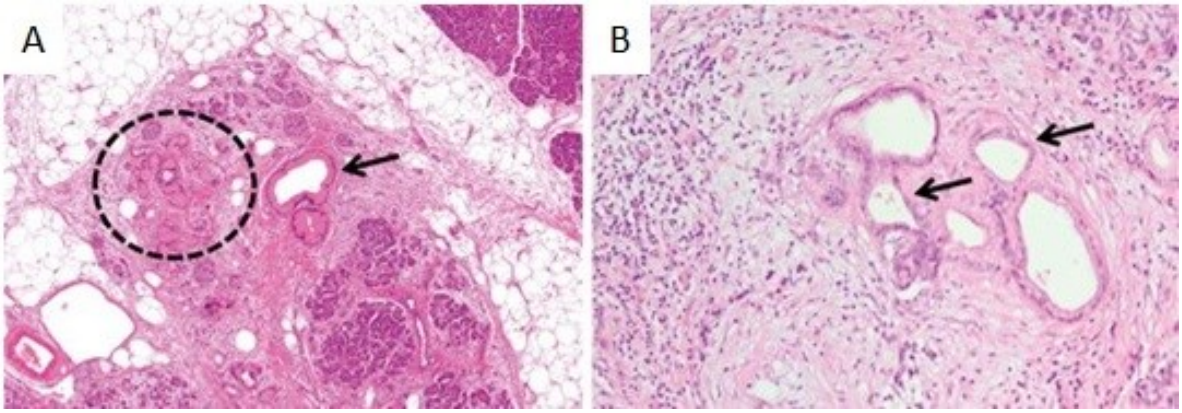


Figure 4: Morphology atypical flat lesion (AFL).
(A) Area of ADM in a patient with a familiar predisposition of PDAC. The ADM area shows mucinous tubular complexes (circle) and a PanIN low grade (arrow). (B) Atypical flat lesion (arrows) in an ADM area. The surrounding stroma is loose myxoid and cell-rich. (Figure is adapted from Esposito *et al.* 2012[58]).

It is discussed as another precursor lesion of PDAC and seems to represent the link between ADM and ductal carcinogenesis[47].

Most of the observations are from mouse models. In the *Ptf1a*^{CreER}; *LSL-Kras*^{G12D} mice, the expression of different immunomarker was investigated. The acinar marker CPA1 (Carboxypeptidase A1) is not expressed in AFLs. In addition, AFLs have a proliferative character as evidenced by an increased expression of Ki-67 in the nuclei and the marker p53 and pMAPK (mitogen-activated protein kinase) are broadly expressed in nuclei of AFLs[57].

1.2.3 Intraductal papillary mucinous neoplasms

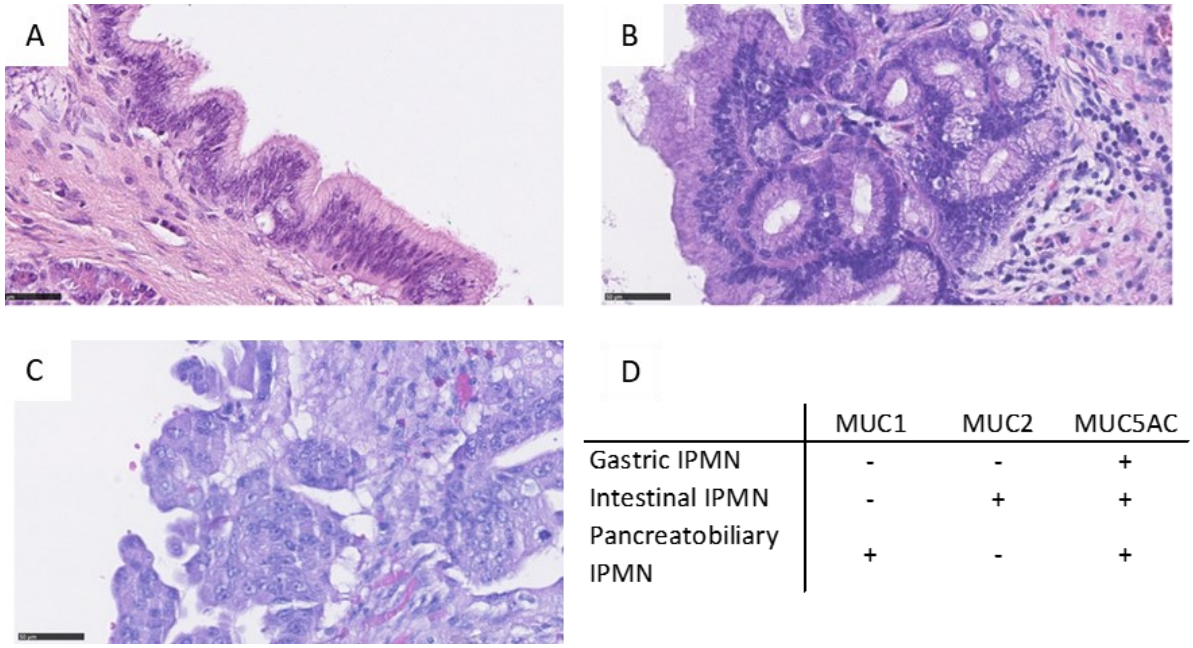
Intraductal papillary mucinous neoplasms (IPMN) are grossly visible and are typically over 0.5 cm in diameter. These lesions arise in the main pancreatic ducts and/or their branches. IPMN consist of mucin-producing cells. The neoplasms are graded according to a two-tiered grading system. The main duct and branch duct IPMN are clinical subtypes and not pathological subtypes. In 80 % of the cases, IPMNs are located in the head of the pancreas. However the IPMN is multicentricity lesion, which means that the lesion processes different centers[59].

The mean age of the patients is between 62-67 years. Clinical symptoms include epigastric pain, chronic pancreatitis, weight loss, diabetes mellitus and jaundice. Branch duct-type IPMNs are often detected incidentally. There are no well-established etiological factors for IPMN and the pathogenesis of IPMN is so far unknown. 45- 60 % of main duct IPMNs harbor an associated invasive carcinoma or a high-grade dysplasia. IPMNs without an invasive carcinoma are often curable, the 5-year survival rates for patients with resected low grade IPMN and high grade IPMN were reported as 100 % and 85 % to 95 %[60–62]. In total, only 15 % of the PDACs arise from mucinous pancreatic cysts, which include IPMN and mucinous cystic neoplasm (MCN)[39].

IPMN lesions can be divided into three different pathological subtypes, namely gastric, intestinal, and pancreatobiliary IPMN (Figure 5 A-C). The different subtypes can be distinguished by immunohistological markers MUC1, MUC2, and MUC5AC. The different morphology and marker expression patterns of the different IPMN subtypes are shown below (Figure 5)[63].

Diagnosis of cysts is possible before the tumor arises. The clinical management is not dependent on the pathological subtype; it depends on the clinical subtype, main or branch duct IPMN. Absolute indications for surgery are established in international guidelines, such as that of the European study group on cystic tumors of the pancreas in 2018. Five different absolute indications for surgery are named, such as positive cytology for malignancy or high grade dysplasia, solid mass formation, jaundice, enhanced mural nodule (≥ 5 mm) or main pancreatic duct dilatation (≥ 10 mm). In addition, some relative indications are named including the growth rate with > 5 mm per year, an increased CA 19- 9 level (> 37 U/mL), a cyst diameter of ≥ 40 mm, new onset of diabetes mellitus, acute pancreatitis or a main pancreatic duct dilatation between 5.0 and 9.9 mm[64]. Without these factors, a follow-up with computed

tomography scan (CT)/magnetic resonance imaging (MRI) or endoscopic ultrasound (EUS) is suggested depending on the size in time intervals of 3 years (cyst size < 1 cm) or every 3 months (cyst size > 3 cm)[65,66].



	MUC1	MUC2	MUC5AC
Gastric IPMN	-	-	+
Intestinal IPMN	-	+	+
Pancreatobiliary IPMN	+	-	+

Figure 5: Morphology and marker expression of the different IPMN subtypes.
Representing picture of (A) gastric IPMN, (B) intestinal IPMN and (C) pancreatobiliary IPMN. (D) Marker expression in the different pathological subtypes of IPMN. (Figure is adapted from Klausen *et al*. 2019 [67] and Painin *et al*. 2014 [63]).

1.2.3.1 Gastric IPMN

The most common subtype of IPMN is the subtype with foveolar phenotype, the gastric IPMN. About 70 % of the IPMNs are of the gastric subtype, which usually occurs in the branch ducts. The gastric IPMN is composed by gastric-type epithelium and forms tall columnar cells with basally-oriented nuclei and pale mucinous cytoplasm, reminiscent of gastric foveolar epithelium. Most of the gastric IPMN lesions are low grade lesions (Figure 5 A)[68,69].

1.2.3.2 Intestinal IPMN

20 % of the IPMNs belong to the intestinal subtype, which typically occurs in the main duct. The intestinal IPMN is characterized by intestinal-type epithelium that forms villous papillae composed of tall columnar cells with cigar-shaped enlarged nuclei and basophilic cytoplasm with a variable amount of apical mucin (Figure 5 B)[70]. Most of the intestinal IPMN are high grade lesions[68,69].

1.2.3.3 Pancreatobiliary IPMN

Pancreatobiliary IPMNs are rare and typically involves the main pancreatic duct. This subtype of IPMN has complex arborizing and interconnecting papillae composed of cuboidal cells with amphophilic cytoplasm, enlarged nuclei, and prominent nucleoli. Due to the architectural complexity and degree of nuclear atypia, most pancreatobiliary type IPMNs are high grade lesions (Figure 5 C)[71]. Focal MUC1 expression is detected in complex thin-branching papillae and highly MUC5AC expression is present in pancreatobiliary subtype of IPMN. This subtype had the worst prognostic of the three IPMN types[72].

1.2.4 Intraductal oncocytic papillary neoplasms

The intraductal oncocytic papillary neoplasm (IOPN) was first described in 1996 and since 2019 is a distinct type of intraductal neoplasm of the pancreas in the WHO guidelines[73]. IOPN, former oncocytic IPMN, are a grossly cystic epithelial neoplasm composed of exophytic nodular projections lined by oncocytic glandular epithelium, which grows within dilated pancreatic ducts (Figure 6). This type of intraductal neoplasm forms large cystically dilated pancreatic ducts, with a mean size of 5.5 cm. 70 % of the IOPNs occur in the head of the pancreas and involve the main duct. 30 % of the IOPN cases are associated with an invasive carcinoma[74]. Only 4.5 % of all intraductal neoplasms are IOPN[75].

The IOPN are more common in females and the mean patient age is 59 years. So far, the epidemiology and etiology are unknown[74]. The pathogenesis is not well understood, the IOPN lacks the typical mutations related to PDAC such as *KRAS* (Kristen rat sarcoma), *GNAS* (Guanine Nucleotide Binding Protein (G Protein) Alpha), and *RNF43* (Ring Finger Protein 43)[76]. Immunohistochemically, IOPNs diffusely label for MUC1 and MUC6 (mucin 6), whereas MUC2 expression is largely restricted to goblet cells[77].

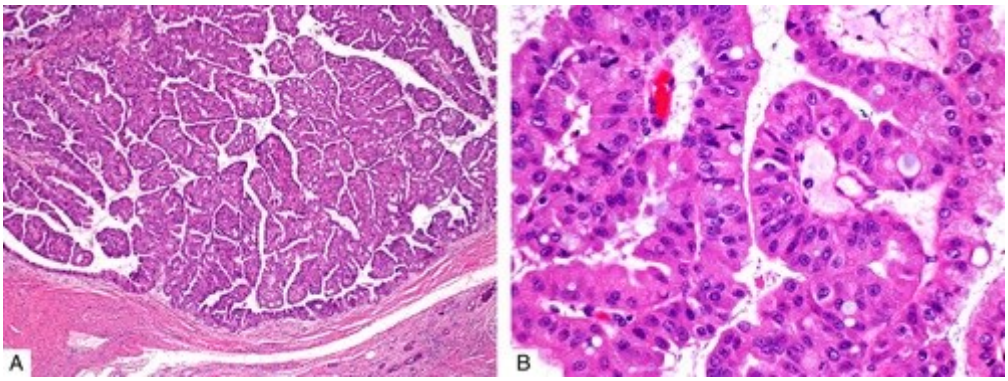


Figure 6: Intraductal oncocystic papillary neoplasms.
The IOPN lesion exhibits papillary architecture with distinct oncocytic cytology and intracytoplasmic lumens. (Figure is adapted from Wang *et al*. 2019 [74]).

1.2.5 Mucinous cystic neoplasm

Mucin cystic neoplasm (MCN) is a cyst-forming and mucin-producing epithelial neoplasm associated with distinctive ovarian-type subepithelial stroma. 8 % of the cystic lesions are MCNs, most commonly located in the body or tail of the pancreas. Over 95 % of MCNs occur in women and the average patient age is 48 years[78]. In a fine needle aspiration (FNA), it is not possible to distinguish between IPMN and MCN.

The cysts of MCNs are lined by epithelium and have underlying ovarian-type stroma. The epithelium is predominantly columnar with mucin-producing cells. The cuboidal cells lack mucin, similar to non-neoplastic ductal cells (Figure 7 A)[79,80]. Estrogen receptor-positive and progesterone receptor-positive ovarian type stroma is typical for an MCN (Figure 7 B)[81].

The mean size of MCN is 6.5 cm, the 5-year survival rate is 96.6 %[82]. Only about 15 % of MCNs have an associated invasive carcinoma component. The clinical management and the indications for surgery are the same as for IPMN[64].

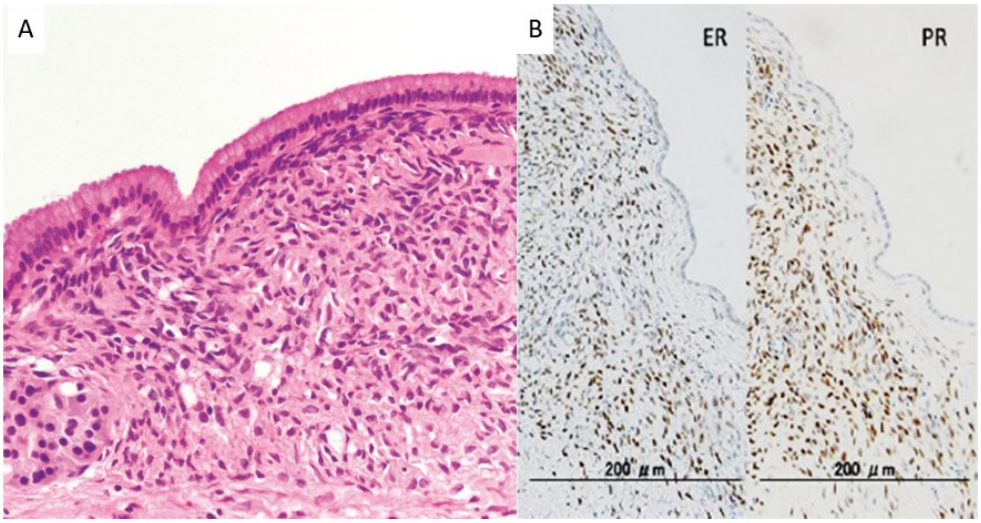


Figure 7: Mucinous cystic neoplasm of the pancreas.
(A) Mucinous cystic neoplasms demonstrate a simple mucinous lining with cellular ovarian-type stroma underneath. (B) immunohistology marker expression of estrogen receptor (ER) and progesterone receptor (PR) in the underlying ovarian type stroma. (Figure is adapted from Pittman *et al*. 2017 [46] and Hijioka *et al*. 2015 [81]).

The development of the MCNs is thought to be in the embryogenesis when the ectopic ovarian stroma incorporated in the pancreas. Other organs may become activated in the setting of hormonal imbalance and growth factors and causing nearby ductal epithelium to proliferate and form cystic neoplasms[83,84]. In rare cases, MCN occurs in male, so another hypothesis is that the ovarian-type stroma represents persistent fetal periductal mesenchyme, which may respond and proliferate in response to a hormonal stimulation[85].

1.3 Genetic alterations involved in PDAC progression

Analyzing the progression of pancreatic cancer, it is important to investigate different possible pathways for the development. The step-wise progression analog to the colon-adenoma-carcinoma sequence, which was postulated from Vogelstein and colleagues in 1988, is the most commonly discussed development pathway[86]. The other possible involved mechanism is the chromothripsis with chromosomal alterations occurring during a catastrophic chromosome event. In addition, DNA-methylation plays a crucial role in the progression[87]. In the next chapters, the principles of the different mechanisms in PDAC will be described shortly.

1.3.1 The stepwise process from normal tissue over precursor lesions to PDAC

In 1988, Vogelstein and colleagues postulated a progression model for colorectal neoplasia in which they hypothesized that the progression from normal colonic epithelium to small adenomatous polyps to infiltrating adenocarcinoma is associated with the activation of oncogenes and the inactivation of tumor suppressor genes[86]. This model is still accepted and is expanded to different tumor entities as for pancreatic cancer.

In 2000, the progression model was adapted for PDAC from Hruban and colleagues. Data from postmortem examination and clinical studies proposed a similar tumor progression model from pancreatic precursors to invasive pancreatic cancer as in the colon-adenoma sequence[88]. This model is well-accepted and various mouse models have reinforced its validity[6,89,90].

Pancreatic carcinogenesis is dependent on various events, including gene alterations (the activation of oncogenes like *KRAS* and the inactivation of tumor suppressor genes like *CDKN2A*, *TP53* and *SMAD4*), environmental insults (pancreatitis and the previously explained risk factors) and including cell types[91]. *KRAS* is a master regulator of pancreatic ductal adenocarcinoma initiation and progression, this alteration takes place as a very early event[92]. Pancreatic carcinogenesis is context-dependent and complicated. It seems that the different known precursors have different cells of origin. The initiation step is unknown in most of the types and the process itself is not well understood[93].

The PanIN lesions develop through the transdifferentiation of acinar cells, the so-called acinar-ductal- metaplasia (ADM)[93]. Metaplasia is a reversible process and takes place during acute and chronic pancreatitis. The initiation of ADM is TGF-beta (transforming growth factor beta) dependent and in recent studies, *KLF4* (kruppel-like factor 4) was found as a required factor

for the metaplasia[94,95]. In addition, AFL lesions develop as well from the centro-acinar compartment via an ADM[58,56]. The estimated time from PanIN to PDAC is 33.6 years. [43].

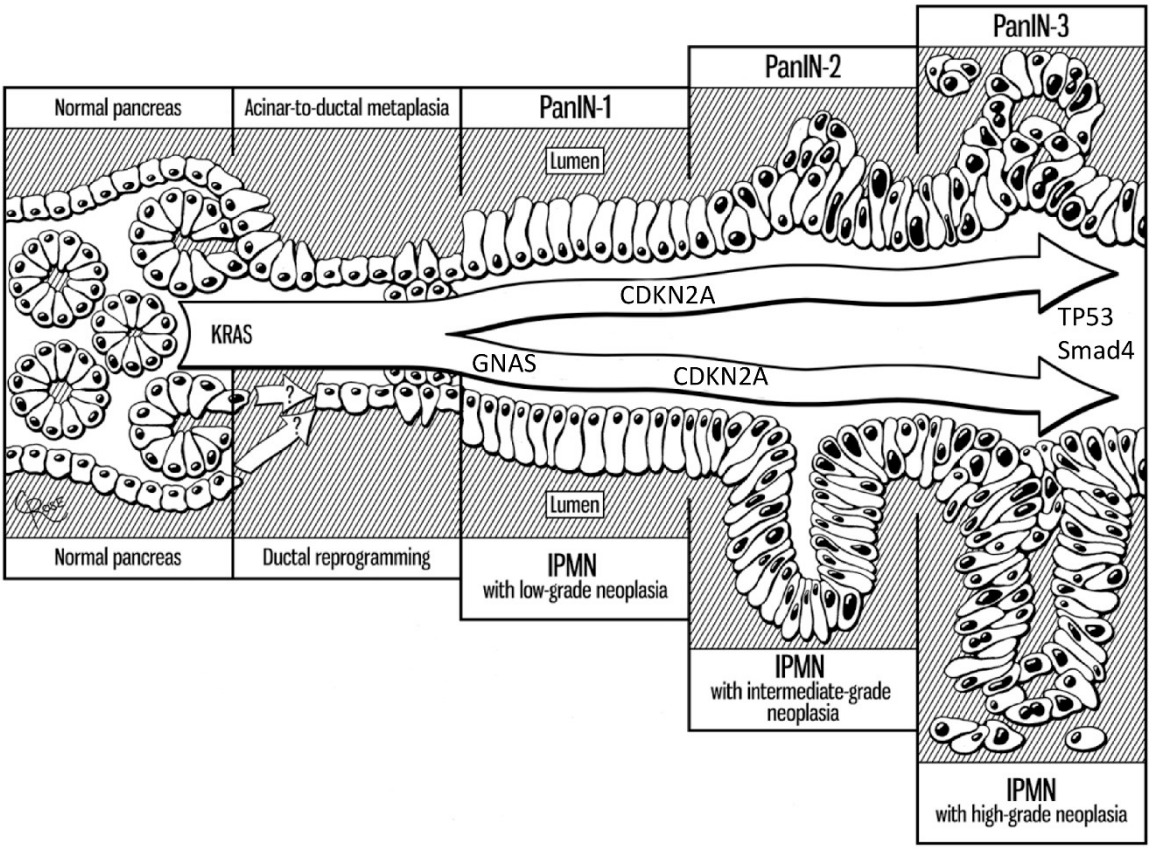


Figure 8: PanIN and IPMN development to invasive carcinoma.
Tumor progression model starting from differentiated endocrine or exocrine cell types resulting in pancreatic ductal adenocarcinoma. Oncogenic *KRAS* is highlighted as the central driver of transformation. Tumor progression is associated with increasing stroma accumulation along with increasing aberrations of epithelial architecture. (Figure adapted from Qiu *et al*. 2016 [96] and from Patra *et al*. 2017 [97]).

IPMN precursor seems to originate from the progenitor compartment of the ductal epithelium[98]. The IPMN pathway is poorly understood and most of the genetic findings and progressions predominantly came via insights from mouse model studies[99].

Besides the two different progression pathways via precursor lesions to pancreatic cancer, a progression model from healthy ductal cells directly to PDAC in a PanIN-independent manner is also described[100]. If the PDAC is established, the estimated time for a PDAC tumor to progress from a T1 to a T4 stage is estimated to be approximately 14 months[44].

An early event in the development and initiation of the PDAC is an alteration in the gene *KRAS* (Kristen rat sarcoma*)*. It is an oncogene, which is altered in 30 % of all cancers in humans[101,102]. The proto-oncogene encodes for a small GTPase transducer protein, belonging to a group of small guanosine triphosphate (GTP) binding proteins, which belong to the highly conserved *RAS* superfamily[103].

A point mutation in *KRAS* is found in about 90 % of all pancreatic cancers, most of the mutations appear in codon 12 at exon 2, as well as in codon 13 and 61 and less frequency occur in codon 63, 117, 119 and 146[104]. The G12D variation is the most common one with about 35 % of all *KRAS* mutations in PDAC, followed by G12V and G12R[105]. The G12V mutation is associated with a worse prognosis than G12D in colorectal and lung cancer[101].

All these mutations are a form of gain-of-function mutation. The conformational change of *KRAS* changed to an active state, the *KRAS* binds GTP (guanosine triphosphate). These GTP-bound state (continuously active form) leading to prolonged activation of its downstream associated pathways[106].

In detail, the changes in KRAS affecting the interaction with GTPase activation proteins, which amplify the GTPase activity of RAS proteins around 100,000-fold and affecting the interactions with guanine nucleotide exchange factors (GEFs) and promoting the release of GTP[101,106,107]. These lead to uncontrolled cellular growth, evasion of apoptotic signals, metabolic reprogramming and remodeling of the tumor microenvironment[108].

Another alteration, which is associated with the early phase of carcinogenesis is the alteration in *CDKN2A* (cyclin-dependent kinase Inhibitor 2A) gene. The somatic mutations of *CDKN2A* are present in up to 95 % of pancreatic tumors[109]. This gene encodes for p16, which is a recognized tumor suppressor because of its role in preventing progression through the G1 cell cycle checkpoint[110]. Alterations of *CDKN2A* are present in PanIN and IPMN lesions[111].

Another early event, mainly in the IPMN, is the alteration in the *GNAS* complex locus. It is the alpha subunit belonging to guanine nucleotide-binding protein (G protein $G\alpha_s$). The oncogene is predominantly mutated in codon 201 in IPMN with 41-75 % of the cases[112,113]. *GNAS* is involved in ligand binding to seven transmembrane receptors activated adenylyl cyclase (AC) to produce the second messenger cyclic adenosine monophosphate (cAMP)[114,115]. *GNAS* mutation is present in various endocrine tumors and fibrous dysplasia of bone, and in a more widespread distribution in patients with McCune-Albright syndrome[116].

The most common alteration in a late phase of carcinogenesis is the alteration in the *TP53* gene (tumor protein p53). *TP53* is the most commonly inactivated tumor suppressor in PDAC. Approximately 70 % of patients with PDAC harbor alterations in the *TP53* gene[117,118]. In a tumor suppressor gene, alterations are a loss-of-function mutation and are spread over the whole gene with different hotspots. *TP53* is a transcription factor that regulates the expression of several genes and its biological functions include the inhibition of cell

proliferation by inducing *p21* (*CDKN1A;* cyclin dependent kinase inhibitor 1A) expression, promoting the apoptosis of tumor cells, maintaining genetic stability, and inhibiting tumor vascularity[119].

The other event in the late phase of carcinogenesis is the alteration in the SMAD family member 4 (*SMAD4*) gene. *SMAD4* is a tumor suppressor gene that is inactivated in more than 50 % of pancreatic cancer cases[120]. Many studies demonstrated that the loss of SMAD4 protein expression was associated with tumor progression, lymph node metastasis and pattern of failure in pancreatic cancer patients[121–123].

Mutations in the tumor suppressor *SMAD4* are present in about 30 % of the high grade PanIN lesions, but they were absent in the low grade lesions[124,125]. Recent studies show, that the *SMAD4* mutations are detected very rare in PanIN high grade lesions without an existing tumor[126]. In mouse models, it was shown that the loss of *SMAD4* alone is not suitable to initiate tumor formation or changing the normal pancreatic phenotype[127]

1.3.2 Chromothripsis

To develop from a healthy cell to a cancer cell, typically more than one single event is necessary. The stepwise process like the colon-adenoma sequence from Vogelstein is an example. The process includes several DNA alteration events, which results in incremental development and progression of cancer[128].

Chromothripsis on the other hand is a one-hit event. It is a single catastrophic event, which is driven by multiple double-strand breaks (DSBs)[129]. This event is limited to one or a few chromosomes. The chromosomes break and reassemble the DNA in random order and orientation is followed. During this event, complex derivative chromosomes are built, large segments are lost, wrong orientation, and wrong order appear[130]. Chromothripsis acts on chromosomes in association with mitosis, which may explain the highly localized nature of DNA breakpoints on a single (or few) chromosomes[131]. For the first time in 2011, Stephens *et al*. described these phenomes and called it chromothripsis[132].

The main repair mechanism after the catastrophic chromosome event is the classical non-homologous end joining (c-NHEJ) mechanism. This mechanism repairs double-strand breaks in the DNA. Chromothripsis can lead to an abrupt change in the characteristics of its carrier cell in a single catastrophic event[133].

In 2013, Zack *et al.* performed a metadata analysis from a pan-cancer study with 4934 cancer patients from the cancer genome atlas pan-cancer data set (TCGA- Pan-Cancer). The main finding was that about 5 % of all cancer patients show hints of chromothripsis. The group found 140 existing regions with somatic copy number alteration over the whole-genome, 102 regions without any known oncogenes or tumor suppressor genes[134].

In pancreatic cancer, a study from Notta *et al.* in 2016 shows that 2/3 of pancreatic cancer harbored a complex chromothripsis-like pattern[135]. Even in the precursor lesion PanIN, 40 % of the lesions were affected by chromothripsis-like events, a majority in high grade lesions[136]. Another meta-analysis study from 2020 also found that 60- 65 % of pancreatic cancer patients show chromothripsis-like rearrangement by using WGS (whole-genome sequencing) data[137]. They describe chromothripsis as a major process that drives genome evolution in human cancer. In both studies, there was no data about the presence of chromothripsis in other precursor lesions such as MCN or IPMN.

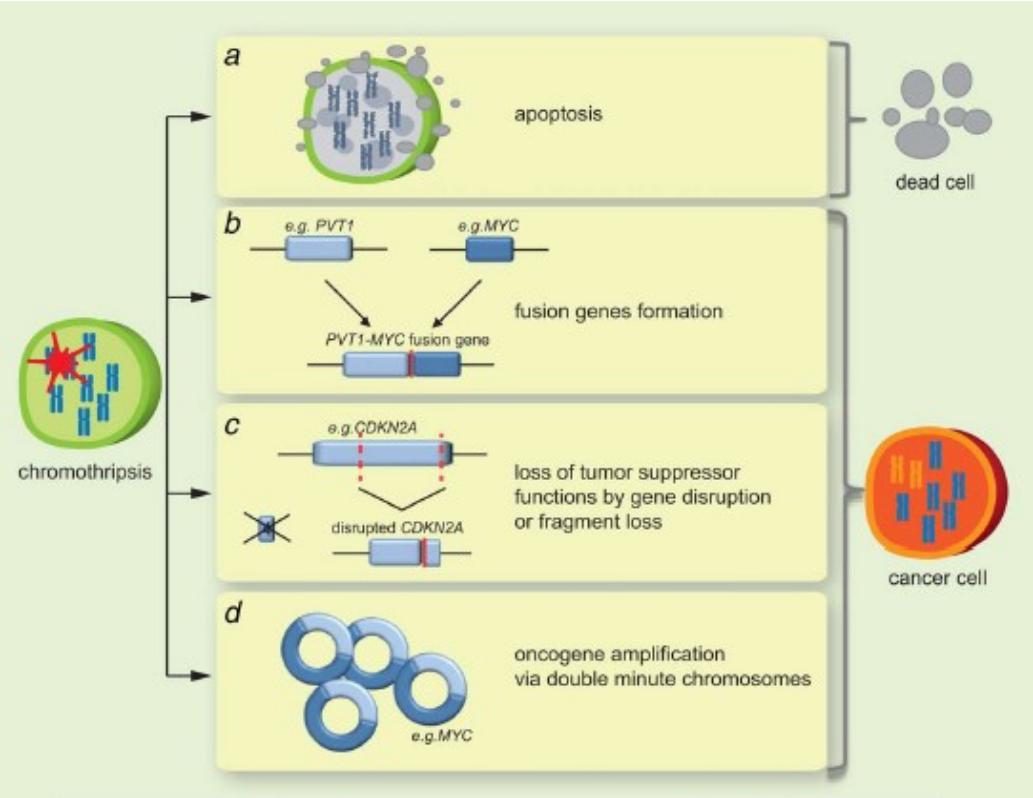


Figure 9: Mechanisms of cancer progression via chromothripsis.
Different possible mechanisms, how chromothripsis can change the status of the carrier cell. The cell goes into apoptosis after chromothripsis (a). Fusion genes, loss of tumor suppressor genes or amplification of oncogenes can change the cell status to a cancer cell (b-d). (Figure is adapted from Rode *et al*. 2016 [138]).

The chromothripsis event can lead to different outcomes and can prolong the development of cancer in different ways. The different mechanisms are summarized in the figure above (Figure 9).

The majority of cells undergoing chromothripsis are eliminated by apoptosis or clonal outgrowth (Figure 9 A). A vast disruption of chromosomes leads directly to cell death. Chromothripsis can lead to fusion genes by a random rejoining of shattered chromosome fragments. These fusion genes can lead to the activation of oncogenes or the generation of proteins with new oncogenic function (Figure 9 B). Or the disruption of tumor suppressor genes can occur. The event results in massive DNA loss. The loss of a tumor suppressor or a part of the gene can lead to an inactivation of the tumor suppressor gene (Figure 9 C). The opposite is also possible, the non-joining material can form a so-called double-minute chromosome. The chromosome segment can readily amplify and with an oncogene inside it can affect the cell negatively. All these possible effects of chromothripsis can lead to a cancer cell (Figure 9 D).

In PDAC, chromothripsis leads to significant genetic instability and, subsequently, worse clinical outcomes for patients whose tumors had at least one such event. The main effects are the loss of tumor suppressor genes like *CDKN2A, TP53*, and *SMAD4*[139].

1.3.3 DNA Methylation in cancer

DNA methylation is an epigenetic process, in which the DNA itself is changed, in contrast to histone modifications, in which the genetic regulation takes place by changing the histone proteins, and thus, the chromatin structure.

The DNA methylation consists of the reversible (covalent binding) addition of a methyl group to the carbon 5'-position of cytosine[140]. In a mammalian genome, the binding only occurs, when the cytosine base is followed by a guanine base, a so-called CpG site (5'-C-phosphate-G-3' site)[141]. The enzyme family of DNA-methyltransferase (DNMTs) catalyzes the transfer of a methyl group to cytosine (Figure 10 A). The process is associated with stable gene silencing and regulation of gene expression.

The location of the CpG sites is divided into 4 different types. The CpG islands (CGIs) are defined as regions > 500 bp and with > 55 % GC content, consist of usually unmethylated CpG-dinucleoties[142]. In general, CGIs are enriched in promoter regions and/or the first exon region. 72 % of human promoters are characterized by high-CpG content[143], the CpG density itself does not influence gene expression. In these regions, for example, the 5' end of housekeeping genes are located, so the gene can be constitutively active.

The CpG shores are the transition regions and are about ~2 Kb from islands. More than 75 % of the tissue-specific differentially methylated regions are found on the shores. Methylation in shores shows a higher correlation with gene expression than CpG islands[144].

The CpG shelves are ~4 Kb away from the CpG Islands and are usually methylated. The region of the ocean covers the biggest part of the genome, where the CpG density is low. However, a few CpG sites in the ocean are normally strongly methylated and belong to repetitive elements or transposons[145].

DNA-methylation plays a role in the tumor progression. Epigenetic alterations are flexible alternatives that can be maintained over several cell divisions to adapt to the cellular phenotype. Genetic changes and epigenetic changes influence each other and establish the typical characteristics of tumor cells[146].

Different mechanisms of DNA-methylation are involved in the tumor progression (Figure 10 B). The hypomethylation of the region in heterochromatin structures can lead to genomic

instability. The next mechanism is the hypermethylation of promoter regions, which leads to transcriptional repression and imbalance in the cell. Therefore, an increased expression of DNA methyltransferases may contribute to tumor progression through a methylation-mediated gene inactivation in various human cancers[147].

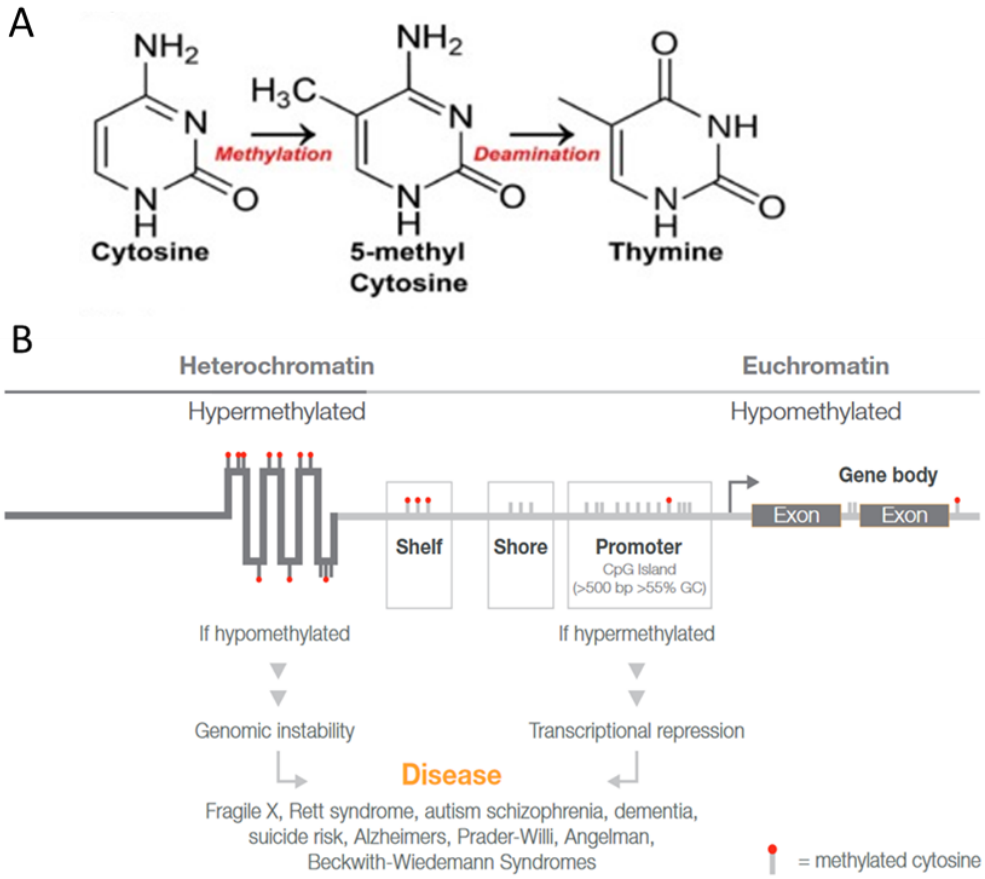


Figure 10: Mechanism of cytosine methylation and perturbation of methylation.
(A) Transfer of a methyl group (-CH_3) to the 5' carbon atom of a cytosine. This reaction is catalyzed by DNA methyltransferase (DNMT). Deamination is the removal of a methyl group of cytosine, resulting in thymine. (B) The localization of methylated CpGs is very important for the resulting effect in the cell. Hyper- or hypomethylation in a defined region can lead to genomic instability or transcriptional repression. (Figure adapted from Menezo *et al.* 2016 [148] and Illumina 2016 [149].

The other involved mechanism is the deamination, which can indirectly cause point mutation in the DNA. The process of deamination is the removal of a methyl group of a cytosine. When a deaminase spontaneously removes the amino group of a 5-methyl-cytosine by hydrolyses, it becomes a thymine. A mismatch between thymine and guanine is established. When the repair mechanism does not repair the mismatch, in the next cell cycle a point mutation is established. In rare cases, it is possible that an unmethylated cytosine is deaminated and becomes uracil. The mismatch between uracil and guanine is much often repaired under normal conditions because the uracil does not naturally occur in the DNA structure. It is estimated that this form of epigenetic mutation in human tumor diseases affects a quarter of

all *TP53* mutations[150]. In addition, the methylation status is associated with drug sensitivity in different cancer types like B-cell lymphoma and glioma[151].

The different DNA-methylation pattern in cancer versus normal tissue can be used as a biomarker. The first observation of different methylation in cancer versus normal tissue was done by Feinberg and Vogelstein in 1983. They showed that in cancer cells a lower global DNA methylation is detected than in healthy cells[152].

Nowadays, there exists a specific methylation pattern for different cancer types. For example, for colorectal cancer, a pattern of methylation can be used for the detection of precursor lesions from stool or liquid biopsy samples[153]. For the detection of pancreatic cancer some methylation profiles are available[154]. Cell-free tumor DNA from blood samples of pancreatic cancer patients is analyzed. These biomarkers can be used for the detection of pancreatic cancer in an early stage. It is possible to screen patients with high risk with these methylation patterns. The detection of precursor lesions, before the progression to cancer is completed, is not possible. For that the precursor lesions PanIN, IPMN, and MCN have to be further investigated molecularly.

2 Aim

The aim of this study was a detailed characterization of the common precursor lesions of the pancreatic ductal adenocarcinoma. Mutational profile, transcriptome profile, methylome profile, and copy number profile were performed for the investigation of the most common precursor lesions of the PDAC. Gastric IPMN, intestinal IPMN, and PanIN lesions were in the focus of the multiomic study in comparison to each other.

The first part of this project was the distinction of PanIN and gastric IPMN, since it is still unknown whether PanIN and gastric IPMN represent a small/large version of the same lesion or distinct lesions. PanIN and gastric IPMN are mostly localized in the peripheral duct system and are mainly distinguished based on their size. A further part of the project was the identification of a protein marker to distinguish the different precursor lesions. The clinical outcome is different in the different subtypes. The characterization was performed with targeted-NGS, whole-genome low coverage sequencing, RNASeq, and Methylation Array.

The goal was to improve the understanding of the progression to PDAC and to find a detection tool for the precursor lesions. The marker should also be suitable for distinguishing the different precursor lesions. A suitable biomarker will give advantages in clinical practice and improve the information for the patient according to the survival rate and prognosis.

3 Materials and Methods

3.1 Equipment

The used equipment, during this experimental study, are listed in the following table (Table 1).

Table 1: Equipment.

Device	Company
DEPArray™ System	Menarini silicon biosystems, Castel Maggiore, Italy
Eppendorf Bio-Photometer 6131	Eppendorf, Hamburg, Germany
Eppendorf Centrifuge 5417R	Eppendorf, Hamburg, Germany
Eppendorf Mastercycler PCR device	Eppendorf, Hamburg, Germany
Eppendorf Minispin Plus centrifuge	Eppendorf, Hamburg, Germany
Eppendorf Research Pipettes	Eppendorf, Hamburg, Germany
Eppendorf Thermomixer 5436	Eppendorf, Hamburg, Germany
FlexCycler PCR device	Jena Analytik AG, Jena, Germany
Fragment Analyzer systems	Agilent technologies, Santa Clara, USA
Ion OneTouch ES	ThermoFisher Scientific, Waltham, USA
Ion OneTouch™ 2 Instrument	ThermoFisher Scientific, Waltham, USA
Ion Torrent S5	ThermoFisher Scientific, Waltham, USA
LEICA CM1950 Cryostat	Leica Biosystems, Wetzlar, Germany
Magnetic Stirrer MS 3000	Biosan, Riga, Latvia
Microwave Oven	Commercial
NovaSeq6000 S4 Sequencing	Illumina, San Diego, USA
Palm Microbeam	Carl Zeiss, Jena, Germany
PH-Meter HI2020 edge®	Hanna instruments, Woonsocket, USA
QUANTUM ST5 gel documentation	PEQLAB Biotechnologie GmbH, Erlangen, Germany
Qubit® 2.0 Fluorometer	ThermoFisher Scientific, Waltham, USA
Semi-automatic precision microtome CUT 5062	Slee medical GmbH, Mainz, Germany

Shaking Water Bath	GFL Gesellschaft für Labortechnik mbH, Burgwald, Germany
Sprout® Mini-Centrifuge 100-240VAC	Heathrow Scientific, Vernon Hills, Germany
StepOnePlus™ Real-Time PCR System	Applied Biosystem, Forster City, USA
TissueLyser LT	Qiagen, Hilden, Germany
Tissue-Tek Film® Coverslipper E2	Sakura Finetek Germany GmbH, Staufen, Germany
Tissue-Tek Prisma® Slide Stainer E2S	Sakura Finetek Germany GmbH, Staufen, Germany
Vacuum Centrifuge	Eppendorf, Hamburg, Germany
Vortex Genie 2	Bender&Hobein AG, Zürich, Switzerland
Zeiss Axio Imager M1 fluorescence microscope	Carl Zeiss, Jena, Germany

3.2 Kits and enzymes

The molecular biological kits and enzymes, which was used during this thesis are listed in the following table (Table 2).

Table 2: Kits and enzymes.

Name	Company	Article-number
Agencourt AMPure XP	Beckman Coulter, Brea, USA	A63880; A63881
Ampli1 Low-pass Kit	Menarini silicon biosystems, Castel Maggiore, Italy	WGLPTA
Ampli1 QC Kit	Menarini silicon biosystems, Castel Maggiore, Italy	WGQC4
Ampli1 WGA Kit	Menarini silicon biosystems, Castel Maggiore, Italy	WG001R
GeneRead DNA FFPE Kit	Qiagen, Hilden, Germany	180134
Human Genomic DNA	Promega Coporation, Madison, USA	G304A
Ion 520™ & Ion 530™ Kit-OT2	ThermoFisher Scientific, Waltham, USA	A27751
Ion 520™ Chip Kit	ThermoFisher Scientific, Waltham, USA	A27762
Ion 530™ Chip Kit	ThermoFisher Scientific, Waltham, USA	A27764
Ion AmpliSeq™ Library Kit 2.0	ThermoFisher Scientific, Waltham, USA	4475345; 4480441
Ion Library TaqMan™ Quantitation Kit	ThermoFisher Scientific, Waltham, USA	4468802
MseI	New England Biolabs GmbH, Frankfurt/Main, Germany	R0525S
Power SYBR™ Green PCR Master Mix	Applied Biosystem, Forster City, USA	4367660
QIAamp DNA Micro Kit	Qiagen, Hilden, Germany	56304

Qubit™ dsDNA BR Assay Kit	ThermoFisher Scientific, Waltham, USA	Q32850
Qubit™ dsDNA HS Assay Kit	ThermoFisher Scientific, Waltham, USA	Q32851
Qubit™ RNA HS Assay Kit	ThermoFisher Scientific, Waltham, USA	Q32852
RevertAid H Minus First Strand cDNA Synthesis Kit	ThermoFisher Scientific, Waltham, USA	K1632
RNAlater Stabilization Solution	ThermoFisher Scientific, Waltham, USA	AM7020
RNaseOUT	ThermoFisher Scientific, Waltham, USA	10777019
RNeasy FFPE Kit	Qiagen, Hilden, Germany	73504
RNeasy Plus Micro Kit	Qiagen, Hilden, Germany	74034
RNeasy Plus Mini Kit	Qiagen, Hilden, Germany	74136
SPRIselect beads Reagent	Beckman Coulter, Brea, USA	B23317
TaqMan® RNase P Detection Reagents Kit	ThermoFisher Scientific, Waltham, USA	4316831
UroVysionTM™ multi-colour FISH probe	Abbott GmbH, Wiesbaden, Germany	02J27-020
VECTASTAIN® EliteABC HRP Kit (Peroxidase, Rabbit IgG)	Vector Laboratories Inc., Burlingame, USA	PK-6101

3.3 Chemicals/ Reagents

All used chemicals and reagents used during this study are listed below (Table 3).

Table 3: Chemicals and Reagents.

Reagent	Company
Agarose Gel Ultra Pure	ThermoFisher Scientific, Waltham, USA
Albumin Fraction V	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
BSA	Sigma Aldrich, St. Louis, USA
Citric acid	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Collagenase II	Merck, Darmstadt, Germany
Crystal violet	Sigma Aldrich, St. Louis, USA
DAB Substrate Kit	Abcam, Cambridge, United Kingdom
DAPI	Abbott GmbH, Wiesbaden, Germany
DEPC (Diethylpyrocarbonat)	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
DNA/RNA dye, peqGREEN	VWR International, Radnor, USA
EDTA disodium salt	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Eosin	Sigma Aldrich, St. Louis, USA
Ethanol	VWR International, Radnor, USA
Ethanol 70 %	VWR International GmbH, Darmstadt
Ethanol 99.5 % denatured with 1 % MEK	VWR International GmbH, Darmstadt
Eukitt® Quick-hardening mounting medium	Merck, Darmstadt, Germany
Formaldehyde solution 4 % phosphate buffered	NeoFroxx GmbH, Einhausen, Germany
Glacial / Acetic acid	Merck, Darmstadt, Germany
Hematoxilin	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Hydrogen peroxide	Merck, Darmstadt, Germany
Low TE buffer	ThermoFisher Scientific, Waltham, USA
Meyer's hemalum solution	Merck, Darmstadt, Germany

Normal Goat Serum	Abcam plc, Cambridge (UK)
OCT embedding medium for frozen section	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Paraffin	Engelbrecht GmbH, Edermünde, Germany
PBS	Gibco, Carlsbad, USA
Pretreatment buffer	Abbott GmbH, Wiesbaden, Germany
Protease I	Abbott GmbH, Wiesbaden, Germany
Proteasebuffer	Abbott GmbH, Wiesbaden, Germany
RPMI	Gibco, Carlsbad, USA
Sodium citrate	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Tris	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Tween® 20	Merck, Darmstadt, Germany
Water	Braun GmbH, Kronberg, Germany
Xylene	VWR International, Radnor, USA

3.4 Buffer

The receipts of different buffers and solutions are listed in the following table (Table 4).

Table 4: Buffer.

Name	Composition
100x DNA staining solution	1 mM DAPI in ddH2O
Dissociation solution	1 mg/mL Collagenase solved in RPMI
10x HIAR	0.1 M sodium citrate solution
	0.1 M citric acid solution to adjust the pH to 6.40
50x TAE	50 mM EDTA disodium salt
	2 M Tris
	1 M glacial / acetic acid
	adjust to pH 8.0
PBS-T (washing buffer)	PBS + 0.05 % Tween
PBATw buffer	10 mg/L BSA solved in PBST
DEPC-H20	0.1 % (w/v) DEPC in H2O overnight stirring, autoclaved
1 % Crystal violet solution	1 % crystal violet (v/W) in 50 % ethanol
20x SCC buffer	3M NaCL
	300mM Trisodium citrate
	adjust to pH 7.0
Post hybridisation buffer	2xSCC Buffer
	0.3 % NP40
Citrate Buffer	10 mM Citric acid
	adjust to pH 6.0
TBS (10x)	1400 mM NaCl
	250 mM TRIS
	adjust pH to 7.6
TE buffer	10 mM Tris
	1 mM EDTA
	adjust pH to 9.0

3.5 Primer

The self-designed primer are listed below (Table 5). A detailed list of the different amplicons of the two self-designed NGS Panels are in the attachment (chapter 11.1 and 11.2).

Table 5: Primer.

Name	Sequence (5'-3')
HML-2_for	AAACGCCAATCCTGAGTGTC
HML-2_rev	CATAGCTCCTCCGATTCCAT
PDAC#1 Panel	IAD111237_1 (194 amplicons) (see attachment 11.1)
PDAC#2 Panel	IAD111237_231 (217 amplicons) (see attachment 11.2)

3.6 Antibodies

The antibodies, which are used during this study are listed below (Table 6).

Table 6: DEPArray antibodies.

Name	Isoform	Article-number	Company
anti-Cytokeratin MNF116	IgG1	M082101	Agilent Dako, Santa Clara, USA
anti-Cytokeratin AE1/AE3	IgG1	MAB3412	Merck Millipore, Burlington, USA
anti-Vimentin Vim 3B4	IgG2A	M702001	Agilent Dako, Santa Clara, USA
Alexa Fluor 488®	IgG1	A-21121	ThermoFisher Scientific, Waltham, USA
Alexa Fluor 647®	IgG2A	A-21241	ThermoFisher Scientific, Waltham, USA

Table 7: IHC antibodies.

Antibody	Clone	Company
Anti-GNAS	ab58916	Abcam, Cambridge, United Kingdom
Anti-KLF4	EPR19590	Abcam, Cambridge, United Kingdom
Anti-Estrogen Inducible Protein pS2	EPR3972	Abcam, Cambridge, United Kingdom
Anti-MUC1	695	Biocare Medical, Pacheco, USA
Anti-MUC2	CCP58	Agilent Dako, Santa Clara, USA
Anti-MUC5AC	CLH2	Chemicon, Temecula, USA
Anti-MUC6	DCM-20	DCS, Hamburg, Germany
Anti-CDX2	CDX-88	BioGenex, Fremont, USA
Anti-Ki-67	MIB-1	Agilent Dako, Santa Clara, USA

3.7 Software and databases

Different software's and databases are used during this study. A detailed list is present in the table below (Table 8).

Table 8: Software.

Software/ Database	Company/ Website
Bowtie2 (v2.3.4.1)	https://ccb.jhu.edu/software/hisat2/index.shtml
cBioportal for cancer genomics	https://www.cbioportal.org/
ClustVis	https://biit.cs.ut.ee/clustvis/
CTLPScanner (© 2018 Cai Laboratory)	http://47.88.3.162/CTLPScanner/index.php
Enrichr	https://amp.pharm.mssm.edu/Enrichr/
Ensembl database	http://www.ensembl.org/index.html
FASTQC (v0.11.7)	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Gene List Venn Diagram	https://www.bioinformatics.org/gvenn/index.htm
Graphpad Prism 5.0	GraphPad Software Inc, San Diego, USA
HISAT2 (v2.1.0)	https://ccb.jhu.edu/software/hisat2/index.shtml
Ion AmpliSeq™ Desinger	https://ampliseq.com
Ion Reporter Software (v5.12.0.0)	ThermoFisher Scientific, Waltham, USA
KEGG 2019 human	https://www.kegg.jp/
Mendeley	Elsevier, Amsterdam, Netherland
MTR Viewer	http://mtr-viewer.mdhs.unimelb.edu.au/mtr-viewer/
Panther (v14.1)	http://pantherdb.org/
PERViewer	http://per.broadinstitute.org/
R (v3.5.1)	https://www.r-project.org/
R package TCC-GUI	https://github.com/swsoyee/TCC-GUI
R studio (v1.1.456)	https://rstudio.com/
ShinyCircos	https://yimingyu.shinyapps.io/shinycircos/
STRING (v11.0)	https://string-db.org/
StringTie (v1.3.4d)	https://ccb.jhu.edu/software/stringtie/

Trimmomatic (v0.38)	http://www.usadellab.org/cms/?page=trimmomatic
UCSC Genome Browser	https://genome.ucsc.edu/
UCSC Xena Functional Genomics Explorer	https://xenabrowser.net/
VarSome ACMG Implementation (v7.3.7)	https://varsome.com/
Aperio ImageScope (v12.3.2.8013)	Leica Biosystems, Wetzlar, Germany

3.8 Patient cohort

The tissue collection was obtained from 120 patients with precursor lesions of PDAC. Patients underwent pancreatic surgery between 2008 and 2019 and were diagnosed at the University Hospital in Duesseldorf. All cases were re-evaluated and immunohistochemistry was performed (MUC1, MUC2, MUC5AC; Figure 11)[155].

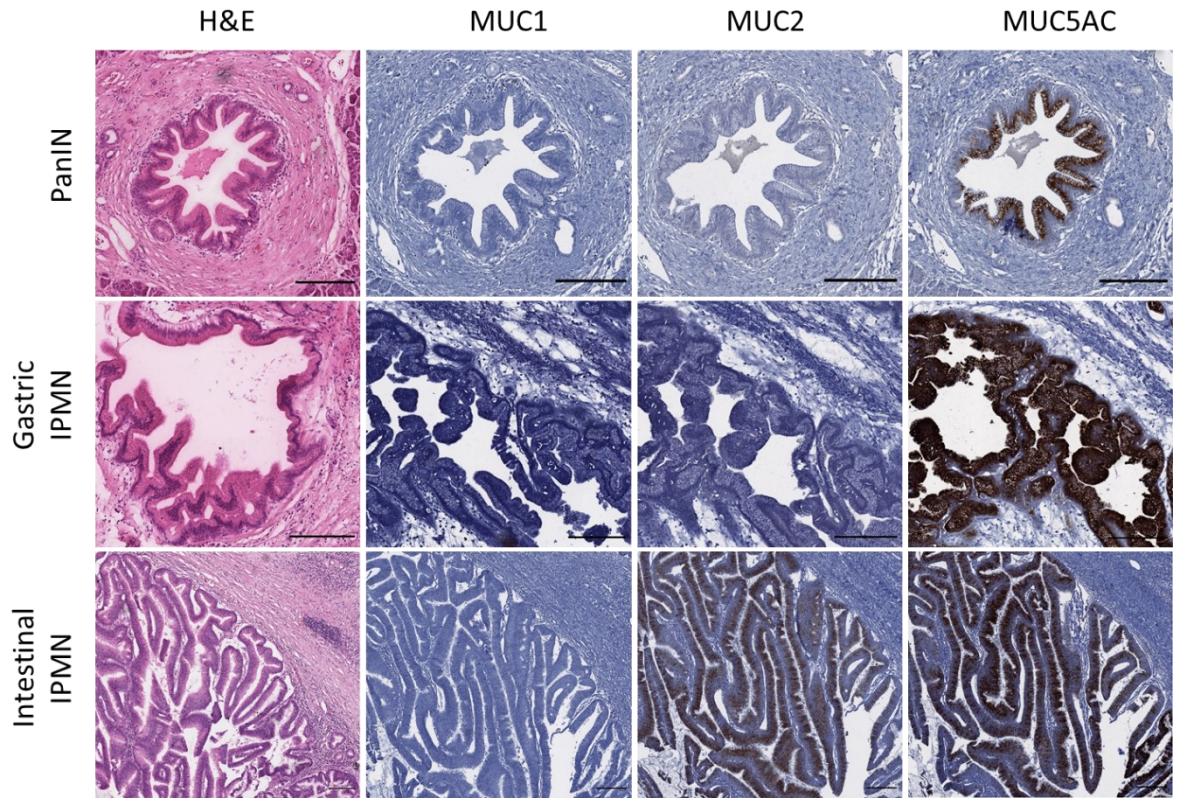


Figure 11: Morphology and immunohistology profile of gastric IPMN and intestinal IPMN.
Representative PanIN and gastric IPMN are distinguished according to morphology and size and display an identical immunohistochemical profile with diffuse positivity for MUC5AC, no expression of MUC1 and MUC2. Intestinal IPMN are clearly distinct lesions, both on the morphological and immunohistochemical level, characterized by positivity for MUC2 and MUC5AC. (scale bar= 200µm).

The ethics committee of the Heinrich-Heine University and the University Hospital of Duesseldorf approved the use of formalin-fixed, paraffin-embedded (FFPE) tissues for retrospective analyses (ethic no. 3821).

The study cohort consisted of 139 different precursor lesions and includes both PanIN (n= 56) and IPMN (n= 62). Fourteen AFL (11 %) were included as well (Table 9).

Table 9: Study cohort.

Diagnosis	Number of cases [%]	Degree of dysplasia		PDAC
		low grade	high grade	
PanIN	56/125 (44.8 %)	44/56 (78.6 %)	12/56 (21.4 %)	36/56 (64.3 %)
IPMN gastric	38/125 (30.4 %)	31/38 (81.6 %)	7/38 (18.4 %)	15/38 (39.5 %)
IPMN intestinal	21/125 (16.8 %)	11/21 (52.4 %)	10/21 (47.6 %)	7/21 (33.3 %)
IPMN pancreatobiliary	3/125 (2.4 %)	1/3 (33.3 %)	2/3 (66.7 %)	2/3 (66.6 %)
IPMN mixed histological subtype	7/125 (5.6 %)			1/7 (14.3 %)

3.9 Histology Methods

3.9.1 Tissue microarray cohort

Tissue microarrays (TMA) were done by Marie Schulte (medical student) for her thesis. The cohort consisted of some new patients and some patients from the consisting cohort. The following figure shows the distribution of the precursor lesion in this cohort.

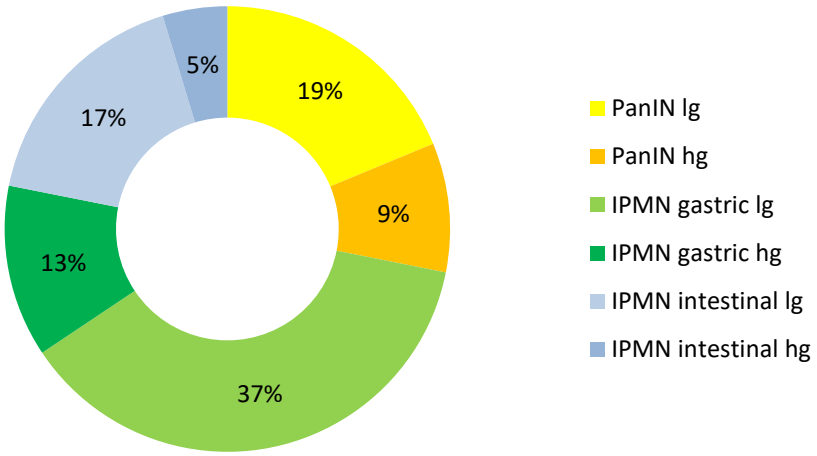


Figure 12: Precursor lesion distribution of the tissue microarray.
The main lesion with 50 % is the gastric IPMN, followed by PanIN with 28 % and intestinal IPMN with 22 %. (n= 61).

The cohort consisted of gastric IPMN (n= 32; low grade n= 24; high grade n= 8), intestinal IPMN (n= 11; low grade n= 3; high grade n= 8) and PanIN lesions (n= 18; low grade n= 12; high grade n= 6). These samples were only used for immunohistology staining (see chapter 3.9.4).

The TMA was done with a two-millimeter punch in diameter. The space between each sample was 0.5 mm. Four different TMAs were done with up to 66 samples on one block. As control samples, normal tissue, tumor central, and tumor periphery were used from each patient.

3.9.2 Hematoxylin and Eosin (H&E) Staining

Hematoxylin and Eosin (H&E) staining were used for the basic evaluation of tissue histology. For the staining, a 4 µm thick FFPE tissue section was prepared on a glass slide. To reduce the paraffin from the section, the slides were incubated for about 1 hour at 60 °C. The H&E staining was done by an automated staining system from Sakura with the following standard protocol.

First, the paraffin tissue was incubated 2 times for 3 minutes in xylol for deparaffinization. Then, the slides were rehydrated in a decreasing alcohol series with 100 %, 95 % and 70 % ethanol for 30 seconds per concentration. An incubation in water for 30 seconds was the last step of the series. Subsequently, the slides were stained twice with hematoxylin each for 3 minutes. A short water incubation followed by 30 seconds. After that, the counterstaining with Eosin was done for 2:45 minutes.

After the two staining steps, an increasing alcohol series with 70 %, 95 % and 100 % ethanol followed. The last step was incubation in xylol twice for 30 seconds and 2 minutes. After that, the slides were automatically covered with a cover film from Sakura.

3.9.3 Crystal violet staining

Crystal violet staining was used to stain the tissue for the microdissection with the Palm Microbeam (see chapter 3.9.6). Crystal violet reduces the damage to the integrity of RNA and DNA during the staining procedure[156]. For isolation, crystal violet staining was used instead of the standard H&E staining. For the crystal violet staining, 8 µm thick paraffin sections were prepared.

The staining started with incubation at 60 °C for one hour. Then, the paraffin sections were incubated 2 times for 5 minutes in xylol. Subsequently, the slides were incubated in a decreasing alcohol series with ethanol concentrations of 100 %, 96 % and 70 %, per concentration for 1 minute. The next step was incubation in 1 % crystal violet solution for 30 seconds. This was followed by a short increasing alcohol series with 70 % and 100 % ethanol each for 15 seconds. In the end, the slides were air-dried shortly. Freshly stained slides were

used directly at the Palm Microbeam for microdissection. All solutions were prepared with DEPC-treated water to reduce the damage to the RNA.

3.9.4 Immunohistochemistry

The immunohistology markers were used on TMA. For this, 1.5 µm thick paraffin sections were incubated overnight at 60 °C in the oven. Afterwards, the sections were deparaffinized in xylol and rehydrated in a descending alcohol series ending in water. Antigen retrieval was performed in sodium citrate buffer with a pH of 6.0 or TE buffer with a pH of 9.0 in a pressure cooker for 10 minutes. Afterwards, the slides were incubated in the buffer at room temperature for 30 minutes to cool down. The sections were washed in TBS-T twice for 5 minutes and were incubated in 3 % H_20_2 to block the endogenous peroxidase, followed by two washing steps with TBS-T each for 5 minutes. To avoid non-specific cross-linking of the antibody, the slides were incubated for 30 minutes in blocking solution, followed by a washing step.

The next step was incubation for 1 hour in the primary antibody, which was dissolved in blocking solution, followed by two washing steps. The antibodies were diluted according to table 10. Afterwards, the secondary antibody was incubated for 30 minutes. The slides were washed again two times and the detection system from the ABC-Vectastain kit was added for 30 minutes. The slides were washed again twice in TBS-T and incubated with DAB for a duration of 30-45 seconds. The color reaction was stopped by washing the slides in tap water and counterstained with Meyer's hematoxylin for 20 seconds. The slides were subsequently rinsed with running tap water. Finally, slides were dehydrated in an ascending alcohol series and coverslips were mounted onto slides using a mounting medium.

Table 10: Application of IHC antibodies.

Antibody	Type	Dilution	Antigen Demasking
Anti-GNAS	Rb Poly	1:200	Citrate buffer
Anti-KLF4 antibody [EPR19590]	Rb Mono	1:1000	TE buffer
Anti-Estrogen Inducible Protein pS2 antibody [EPR3972]	Rb Mono	1:200	Citrate buffer
Anti-MUC1	Mo Mono	1:100	CC1 buffer
Anti-MUC2	Mo Mono	1:100	CC1 buffer

Anti-MUC5AC	Mo Mono	1:1000	CC1 buffer
Anti-MUC6	Mo Mono	1:100	CC1 buffer
Anti-CDX2	Mo Mono	1:40	CC1 buffer
Anti-Ki-67	Mo Mono	1:100	CC1 buffer

*Rb: rabbit.Mo: mouse Mono: monoclonal. Poly: polyclonal. CC1: Cell Conditioning 1 (Ventana Medical System Inc., Tucson, USA).

Marker MUC1, MUC2, MUC5AC, MUC6 and CDX2 were performed on whole specimens, the marker Ki-67 and some MUC6 staining's were performed on TMAs with the XT ultraView™ DAB v3 Detection Kit on the Ventana BenchMark ultra-automated IHC/ISH slide staining system (Ventana Medical Systems Inc., Tucson, USA) according to the established protocol for the routine diagnostic of the Institute of Pathology at the University Hospital Duesseldorf.
The immunohistochemistry score of the different markers was done according to the percentage of positively stained cells, a score 0-4 (0= negative; 1= <10 %, 2= 10- 50 %, 3= 51- 80 % and 4= >80 %) was assigned. The staining intensity was evaluated according to a three-tiered system (1+, weak; 2+, moderate and 3+, strong). The immune reactive score (IRS) was obtained by multiplying the two scores and ranged from 0 to 12 (IRS 12= strong intensity in >80 % of the cells). For the GNAS antibody, only the staining intensity was evaluated (weak, moderate, or strong).

3.9.5 Microdissection (manual)

For the manual microdissection, 8 µm thick paraffin tissue sections were prepared on glass slides without any preparation. One HE-stained slide was used as a control with marked the lesions by a pathologist. This template was used to scratch manual with sterile filter tips the lesion from the glass slide. The scratched paraffin with tissue was collected in a 1.5 mL reaction tube with deparaffination solution from the DNA or RNA Isolation kit (Qiagen).

3.9.6 Microdissection with Palm Microbeam

For the microdissection with the Palm Microbeam (Zeiss), 6- 8 µm thick paraffin tissue sections were prepared and previously stained with H&E or with crystal violet (see 3.9.2 and 3.9.3).
For the isolation of DNA or RNA, normal glass slides with stained paraffin sections were used. The dry collection method from Zeiss was performed. The glass slides were scratched off by

laser impulses. This method is called Laser Pressure Catapulting (LPC). During this method, the tissue was cut into small fragments. With the CapMover II, adhesive caps were moved directly over the slides while cutting so that the small fragments were collected directly by catapulted in the cap of the 0.2 mL reaction tube with the special adhesive cap. The tissue was cut with the 5x or 20x objective of the microscope.

3.10 Molecular biology methods

To investigate the molecular pattern from different precursor lesions, several different molecular biology methods had to be carried out. In the following chapter, the methods and the procedure will be explained.

3.10.1 Targeted Gene Sequencing by Next-Generation-Sequencing (NGS)

3.10.1.1 DNA Isolation from FFPE tissue

To isolate the DNA from the FFPE section, the GeneRead FFPE Kit from Qiagen was used. The procedure was done according to the manufacture's guide.

To yield enough genomic DNA for further downstream analysis, 5-8 slides with 8 µm thick tissue sections were used as starting material. Manual microdissection or dissection with the PALM Microbeam microscope from Zeiss was done (see chapter 3.9.6). By the manual dissection, the scratched tissue section was directly transferred with a tip to the deparaffinization solution from the DNA GeneRead kit.

A xylol free, deparaffinization step was done first, followed by a proteinase k digestion at 56 °C for 1 to 2 hours. For termination of the digestion step, incubation for 1 hour at 90 °C was done in a thermoblock. In addition to other genomic FFPE DNA isolation Kits, this kit worked with a uracil-DNA glycocylase (UDG), which reduce the C>T (and G>A) sequence artefacts. In DNA from FFPE samples, cytosine deamination to uracil is a common DNA damage. For this, incubation for 1 hour was performed with the UDG enzyme at 50 °C.

The digestion was followed by a column-based cleanup. First, the column was loaded with a suspension out of AL buffer, lysis product, and ethanol. Afterwards, the columns were washed with the AW1 and AW2 buffers from Qiagen. The last washing step was with ethanol absolute. The elution was done with an ATE buffer. Photometric measurement of the isolated genomic DNA was done at a wavelength of 260 nm for dsDNA to determine the concentration and purity. With this measurement, the contamination from salts and proteins can be measured at a wavelength of 230 nm and 280 nm. The ratio of the values gives a hint of purity.

3.10.1.2 Quality control of genomic DNA from FFPE

Before the preparation of the library was done, the isolated genomic DNA from FFPE samples was checked via quantitative real-time polymerase chain reaction (qRT-PCR).

The commercially available TaqMan™ RNase P Detection Reagents Kit was used to determine the quality and quantity of the genomic human DNA from FFPE samples. This kit was used to

amplify the Ribonuclease P *RNA component H1* (*H1RNA*)-gene (*RPPH1*) on chromosome 14. The produced fragment is about 98bp long.

Furthermore, a self-designed Primerset HML-2 (human mammary tumor virus like-2; human endogenous retroviruses) was used for quantification of the DNA concentration from FFPE samples from humans. These Primers were designed to produce an amplification product of about 120 bp, which reduced the false produced libraries afterwards. For each qRT-PCR, a standard curve with human control DNA was prepared with the concentrations of 1 ng, 5 ng and 20 ng. In the following table (Table 11), the qRT- PCR programs are displayed for the two different assays.

Table 11: qPCR program of the two different quantification programs for the DNA quality.

RNase P (TaqMan)	HML-2 (SYBR Green)
50 °C 2 min	95°C 10min
95 °C 10 min	40 X 95°C 15 sec
40 X 95 °C 15 sec	57°C 20 sec
60 °C 1 min	72°C 30 sec
Melting curve	Melting curve

3.10.1.3 Amplicon-based library preparation

The concentration of amplifiable DNA was very low. For this reason, the volume was reduced by a vacuum centrifuge, to increase the used genomic DNA concentration for library preparation.

To produce a library for next-generation sequencing (NGS), about 10 ng of good quality DNA was necessary for each Primer Pool. The AmpliSeq™ Library 2.0 kit (ThermoFisher Scientific) was used for the preparation of all amplicon-based NGS libraries.

During this study, two self-designed panels were used: PDAC#1 and PDAC#2. Both were designed with the Ion AmpliSeq™ Designer (v5.6, ThermoFisher Scientific). The first panel was designed with four primer pools with 194 amplicons in total, the second panel was designed with two primer pools with 217 amplicons. In the following table is the list of genes, which are covered within the panel (Table 12).

Table 12: List of genes involved in the PDAC gene panel.
The grey marked genes are nearly completely covered by the amplicons. The coverage of the amplicons is displayed in the attachment for panel PDAC#1 and PDAC#2.

ARID1A	*FBXW7*	*KRAS*
NRAS	*APC*	*IDH2*
ALK2	*EGFR1*	*TP53*
IDH1	*BRAF*	*RNF43*
VHL	*CDKN2A*	*SMAD4*
CTNNB1	*PTEN*	*STK11*
PIK3CA	*FGFR2*	*GNAS*

The covering of each gene of all panels is in the attachment (chapter 11.1 and 11.1). The genes *TP53*, *RNF43,* and *ARID1A* (AT-Rich Interaction Domain 1A) were nearly completely covered in PDAC#1 and PDAC#2.

The preparation of an NGS library began with a multiplex PCR with two or four primer pools depending on the used panel. For each primer pool, a 10 µL approach was used. In the following list, the 10 µL approach is shown.

10 µL library preparation: 5 µL Ion AmpliSeq™ Primer Pool
3 µL DNA
2 µl Ion AmpliSeq™ HiFi Mix

Table 13: PCR program NGS.

Step	Temperature	Time	Cycle
Activation	99°C	2 min	
Denaturation	99°C	15 sec	x 25 Cycles
Annealing	60°C	4 min	
	10°C	hold	

After the PCR (Table 13), the partial amplicon digestion was done with the enzyme FuPa. This enzyme was used to digest the primer sequences and phosphorylated the amplicons for the adaptor ligation in the next step. At each individual sample, the adapter P1 and a barcode between 1 and 96 were added. The P1 adaptor is responsible for the Ion Sphere particles (ISPs) and the barcode is for determining the different samples later in the library pool. Following, the sequence for the adapter P1 is shown.

Adapter Sequence P1: 5’ - CCTCTCTATGGGCAGTCGGTGAT - 3’

The next step was the library purification with a bead-based system (Agencourt AMPure XP; Beckman Coulter). To remove the primer waste and make a size selection, the library was cleaned with magnetic beads. For this, 1.5x volume was used for purification.

At the end of the library preparation, the quality and quantity were evaluated with Ion Library TaqMan™ Quantitation kit with a standard curve of 0.68 pM.

Afterwards, the library was diluted dependent on the number of amplicons to a final concentration between 6 pM and 8 pM and pooled with other barcoded libraries together.

Production of ISP particle with the Ion OneTouch™ 2 Instrument was performed by the technical staff of the molecular pathology. The procedure was done according to the manufacture’s manual of Ion 520™ & Ion 530™ Kit-OT2. The samples were loaded onto an Ion 520™ or 530™ Chip and sequenced by the Ion S5™ System in the molecular pathology of the University Hospital Duesseldorf.

3.10.2 Low-coverage whole-genome sequencing

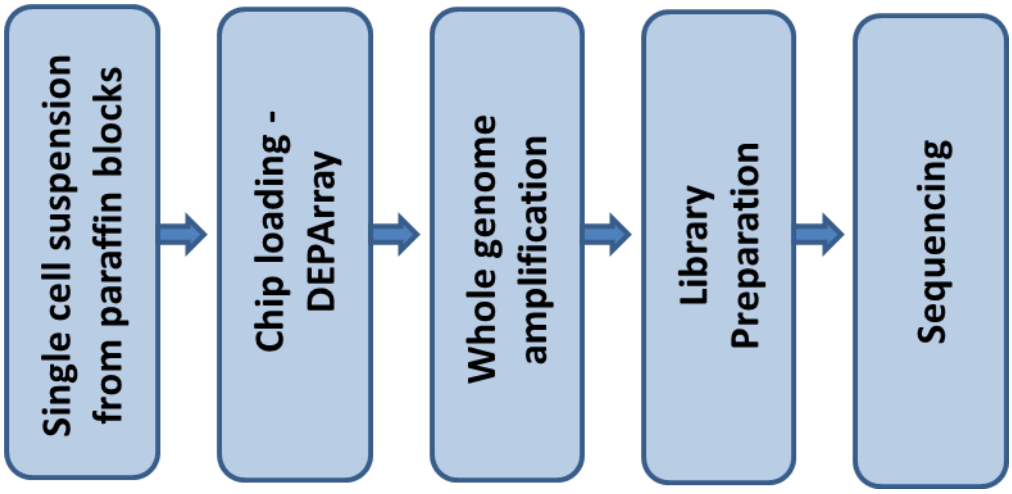


Figure 13: Pipeline Low-pass Sequencing.
The pipeline starts from DNA isolated from single cells sorted from the DEPArray or genomic DNA from manual microdissection from paraffin sections.

The low-pass sequencing is a workflow for the whole procedure of single cell whole-genome low-coverage sequencing (from Menarini). The pipeline of the workflow of the low-pass sequencing is shown in the picture above (Figure 13). The pipeline consists of 5 parts, (1) the single-cell suspension out of paraffin blocks, (2) the DEPArray™ cell sorting, (3) the whole-genome amplification, (4) the library preparation, and (5) the sequencing on the Ion torrent.

3.10.2.1 Single-cell suspension from paraffin blocks

The isolation of single cells from FFPE tissue was done with a modified protocol from Menarini diagnostic called “From FFPE tissue to DEPArray system” (version 1).

The starting material was a punch with a diameter of 1 mm from the targeted region. H&E staining was used to find the region of interest on the paraffin block. The punch was transferred into a nylon bag, which was closed via heating. The first step was the deparaffinization with 3 times Xylol each for 10 minutes. The nylon bag was in a 50 mL conical centrifuge tube and the surrounding solutions were always changed. This was followed by a decreasing alcohol series (100 %, 70 % and 50 %) 3 times each for 10 minutes. The last step of the alcohol series was incubation in water for 5 minutes. Then, HIAR buffer was used for 5 minutes at RT, followed by incubation for 1 hour at 80 °C in a water bath in preheated HIAR buffer.

After the incubation, the sample was cooling down for at least 20 minutes at RT. Afterwards, the nylon bag in the centrifuge tube was washed 3 times with RPMI medium without any

supplements under sterile conditions. The digestion of the tissue was done by collagenase digestion (1 mg/ml) in RPMI medium in a 37 °C shaking water bath overnight. On the next day, the cells were dissociated. To harvest the cells, the RPMI medium with cells inside were filtered through a nylon mesh with pores of 70 µm. The tubes and filter were washed with PBATw 3 times, centrifuged at 500 g for 5 minutes. The cells were resuspended in 500 µL PBATw buffer and transferred to a reaction tube. Isolated cells were stored at 4 °C until the staining procedure.

3.10.2.2 DEPArray™

The single-cell suspension was used further for the cell sorting with DEPArray™. The cell number was determined with a Neubauer counting chamber. About 500, 000 cells were used for the following staining procedure.

Antibody cocktail I:	anti-cytokeratin MNF116	3.2 µg/mL
	Anti-cytokeratin AE1/AE3	10 µg/mL
	Anti-vimentin Vim 3B4	3.1 µg/mL

The antibody cocktail I was incubated for 30 minutes at 4 °C, every 15 minutes the reaction tube was vortexed gently for 5 seconds. After the incubation, the cells were washed with PBATw buffer.

Antibody cocktail II:	AlexaFluor® 488 goat	2.5 µg/mL
	AlexaFluor® 647 goat	2.5 µg/mL

Then, the antibody cocktail II with the secondary antibodies was incubated for 1 hour at 4 °C in the dark, every 15 minutes the reaction tube was vortex gently for 5 seconds. Again, the cells were washed with PBATw buffer. In the last staining step, the nuclei staining with DAPI was done. For this, the DNA staining solution was incubated for 30 minutes at 37 °C in the dark. After the final staining step, the cells were washed 2 times with PBATw buffer and counted again in the Neubauer counting chamber. Until the final DEPArray™ sorting, the cells were stored at 4 °C for up to one week.

Loading the cartridge and run the DEPArray™ system was done by Elina Bongers, a technical assistant from the working group Experimental Surgical Oncology of Prof. Stoecklein at the University Hospital Duesseldorf.

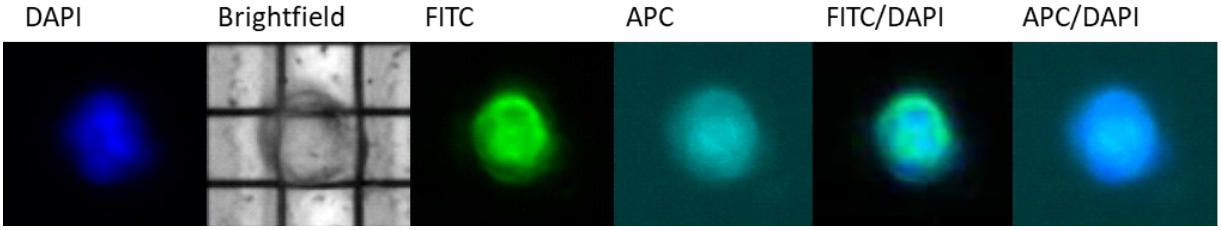


Figure 14: One example of the double positive recruited cells of the DEPArray™.
Single cell observation, recruiting and analysis is possible with the DEPArray™ system.

The upper picture (Figure 14) is shown as an example of a positively selected cell from the DEPArray™ system. The nuclei staining was represented by the DAPI staining (blue), the size and morphology were shown in the brightfield and the antibody positivity was shown by fluorescein isothiocyanate (FITC) emission for cytokeratin (green) and allophycocyanin (APC) emission for vimentin (cyan).

One single cell (or a small cluster) was flushed out into a single reaction tube. The cells could be stored at -20°C.

3.10.2.3 Whole-genome amplification

The whole-genome amplification (WGA) was done with the Ampli1™ WGA Kit from Menarini Biosystems. The Kit was used for the amplification of DNA from a single cell. The kit was also used for small lesions with more than one cell from FFPE samples.

The protocol was done according to the manufacture's protocol. A single cell or 2 µL genomic DNA was used as starting material.

The first step was the cell lysis, therefore a reaction mixed for at least 10 samples was prepared according to the following table (Table 14). For each sample, 2 µL cell lysis mix was added and incubated overnight at 42 °C for the single cells, and for the genomic DNA as starting material, the incubation was reduced to 1 hour. After the one hour or overnight incubation, 30 minutes incubation at 65°C followed. The cell lysis was terminated by incubation at 80 °C for 15 minutes.

Table 14: Cell lysis of the low-pass workflow.

Reagent	Volume per 10 reactions [µL]
Ampli1™ Reaction Buffer 1	2.0
Ampli1™ Reagent 2	1.3
Ampli1™ Reagent 3	1.3
Ampli1™ Enzyme 1	2.6
Water	12.8

After the cell lysis, DNA digestion was performed. For this step, the reagent was mixed according to the following table (Table 15). Per sample 2 µL of this mix was added, followed by a 5-minute incubation at 37 °C and 5 minutes at 65 °C.

Table 15: Preparation of Digestion reaction mix.

Reagent	Volume per reaction [µL]
Ampli1™ Reaction Buffer 1	0.2
Ampli1™ Enzyme 2	0.2
Water	1.6

In the meantime, the preanneling mixture was prepared according to the following table (Table 16). This reaction mixture was incubated in the thermocycler starting at 65 °C by decreasing 1 °C per minute until 15 °C was reached.

Table 16: Preannealing Reaction mixture.

Reagent	Volume per reaction [µL]
Ampli1™ Reaction Buffer 1	0.5
Ampli1™ Reagent 4	0.5
Ampli1™ Reagent 5	0.5
Water	1.5

The next step was the ligation, therefore the preannealing mixture was used. The reagents were added into the preannealing mixture, not to the samples. Afterwards, 5 µL from the ligation mixture was added to each sample and incubated overnight at 15 °C.

Table 17: Ligation Mixture.

Reagent	Volume per reaction [µL]
Ampli1™ Reagent 6	1.0
Ampli1™ Enzyme 3	1.0
Preannealing mixture	3.0

The final step of the WGA was the primary PCR. According to the following table (Table 18), the reagents were added to each sample.

Table 18: Primary PCR reaction mixture.

Reagent	Volume per reaction [µL]
Ampli1™ Reaction Buffer 7	3.0
Ampli1™ Reagent 8	2.0
Ampli1™ Enzyme 4	1.0
Water	34.0

The program of the PCR was done according to table 19. The final volume of each sample was 50 µL.

Table 19: PCR Thermal incubation profile of primary PCR reaction.

Cycle number	Temperature [°C]	Time	Additional time and temperature
1	68 °C	3 min	
14	94 °C 57 °C 68 °C	40 sec 30 sec 90 sec*	 *= +1 sec/cycle
8	94 °C 57 °C** 68 °C	40 sec 30 sec 105 sec*	 **= +1 °C/cycle *= +1 sec/cycle
22	94 °C 65 °C 68 °C	40 sec 30 sec 113 sec*	 *= +1 sec/cycle
1	68 °C 4 °C	220 sec ∞	

After the PCR, the samples were stored at -20 °C or it was immediately prolonged with the low-pass library preparation.

3.10.2.4 Low-pass library preparation

The libraries for the low-pass whole-genome sequencing were done with the Ampli1 low-pass Kit from Menarini Biosystems. This kit is built on the WGA product of the previously used kit. The library preparation was done according to the manufacture's guidelines. 10 µL of the Ampli1 WGA product was used to clean up with 1.8 x SPRIselect beads (Beckman Coulter). Beyond, low-coverage whole-genome libraries were prepared with the Ampli1™ low-pass Kit (Menarini Silicon Biosystems) according to the manufacturer's instructions. 1.2 µL of the cleaned WGA product were used for the barcode reaction (Table 20). The PCR program of the low-pass barcode reaction was done according to the following table (Table 21).

Table 20: Barcode reaction.

Reagent	Volume per reaction [µL]
Water	17.5
Ampli1^{TM} Reaction Buffer	2.5
Ampli1^{TM} dNTPs	0.9
Ampli1^{TM} BSA	0.65
P1	1.0
Ampli1^{TM} Taq Polymerase	0.25
Ampli1^{TM} Adapter Barcode	1.0
WGA product	1.2

Table 21: PCR Program Barcode reaction.

Cycles	Temperature [°C]	Time	Additional time and temperature
1	95 °C	4 min	
1	95 °C 60 °C 72 °C	30 sec 30 sec 2 min	
10	95 °C 60 °C 72 °C	30 sec 30 sec 2 min*	*= +20 sec/cycle
1	72 °C 4 °C	7 min ∞	

The cleanup of the libraries was done with 1.8 x SPRIselect beads (Beckman Coulter) with a final volume of 25 µL low TE buffer. The final libraries were measured with the Qubit HS DNA Assay. Afterwards, every single library was checked at the fragment analyzer system DNA HS Assay in the BMFZ (german: Biomedizinisches Forschungszentrum) at the Heinrich-Heine university Duesseldorf.

The final library concentration was determined on the fragment analyzer system with the highly sensitive genomic DNA analysis kit. A gel-size selection was not done with the samples. In the end, the libraries were set to a 100 pM equimolar library pool.

3.10.2.5 Low-pass sequencing

The pooled libraries were used for the emulsion PCR amplification (400bp). The template positive Ion Sphere Particles (ISPs) were enriched using the Ion OneTouch™2 System. Sequencing was performed using the Ion 530™ chip on the Ion S5™ System with 525 flows from ThermoFisher Scientific. On a 530^{TM}Chip, a maximum of 12 samples were loaded. The whole coverage was in an average of 0.8 x-1.3 x. Each sample has between 750,000-2,500,000 reads. The median of the absolute values of all pairwise differences (MAPD) quality value was at least below 0.30. This value measured the absolute difference between the log2 copy number ratios of neighboring bins and then calculates the median across all bins. Larger MAPD values indicate greater noise.

3.10.3 Whole transcriptome sequencing by NGS

3.10.3.1 RNA Isolation from FFPE tissue

To isolate the RNA from FFPE tissue, 5-8 slides with 8 µm thick tissue section were prepared. The slides were stained with a crystal violet/Eosin staining to distinguish the different cell types. For that procedure, the slides were incubated for 30 seconds each in a decreasing alcohol series.

The staining solution consists of a crystal violet/Eosin staining (70v/v:30v/v) for 30 seconds, afterwards an increasing ethanol series with two times 95 % ethanol and 2 times 100 % ethanol each for 30 seconds. In the end, the slides were incubated two times for 5 minutes in xylene. After the staining, the slides were stored at RT until the dissection with the laser microscope started. The microdissection with the PALM Microbeam microscope from Carl Zeiss was described in chapter 3.9.6 in detail. The paraffin sections were mounted on a glass slide and via manual microdissection, the tissue was scratched and directly transfer into a reaction tube by a tip. For total RNA isolation, the RNeasy FFPE Kit from Qiagen was used. The procedure was done according to the manufacturer's handbook.

6-12 slides with a thickness of 8 µm were used for the isolation of RNA from paraffin sections. The scratched areas were transferred into a reaction tube with a 160 µL deparaffinization solution, followed by incubation at 56 °C for 3 minutes to remove the remaining paraffin.

150 µL PKD buffer was added to each sample and centrifuged for 1 minute at 11.000 x g at RT. To the lower clear phase, 10 µL proteinase K was added. The digestion took place at 56 °C in a heating block for 15 minutes, followed by an incubation at 80 °C for 15 minutes to stop the enzyme reaction. The lower clear phase was transferred into a new reaction tube and incubated for 3 minutes on ice. Afterwards, the reaction tube was centrifuged for 15 minutes at 20.000 x g. The supernatant was transferred to a new reaction tube. DNase I digestion was done for 15 minutes at RT. After the incubation, 320 µL RBC buffer and 720 µL ethanol (absolute) were added. Up to 700 µL was transferred to a column and centrifuged for 15 sec at 8.000 x g, this was repeated until the sample was completely loaded on the column. Then, the column was washed twice with 500 µL RPE buffer. The column was set on a new collection tube and was centrifuged for 5 minutes at full speed to dry the column. Afterwards, the column was placed in a new labeled 1.5 mL reaction tube. The final elution was done by adding 16 µL of water to the center of each column and centrifuged for 1 minute at full speed. The RNA was stored at -80 °C with 1 µL RNaseOUT inside.

3.10.3.2 RNA quantification

The concentration of the RNA was measured with the Qubit RNA HS Assay. For the RNASeq library preparation, 1 µg RNA was necessary. The quality of the RNA samples was determined with the bioanalyzer at the core facility BMFZ from the Heinrich-Heine University of Duesseldorf. The RNA integrity number (RIN) should be over 7.0 for high-quality RNA sequencing. The used RNA was from FFPE tissue samples with degraded RNA but with a high fragment size (70 % above 200 bp.). The measured RIN from the used samples were between 2.3 -4.0.

In addition to the bioanalyzer, a quantitative real-time PCR was done for a quality check to test the concentration of amplifiable RNA. For this, the HML-2 quantitative real-time PCR from chapter 3.10.1.2 was done on first-strand synthesis cDNA. The cDNA was prepared with the RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher).

3.10.3.3 RNASeq

The isolated total RNA from human FFPE tissue was sent to the Korean company, Macrogen. The library preparation with the TrueSeq mRNA library kit from Illumina and the NGS run on the NovaSeq 6000 with 2 x 100 bp and 30 Mio reads per sample was done by the company with the standard manufactured protocols.

3.10.4 Genome-wide DNA Methylation

3.10.4.1 Illumina Infinium MethylationEpic Array

For the genome-wide DNA methylation Array, FFPE genomic DNA was used. The DNA was isolated with the GeneRead FFPE Kit from Qiagen (chapter 3.10.1.1). The array needed about 500 ng genomic DNA in total. The concentration was measured with the Qubit BR DNA Assay. The Infinium DNA Methylation Kit from Illumina was used. The procedure was done accordingly to the manufacture's instruction. The whole procedure and array run were done in Heidelberg in the labs of the Genomic and Proteomics Core Facility of the German cancer research center (German: deutsches Krebsforschungszentrum, DKFZ).

The analysis was done in the lab of Prof Siveke from the University Hospital Essen at the department of internal medicine.

3.10.5 FISH

The fluorescence *in situ* hybridization (FISH) was done with the UroVysion probe (Abbott). It is an FDA-approved probe for bladder cancer from cytological smears. The kit consisted out of four different probes. Three probes are chromosome enumeration probes (CEP) for chromosome 3, 7 and 17, the fourth probe binds at chromosome 9 at the region of gene *CDKN2A* and *CDKN2B* (Figure 15).

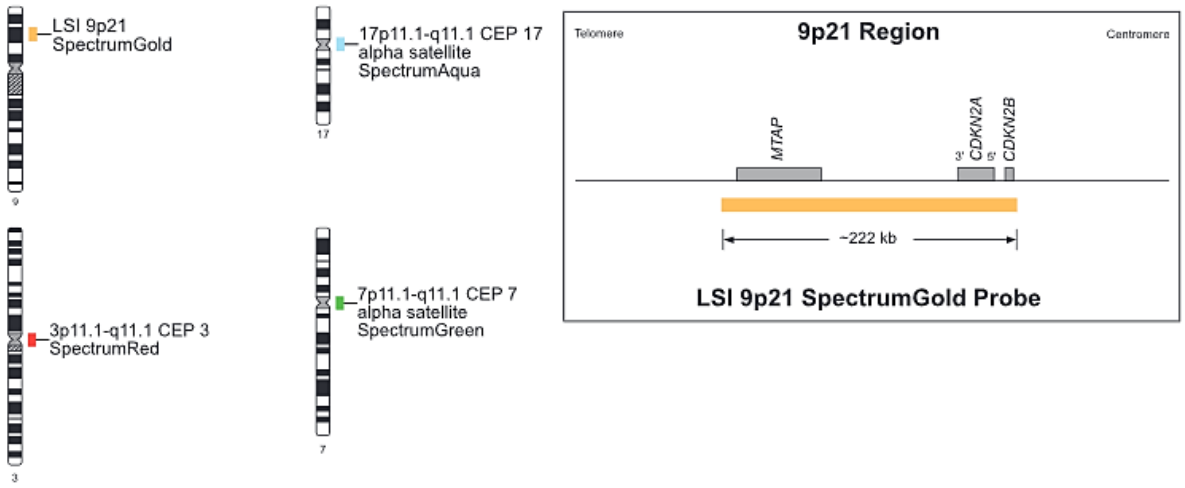


Figure 15: Probes of the UroVysion kit from Abbott.
Three Probes are CEPs for Chromosome 3, 7 and 17. The fourth probe binds at chromosome 9 at the region of genes MTAP, CDKN2A and CDKN2B. (figure is adapted from [157]).

The probes were used at single-cell suspension from the paraffin section (preparation described in chapter 3.10.2.1). The procedure was done according to the manufacture's manual in cytopathology from Dr. Bruno Eduardo Silva de Araujo.

In the next step, the protocol of the FISH was adapted to the paraffin section from the technical stuff of the molecular pathology. The protocol was done according to a modified FISH protocol from the company Abbott. A section of bladder cancer was used as a control section for the procedure.

For the FISH analysis, 1 µm thick tissue sections were used. To remove the paraffin, the sections were incubated at least for 2 hours at 60 °C. After the incubation in the oven, the sections were incubated four times in xylol for 5 minutes, followed by two times incubation in ethanol absolute for 1 minute. The next step was incubation in HCl (0.2 M) for 20 minutes, followed by an incubation in H_2O for 2 minutes. Afterwards, the sections were washed in 2x SCC buffer for 3 minutes. Incubation at 80 °C in a prewarmed pretreatment buffer was followed for 30 minutes. Then, the sections were washed in H_2O for 1 minute.

Two washing steps with 2xSCC buffer for 5 minutes each followed. Then, the sections were incubated in prewarmed protease digestion buffer with 25 mg/mL protease I for 45 minutes at 37 °C in the water bath. Afterwards, the sections were washed two times in 2x SCC buffer each for 5 minutes. Subsequently, the fixation in 4 % formalin for 10 minutes was done. Then, the sections were washed 2 times with 2x SCC buffer each for 5 minutes. The probes mixture was diluted 1:5 and pipetted on the section. The sections were covered by a thin glass slide and incubated for 10 minutes at 85 °C followed by overnight incubation in the hybridization oven at 37 °C. To remove the probe mixture, the sections were washed with 2x SCC buffer for 5 minutes until the glass slides could be removed. Afterwards, the sections were treated with a prewarmed post hybridisation buffer for 2 minutes in the water bath at 72 °C. The next step was an additional washing step with 2x SCC buffer. The DAPI solution was added and the sections were covered again with a glass slide. Until evaluation, the slides were stored at 4 °C. The evaluation of the FISH samples was done in the cytopathology of the University Hospital Duesseldorf. At least 60 single cells were evaluated per sample.

3.10.6 Bioinformatic methods

During this study, different methods, tools and databases were used to investigate the generated data from DNA Panel-Sequencing, Infinium MethylationEpic array, Transcriptome data and whole-genome low-coverage sequencing.

3.10.6.1 NGS analysis

The results of the NGS sequencing from the Ion Torrent S5 were aligned to the human reference genome (GRCh37/hg19) using the S5 Ion Torrent Server VM (ThermoFisher Scientific, Germany). Ion Reporter software (Version 5.12.0.0) was used for variant calling and annotations of the DNA panel sequencing. The parameters for variant calling were set for all samples equal. The minimum allele frequency was 3 % with a minimum coverage of 500 and a Phred Score of > 30. Detected variants were validated with different databases and tools to prove the clinical significance and decide to be pathogenic. The main databases were the ClinVar database from NIH, UCSC Genome Browser, COSMIC, and FATHMM.

For some variants, which were not presented in the databases, the judgment according to the ACMG guidelines was done[158]. For this own judgement, the ACMG database (varsome.com; v7.3.7) was used. And in addition, two web-based tools PERViewer and MTR Viewer were used to show the prediction of the unknown variant[159,160].

PERViewer was used to identify pathogen enriched regions (PER), conserved amino acids among gene family members were considered. The MTR Viewer used a so-called missense tolerance ratio (MTR), which summarizes available human standing variation data within genes to encapsulate population-level genetic variation.

3.10.6.2 Analysis of copy number variations

The low-pass sequencing data was mapped with the Ion Reporter software using the low-pass whole-genome aneuploidy w1.0 pipeline. The resulting list consisted of the copy number at defined bins. As a control, six normal acinar tissue samples were measured. With the pooled acinar tissue samples and the lesion sample, the log2 value was determined. A copy number variation (CNV) was set to be at a $\log_2$ value of +/- 0.2. The circus plots of the CNV distribution were generated with the shinycircos software[161].

To inspect the values for chromothripsis-like regions, the CTLPScanner was used[162,163]. The $\log_2$ values of the methylome data were also investigated with this method. The settings were set as followed, copy number aberration status changes events of ≥ 11 and $\log_{10}$ of likelihood ration of ≥ 8. The chromosomal segments with a size of ≥ 50 kb were considered to exclude artifacts[136].

3.10.6.3 Analysis pipeline RNA-Seq Data

The raw data processing of the transcriptome data from the sequencing with the Illumina Nova Seq6000 was done by the company Macrogen. The workflow started with the packages FASTQC (v0.11.7) and Trimmomatic (v0.38) to improve the reading quality by trimming the data. The programs HISAT2 (v2.1.0) and Bowtie2 (v2.3.4.1) were used to map the next-generation sequencing reads to the human genome (GRCh38/hg38). The last step, which was performed by Macrogen, was the analysis with the program StringTie (v1.3.4d) to assemble the RNA-Seq alignments into potential transcripts and splice variants for each gene locus[97–101]. The StringTie packages spend the read counts of each gene or transcript for each sample. The read counts were used for the R package TCC-GUI[169,170]. DESeq was used as the normalization method and DESeq2 to identify differentially expressed genes (DEGs) with a p-value of < 0.05[171,172]. There was no threshold for low counts genes and the number of iterations was set at three. The FDR cut-off was 0.1[173].

3.10.6.4 Methylome

The analysis of the Illumina Methylation Epic Array was done at the Department of internal medicine at the university hospital in Essen by Dr. Sven T. Liffers and Laura Godfrey from the working group of Prof Jens Siveke. A brief description was written by Dr. Sven Liffers.

DNA methylation profiles were measured with the Infinium Methylation EPIC BeadsChips (Illumina, San Diego, USA) at the Genomics and Proteomics Core Facility of the German Cancer Research Center (DKFZ) Heidelberg. Methylation analysis was carried out using the R Bioconductor package ChAMP (v2.14.0). Briefly, IDAT files were loaded into the ChAMP and preprocessed. In the first step, all probes with a p-value > 0.01 were excluded. Followed by the exclusion of probes with a bead count > 3 in at least 5 % of the samples, non-cg probes, SNP containing probes and sex probes were also filtered. Filtered datasets were normalized using the BMIQ method and batch corrected before differential analysis. Differential probes were defined with a delta of 0.2 and an adjusted p-value (BH) of >= 0.05. The phylogenetic tree was plotted using the R-package ape (v.5.3).

3.10.6.5 Pathway Analysis

An enrichment analysis based on the functional annotation of genes or the belonging pathways were used to investigate the resulting gene list from a differentially methylated pattern (DMP), DEG and regions from the CNVs. This method worked by comparing the

frequency of individual annotations in the gene list with a reference list (genome) [174,175]. The analysis was done using the online tool Enrichr (https://amp.pharm.mssm.edu/Enrichr/). The tools Panther, Webgestalt, and DAVID were used in addition to comparing the results and check the differences from the different algorithms. In addition, a STRING analysis was done for known and predicted protein-protein interactions [176,177].

3.10.6.6 Statistical analysis

Statistical analysis was performed using the GraphPad Prism 5 software (GraphPad Software Inc., USA) and statistical significance was determined by the Kruskal-Wallis test. Results were presented as means ± standard error of the mean (SEM). P values less than 0.05 were considered statistically significant (* $P< 0.05$; ** $P< 0.01$; *** $P< 0.001$).

4 Results

For a detailed molecular characterization of the common precursor lesions of the PDAC, a retrospective cohort analysis including mutational profile, transcriptome profile, methylation profile and copy number profile were performed during this project.

4.1 Mutation profile of precursor lesions

In this project, 59 samples from 51 different patients were sequenced with two self-designed PDAC-Panels, including 7 control samples of normal acinar tissue. 24 samples from female and 28 samples from male patients are included. The following oncoprint (Figure 16) shows the results of the next-generation sequencing.

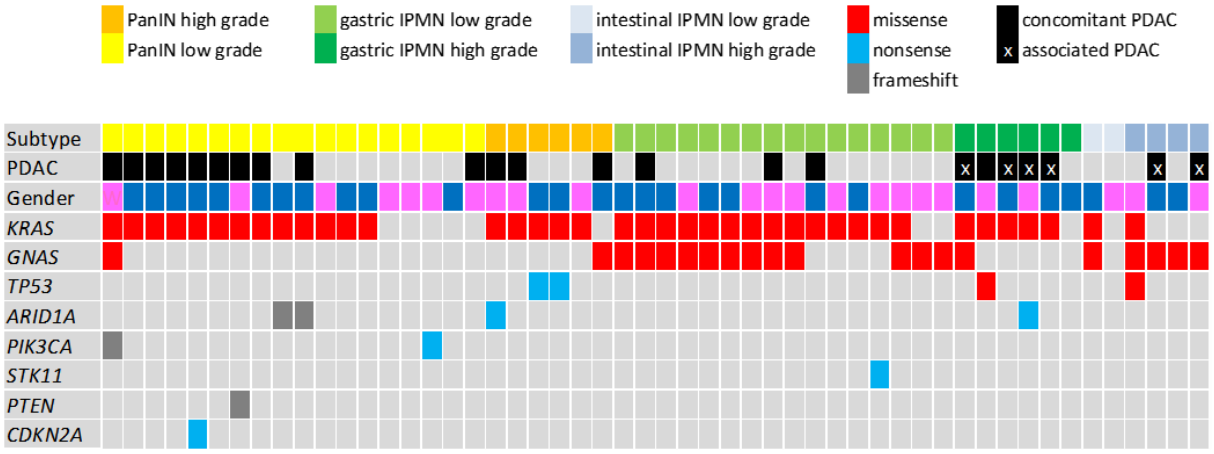


Figure 16: Mutation profile of targeted NGS.
Low and high grade PanIN, gastric IPMN and intestinal IPMN were included in the analysis. Cases with a concomitant PDAC are indicated with a black square and associated PDACs are indicated in addition with a white X. Labeled mutations represent pathogenic mutations according to the ClinVar database and/ or ACMG guidelines with an allele frequency of ≥ 3 %. Red squares represent missense mutations, blue squares are nonsense mutations and grey squares are frameshift mutations. Empty squares indicate the absence of pathogenic mutations. Analysis was performed using a 21-gene customer-designed panel on the S5 Ion Torrent platform. (Phred Score ≥ 30, coverage ≥ 500).

In the first row, the samples were sorted by the precursor lesion and were separated into low and high grade according to the grade of dysplasia. The second row of the oncoprint shows whether an associated or a concomitant PDAC was present in the patient or not. Labeled mutations represented pathogenic mutations according to the ClinVar database[178] and/or the American College of Medical Genetics and Genomics (ACMG) guidelines[158] with an allele frequency of ≥ 3 %. The quality threshold was set with the Phred Score at ≥ 30 and at least a coverage of 500.

The predominant pathogenic mutation was a mutation in the gene *KRAS*, which was present in all different precursor lesions. In detail, 72.22 % in PanIN low grade and 83.33 % high grade

PanIN and in 87.5 % gastric IPMN low grade and 83.33 % in high grade lesions. The amount in intestinal IPMN was lower, with 50 % in low grade and 25 % in high grade lesions.

The pathogenic *GNAS* mutation was mainly present in IPMNs, and in one PanIN low grade and one PanIN high grade lesion. In 50 % of low grade intestinal IPMN, the mutation was present and in all intestinal IPMN high grade. In the gastric IPMN low grade, 75 % consisted of this mutation and it was shown in 16.67 % of gastric IPMN high grade.

TP53 pathogenic missense mutation was only present in one sample intestinal IPMN high grade and one gastric IPMN high grade. Two nonsense *TP53* mutations in PanIN high grade lesions were detected.

In addition, *ARID1A*, *PIK3CA* (Phosphatidylinositol-4,5-Bisphosphate 3-kinase catalytic subunit alpha), *STK11*, *PTEN* (Phosphatase and Tensin homolog*),* and *CDKN2A* nonsense and frameshift mutations were present. A *CDKN2A* nonsense mutation was only in one PanIN low grade sample, the other investigated samples do not possess a *CDKN2A* mutation.

The following table (Table 22) represents the percentage of the found mutations in the cohort. For a better overview, only the genes with mutations were presented.

Table 22: Overview of the mutation distribution of all genes in the panels.
The precursor lesions were separated into a low grade (lg) and high grade (hg). The values were the percentage of the samples, which consisted of the mutations. PanIN n= 24 (lg n= 18; hg n= 6); gastric IPMN n= 22 (lg n= 16; hg n= 6); Intestinal IPMN n= 6 (lg n= 2; hg n= 4).

	PanIN		IPMN gastric		IPMN intestinal	
	lg [%]	hg [%]	lg [%]	hg [%]	lg [%]	hg [%]
KRAS	72.22	83.33	87.50	83.33	50.00	25.00
GNAS	5.56	16.67	75.00	16.67	50.00	100.00
TP53	0.00	33.33	0.00	16.67	0.00	25.00
ARID1A	11.11	16.67	0.00	16.67	0.00	0.00
PIK3CA	11.11	0.00	0.00	0.00	0.00	0.00
STK11	0.00	0.00	6.25	0.00	0.00	0.00
PTEN	5.56	0.00	0.00	0.00	0.00	0.00
CDKN2A	5.56	0.00	0.00	0.00	0.00	0.00

A closer look at the *KRAS* mutation showed significant different variant allele frequencies (VAFs) along with the different precursor lesions (Figure 17 A).

The VAFs of *KRAS* and *GNAS* normalized to the tumor cell content were evaluated. All VAFs of 3 % or higher were considered.

PanIN lesions showed a mean VAF of 11.09 ± 5.17 % (n= 18), gastric IPMN had a mean of 24.94 ± 9.65 % (n= 19), which was significantly higher than in the PanIN lesions ($p < 0.001$). The

highest mean VAF of *KRAS* was found in intestinal IPMN with 35.88 ± 2.93 % (n= 2) and were significantly higher than the PanIN lesions ($p< 0.05$).

The *KRAS* mutation was in 99.93 % the G12 (glycine on codon 12) mutation on exon 2. The amino acid change (aa change) of the *KRAS* mutation G12 is presented in the following figure (Figure 17 B). The most common aa change in gastric IPMN samples was the change from glycine (Gly) to valine (Val) in 57.9 % of the specific *KRAS* mutation. In comparison to PanIN lesions, the most common aa changes were from Gly to aspartic acid (Asp) with 50 % of the mutations. In both entities, the aa change from Gly to arginine (Arg) was present. The aa change from Gly to Arg was present in 16.7 % of the PanIN lesion and only in 5.3 % of the gastric IPMNs.

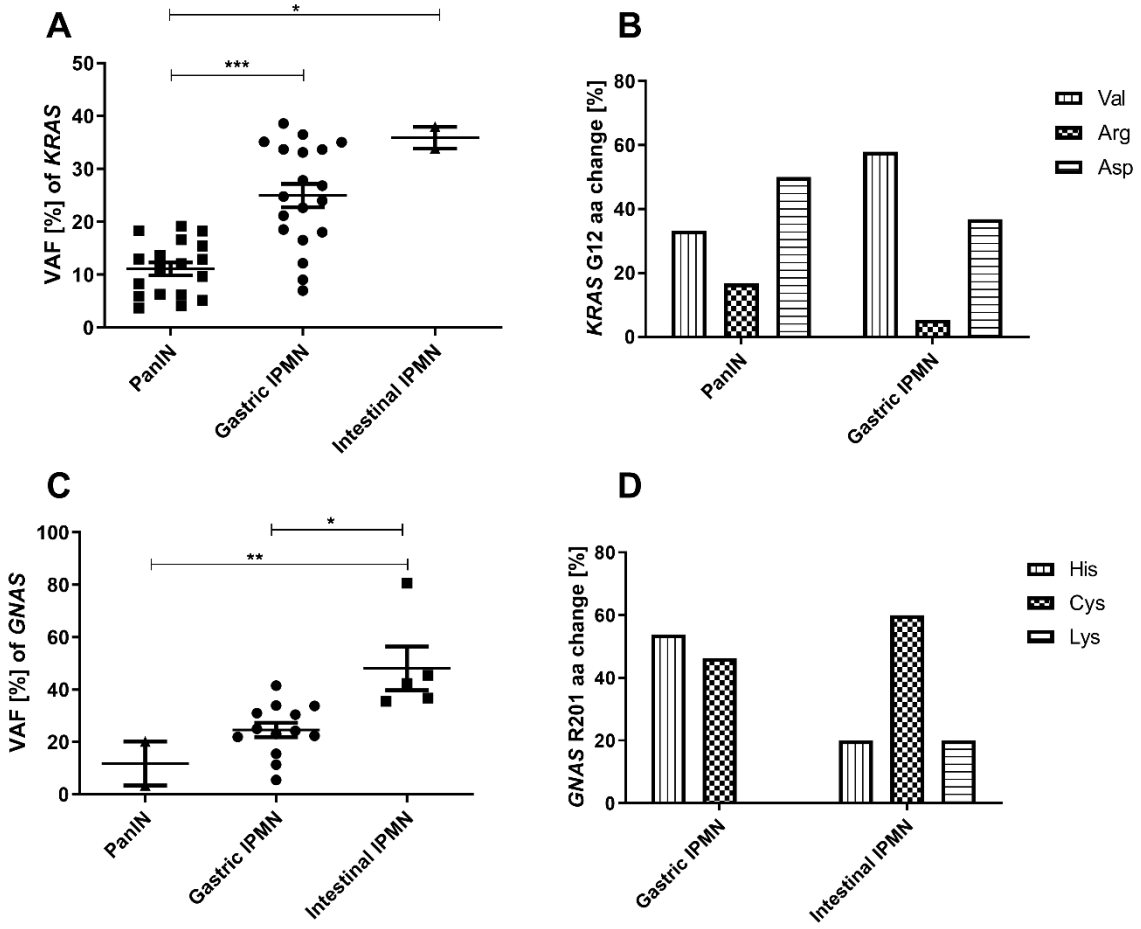


Figure 17: Variant allele frequency of *KRAS* and *GNAS* mutations in different precursor lesions.
(A) VAF of *KRAS* mutations (B) Amino acid changes from *KRAS* G12 mutations in gastric IPMN and PanIN lesions (C) VAF of *GNAS* R201 mutations (D) Amino acid changes from *GNAS* mutations in gastric IPMN and intestinal IPMN. (Kruskal-Wallis-test * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

The VAF of *GNAS* mutations had a similar distribution as the VAF of the *KRAS* mutations, being 24.63± 9.87 % in gastric IPMN, 48.14± 18.59 % in intestinal IPMN and 11.80± 11.84 %. in PanIN lesion.

The *GNAS* allele frequency was significantly higher in intestinal IPMN (mean 48 %) in comparison to gastric IPMN samples (mean 26 %) (Figure 17 C), with no differences between low grade and high grade lesions. 95.2 % of the detected *GNAS* mutations affected the arginine residue on codon 201 of exon 8. The aa change distribution is shown in figure 17 (Figure 17 D). The most common aa change in gastric IPMN was from Arg to histidine (His) with a mean of 53.8 % and in intestinal IPMN the most common aa change was from Arg to cystine (Cys) with a mean of 60 %. The aa change from Arg to lysine (Lys) was only present in intestinal IPMN with a mean of 20 %.

The Pearsons' correlation analysis of double mutated (*GNAS* and *KRAS*) gastric IPMN samples confirmed a positive correlation between the variant allele frequencies of the two mutations *KRAS* and *GNAS* (r= 0.979523, p= < 0.0001, Figure 18) when the *KRAS* G12 and the *GNAS* R201 mutations were considered.

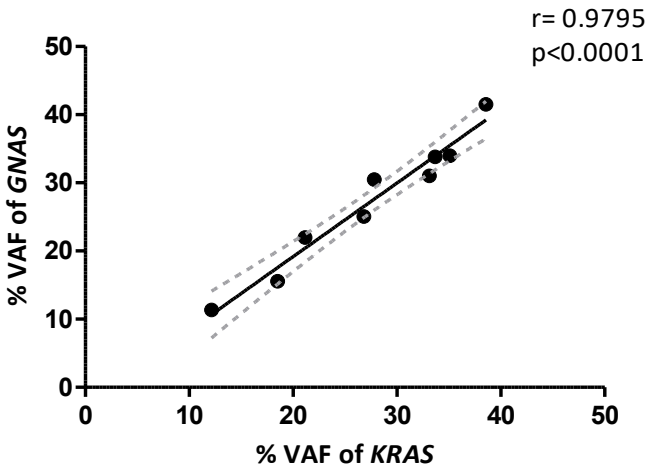


Figure 18: Correlation of the VAF of *KRAS* and *GNAS* in gastric IPMN samples.
A scatterplot shows the VAF of the *KRAS* mutations (G12) against that of the *GNAS* mutations (R201) detected in gastric IPMN low and high grade lesions (n= 9). The grey lines represent the 95 % confidence interval of Pearson's correlation coefficient r.

4.2 Detection of copy number variations

Whole-genome low coverage sequencing was performed to identify regions with copy number variations and to find a hint for chromothripsis in precursor lesions. In recent years, evidence was found that pancreatic cancer can occur through two chromosomal rearrangements via chromothripsis. In that study, the number of PDACs, which arise from chromothripsis events were estimated at about 60 %[135].

In this study, 47 samples were sequenced with the low-pass method, including 11 PanIN lesions (n= 9 low grade; n= 2 high grade), 14 gastric IPMN (n= 9 low grade and n=5 high grade), seven AFL samples, five intestinal IPMN (n= 3 low grade and n= 2 high grade), four PDAC and six normal acinar tissue samples as a control group. To compare the CNV pattern, a Log_2 value were used with normal acinar tissue samples as control.

In only 18.2 % of PanIN (2/11) lesions, copy number variations were present. In PanIN lesions, no repeated regions were found with copy number variations. One PanIN high grade sample showed similar regions involved as in the gastric IPMN, on chromosome 1 and chromosome 9 at the *CDKN2A* location (Figure 19 A). In 42.8 % of the gastric IPMNs (6/14), copy number variations were present. Repeated regions on chromosome 9 at the region of *CDKN2A* and on chromosome 17 at the region of *TP53* were affected in three cases (Figure 19 B). In both regions, the copy number variation was a loss of copy number. Also, the region on chromosome 1 was affected in 4 samples, here, no gene related to tumorigenesis or to the pancreas was identified. 1126 genes were involved in the region of chromosome 1. 221 genes were at least 2 times involved. Enrichment analysis with an online webtool enrichr of these 221 genes showed that the genes were related to drug metabolism and cell adhesion molecules (see attachment 11.3).

The highest percentage of affected samples (80 %, 4/5) belonged to the group of intestinal IPMN. In intestinal IPMN samples, large areas of chromosomes were affected, mainly with a copy number gain. For example, on chromosome 7, the area of the gene *BRAF* (B-Raf Proto-Oncogene, Serine/Threonine kinase) was affected in all samples (Figure 19 C). In 57.1 % of the AFL lesions (4/7), CNVs were detected. In comparison to the other precursor lesions, the AFL samples were affected with copy number variations only in small areas. No specific regions were affected several times (Figure 19 D).

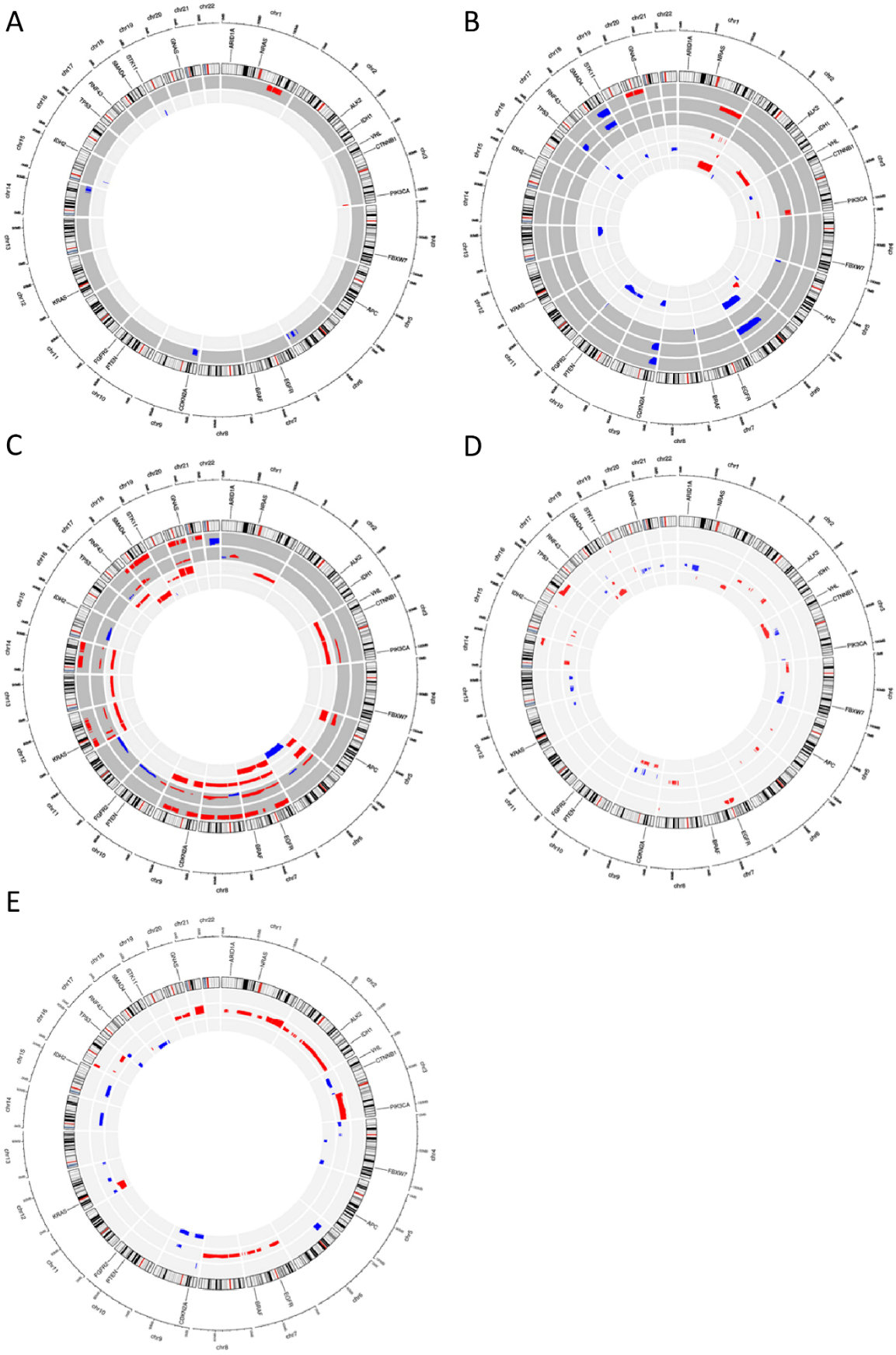

A
B
C
D
E

Figure 19: Copy number variations in PanIN, gastric IPMN, intestinal IPMN, AFL and PDAC.
Copy number variations were detected in PanIN, IPMN, AFL and PDAC over the whole-genome by low-coverage sequencing. Regions in red show copy number gains and regions in blue represent copy number losses. (A) PanIN (n= 11, 9 lg and 2 hg) did not possess repeated or larger regions of CNV; (B) gastric IPMN (n= 14, 9 lg and 2 hg lesions) reveal three distinct repeated regions of copy number loss at chromosome 6, 9 and 18; (C) intestinal IPMN (n= 5, 3 lg and 2 hg) had the highest frequency of chromosomal alterations. The broad genomic alterations generally involved entire chromosomes and were mostly located on chromosome 7, 8, 12, 18 and 20; (D) Copy number variations detected in AFL lesions with low-coverage whole-genome sequencing. Five samples from four different patients, consisting of small 3-5 cell clusters were sorted by DEPArray (n= 5). Two further samples were microdissected from tissue (n= 2). (Two grey shadows show two different patients); (E) PDAC (n= 4); (Dark grey background= high grade lesions, light grey background= low grade lesions; log2 value, threshold +/- 0.2).

In 75 % of PDAC samples (3/4), CNVs were present. No clear distribution pattern was observable. One sample was affected with copy number gains at eight different chromosomes. The regions were covering almost the whole chromosomes (Figure 19 E).

During this study, a cohort of PanIN, gastric IPMN, and intestinal IPMN were tested with the Illumina Infinium MethylationEpic Array to generate the DNA methylation profile. These array data was also used for copy number detection. The resolution was much higher as that of the whole-genome low-coverage sequencing. The bin size of the methylome data for the CNV detection was 50-150 kbp, this was 13 to 40 times smaller than the bins of the whole-genome low-coverage sequencing. The circus plots showed the CNV pattern of the precursor lesions PanIN, gastric IPMN, and intestinal IPMN (Figure 20). The CNV analysis of the methylome data showed a similar pattern as the data from the whole-genome low coverage sequencing.

In the PanIN samples, there was only one PanIN (low grade) with an abnormal scattering of the copy numbers, without a specific involved chromosome or region. So only 16.7 % (1/6) showed copy number variations in the PanIN low grade samples. In the PanIN high grade samples, 50 % (1/2) showed a copy gain on chromosome 7 in the area of *MYC* (MYC Proto-Oncogene, BHLH Transcription Factor) and on chromosome 10 in the area of *PTEN*.

21.4 % (3/14) of the gastric IPMN low grade showed some copy number variations. In one sample, the affected chromosomes were chromosome 9 and 10 with a copy loss. This was the area of *CDKN2A* and *PTEN*. The other two samples did not have a defined region, which was affected, the whole scatter of the copy numbers was high. In the gastric IPMN high grade, 50 % (2/4) of the samples showed copy number variations. One sample on chromosome 1 in the area of *MCL1* (MCL1 apoptosis regulator, BCL2 family member) and on chromosome 6 was affected. The other affected samples showed copy variations on chromosome 9 at the locus of *CDKN2A* and chromosome 10 at the locus of *PTEN* and chromosome 16 at the locus of *TP53*.

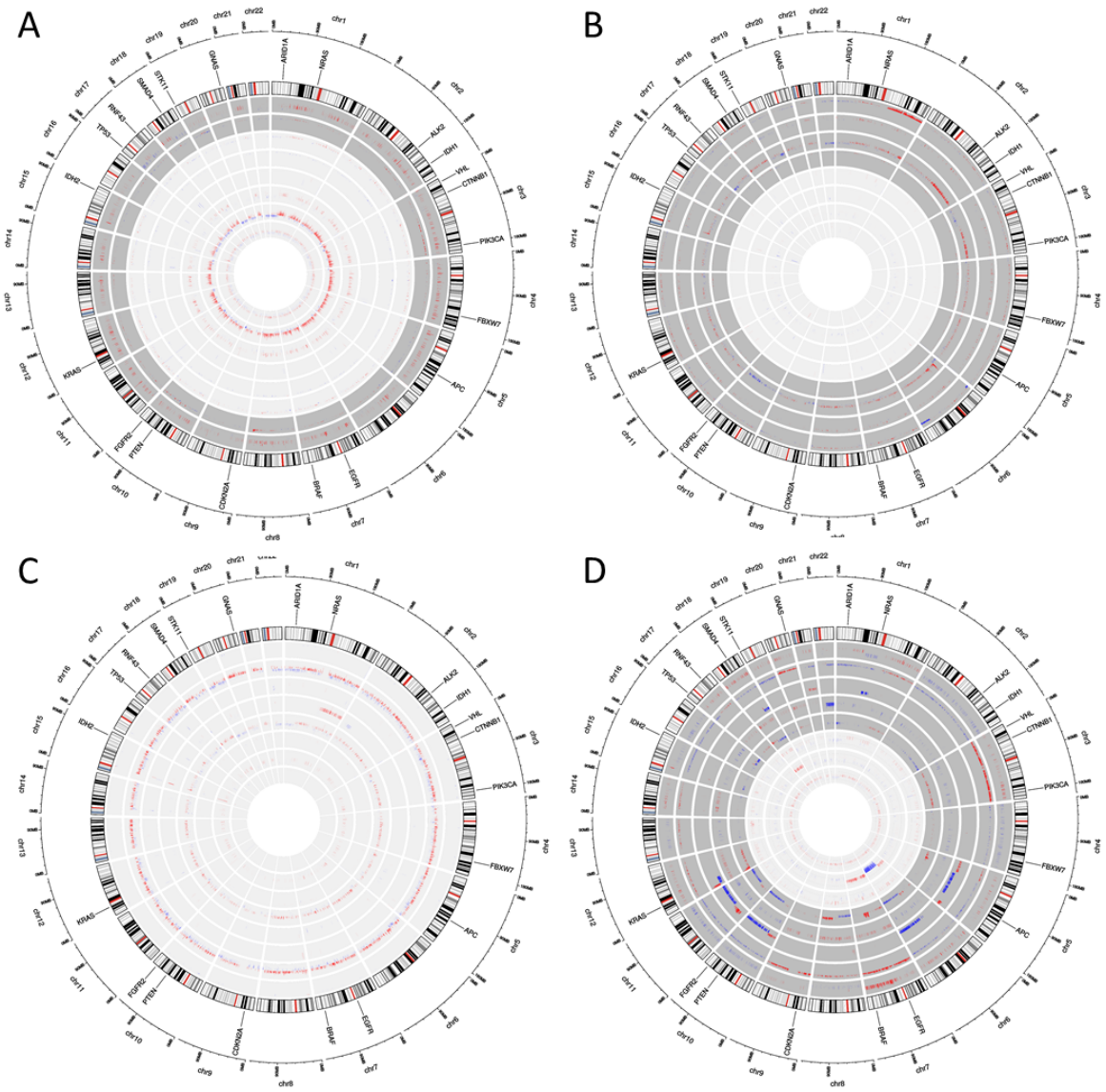


Figure 20: Copy number variations of the array-based Illumina data.
Copy number variations were detected in PanIN and IPMN over the whole-genome by the Illumina Infinium MethylationEpic Array. Regions in red show copy number gains and regions in blue represent copy number losses. (A) PanIN (n= 8, 6 lg and 2 hg); (B) + (C) gastric IPMN (n= 18, 14 lg and 4 hg); (D) Intestinal IPMN (n= 10, 5 lg and 5 hg); (Threshold is set at +/- 0.3, dark grey background= high grade lesions; light grey background= low grade lesions; red= copy number gain; blue= copy number loss; n= 36).

The intestinal IPMN low grade samples showed a copy number loss on chromosome 6 in 40 % (2/5) of the samples. In the intestinal IPMN high grade, 40 % (2/5) showed a copy number gain on chromosome 7 in the area of *EGFR* (epidermal growth factor receptor) and also 40 % of the samples showed copy number loss on chromosome 10 and 11 with *PTEN* and *CCND1* (Cyclin D1) as affected gene areas.

In the next step, the CTLPScanner was used to find evidence for chromothripsis-like regions in the different precursor lesions. Chromothripsis was defined as the shattering and reassembly of one or more chromosomes and was detected using the CTLPScanner algorithm[179]. This

algorithm was originally designed for the detection of chromothripsis-like events from array data. This algorithm measured the copy number changes between neighboring bins, in the whole-genome low coverage sequencing, the bins were large with a size of 2 Mbp. Thus, only catastrophic events in these data could be detected. In all whole-genome low-coverage samples (n= 49), no evidence of chromothripsis-like events was found with the CTLPScanner.
The data from the Illumina Methylome Epic array was also used with the CTLPScanner, these data had a higher resolution, and the bin sites were smaller compared to the sequencing data. In these samples, 11 samples were positively tested for chromothripsis-like events (Figure 21 B). Among all subtypes, evidence of chromothripsis-like events was present, according to the often-changing copy numbers between adjuvant bins. In the first group of tested precursors were the PanIN lesions. In 37.5 % of the PanIN samples (3/8) evidence were found for chromothripsis-like events. The affected chromosomes were chromosome 1 and 12. Again, in PanIN lesions, chromosome 1 was affected in two different samples. In 22.2 % of gastric IPMN (4/18), chromothripsis-like events were found on chromosome 1, 4, 6 and 9. Chromosome 1 was affected in two different samples. The following figure is one example of a gastric IPMN high grade with evidence on chromosome 1 and 4 (Figure 21 A).
In intestinal IPMN, 40.0 % of the samples (4/10) consisted of chromothripsis-like events. The affected areas in intestinal IPMNs were the chromosome 2, 3, 6 and 7, even though chromosome 7 was involved in more than one sample.

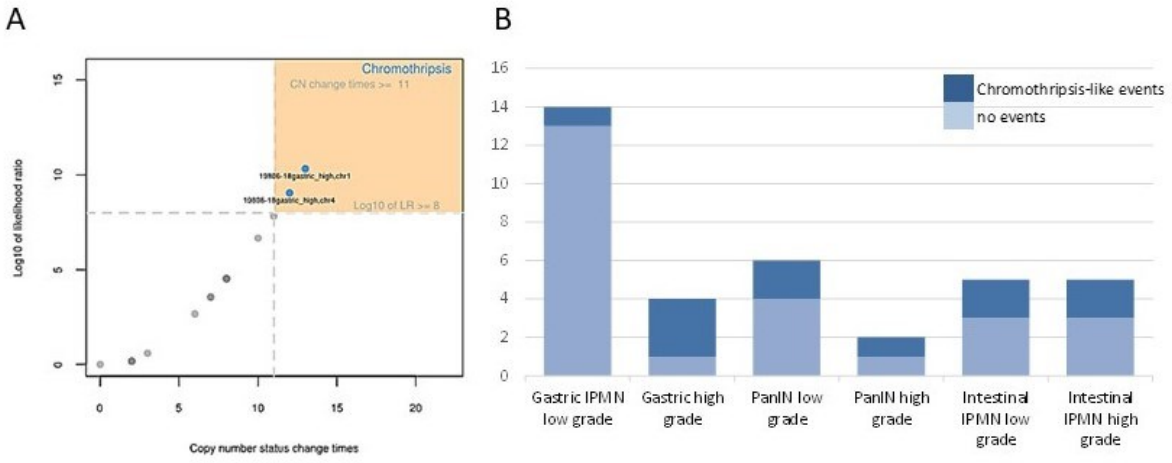


Figure 21: Chromothripsis-like events in precursor lesions of the PDAC.
(A) The gastric IPMN high grade samples showed two chromothripsis-like events on chromosome 1 and chromosome 4. The plot represents copy number status changes versus log10 of the likelihood ratio. (B) Overview of all samples for chromothripsis-like events. The blue box shows the number of samples with chromothripsis-like events and grey without any event. (n= 36 (PanIN n= 8, gastric IPMN n= 18 and intestinal IPMN n= 10)).

In addition, a second technic for aneuploidy was done, the fluorescence *in situ* hybridization (FISH). The FISH probes were binding at the centromere so that the probes worked as chromosome counters. With this technic, it is possible to detect aneuploidy of chromosome 3, 7 and 17 and in addition an abnormal copy number in the loci of *CDKN2A* and *CDKN2B* at chromosome 9. The technic was used for the validation of the whole-genome low-coverage sequencing and the AFL samples with limited DNA material.

A small number of samples were tested for chromosome abnormalities. In total, six samples were tested, four AFL samples, two PanIN low grade. All samples showed no evidence of aneuploidy. The figures of the cells represented a diploid situation for both cells (Figure 22). The cell from figure A was from a cell suspension of an AFL and figure B was from a paraffin section of a PanIN low grade. For each sample, a minimum of 60 single cells was evaluated to determine the diploid status.

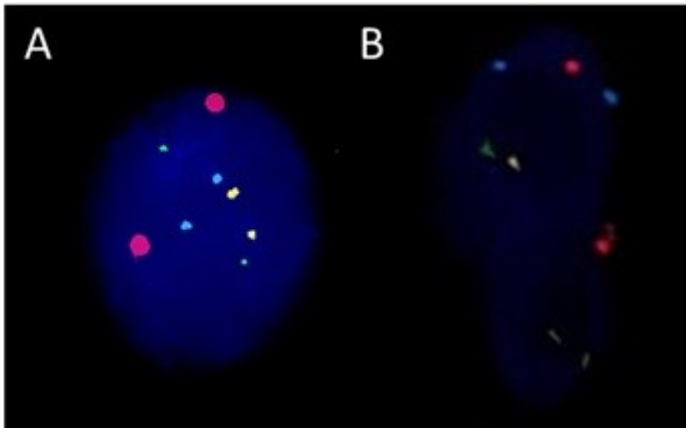


Figure 22: Two examples of a cell with a diploid manner.
(A) One cell of the single cell suspension of an AFL sample; (B) one cell of a paraffin section of a PanIN sample; (red= chr3; green= chr7; gold= *CDKN2A* (chr9); aqua= chr17; DAPI staining= dark blue background).

4.3 Proliferation rate

The marker of proliferation Ki-67 is a nuclear cell marker, which is associated with cellular proliferation rate. Ki-67 protein is present during all active phases of the cell cycle (G1, S, G2, and mitosis), but is absent in resting (quiescent) cells[180]. The marker is well-known and established as a proliferation marker.

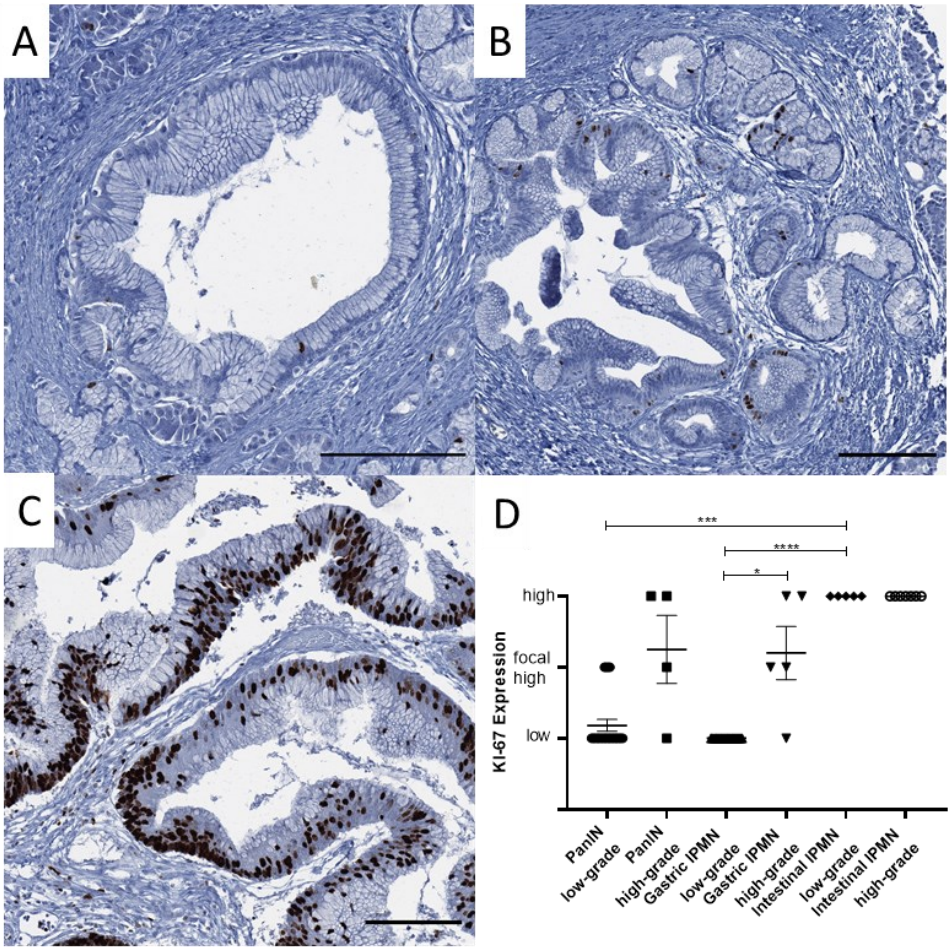


Figure 23: Proliferation rate of PanIN, gastric IPMN and intestinal IPMN.
(A) PanIN low grade lesion with low Ki-67 expression. (B) Gastric IPMN low grade with low expression of Ki-67. (C) Intestinal IPMN low grade with high Ki-67 expression. (D) Quantitative analysis of Ki-67 expression in precursor lesions on TMAs. (n= 70; scale bar= 200 µm, Kruskal-Wallis-test * $P < 0.05$; *** $P<0.001$, **** $P<0.0001$).

In this study, the expression of Ki-67 was determined on 70 different precursor lesions on TMA sections with a MIB-1 antibody to determine the proliferation rate. PanIN and gastric IPMN low grade showed a low expression of Ki-67 in the epithelia cells, with increasing grade of dysplasia the expression of Ki-67 was also increasing. The differences between gastric IPMN low and high grade were significant (Figure 23 D). The expression of Ki-67 in intestinal IPMN was in low and high grade lesions high, there were no differences between the dysplasia grades observable. Intestinal low grade lesions had a significantly higher expression of Ki-67 than gastric IPMN low grade and PanIN low grade.

4.4 Status of the *GNAS* gene

In the mutational profile, the *GNAS* mutation was one of the most common differences between the precursor lesions PanIN and gastric IPMN. To analyze the protein expression of GNAS, two TMAs were generated within a total of 61 different samples. The TMA consisted of PanIN lesions (n= 18; low grade n= 12; high grade n= 6), gastric IPMN (n= 32; low grade n= 24; high grade n= 8) and intestinal IPMN (n= 11; low grade n= 3; high grade n= 8). The expression of GNAS was categorized as negative/weak, moderate or strong. Representative images of the GNAS staining in PanIN, gastric IPMN, and intestinal IPMN are shown below (Figure 24).

PanIN and gastric IPMN lesions showed a significantly higher expression of GNAS than the intestinal IPMNs. There were no differences between low and high grade samples in one entity. Between PanIN and gastric IPMN, no significant difference in the protein expression of GNAS was observed. 24.5 % (15/61) of the samples were included in the sequenced samples. A correlation of the mutation status of *GNAS* and the GNAS protein expression was not detected.

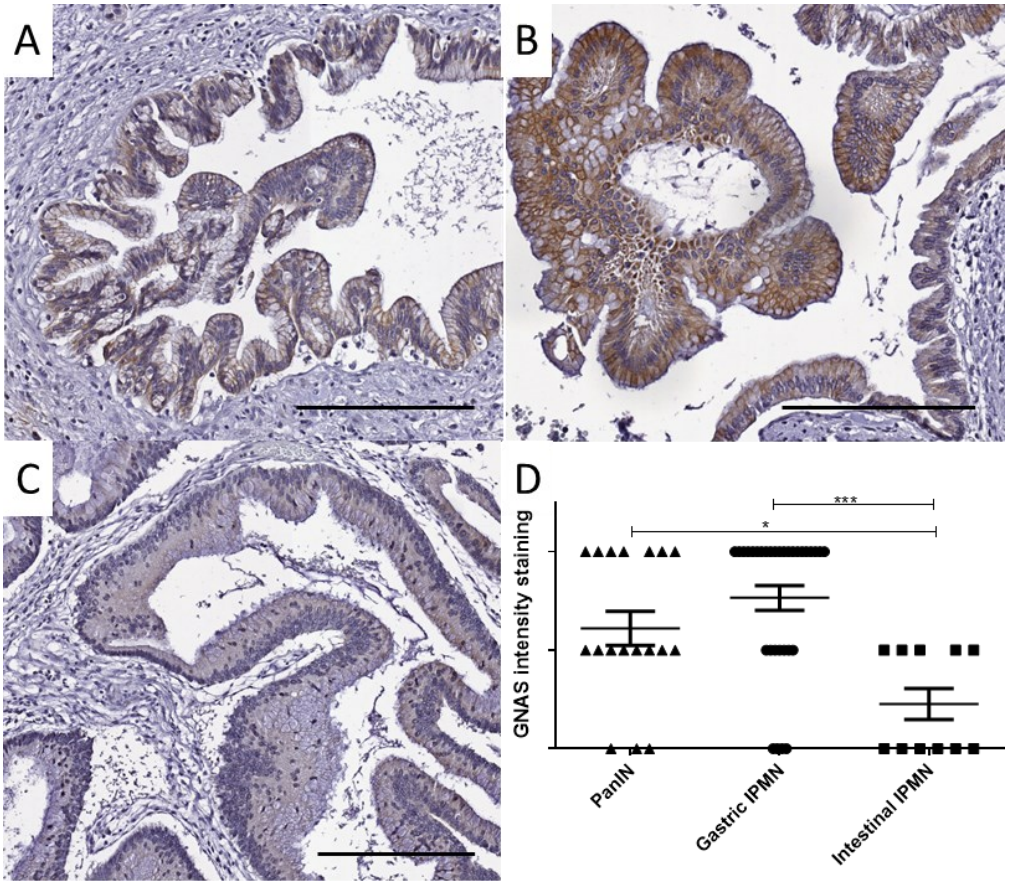


Figure 24: GNAS expression in gastric IPMN, PanIN and intestinal IPMN.
The expression pattern from TMA analysis of 61 samples is presented. Representative images of (A) moderate stained PanIN lesion, (B) strong stained gastric IPMN and (C) weak/ no stained intestinal IPMN. (D) The quantitative analysis of GNAS staining intensity in PanIN, gastric IPMN and intestinal IPMN lesions; (scale bar= 200 µm, Kruskal-Wallis-test * $P < 0.05$; *** $P<0.001$; n= 61).

During this study, the transcriptional profile of different precursor lesions as well the methylation profile was done.

The transcription data of *GNAS* showed no differences in the expression between PanIN, gastric IPMN, and PDAC. The expression in acinar tissue tended to be lower than in the other groups (Figure 25 A). The mutation status of these samples was known, all gastric IPMN harbored a *GNAS* mutation with a VAF of about 3 %, and in three from four PanIN lesions no *GNAS* mutation was detected, the status of the fourth lesion is unknown.

The status of *GNAS* in the methylome dataset could be used for the separation of intestinal IPMN, PanIN and gastric IPMN. However, the status of GNAS between PanIN and gastric IPMN showed no significant different.

The right panel of the following figure showed the methylation profile of all subtypes in an average beta-value plot (Figure 25 B). The methylation threshold was 0.4. The CpGs with a red color were methylated, the CpGs in bluish colors were unmethylated. The displayed samples showed the average value of each sample type. The clustering of the rows was according to the Euclidean algorithm.

The clustering at the left side of figure B showed the near relationship between gastric IPMN samples and the low grade PanIN samples, which was nearer to the gastric IPMN as to the PanIN high grade lesions. The most different methylation patterns were shown in the intestinal IPMN and the beta cell cluster. These patterns reflected the distribution of the MDS plot (Figure 33). All precursors were high methylated at the transcription start site (TSS).

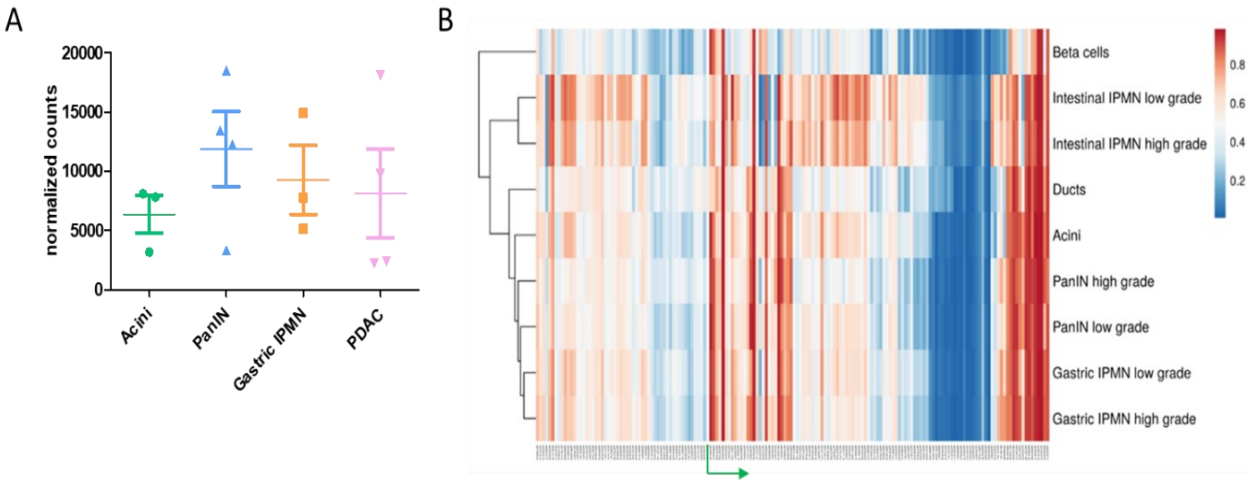


Figure 25: Transcription profile and methylation pattern of *GNAS*.
(A) The normalized read counts of *GNAS* from the transcription data was shown. As normalization methods, DESeq and DEseq2 were used. 3 acinar tissue samples, 4 PanIN, 3 gastric IPMN and 4 PDAC samples were used for the analysis (n= 14). (B) The heatmap represents the methylation pattern in different precursor lesions of the gene locus *GNAS*. (Clustering rows according to Euclidean; average beta value per subtype (n= 49 samples), red= strong methylation, blue= weak methylation; green arrow indicates transcription start site).

4.5 Transcriptome Data

The whole-genome transcription data was generated with 14 different FFPE samples. This cohort consisted of 3 gastric IPMN low grade, 4 PanIN low grade, 3 acinar tissue samples and 4 PDAC samples. As a first step, a principal component analysis (PCA) was done to reduce the dimension and to show the distribution among the different subtypes.

Here, the clustering, according to the 100 most variable genes of all transcriptomic samples (n= 14) of this study is shown (Figure 26).

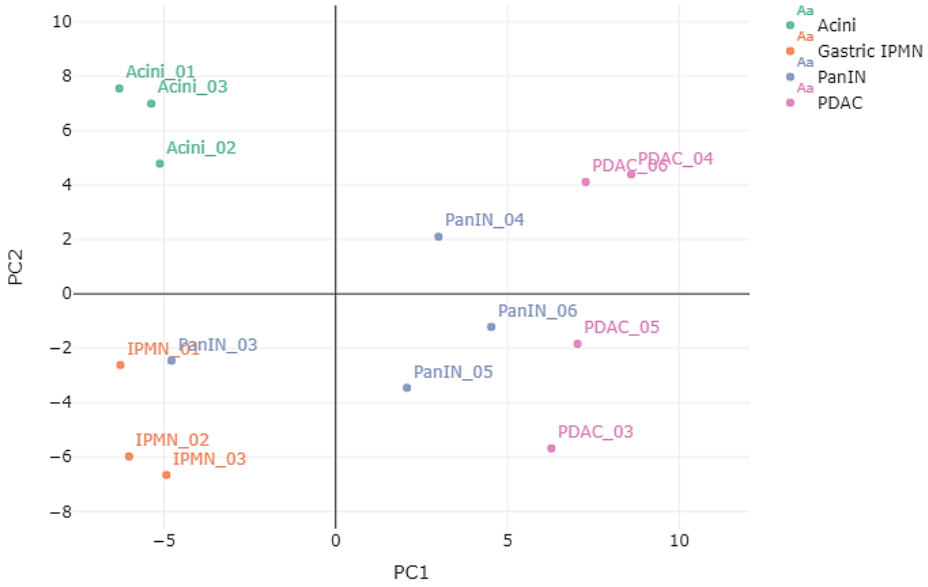


Figure 26: PCA Plot (2D) of all FFPE transcriptome samples.
PCA was performed on 100 top genes selected by the most variable genes. AS value of 0.106. Log(x+1) transformation, centered and scaled. PC1 (35.95 %) and PC2 (23.95 %) – the proportion of variance. TCC-GUI. n= 14, four groups.

Each group clustered in a defined area without overlapping. The acinar tissue samples (green) clustered in the upper left corner. The gastric IPMN samples (orange) clustered in the lower-left corner, one PanIN samples was closer to the gastric IPMN cluster than to the PanIN cluster. Also, the PDAC samples (violet) clustered in one area. So that all different morphological subtypes could be distinguished according to the PCA plot. Gastric IPMN were closer to the acinar tissue samples than to the PanIN samples (blue). However, the distance at the PC2-axis was smaller as at the PC1-axis. The average silhouettes (AS) value was 0.106 for all samples. The value ranged from 1.0 to -1.0, higher values indicated a higher degree of group separation and therefore a higher percentage of differently expressed genes (DEGs)[181].

As a normalization method, the DESeq and DESeq2 algorithms from the TCC-GUI R package were used. 1547 genes were DEGs with a q-value < 0.01 among all samples.

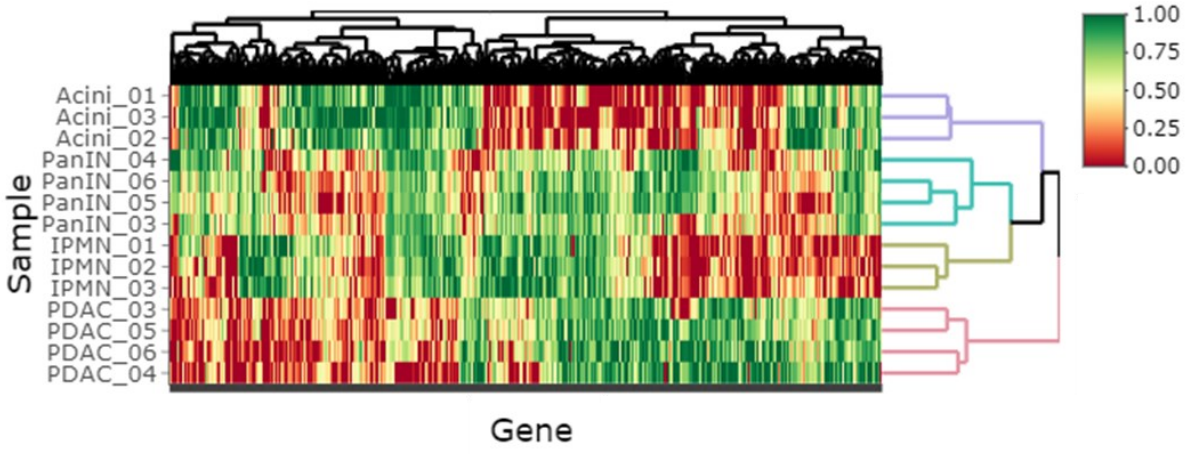


Figure 27: Heatmap of DEGs from the transcriptome data of FFPE samples.
The heatmap represents 1547 significantly different expressed genes (q-value < 0.01) in acinar tissue, gastric IPMN, PanIN and PDAC. In rows are the samples and in columns are the different genes. The hierarchy clustering is done with the Euclidean algorithm. (green= high expressed, red= low expressed, normalized; log (1+x) transformed, min-max, (n= 14)).

The heatmap (Figure 27) shows the DEGs with a q-value of < 0.01. All samples clustered according to the belonging groups by using the Euclidean algorithm in rows. The four different precursor subtypes were well separated.

There were some genes, which were downregulated in the precursor and PDAC in comparison to the control acinar tissue samples. However, there were also some genes upregulated to the controls. PanIN and gastric IPMN were closer to each other than to acinar tissue or PDAC, which was shown in the hierarchy clustering at the right site (Figure 27). Also, the precursors were closer to the acinar tissue than to PDAC in this dataset of genes.

A gene list analysis with PANTHER was done to identify the involved molecular function. In total, 1386 genes were found in the PANTHER database and 919 genes consisted of a known molecular function. The two main molecular functions of these genes belonged to binding (39.8 %) and catalytic activity (34.5 %). The other functions were divided into transporter activity (7.4 %), molecular function regulator (5.5 %), transcription (5.4 %), translation regulator activity (0.4 %), molecular transducer activity (4.9 %) and structural molecule activity (2.0 %).

KEGG 2019 Human pathway library was used to investigate the pathways, which were involved in the 1547 DEGs. The three main involved pathways according to the KEGG pathway analysis were the pancreatic secretion pathway (31/98 genes), protein digestion and absorption (24/90 genes) and regulation of actin cytoskeleton (40/214 genes). These three pathways had the highest overlapping with the DEGs. The p-values of these different pathways were $4.99e^{-12}$, 2.32^{-9} and $1.49e^{-7}$, respectively.

Table 23: DEGs and AS value of each group of the transcription data.

	AS value	No. of DEGs
Acinar tissue - PDAC	0.321	1160
Acinar tissue - gastric IPMN	0.263	761
Acinar tissue - PanIN	0.273	951
PDAC – gastric IPMN	0.122	364
PDAC - PanIN	0.251	894
PanIN – gastric IPMN	0.183	595

The average silhouettes and the number of differently expressed genes between each group were represented in the table (q-value < 0.01, n= 14) (Table 23).

The table showed the AS values of each group to each other. The best separation according to the AS value was between acinar tissue and PDAC samples with an AS value of 0.321 and 1160 DEGs. The worst separation in this dataset according to the cluster overlapping was the gastric IPMN and PDAC with an AS value of 0.122 and 364 DEGs (q-value < 0.01).

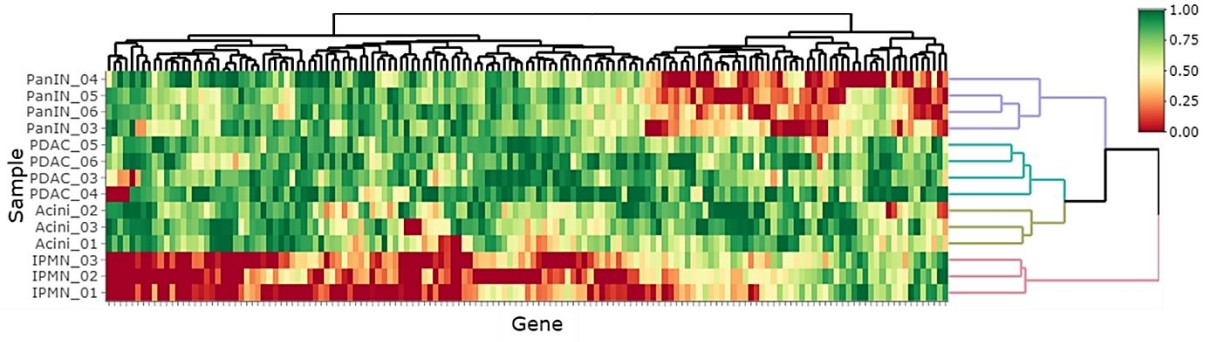


Figure 28: Unique downregulated genes in gastric IPMN and PanIN lesions of the DEGs.
The heatmap represents the unique downregulated genes from gastric IPMN and PanIN. These genes belong to the 1547 DEGs from the whole dataset. Only downregulated in PanIN (53 genes) and only downregulated in gastric IPMN (94 genes) are shown. (q-value < 0.01, green= high expressed, red= low expressed, clustering with Euclidean algorithm; normalized, log (1+x) transformed (n= 14)).

The DEG heatmap between all samples (Figure 27) showed two clusters, in which only PanIN or gastric IPMN were involved. In these two clusters, the precursor was downregulated in comparison to all other groups. The two clusters were presented in the heatmap above (Figure 28). The delimitation in the PanIN samples was not so strong as in the gastric IPMN cluster. For the PanIN cluster, there were 53 genes involved. For these genes, no significant involved pathway was detected with the webtool enrichr.

The delimitation from the gastric IPMN cluster was much better. Here 94 genes were involved. The KEGG pathway analysis showed that the cell adhesion molecules (CAMs) pathway with a p-value of $6.113e^{-4}$ was mainly involved.

A PCA plot only with the two precursor lesions gastric IPMN and PanIN is shown below (Figure 29). The separation value AS was 0.183. The sample PanIN_03 was closer to the cluster of gastric IPMN than to the cluster of PanIN samples.

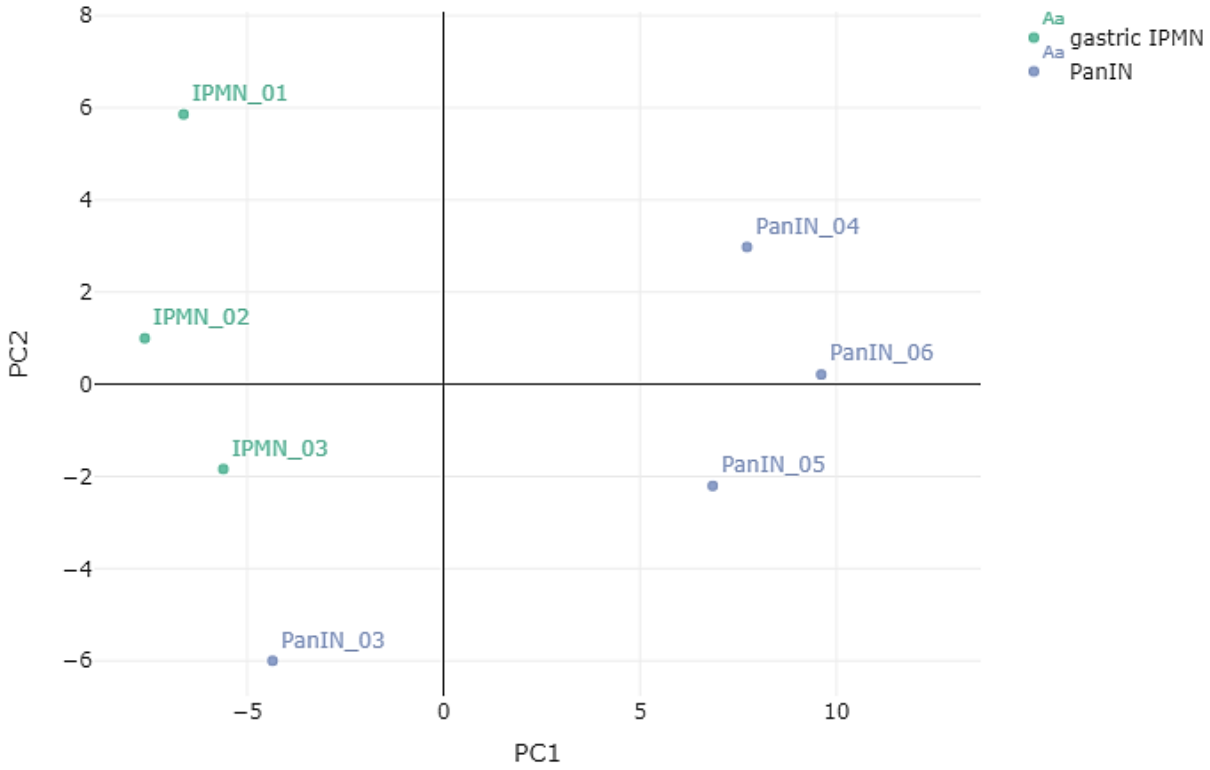


Figure 29: PCA Plot with gastric IPMN and PanIN lesions.
The PCA plot shows the clustering of the 100 most variable genes. The AS value was 0.183. (green= gastric IPMN, blue= PanIN; log (x+1) transform; PC1: 58.495 %; PC2:14.748 %; (n= 7)).

The DESeq2 algorithm was used to determine the DEGs between PanIN and gastric IPMN. 595 DEGs between PanIN and gastric IPMN with a q-value < 0.01 (FDR rate 1 %) were found.

A volcano plot was generated to show the up- and downregulated genes in the two precursor lesions (Figure 30). The green dots represented the downregulated genes in PanIN in comparison to gastric IPMN. The red dots represented the upregulated genes in PanIN in comparison to gastric IPMN.

The cut-off of the fold change (x-axis) was +/- 1 and the p-value cut-off (y-axis) was < 0.01. In PanIN samples, 526 genes were downregulated, and 1273 genes were upregulated in comparison to gastric IPMN.

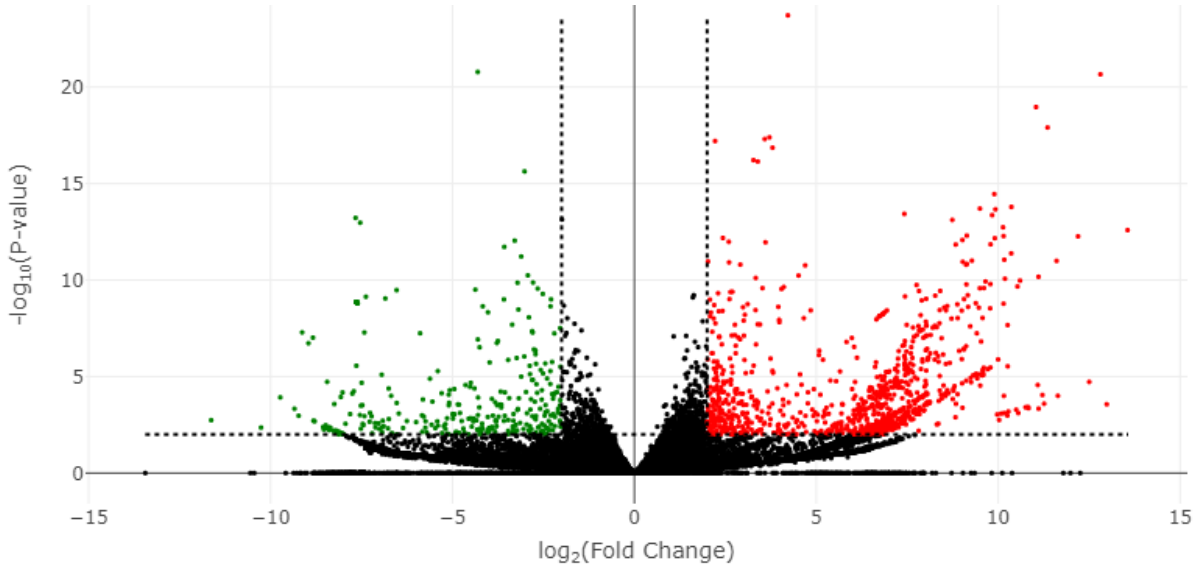


Figure 30: Volcano plot of different expressed genes in PanIN lesions in comparison to gastric IPMN.
526 genes were downregulated (green) in PanIN and 1273 genes were upregulated (red) in PanIN in comparison to gastric IPMN. (Fold change: -1 and 1; q-value 0.01 (n= 7)).

A list of 237 downregulated genes with a q-value < 0.001 was used for a pathway analysis with the webtool enrichr. The main involved KEGG pathway was mucin-type O-glycan biosynthesis with a p-value of $4.704e^{-4}$. In the Panther database, 186 genes were present, and from 121 genes the molecular function was known. The main molecular functions of the involved genes were belonging to binding (45 genes; 45.5 %) and catalytic activity (32 genes; 32.3 %).
The list of 282 upregulated genes with a q-value < 0.001 was used for a pathway analysis as well. The two mainly involved KEGG pathways were ribosome biogenesis in eukaryotes and ribosome with p-values of $8.713e^{-13}$ and $5.877e^{-11}$, respectively. In the Panther database, 175 genes were present, and from 99 genes, the molecular function was known. The main molecular functions of these genes were connected to binding (44 genes; 36.4 %) and catalytic activity (36 genes; 29.8 %).

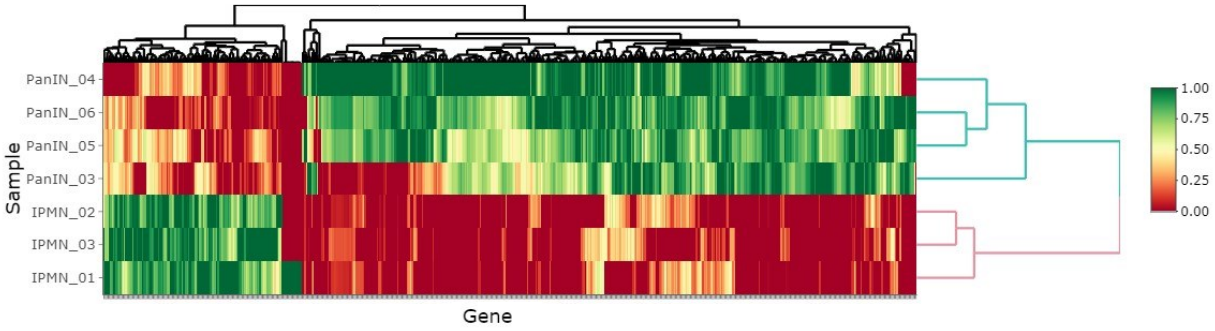


Figure 31: Heatmap of DEGs between gastric IPMN and PanIN.
The heatmap represents 595 DEGs between gastric IPMN and PanIN lesions. In rows are the samples and in columns are the different genes. The hierarchy clustering in rows and columns were done with the Euclidean algorithm. (green= high expressed, red= low expressed, normalized; log (1+x) transformed, q-value < 0.01, (n= 7)).

In PanIN and gastric IPMN, 595 different expressed genes with a p-value < 0.01 were present (Figure 31). So, 1.27 % (595/46697) of all sequenced genes were DEGs in PanIN and gastric IPMN. The molecular functions were determined. PANTHER analysis was done with 595 genes, only 416 genes were present in the PANTHER database and 250 out of it have a known molecular function. The main molecular function of these involved genes belonged to binding (98 genes; 39.2 %) and catalytic activity (82 genes; 32.8 %).

Pathway analysis with the webtool enrichr showed three mainly involved KEGG pathways between gastric IPMN and PanIN. The Ribosome biogenesis in eukaryotes (16/101 genes) pathway with a p-value of $4.741e^{-8}$, the Ribosome (17/153 genes) pathway with a p-value of $3.235e^{-6}$ and the Proximal tubule bicarbonate reclamation (4/23 genes) pathway with a p-value $4.37e^{-3}$ were the pathways with the highest overlapping and p-value from the KEGG pathways.

4.6 DNA Methylation pattern

The methylation patterns of 49 samples were investigated during this study. The pie chart shows the distribution of the used precursor lesions and control samples (Figure 32). The largest group was the gastric IPMN with 18 samples (14 low grade and 4 high grade). The intestinal IPMN group consisted of 10 samples (5 low grade and 5 high grade). The PanIN lesion was involved with 8 samples (6 low grade and 2 high grade). As control groups, three ductal cell samples and three beta cells samples were included. In addition, seven acinar tissue samples were involved with matching precursor lesions from the same patients.

A multidimensional scaling (MDS) plot with all samples of the whole-genome DNA methylation array was presented below (Figure 33). There were six different groups, as the control groups were beta cells, ductal cells, and acinar tissue samples. As a comparison of different PDAC precursor lesions, there were PanIN, gastric IPMN, and intestinal IPMN lesions. The brownish cluster with beta cells was far away from all other groups. The intestinal IPMNs showed a second well-separated cluster in blue. Acinar tissue showed a tight cluster in red. The PanIN samples in pink/violet overlapped the acinar tissue (red) and the gastric IPMN (green) cluster. The cluster of duct cells (khaki) showed in the direction from acinar tissue to the gastric IPMN.

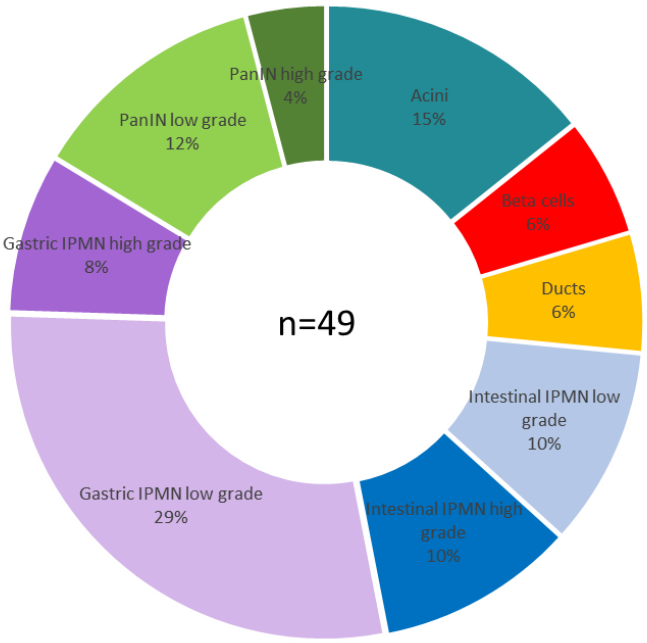


Figure 32: Pie chart of the precursor distribution for the methylome analysis.
For the methylome analysis, a tissue collection of 49 samples was used (n= 49). The collections consisted out of 8 PanIN (6 low grade and 2 high grade), 18 gastric IPMN (14 low grade and 4 high grade), 10 intestinal IPMN (5 low grade and 5 high grade), 7 acinar tissue samples, 3 beta and 3 ducts cells samples.

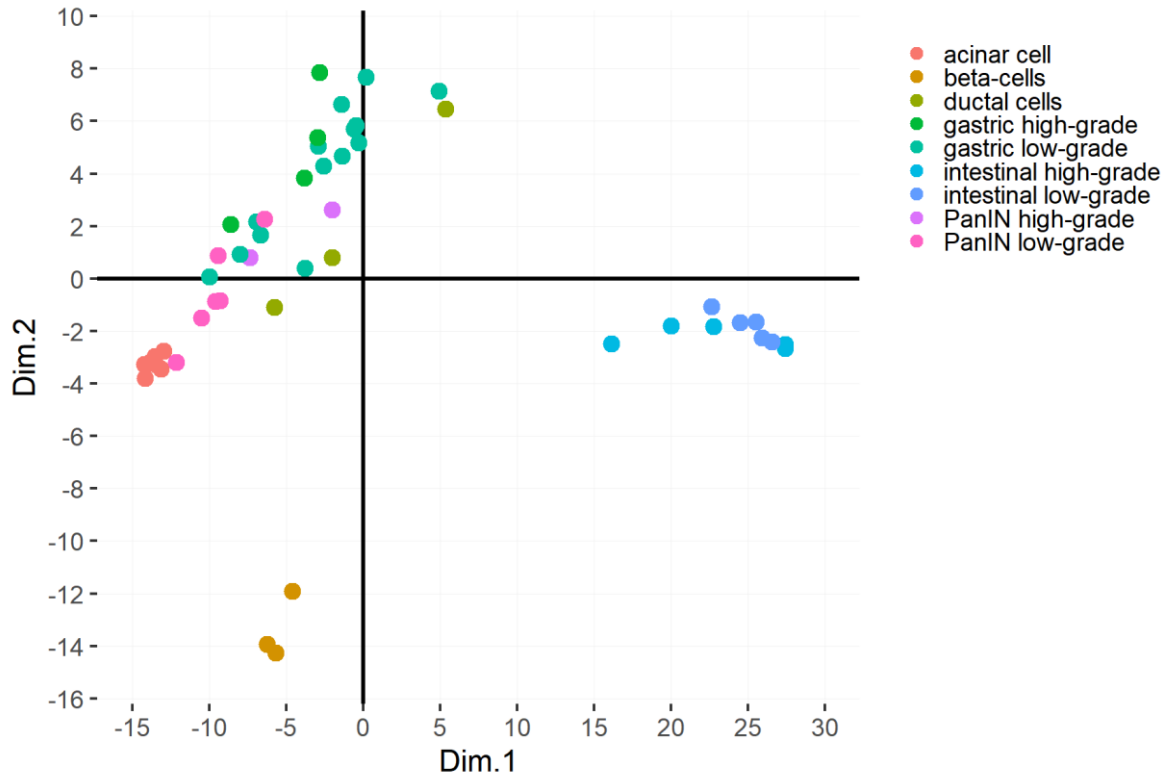


Figure 33: Multidimensional scaling plot of the DNA methylation data.
MDS shows the 100 most variable positions from the array analysis. 6 different groups were included. (brownish= beta cells, khaki= ducts, red= acini, pink/violet= PanIN (light= low grade; dark= high grade), green= gastric IPMN (light= low grade; dark= high grade), and blue= intestinal IPMN (dark= low grade; light= high grade); (n= 49)).

In total, the Illumina Infinium MethylationEpic Array consisted of 865918 cg sites distributed over the whole-genome, after ChAMP and ComBAT analysis, 727809 loci (84.05 %) were included in the following analysis.

The volcano plot (Figure 34) represents the differently methylated cg-sites between each subtype. The blue dots showed higher methylation in the subtype on the y-axis and red dots represented higher methylation in the subtypes on the x-axis. The black dots showed a similar methylation pattern between the two involved subtypes.

According to the MDS, the beta cells showed the highest number of differentially methylated genes between each subtype. In addition, the methylation profile of intestinal IPMN was clearly differently methylated to the other groups.

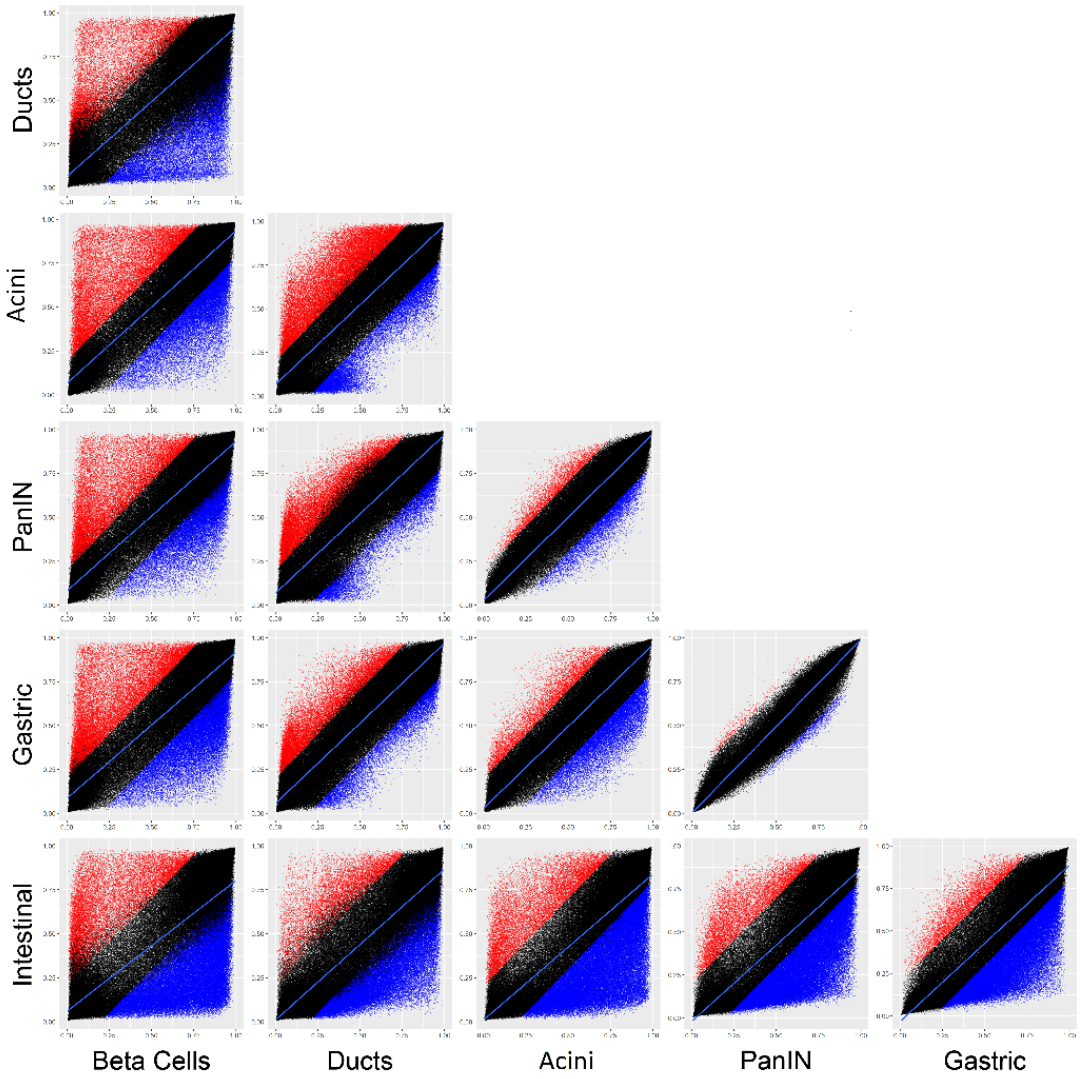


Figure 34: Volcano plots represent the difference methylated genes of each sample type.
The volcano plot represents the beta values from two different groups. Black dots represent a similar methylation pattern of the cg sites. The red dots show strong methylated of the group on the y-axis and the blue dots show strong methylated cg sites of the group on the x-axis. (n= 49, 6 different groups).

The pairwise comparison of the control groups revealed that 14 % (acinar vs. ductal cells) to 15 % (ductal and beta-cells) were differentially methylated. The highest degree of differential methylation was detected between intestinal IPMNs and beta-cells (27.4 %), whereas the smallest change with only 0.6 % (4066 cg sites) differentially methylated probes was observed between PanIN and gastric IPMNs lesions. The 4066 cg sites included 1779 genes. These genes were significantly differently methylated at least at one position.

The gene list of 1779 genes was evaluated with the KEGG pathways to analyze the involved pathways. The involved pathways were the circadian entrainment (23/97 genes) with a p- value of $1.2E^{-5}$. The other two significant involved KEGG pathways were morphine addiction (19/91 genes) with a p-value of $3.49E^{-4}$ and the inflammatory mediator regulation of transient receptor potential (TRP) channels (19/100 genes) with a p-value of 0.0012.

The methylome data was also used for the generation of a phylogenetic tree. The phylogenetic tree displayed the relationship between precursor lesions and cell types based on the methylation profile. The length of the arms of the unrooted tree represented the distance between each sample based on the different methylation profile.

Figure 35: Distance-based, neighbor joining tree calculated from methylation status of 49 different precursor samples. Unrooted Neighbor-joining tree (phylogenetic tree); Phylogenetic tree displays the relationship between precursor lesions and cell types based on methylation data (n= 49).

In the middle of the tree were the beta cells, ductal cells, and acinar tissue, all three groups clustered next to each other. On the left side of the ductal and beta cells, the intestinal IPMN were the next corresponding samples. On the right side of the acinar tissue, first, the PanIN samples and then the gastric IPMN were displayed. This phylogenetic tree could represent the hypothetic origin of the precursor lesions, only based on the methylation profile.

The distribution of the differently methylated genes (DMGs) between PanIN and gastric IPMN were evaluated. The DMGs were classified into two groups, hypermethylated and hypomethylated. The cut-off was set for the adjusted p-value below 0.05 and to the Δbeta value > 0.2 or < (-0.2). These cutoffs included 4066 CpG sites, which covered 1779 different genes.

The distribution of the differentially methylated sites was examined in various functional genomic regions. Within the hypermethylated genes, the rate was higher in the body and near the transcription start site (TSS), such as 5'UTR and TSS1500, while within the hypomethylated genes, the rate was higher in the first exon and TSS200 (Figure 36 A). The body and the

intergenic region (IGR) showed the largest number of hypermethylated genes in the DMGs between PanIN and gastric IPMN.

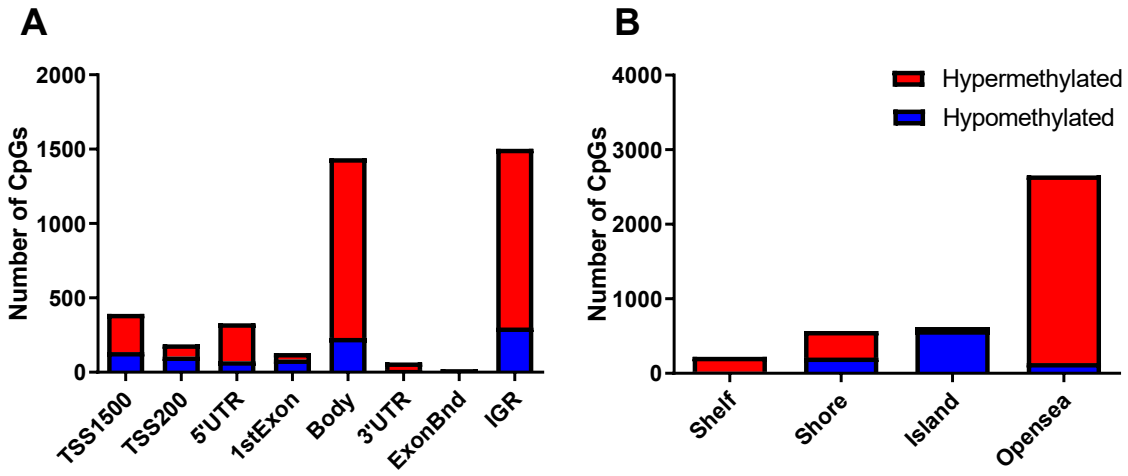


Figure 36: CpGs differently methylated characterize according to the CG annotation between PanIN and gastric IPMN. (A) Distribution of CpGs in different genomic regions divided into hypomethylated (blue) and hypermethylated (red). 4347 CpGs were differently methylated in PanIN in comparison to gastric IPMN. (B) CpGs in different regions related to CGIs features divided into hypomethylated (blue) and hypermethylated (red).

The most important CpG sites were in the promoter region. The methylation in these regions mainly affected the transcription rate. Here, the promoter region was defined as TSS1500, TSS200 or 5'UTR. 991 differently methylated CpGs were located in the promoter region by comparing PanIN and gastric IPMN. These CpGs covered 698 different genes, among them 477 were hypermethylated and 227 were hypomethylated genes in PanIN in comparison to gastric IPMN. Six of these genes consisted of hyper- and hypomethylated CpGs in the promoter region.

Enrichment analysis of the 698 genes showed three significant enriched pathways. The tight junction (13/170 genes; p-value 0.007), circadian entrainment (7/97 gens; p-value 0.007) and phosphonate and phosphinate metabolism (2/7 genes; p-value 0.02).

Correlation analysis of the methylation and expression of all DMGs and DEGs by comparing PanIN and gastric IPMN was done in the next part as an integrative analysis.

4.7 Integrative analysis: gene expression and methylation

In the next step, the gene expression and methylation data were combined in an integrative analysis to identify PanIN- and gastric IPMN-specific markers. Integrative analysis of DNA methylation and gene expression was conducted by determining the intersection between the differently methylated genes and differentially expressed genes in PanIN lesions in comparison to gastric IPMN. The genes were classified into four groups based on the intersection between DMGs and DEGs: hypermethylated-upregulated, hypermethylated-downregulated, hypomethylated-upregulated and hypomethylated-downregulated genes.

A gene was considered to be differently methylated if the adjusted p-value was below 0.05 and the Δbeta value was > 0.2 or < (-0.2). So, 1626 genes were differently methylated between PanIN and gastric IPMN, including 1274 hypermethylated genes and 352 hypomethylated genes. In addition, a gene was considered to be differentially expressed if the adjusted p-value was below 0.05 and possessed a fold change of > 1 or < (-1). Consequently, 660 genes were classified as differentially expressed, including 176 downregulated genes and 484 upregulated genes in PanIN in comparison to gastric IPMN.

From the 1626 DMGs and 660 DEGs, the integrative analysis resulted in a list of 84 genes, in which the expression and the methylation were significantly different in PanIN in comparison to gastric IPMN (Figure 37).

Since the transcriptional silencing of tumor suppressor genes by aberrant hypermethylation of promoters is one of the most frequently observed alterations in cancers, the focus of the further analysis was set on 45 hypermethylated-downregulated genes, including 18 genes with hypermethylated CpGs in the promoter region. Gene Ontology (GO) analysis of the 45 hypermethylated-downregulated genes revealed a significant enrichment of genes involved in digestion ($p < 0.0001$), carbohydrate metabolic process ($p = 0.007$) and O-glycan processing ($p = 0.008$).

KEGG pathway analysis demonstrated three significant enriched pathways, one related to mucin O-glycan biosynthesis ($p = 0.002$), one to glycerol lipid metabolism ($p = 0.008$), and one to phenylalanine metabolism ($p = 0.03$). Importantly, O-glycan biosynthesis had been reported to be associated with several pathological conditions, including cancer progression[182,183].

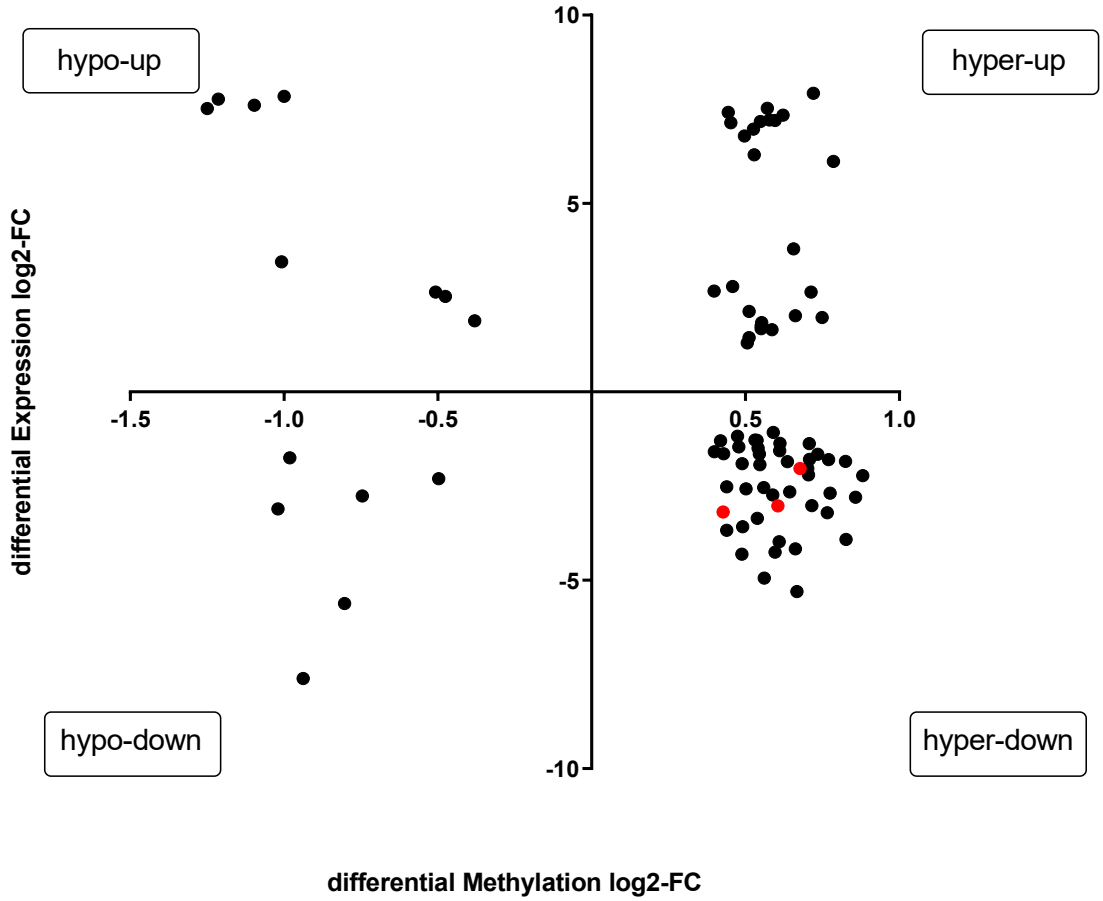


Figure 37: Integrative analysis of DNA methylation and gene expression between PanIN and gastric IPMN.
Differently methylated and differentially expressed genes are shown in the scatterplot (n= 84). The red marked genes are KLF4, MUC6 and TFF1. The highest deltabeta value in the gene was considered. Threshold: Δbeta value >0.2 or <-0.2 and q<0.05 for differently methylated genes and p<0.05 and a fold change of >1 or <-1 for differently expressed genes.

A protein-protein-interaction analysis via the string-db.org platform was done for investigating the predicted interaction between the hyper-down genes.

The 45 hyper-down genes were included without any further interaction partner. All nodes without a connection were removed and the interaction score was set a >0.4 (Figure 38). From the resulting network of 15 genes, three hub genes were selected.

Kruppel Like Factor 4 (KLF4), Trefoil Factor 1 (TFF1) and Mucin 6 (MUC6) were selected. These genes were chosen according to the high interaction score of 0.6 or higher and the known function as a tumor suppressor (KLF4 and TFF1) and differently methylated in the promoter region (TFF1), as well as the known high expression in PDAC in comparison to 39 different cancer types (TCGA, n= 10580) of MUC6. In addition, the MUC6 gene belongs to the O-glycan mucin biosynthesis pathway, which was enriched in the pathway analysis of the genes from the integrative analysis.

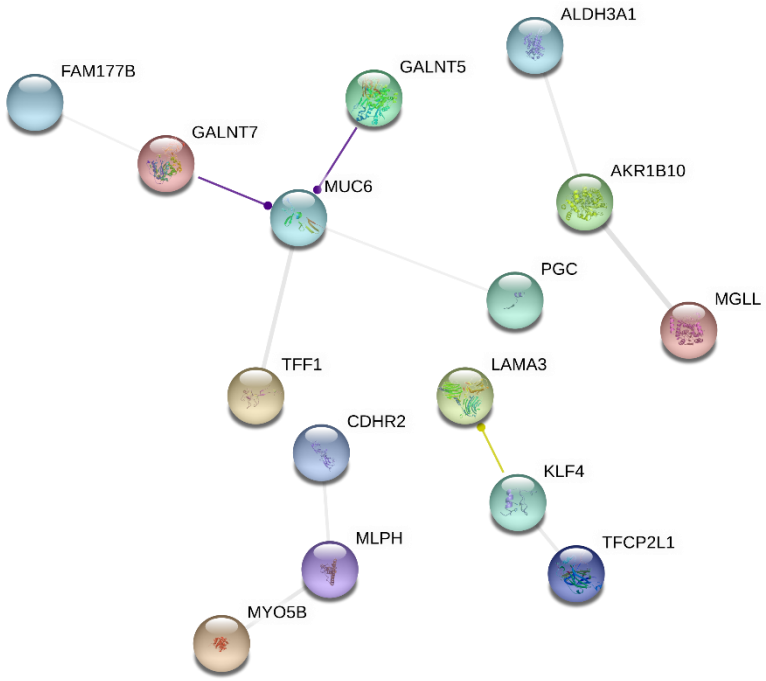


Figure 38: Protein-protein interaction network of hypermethylated and downregulated genes.
The PPI network of hyper-down genes (45 genes) with a confidence > 0.4 represents the interaction between the different proteins. The lines represent the action type, purple means catalyzed and yellow lines show transcriptional regulation. The dots at the end represent unspecific action effect. The nodes without any connection were hidden.

A detailed investigation of the molecular pattern of *MUC6*, *TFF1* and *KLF4* were done. The mRNA expression and the DNA methylation were observed. For all three different genes, the RNA expression was significantly higher in gastric IPMN in comparison to the PanIN lesions. The heatmaps of *MUC6* did not show any obvious differences in the methylation pattern in the entire gene locus of *MUC6*. The intestinal IPMN showed reduced DNA methylation over the whole gene locus of *MUC6*.

The methylation pattern of *TFF1* was different over the whole locus between gastric IPMN and PanIN, gastric IPMN showed less DNA methylation than PanIN lesions. Intestinal IPMN did not show any methylation of the gene locus of *TFF1*. The last hub gene *KLF4* was completely unmethylated in all three precursor lesions. Only the body of the gene of PanIN lesions was methylated at two CpG sites, the same as in the acinar tissue control group.

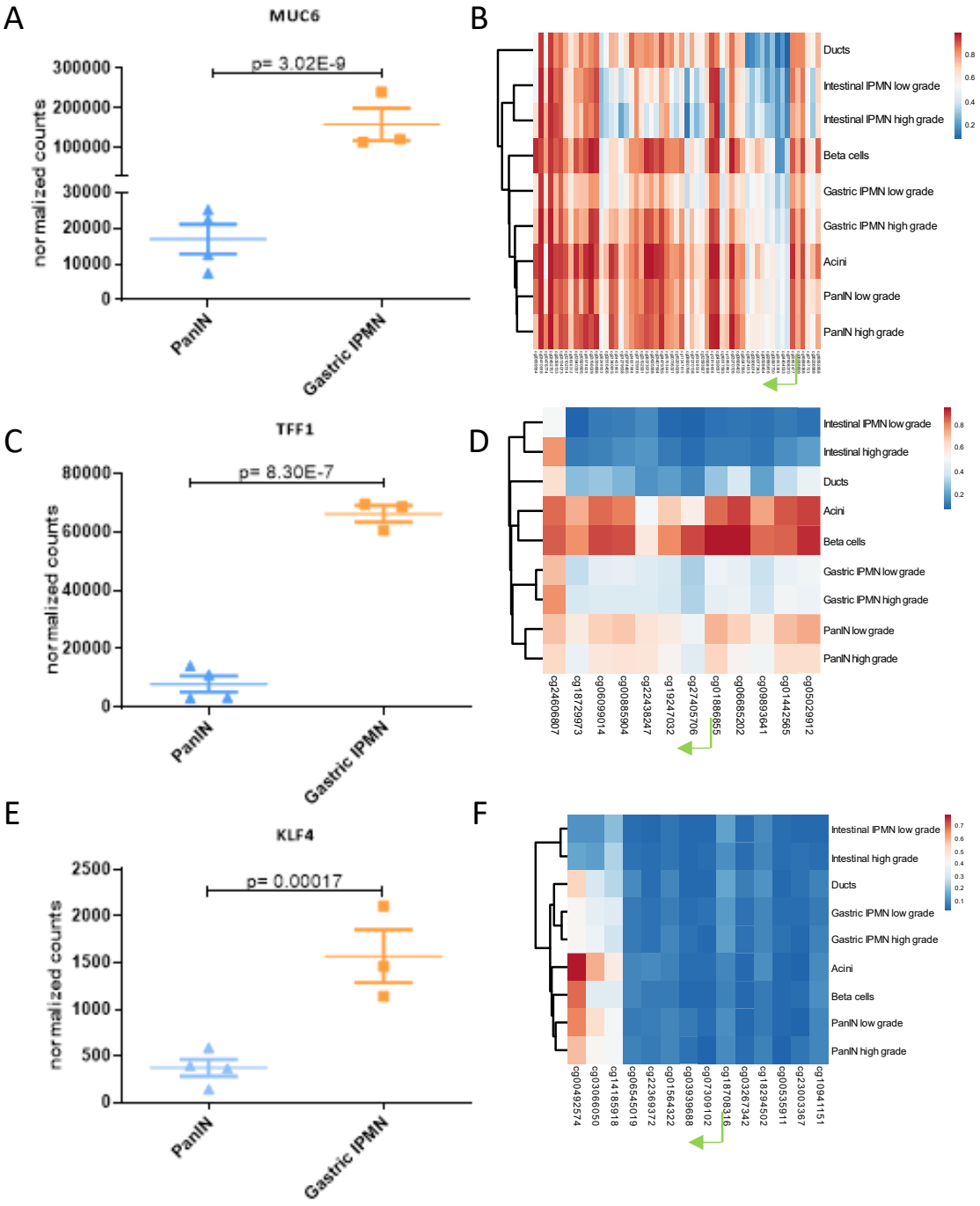


Figure 39: Expression and methylation pattern of MUC6, TFF1 and KLF4 in different precursors.
(A) The normalized readcounts of MUC6, (C) TFF1 and (E) KLF4 from the transcription data are shown. As normalization methods, DESeq and DEseq2 were used. 3 acinar tissue samples, 4 PanIN, 3 gastric IPMN and 4 PDAC samples were used for the analysis (n= 14). (B) The heatmap represents the methylation pattern in different precursor lesions of the gene locus MUC6; (D) gene locus of TFF1 and (F) gene locus of KLF4. (Clustering rows according to Euclidean; average beta value per subtype (n= 49 samples), red= strong methylation, blue= weak methylation; green arrow indicates transcription start site).

To further investigate the expression of the three identified genes, immunohistochemistry was performed using TMAs containing 27 PanIN, 33 gastric IPMN, 16 intestinal IPMN, 38 PDAC tissue cores, and 66 acinar tissue control tissue cores obtained from 71 different patients (Figure 40). In the normal pancreas, MUC6 was expressed in centroacinar cells in 48/65 cases (73.8 %), whereas acinar cells were always negative. MUC6 was expressed with comparable frequency in PanIN (mean 44.8 %) and gastric IPMN (mean 46.8 %), where it was mostly confined to the non-papillary site of the lesions. Intestinal IPMN (mean 15.8 %) were significantly less frequently positive for MUC6 compared to gastric IPMN (p= 0.0228). PanIN and gastric IPMN showed a wide range of expression pattern for MUC6 positive cells. In PDAC, about 10 % of the tumor cells of all cases were stained positive, which were significantly reduced to gastric IPMN (p= 0.0388) (Figure 40, and Figure 41 A). In colloid carcinoma the expression is reduced compared to PDAC.

KLF4 was not expressed in the normal pancreas. Instead, it was strongly expressed in intestinal IPMN (mean IRS 9.4), whereas a significant lower expression in PanIN (mean IRS 3.8; p= 0.0138) was determined. Gastric IPMN (mean IRS 4.5) displayed a comparably weak expression than PanIN without any significances. In conventional PDAC, a wide range from weak to strong expression was detected with a mean IRS of 5.8. In colloid carcinoma, a diffuse, strong expression (IRS 12) was detected in all four samples (Figure 40, Figure 41 B).

TFF1 was not expressed in the normal pancreas, but it was strongly positive in all three precursor lesions. Here, the expression was lower in PanIN than in gastric IPMN. In comparison to PanIN and gastric IPMN, TFF1 expression in conventional PDAC was significantly reduced (mean 47 %, p= 0.03 and p= 0.005). In contrast, the expression in colloid carcinoma was strong and diffuse in all 4 samples (Figure 40, Figure 41 C).

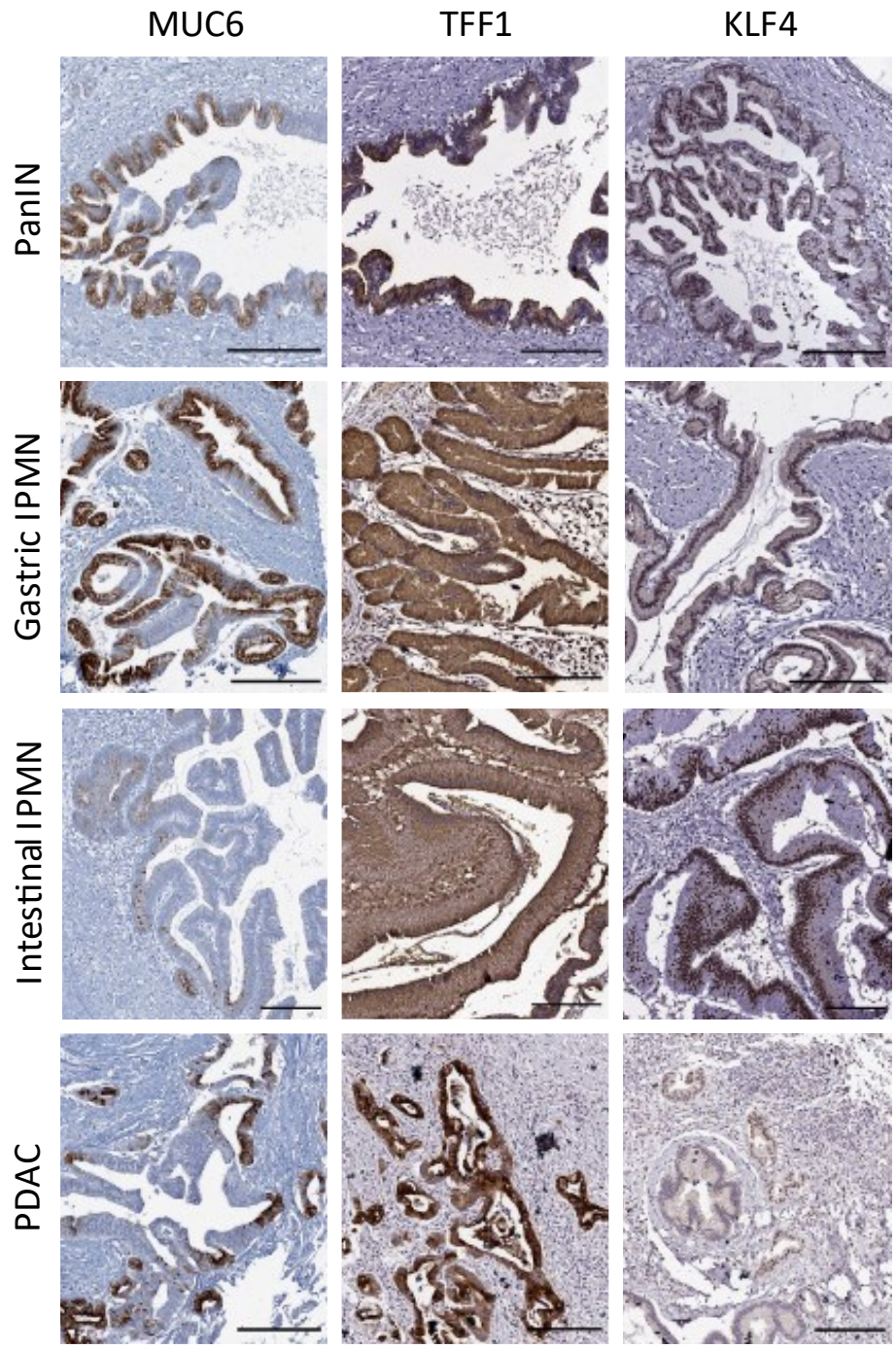


Figure 40: Expression profile of KLF4, MUC6 and TFF1 on precursors of pancreatic cancer.
IHC staining of the indicated proteins in a tissue microarray containing 27 PanIN, 33 gastric IPMN, 16 intestinal IPMN, 66 acinar tissue samples and 38 PDAC tissue. Representative images of each staining in all subgroups are shown in the panels. (Scale bar= 200 µm; n= 71).

No differences in the expression between low and high grade lesions were observed, except for the significantly higher expression of TFF1 in low versus high grade intestinal IPMN (p= 0.0152).

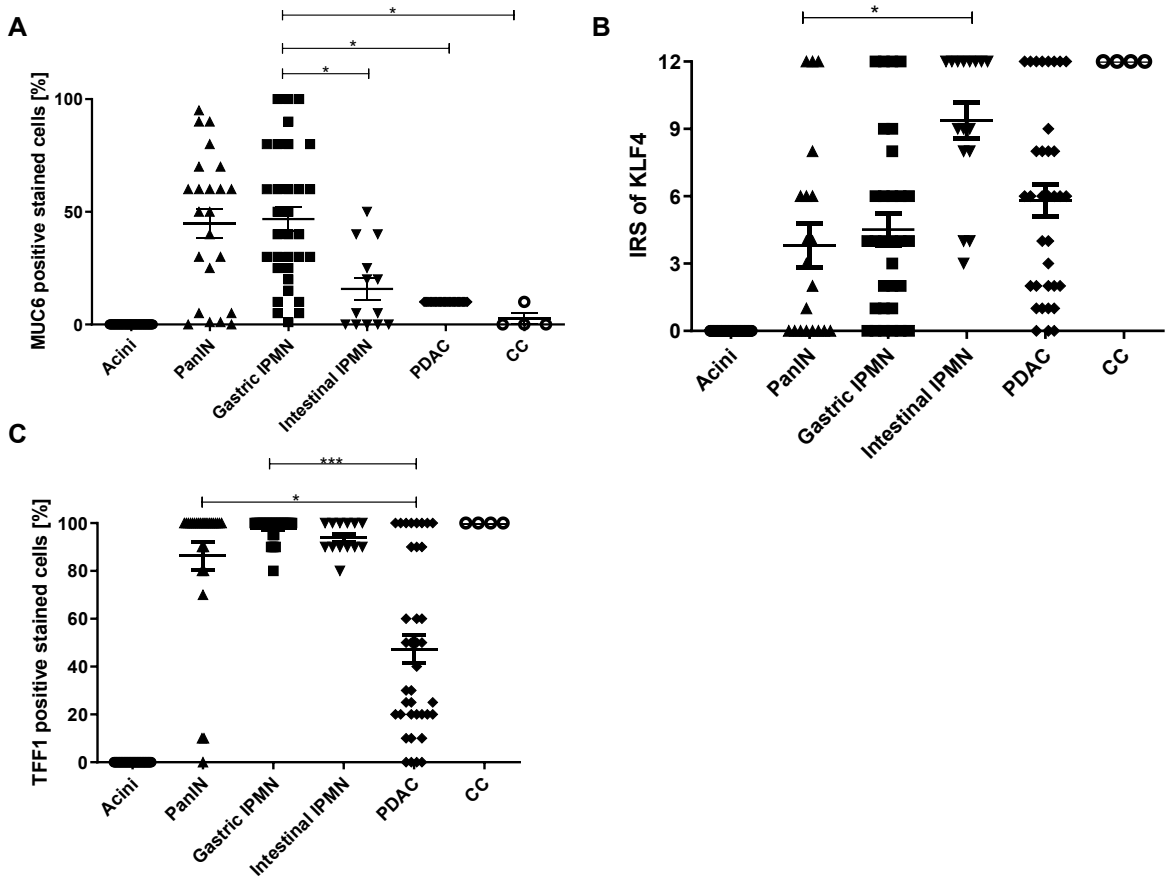


Figure 41: Quantitative analysis of MUC6, KLF4 and TFF1 in precursor lesions, acinar tissue and PDAC tissue.
IHC staining of the indicated proteins in TMAs containing 66 acinar tissue samples, 27 PanIN, 33 gastric IPMN, 16 intestinal, 38 PDAC samples and 4 colloid carcinoma in whole tissue sections. Quantitative analysis of the IHC marker are shown in (A) MUC6 staining, (B) KLF4 staining and (C) TFF1 staining. The immune reaction score or the % of positively stained cells are shown (* p< 0.01, ***p<0.001, Kruskal-Wallis test).

4.7.1 O-glycan Mucin biosynthesis

MUC6 as a gene involved in the O-glycan mucin biosynthesis, the pathway was investigated in detail. The O-glycan mucin biosynthesis is one of the common post-translational modification processes. During this process, the newly synthesized proteins are altered with glycans. It is known that in many different cancer types, glycolysis is involved in the carcinogenesis[184].

The pathway consisted of 50 genes (KEGG pathway). 7/50 genes (14 %) were DEGs between PanIN and gastric IPMN, all were highly overexpressed in gastric IPMN compared to PanIN. The DEGs of the enriched pathway mucin O-glycan biosynthesis are shown in the figure below (Figure 42).

The N-acetylgalactosaminyltransferases *GALNT5*, *GALNT7,* and *GALNT10*, which play a role in the initiation of O-glycosylation by adding N-acetylgalactosamine on the serine/threonine backbone of proteins[167], were significantly overexpressed at the mRNA level in gastric IPMN compared to PanIN. The *GALNT5*, *GALNT7* and *MUC6* were also included in the hyper-down genes from the integrative analysis. Thus, the methylation and expression of these genes were significantly different between PanIN and gastric IPNM.

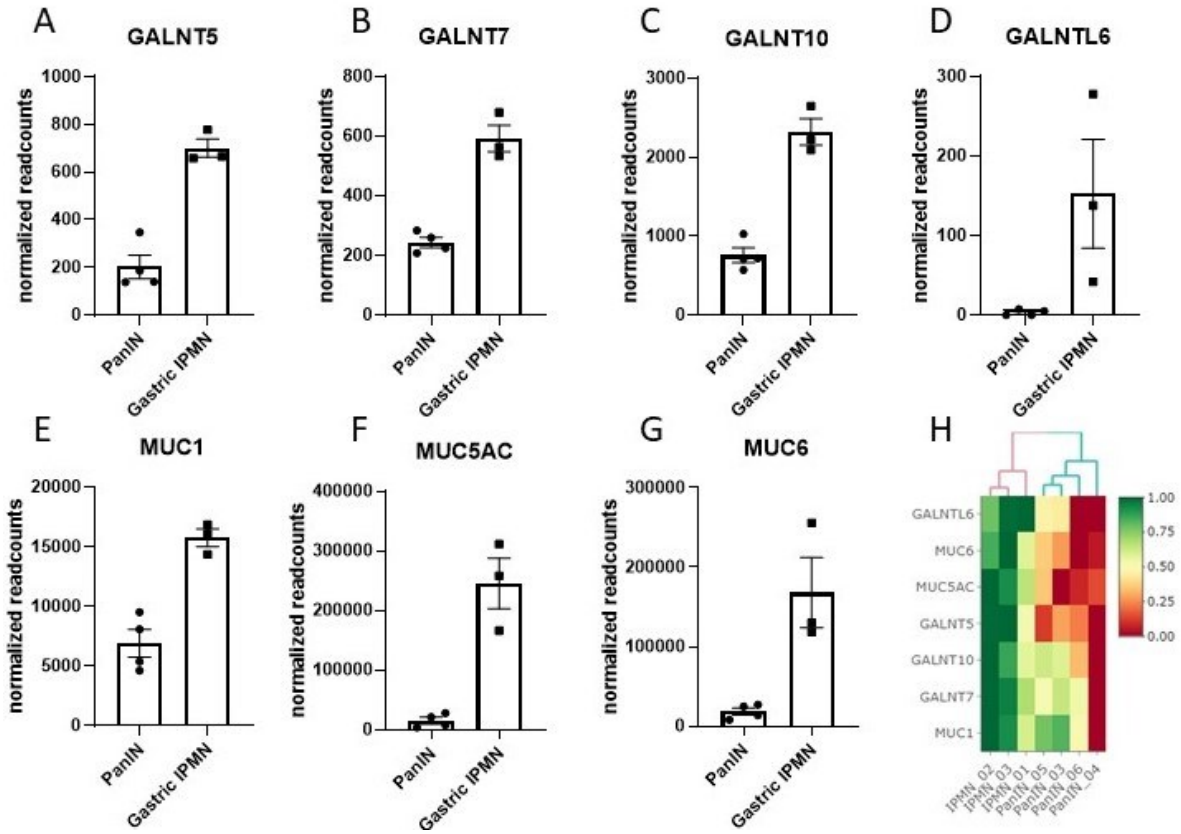


Figure 42: DEGS from O-glycan mucin biosynthesis in gastric IPMN and PanIN.
The graphs are showing the expression of the O-glycan pathway genes in gastric IPMN and PanIN. (A) *GALNT5* expression, (B) *GALNT7* expression, (C) *GALNT10* expression, (D) *GALNTL6* expression, (E) *MUC1* expression, (F) *MUC5AC* expression and (G) *MUC6* expression. The DESeq2 algorithm was used to determine the significant ($p<0.001$). (H) Heatmap of O-glycan mucin biosynthesis DEGs in PanIN and gastric IPMN. (normalized; log (1+x) transformed, q-value < 0.01, (n= 7)).

5 Discussion

During the last decades, the improvement in therapy of pancreatic cancer is unsatisfying. The 5-year survival rate reached only 10 % in 2019. In comparison to the 1970s, the survival rate was at a low level of 3 %[11]. The bad prognosis is due to the fact that an early diagnosis is not possible. Most of the early-stage tumors diagnosed are incidental findings.

The best option for a successful therapy and prognosis would be to screen patients for precursor lesions of pancreatic cancer. Currently, the only curative treatment is the resection of the whole tumor. During this study, pancreatic cancer precursors were analyzed using multiple targeted and genome-based approaches to address their molecular heterogeneity and to find targets for possible biomarkers. The aim of this study was the investigation of the molecular landscape in PDAC preneoplastic lesions by an integrated multiomics analysis.

One focus of this project was the distinction of PanIN and gastric IPMN because it is still unknown whether PanIN and gastric IPMN represent a small/large version of the same lesion or distinct lesions. PanIN and gastric IPMN are predominantly localized in the peripheral duct system and are mainly distinguished by their size. The size range is specified in the classification guideline from the WHO, where a lesion below 5 mm is defined as a PanIN lesion and a lesion above 5 mm as an IPMN[1].

5.1 Characterization of mutation patterns of precursor lesions

IPMN and PanIN are well-known PDAC precursors and have been characterized concerning their morphology and their immunohistochemical profile. The first important result of our investigations is the identification of the involvement of different compartments with different transformation potentials, namely the main and the peripheral duct system. PanIN and gastric IPMN are predominantly localized in the peripheral duct system (branch duct) and are mainly distinguished based on their size and intestinal IPMN are mostly main-duct lesions. In fact, both low and high grade intestinal and gastric IPMN frequently show either two driver mutations (*KRAS* and/or *GNAS*) in 26/28 (92.9 %) of the cases. The PanIN lesions primarily (18/24, 75 %) have one driver mutation (*KRAS*).

The variant allele frequency is higher in gastric IPMN and intestinal IPMN in comparison to PanIN lesions. One fact is concerning the technical procedure. The manual microdissection is more precise in larger lesions. Therefore, the contamination with the surrounding area in PanIN tends to be higher as in IPMNs resulting in lower sample purity and thus resulting in lower tumor content. The estimated sample purity in all samples is at least > 80 % with a deep

sequencing rate > 4000 reads should be suitable for this purpose. False positive results can be neglected. Furthermore, the different VAF can be a hint for the different heterogenetic level. Tumor heterogeneity is referred to the hypothesis that not all malignant cells within a tumor are the same. It is believe that the progression of a tumor is a clonal process beginning with a single transformed cell[185].

Human cancers accumulate various mutations during their development and consist of highly heterogeneous cell populations. This phenomenon is called intratumoral heterogeneity (ITH). ITH is known to be involved in tumor growth, progression, invasion, and metastasis[186]. Kikutake *et al.* showed in a pan-cancer analysis that genetic characteristics, like a mutation, show a high correlation with certain shapes of VAF distributions. These shapes can be used as prognostic predictors or diagnostic markers[187]. So, the different *KRAS* allele frequency distribution in gastric IPMN can represent another prognostic prediction in comparison to PanIN and intestinal IPMN. It is known that gastric IPMN has a better prognosis than the intestinal IPMN subtype[188].

A further interesting observation was made in *KRAS*- and *GNAS*-mutated IPMN gastric low and high grade cases. The allele frequencies of both mutations showed a nearly perfect correlation (Figure 18; r= 0.98), supporting the theory that the two different mutations originated from the same tumor clone[189].

Fischer *et al.* 2019 showed that the stepwise progression to PDAC can be explained using a clonal evolution model. In this model, multiple independent clones arise with distinct mutations in early driver genes. The authors postulated that in early tumorigenesis, some selective pressures lead to the expansion of a dominant clone. In the late phases, eventually, a dominant clone invades and gives rise to PDAC, which is also described for cancer of the gastrointestinal tract[190,191,192]. In addition, the presented data is limited to one region of a lesion without a possible conclusion about mono or polyclonal origin. The correlation of *KRAS* and *GNAS* shows a typical monoclonal process. However, only one part of the lesion was sequenced. It was not checked for different clones in the same lesion at different locations. Possibly another clone will be dominant and give rise to PDAC during further progression. Statistically, the *GNAS* mutated clone does not seem to be the dominant clone and will not give rise to PDAC, because only about 4 % of the PDAC harbor a *GNAS* mutation according to the TCGA dataset. Therefore, another clone in a distinct region eventually will give rise to PDAC.

Besides the driver mutation, no further hotspot mutations are found in the early precursors. Some lesions have *TP53*, *ARID1A* and *PIK3CA* mutations. These results fit the fact that the PDAC is a tumor with a low mutational frequency[193,194].

Another interesting aspect is the lack of *RNF43* mutations in IPMNs. In the study by Tan *et al.* 2015, *RNF43* mutations are found with a mutation prevalence of about 10 % of the low grade IPMNs and about 20-75 % of the high grade lesions. All mutations are defined as inactive or LOH mutation (loss of heterozygosity)[195]. In our study, some mutations with uncertain significance were found without any hint of pathogenicity. Therefore, these mutations were not involved in further analysis. Sakamoto *et al*. showed that *RNF43* mutations might cause a reduced expression of ring finger protein 43 and play a crucial role and associate synergistically with *GNAS* mutation in the development of IPMN[196].

In PDAC, there are four main mutated genes present: *KRAS*, *TP53*, *CDKN2A,* and *SMAD4*[197]. Interestingly, according to the TCGA dataset, the genes *SMAD4* (21.4 %) and *CDKN2A* (14.8 %) were present with a high prevalence in PDAC. However, they were not present in the early precursor PanIN and IPMN in this study, which suggests that these mutations seem to be a part of a late event in the tumor progression[123]. Mutations in tumor suppressor *SMAD4* were present in about 30 % of the high grade PanIN lesions, but they were absent in the low grade lesions[124,125]. In recently published studies, the SMAD4 mutations were detected very rare in PanIN high grade lesions without an existing tumor[126]. The stepwise process via an accumulation of specific mutations to PDAC, is being challenged by modern sequencing methods. In mouse models, it was shown that the loss of *SMAD4* alone is not suitable to initiate tumor formation or changing the normal pancreatic phenotype[127]. In addition, the combination of *KRAS* and *SMAD4* mutation in mice leads to IPMN formation and progression to PDAC[198].

5.1.1 Impact of *GNAS*

During the analysis of gastric IPMN and PanIN, the influence of *GNAS* was very interesting. Both of these lesions differ at first glance only in size and in occurring *GNAS* mutation in gastric IPMN. It is important to investigate its possible function and whether this mutation could have an impact on the clinical relevance or the prognosis. G-proteins, in general, play a major role in tumor-induced angiogenesis, inflammation, immune tolerance, proliferation and survival, invasion and metastasis, which included four known hallmarks of cancer[199].

Gastric IPMN harbors *GNAS* mutations in about 40 % of the cases. In addition, intestinal IPMN with a possible higher transformation potential harbors *GNAS* in about 50-70 % of the cases[113]. The classical *GNAS* mutation R201 leads to a constitutively activated GTP-bound form. The active *GNAS* stimulates adenylyl cyclase, leading to the elevated synthesis of the second messenger, cyclic AMP (cAMP)[114,115]. It acts through multiple effectors including activating protein kinase A, and *EPAC1* and *2*, which are nucleotide exchange factors for the Rap subfamily of RAS-like small GTPases, as well as regulating the opening of cyclic nucleotide-gated ion channels[97]. These channels are known to play a part in resistance to colon cancer mediated by bacterial enterotoxins and are associated with a worse prognosis and clinical outcome[200,201]. The direct function of *GNAS* during the process of tumorigenesis is unknown. In mouse models, it was shown that cooperation of *GNAS* R201 mutation alone failed to show tumorigenesis, the co-mutated KRAS is essential to establish tumorigenesis. Both of these mutations lead to IPMN like lesions in mice[202,203].

During this study, one interesting observation was made regarding intestinal IPMN. It is known that *GNAS* mutation is particularly present in 50-70 % of the intestinal IPMN cases[113]. The mutation of R201 on exon 8 leads to an active form of *GNAS*. Maybe as a compensatory effect or feedback loop, the methylation is increased upstream of the transcription start site of intestinal IPMN in comparison to PanIN and gastric IPMN samples (Figure 25). These hypotheses would explain the lower expression level of GNAS in intestinal IPMN on the protein level compared to PanIN or gastric IPMN (Figure 24).

Innamorati *et al.* showed that the active form of *GNAS* acts like an oncogene in multifactorial transformation processes in low grade or benign neoplasia. This function is associated with papillary morphology and high mucin secretion. In their cohort, the expression on mRNA level of mucins (MUC1, MUC5AC and MUC6) was significantly higher in gastric IPMN in comparison to PanIN lesions. High *GNAS* activity may interfere with signaling in immature stages, but is not sufficient to progress to invasive carcinoma[204]. On the other hand, there is no correlation between *GNAS* status and IPMN histologic grade or clinical characteristics[205].

5.2 The potential to progress from a precursor lesion to PDAC

In 5 % of all patients, chromothripsis-like pattern was found resulting from a retrospective pan-cancer study from Zack *et al*. with 4934 patients. The two most affected chromosomes, regardless of the tumor entity, were chromosomes 9 and 12[134]. In the last years, much more chromothripsis-like events were detected as mentioned before, which is due to the fact that deep sequencing as a new technical approach is widely used. To find chromothripsis, oscillation of copy numbers, random joins, retention of heterozygosity and breakage–fusion–bridge have to be investigated[206]. The whole-genome low-coverage sequencing method is not suitable for the detection of chromothripsis events but can show the oscillation of copy numbers as a hint for chromothripsis-like pattern[137].

Notta *et al*. described chromothripsis in pancreatic cancer in 2016. In a retrospective analysis of 107 patients, the authors have found out that 2/3 of the PDACs harbor chromothripsis-like patterns. Another study from Hata *et al.* demonstrated that already 40 % of the high grade PanIN lesions harbor a chromothripsis-like event[136].

An aspect of this project was to investigate the distribution of chromothripsis-like events in early low and high grade precursor lesions from PDAC. To test the presence or impact of chromothripsis-like events in tumorigenesis, different molecular biological methods were applied.

Low-coverage whole-genome sequencing was performed to investigate the precursor lesions for chromothripsis-like events. A quite high frequency of recurrent CNVs in IPMN 10/19 (52 %) cases was detected, whereas most intestinal IPMN display recurrent alterations. Even in low grade lesions, gastric IPMN possibly represent a heterogeneous group consisting of clearly neoplastic and more indolent lesions.

Going into more detail, intestinal IPMN are mainly characterized by recurrent chromosomal gains, sometimes involving the whole chromosome or large fragments, in 4/5 (80 %) cases in regions containing relevant genes. Gastric IPMN instead shows recurrent CNVs in a lower percentage of the cases (6/14, 42 %) and in different localizations. PanIN possessed very few and non-recurrent CNVs. The variations were detected only in 2/11 (18 %) of the cases and involved small areas on chromosome 1 and 3 (gains) and chromosome 6, 9, 14, 15 and 19 (losses), thus displaying relevant overlap with gastric IPMN (Figure 19). The AFL samples showed a different profile with very small and not recurrent CNV regions over the whole-genome but compared to PanIN more frequently in 4/5 (80 %) cases. The most frequently

detected copy number variations in pancreatic cancer are amplifications (gain) in the region of chromosome 7p22, 8q12, 13q and 17q12 and deletions (loss) in the region of chromosome 1, 9p24 and 18q[207–209].

In addition, the copy number data was analyzed with the open-source software CTLPScanner, which is provided from Cai Lab, College of Life Sciences (Sichuan University, China)[163]. The program assesses the oscillation of copy numbers within a chromosome, which is typical for chromothripsis[132]. The CTLPScanner cannot prove the presence of chromothripsis but can give a hint for chromothripsis-like patterns. The low-coverage data was not suitable according to the big bin size of 2 Mbp and bigger. The methylome array data was used with an average bin size of 50-150 kbp.

The highest number of chromothripsis-like patterns was found in intestinal IPMN with 40 % of the cases (4/10). The affected chromosomes are 2, 3, 6 and 7. An amplification (gain) of arm q from chromosome 3 is known from a previous karyotyping study from Durante *et al.* 2016. The authors showed that the gain of the arm 3q can be used as a prognostic marker in IPMN. On parts of the chromosome several known oncogenes such as *PIK3CA*, *GATA2* (GATA binding protein 2) and *TERC* (telomerase RNA component) are present. These genes are involved in a pathway that regulates cell growth and promotes diseases progression[210]. In contrast to that, only 22.2 % of the gastric IPMN cases (4/18) showed a chromothripsis-like pattern on chromosome 1, 4, 6 and 9.

No publications are available, which are shown chromothripsis or chromothripsis-like pattern in IPMN. The present study was the first project, which investigated the chromothripsis-like pattern in IPMN and compared it to other precursor lesions as PanIN. The presence of this event in PanIN, mostly in high grade lesions are described by several studies[135,136]. In this study, the PanIN samples showed hints for chromothripsis-like events in 37.5 % of the PanIN lesions (3/8), in which chromosome 1 and 12 are affected. In general, chromothripsis-like events are more present in high grade lesions than in low grade lesions, and the intestinal IPMN show more copy number variations and chromothripsis-like events than the gastric IPMN.

These results of copy number variations and chromothripsis-like pattern analyses suggest a lower transformation potential of gastric IPMN and PanIN compared to intestinal IPMN, which could be due to a different intrinsic susceptibility or, most probably, to less frequent exposure

to carcinogens[138]. These results give only a hint for the present of chromothripsis-like events, to further analyze these findings, suitable genetic methods have to be done.

The localization of the peripheral ductal compartment versus the main duct could also support these results. PanIN and gastric IPMN are mostly located in branched ducts and intestinal IPMN is typically located in the main pancreatic duct.

Adachi *et al.* 2006 showed in an animal experiment with hamsters, that the bile reflux into the main pancreatic ducts is associated with tumor development. The bile is respnosible for the acceleration of epithelial cell kinetics of the main pancreatic duct[211]. In addition, important risk factors fot the PDAC formation are also the anatomic variations at the pancreatobiliary maljunctions[212].

Additionally, the main pancreatic duct(s) possibly represent the direct target of an altered oral, gastric, and intestinal microbiome, which has been shown to be associated with pancreatic cancer[213].

The higher expression of the proliferation marker Ki-67 in intestinal IPMN, compared to gastric IPMN and PanIN, already shows in low grade lesions a higher proliferation rate and also indicates a higher transformational potential for this lesion. In gastric IPMN and PanIN lesions, the transformation potential increases with the increasing grade of dysplasia, along with an increasing Ki-67 expression. In 2002, the study from Klein *et al.* showed an increasing Ki-67 proliferation index with increasing grades of PanIN[53].

In another study, colorectal tumors showed that the percentage of Ki-67 expression correlates with malignant potential and prognosis. An increased expression of Ki-67 is a poor prognostic marker[214]. The assumption that intestinal IPMN has a higher transformation potential than gastric IPMN and PanIN, fits our results.

In combination with the CNV, CTLPScanner analysis and Ki-67 staining data, the different transformation potential of intestinal IPMN in comparison to gastric IPMN and PanIN can be well explained.

The different transformation potential of PanIN and gastric IPMN cannot be elucidated based on the aforementioned analyses. However, if the results of the integrative analysis are taken into account, new aspects open up. The differential methylation and expression of the hypermethylated and downregualted genes in PanIN compared to gastric IPMN points to differences between the two lesions and support the concept of higher neoplastic potential of gastric IPMN. This can be further highlighted by analyzing the functions of the single

proteins including tumor suppressor genes like TFF1 and KLF4. In addition, gastric IPMN showed more copy number variations than PanIN. However, PanIN shows more hints for chromothripsis-like events.

The general argumentation is that the loss of function alterations in tumor suppressor genes are favorable for the progression to the tumor, which was postulated in 2000 from Hanahan and Weinberg in the cancer hallmarks. These mainly involved processes are generally accepted and are true for all cancer processes and supports our hypothesis that gastric IPMN has higher transformation potential[215].

PanIN high grade lesions are viewed as the main precursors of PDAC and are classified as clinically relevant in the absence of an established invasive carcinoma by the WHO, especially if present in the resection margin[1]. PanIN low grade lesions are classified not as clinically relevant in molecular pathology diagnostic.

In conclusion, it is still unknown how many PDAC derive from PanIN and if the process is the main process in the development. Our results of the neoplastic potential suggest a minor role of PanIN low grade in the carcinogenesis and a higher impact of IPMN with an increased transformational potential compared to PanIN lesions.

5.2.1 Cell of origin

The focus of this study was not to determine the possible cell of origin, but during the analysis of our data, some interesting hints for discussion were found.

The genome-wide DNA methylation data showed in the MSD Plot a clustering of the control groups of acinar tissue and ductal cells (Figure 33). The PanIN and gastric IPMN lesions are overlapping both clusters. It seems that the PanIN and gastric IPMN have the methylation signature from acinar to ductal cells, maybe through the process of metaplasia[216]. Furthermore, the resulting phylogenetic tree suggests the acinar cell compartment as the most probable origin of the analyzed gastric IPMN and some PanIN lesions. Here also are some PanIN lesions closer to the ductal compartment. Maybe the PanIN lesions closer to the ductal compartment are further developed through the process of metaplasia.

Another hypothesis is from Ferreira *et al*., which postulate a number of “bystander” PanIN lesions originated from invasive carcinoma. These bystander PanIN lesions are not involved in the carcinogenesis itself[100].

The study by Ferreira *et al*. reported that there exist PDACs, which can derive from acinar cells in a PanIN-derived process or ductal cells in a PanIN independent process. The expression of

AGR2 is used as a marker for the differentiation between the two cancer-derived types[100]. In our cohort, the expression of AGR2 was higher in gastric IPMN on mRNA level compared to PanIN and PDAC. The ARG2 positive PDACs is according to the study a duct derived tumor. These "bystander" PanIN possess different molecular pattern as the normal PanIN. However, the morphology of both PanIN types is the same. Maybe the PanIN lesions which are closer to the ductal compartment are "bystander" PanIN lesions from the invasive carcinoma. This hypothesis has to be investigated in further studies.

In contrast to the gastric IPMN, it seems that the intestinal IPMN have their cell of origin in the ductal compartment. A variety of papers postulate the ductal compartment as a cell of origin for IPMN. However, these studies do not distinguish the subtypes of IPMN, whereas our results postulate, as the first study, a ductal origin for intestinal IPMN and the acinar cell compartment for gastric IPMN may originate through a process of ADM. Different genetic mouse models showed that both the ductal compartment and the acini can give rise to PDAC through metaplasia (ADM)[100,217–221]. However, the cell of origin for humans is still controversially discussed[216,219,222].

Espinet *et al*. showed that the pancreatic cancer epigenomes maintain cell-of-origin signatures. The study investigated PDAC and pure cell population of acinar, dedifferentiated acinar, and ductal cells in the DNA methylation profile. The authors found two distinct clusters of PDACs. One correlates with the ductal cell and has a worse prognosis and the other cluster correlates with the acinar and dedifferentiated acinar cells[223]. These findings suggest different cells of origins for the PDAC tumor. This finding fits our results that the duct derived intestinal IPMN have a higher transformational potential and could explain also the worse prognosis compared to gastric IPMN. The PanIN and gastric IPMN derived from acinar cells and dedifferentiated acinar cells with a better prognosis.

5.2.2 Gastric IPMN versus intestinal IPMN

Besides the focus on gastric IPMN and PanIN lesions, several results of this study demonstrate, the different biology of gastric IPMN and intestinal IPMN. The two different subtypes of IPMN were distinguishable according to the immunohistochemical profile with the marker MUC1, MUC2, MUC5AC and CDX2.

In many studies about precursors, IPMNs were analyzed as one group of precursors. One reason was the similar mutation profile. *KRAS* and *GNAS* were the most common mutations, which was also demonstrated in this study. On the other hand, the DNA methylation profile is

completely different between the two lesions. This was the first study with a genome-wide DNA methylation profile of the precursor lesions of PDAC. The intestinal IPMN showed more CNVs and a higher proliferation rate, which both fits the hypothesis that intestinal IPMN has a higher transformational potential than gastric IPMN. In addition, the finding of the DNA methylation data of this study suggests a different cell of origin of these two lesions according to the different relations to the acinar and ductal control groups.

Omori *et al.* postulated that the intestinal IPMN emerges from gastric IPMN driven by *CDX2* (Caudal Type Homeobox 2). The postulated process is through that the mutant *GNAS* induced overexpression of the CDX2 protein with some other molecules, maybe activation of *PKA* (Protein Kinase CAMP-Activated Catalytic Subunit Alpha) and *ERK1/2* (Mitogen-Activated Protein Kinase 3) and certain unknown microenvironment factors[224]. The mechanism, how CDX2 is induced by *GNAS* mutations is still unknown. Innamorati *et al*. showed that the gain-of-function *GNAS* mutation interfere with signals in the microenvironment of permissive tissues and lead to different pathological molecular phenotypes[204].

In this study, we could not find any indication in the data that supports this hypothesis. Quite the opposite, our results show major differences in DNA methylation and at the genomic level.

The last interesting aspect is the different clinical outcome. Gastric-type IPMN has an overall favorable prognosis, exhibiting a very low risk of developing into carcinomas and with longer survival compared to other subtypes [72,225,226]. The survival rate between gastric and intestinal IPMN is significantly different. Gastric IPMN show significantly better survival than the intestinal IPMN[188]. Our data strongly suggests that studies concerning precursor lesions should differentiate the different entities and should not consider them as one group.

5.3 Investigation of new biomarkers

The integrative analysis was used to discover a potential biomarker, which can be used to distinguish the precursor lesions PanIN and gastric IPMN. Several studies with different cancer entities used integrative analysis of DNA methylation and gene expression to identify diagnostic markers[227]. Since CpG islands are mostly found within promoter regions, their methylation statuses often determine whether a particular gene is transcribed or silenced[228]. A restriction of the present analysis is that the used CpGs were not exclusively in the promoter region. However, by combining methylome and transcriptome data, a few relevant metabolic pathways were identified, which were significantly hypermethylated and downregulated in PanIN compared to gastric IPMN.

The correlation analysis revealed that methylation was negatively correlated with the local regulation of gene expression. This observation is consistent with previous studies showing that hypermethylation is associated with transcriptional silencing of genes[229–231]. Interestingly, a negative correlation between methylation and gene expression was observed in hypermethylated genes and a positive correlation was found in hypomethylated genes[227].

In this study, the hypermethylated genes with a negative correlation between methylation and distant gene expression were used. The three *hub* genes *KLF4*, *MUC6* and *TFF1* were identified, which all belong to a group of extrapancreatic foregut markers and have been previously described to be overexpressed in PanIN compared to normal pancreatic ducts[232]. However, these markers were highly upregulated in gastric IPMN compared to PanIN lesions. The differential methylation and expression of these markers in PanIN and gastric IPMN points out the differences between the two lesions. Furthermore, these results support the concept of a higher neoplastic potential of gastric IPMN, which can be further highlighted by analyzing the functions of the single proteins.

5.3.1 Krüppel-like factor 4

The *hub* gene *KLF4* encodes a zinc-finger containing a transcription factor, which plays an important role in the homeostasis of tissues and organs and exerts tissue-dependent tumor-promoting or tumor-suppressive functions[233]. In gastrointestinal cancer, *KLF4* regulates negatively cell proliferation, promotes tissue differentiation and low expression is associated with poor survival[234]. Furthermore, in mice, the gene *KLF4* plays an essential role in the maintenance of genomic stability[235]. In the pancreas, it has been shown to be required for the

process of acinar-ductal metaplasia and the initial phases of *KRAS* associated carcinogenesis in mice[94].

Another aspect, where KLF4 is involved in the tumorigenesis, is the activation of the TGF-β signaling pathway. It has been shown that the degradation of KLF4 is related to the activation of TGF-β, a well-known strong inducer of EMT. This is a relevant mechanism of inhibition of the TGF-β pathway in PDAC and its precursor lesions and is related to *KLF4* overexpression[236–239].

Our data showed significant overexpression of *KLF4* on mRNA level in gastric IPMN compared to PanIN lesions, acinar tissue and PDACs. Moreover, they showed an overexpression on the protein level in PanIN, gastric and intestinal IPMN as well as in PDAC samples compared to normal pancreatic cells. This supports an oncogenic role of *KLF4* at least in the early phase of the carcinogenesis process.

In contrast to our findings, the study from Zammarchi *et al.* showed that about 70 % of the PDAC did not have *KLF4* expression on mRNA and protein level. In addition, about 65 % of the PDAC samples lost an 8-Mbp region on chromosome 9, including the gene locus of *KLF4*[240]. In our data, the loss of a region on chromosome 9 could be detected in 50 % of the PDAC cases (2/4). The precursor lesion, gastric IPMN and PanIN did not show any alteration in this region in the copy number profile and the intestinal IPMN showed a gain in 60 % of the cases (3/5).

Recently, *KLF4* has been identified as a genetic driver of IPMN, where four different hotspot mutations have been found in > 50 % of the cases, mostly associated with low grade dysplasia[241]. Altogether, the data points out an oncogenic effect of *KLF4* in the early pancreatic tumorigenesis, including the acinar to ductal metaplasia and the progression to invasive carcinoma. However, KLF4 is not suitable for a marker to distinguish the precursor lesions of PDAC.

5.3.2 Mucin 6 (*MUC6*) and Trefoil factor 1 (*TFF1*)

The genes *Muc6* and *TFF1* both are related to the mucin O-glycan biosynthesis pathway. Mucin-type O-glycosylation is an evolutionarily conserved protein modification present on membrane-bound and secreted proteins. The mucin O-glycan biosynthesis pathway has been shown to affect relevant processes of cancer progression and metastasis and is involved in pancreatic cancer as well[184,242]. The N-acetylgalactosaminyltransferases *GALNT5* and *GALNT7*, which play a role in the initiation of O-glycosylation by adding N- acetylgalactosamine on the serine/threonine backbone of proteins, were significantly overexpressed at the mRNA level in

gastric IPMN compared to PanIN, possibly suggesting a stronger neoplastic potential of gastric IPMN compared to PanIN.

The TCGA dataset of 185 PDAC patients showed that *GALNT7* can be used as a prognostic marker, the overall survival is significant increased by low mRNA expression. Changes in O-glycosylation, as well as inactive mutations in the genes controlling O-glycosylation, are also associated with tumor formation and progression[243]. Glycan structures play a role in cancer progression by affecting tumor cell invasiveness, the ability to disseminate through the blood circulation and to metastasize in distant organs[182,244].

The effects of glycosylation on various markers have been shown to influence proliferation, migration, invasion and metastasis[184]. The overexpression of the process of glycosylation ends up in truncated O-GalNAc glycans, which ends up in glycogen expression, mislocalization of glycosyltransferase, abnormal Golgi pH, or altered tumor microenvironment[182].

The expression of *MUC6* in PDAC and PanIN was significantly downregulated compared to gastric IPMN on mRNA level, which indicated that the overexpression of the pathway could be important during metaplasia, but maybe not in the carcinoma itself. On the protein level, the expression of MUC6 is strongly correlated with gastric differentiation, being significantly upregulated in gastric IPMN compared to intestinal IPMN and its associated colloid-type cancer and the classical PDAC. These findings were also observed by Ohya *et al.* The colleagues showed that MUC6 expression decreased with an increasing grade of dysplasia[245]. This postulated a higher transformational potential of intestinal IPMN with lower MUC6 expression comparable with PDAC and colloid carcinoma.

Furthermore, other mucins are influenced by glycosylation. An overexpression of MUC1 is positively correlated with increased metastasis and poor patient survival in PDAC[246]. The expression of MUC1 is not detected on the protein level in PanIN, gastric IPMN or intestinal IPMN. The expression of MUC1 is classically detected in the pancreatobiliary IPMN subtype[1]. On the other hand, the mRNA expression of *MUC1* is significantly higher in gastric IPMN compared to PanIN.

In general, mucin overexpression is related to pancreatic cancer[247–249]. MUC5AC for example is highly overexpressed in tumoral tissues compared to tumor surrounding tissue[250]. The downregulation of *MUC5AC* could also result in decreased growth and metastasis, while upregulation of *MUC5AC* accelerated high grade PanIN to invasive cancer[251]. In our data, the mRNA expression of *MUC5AC* is highly overexpressed in gastric IPMN compared to PanIN

lesions. These data support the higher neoplastic potential for gastric IPMN compared to PanIN lesions. In colon cancer and biliary tract cancer, the expression of *MUC5AC* and *TFF1* was colocalized and these genes could be potential diagnostic markers[252,253].

TFF1, another *hub* gene in this study, is also indirectly related to the O-glycan biosynthesis pathway. The family of TFFs protect the mucosal epithelium by increasing the viscoelasticity of mucus and enhancing epithelial restitution rates. *TFF* overexpression is prevalent in adenocarcinomas and a hallmark of chronic inflammatory diseases of the respiratory tract[254]. Although some functional data ascribe *TFF1* a tumor-suppression function in pancreatic cancer, possibly due to its role in the prevention of epithelial-mesenchymal transition. In our study, four different CpGs in the promoter region were hypermethylated and one additional in the first exon in PanIN lesions in comparison to gastric IPMN. This data showed a positive correlation with the expression, which is significantly downregulated in PanIN. In PanIN lesions, the decreased expression of *TFF1* showed a higher neoplastic potential compared to gastric IPMN. The phenomenon of *TFF1* expression loss is frequently described in gastric carcinoma, probably through the mechanism of DNA methylation, and is therefore considered as a tumor suppressor gene[255]. The initiation phase of antral gastric cancer is associated with progressive epigenetic silencing of *TFF1*, which can be suppressed by the hormone gastrin[256]. In pancreatic cancer, the tumor suppressor function of *TFF1* is described by inhibiting EMT and the invasive transformation of PanIN. When the expression of *TFF1* is suppressed in vitro, pancreatic cancer cell lines showed enhanced invasive ability and features of epithelial-mesenchymal transition. The expression was specifically observed in PanIN and was lost in PDAC[257,258]. In mice model, the loss of *TFF1* resulted in the expansion of PanIN lesions, an EMT phenotype in PanIN cells and an accumulation of cancer-associated fibroblasts, resulting in the development of invasive adenocarcinoma[257]. In this study, the expression of *TFF1* was also significantly reduced in PDAC compared to PanIN and gastric IPMN, which showed the tumor suppressor function also in this cohort.

Furthermore, *TFF1* has been shown to stimulate the motility of pancreatic cancer cells and pancreatic stellate cells suggesting a role in invasion and metastasis[259]. *TFF1* and the paralog *TFF2* have been recently shown to be elevated in the serum of patients with early-stage PDAC, suggesting their possible role as early tumor biomarkers[257,260].

5.4 Conclusion

In conclusion, we were able to improve the characterization of the precursor lesions of pancreatic cancer. PDAC precursors are heterogeneous in terms of morphology, genetics and epigenetics. PanIN and gastric IPMN have similar molecular profiles. However, the lesions show intra-lesional heterogeneity, which determines their biological behavior. The data showed that the cell of origin is probably the same for gastric IPMN and PanIN, but these two lesions show different transformational potentials regarding the copy number profile and the activation of different pathways on the transcriptome level. PanIN and IPMN are different entities, with partly similar behavior but also with a variety of differences in their molecular patterns and their clinical outcome. Despite their similarities, the distinction between PanIN and gastric IPMN is relevant due to their different levels of transformational potential which should be pursued with available (e.g. *GNAS* mutations) and, possibly, newly established markers.

The most interesting finding of this study was the completely different behavior of intestinal IPMN compared to gastric IPMN and PanIN lesions. Intestinal IPMN already shows a distinct methylome profile and higher neoplastic potential in low grade lesions. In this study, we postulated a different cell of origin for intestinal and gastric IPMN. Immunophenotypical subtyping is fundamental in identifying PDAC precursors with a different risk of progression and should be included as a core element in pathology reports.

The study has some limitation, the possible analysis of archived paraffin material is mainly dependent on the quality of DNA and RNA. The sequencing data have to be evaluated carefully to detect artefacts according to the fixed samples. The comparison to other studies with fresh samples is difficult and need high bioinformatic effort. These facts restricted the number of analyzed samples and used methods. However, our data showed robust sequencing results with targeted and genome-wide sequencing from FFPE derived material[261]. This study is the first molecular characterization of precursor lesions of the PDAC, which showed the molecular heterogeneity of the different precursor lesions.

Therefore, we strongly believe that studies addressing precursor lesions should differentiate between the different entities and should not consider them as one group. Further studies are needed to identify reliable, possibly protein-based markers, which can help us to spot among morphologically similar lesions those with stronger neoplastic potential to better direct clinical management. This project will be continued by increasing the number of transcriptome

analysis to find markers which can be validated and stronger supported by the data. Moreover, intestinal IPMN will be included in the RNASeq analysis to look for new possible biomarkers.

6 References

1. Gill, A., Klimstra, D., Lam, A. & Washington, M. Tumours of the pancreas. in *WHO classification of Tumours of the Digestive System* (eds. Bosman, F., Hruban, R. & Carneiro, F.) **5**, 295–372 (International Agency for Research on Cancer, 2019).

2. Haeberle, L. & Esposito, I. Pathology of pancreatic cancer. *Transl. Gastroenterol. Hepatol.* **4**, 1–12 (2019).

3. Ryan, D. P., Hong, T. S. & Bardeesy, N. Pancreatic adenocarcinoma. *N. Engl. J. Med.* **371**, 1039–1049 (2014).

4. Meng, Z. *et al.* Tumor location as an indicator of survival in T1 resectable pancreatic ductal adenocarcinoma: A propensity score-matched analysis. *BMC Gastroenterol.* **19**, 1–9 (2019).

5. Schlitter, A. M. *et al.* Molecular, morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes. *Sci. Rep.* **7**, 1–12 (2017).

6. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).

7. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019 (US statistics). *CA. Cancer J. Clin.* **69**, 7–34 (2019).

8. Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J. Oncol.* **10**, 10–27 (2019).

9. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).

10. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.

11. Quaresma, M., Coleman, M. P. & Rachet, B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: A population-based study. *Lancet* **385**, 1206–1218 (2015).

12. Elbanna, K. Y., Jang, H. J. & Kim, T. K. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review. *Insights Imaging* **11**, (2020).

13. Kunovsky, L. *et al.* The Use of Biomarkers in Early Diagnostics of Pancreatic Cancer. *Can. J. Gastroenterol. Hepatol.* **2018**, 1–10 (2018).

14. Malka, D. *et al.* Risk of pancreatic adenocarcinoma in chronic pancreatitis. *Gut* **51**, 849–852 (2002).

15. Matsubayashi, H., Ishiwatari, H., Sasaki, K., Uesaka, K. & Ono, H. Detecting Early

Pancreatic Cancer: Current Problems and Future Prospects. *Gut Liver* **14**, 30–36 (2020).

16. Bosetti, C. *et al.* Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium. *Ann. Oncol.* **25**, 2065–2072 (2014).

17. Klein, A. P. *et al.* Prospective Risk of Pancreatic Cancer in Familial Pancreatic Cancer Kindreds. *Cancer Res.* **64**, 2634–2638 (2004).

18. Benzel, J. & Fendrich, V. Familial pancreatic cancer. *Oncol. Res. Treat.* **41**, 611–618 (2018).

19. Zhen, D. B. *et al.* BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: A PACGENE study. *Genet. Med.* **17**, 569–577 (2015).

20. Solomon, S., Das, S., Brand, R. & Whitcomb, D. C. Inherited pancreatic cancer syndromes. *Cancer J. (United States)* **18**, 485–491 (2012).

21. Bosetti, C. *et al.* Cigarette smoking and pancreatic cancer: An analysis from the International Pancreatic Cancer Case-Control Consortium (PANC4). *Ann. Oncol.* **23**, 1880–1888 (2012).

22. Wang, Y. T., Gou, Y. W., Jin, W. W., Xiao, M. & Fang, H. Y. Association between alcohol intake and the risk of pancreatic cancer: A dose-response meta-analysis of cohort studies. *BMC Cancer* **16**, 1–11 (2016).

23. Larsson, S. C. & Wolk, A. Red and processed meat consumption and risk of pancreatic cancer: Meta-analysis of prospective studies. *Br. J. Cancer* **106**, 603–607 (2012).

24. Aune, D. *et al.* Body mass index, abdominal fatness and pancreatic cancer risk: A systematic review and non-linear dose-response meta-analysis of prospective studies. *Ann. Oncol.* **23**, 843–852 (2012).

25. American Cancer Society. *Cancer Facts & Figures 2020*. *American Cancer Society* (2020).

26. Yachida, S. *et al.* Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clin. Cancer Res.* **18**, 6339–6347 (2012).

27. Lacobuzio-Donahue, C. A. *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *J. Clin. Oncol.* **27**, 1806–1813 (2009).

28. Ducreux, M. *et al.* Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26**, v56–v68 (2015).

29. P, H. *et al.* PARP inhibition in treatment of pancreatic cancer. *Expert Rev. Anticancer Ther.* **20**, 939–945 (2020).

30. Zeng, S. *et al.* Chemoresistance in pancreatic cancer. *Int. J. Mol. Sci.* **20**, 1–19 (2019).

31. Oberstein, P. E. & Olive, K. P. Pancreatic cancer: Why is it so hard to treat? *Therap. Adv. Gastroenterol.* **6**, 321–337 (2013).

32. Ho, W. J., Jaffee, E. M. & Zheng, L. The tumour microenvironment in pancreatic cancer — clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* (2020). doi:10.1038/s41571-020-0363-5

33. Azizian, A. *et al.* CA19-9 for detecting recurrence of pancreatic cancer. *Sci. Rep.* **10**, 1–10 (2020).

34. Goggins, M. Molecular markers of early pancreatic cancer. *J. Clin. Oncol.* **23**, 4524–4531 (2005).

35. Sogawa, K. *et al.* Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein α-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags. *Br. J. Cancer* **115**, 949–956 (2016).

36. Joergensen, M. T., Heegaard, N. H. H. & Schaffalitzky de Muckadell, O. B. Comparison of Plasma Tu-M2-PK and CA19-9 in Pancreatic Cancer. *Pancreas* **39**, 243–247 (2010).

37. Baine, M. J. *et al.* Differential gene expression analysis of peripheral blood mononuclear cells reveals novel test for early detection of pancreatic cancer. *Cancer Biomarkers* **11**, 1–14 (2012).

38. Schultz, N. A. *et al.* MicroRNA Biomarkers in Whole Blood for Detection of Pancreatic Cancer. *JAMA* **311**, 392 (2014).

39. Singhi, A. D., Koay, E. J., Chari, S. T. & Maitra, A. Early Detection of Pancreatic Cancer: Opportunities and Challenges. *Gastroenterology* **156**, 2024–2040 (2019).

40. Rhim, A. D. *et al.* Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. *Gastroenterology* **146**, 647–651 (2014).

41. Kulemann, B. *et al.* Circulating tumor cells found in patients with localized and advanced pancreatic cancer. *Pancreas* **44**, 547–550 (2015).

42. Iacobuzio-Donahue, C. A. Genetic evolution of pancreatic cancer: lessons learnt from the pancreatic cancer genome sequencing project. *Gut* **61**, 1085 (2012).

43. MLB, P. *et al.* Progression to pancreatic ductal adenocarcinoma from pancreatic intraepithelial neoplasia: Results of a simulation model. *Pancreatology* **18**, 928–934 (2018).

44. Yu, J., Blackford, A. L., Dal Molin, M., Wolfgang, C. L. & Goggins, M. Time to progression of pancreatic ductal adenocarcinoma from low-to-high tumour stages. *Gut* **64**, 1783–1789 (2015).

45. Basturk, O. *et al.* A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas. *Am. J. Surg. Pathol.* **39**, 1730–1741 (2015).

46. Pittman, M. E., Rao, R. & Hruban, R. H. Classification, Morphology, Molecular Pathogenesis, and Outcome of Premalignant Lesions of the Pancreas. *Arch. Pathol. Lab. Med.* **141**, 1606–1614 (2017).

47. Esposito, I., Segler, A., Steiger, K. & Klöppel, G. Pathology, genetics and precursors of

human and experimental pancreatic neoplasms: An update. *Pancreatology* **15**, 598–610 (2015).

48. De Wilde, R. F., Hruban, R. H., Maitra, A. & Offerhaus, G. J. A. Reporting precursors to invasive pancreatic cancer: Pancreatic intraepithelial neoplasia, intraductal neoplasms and mucinous cystic neoplasm. *Diagnostic Histopathol.* **18**, 17–30 (2012).

49. Konstantinidis, I. T. *et al.* Incidentally discovered pancreatic intraepithelial neoplasia: What is its clinical significance? *Ann. Surg. Oncol.* **20**, 3643–3647 (2013).

50. Andea, A., Sarkar, F. & Adsay, V. N. Clinicopathological Correlates of Pancreatic Intraepithelial Neoplasia: A Comparative Analysis of 82 Cases With and 152 Cases Without Pancreatic Ductal Adenocarcinoma. *Mod. Pathol.* **16**, 996–1006 (2003).

51. Cubilla, A. L. & Fitzgerald, P. J. Morphological Lesions Associated with Human Primary Invasive Nonendocrine Pancreas Cancer. *Cancer Res.* **36**, (1976).

52. Yonezawa, S., Higashi, M., Yamada, N. & Goto, M. Precursor Lesions of Pancreatic Cancer. *Gut Liver* **2**, 137–154 (2008).

53. Klein, W. M., Hruban, R. H., Klein-Szanto, A. J. P. & Wilentz, R. E. Direct Correlation between Proliferative Activity and Dysplasia in Pancreatic Intraepithelial Neoplasia (PanIN): Additional Evidence for a Recently Proposed Model of Progression. *Mod. Pathol.* **15**, 441–447 (2002).

54. Zińczuk, J. *et al.* P16, P21, and P53 Proteins Play an Important Role in Development of Pancreatic Intraepithelial Neoplastic. *Ir. J. Med. Sci.* **187**, 629–637 (2018).

55. W, H. *et al.* Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4. *J. Pathol.* **242**, 16–23 (2017).

56. Aichler, M. *et al.* Origin of pancreatic ductal adenocarcinoma from atypical flat lesions: a. *J Pathol* **226**, 723–734 (2012).

57. von Figura, G. *et al.* Atypical flat lesions derive from pancreatic acinar cells. *Pancreatology* **17**, 350–353 (2017).

58. Esposito, I., Konukiewitz, B., Schlitter, A. M. & Klöppel, G. Neue Einblicke in die Entstehung des Pankreaskarzinoms. *Pathologe* **33**, 189–193 (2012).

59. Berger, A. W., Seufferlein, T. & Kleger, A. Zystische Pankreastumoren: Diagnostik und neue Biomarker. *Chirurg* **88**, 905–912 (2017).

60. Schnelldorfer, T. Experience With 208 Resections for Intraductal Papillary Mucinous Neoplasm of the Pancreas. *Arch. Surg.* **143**, 639 (2008).

61. Nara, S., Shimada, K., Kosuge, T., Kanai, Y. & Hiraoka, N. Minimally Invasive Intraductal Papillary-mucinous Carcinoma of the Pancreas: Clinicopathologic Study of 104 Intraductal Papillary-mucinous Neoplasms. *Am. J. Surg. Pathol.* **32**, 243–255 (2008).

62. Nakagohri, T., Asano, T., Kenmochi, T., Urashima, T. & Ochiai, T. Long-term Surgical Outcome of Noninvasive and Minimally Invasive Intraductal Papillary Mucinous

Adenocarcinoma of the Pancreas. *World J. Surg.* **26**, 1166–1169 (2002).

63. Paini, M. *et al.* Molecular pathology of intraductal papillary mucinous neoplasms of the pancreas. *World J. Gastroenterol.* **20**, 10008–10023 (2014).

64. Del Chiaro, M. *et al.* European evidence-based guidelines on pancreatic cystic neoplasms. *Gut* **67**, 789–804 (2018).

65. Tanaka, M. *et al.* International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. *Pancreatology* **12**, 183–197 (2012).

66. Crippa, S., Arcidiacono, P. G., De Cobelli, F. & Falconi, M. Review of the diagnosis and management of intraductal papillary mucinous neoplasms. *United Eur. Gastroenterol. J.* **8**, 249–255 (2020).

67. Klausen, P. *et al.* Subtyping of intraductal papillary mucinous neoplasms – pitfalls of <scp>MUC</scp> 1 immunohistochemistry. *APMIS* **127**, 27–32 (2019).

68. Tanaka, M. *et al.* International Consensus Guidelines for Management of Intraductal Papillary Mucinous Neoplasms and Mucinous Cystic Neoplasms of the Pancreas. *Pancreatology* **6**, 17–32 (2006).

69. Furukawa, T. *et al.* Classification of types of intraductal papillary-mucinous neoplasm of the pancreas: a consensus study. *Virchows Arch.* **447**, 794–799 (2005).

70. Ban, S. *et al.* Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas: Its Histopathologic Difference Between 2 Major Types. *Am. J. Surg. Pathol.* **30**, 1561–1569 (2006).

71. Shi, C. & Hruban, R. H. Intraductal papillary mucinous neoplasm. *Hum. Pathol.* **43**, 1–16 (2012).

72. Furukawa, T. *et al.* Prognostic relevance of morphological types of intraductal papillary mucinous neoplasms of the pancreas. *Gut* **60**, 509–516 (2011).

73. Adsay, N. V., Adair, C. F., Heffess, C. S. & Klimstra, D. S. Intraductal Oncocytic Papillary Neoplasms of the Pancreas. *Am. J. Surg. Pathol.* **20**, 980–994 (1996).

74. Wang, T. *et al.* Intraductal Oncocytic Papillary Neoplasms. *Am. J. Surg. Pathol.* **43**, 656–661 (2019).

75. D'Onofrio, M. *et al.* Oncocytic Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Pancreas* **45**, 1233–1242 (2016).

76. Basturk, O. *et al.* The oncocytic subtype is genetically distinct from other pancreatic intraductal papillary mucinous neoplasm subtypes. *Mod. Pathol.* **29**, 1058–1069 (2016).

77. Basturk, O. *et al.* Distinct pathways of pathogenesis of intraductal oncocytic papillary neoplasms and intraductal papillary mucinous neoplasms of the pancreas. *Virchows Arch.* **469**, 523–532 (2016).

78. REDDY, R. *et al.* Pancreatic mucinous cystic neoplasm defined by ovarian stroma: Demographics, clinical features, and prevalence of cancer. *Clin. Gastroenterol. Hepatol.*

2, 1026–1031 (2004).

79. Thompson, L. D. R., Becker, R. C., Przygodzki, R. M., Adair, C. F. & Heffess, C. S. Mucinous cystic neoplasm (mucinous cystadenocarcinoma of low-grade malignant potential) of the pancreas: A clinicopathologic study of 130 cases. *Am. J. Surg. Pathol.* **23**, 1–16 (1999).

80. Zamboni, G. *et al.* Mucinous cystic tumors of the pancreas: Clinicopathological features, prognosis, and relationship to other mucinous cystic tumors. *Am. J. Surg. Pathol.* **23**, 410–422 (1999).

81. Hijioka, S. *et al.* Morphological differentiation and follow-up of pancreatic cystic neoplasms using endoscopic ultrasound. *Endoscopic Ultrasound* **4**, 312–318 (2015).

82. Yamao, K. *et al.* Clinicopathological features and prognosis of mucinous cystic neoplasm with ovarian-type stroma: A multi-institutional study of the Japan pancreas society. *Pancreas* **40**, 67–71 (2011).

83. Erdogan, D. *et al.* Cystadenomas with Ovarian Stroma in Liver and Pancreas: An Evolving Concept. *Dig. Surg.* **23**, 186–191 (2006).

84. Erdogan, D. *et al.* Mucinous Cystadenomas in Liver: Management and Origin. *Dig. Surg.* **27**, 19–23 (2010).

85. Sethi, V., Giri, B., Saluja, A. & Dudeja, V. Insights into the Pathogenesis of Pancreatic Cystic Neoplasms. *Dig. Dis. Sci.* **62**, 1778–1786 (2017).

86. Vogelstein, B. *et al.* Genetic Alterations during Colorectal-Tumor Development. *N. Engl. J. Med.* **319**, 525–532 (1988).

87. Toyota, M., Itoh, F. & Imai, K. DNA methylation and gastrointestinal malignancies: Functional consequences and clinical implications. *Journal of Gastroenterology* **35**, 727–734 (2000).

88. Hruban, R. H., Goggins, M., Parsons, J. & Kern, S. E. Progression model for pancreatic cancer. *Clin. Cancer Res.* **6**, 2969–2972 (2000).

89. Aguirre, A. J. Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev.* **17**, 3112–3126 (2003).

90. Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).

91. Yamaguchi, J., Yokoyama, Y., Kokuryo, T., Ebata, T. & Nagino, M. Cells of origin of pancreatic neoplasms. *Surg. Today* **48**, 9–17 (2018).

92. Morris, J. P., Wang, S. C. & Hebrok, M. KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. *Nat Rev Cancer* **10**, 683–695 (2010).

93. Murphy, S. J. *et al.* Genetic Alterations Associated With Progression From Pancreatic Intraepithelial Neoplasia to Invasive Pancreatic Tumor. *Gastroenterology* **145**, 1098-

1109.e1 (2013).

94. Wei, D. *et al.* KLF4 Is Essential for Induction of Cellular Identity Change and Acinar-to-Ductal Reprogramming during Early Pancreatic Carcinogenesis. *Cancer Cell* **29**, 324–338 (2016).

95. Chuvin, N. *et al.* Acinar-to-Ductal Metaplasia Induced by Transforming Growth Factor Beta Facilitates KRASG12D-driven Pancreatic Tumorigenesis. *Cmgh* **4**, 263–282 (2017).

96. Qiu, W. *et al.* Loss of Activin Receptor Type 1B Accelerates Development of Intraductal Papillary Mucinous Neoplasms in Mice with Activated KRAS. *Gastroenterology* **150**, 218-228.e12 (2016).

97. Patra, K. C., Bardeesy, N. & Mizukami, Y. Diversity of Precursor Lesions For Pancreatic Cancer: The Genetics and Biology of Intraductal Papillary Mucinous Neoplasm. *Clin. Transl. Gastroenterol.* **8**, e86 (2017).

98. Reichert, M., Blume, K., Kleger, A., Hartmann, D. & Von Figura, G. Developmental Pathways Direct Pancreatic Cancer Initiation from Its Cellular Origin. *Stem Cells Int.* **2016**, 1–8 (2016).

99. Tsuda, M., Fukuda, A., Takaori, K. & Seno, H. Genetics and biology of pancreatic cancer and its precursor lesions: lessons learned from human pathology and mouse models. *Ann. Pancreat. Cancer* **2**, 15–15 (2019).

100. Ferreira, R. M. M. *et al.* Duct- and Acinar-Derived Pancreatic Ductal Adenocarcinomas Show Distinct Tumor Progression and Marker Expression. *Cell Rep.* **21**, 966–978 (2017).

101. Liu, P., Wang, Y. & Li, X. Targeting the untargetable KRAS in cancer therapy. *Acta Pharm. Sin. B* **9**, 871–879 (2019).

102. Stephen, A. G., Esposito, D., Bagni, R. G. & McCormick, F. Dragging ras back in the ring. *Cancer Cell* **25**, 272–281 (2014).

103. Simanshu, D. K., Nissley, D. V & Mccormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17–33 (2017).

104. Diane E. Handy Rita Castro Joseph Loscalzo. RAS oncogenes: weaving a tumorigenic web Yuliya. *Nat Rev Cancer* **11**, 761–774 (2011).

105. Bryant, K. L., Mancias, J. D., Kimmelman, A. C. & Der, C. J. KRAS: feeding pancreatic cancer proliferation. *Trends Biochem. Sci.* **39**, 91–100 (2014).

106. Scheffzek, K. The Ras-RasGAP Complex: Structural Basis for GTPase Activation and Its Loss in Oncogenic Ras Mutants. *Science (80-.).* **277**, 333–338 (1997).

107. Vatansever, S., Erman, B. & Gümüş, Z. H. Oncogenic G12D mutation alters local conformations and dynamics of K-Ras. *Sci. Rep.* **9**, 1–13 (2019).

108. Adjei, A. A. Blocking oncogenic Ras signaling for cancer therapy. *J. Natl. Cancer Inst.* **93**, 1062–1074 (2001).

109. Schutte, M. *et al.* Abrogation of the Rb/p16 Tumor-suppressive Pathway in Virtually All

Pancreatic Carcinomas. *Cancer Res.* **57**, (1997).

110. J, L. *et al.* Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375**, 503–506 (1995).

111. S, Y., M, H., N, Y. & M, G. Precursor lesions of pancreatic cancer. *Gut Liver* **2**, 137–154 (2008).

112. Ideno, N. *et al.* Clinical significance of GNAS mutation in intraductal papillary mucinous neoplasm of the pancreas with concomitant pancreatic ductal adenocarcinoma. *Pancreas* **44**, 311–320 (2015).

113. Tanaka, M. *et al.* Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. *Pancreatology* **17**, 738–753 (2017).

114. Plagge, A., Kelsey, G. & Germain-Lee, E. I. Physiological functions of the imprinted Gnas locus and its protein variants Gαs and XLαs in human and mouse. *J. Endocrinol.* **196**, 193–214 (2008).

115. Landis, C. A. *et al.* GTPase inhibiting mutations activate the α chain of Gs and stimulate adenylyl cyclase in human pituitary tumours. *Nature* **340**, 692–696 (1989).

116. Weinstein, L. S., Liu, J., Sakamoto, A., Xie, T. & Chen, M. Minireview: GNAS: Normal and abnormal functions. *Endocrinology* **145**, 5459–5464 (2004).

117. Knudsen, E. S., O'Reilly, E. M., Brody, J. R. & Witkiewicz, A. K. Genetic Diversity of Pancreatic Ductal Adenocarcinoma and Opportunities for Precision Medicine. *Gastroenterology* **150**, 48–63 (2016).

118. Qian, Z. R. *et al.* Association of alterations in main driver genes with outcomes of patients with resected pancreatic ductal adenocarcinoma. *JAMA Oncol.* **4**, e173420–e173420 (2018).

119. Sullivan, K. D., Galbraith, M. D., Andrysik, Z. & Espinosa, J. M. Mechanisms of transcriptional regulation by p53. *Cell Death Differ.* **25**, 133–143 (2018).

120. SA, H. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).

121. Javle, M. *et al.* Biomarkers of TGF-β Signaling Pathway and Prognosis of Pancreatic Cancer. *PLoS One* **9**, (2014).

122. Jiang, H. *et al.* RhoT1 and Smad4 Are Correlated with Lymph Node Metastasis and Overall Survival in Pancreatic Cancer. *PLoS One* **7**, (2012).

123. Yamada, S. *et al.* SMAD4 Expression Predicts Local Spread and Treatment Failure in Resected Pancreatic Cancer. *Pancreas* **44**, 660–664 (2015).

124. Wilentz, R. E. *et al.* Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. *Cancer Res.* **60**, 2002–6 (2000).

125. Duff, E. & Clarke, A. Smad4 (DPC4) - a potent tumour suppressor? *Br. J. Cancer* **78**, 1615–

1619 (1998).

126. W, H. *et al.* Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4. *J. Pathol.* **242**, 16–23 (2017).

127. Bardeesy, N. *et al.* Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev.* **20**, 3130–3146 (2006).

128. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).

129. Koltsova, A. S. *et al.* On the Complexity of Mechanisms and Consequences of Chromothripsis: An Update. *Front. Genet.* **10**, (2019).

130. Pellestor, F. Chromoanagenesis: Cataclysms behind complex chromosomal rearrangements. *Mol. Cytogenet.* **12**, 1–12 (2019).

131. Jones, M. J. K. & Jallepalli, P. V. Chromothripsis: Chromosomes in Crisis. *Dev. Cell* **23**, 908–917 (2012).

132. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).

133. Storchova, Z. & Kloosterman, W. P. The genomic characteristics and cellular origin of chromothripsis. *Cell Biol.* **40**, 106–113 (2016).

134. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

135. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).

136. Hata, T. *et al.* Genome-Wide Somatic Copy Number Alterations and Mutations in High-Grade Pancreatic Intraepithelial Neoplasia. *Am. J. Pathol.* **188**, 1723–1733 (2018).

137. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* **52**, 331–341 (2020).

138. Rode, A., Maass, K. K., Willmund, K. V., Lichter, P. & Ernst, A. Chromothripsis in cancer cells: An update. *Int. J. Cancer* **138**, 2322–2333 (2016).

139. Dreyer, S. B., Chang, D. K., Bailey, P. & Biankin, A. V. Pancreatic cancer genomes: Implications for clinical management and therapeutic development. *Clin. Cancer Res.* **23**, 1638–1646 (2017).

140. Daniel, F. I., Cherubini, K., Yurgel, L. S., De Figueiredo, M. A. Z. & Salum, F. G. The role of epigenetic transcription repression and DNA methyltransferases in cancer. *Cancer* **117**, 677–687 (2011).

141. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).

142. Kulis, M. & Esteller, M. DNA Methylation and Cancer. *Adv. Genet.* **70**, 27–56 (2010).

143. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1412–1417 (2006).

144. Edgar, R., Tan, P. P., Portales-Casamar, E. & Pavlidis, P. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin* **7**, 28 (2014).

145. Wilson, A. S., Power, B. E. & Molloy, P. L. DNA hypomethylation and human diseases. *Biochimica et Biophysica Acta - Reviews on Cancer* **1775**, 138–162 (2007).

146. Herceg, Z. & Hainaut, P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. *Molecular Oncology* **1**, 26–41 (2007).

147. Varela-Rey, M., Woodhoo, A., Martinez-Chantar, M.-L., Mato, J. M. & Lu, S. C. Alcohol, DNA methylation, and cancer. *Alcohol Res.* **35**, 25–35 (2013).

148. Menezo, Y. J. R., Silvestris, E., Dale, B. & Elder, K. Oxidative stress and alterations in DNA methylation: two sides of the same coin in reproduction. *Reproductive BioMedicine Online* **33**, 668–683 (2016).

149. Field Guide to Methylation Methods. *Illumina Sequencing Methods* Pub. No. 270-2013-001 (2016). Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/field_guide_methylation.pdf. (Accessed: 3rd November 2020)

150. Olivier, M., Hollstein, M. & Hainaut, P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008–a001008 (2010).

151. Cheng, Y. *et al.* Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* **4**, 1–39 (2019).

152. Andrew P, F. & Bert, V. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).

153. Liu, R. *et al.* A systematic review and quantitative assessment of methylation biomarkers in fecal DNA and colorectal cancer and its precursor, colorectal adenoma. *Mutat. Res. Mutat. Res.* **779**, 45–57 (2019).

154. Eissa, M. A. L. *et al.* Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood. *Clin. Epigenetics* **11**, 1–10 (2019).

155. Sierzega, M., Młynarski, D., Tomaszewska, R. & Kulig, J. Semiquantitative immunohistochemistry for mucin (MUC1, MUC2, MUC3, MUC4, MUC5AC, and MUC6) profiling of pancreatic ductal cell adenocarcinoma improves diagnostic and prognostic performance. *Histopathology* **69**, 582–591 (2016).

156. Clément-Ziza, M., Munnich, A., Lyonnet, S., Jaubert, F. & Besmond, C. Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. *RNA* **14**, 2698–2704

(2008).

157. Abbott Laboratories. UroVysion Bladder Cancer Kit. (2014).

158. Richards, S. *et al.* Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology Sue. *Genet. Med.* **17**, 405–424 (2015).

159. Pérez-Palma, E. *et al.* Identification of pathogenic variant enriched regions across genes and gene families. *Genome Res.* **30**, 62–71 (2020).

160. Traynelis, J. *et al.* Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. *Genome Res.* **27**, 1715–1729 (2017).

161. Yu, Y., Ouyang, Y. & Yao, W. ShinyCircos: An R/Shiny application for interactive creation of Circos plot. *Bioinformatics* **34**, 1229–1231 (2018).

162. Cai, H. *et al.* Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *BMC Genomics* **15**, 1–13 (2014).

163. Yang, J. *et al.* CTLPScanner: a web server for chromothripsis-like pattern detection. *Nucleic Acids Res.* **44**, W252–W258 (2016).

164. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

165. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

166. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

167. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

168. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).

169. Sun, J., Nishiyama, T., Shimizu, K. & Kadota, K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **14**, 219–232 (2013).

170. Su, W., Sun, J., Shimizu, K. & Kadota, K. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res. Notes* **12**, 133–139 (2019).

171. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–571 (2014).

172. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, 106–118 (2010).

173. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science (80-.).* **321**, 956–960 (2008).

174. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

175. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).

176. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

177. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

178. MJ, L. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

179. Pea, A. *et al.* Genetic Analysis of Small Well-differentiated Pancreatic Neuroendocrine Tumors Identifies Subgroups With Differing Risks of Liver Metastases. *Ann. Surg.* **271**, 566–573 (2020).

180. Bruno, S. & Darzynkiewicz, Z. Cell cycle dependent expression and stability of the nuclear protein detected by Ki-67 antibody in HL-60 cells. *Cell Prolif.* **25**, 31–40 (1992).

181. Zhao, S., Sun, J., Shimizu, K. & Kadota, K. Silhouette Scores for Arbitrary Defined Groups in Gene Expression Data and Insights into Differential Expression Results. *Biol. Proced. Online* **20**, 1–12 (2018).

182. Chugh, S. *et al.* Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochim. Biophys. Acta - Rev. Cancer* **1856**, 211–225 (2015).

183. Burchell, J. M., Beatson, R., Graham, R., Taylor-Papadimitriou, J. & Tajadura-Ortega, V. O-linked mucin-type glycosylation in breast cancer. *Biochem. Soc. Trans.* **46**, 779–788 (2018).

184. Gupta, R., Leon, F., Rauth, S., Batra, S. K. & Ponnusamy, M. P. A Systematic Review on the Implications of O-linked Glycan Branching and Truncating Enzymes on Cancer Progression and Metastasis. *Cells* **9**, 446 (2020).

185. J, P. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).

186. Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Cancer Research* **72**, 4875–4882 (2012).

187. Kikutake, C., Yoshihara, M., Sato, T., Saito, D. & Suyama, M. Pan-cancer analysis of intratumor heterogeneity associated with patient prognosis using multidimensional measures. *Oncotarget* **9**, 37689–37699 (2018).

188. Chang, X. Y., Wu, Y., Li, Y., Wang, J. & Chen, J. Intraductal papillary mucinous neoplasms

of the pancreas: Clinical association with KRAS. *Mol. Med. Rep.* 8061–8068 (2018). doi:10.3892/mmr.2018.8875

189. Choi, M. H. *et al.* Mutation analysis by deep sequencing of pancreatic juice from patients with pancreatic ductal adenocarcinoma. *BMC Cancer* **19**, 1–12 (2019).

190. Fischer, C. G. & Wood, L. D. From somatic mutation to early detection: insights from molecular characterization of pancreatic cancer precursor lesions. *J. Pathol.* **246**, 395–404 (2018).

191. Fischer, C. G. *et al.* Intraductal Papillary Mucinous Neoplasms Arise From Multiple Independent Clones, Each With Distinct Mutations. *Gastroenterology* **157**, 1123–1137 (2019).

192. Alison, M. R. The cellular origins of cancer with particular reference to the gastrointestinal tract. *Int. J. Exp. Pathol.* **101**, 132–151 (2020).

193. Pearce, M. M. P. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).

194. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).

195. Tan, M. C. *et al.* GNAS and KRAS mutations define separate progression pathways in intraductal papillary mucinous neoplasm-associated carcinoma. *J. Am. Coll. Surg.* **220**, 845-854.e1 (2015).

196. Sakamoto, H. *et al.* Clinicopathological significance of somatic RNF43 mutation and aberrant expression of ring finger protein 43 in intraductal papillary mucinous neoplasms of the pancreas. *Mod. Pathol.* **28**, 261–267 (2015).

197. Castelli, G., Pelosi, E., Testa, U., Castelli, G. & Testa, U. Pancreatic Cancer: Molecular Characterization, Clonal Evolution and Cancer Stem Cells. *Biomedicines* **5**, (2017).

198. Zhao, M., Mishra, L. & Deng, C.-X. The role of TGF-β/SMAD4 signaling in cancer. *Int. J. Biol. Sci.* **14**, 111–123 (2018).

199. O'Hayre, M., Degese, M. S. & Gutkind, J. S. Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr. Opin. Cell Biol.* **27**, 126–135 (2014).

200. Lane Brown, R., Strassmaier, T., Brady, J. & Karpen, J. The Pharmacology of Cyclic Nucleotide-Gated Channels: Emerging from the Darkness. *Curr. Pharm. Des.* **12**, 3597–3613 (2006).

201. Phan, N. N., Huynh, T. T. & Lin, Y.-C. Hyperpolarization-activated cyclic nucleotide-gated gene signatures and poor clinical outcome of cancer patient. *Transl. Cancer Res.* **6**, (2017).

202. Taki, K. *et al.* GNASR201H and KrasG12D cooperate to promote murine pancreatic tumorigenesis recapitulating human intraductal papillary mucinous neoplasm. *Oncogene* **35**, 2407–2412 (2016).

203. Ideno, N. *et al.* GNAS R201C Induces Pancreatic Cystic Neoplasms in Mice That Express

Activated KRAS by Inhibiting YAP1 Signaling. *Gastroenterology* **155**, 1593-1607.e12 (2018).

204. Innamorati, G. *et al.* The curious case of Gαs gain-of-function in neoplasia. *BMC Cancer* **18**, 1–15 (2018).

205. Tamura, K. *et al.* Assessment of clonality of multisegmental main duct intraductal papillary mucinous neoplasms of the pancreas based on GNAS mutation analysis. *Surgery* **157**, 277–284 (2015).

206. Pearce, M. M. P., Spartz, E. J., Hong, W., Luo, L. & Kopito, R. R. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).

207. Aguirre, A. J. *et al.* High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc. Natl. Acad. Sci.* **101**, 9067–9072 (2004).

208. Harada, T. *et al.* Identification of genetic alterations in pancreatic cancer by the combined use of tissue microdissection and array-based comparative genomic hybridisation. *Br. J. Cancer* **96**, 373–382 (2007).

209. Nowak, N. J. *et al.* Genome-wide aberrations in pancreatic adenocarcinoma. *Cancer Genet. Cytogenet.* **161**, 36–50 (2005).

210. Durante, S. *et al.* Copy number gain of chromosome 3q is a recurrent event in patients with intraductal papillary mucinous neoplasm (IPMN) associated with disease progression. *Oncotarget* **7**, 74797–74806 (2016).

211. Adachi, T. *et al.* Bile-Reflux into the Pancreatic Ducts is Associated with the Development of Intraductal Papillary Carcinoma in Hamsters. *J. Surg. Res.* **136**, 106–111 (2006).

212. Muraki, T. *et al.* Variant anatomy of the biliary system as a cause of pancreatic and peri-ampullary cancers. *HPB* **22**, 1675–1685 (2020).

213. Kiss, B. *et al.* Oncobiosis and microbial metabolite signaling in pancreatic adenocarcinoma. *Cancers* **12**, (2020).

214. Luo, Z.-W. *et al.* Increased expression of Ki-67 is a poor prognostic marker for colorectal cancer patients: a meta analysis. *BMC Cancer 2019 191* **19**, 1–13 (2019).

215. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

216. Storz, P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 296–304 (2017).

217. Guerra, C. *et al.* Chronic Pancreatitis Is Essential for Induction of Pancreatic Ductal Adenocarcinoma by K-Ras Oncogenes in Adult Mice. *Cancer Cell* **11**, 291–302 (2007).

218. Navas, C. *et al.* EGF Receptor Signaling Is Essential for K-Ras Oncogene-Driven Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **22**, 318–330 (2012).

219. Lee, A. Y. L. *et al.* Cell of origin affects tumour development and phenotype in pancreatic ductal adenocarcinoma. *Gut* **68**, 487–498 (2019).

220. Bailey, J. M. *et al.* P53 mutations cooperate with oncogenic Kras to promote adenocarcinoma from pancreatic ductal cells. *Oncogene* **35**, 4282–4288 (2016).

221. Choi, B. *et al.* The chromatin regulator Brg1 suppresses formation of intraductal papillary mucinous neoplasm and pancreatic ductal adenocarcinoma. *Nat Cell Biol.* **16**, 255–267 (2014).

222. Rooman, I. & Real, F. X. Pancreatic ductal adenocarcinoma and acinar cells: a matter of differentiation and development? *Gut* **61**, 449–458 (2012).

223. Espinet, E. *et al.* Aggressive PDACs show hypomethylation of repetitive elements and the execution of an intrinsic IFN program linked to a ductal cell-of-origin. *Cancer Discov.* **0**, CD-20-1202 (2020).

224. Omori, Y. *et al.* How does intestinal-type intraductal papillary mucinous neoplasm emerge? CDX2 plays a critical role in the process of intestinal differentiation and progression. *Virchows Arch.* **477**, 21–31 (2020).

225. Kang, M. J. *et al.* Disease Spectrum of Intraductal Papillary Mucinous Neoplasm With an Associated Invasive Carcinoma. *Pancreas* **42**, 1267–1274 (2013).

226. Mino-Kenudson, M. *et al.* Prognosis of invasive intraductal papillary mucinous neoplasm depends on histological and precursor epithelial subtypes. *Gut* **60**, 1712–1720 (2011).

227. Xu, W. *et al.* Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. *Signal Transduct. Target. Ther.* **4**, (2019).

228. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

229. Stirzaker, C., Song, J. Z., Davidson, B. & Clark, S. J. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.* **64**, 3871–3877 (2004).

230. Tan, A. C. *et al.* Characterizing DNA methylation patterns in pancreatic cancer genome. *Mol. Oncol.* **3**, 425–438 (2009).

231. Mishra, N. K. & Guda, C. Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget* **8**, 28990–29012 (2017).

232. Prasad, N. B. *et al.* Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. *Cancer Res.* **65**, 1619–1626 (2005).

233. Cui, J., Shi, M., Quan, M. & Xie, K. Regulation of EMT by KLF4 in Gastrointestinal Cancer. *Curr. Cancer Drug Targets* **13**, 986–995 (2014).

234. Wei, D., Kanai, M., Huang, S. & Xie, K. Emerging role of KLF4 in human gastrointestinal cancer. *Carcinogenesis* **27**, 23–31 (2006).

235. El-Karim, E. A., Hagos, E. G., Ghaleb, A. M., Yu, B. & Yang, V. W. Krüppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts. *Mol. Cancer* **12**, 1–12 (2013).

236. Hu, D. & Wan, Y. Regulation of Krüppel-like factor 4 by the anaphase promoting complex pathway is involved in TGF-β signaling. *J. Biol. Chem.* **286**, 6890–6901 (2011).

237. Kleeff, J. *et al.* The TGF-β signaling inhibitor Smad7 enhances tumorigenicity in pancreatic cancer. *Oncogene* **18**, 5363–5372 (1999).

238. Kuang, C. *et al.* In vivo disruption of TGF-β signaling by Smad7 leads to premalignant ductal lesions in the pancreas. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1858–1863 (2006).

239. Li, X. M. *et al.* KLF4 suppresses the tumor activity of cutaneous squamous cell carcinoma (SCC) cells via the regulation of SMAD signaling and SOX2 expression. *Biochem. Biophys. Res. Commun.* **516**, 1110–1115 (2019).

240. Zammarchi, F. *et al.* KLF4 is a novel candidate tumor suppressor gene in pancreatic ductal carcinoma. *Am. J. Pathol.* **178**, 361–372 (2011).

241. Fujikura, K. *et al.* Multiregion whole-exome sequencing of intraductal papillary mucinous neoplasms reveals frequent somatic KLF4 mutations predominantly in low-grade regions. *Gut* gutjnl-2020-321217 (2020). doi:10.1136/gutjnl-2020-321217

242. Pan, S., Brentnall, T. A. & Chen, R. Glycoproteins and glycoproteomics in pancreatic cancer. *World Journal of Gastroenterology* **22**, 9288–9299 (2016).

243. Tran, D. T. & Ten Hagen, K. G. Mucin-type o-glycosylation during development. *Journal of Biological Chemistry* **288**, 6921–6929 (2013).

244. EV, C., DMarathe, D., Neelamegham, S., Lau, J. Y. & Matta, K. L. A High-level Over-Expression of N-and O-Glycan Glycosyltransferases in Pancreatic Tumors and Diabetic Neutrophils: An Unique Pathological Situation in Pancreatic Cancer and Diabetic Retinopathy. *J. Glycobiol.* **8**, 1–15 (2019).

245. Ohya, A., Yamanoi, K., Shimojo, H., Fujii, C. & Nakayama, J. Gastric gland mucin-specific O-glycan expression decreases with tumor progression from precursor lesions to pancreatic cancer. *Cancer Sci.* **108**, 1897–1902 (2017).

246. Caffrey, T. *et al.* The glycoprotein mucin-1 negatively regulates GalNAc transferase 5 expression in pancreatic cancer. *FEBS Lett.* **593**, 2751–2761 (2019).

247. Remmers, N. *et al.* Aberrant expression of mucin core proteins and O-linked glycans associated with progression of pancreatic cancer. *Clin. Cancer Res.* **19**, 1981–1993 (2013).

248. Munkley, J. The glycosylation landscape of pancreatic cancer (Review). *Oncol. Lett.* **17**, 2569–2575 (2019).

249. Taniuchi, K. *et al.* Overexpression of GalNAc-transferase GalNAc-T3 promotes pancreatic cancer cell growth. *Oncogene* **30**, 4843–4854 (2011).

250. Ohuchida, K. *et al.* Quantitative analysis of MUC1 and MUC5AC mRNA in pancreatic juice for preoperative diagnosis of pancreatic cancer. *Int. J. Cancer* **118**, 405–411 (2006).

251. Hoshi, H. *et al.* Tumor-associated MUC5AC stimulates in vivo tumorigenicity of human pancreatic cancer. *Int. J. Oncol.* **38**, 619–627 (2011).

252. Khaidakov, M. *et al.* Gastric proteins MUC5AC and TFF1 as potential diagnostic markers of colonic sessile serrated adenomas/polyps. *Am. J. Clin. Pathol.* **146**, 530–537 (2016).

253. Sicklick, J. K. & Fanta, P. T. Molecular pathogenesis of biliary tract cancer. in *Blumgart's Surgery of the Liver, Biliary Tract and Pancreas, 2-Volume Set* **1**–**2**, 150-159.e2 (Elsevier, 2017).

254. Järvå, M. A. *et al.* Trefoil factors share a lectin activity that defines their role in mucus. *Nat. Commun.* **11**, 1–9 (2020).

255. Feng, G. *et al.* DNA methylation of trefoil factor 1 (TFF1) is associated with the tumorigenesis of gastric carcinoma. *Mol. Med. Rep.* **9**, 109–117 (2014).

256. Tomita, H. *et al.* Inhibition of gastric carcinogenesis by the hormone gastrin is mediated by suppression of TFF1 epigenetic silencing. *Gastroenterology* **140**, 879-891.e18 (2011).

257. Yamaguchi, J. *et al.* Trefoil factor 1 inhibits epithelial-mesenchymal transition of pancreatic intraepithelial neoplasm. *J. Clin. Invest.* **128**, 3619–3629 (2018).

258. Klett, H. *et al.* Identification and Validation of a Diagnostic and Prognostic Multi-Gene Biomarker Panel for Pancreatic Ductal Adenocarcinoma. *Front. Genet.* **9**, 1–14 (2018).

259. Arumugam, T. *et al.* Trefoil Factor 1 Stimulates Both Pancreatic Cancer and Stellate Cells and Increases Metastasis. *Pancreas* **40**, 815–822 (2011).

260. Jahan, R. *et al.* Trefoil factor(s) and CA19.9: A promising panel for early detection of pancreatic cancer. *EBioMedicine* **42**, 375–385 (2019).

261. De Schaetzen Van Brienen, L. *et al.* Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. *BMC Med. Genomics* **13**, (2020).

7 List of Tables

Table 1: Equipment 28
Table 2: Kits 30
Table 3: Chemicals and Reagents. 32
Table 4: Buffer. 34
Table 5: Primer. 35
Table 6: DEPArray antibodies. 36
Table 7: IHC antibodies. 36
Table 8: Software 37
Table 9: Study cohort. 40
Table 10: Application of IHC antibodies. 42
Table 11: qPCR program of the two different quantification programs for the DNA quality. 46
Table 12: List of genes involved in the PDAC gene panel. 47
Table 13: PCR program NGS. 47
Table 14: Cell lysis of the low-pass workflow. 52
Table 15: Preparation of Digestion reaction mix. 52
Table 16: Preannealing Reaction mixture. 52
Table 17: Ligation Mixture. 53
Table 18: Primary PCR reaction mixture. 53
Table 19: PCR Thermal incubation profile of primary PCR reaction. 53
Table 20: Barcode reaction. 54
Table 21: PCR Program Barcode reaction. 54
Table 22: Overview of the mutation distribution of all genes in the panels. 64
Table 23: DEGs and AS value of each group of the transcription data. 78
Table 24: NGS Panel #1. 133
Table 25: NGS Panel #2. 139
Table 26: 84 genes from the Integrative analysis between gastric IPMN and PanIN. 146

8 List of Figures

Figure 1: Macroscopy and morphology of the PDAC. 7
Figure 2: The 5-year survival in different cancer types between 1971 and 2011. 9
Figure 3: Morphology of a PanIN lesion and the immunohistology pattern. 12
Figure 4: Morphology atypical flat lesion (AFL). 13
Figure 5: Morphology and marker expression of the different IPMN subtypes. 15
Figure 6: Intraductal oncocystic papillary neoplasms. 16
Figure 7: Mucinous cystic neoplasm of the pancreas. 17
Figure 8: PanIN and IPMN development to invasive carcinoma. 19
Figure 9: Mechanisms of cancer progression via chromothripsis. 23
Figure 10: Mechanism of cytosine methylation and perturbation of methylation. 25
Figure 11: Morphology and immunohistology profile of gastric IPMN and intestinal IPMN. . 39
Figure 12: Precursor lesion distribution of the tissue microarray. 40
Figure 13: Pipeline Low-pass Sequencing. 49
Figure 14: One example of the double positive recruited cells of the DEPArray™. 51
Figure 15: Probes of the UroVysion kit from Abbott. 58
Figure 16: Mutation profile of targeted NGS. 63
Figure 17: Variant allele frequency of *KRAS* and *GNAS* mutations in different precursor lesions. 65
Figure 18: Correlation of the VAF of *KRAS* and *GNAS* in gastric IPMN samples. 66
Figure 19: Copy number variations in PanIN, gastric IPMN, intestinal IPMN, AFL and PDAC. 69
Figure 20: Copy number variations of the array-based Illumina data. 70
Figure 21: Chromothripsis-like events in precursor lesions of the PDAC. 71
Figure 22: Two examples of a cell with a diploid manner. 72
Figure 23: Proliferation rate of PanIN, gastric IPMN and intestinal IPMN. 73
Figure 24: GNAS expression in gastric IPMN, PanIN and intestinal IPMN. 74
Figure 25: Transcription profile and methylation pattern of *GNAS*. 75
Figure 26: PCA Plot (2D) of all FFPE transcriptome samples. 76
Figure 27: Heatmap of DEGs from the transcriptome data of FFPE samples. 77
Figure 28: Unique downregulated genes in gastric IPMN and PanIN lesions of the DEGs. 78
Figure 29: PCA Plot with gastric IPMN and PanIN lesions. 79

Figure 30: Volcano plot of different expressed genes in PanIN lesions in comparison to gastric IPMN. 80
Figure 31: Heatmap of DEGs between gastric IPMN and PanIN. 81
Figure 32: Pie chart of the precursor distribution for the methylome analysis. 82
Figure 33: Multidimensional scaling plot of the DNA methylation data. 83
Figure 34: Volcano plots represent the difference methylated genes of each sample type... 84
Figure 35: Distance-based, neighbor joining tree calculated from methylation status of 49 different precursor samples. 85
Figure 36: CpGs differently methylated characterize according to the CG annotation between PanIN and gastric IPMN. 86
Figure 37: Integrative analysis of DNA methylation and gene expression between PanIN and gastric IPMN. 88
Figure 38: Protein-protein interaction network of hypermethylated and downregulated genes. 89
Figure 39: Expression and methylation pattern of MUC6, TFF1 and KLF4 in different precursors. 90
Figure 40: Expression profile of KLF4, MUC6 and TFF1 on precursors of pancreatic cancer. . 92
Figure 41: Quantitative analysis of MUC6, KLF4 and TFF1 in precursor lesions, acinar tissue and PDAC tissue 93
Figure 42: DEGS from O-glycan mucin biosynthesis in gastric IPMN and PanIN. 94
Figure 43: Enrichment analysis with the webtool enrichr. 145

9 Communications

Haeberle L, Schramm M, Goering W, Frohn L, Driescher C, Hartwig W, Preissinger-Heinzel K-H, Beyna T, Neuhaus H, Fuchs K, Keitel-Anselmino V, Knoefel W. T, Esposito I: Molecular analysis of cyst fluids improves the diagnostic accuracy of pre-operative assessment of pancreatic cystic lesions. Sci. Rep. 11, (2021).

Driescher C, Fuchs K, Haeberle L, Goering W, Frohn L, Opitz F. V, Haeussinger D, Knoefel W. T, Keitel V, Esposito I: Bile-Based Cell-Free DNA Analysis Is a Reliable Diagnostic Tool in Pancreatobiliary Cancer. Cancers (Basel). 13, 39 (2020).

Presentations

- Oral Presentation 49th EPC (European pancreas club) May 2018, Berlin (Germany). Molecular characterization of early precursor lesions of the PDAC. *Frohn L, Häberle L, Göring W, Esposito I.*
- Poster Presentation DGP (Deutsche Gesellschaft für Pathologie) May 2018, Berlin (Germany). Molecular characterization of early precursor lesions of the PDAC. *Frohn L, Häberle L, Göring W, Esposito I.*
- Poster Presentation at 50th EPC (European pancreas club) June 2019, Bergen (Norway). Molecular characterization of early precursor lesions of the PDAC. *Frohn L, Häberle L, Göring W, Esposito I.*
- Oral Presentation DPC (Deutscher Pankreasclub) February 2020, Halle (Germany). Molecular characterization of early precursor lesions of the PDAC. *Frohn L, Häberle L, Göring W, Liffers S-T, Godfrey L, Siveke JT, Esposito I.*

Travel grand

- 50th EPC (European pancreas club) June 2019; Bergen (Norway)
- DPC (Deutscher Pankreasclub) February 2020; Halle (Germany)

10 Eidesstattliche Versicherung

Ich, Lisa Lotte Frohn, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Duisburg, 26.10.2021 _____

Unterschrift

11 Attachment

11.1 Self-designed NGS Panel #1

Table 24: NGS Panel #1.

Gene symbol	Chr.	Ion AmpliSeq Fwd Primer (5'-3')	Ion AmpliSeq Rev Primer (5'-3')	Amplicaon ID
ALK	chr2	TCTCTCGGAGGAAGGACTTGAG	GCCCAGACTCAGCTCAGTTAAT	AMPLCHP2_ALK_1
ALK	chr2	ACAGGGTACCAGGAGATGATGTAAG	GGAAGAGTGGCCAAGATTGGA	AMPLCHP2_ALK_2
APC	chr5	GAGAGAACGCGGAATTGGTCTA	GTATGAATGGCTGACACTTCTTCCA	AMPLCHP2_APC_1
APC	chr5	AGCACTGATGATAAACACCTCAAGTT	ATCTTCTTGACACAAAGACTGGCT	AMPLCHP2_APC_2
APC	chr5	TTCATTATCATCTTTGTCATCAGCTGAA	TTTGGTTCTAGGGTGCTGTGAC	AMPLCHP2_APC_3
APC	chr5	GCAGACTGCAGGGTTCTAGTT	GTGAACTGACAGAAGTACATCTGCT	AMPLCHP2_APC_4
APC	chr5	AGCCCCAGTGATCTTCCAGATA	CCCTCTGAACTGCAGCATTTACT	AMPLCHP2_APC_5
APC	chr5	AGAGGGTCCAGGTTCTTCCA	TCATTTTCCTGAACTGGAGGCATT	AMPLCHP2_APC_6
APC	chr5	ATGAAACAGAATCAGAGCAGCCTAAA	CGTGATGACTTTGTTGGCATGG	AMPLCHP2_APC_7
ARID1A	chr1	GGGCTACCAGGGCTACC	GCGTGGTGGCTCCTTTGTT	AMPL415290468
ARID1A	chr1	GGACATGGCCTCGCAGT	CCTCCCCACTCAGCTGTGTA	AMPL415290528
ARID1A	chr1	CTCTCCGCGGACGAGAC	GGCGGCTGCCTTCATTTC	AMPL416196718
ARID1A	chr1	GCCCTCGGAGCTGAAGA	CCGTCCTGCAGCTCCTT	AMPL416196984
ARID1A	chr1	CCGCTGGGAAAGGAGCTG	CGGCTCCGTGAGGTTATTGTT	AMPL416197271
ARID1A	chr1	CCGGACCTGAAGAACTCGAA	GGTTGCCCGAAGCCGTA	AMPL416197501
ARID1A	chr1	CTCTCACGACCACGGCTT	CGCTGCTGAGCGAAGGA	AMPL418372580
ARID1A	chr1	CGCCTCCTCGTCGTCTT	GGCCCGCCACTGTAGTC	AMPL418375392
ARID1A	chr1	GCCGTCTTCCACCAACAACA	GGGTAGTAGGAGTTGTACTGGTG	AMPL7153047556
ARID1A	chr1	GGAAGCATGTGACAACAGCAGA	TCTCCTTGATGGCCTCTGAACT	AMPL7153047925
ARID1A	chr1	TGAACAATAACCTCACGGAGCC	GTCGTGAGAGTTCTGCTGGG	AMPL7153107115
ARID1A	chr1	GCCCTCGGAGCTGAAGAAA	GTTCTTCAGGTCCGGCTCC	AMPL7153107128
ARID1A	chr1	GTGAGTTGCTAGTGAGTGACTAACC	CCAGGCTGTCCATGCATTTGA	AMPL7153198771
ARID1A	chr1	AGCCATTTCTAGCTCTGAATTAACTTCC	TCGATCTTGGGCAATGCTTGAT	AMPL7153198794
ARID1A	chr1	GTTTAAGGAAAATGCTAAGCAAGTAGTAGG	GGAAATCCCTGATGTGCTCACT	AMPL7153220108
ARID1A	chr1	CCATCACAGCTTTTGTTTTTCTTGTTG	ACCTTTCAGAAGGTGCAGAAATACT	AMPL7153319997
ARID1A	chr1	CCCAAAGAGATTCTGGGTCGTT	CTTTAAGTCTCCTCCTTCTGCCAT	AMPL7154253861
ARID1A	chr1	CTCTGTACGAATCACGGCTGTT	AGGGCAACAGTCAGTTTCTAAGTTC	AMPL7155872860
ARID1A	chr1	GGCCACTTTTCTCCCTTAATTTATTTCCT	TGAACCTCCCAGGATCCAGTAG	AMPL7158830024
ARID1A	chr1	CGGGAGATATACCTCGACTCCTTT	CCCGTTCAATCTTGCATTCAAAGG	AMPL7158830077
ARID1A	chr1	CCAGTATGGCAATGCCTATCCT	CATGTTTTGCTGGGCATTGGT	AMPL7158830135
ARID1A	chr1	ACAGCTAAACTTACTGGACTTGAGAATTTT	AGTCAAGACAAAAATCACTACCTTGGAT	AMPL7158830192
ARID1A	chr1	GGAGAACCGCACCTCTCCTA	AGCCAGGTGGGAAGGTGATA	AMPL7158830197
ARID1A	chr1	TGGTTCAATAGATGACCTCCCCAT	GGGAGGTATGAGGAGAGAAAGG	AMPLP222485734
ARID1A	chr1	AGCAGAGTAATCCAGCTCAGTCT	CCCTAAAGCCCAGCACTTGA	AMPLP222488467
ARID1A	chr1	GCCCCCTCCATCTAACTACCAG	ATTTTCATCCCAGAGTGCAGGAA	AMPLP222552803
ARID1A	chr1	AGTCGGACAGCATCATGCATC	GGTCAAACAGCTCTCCAAAAGCTA	AMPLP222566968
ARID1A	chr1	CAGTACCTGCCTCGCACATAG	GGCAGATTAGGCAACCGAATGA	AMPLP222573315
ARID1A	chr1	CCAGGCTTGTCAACTTACCAGTTT	CATCGCCACATTTCCTAGCC	AMPLP222657782
ARID1A	chr1	GGTGACCCCTACAGTCGTG	AGAGGCAGATTGAGCCCACTATA	AMPLP222666532

ARID1A	chr1	CTCATCATCAGTGCATAGCTTCTCA	GGTGAGGACTTTGCTGGTTGT	AMPLP222689454
ARID1A	chr1	CAACAGGGCCAGACTCCATATT	CTGGGAGTATGGAGGCTGAGA	AMPLP222713071
ARID1A	chr1	CCACCACAGCTCCAGTCC	GAGGTTGCTGCTGCTGGTA	AMPLP222720125
ARID1A	chr1	CAGCCACAGGCTCAGTCTC	GCAGGAGGCAGGGATATCTT	AMPLP222726251
ARID1A	chr1	CCCAGGATAAGGATGGAGAGCAT	GTGTGTATCTGTCCTCCGGAAG	AMPLP222870435
ARID1A	chr1	GACTCCTGCGTGTCCTTTGTTA	TGATACCCAGGGTTTGGAGTCA	AMPLP222879003
ARID1A	chr1	CAGCCTTATCTCCGCGTCA	ACTGTTTTCTCCTCTCACCCGTAT	AMPLP222881883
ARID1A	chr1	ACATTCCAGAAGCGGAATTCCA	GAGCACAAGTTCAAATAGCAATCAGAT	AMPLP222906595
ARID1A	chr1	CAGACTCCCCAGTCAACCAG	CAAGGAGTTCCCATGCACTTATCTT	AMPLP223141663
ARID1A	chr1	GTGGATAGACGACATGGAGGTT	ATCATACAGCAGGATGTTGATGGTATC	AMPLP223180129
ARID1A	chr1	GCAGAGAGCACATGGGCATTA	AGAAGATCCCAAACCCTCTCAATCT	AMPLP223193788
ARID1A	chr1	GCCAGCTCCTTGAAAAAGCAGTAT	CCGCTGACCCCATCCTTAC	AMPLP223289373
ARID1A	chr1	ACCACCCCAGAGTAAGAAGCTTTA	AACTTGGTGATCTTCTCATTGGTTGT	AMPLP223443770
ARID1A	chr1	GAGGCCATCAAAGCTCAGGTTAA	GGGTACCCATGTCCTTGCTG	AMPLP223455681
ARID1A	chr1	CAACAGGGACCTCCGTCAG	AACAAGGGTCAAGGTAATCACAATCA	AMPLP223467808
ARID1A	chr1	ACCTTCCAGAAATCCAGTTCTTCTACT	AGAGGTCCAGAGGTTTCCTACC	AMPLP223482884
ARID1A	chr1	CATGACAAATCTGCCTGCTGTG	CATGAAGCCAGTGAGTACCTAGAAAG	AMPLP223507169
ARID1A	chr1	AGGGTTGCTAGAGCTCCTTGTA	GGAGCTGGACTAGACACCTTG	AMPLP223518210
ARID1A	chr1	CACTGAGCATATCCAGACCCA	TGGTCTGTTGTCCCTGGTGTA	AMPLP223518218
ARID1A	chr1	CTACTGGATCCTGGGAGGTTCA	CACTATTCTCTGAAGCTGGCTTGT	AMPLP223532950
ARID1A	chr1	AGGAGATAGCCTTTTCAGGCAAG	CTTATCTGAGCAGTCCACCACAAAT	AMPLP223539842
ARID1A	chr1	CTTGACCGAGGATGGAGCTAA	GTCCAGAAGGGTACACAGTGG	AMPLP223543832
ARID1A	chr1	GCTTCCAGTAAAGATCGTACAGAAGAAT	AGCTCTGTCTTGCTCTCGAAG	AMPLP223551163
ARID1A	chr1	CGAACCCCACAGTAAGGATGAGA	AAGTCATTGCCTGGCACAAATG	AMPLP223561744
ARID1A	chr1	GCCCACCCTAATCCTGTGTT	CCAGTCCATTCCCAGCTACC	AMPLP223571185
ARID1A	chr1	CCAGCACCCTGAAGCTATAGT	AGGAGAATACATCCCCGAGTCTG	AMPLP223578223
ARID1A	chr1	GTCCAATACCATTCGAAGCCTGT	CTCACCCCTTGGTCCTGTTC	AMPLP223579249
ARID1A	chr1	CGAATCTCATGCCTTCCAACC	GTGGCGAAGCCTGATCCATA	AMPLP223591659
ARID1A	chr1	CAGGCACCACTAACTTATGAAAAGGA	AGATGTTGGCGAGTGTAACCAAG	AMPLP223601702
ARID1A	chr1	CATGGCTGGAGGCATAAACC	GGACACATCCCTGACCCAAC	AMPLP224713333
ARID1A	chr1	ATGGCCAATATGCCACCTCA	CTCTGCTTCCCAGGCCTTAC	AMPLP224726605
ARID1A	chr1	GTGCCATTGCAGTGCAGAAG	CGCCGCATCATGTCCACA	AMPLP224973534
ARID1A	chr1	CTTGGAGATGCTCCGGGAAAA	CAAACTGCCCAGTGTAGGAGTC	AMPLP224979763
ARID1A	chr1	CCCACCCTTTGAGCCAACTA	TGACTTGTGAAACCAATGAGTTCATCA	AMPLP224984253
ARID1A	chr1	TCCATACCCCGAGAGCATTTG	ACATTGTTGTCCTGGATGCTGAG	AMPLP224985943
ARID1A	chr1	GGTCTTGGAAACCCTCAGCAAA	CGGGTTCTTTCGGTCACTGA	AMPLP224997894
ARID1A	chr1	TTGTATAGCACTATGGTGCGCTT	CCAGGAGGTTGCCGATACTG	AMPLP225009187
ARID1A	chr1	GGCGTGAACCGAACAGATGA	AATGTGATTCTGCATGCTTGGTG	AMPLP225137867
ARID1A	chr1	CTTTATGTCCCTGAGTGCAGAGTAT	CAACTGCTGCTGCTGAGG	AMPLP225144662
ARID1A	chr1	GGAGATGTACAGCGTGCCATA	GATAGGCATTGCCATACTGGTTG	AMPLP225154888
ARID1A	chr1	CCTTCCCCTCAGCAAGATGTATA	ATGTTTTGCTGGGCATTGGTG	AMPLP225160828
ARID1A	chr1	ACAGCACTATTTGGCTCCAGTT	CAGCCATACTATTAATCCCTTGTCCATA	AMPLP225164710
ARID1A	chr1	CCATTCCAGTTTGGCCGAGA	CCTGTTGGCATAATTATAGGTCATGTCA	AMPLP225170797
ARID1A	chr1	CATACAGGCATCAGCTGAGGTT	CCTCTGATCTGTGTGCAGCATT	AMPLP225173327
ARID1A	chr1	TGATGGGAACTGGACCTCCT	TGACTCTTGAAGAAATCCCTAGTGAGAA	AMPLP225180821
BRAF	chr7	CATACTTACCATGCCACTTTCCCTT	TTTCTTTTTCTGTTTGGCTTGACTTGA	AMPLCHP2_BRAF_1

BRAF	chr7	CCACAAAATGGATCCAGACAACTGT	GCTTGCTCTGATAGGAAAATGAGATCTA	AMPLCHP2_BRAF_2
CDKN2A	chr9	CACCAGCGTGTCCAGGAA	CCCTGGCTCTGACCATTCTGT	AMPLCHP2_CDKN2A_1
CDKN2A	chr9	CATCTATGCGGGCATGGTTACT	ACACGCTGGTGGTGCTG	AMPLCHP2_CDKN2A_2
CTNNB1	chr3	ACTGTTTCGTATTTATAGCTGATTTGATGGA	CCTCTTCCTCAGGATTGCCTTT	AMPLCHP2_CTNNB1_1
EGFR	chr7	CCTCATTGCCCTCAACACAGT	TCAGTCCGGTTTTATTTGCATCATAGTT	AMPLCHP2_EGFR_1
EGFR	chr7	CACCACGTACCAGATGGATGT	CCCAAAGACTCTCCAAGATGGGATA	AMPLCHP2_EGFR_2
EGFR	chr7	AGACATGCATGAACATTTTTCTCCAC	TCCAGACCAGGGTGTTGTTTTC	AMPLCHP2_EGFR_3
EGFR	chr7	TGTGGAGCCTCTTACACCCA	GTGCCAGGGACCTTACCTTATAC	AMPLCHP2_EGFR_4
EGFR	chr7	ACGTCTTCCTTCTCTCTCTGTCA	CTGAGGTTCAGAGCCATGGA	AMPLCHP2_EGFR_5
EGFR	chr7	CATGCGAAGCCACACTGAC	CGGACATAGTCCAGGAGGCA	AMPLCHP2_EGFR_6
EGFR	chr7	GACTATGTCCGGGAACACAAAGA	CCCCATGGCAAACTCTTGCTA	AMPLCHP2_EGFR_7
EGFR	chr7	CGCAGCATGTCAAGATCACAGAT	GCATGTGTTAAACAATACAGCTAGTG	AMPLCHP2_EGFR_8
FBXW7	chr4	TGACAATGTTTAAAGGTGGTAGCTGTT	ACTCATTGATAGTTGTGAACCAACACA	AMPLCHP2_FBXW7_1
FBXW7	chr4	CCTGTGACTGCTGACCAAACTTTTA	CACATCTTTCTTATAGGTGCTGAAAGG	AMPLCHP2_FBXW7_2
FBXW7	chr4	CCCAACCATGACAAGATTTTCCC	GGTCATCACAAATGAGAGACAACATCA	AMPLCHP2_FBXW7_3
FBXW7	chr4	ACTAACAACCCTCCTGCCATCATA	TCTGCAGAGTTGTTAGCGGTT	AMPLCHP2_FBXW7_4
FBXW7	chr4	GTAGAATCTGCATTCCCAGAGACAA	TCTCTTGATACATCAATCCGTGTTTGG	AMPLCHP2_FBXW7_5
FGR2	chr10	CATCACTGTAAACCTTGCAGACAAAC	TGGTCTCTCATTCTCCCATCCC	AMPLCHP2_FGFR2_1
FGR2	chr10	CATCCTCTCTCAACTCCAACAGG	AGTGGATCAAGCACGTGGAAAA	AMPLCHP2_FGFR2_2
FGR2	chr10	GCTTCTTGGTCGTGTTCTTCATT	CTCCTCCTGTGATCTGCAATCT	AMPLCHP2_FGFR2_3
FGR2	chr10	TGGAAGCCCAGCCATTTCTAAA	GATGATGAAGATGATTGGGAAACACAAG	AMPLCHP2_FGFR2_4
GNAS	chr20	TTGGTGAGATCCATTGACCTCAATTT	TGAATGTCAAGAAACCATGATCTCTGTT	AMPLCHP2_GNAS_1
GNAS	chr20	CCTCTGGAATAACCAGCTGTCC	TGATCCCTAACAACACAGAAGCAA	AMPLCHP2_GNAS_2
IDH2	chr15	ACCCTGGCCTACCTGGTC	AGTTCAAGCTGAAGAAGATGTGGAA	AMPLCHP2_IDH2_1
IDH2	chr2	CCAACATGACTTACTTGATCCCCAT	ATCACCAAATGGCACCATACGA	AMPLCHP2_IDH1_1
KRAS	chr12	CAAAGAATGGTCCTGCACCAGTAATAT	AGGCCTGCTGAAAATGACTGAATATAA	AMPLCHP2_KRAS_1
KRAS	chr12	TCCTCATGTACTGGTCCCTCATT	GTAAAAGGTGCACTGTAATAATCCAGACT	AMPLCHP2_KRAS_2
KRAS	chr12	CAGATCTGTATTTATTTCAGTGTTACTTACCT	GACTCTGAAGATGTACCTATGGTCCTA	AMPLCHP2_KRAS_3
NRAS	chr1	CCTCACCTCTATGGTGGGATCATAT	GTTCTTGCTGGTGTGAAATGACTG	AMPLCHP2_NRAS_1
NRAS	chr1	TTCGCCTGTCCTCATGTATTGG	CACCCCCAGGATTCTTACAGAAAA	AMPLCHP2_NRAS_2
NRAS	chr1	GCACAAATGCTGAAAGCTGTACC	CAAGTGTGATTTGCCAACAAGGA	AMPLCHP2_NRAS_3
PIK3CA	chr3	CCATAAAGCATGAACTATTTAAAGAAGCAAGA	GGTTGAAAAAGCCGAAGGTCAC	AMPLCHP2_PIK3CA_1
PIK3CA	chr3	TGGAATGCCAGAACTACAATCTTTTGAT	GTGGAAGATCCAATCCATTTTTGTTGTC	AMPLCHP2_PIK3CA_10
PIK3CA	chr3	TGGATCTTCCACACAATTAAACAGCAT	TGCTGTTCATGGATTGTGCAATTC	AMPLCHP2_PIK3CA_11

PIK3CA	chr3	CCCTTTTTAAAAGTAATTGAACCAGTAGGC	TTTAAGATTACGAAGGTATTGGTTTAGACAGAA	AMPLCHP2_PIK3CA_2
PIK3CA	chr3	GACGCATTTCCACAGCTACAC	AGCATCAGCATTTGACTTTACCTTATCA	AMPLCHP2_PIK3CA_3
PIK3CA	chr3	CATAGGTGGAATGAATGGCTGAATTATG	TCAATCAGCGGTATAATCAGGAGTTTTT	AMPLCHP2_PIK3CA_4
PIK3CA	chr3	TCCCATTATTATAGAGATGATTGTTGAATTTTCCT	CAAACAAGTTTATATTTCCCCATGCCA	AMPLCHP2_PIK3CA_5
PIK3CA	chr3	GCTTTGAATCTTTGGCCAGTACCT	CATAAGAGAGAAGGTTTGACTGCCATA	AMPLCHP2_PIK3CA_6
PIK3CA	chr3	CAGAGTAACAGACTAGCTAGAGACAATGA	GCACTTACCTGTGACTCCATAGAAA	AMPLCHP2_PIK3CA_7
PIK3CA	chr3	CACGATTCTTTTAGATCTGAGATGCACA	CCTTTTGTGTTTCATCCTTCTTCTCCTG	AMPLCHP2_PIK3CA_8
PIK3CA	chr3	GATGCAGCCATTGACCTGTTTAC	AGAAAACCATTACTTGTCCATCGTCT	AMPLCHP2_PIK3CA_9
PTEN	chr10	GCCATCTCTCTCCTCCTTTTTCTT	GCCGCAGAAATGGATACAGGTC	AMPLCHP2_PTEN_1
PTEN	chr10	TGTTAATGGTGGCTTTTTGTTTGTTTGT	TCTACCTCACTCTAACAAGCAGATAACT	AMPLCHP2_PTEN_2
PTEN	chr10	CCATAACCCACCACAGCTAGAA	TGCCCCGATGTAATAAATATGCACAT	AMPLCHP2_PTEN_3
PTEN	chr10	GGCTACGACCCAGTTACCATAG	TGCCACTGGTCTATAATCCAGATGAT	AMPLCHP2_PTEN_4
PTEN	chr10	TGAGATCAAGATTGCAGATACAGAATCC	ACCTTTAGCTGGCAGACCAC	AMPLCHP2_PTEN_5
PTEN	chr10	AGGTGAAGATATATTCCTCCAATTCAGGAC	TTGGATATTTCTCCCAATGAAAGTAAAGTAC	AMPLCHP2_PTEN_6
PTEN	chr10	CACTTTTGGGTAAATACATTCTTCATACCAGGA	CACGCTCTATACTGCAAATGCTATCGA	AMPLCHP2_PTEN_7
PTEN	chr10	GCAGTATAGAGCGTGCAGATAATGA	CATCACATACATACAAGTCAACAACCC	AMPLCHP2_PTEN_8
RNF43	chr17	CACGTACTCCTTCCTTCTCCCTA	GTGTGGATCCTAATGACAGTGGT	AMPL7153209699
RNF43	chr17	GGGCAGAGAGGCTGGATTTT	GTCACCAGATCCAACTCAGCA	AMPL7153209729
RNF43	chr17	CTGCTGAGTTGGATCTGGTGA	ACCACCACTACAAAAAGCGGTT	AMPL7153209735
RNF43	chr17	GGATCTGGTGACTTGCTGATCA	GGTTTCCAGCCATGTCCACTAC	AMPL7153209741
RNF43	chr17	TGGAACCGCTTTTTGTAGTGGT	CCTAGTGTACTGCAGCCCTAAAG	AMPL7153209743
RNF43	chr17	CTGTGAGGTGGATTGGAGGTG	ACGACCTGGTCCCTTCCTG	AMPL7153209759
RNF43	chr17	ACAAGAGAGCACAGGGTCTTCT	GGATTCATCAGCATCGTCAAGCT	AMPL7153209793
RNF43	chr17	TGTAGGGCGAAGTGTGAGTCTA	CCCTAACCCAAGTCTGTCTCTCT	AMPL7153539400
RNF43	chr17	TGGCCCATCTGCCAGGTAC	CCCTATGCACAAGGCTGGG	AMPL7154058091
RNF43	chr17	GGTAGTGGACATGGCTGGAAAC	GTTCTACTTTCTGCAGCTCCCTAA	AMPL7154058092
RNF43	chr17	GGACCAAGGATATGCCACACT	CCAGCAGTCTGTTCAACTTGC	AMPL7154058093
RNF43	chr17	CACCCACTTCCCTCTGAAAACT	CAAATTCACAGCCAGTGTGGTT	AMPL7154058096
RNF43	chr17	TGCATTAATTTTCCTTCTGCTGGAGTTAT	GACGCACAGGACTGGTACT	AMPL7154392234
RNF43	chr17	CCGAAGCCAGGATGATCACAAA	CTGTGTTTCCTCCCGAATTGAC	AMPL7154392238
RNF43	chr17	GGTGTCAGAACTCCATTCAGAAGG	ACAAGAGGCTGCTACCAGAAAC	AMPL7154392240
RNF43	chr17	GAGGCTGGATTTTTGCAAGTTGA	AGAGCCACCTTCTCCTGATCA	AMPL7154392243
RNF43	chr17	CCTTTAGGGCTGCAGTACACTAG	TCATGGCTCTTCCAGTGACTCT	AMPL7154392244
RNF43	chr17	GTCTGGAGGTCTAGTGTGCTTAC	CTGGCCAGTGGTGTTGATCT	AMPL7154395019
RNF43	chr17	CCCACCTTGAACACGCAAATG	CCTCAGCCCAACCTCTACTGT	AMPL7154412295
RNF43	chr17	TAGGCAGCAGGGAGGTGGTAG	GTTAAACATCTGTGCTCTTGGTTCTTTTT	AMPL7154412297
RNF43	chr17	CCTCCTACCTGTGATGTTGAACAT	CTATGGCTACAGAAGCCTTTGGT	AMPL7156001482

RNF43	chr17	CCCTCAGAGAACTCCTCCAGA	CCCATTCCTCTTCCCAGGATC	AMPL7156565905
RNF43	chr17	AAGCGGTGATGCCGAGG	GGTCGAAGACTCCACCTCATTC	AMPL7156565906
RNF43	chr17	AGGTGAACCACAAGACCTGC	CTGGCCACCAGGAGGTACCA	AMPL7157047061
RNF43	chr17	GCTTTCTGTTCTGCTGATCTTTCAG	AGTATGTATGGTTGAAGTGCATTGCT	AMPL7158269366
RNF43	chr17	CCATCAGCTTCTCAGCGTCATT	GGGCTGGATGGAGGAAGATAAAG	AMPL7158680324
RNF43	chr17	CCATCAACCCTTTGTTGGCTGA	GGGCCTTCTGAATGGAGTTCTG	AMPL7158680326
RNF43	chr17	ACAACCACACTGGCTGTGAAT	CACCCAGCTTGCCAGATTTTTC	AMPL7158680327
RNF43	chr17	CGGCAAAAAGAATGGTGTTTGCT	TCTAAATACAGGCTCAGAGGAGGAA	AMPL7158680331
RNF43	chr17	AGTAGAACTGCTGCCATGGAC	TGACAGCAGTGGATCTGGAGA	AMPL7158680332
RNF43	chr17	TGGGAGACAAAAGAAGAAAGACATATTTCA	CACTTTGGAAGGTGTGTTTGCTG	AMPL7159410737
RNF43	chr17	CATCTCACACAGCCTGTTCACA	CTTTCCAGACAGTCTCTGCTTCTC	AMPL7163079287
RNF43	chr17	TATGGTGGCAGTTCTGCTTTCTC	CTGGTTACATCAGCATCGGACTT	AMPL7163079318
RNF43	chr17	CCAGGCTGTTCTCTGCTGAAG	GAGACGAACTCCTAAGGATGTTCAAC	AMPL7163079324
SMAD4	chr18	CTCATGTGATCTATGCCCGTCT	AGTCTACTTACCAATTCCAGGTGATACA	AMPLCHP2_SMAD4_1
SMAD4	chr18	TGCTACTTCTGAATTGAAATGGTTCA	GATTACCTACCATTACTCTGCAGTGTT	AMPLCHP2_SMAD4_2
SMAD4	chr18	ATGGTGAAGGATGAATATGTGCATGA	GCTGGTAGCATTAGACTCAGATGG	AMPLCHP2_SMAD4_3
SMAD4	chr18	GTGAAGGACTGTTGCAGATAGCAT	AAGGCCCACATGGGTTAATTTG	AMPLCHP2_SMAD4_4
SMAD4	chr18	TTTCTTTAGGGCCTGTTCACAATGA	CTGAGAAGTGACCCCATAATTCCATT	AMPLCHP2_SMAD4_5
SMAD4	chr18	GCTCCTGAGTATTGGTGTTCCAT	CCTGTGGACATTGGAGAGTTGA	AMPLCHP2_SMAD4_6
SMAD4	chr18	TGTAATTTCTTTTTTCTTCCTAAGGTTGCACATAG	ACTTGGGTAGATCTTATGAACAGCAT	AMPLCHP2_SMAD4_7
SMAD4	chr18	AGGTCTTTGATTTGCGTCAGTGT	GCTGGAGCTATTCCACCTACTG	AMPLCHP2_SMAD4_8
SMAD4	chr18	GCTGCTGGAATTGGTGTTGATG	AGTACTTCGTCTAGGAGCTGGAG	AMPLCHP2_SMAD4_9
STK11	chr19	GAGCTGATGTCGGTGGGTAT	CTCCGAGTCCAGCACCTC	AMPLCHP2_STK11_1
STK11	chr19	CTCCCAGGCAGCTGCAA	CCGGTGGTGAGCAGCAG	AMPLCHP2_STK11_2
STK11	chr19	CCGGTGGCACCCTCAAA	CTGGTCCGGCAGGTGTC	AMPLCHP2_STK11_3
STK11	chr19	AACATCACCACGGGTCTGTAC	GATGAGGCTCCCACCTTTCAG	AMPLCHP2_STK11_4
STK11	chr19	GAAGAAACATCCTCCGGCTGAA	ACCGTGAAGTCCTGAGTGTAGA	AMPLCHP2_STK11_5
TP53	chr17	TCCACTCACAGTTTCCATAGGTCT	GTTGGAAGTGTCTCATGCTGGAT	AMPLCHP2_TP53_1
TP53	chr17	GGCTGTCCCAGAATGCAAGAA	GATGAAGCTCCCAGAATGCCA	AMPLCHP2_TP53_2
TP53	chr17	TGCACAGGGCAGGTCTTG	CCGTCTTCCAGTTGCTTTATCTGT	AMPLCHP2_TP53_3
TP53	chr17	ACCAGCCCTGTCGTCTCT	GTGCAGCTGTGGGTTGATTC	AMPLCHP2_TP53_4
TP53	chr17	CCAGTTGCAAACCAGACCTCA	AGGCCTCTGATTCCTCACTGAT	AMPLCHP2_TP53_5
TP53	chr17	GGCTCCTGACCTGGAGTCTT	CTCATCTTGGGCCTGTGTTATCTC	AMPLCHP2_TP53_6
TP53	chr17	CGCTTCTTGTCCTGCTTGCT	TTCTCTTTTCCTATCCTGAGTAGTGGT	AMPLCHP2_TP53_7
TP53	chr17	GGAAGGGGCTGAGGTCACT	CCCCTCCTCTGTTGCTGC	AMPLCHP2_TP53_8
VHL	chr3	CTCCCAGGTCATCTTCTGCAAT	GTACCTCGGTAGCTGTGGATG	AMPLCHP2_VHL_1
VHL	chr3	GTGGCTCTTTAACAACCTTTGCT	GTCAGTACCTGGCAGTGTGATA	AMPLCHP2_VHL_2

VHL	chr3	GGCAAAGCCTCTTGTTCGTTC	TGACGATGTCCAGTCTCCTGTAAT	AMPLCHP2_VHL_3

11.2 Self-designed NGS Panel #2

Table 25: NGS Panel #2.

Gene Symbol	Chr	Ion AmpliSeq Fwd Primer (5'-3')	Ion AmpliSeq Rev Primer (5'-3')	Amplicon ID
ALK	chr2	TCTCTCGGAGGAAGGACTTGAG	GCCCAGACTCAGCTCAGTTAAT	CHP2_ALK_1
ALK	chr2	ACAGGGTACCAGGAGATGATGTAAG	GGAAGAGTGGCCAAGATTGGA	CHP2_ALK_2
APC	chr5	GAGAGAACGCGGAATTGGTCTA	GTATGAATGGCTGACACTTCTTCCA	CHP2_APC_1
APC	chr5	AGCACTGATGATAAACACCTCAAGTT	ATCTTCTTGACACAAAGACTGGCT	CHP2_APC_2
APC	chr5	TTCATTATCATCTTTGTCATCAGCTGAA	TTTGGTTCTAGGGTGCTGTGAC	CHP2_APC_3
APC	chr5	GCAGACTGCAGGGTTCTAGTT	GTGAACTGACAGAAGTACATCTGCT	CHP2_APC_4
APC	chr5	AGCCCCAGTGATCTTCCAGATA	CCCTCTGAACTGCAGCATTTACT	CHP2_APC_5
APC	chr5	AGAGGGTCCAGGTTCTTCCA	TCATTTTCCTGAACTGGAGGCATT	CHP2_APC_6
APC	chr5	ATGAAACAGAATCAGAGCAGCCTAAA	CGTGATGACTTTGTTGGCATGG	CHP2_APC_7
ARID1A	chr1	CAAAATGAACAACAAGGCAGATGGG	TCAGAGACTATCTAGTCCGGTGTC	ARID1A_10.112972
ARID1A	chr1	CAGCTAAACTTACTGGACTTGAGAATTTTT	GAGTCAAGACAAAAATCACTACCTTGG	ARID1A_10.135473
ARID1A	chr1	CATGATGGGAACTGGACCTCCTTA	TTAGCTGTGATGTGACTCTTGAAGAAAT	ARID1A_10.143283
ARID1A	chr1	CCCCCAGCCTACGGCTTC	CCCCGCGTAGGGCTCCA	ARID1A_1.1.15178
ARID1A	chr1	CCCTAGGCCCGCCCTGA	GGCTCCGGCCGTAGGGT	ARID1A_1.1.16654
ARID1A	chr1	CAGTCAAGAGACTTCTGAGACCCTTA	CAGATAACGGTCCACCCACATC	ARID1A_11.181180
ARID1A	chr1	CCGCTGGGAAAGGAGCTG	GCCTAGGGCCCGCGTTC	ARID1A_1.1.20289
ARID1A	chr1	CTATCGCCTCTATGTGTCTGTGAAG	GTACCACATGAAGCCAGTGAGTAC	ARID1A_11.248116
ARID1A	chr1	ACAACTCCTACTACCCCAACC	CTGCTGAGCGAAGGACGA	ARID1A_1.1.2481
ARID1A	chr1	CTTCCAGAAATCCAGTTCTTCTACTACA	ATAGAGGTCCAGAGGTTTCCTACC	ARID1A_11.279375
ARID1A	chr1	CTCAGCAGCGCTTCGGG	GGGCCCGCCACTGTAGT	ARID1A_1.1.36612
ARID1A	chr1	CTCGGAGCTGAAGAAAGCCG	GCTCTCGGCCCCGTCCT	ARID1A_1.1.38056
ARID1A	chr1	GAGCCCGTCTGCCGTCG	GGAGTTGTACTGGTGGTTGGG	ARID1A_1.1.42139
ARID1A	chr1	GGCCCCAGCAGAACTCTCAC	AGCCCGGAGTGCCACCTC	ARID1A_1.1.52554
ARID1A	chr1	GGCTGCCGGCTCCAAGC	GCTGGGCGACGTGAGCA	ARID1A_1.1.54514
ARID1A	chr1	GGGATCATGGCCGCGCA	CCGGCGGCTGCCTTCAT	ARID1A_1.1.54590
ARID1A	chr1	TTATCTGGCCTTCACTGAGGAGAA	CTCACCTGAGTCAATCCACCAAT	ARID1A_11.550938
ARID1A	chr1	AGCCGGACCTGAAGAACTCG	GGCCGCGGCTGAGTGAG	ARID1A_1.1.6484
ARID1A	chr1	CTCGCCCGGACCCCTCAG	GCCAGACAATGGCAGCTCC	ARID1A_1.2.19161
ARID1A	chr1	GGGCTACCAGGGCTACCC	GGGCTCATGGGCGCGTG	ARID1A_1.2.26067
ARID1A	chr1	GATATACCTCGACTCCTTTGGTTTGG	AGGGTCTTCTCCCCGTTCAAT	ARID1A_12.293039
ARID1A	chr1	GCCAGCTCCTTGAAAAAGCAGTATATC	GACCCCATCCTTACCAGGAGAG	ARID1A_12.311881
ARID1A	chr1	AGACATCTTTGCAGCTGCTGATT	CACAGATCCTTGGCATATCCTGTTG	ARID1A_12.73402
ARID1A	chr1	CCGGCGGACATGGCCTC	CCTCCCCACTCAGCTGTGTA	ARID1A_1.2.9363
ARID1A	chr1	CTCAACTTGTATCTCTGTCCACAGC	CTGCTCTTGGCCTTACCTCATG	ARID1A_13.224100
ARID1A	chr1	CTCCTGCGTGTCCTTTGTTATATTGG	TGGAGTCATGGAATTCCGCTT	ARID1A_13.228066
ARID1A	chr1	GAGGAGACTTAAAGCCACCAACTC	CAAGGAGTTCCCATGCACTTATCT	ARID1A_13.262576
ARID1A	chr1	GCCTTGTAGATCCTCTGCTAAGAAG	GCCCCTGCATAGATCCTGATCC	ARID1A_13.286741
ARID1A	chr1	CTTTAATGATGGAAGTGACTCCACATTC	CAAGTTCAAATAGCAATCAGATCAGTCA	ARID1A_14.234479
ARID1A	chr1	TGACTCCAAACCCTGGGTATCA	CATTTCACTGGCCCTGTCTTTACG	ARID1A_14.440936
ARID1A	chr1	GACCACGACAGCACTATCCCTA	TCATGTTTCCCTCAGGCCCTATT	ARID1A_15.209989
ARID1A	chr1	TCACCGCTTGCCTTTCTACG	TCACTCTGTCATAAGGACCTCCA	ARID1A_15.321878

ARID1A	chr1	CCAATTTTGTTTAGGACGGAGCCT	CACCGAGACCAGGCTTTACTC	ARID1A_15.99688
ARID1A	chr1	CTAATCCTGTGTTTCTTTGCCTCCT	TTTTCAAGGCGAACCTGCATG	ARID1A_16.147847
ARID1A	chr1	GGATGTATTCTCCTAGCCGCTAC	TTGGGTGGAGAACTGATTGCCATA	ARID1A_16.243588
ARID1A	chr1	AGCGTGCCATACAGCACT	GGCAGTGGCAGGATAGGCA	ARID1A_18.122838
ARID1A	chr1	AACCGCACCTCTCCTAGC	TCCCGCCGAATCATGGG	ARID1A_18.17117
ARID1A	chr1	CAGATGAAATGCTGCACACAGATC	GATACCTGAGGAATGTGATTCTGCAT	ARID1A_18.249269
ARID1A	chr1	CAGGTATCCAGCCCTGCTC	TGCTATGTGCGAGGCAGGT	ARID1A_18.260793
ARID1A	chr1	CCACTGCCACAGCTGCTAC	GCTGAGCAACCTCAGCTGAT	ARID1A_18.303487
ARID1A	chr1	AAGGCTCGTGGCCTTCCC	GTGCGGTTCTCCATTGGC	ARID1A_18.33212
ARID1A	chr1	CTGTGTCCACCAAGCATCTGG	GGCACGCTGTACATCTCC	ARID1A_18.457891
ARID1A	chr1	GCAAAACATGCCACCACAAATGATG	TGTTCGGTTCACGCCATGATAG	ARID1A_18.536845
ARID1A	chr1	GCCTTCCCCTCAGCAAGATGTATA	GGTCTCGGCCAAACTGGAATG	ARID1A_18.584475
ARID1A	chr1	ACATAGCACCTGCCCCTGT	GGGCAGATTAGGCAACCGAATG	ARID1A_18.63843
ARID1A	chr1	TGCTCAGCAAGGCACCATG	CGAGCCTTCGTGGTTGG	ARID1A_18.820768
ARID1A	chr1	TTGTCTCTGCCTTAGAATTACAAGCG	GCTGGGCAGCTTGTTGCT	ARID1A_18.880618
ARID1A	chr1	AGACGACATGGAGGTTTATTTCAGG	CCCCAGGCACTGATACTCA	ARID1A_19.54023
ARID1A	chr1	ATCTTCAGAGTAGCTTCACTGATGGG	GTTGATGGTATCTAATGCCCATGTG	ARID1A_19.79292
ARID1A	chr1	CAACATCCTGCTGTATGATGACAAC	GGCATGGAAGATATCTACAAGAGAGAAA	ARID1A_19.96133
ARID1A	chr1	ATCCTGGGAGGTTCAGCAA	GAAGCTGGCTTGTCCTTGC	ARID1A_20.1.247050
ARID1A	chr1	CCAGCTTCAGAGAATAGTGAGGA	GCAGCAGGCCACTGTCAAA	ARID1A_20.1.351511
ARID1A	chr1	CCCTCGGAAGCATGTGACAAC	CTTGATGGCCTCTGAACTCTTAGC	ARID1A_20.1.374770
ARID1A	chr1	CCTGCTGCACTGGCGGAT	GGCCCCTCCTGGTCTGTTG	ARID1A_20.1.397870
ARID1A	chr1	CGAAGCCTGTCATTTGTGCCA	GTTAGTGGTGCCTGCTTCCG	ARID1A_20.1.400147
ARID1A	chr1	CTGTTCTTAGGCCACTTTTCTCC	CCCAGGATCCAGTAGCGTT	ARID1A_20.1.464959
ARID1A	chr1	GAGGAAGTAGTTGAAAATGATGAGGAGA	TCCACCACAAATGGATCATTCTTCTGTA	ARID1A_20.1.565326
ARID1A	chr1	GCAGCAAGTTTCCATTTGGCATTAG	AGGCTTCGAATGGTATTGGACAC	ARID1A_20.1.612574
ARID1A	chr1	GCCTGATTGAGATCTTTGGCATTTTAAA	ACTTCCTCTTCTTCTTCCTCTTCTAGTTTA	ARID1A_20.1.637352
ARID1A	chr1	GGGCCCCCACCTGATGGA	GTTCCGGTGGCTCTGTGC	ARID1A_20.1.733579
ARID1A	chr1	GTGGTGGACTGCTCAGATAAGCT	AGCTCTGTCTTGCTCTCGAAGT	ARID1A_20.1.787955
ARID1A	chr1	ACCGGAACATCAAGATCCTAGAG	CCCTGGGTGTTTGGACATC	ARID1A_20.1.90003
ARID1A	chr1	ATGGTGCGCTTCCTCAGT	CGGCAAGGCTGTCCTCTAG	ARID1A_20.2.152846
ARID1A	chr1	CAGTGCAGAAGGGCAGTATCG	CCGCATCATGTCCACACTAGTTG	ARID1A_20.2.222602
ARID1A	chr1	CCACTAACTTATGAAAAGGAGGAGGAA	CCCGAGATGTTGGCGAGTGTA	ARID1A_20.2.247152

ARID1A	chr1	CCTTGCCGCCACACAGTT	CAACAGCCGTGATTCGTACA	ARID1A_20.2.313338
ARID1A	chr1	CTTGGAGATGCTCCGGGAA	AGGGCAAACTGCCCAGTGTA	ARID1A_20.2.409280
ARID1A	chr1	CTTCAGCTGAAGCCCAGGAC	CGCACCATAGTGCTATACAACTTCT	ARID1A_20.2.432258
ARID1A	chr1	AAACTCAGCATCCAGGACAACAAT	GCCAGGTTGGCCAGCAGTA	ARID1A_20.2.5295
ARID1A	chr1	ACCCAGGGCTGCTGCTCAT	TCTCCAAGCAGTCCCACCA	ARID1A_20.2.54967
ARID1A	chr1	GGCTGTTGGACATCTCGGT	GTTTTGCATAAATAAAGGGCAACAGTC	ARID1A_20.2.603909
ARID1A	chr1	GGGCAGTTGGACCTATCTCCATAC	CTGAGTTTGCTGAGGGTTTCCAA	ARID1A_20.2.613243
ARID1A	chr1	TGGACGAGAACCACTCAGAGTTTAC	GCTGTCATGACTGGCCAATCAAAA	ARID1A_20.2.760156
ARID1A	chr1	CATGGGCGGCCTCTCTTATAC	TAGTAGCACTCTGTAATTAACTGAGCCA	ARID1A_2.198073
ARID1A	chr1	CTAACCCATACTCGCAGCAACA	TCACAATCACCATCTACCTGCTG	ARID1A_2.263887
ARID1A	chr1	GCCATCCAGTCCAATGGATCAG	CCTGCATGGTCATCGGGTAC	ARID1A_2.310812
ARID1A	chr1	AAACCTGTGTACTTGGGTTATATATTCAGT	CCATATGGCTGAGGTCTCATCTTG	ARID1A_2.6808
ARID1A	chr1	AGTCCCAGCAAACTGCCTATTC	ACCCAGAGTTTAATTGGTCTTTAAGTG	ARID1A_3.115759
ARID1A	chr1	CAGCAAAGTCCTCACCCTCAG	GGGATGGCTGCTGGGAGTAT	ARID1A_3.189854
ARID1A	chr1	CAGCCTCCACATCAGCAGTC	AGCCTGCTGGGAGAGCGT	ARID1A_3.203145
ARID1A	chr1	CAGGCTCAGTCTCCTTACCA	GCAGGAGGCAGGGATATCTT	ARID1A_3.210697
ARID1A	chr1	TGCTTTCTATACTCATCATCAGTGCAT	CTTTGCTGGTTGTAATATGGAGTCTG	ARID1A_3.663071
ARID1A	chr1	TTTTCCTTTCCTACAGATTCCTCCTT	CTGCTGCTGATACGAAGGTTG	ARID1A_3.701387
ARID1A	chr1	AGCAGCAGCCACAGTCTCAA	TGAGCCTGTGGCTGTGAGTA	ARID1A_3.77267
ARID1A	chr1	CCATCACAGCTTTTGTTTTTCTTGTTGTAG	ACCTTTCAGAAGGTGCAGAAATACT	ARID1A_4.173246
ARID1A	chr1	CTGGCCTTCACATAATACTTTTCGC	GATGCCTGAGACCCAAATGAATC	ARID1A_4.215850
ARID1A	chr1	GAGGGCAAGAAGATATGAACCTGAG	AGGTCAAAATTAGCTAAACTTCCAACCA	ARID1A_4.255932
ARID1A	chr1	GAGTCCTGGAGTGAGCACATC	CGAGAGTGGTCCTGAGCGA	ARID1A_5.220317
ARID1A	chr1	AGAATCTTTCTGCCTAATATTACTAATCCATG	AGGAGACTGAGCTGGATTACTCT	ARID1A_5.34011
ARID1A	chr1	AGTCTCCTTTCTCTCCTCATACCT	CAGTCACCTTTCCCTCTCCCTAA	ARID1A_5.70816
ARID1A	chr1	GATATGCTTATGTTGTTCTTTGTCTGGA	CACTCAAATGTCTGCCCTAGCTC	ARID1A_6.227771
ARID1A	chr1	AGCCATTTCTAGCTCTGAATTAACTTCC	TCGATCTTGGGCAATGCTTGAT	ARID1A_6.45476
ARID1A	chr1	CAGCCTTATCTCCGCGTCAG	ACTGTTTTCTCCTCTCACCCGTAT	ARID1A_7.149596
ARID1A	chr1	CAGGATAAGGATGGAGAGCATTTGTTC	TGTGTGTATCTGTCCTCCGGAA	ARID1A_7.152412
ARID1A	chr1	CATGGCCAATATGCCACCTCA	ATAATACATTTTCTTGCACTGACACCCT	ARID1A_8.210031
ARID1A	chr1	CCAATGCCAACTACCCCAGTG	GGCCATGTTAGGGCCATAAGG	ARID1A_8.229785
ARID1A	chr1	GTTGCTAGTGAGTGACTAACCAAGTC	GGCTGTCCATGCATTTGACCTC	ARID1A_8.517555
ARID1A	chr1	AGGATGAGTCACGCCTCCATG	GGCCTTACCTGTTTTGGATAGAGTTG	ARID1A_8.91537
ARID1A	chr1	AGCACTATTTGGCTCCAGTTCAAATC	GGTTGATCATGCCAGCCATACTATTAA	ARID1A_9.65769
BRAF	chr7	CATACTTACCATGCCACTTTCCCTT	TTTCTTTTTCTGTTTGGCTTGACTTGA	CHP2_BRAF_1
BRAF	chr7	CCACAAAATGGATCCAGACAACTGT	GCTTGCTCTGATAGGAAAATGAGATCTA	CHP2_BRAF_2
CDKN2A	chr9	CACCAGCGTGTCCAGGAA	CCCTGGCTCTGACCATTCTGT	CHP2_CDKN2A_1
CDKN2A	chr9	CATCTATGCGGGCATGGTTACT	CGCTGGTGGTGCTG	CHP2_CDKN2A_2

CTNNB1	chr3	ACTGTTTCGTATTTATAGCTGATTTGATGGA	CCTCTTCCTCAGGATTGCCTTT	CHP2_CTNNB1_1
EGFR	chr7	CCTCATTGCCCTCAACACAGT	TCAGTCCGGTTTTATTTGCATCATAGTT	CHP2_EGFR_1
EGFR	chr7	CACCACGTACCAGATGGATGT	CCCAAAGACTCTCCAAGATGGGATA	CHP2_EGFR_2
EGFR	chr7	AGACATGCATGAACATTTTTCTCCAC	TCCAGACCAGGGTGTTGTTTTC	CHP2_EGFR_3
EGFR	chr7	TGTGGAGCCTCTTACACCCA	GTGCCAGGGACCTTACCTTATAC	CHP2_EGFR_4
EGFR	chr7	ACGTCTTCCTTCTCTCTCTGTCA	CTGAGGTTCAGAGCCATGGA	CHP2_EGFR_5
EGFR	chr7	CATGCGAAGCCACACTGAC	ACATAGTCCAGGAGGCA	CHP2_EGFR_6
EGFR	chr7	GACTATGTCCGGGAACACAAAGA	CCCCATGGCAAACTCTTGCTA	CHP2_EGFR_7
EGFR	chr7	CGCAGCATGTCAAGATCACAGAT	GCATGTGTTAAACAATACAGCTAGTG	CHP2_EGFR_8
FBXW7	chr4	TGACAATGTTTAAAGGTGGTAGCTGTT	ACTCATTGATAGTTGTGAACCAACACA	CHP2_FBXW7_1
FBXW7	chr4	CCTGTGACTGCTGACCAAACTTTTA	CACATCTTTCTTATAGGTGCTGAAAGG	CHP2_FBXW7_2
FBXW7	chr4	CCCAACCATGACAAGATTTTCCC	GGTCATCACAAATGAGAGACAACATCA	CHP2_FBXW7_3
FBXW7	chr4	ACTAACAACCCTCCTGCCATCATA	TCTGCAGAGTTGTTAGCGGTT	CHP2_FBXW7_4
FBXW7	chr4	GTAGAATCTGCATTCCCAGAGACAA	TCTCTTGATACATCAATCCGTGTTTGG	CHP2_FBXW7_5
FGFR2	chr10	CATCACTGTAAACCTTGCAGACAAAC	TGGTCTCTCATTCTCCCATCCC	CHP2_FGFR2_1
FGFR2	chr10	CATCCTCTCTCAACTCCAACAGG	AGTGGATCAAGCACGTGGAAAA	CHP2_FGFR2_2
FGFR2	chr10	GCTTCTTGGTCGTGTTCTTCATT	CTCCTCCTGTGATCTGCAATCT	CHP2_FGFR2_3
FGFR2	chr10	TGGAAGCCCAGCCATTTCTAAA	GATGATGAAGATGATTGGGAAACACAAG	CHP2_FGFR2_4
GNAS	chr20	TTGGTGAGATCCATTGACCTCAATTT	TGAATGTCAAGAAACCATGATCTCTGTT	CHP2_GNAS_1
GNAS	chr20	CCTCTGGAATAACCAGCTGTCC	TGATCCCTAACAACACAGAAGCAA	CHP2_GNAS_2
IDH1	chr2	CCAACATGACTTACTTGATCCCCAT	ATCACCAAATGGCACCATACGA	CHP2_IDH1_1
IDH2	chr15	ACCCTGGCCTACCTGGTC	AGTTCAAGCTGAAGAAGATGTGGAA	CHP2_IDH2_1
KRAS	chr12	CAAAGAATGGTCCTGCACCAGTAATAT	AGGCCTGCTGAAAATGACTGAATATAA	CHP2_KRAS_1
KRAS	chr12	TCCTCATGTACTGGTCCCTCATT	GTAAAAGGTGCACTGTAATAATCCAGACT	CHP2_KRAS_2
KRAS	chr12	CAGATCTGTATTTATTTCAGTGTTACTTACCT	GACTCTGAAGATGTACCTATGGTCCTA	CHP2_KRAS_3
NRAS	chr1	CCTCACCTCTATGGTGGGATCATAT	GTTCTTGCTGGTGTGAAATGACTG	CHP2_NRAS_1
NRAS	chr1	TTCGCCTGTCCTCATGTATTGG	CACCCCCAGGATTCTTACAGAAAA	CHP2_NRAS_2
NRAS	chr1	GCACAAATGCTGAAAGCTGTACC	CAAGTGTGATTTGCCAACAAGGA	CHP2_NRAS_3
PIK3CA	chr3	CCATAAAGCATGAACTATTTAAAGAAGCAAGA	GGTTGAAAAAGCCGAAGGTCAC	CHP2_PIK3CA_1
PIK3CA	chr3	TGGAATGCCAGAACTACAATCTTTTGAT	AAGATCCAATCCATTTTTGTTGTC	CHP2_PIK3CA_10
PIK3CA	chr3	TGGATCTTCCACACAATTAAACAGCAT	TGCTGTTCATGGATTGTGCAATTC	CHP2_PIK3CA_11
PIK3CA	chr3	CCCTTTTTAAAAGTAATTGAACCAGTAGGC	TTTAAGATTACGAAGGTATTGGTTTAGACAGAA	CHP2_PIK3CA_2
PIK3CA	chr3	GACGCATTTCCACAGCTACAC	AGCATCAGCATTTGACTTTACCTTATCA	CHP2_PIK3CA_3
PIK3CA	chr3	CATAGGTGGAATGAATGGCTGAATTATG	TCAATCAGCGGTATAATCAGGAGTTTTT	CHP2_PIK3CA_4
PIK3CA	chr3	TCCCATTATTATAGAGATGATTGTTGAATTTTCCT	CAAACAAGTTTATATTTCCCCATGCCA	CHP2_PIK3CA_5
PIK3CA	chr3	GCTTTGAATCTTTGGCCAGTACCT	CATAAGAGAGAAGGTTTGACTGCCATA	CHP2_PIK3CA_6
PIK3CA	chr3	CAGAGTAACAGACTAGCTAGAGACAATGA	GCACTTACCTGTGACTCCATAGAAA	CHP2_PIK3CA_7
PIK3CA	chr3	CACGATTCTTTTAGATCTGAGATGCACA	CCTTTTGTGTTTCATCCTTCTTCTCCTG	CHP2_PIK3CA_8
PIK3CA	chr3	GATGCAGCCATTGACCTGTTTAC	AGAAAACCATTACTTGTCCATCGTCT	CHP2_PIK3CA_9
PTEN	chr10	GCCATCTCTCTCCTCCTTTTTCTT	GCCGCAGAAATGGATACAGGTC	CHP2_PTEN_1
PTEN	chr10	TGTTAATGGTGGCTTTTTGTTTGTTTGT	TCTACCTCACTCTAACAAGCAGATAACT	CHP2_PTEN_2
PTEN	chr10	CCATAACCCACCACAGCTAGAA	TGCCCCGATGTAATAAATATGCACAT	CHP2_PTEN_3
PTEN	chr10	GGCTACGACCCAGTTACCATAG	TGCCACTGGTCTATAATCCAGATGAT	CHP2_PTEN_4

PTEN	chr10	TGAGATCAAGATTGCAGATACAGAATCC	ACCTTTAGCTGGCAGACCAC	CHP2_PTEN_5
PTEN	chr10	AGGTGAAGATATATTCCTCCAATTCAGGAC	TTGGATATTTCTCCCAATGAAAGTAAAGTAC	CHP2_PTEN_6
PTEN	chr10	CACTTTTGGGTAAATACATTCTTCATACCAGGA	TATACTGCAAATGCTATCGA	CHP2_PTEN_7
PTEN	chr10	GCAGTATAGAGCGTGCAGATAATGA	CATCACATACATACAAGTCAACAACCC	CHP2_PTEN_8
RNF43	chr17	CCAAACACATCTGGAGCACACT	GCCTGACCCTCAATGACCTCTT	RNF43_1.100174
RNF43	chr17	CCGCTTTTTGTAGTGGTGGT	TGACTTTGACCCCCTAGTGTACT	RNF43_2.1.213215
RNF43	chr17	ACAACCACACTGGCTGTGAA	GCACCCAGCTTGCCAGATT	RNF43_2.1.22754
RNF43	chr17	CGGTGTCAGAACTCCATTCAGAAG	GACAAGAGGCTGCTACCAGAAA	RNF43_2.1.264668
RNF43	chr17	CTCTCCCTACCACACCCACTT	GTGGTTGTGCCTGACTCCTC	RNF43_2.1.277654
RNF43	chr17	CTGGGTGCACAGTTGCATC	CCCTGGCCCAGTTGACG	RNF43_2.1.308947
RNF43	chr17	GAAACCTGGGTTTCCCCTGT	GGGTCCATGGCAGCAGTTC	RNF43_2.1.342797
RNF43	chr17	AAAGTCACTGCTTAGGGAGCT	AGAAAGCTATTGCACAGAACGC	RNF43_2.1.4249
RNF43	chr17	GGGACCAAGGATATGCCACACT	TGCAAAAATCCAGCCTCTCTGC	RNF43_2.1.479773
RNF43	chr17	GGGCACTGTGGGTTAGAGAG	AAAAGCGGTTCCAGTGGCA	RNF43_2.1.483268
RNF43	chr17	GTGACTTGCTGATCAGGAGAAGGT	GTTTCCAGCCATGTCCACTACC	RNF43_2.1.554471
RNF43	chr17	TTTTTGCAAGTTGAACAGACTGCT	CAAGTCACCAGATCCAACTCAGC	RNF43_2.1.716452
RNF43	chr17	CTCCAGATCCACTGCTGTCA	TTCCCCAGAGCTGCACATC	RNF43_2.2.168519
RNF43	chr17	GTAGGCTGATGTCCGTGCAG	GCTTGCCCAGTGCCCCTA	RNF43_2.2.335644
RNF43	chr17	GTGCTGTGAGGTGGATTGGAG	CCCACGACCTGGTCCCTT	RNF43_2.2.348341
RNF43	chr17	GTGATGCCGAGGGCCCAT	CAGGTCGAAGACTCCACCTC	RNF43_2.2.359855
RNF43	chr17	AGGTGGTAGTGGGCATGGC	TGTCTTTCTGAATGCATTCTCTGTAGG	RNF43_2.2.62203
RNF43	chr17	CCTCCTACCTGTGATGTTGAACATG	CCTGATTCCTGGCAATTCCTATGG	RNF43_3.144168
RNF43	chr17	AAGCCACATTCTAGACCTGTCTG	CTCTTTTTCTCCAGGAGCTACGG	RNF43_3.9926
RNF43	chr17	TCCTCCAGACAGATGGCACA	CCCAATCTGAGCCCCATTCCT	RNF43_4.332752
RNF43	chr17	TTCAATCTCCCCAGTCTGGTCAT	AGCTGGCCACCAGGAGGTA	RNF43_4.381754
RNF43	chr17	TACTCCTTCCTTCTCCCTAACCAC	ATGATGTGTGGATCCTAATGACAGT	RNF43_5.252108
RNF43	chr17	AAGCCAGGATGATCACAAAGATGG	CTCAAGGGAACCTCCAGTTAGCTAT	RNF43_5.9080
RNF43	chr17	CCCTGAGAGCTTTATCTTCCTCCATC	GACCTCAGCCCAACCTCTACT	RNF43_6.126819
RNF43	chr17	GTCTGGAGGTCTAGTGTGCT	TGGGCACTTTCCCCCTGTA	RNF43_6.251138
RNF43	chr17	ATCAGCTTCTCAGCGTCATTACC	CTGGATGGAGGAAGATAAAGCTCTCA	RNF43_6.62210
RNF43	chr17	CACAGGACAAAGTAGGGCTAAGTG	AAGCTGATGGAGTTTGTGTACAAGAA	RNF43_6.83735
RNF43	chr17	CCAGCTTGACGATGCTGATGAAT	TGGTACCTCCCTAGAAAAATGGAGAG	RNF43_8.145848
RNF43	chr17	GTGTAGGGCGAAGTGTGAGTC	CCTAACCCAAGTCTGTCTCTCTCTG	RNF43_8.334246
RNF43	chr17	ATTTCCACTTCTCTCAGACCAGTCAT	CCTGTCACTGGCTAGCAAGGTA	RNF43_8.99948
RNF43	chr17	GCAAACACACCTTCCAAAGTGAGATT	TGGACGCACAGGACTGGTAC	RNF43_9.251835
RNF43	chr17	ACAAAAGAAGAAAGACATATTTCAAACAGATG	TTATCAGAGTGATCCCCTTGAAAATGG	RNF43_9.43347
RNF43	chr17	AGCTTTCTGTTCTGCTGATCTTTCA	GTATGTATGGTTGAAGTGCATTGCTG	RNF43_9.83487
SMAD4	chr18	CTCATGTGATCTATGCCCGTCT	AGTCTACTTACCAATTCCAGGTGATACA	CHP2_SMAD4_1
SMAD4	chr18	TGCTACTTCTGAATTGAAATGGTTCA	GATTACCTACCATTACTCTGCAGTGTT	CHP2_SMAD4_2
SMAD4	chr18	ATGGTGAAGGATGAATATGTGCATGA	GCTGGTAGCATTAGACTCAGATGG	CHP2_SMAD4_3
SMAD4	chr18	GTGAAGGACTGTTGCAGATAGCAT	AAGGCCCACATGGGTTAATTTG	CHP2_SMAD4_4
SMAD4	chr18	TTTCTTTAGGGCCTGTTCACAATGA	CTGAGAAGTGACCCCATAATTCCATT	CHP2_SMAD4_5
SMAD4	chr18	GCTCCTGAGTATTGGTGTTCCAT	CCTGTGGACATTGGAGAGTTGA	CHP2_SMAD4_6
SMAD4	chr18	TGTAATTTCTTTTTTCTTCCTAAGGTTGCACATAG	ACTTGGGTAGATCTTATGAACAGCAT	CHP2_SMAD4_7
SMAD4	chr18	AGGTCTTTGATTTGCGTCAGTGT	GCTGGAGCTATTCCACCTACTG	CHP2_SMAD4_8

SMAD4	chr18	GCTGCTGGAATTGGTGTTGATG	AGTACTTCGTCTAGGAGCTGGAG	CHP2_SMAD4_9
STK11	chr19	GAGCTGATGTCGGTGGGTAT	CTCCGAGTCCAGCACCTC	CHP2_STK11_1
STK11	chr19	CTCCCAGGCAGCTGCAA	CCGGTGGTGAGCAGCAG	CHP2_STK11_2
STK11	chr19	CCGGTGGCACCCTCAAA	CTGGTCCGGCAGGTGTC	CHP2_STK11_3
STK11	chr19	AACATCACCACGGGTCTGTAC	GATGAGGCTCCCACCTTTCAG	CHP2_STK11_4
STK11	chr19	GAAGAAACATCCTCCGGCTGAA	ACCGTGAAGTCCTGAGTGTAGA	CHP2_STK11_5
TP53	chr17	TCCACTCACAGTTTCCATAGGTCT	GTTGGAAGTGTCTCATGCTGGAT	CHP2_TP53_1
TP53	chr17	GGCTGTCCCAGAATGCAAGAA	GATGAAGCTCCCAGAATGCCA	CHP2_TP53_2
TP53	chr17	TGCACAGGGCAGGTCTTG	CCGTCTTCCAGTTGCTTTATCTGT	CHP2_TP53_3
TP53	chr17	ACCAGCCCTGTCGTCTCT	GTGCAGCTGTGGGTTGATTC	CHP2_TP53_4
TP53	chr17	CCAGTTGCAAACCAGACCTCA	AGGCCTCTGATTCCTCACTGAT	CHP2_TP53_5
TP53	chr17	GGCTCCTGACCTGGAGTCTT	CTCATCTTGGGCCTGTGTTATCTC	CHP2_TP53_6
TP53	chr17	CGCTTCTTGTCCTGCTTGCT	TTCTCTTTTCCTATCCTGAGTAGTGGT	CHP2_TP53_7
TP53	chr17	GGAAGGGGCTGAGGTCACT	CCCCTCCTCTGTTGCTGC	CHP2_TP53_8
VHL	chr3	CTCCCAGGTCATCTTCTGCAAT	GTACCTCGGTAGCTGTGGATG	CHP2_VHL_1
VHL	chr3	GTGGCTCTTTAACAACCTTTGCT	GTCAGTACCTGGCAGTGTGATA	CHP2_VHL_2
VHL	chr3	GGCAAAGCCTCTTGTTCGTTC	TGACGATGTCCAGTCTCCTGTAAT	CHP2_VHL_3

11.3 Enrichment Analysis

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Drug metabolism	0.001276	0.2169	5.38	35.88
2	Cell adhesion molecules	0.02443	0.9195	3.18	11.80
3	African trypanosomiasis	0.06289	0.9195	5.15	14.25
4	Lipid and atherosclerosis	0.09065	0.9195	2.16	5.18
5	Carbohydrate digestion and absorption	0.09512	0.9195	4.00	9.42
6	Chagas disease	0.1036	0.9195	2.74	6.20
7	Sulfur metabolism	0.1052	0.9195	9.98	22.49
8	Malaria	0.1055	0.9195	3.75	8.44
9	Endocrine and other factor-regulated calcium reabsorption	0.1163	0.9195	3.53	7.60
10	Regulation of lipolysis in adipocytes	0.1235	0.9195	3.40	7.11

Figure 43: Enrichment analysis with the webtool enrichr.
KEGG 2021 Human Pathway database was used to investigate 221 genes on chromosome 1 in gastric IPMN. These regions were affected multiple.

11.4 Integrative Analysis gene list

Table 26: 84 genes from the Integrative analysis between gastric IPMN and PanIN.

Gene_id	a.value	m.value	p.value	q.value	Expression_change	Methylation_change
ABHD17C	8.406836	-1.905625	0.000414	0.025095	downregulated	hypermethylated
ADAMTS8	6.882026	2.138768	0.000151	0.011398	upregulated	hypermethylated
AKAP7	7.810693	3.452716	0.000000	0.000006	upregulated	hypomethylated
AKR1B10	9.538986	-2.735876	0.000201	0.014379	downregulated	hypermethylated
ALDH3A1	9.231040	-3.359306	0.000000	0.000006	downregulated	hypermethylated
ANKRD22	10.130504	-2.574107	0.000005	0.000663	downregulated	hypermethylated
ANKRD62	-3.246601	7.417131	0.000000	0.000051	upregulated	hypermethylated
ANO7	9.924905	-3.576798	0.000000	0.000000	downregulated	hypermethylated
ANXA10	10.323523	-2.795399	0.000000	0.000014	downregulated	hypermethylated
AQP1	11.001677	2.530419	0.000996	0.049105	upregulated	hypomethylated
ARID5A	7.776322	2.676413	0.000000	0.000018	upregulated	hypermethylated
ASCL2	-3.246601	7.604881	0.000005	0.000704	upregulated	hypomethylated
B4GALNT3	9.909873	-1.843356	0.000000	0.000003	downregulated	hypermethylated
BAIAP2L1	10.334025	-1.497507	0.000611	0.034304	downregulated	hypermethylated
BCAR3	7.250907	-1.652941	0.000149	0.011312	downregulated	hypermethylated
C1orf94	-3.246601	-7.610273	0.000000	0.000001	downregulated	hypomethylated
C1QTNF1	7.585663	3.798318	0.000000	0.000000	upregulated	hypermethylated
C5orf38	-3.246601	7.520667	0.000629	0.035064	upregulated	hypomethylated
C6	9.052273	2.797611	0.000078	0.006737	upregulated	hypermethylated
CAPN8	12.116386	-3.018237	0.000000	0.000000	downregulated	hypermethylated
CAPN9	9.325554	-4.305547	0.000000	0.000000	downregulated	hypermethylated
CASZ1	8.944525	-1.586625	0.000015	0.001790	downregulated	hypermethylated
CDC25B	8.535133	1.972950	0.000009	0.001152	upregulated	hypermethylated
CDHR2	8.145551	-3.915049	0.000144	0.010999	downregulated	hypermethylated
CLTB	11.649364	-1.643880	0.000001	0.000117	downregulated	hypermethylated
COL26A1	-3.246601	7.218558	0.000034	0.003508	upregulated	hypermethylated
CRMP1	-3.246601	7.841671	0.000485	0.028538	upregulated	hypomethylated
EPS8L3	10.203967	-1.855359	0.000006	0.000816	downregulated	hypermethylated
ERN2	11.322607	-2.761564	0.000000	0.000081	downregulated	hypomethylated
FAM169A	-3.246601	7.173746	0.000004	0.000593	upregulated	hypermethylated
FAM177B	8.849371	-4.166963	0.000000	0.000001	downregulated	hypermethylated
FER1L4	10.871968	-3.110066	0.000000	0.000000	downregulated	hypomethylated
GALNT5	8.557403	-1.789186	0.000083	0.007079	downregulated	hypermethylated
GALNT7	8.569040	-1.279943	0.000009	0.001160	downregulated	hypermethylated
GALNTL6	4.442272	-5.618840	0.000013	0.001581	downregulated	hypomethylated
GNG7	7.348612	1.442794	0.000012	0.001452	upregulated	hypermethylated
GPM6A	-3.246601	6.789297	0.000052	0.004940	upregulated	hypermethylated
HTR2B	-3.246601	7.139080	0.000993	0.049022	upregulated	hypermethylated
ICAM2	8.546207	-2.032928	0.000003	0.000476	downregulated	hypermethylated
ITPKA	7.390167	-2.223571	0.000092	0.007543	downregulated	hypermethylated

KCNN4	9.927649	-2.541290	0.000051	0.004839	downregulated	hypermethylated
KLF4	9.684930	-2.035749	0.000168	0.012439	downregulated	hypermethylated
KPNA7	5.713778	-3.978336	0.000002	0.000309	downregulated	hypermethylated
LAMA3	9.374587	-1.565132	0.000494	0.028929	downregulated	hypermethylated
LINC00504	6.308426	6.119232	0.000550	0.031390	upregulated	hypermethylated
LOC100506142	-3.246601	6.294329	0.000558	0.031722	upregulated	hypermethylated
LOC101928100	-3.246601	7.203176	0.000011	0.001369	upregulated	hypermethylated
LRRC31	6.248793	-4.254222	0.000000	0.000064	downregulated	hypermethylated
MAP2	7.772136	1.750357	0.000003	0.000478	upregulated	hypermethylated
MARCH3	7.311274	1.835402	0.000338	0.021332	upregulated	hypermethylated
MGLL	11.079083	-1.369073	0.000213	0.015041	downregulated	hypermethylated
MLPH	11.896581	-2.517708	0.000000	0.000000	downregulated	hypermethylated
MUC6	15.766144	-3.186602	0.000000	0.000001	downregulated	hypermethylated
MYLK	8.598520	1.647660	0.000024	0.002680	upregulated	hypermethylated
MYO5B	9.407740	-1.375066	0.000710	0.038491	downregulated	hypermethylated
NFIA	9.494979	2.020784	0.000637	0.035329	upregulated	hypermethylated
NR2F6	10.512021	-1.076430	0.000002	0.000376	downregulated	hypermethylated
NRN1	-3.246601	7.766951	0.000169	0.012510	upregulated	hypomethylated
PDE4D	8.682429	-1.462608	0.000240	0.016583	downregulated	hypermethylated
PGC	15.462795	-3.673645	0.000060	0.005533	downregulated	hypermethylated
PI3	-3.246601	7.925157	0.000123	0.009648	upregulated	hypermethylated
PIP5K1B	6.927459	-2.305613	0.000000	0.000001	downregulated	hypomethylated
PRR4	8.806660	-5.297902	0.000073	0.006401	downregulated	hypermethylated
PYGB	11.130636	-1.287239	0.000048	0.004616	downregulated	hypermethylated
REP15	7.279397	-4.944814	0.000895	0.045538	downregulated	hypermethylated
S100P	12.357992	-3.211363	0.000000	0.000000	downregulated	hypermethylated
SH3BP4	9.167832	1.674628	0.000016	0.001879	upregulated	hypermethylated
SH3RF2	9.094173	-1.178038	0.000168	0.012439	downregulated	hypermethylated
SHISA2	-3.246601	7.525944	0.000887	0.045331	upregulated	hypermethylated
SLC44A4	10.955053	-1.753949	0.000058	0.005371	downregulated	hypomethylated
SLC45A3	9.333830	-1.659913	0.000105	0.008389	downregulated	hypermethylated
SULT1C2	10.742830	-2.656306	0.000000	0.000000	downregulated	hypermethylated
SV2B	6.732693	1.882645	0.000713	0.038623	upregulated	hypomethylated
TBC1D14	7.634856	-2.691930	0.000025	0.002715	downregulated	hypermethylated
TFCP2L1	9.366806	-1.930915	0.000960	0.047833	downregulated	hypermethylated
TFF1	14.586772	-3.025607	0.000001	0.000171	downregulated	hypermethylated
TMEM38A	7.172684	-1.799443	0.000140	0.010766	downregulated	hypermethylated
TMEM63A	11.511452	-1.297079	0.000255	0.017316	downregulated	hypermethylated
TNNT2	-3.246601	7.347625	0.000018	0.002054	upregulated	hypermethylated
TSKU	7.651260	2.651734	0.000026	0.002780	upregulated	hypermethylated
TSPAN1	11.237183	-2.199998	0.000000	0.000015	downregulated	hypermethylated
UBE3B	8.028559	1.299038	0.000671	0.036856	upregulated	hypermethylated
VSNL1	-3.246601	6.976913	0.000006	0.000805	upregulated	hypermethylated
VTCN1	8.012627	2.648326	0.000016	0.001879	upregulated	hypomethylated