Molecular heterogeneity and commonalities in pancreatic cancer precursors with gastric and intestinal phenotype

Sven-Thorsten Liffers, Laura Godfrey, Lisa Frohn, Lena Haeberle, Aslihan Yavas, Rita Vesce, Wolfgang Goering, Friederike Opitz, Nikolas Stoecklein, Wolfram Trudo Knoefel, Anna Melissa Schlitter, Guenter Klöppel, Elisa Espinet, Andreas Trumpp, Jens T Siveke, Irene Esposito

Article - Version of Record

OPEN ACCESS

Original research

Molecular heterogeneity and commonalities in pancreatic cancer precursors with gastric and intestinal phenotype

Sven-Thorsten Liffers,[1,2] Laura Godfrey,[1,2] Lisa Frohn,[3] Lena Haeberle,[3] Aslihan Yavas ,[3] Rita Vesce,[3] Wolfgang Goering,[3] Friederike V Opitz ,[3] Nickolas Stoecklein,[4] Wolfram Trudo Knoefel,[4] Anna Melissa Schlitter ,[5] Guenter Klöppel,[5] Elisa Espinet ,[6,7,8] Andreas Trumpp,[6,7,8] Jens T Siveke,[1,2] Irene Esposito [3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-326550).

For numbered affiliations see end of article.

Correspondence to
Professor Irene Esposito, Institute of Pathology, Heinrich-Heine-University Dusseldorf and University Hospital of Dusseldorf, Dusseldorf 40225, Germany;
Irene.Esposito@med.uni-duesseldorf.de

S-TL and LG are joint first authors.
JTS and IE are joint senior authors.

Received 12 November 2021
Accepted 31 July 2022
Published Online First
9 August 2022

To cite: Liffers S-T, Godfrey L, Frohn L, *et al*. *Gut* 2023;**72**:522–534.

ABSTRACT

Objective Due to the limited number of modifiable risk factors, secondary prevention strategies based on early diagnosis represent the preferred route to improve the prognosis of pancreatic ductal adenocarcinoma (PDAC). Here, we provide a comparative morphogenetic analysis of PDAC precursors aiming at dissecting the process of carcinogenesis and tackling the heterogeneity of preinvasive lesions.

Design Targeted and whole-genome low-coverage sequencing, genome-wide methylation and transcriptome analyses were applied on a final collective of 122 morphologically well-characterised low-grade and high-grade PDAC precursors, including intestinal and gastric intraductal papillary mucinous neoplasms (IPMN) and pancreatic intraepithelial neoplasias (PanIN).

Results Epigenetic regulation of mucin genes determines the phenotype of PDAC precursors. PanIN and gastric IPMN display a ductal molecular profile and numerous similarly regulated pathways, including the Notch pathway, but can be distinguished by recurrent deletions and differential methylation and, in part, by the expression of mucin-like 3. Intestinal IPMN are clearly distinct lesions at the molecular level with a more instable genotype and are possibly related to a different ductal cell compartment.

Conclusions PDAC precursors with gastric and intestinal phenotype are heterogeneous in terms of morphology, genetic and epigenetic profile. This heterogeneity is related to a different cell identity and, possibly, to a different aetiology.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Intraductal papillary mucinous neoplasms (IPMN) and pancreatic intraepithelial neoplasias (PanIN) are well-known pancreatic ductal adenocarcinoma (PDAC) precursors and have been characterised concerning their morphology and their immunohistochemical and genetic profile.
⇒ PanIN and gastric IPMN are mostly localised in the peripheral duct system and are mainly distinguished based on their size.
⇒ Intestinal IPMN are mostly main duct lesions with high frequency of *GNAS* mutations.

WHAT THIS STUDY ADDS

⇒ PanIN and gastric IPMN have a very similar genetic and epigenetic profile traceable to the ductal cell compartment.
⇒ Differential epigenetic regulation and expression of mucin-like 3 (MUCL3) and the presence of recurrent copy number variation (mainly deletions) in gastric IPMN may indicate a higher potential for progression in these lesions.
⇒ Intestinal IPMN display a distinct genetic landscape and higher level of genomic instability with higher proliferation rates already in low-grade lesions, suggesting a higher susceptibility for progression compared with PanIN and gastric IPMN.
⇒ Intestinal IPMN show an upregulation of genetic signatures related to mucin secretion and a clearly distinct epigenetic profile based on DNA methylation patterns compared with PanIN and gastric IPMN, relating them to a different adult cell type within the ductal compartment.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive human neoplasms and represents the fourth most common cause of cancer-related deaths in Western countries.[1] A curative surgical approach is feasible only in about 20% of the patients.[2] Despite numerous progresses in the last years, the number of PDAC patients surviving longer than 5 years is disappointingly low and most patients will succumb to their disease.[3–5] Current treatment strategies focus on fighting the advanced disease, present in about half of the cases at diagnosis, by combining standard chemotherapy with targeted and immune-based therapies with limited benefit so far.[6 7]

In contrast to other solid tumours like lung, breast or colon cancer, there are only a few known

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Immunophenotypical and, when necessary, molecular subtyping is fundamental in correctly classifying PDAC precursors with different risk of progression and its use in pathology reports should be enforced.
⇒ Despite their similarities, distinction between PanIN and gastric IPMN is relevant and should be pursued with available (eg, *GNAS* mutations) and, possibly, newly established markers, such as MUCL3.
⇒ PDAC precursor subtype-directed lineage and aetiological factor-based studies will allow identification and evaluation of lesion-specific prevention strategies.

factors that increase the risk of developing PDAC.[6 8] Among at least partly modifiable factors, cigarette smoking, obesity, long-standing diabetes and non-hereditary chronic pancreatitis are associated with a threefold to sixfold increased lifetime risk for PDAC. Other factors include hereditary causes with identified or still unknown gene alterations, for which effective screening strategies are still missing.[9 10] It seems therefore that the only effective strategy to substantially change the prognosis of this dismal disease is to detect and treat it in a very early stage, possibly at the stage of its precursor lesions.[11]

Intraductal and cystic lesions belong to the currently recognised precursors of PDAC and include pancreatic intraepithelial neoplasias (PanIN), intraductal papillary mucinous neoplasms (IPMN), intraductal oncocytic papillary neoplasms, intraductal tubulo-papillary neoplasms and mucinous cystic neoplasms.[12 13] Among them, PanIN and IPMN are the most relevant due to their frequency (PanIN) and the clinical challenge related to their treatment (IPMN). PanIN represent the longest-known and best-characterised precursors of classical PDAC, although they progress with low frequency.[14 15] IPMN encompass three histopathological subtypes, namely intestinal, gastric and pancreatobiliary, with different morphology, immunophenotype and, partly, biological behaviour.[13] Despite these differences, numerous overlapping features exist, and distinction is not always clear-cut. For example, although intestinal IPMN are usually localised in the main pancreatic duct, they may extend, or seldom even occur, in peripheral branch ducts. The same holds true for branch-duct gastric IPMN, which can extend to or occur in the main pancreatic duct. In addition, mixed immune phenotypes are detected in about 5% of the cases and recently, a possible origin of intestinal IPMN from gastric IPMN has been proposed as well.[16 17] In addition, the distinction between PanIN and gastric IPMN is mainly based on their size (<0.5 cm and >1 cm, respectively) and on the different frequency of *GNAS* mutations. However, they share a common localisation and an identical morphology and immune profile (figure 1A), rendering distinction not always straightforward in cases of small ('incipient') IPMN.[18] This distinction may be of clinical relevance, for example, in the setting of intraoperative examination of the pancreatic neck margin: whereas leaving behind a PanIN will not have any consequences in most cases, a residual gastric IPMN might bear a higher risk of recurrence.[19 20]

Recent studies have tried to shed light on the progression of precursor lesions to invasive cancer. Accordingly, the genetic evolution of PanIN has been quantified, revealing a period of about 7 years necessary for an initiating cell to develop into metastatic cancer.[21] However, this model is often contradicted by the clinical observation of rapidly progressive disease with systemic dissemination preceding clinical appearance, thus suggesting the possibility of additional, not yet fully elucidated more rapid progression models such as chromothripsis rather than or in addition to linear stepwise genetic evolution.[22 23] A stepwise progression model is thought to play a role in IPMN as well, but the natural history of these lesions especially taking into consideration the different subtypes, remains largely unknown.[24]

It is therefore apparent that numerous questions regarding the development of pancreatic precursors from their cell of origin to high-grade and invasive lesions and their relation to each other remain open. In this work, we provide the first extensive genetic and epigenetic characterisation of PDAC precursors focusing on the molecular comparison between intestinal and gastric IPMN and PanIN.

MATERIALS AND METHODS

Additional protocols and complete procedures are described in the online supplemental material and methods section.

Study cohort

A tissue collection of precursor lesions of PDAC obtained from 154 patients operated on in the years 2008–2021 was established. The study cohort consists of 132 different precursor lesions and includes PanIN (n=55) and IPMN (n=77); 59 lesions (44.6%) occurred in the context of PDAC (table 1, online supplemental figure 1). All lesions were re-classified by reviewing all slides according to current criteria and nomenclature.[12 25] Only PanIN in pre-existent ducts were included. Representative slides from all lesions were stained with antibodies for mucin 1 (MUC1), mucin 2 (MUC2), mucin 5 (MUC5AC) and caudal type homeobox 2 (CDX2) for histopathological subtyping.[13] Only morphology and immunohistochemistry were considered for the distinction between PanIN and gastric IPMN and for IPMN subtyping. Diagnoses were performed by a pathologist with over 20 years' experience in pancreatic pathology (IE); difficult cases were discussed with another pathologist with over 50 years' experience in pancreatic pathology (GK). In cases with IPMN and PDAC, PDAC was defined as 'associated' if there was clear morphological evidence of its origin from the IPMN (ie, the invasive component originated from the intraductal lesion). If the IPMN was not spatially related to the PDAC, this was considered a 'concomitant' PDAC. A collective of 79 precursor lesions (66 of them being used in this study also for molecular analyses) and of 24 PDAC specimens was analysed by whole-slide immunohistochemistry to test the expression of trefoil factor 3 (TFF3) and mucin-like 3 (MUCL3) proteins (online supplemental table 1a,b).

Estimation of DNA copy number variation by low-coverage whole-genome sequencing and methylation data

For this analysis, 36 PanIN (28 low-grade and 8 high-grade), 38 gastric IPMN (29 low-grade and 9 high-grade) and 21 intestinal IPMN (8 low-grade and 13 high-grade) were used. Twenty-eight lesions (11 PanIN, 13 gastric IPMN and 4 intestinal IPMN) were analysed by low-coverage whole-genome sequencing (WGS). Briefly, isolated genomic DNA from formalin-fixed, paraffin-embedded (FFPE) samples was amplified with the Ampli1 WGA kit (Menarini Silicon Biosystems, Bologna, Italy) according to the manufacturer's instructions. Ten microlitres of the Ampli1 WGA product were used to clean up with 1.8× SPRIselect beads (Beckman Coulter, Lahntal, Germany). After that, low-coverage whole-genome libraries were prepared with the Ampli1 low-pass kit (Menarini Silicon Biosystems) according to the manufacturer's

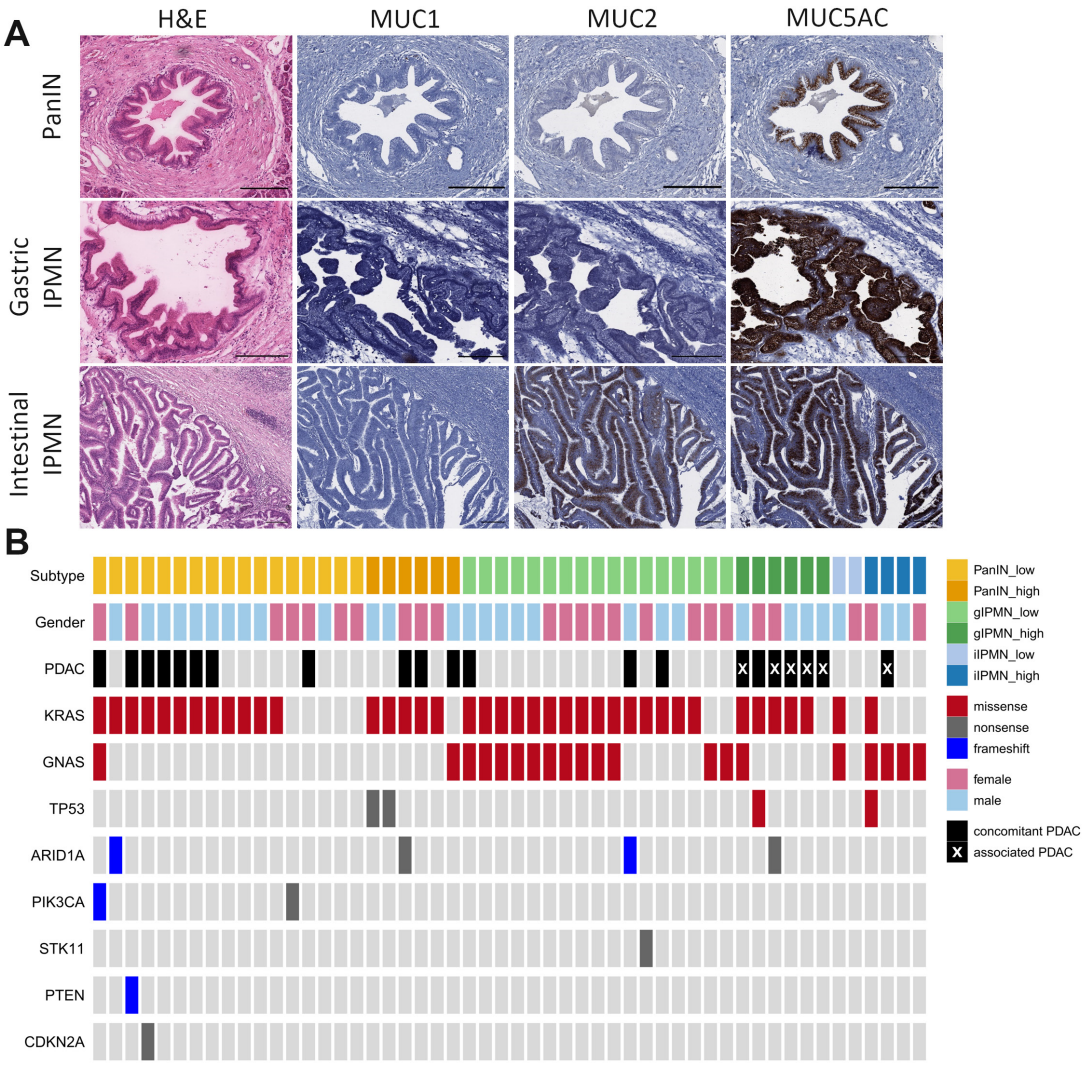

Figure 1 Intraductal precursors of pancreatic cancer: morphology and genetics. (A) Morphology and immunohistochemistry: PanIN and gastric IPMN are distinguished according to morphology and size and display an identical immunohistochemical profile with diffuse positivity for MUC5AC and no expression of MUC1 and MUC2. Intestinal IPMN are clearly distinct lesions, both on the morphological and immunohistochemical level, characterised by positivity for MUC2 and MUC5AC. H&E and immunohistochemistry (see 'Materials and methods' section; scale bar=200 µm). (B) Targeted-next-generation sequencing analysis: low-grade 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 those with associated PDACs are marked in addition with a white X. Labelled mutations represent pathogenic mutations according to the ClinVar database and/or the American College of Medical Genetics and Genomics guidelines with an allele frequency of ≥3%. Red squares represent missense mutations, grey squares are nonsense mutations and blue squares are frameshift mutations. Empty squares indicate absence of pathogenic mutations. Analysis was performed using a 21-gene custom panel on the S5 Ion Torrent platform (Phred score ≥30, coverage ≥500). IPMN, intraductal papillary mucinous neoplasms; MUC1, mucin 1; MUC2, mucin 2; MUC5AC, mucin 5; PanIN, pancreatic intraepithelial neoplasias; PDAC, pancreatic ductal adenocarcinoma.

Table 1 Study cohort

Diagnosis	Number of cases (%)	Degree of dysplasia		Cases with PDAC
		Low-grade	High-grade	
PanIN	55/132 (41.7%)	43/55 (78.2%)	12/55 (21.8%)	34/55 (61.8%)
IPMN gastric	46/132 (34.8%)	35/46 (76%)	11/46 (23.9%)	15/46 (32.6%)
IPMN intestinal	21/132 (15.9%)	8/21 (38%)	13/21 (62%)	7/21 (33.3%)
IPMN pancreatobiliary	3/132 (2.3%)	0	3/3 (100%)	2/3 (66.6%)
IPMN mixed	7/132 (5.3%)	7/7 (100%)	0	1/7 (14.3%)

IPMN, intraductal papillary mucinous neoplasms; PanIN, pancreatic intraepithelial neoplasias; PDAC, pancreatic ductal adenocarcinoma.

instructions. The final library concentration was determined on the fragment analyser (Advanced Analytical Technologies, Ames, Iowa, USA) with the Agilent high-sensitivity genomic DNA 50 kb kit (Agilent Technologies, Ratingen, Germany). An additional size selection was not done. A 100 pM equimolar library pool was created and sequenced with the Ion S5 system (ThermoFisher Scientific, Dreieich, Germany) as described in online supplemental methods.

The Ion Browser Suite mapped the sequence to the GRChr37/hg19 genome. The Ion Reporter software (V.5.12.0.0) was used for copy number determination of the low-pass sequencing. The $\log_2$ (tumour/normal) value was calculated for each region. As a control group, six normal tissue samples consisting of acinar tissue were isolated and sequenced with the same method. Copy number regions with a $\log_2$ ratio greater than +0.2 and less than −0.2 were considered. The median of the absolute values of all pairwise differences value was set as <0.35.

Additional 67 lesions (24 PanIN, 26 gastric IPMN and 17 intestinal IPMN) were analysed using the data from the DNA methylation profiles obtained with the Infinium Methylation EPIC BeadChip (see below, "Differential methylation analysis"). Here, the copy number variation (CNV) was estimated with the conumee package (V.1.3)[26] using default settings. As normal control served the combined intensities from the bulk acini samples (n=11); changes of 0.2 and −0.2 in the mean segment value were set as thresholds to define copy number gains and losses. To detect common regions between EPIC and low-pass samples, bed files were generated and compared with BEDTools (V.2.3).[27] A 'common' region was defined if three or more samples within the analysed precursor lesion shared the same loci. For those hits, the annotated curated NCBI RefSeq genes were retrieved from the UCSC Genome Browser (GRCh37/hg19).

Differential methylation analysis

DNA methylation profiles were measured with the Infinium Methylation EPIC BeadChip (Illumina, San Diego, USA) at the Genomics and Proteomics Core Facility of the German Cancer Research Center Heidelberg. Methylation analysis was carried out using the R Bioconductor package ChAMP (V.2.14.0).[28] Briefly, IDAT files were loaded into ChAMP and preprocessed. In the first step, all probes with a detection 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, single nucleotide polymorphism (SNP)-containing probes and sex probes were also filtered. Filtered datasets were normalised using the Beta Mixture Quantile dilation (BMIQ) method and batch corrected before differential analysis. Differentially methylated probes were defined by a delta of 0.2 and an adjusted p value (Benjamini-Hochberg method) of ≤0.05. The phylogenetic tree was plotted using the R-package ape (V.5.3).

RESULTS

Morphology

The sample cohort consisted of 55 PanIN (41.7%), 46 gastric IPMN (34.8%) and 21 intestinal IPMN (15.9%) (table 1). Pancreatobiliary and mixed-type IPMN were excluded from further analyses due to small sample size. PanIN and gastric IPMN displayed the same immunophenotype (figure 1A) and were distinguished according to established criteria.[12 25]

Gene mutations, fusion transcript analysis and chromosome copy number aberrations of PDAC precursors

Targeted next-generation sequencing was performed in 59 samples, including 7 control samples of normal acinar tissue. In detail, 23 PanIN (17 low-grade and 6 high-grade; 12 in cases without PDAC), 23 gastric IPMN (17 low-grade and 6 high-grade) and 6 intestinal IPMN (2 low-grade and 4 high-grade) were sequenced using a custom 21-gene panel (figure 1B; online supplemental table 2).

KRAS G12 mutations on exon 2 were present in 16/23 PanIN (69.5%), 19/23 gastric IPMN (82.6%) and 2/6 (33%) intestinal IPMN. *KRAS* Q61 mutations on exon 3 were found in one PanIN and one gastric IPMN, making up 5.1% of all *KRAS* mutated cases (online supplemental table 3). The pathogenic R201 *GNAS* mutation was present in 17/29 (58.6%) IPMN and in 2/23 PanIN (8.6%), whereas 1/29 IPMN (3.4%) displayed a *GNAS* Q227 mutation. Pathogenic *TP53* mutations were detected in four high-grade lesions (2 IPMN (2.2%) and 2 PanIN (8.7%)). In addition, *ARID1A*, *PIK3CA*, *STK11*, *PTEN* and *CDKN2A* nonsense and frameshift mutations were observed in few individual lesions (figure 1B; online supplemental table 3). The overall frequency of mutations in PanIN in specimens without PDAC was not significantly different from that of PanIN in specimens with PDAC (not shown). IPMN had a significant higher variant allele frequency (VAF) of *KRAS* and *GNAS* than PanIN (online supplemental figure 2A–B), possibly due to contamination by normal tissue in dissected PanIN lesions. Pearson's correlation analysis of double mutated (*KRAS* and *GNAS*) gastric IPMN samples confirmed a positive correlation between the VAFs of the two mutations (r=0.9795, p≤0.0001, online supplemental figure 2C), indicating that these most probably occurred in the same cell.

Six cases without mutations (four PanIN, one gastric IPMN and one intestinal IPMN) were subjected to fusion transcript analysis to check for possible alternative drivers. Five samples revealed no detectable fusion transcripts (not shown); in one case (low-grade PanIN), the analysis was not possible due to insufficient RNA quality. Morphology was not predictive of the genetic status; representative examples of lesions with identical morphology and different genetic changes are shown in online supplemental figure 3.

Next, we assessed CNV by two orthogonal methods: DNA methylation array data and whole-genome low-coverage sequencing. Among the three precursor lesions, PanIN displayed the lowest number of samples affected by genomic losses and gains (n=22, 61%) followed by gastric IPMN (n=29, 76%) and intestinal IPMN (21, 100%) (table 2A). There was no relationship between the degree of dysplasia and the presence/absence of CNV (online supplemental table 4). Although intestinal IPMN showed in general higher numbers of deletions and amplifications per sample, there was a remarkable difference in the median size of deletions for gastric IPMN (4.5 Mb) compared with PanIN (0.7 Mb) and intestinal IPMN (2.3 Mb). CNV values are shown in online supplemental table 5. Furthermore, only gastric IPMN showed a loss of *TP53* (chr.:17) and *CDKN2A* (chr.:9) in multiple samples (table 2B). Deletions on chromosome 11 were solely detected in intestinal IPMN affecting the putative tumour suppressor genes *CTNND1*, *MEN1*, *ATM* and *KMT2A*. Beside this locus, intestinal IPMN where generally affected by amplifications (figure 2B and C). This finding was underpinned by the median size of 5.1 Mb/amplification compared with 1.2 Mb/amplification and 2 Mb/amplification for PanIN and gastric IPMN, respectively (table 2a). In addition, recurrent

Table 2 (A) Overview of DNA copy number variations (CNV) of pancreatic precursor lesions, (B) detailed overview of genomic regions affected by CNV

	PanIN (n=36)		gIPMN (n=38)		iIPMN (n=21)		
Samples with CNV	22	(61%)	29	(76%)	21	(100%)	
Samples with only deletions	8	(22%)	5	(13%)	1	(5%)	
Samples with only amplifications	7	(19%)	12	(32%)	4	(19%)	
Samples with deletions and amplifications	7	(19%)	12	(32%)	16	(76%)	
Median no. of deletions/sample (95% CI)	2	(1 to 3)	2	(2 to 6)	4	(2 to 12)	
Median size of deletions/sample ((95% CI) Mb)	0.7	(0.5 to 1.9)	4.5	(2.3 to 10.3)	2.3	(1.6 to 4.0)	
Median no. of amplifications/sample (95% CI)	3	(2 to 6)	2	(2 to 4)	6.5	(2 to 19)	
Median size of amplifications/sample ((95% CI) Mb)	1.2	(0.8 to 1.9)	2	(1.1 to 3.5)	5.1	(3.9 to 8.0)	
Recurrent chromosomal regions (n≥3)	5		44		229		
Affected putative tumour suppressor genes	0		7		26		
Affected putative oncogenes	0		0		26		

Genomic location	PanIN (n=36)		gIPMN (n=38)		iIPMN (n=21)		Gene symbols
Deleted regions							
chr01:010875000-013052998					3	(14%)	MTOR
chr01:015375000-016825000					3	(14%)	EPHA2
chr06:074175000-074375000			3	(8%)	2	(10%)	EEF1A1
chr06:133664400-143100000			2	(5%)	3	(14%)	TNFAIP3
chr06:143620678-151100000	1	(3%)	3	(8%)	3	(14%)	LATS1
chr09:005958053-023802212	1	(3%)	4	(11%)			PSIP1 | CDKN2A
chr10:071075000-120925000			2	(5%)	4	(19%)	PTEN | TCF7L2
chr10:120925000-125869472			1	(3%)	4	(19%)	FGFR2
chr11:057325000-058807232					4	(19%)	CTNND1
chr11:058807232-069089801					5	(24%)	MEN1 | SF1
chr11:096437584-114325000					5	(24%)	ATM
chr11:114325000-134898258					4	(19%)	KMT2A
chr17:006225000-009675000			3	(8%)	1	(5%)	GPS2 | TP53
chr17:009675000-012500000			3	(8%)	2	(10%)	MAP2K4
chr17:015792977-021566608			3	(8%)	3	(14%)	NCOR1
Amplified regions							
chr01:035225000-037325000					3	(14%)	THRAP3
chr03:176225000-188875000	1	(3%)	2	(5%)	3	(14%)	PIK3CA
chr05:028950000-044925000					5	(24%)	NIPBL
chr06:024125000-033575000			2	(5%)	3	(14%)	HLA-A | HLA-B
chr06:033575000-042725000			1	(3%)	3	(14%)	CDKN1A | PIM1
chr07:000282484-007150000					5	(24%)	RAC1
chr07:054725000-055775000					5	(24%)	EGFR
chr07:061967157-074715724					4	(19%)	GTF2I
chr07:112425436-130154523					5	(24%)	MET
chr07:139404377-142048195					5	(24%)	BRAF
chr07:143397897-154270634					5	(24%)	KMT2C |CUL1
chr08:086726451-089550000					3	(14%)	CNBD1
chr08:127450000-129175000	1	(3%)			6	(29%)	MYC
chr09:001992685-035698318					3	(14%)	PSIP1 | CDKN2A
chr09:070835468-092343416					4	(19%)	GNAQ
chr09:096718222-097575000					4	(19%)	PTPDC1
chr09:097775000-114750000					4	(19%)	PTCH1
chr09:124994207-133073060					3	(14%)	SPTAN1 | PPP6C
chr12:006475000-007169938	1	(3%)			7	(33%)	CHD4
chr12:024993545-028938805	1	(3%)			3	(14%)	KRAS
chr14:020700000-022050000					3	(14%)	CHD8

Continued

Table 2 Continued

Genomic location	PanIN (n=36)	gIPMN (n=38)		iIPMN (n=21)		Gene symbols
chr14:022800000-050175000				3	(14%)	AJUBA | FOXA1
chr14:097258910-107289540				3	(14%)	TRAF3 | AKT1
chr17:061125000-062410760				3	(14%)	CD79B
chr17:062775000-063525000				3	(14%)	GNA13
chr17:068117898-077546461				3	(14%)	SOX9
chr20:008050000-016400000		1	(3%)	6	(29%)	PLCB4
chr20:016625000-021300000		1	(3%)	6	(29%)	ZNF133
chr20:030025000-034897085		1	(3%)	6	(29%)	ASXL1
chr20:036958189-042991501		1	(3%)	5	(24%)	PLCG1
chr20:052650000-061091437		1	(3%)	6	(29%)	GNAS
chr21:032825000-034475000				3	(14%)	SCAF4

CNV alterations were detected by low-coverage sequencing (n=28) and DNA methylation data (n=67), respectively. Values are n (%) unless otherwise indicated.
Gain and loss of genomic regions were detected by low-coverage sequencing (n=28) and DNA methylation data (n=67), respectively. Values are n (%), putative tumour suppressors (green) and putative oncogenes (red).
gIPMN, gastric IPMN; iIPMN, intestinal IPMN; IPMN, intraductal papillary mucinous neoplasms; PanIN, pancreatic intraepithelial neoplasias.

regions containing a putative oncogene were detected only in intestinal IPMN. Among the intestinal IPMN precursors, there was a higher prevalence of amplification for the chromosomes 7 (*EGFR*, *MET*, *BRAF*), 8 (*MYC*), 12 (*CDH4*) and 20 (*GNAS*), which were amplified in five or more cases (table 2b). The significantly higher Ki-67 proliferations rates in intestinal IPMN (online supplemental figure 5) could further support the presence of a higher level of genomic instability in intestinal IPMN.

Genome-wide DNA methylation analysis of PDAC precursors

DNA methylation data were generated in 79 different FFPE samples of precursor lesions, including 27 PanIN (20 low-grade and 7 high-grade), 32 gastric IPMN (24 low-grade and 8 high-grade) and 20 intestinal IPMN (8 low-grade and 12 high-grade). For comparison, we generated DNA methylome profiles from normal acinar, ductal and neuroendocrine cell compartments. Thus, acinar bulk tissue (n=11), main duct (n=11) and branch duct (n=8) cell preparations as well as FACS (fluorescence activated cell sorting)-sorted β-cells (n=3) from healthy pancreatic tissue were included in the analysis. Additionally, we added publicly available samples of FACS-sorted ductal cells (n=4) and FACS-sorted acinus cells (n=4) for internal control.[29 30]

After the data processing, 702 656 probes were used to analyse the cell type and precursor lesion-specific DNA methylation profiles. Despite differences in sample preparation between in-house and publicly available samples (ie, FFPE vs fresh frozen), multidimensional scaling revealed a coherent population of acinar bulk tissue samples and sorted acini, while spreading of ductal cells was larger (online supplemental figure 4A). A hierarchical clustering based on probes associated with known ductal and acinar markers clearly separated the normal cell populations (online supplemental figure 4B). The methylation level of CpGs located in acinar marker genes was higher in ductal cells, whereas ductal markers were hypermethylated in acinar cells. Pairwise comparison of the control groups revealed a similar amount of differentially methylated probes in acinar versus ductal cells (12.4%) and ductal versus β-cells (11.3%). However, based on multidimensional scaling, the distance between β-cells and the other two normal pancreatic cell types was larger than between acinar and ductal cells (figure 3A). The highest degree of significant differential methylation was detected between intestinal IPMN and β-cells (26.8%), whereas no significantly differentially methylated CpG was observed between branch and main ducts and between gastric IPMN and PanIN lesions, respectively (figure 3B). To address potential functional effects of the detected differentially

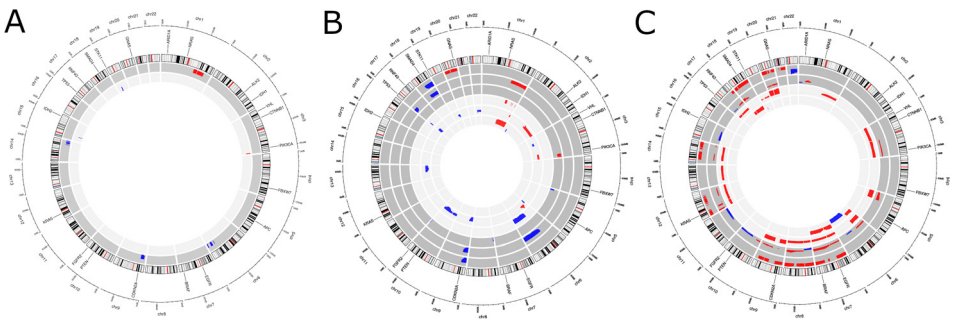

Figure 2 Copy number variation (CNV) in PanIN, gastric and intestinal IPMN. CNVs were detected in PanIN and IPMN over the whole-genome by low-coverage sequencing. Regions in red show copy number gains and regions in blue represent copy number losses. (A) PanIN (n=11, 9 low-grade and 2 high-grade) do not possess repeated or larger regions of CNV; (B) gastric IPMN (n=13, 9 low-grade and 4 high-grade lesions) reveal three distinct repeated regions of copy number loss at chromosome 6, 9 and 18; (C) intestinal IPMN (n=4, 2 low-grade and 2 high-grade) had the highest frequency of chromosomal alterations. The broad genomic alterations generally involve entire chromosomes and are mostly located on chromosome 7, 8, 12, 18 and 20 (dark grey background=high-grade lesions, light grey background=low-grade lesions; $\log_2$ value, threshold±0.2). IPMN, intraductal papillary mucinous neoplasms; PanIN, pancreatic intraepithelial neoplasias.

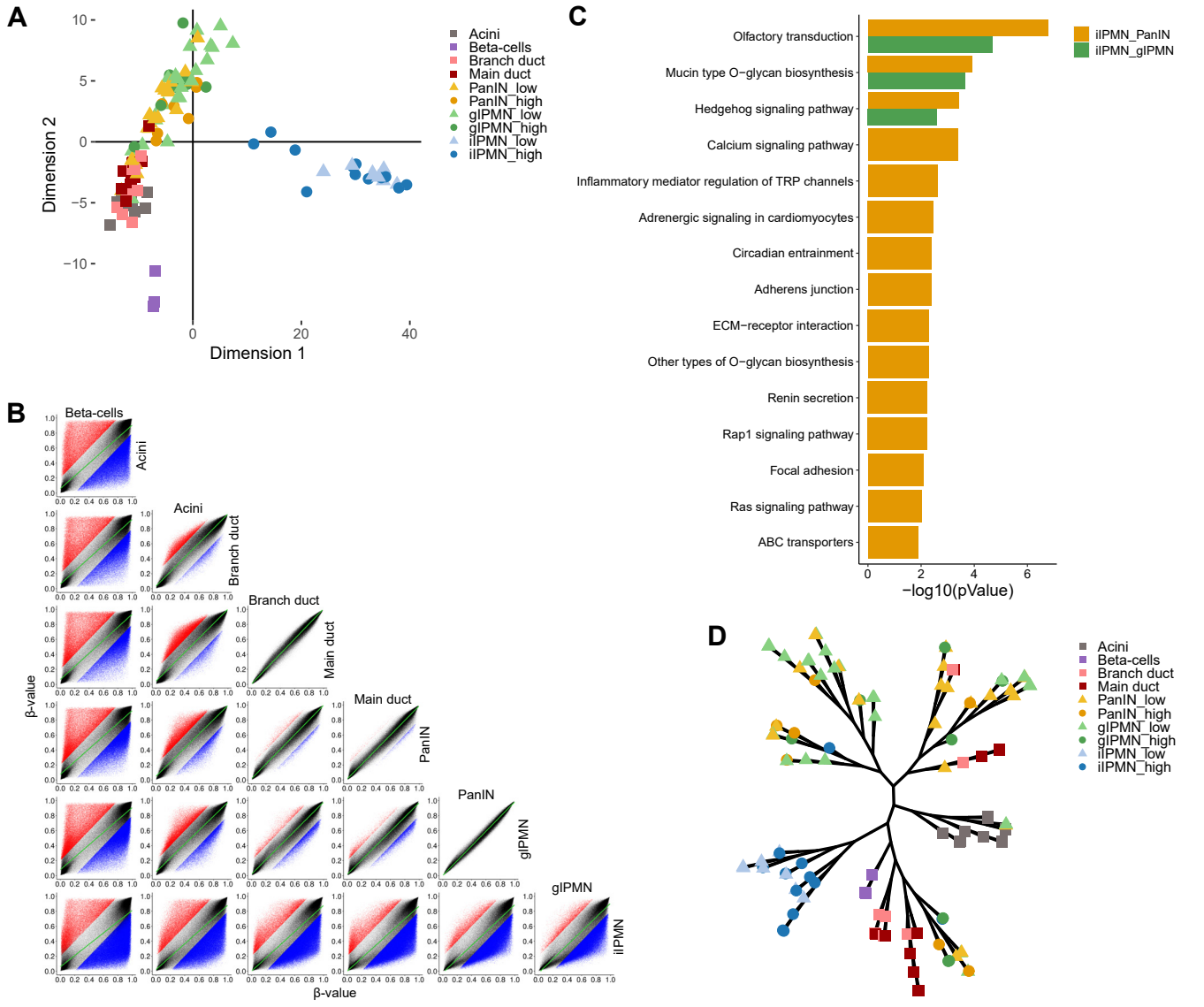

Figure 3 DNA methylation profiling of normal pancreas cells and PDAC precursor lesions. (A) Multidimensional scaling based on the 5000 most variable CpG probes. (B) Scatter plots showing pairwise comparisons of methylated probes between indicated precursor lesions and cell types. Significantly hypermethylated probes (delta beta ≥0.2; adjusted p value ≤0.05) are coloured in red and hypomethylated (delta beta ≤−0.2; adjusted p value ≤0.05) in blue, respectively. (C) KEGG pathway enrichment analysis of differentially methylated probes between IPMN and PanIN. (D) Phylogenetic tree displaying the relationship between precursor lesions and pancreatic cell types based on DNA methylation data. ABC, ATP-binding cassette; ECM, extracellular matrix; gIPMN, gastric IPMN; iIPMN, intestinal IPMN; IPMN, intraductal papillary mucinous neoplasms; PanIN, pancreatic intraepithelial neoplasias; PDAC, pancreatic ductal adenocarcinoma; TRP, transient receptor potential.

methylated probes (DMPs) between the different precursor lesions, we looked for enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) gene sets. This analysis revealed numerous differentially enriched gene sets between PanIN and intestinal IPMN, which involved signalling pathways as well as pathways regulating cell-cell and cell-extracellular matrix interactions (figure 3C). When comparing gastric and intestinal IPMN, only three enriched gene sets were identified, including a differential regulation of the mucin type O-glycan biosynthesis and of the Hedgehog signalling pathway, among others. Notably, due to the low number of DMPs between PanIN and gastric IPMN, no significantly enriched gene set was detected, arguing for a high similarity between the two lesions, as also suggested by the phylogenetic tree analysis (figure 3D).

We next evaluated potential differences in methylation patterns between low-grade and high-grade preneoplastic precursor lesions. Notably, we found no significant DMP between low-grade and high-grade intestinal versus gastric IPMN or between intestinal IPMN versus gastric IPMN and PanIN, respectively. When comparing low-grade and high-grade PanIN lesions, 86 significant DMPs associated with 59 genes were found between PanIN low-grade and high-grade samples, however, no candidate gene or gene network became apparent as promising candidate driver of progression (online supplemental table 6A,B).

Transcriptome analysis of PDAC precursors

To get further insights into the distinguishing features of PanIN and IPMN, transcriptome data were generated from 41 different FFPE samples obtained from 10 PanIN (6 low-grade and 4 high-grade), 12 gastric IPMN (6 low-grade and 6 high-grade), 12 intestinal IPMN

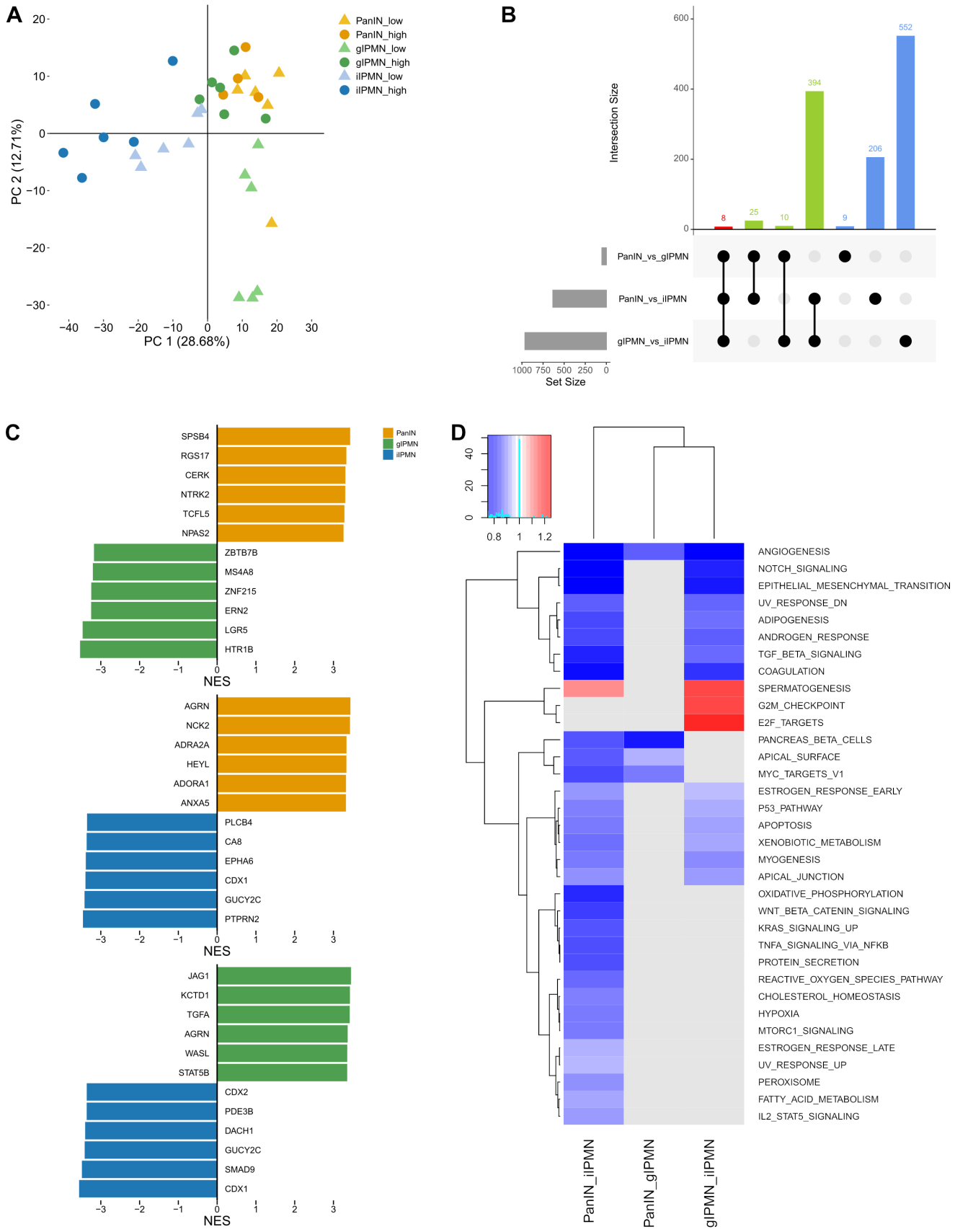

Figure 4 Transcriptomics-based comparative analysis of precursor lesions. (A) Principal component analysis with the 500 most variable genes displaying a precursor-specific clustering. (B) Upset plot summarised the differentially expressed genes between the three precursors. (C) The precursor-specific activation of transcription factors detected by VIPER analysis based on group-wise comparisons. (D) Single sample gene set enrichments analysis indicates precursor-specific activation of *hallmark of cancer* gene sets from the MSigDB collection. gIPMN, gastric IPMN; iIPMN, intestinal IPMN; IPMN, intraductal papillary mucinous neoplasms; NES, normalised enrichment score; PanIN, pancreatic intraepithelial neoplasias.

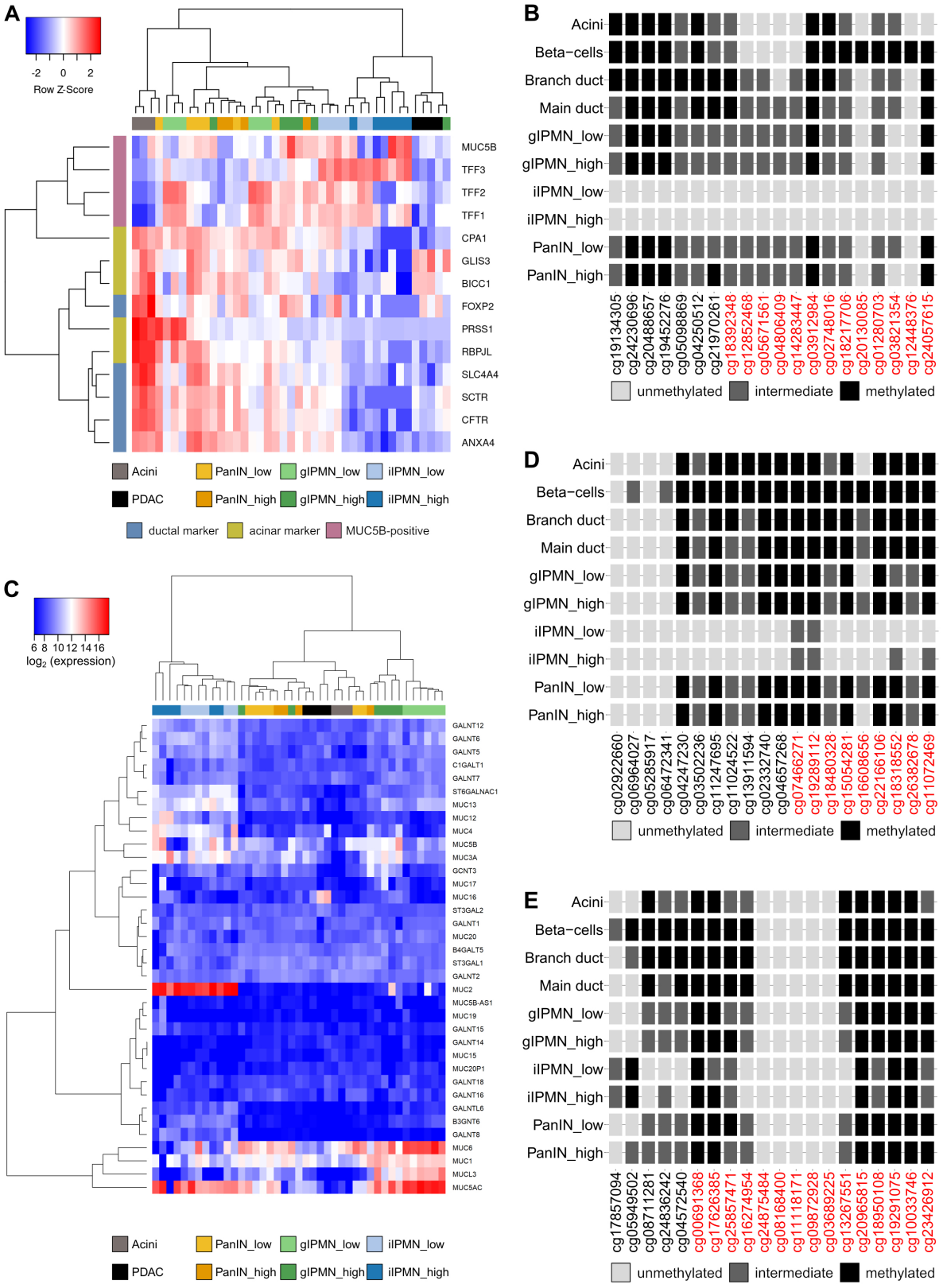

Figure 5 Identification of different precursor subtype-specific markers. (A) Hierarchical clustering of RNA sequencing data based on published marker genes for distinct normal pancreas cell populations.[29] (B) Mean CpG methylation of all *TFF3* annotated probes. CpGs located in the coding region are coloured in red (unmethylated: mean β-value <0.4; intermediate: mean β-value >0.4 and <0.6; methylated: mean β-value >0.6). (C) Hierarchical clustering displaying the expression of genes involved in the Mucin type O-glycan biosynthesis and mucins expressed in precursor lesions. (D, E) Mean CpG methylation of the first 20 *MUC2* (D) and *MUCL3* (E) annotated probes. CpGs located in the coding region are coloured in red (unmethylated: mean β-value <0.4; intermediate: mean β-value >0.4 and <0.6; methylated: mean β-value >0.6). gIPMN, gastric IPMN; iIPMN, intestinal IPMN; IPMN, intraductal papillary mucinous neoplasms; PanIN, pancreatic intraepithelial neoplasias; PDAC, pancreatic ductal adenocarcinoma; TFF3, trefoil factor 3.

(6 low-grade and 6 high-grade), 4 PDAC (unrelated to IPMN) and 3 samples of acinar bulk tissue. The principal component analysis showed a clear separation between intestinal IPMN, and other precursors (figure 4A). The observed pattern was comparable with the results of the DNA methylation data, further underlining the close relationship between PanIN and gastric IPMN also on the transcriptional level. Consistently, only very few genes were differentially expressed between gastric IPMN and PanIN (figure 4B). The Visualization Pipeline for RNAseq) (VIPER) analysis showed a similarity in pathway activation when gastric IPMN and PanIN were compared with intestinal IPMN. This comparison displayed that in both precursor lesions (ie, gastric IPMN and PanIN) genes of the Notch signalling (namely *HEYL*—Hes-Related Family BHLH Transcription Factor With YRPW Motif Like, *JAG1*—Jagged Canonical Notch Ligand 1, *TGFA*—transforming growth factor alpha) are activated compared with intestinal IPMN (figure 4C). The activation of Notch signalling was also observed in single sample gene set enrichments (ssGSEA) based on the *hallmark of cancer* gene sets (figure 4D), arguing for distinct Notch signalling activity in gastric but not intestinal IPMN.

Identification and validation of precursor subtype-specific markers

We further validated our data by comparing them with a recently published comprehensive transcriptomic characterisation of pancreatic cells obtained from healthy organ donors.[31] Here, higher expression levels of digestive enzymes, including *CPA1* (carbopeptydase A1), *PRSS1* (protease serine 1) and of relevant transcription factors, like *FOXP2* (forkhead box P2) and *RBPJL* (Recombination Signal Binding Protein for Immunoglobulin Kappa J Region Like), were found in samples obtained from the normal pancreas (figure 5A). Interestingly, progressively decreasing expression levels were identified in some of the markers (*CPA1*, *RBPJL*), when moving from precursors usually located in the peripheral ductal system, such as PanIN and gastric IPMN, to typical main duct lesions, such as intestinal IPMN. Most interestingly, intestinal IPMN displayed an association with genes related to the subgroup of *MUC5B*-positive ductal cells, which has been identified as 'minor' ductal cell population in the normal pancreas.[31] This subtype is characterised by higher expression levels of genes related to mucous secretion. Accordingly, significantly higher levels of the trefoil factor *TFF3*, whose promoter region was selectively unmethylated, were found in intestinal IPMN compared with all other precursor lesions, both at the messenger RNA and at the protein expression level, respectively (figure 5A–B, online supplemental figure 5). Furthermore, intestinal IPMN displayed highly differentially methylated genes, such as *FOXP2*, *DCLK1* (doublecortin-like kinase 1) and *BICC2* (bicaudal C homolog 2) (online supplemental figure 6), which can be ascribed to progenitor markers, possibly suggesting the presence of a still unidentified progenitor-like cell population in the pancreatic ductal system.

Further focusing on the mucin metabolism, we found an enrichment of the O-linked glycosylation signature enriched in both IPMN subtypes by applying ssGSEA (online supplemental figure 7). As this gene set was already observed at the

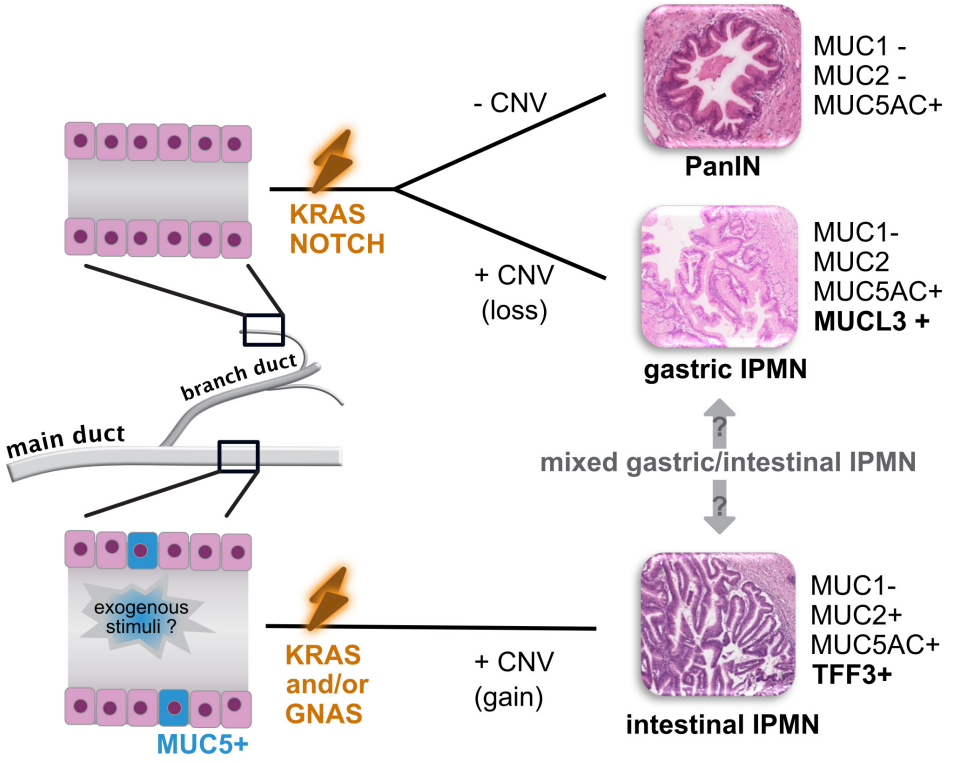

Figure 6 Model of development of pancreatic cancer precursors. *KRAS* mutations induce a gastric phenotype characteristic of mostly peripherally located lesions, such as PanIN and gastric IPMN, which are additionally Notch-dependent. Recurrent deletions occur only in gastric IPMN. These share a very similar mucin profile with PanIN, but they can be distinguished to some extent by different MUCL3 expression, with lack of expression arguing against gastric IPMN. Further stimuli, such as exogenous factors related to a different microenvironment and possibly acting on a minor MUC5B-positive ductal cell population, induce an intestinal phenotype, driven by *KRAS* and/or *GNAS* mutations, with differential regulation of the mucin type O-glycan biosynthesis, expression of MUC2, CDX2 and TFF3 and recurrent amplifications. Mixed phenotypes and/or a transition from a gastric to an intestinal phenotype may also occur. CNV, copy number variation; IPMN, intraductal papillary mucinous neoplasms; MUC1, mucin 1; MUC2, mucin 2; MUC5AC, mucin 5; PanIN, pancreatic intraepithelial neoplasias; TFF3, trefoil factor 3.

methylation level (figure 3C), we further analysed the individual gene expression and evaluated their impact in distinguishing PanIN, gastric and intestinal IPMN from each other. Beside the known precursor-specific expression of *MUC2* (figure 5C–D), *MUC5AC* and *MUC6* (figure 5C), this analysis identified for the first time *MUCL3* as a potential candidate to distinguish PanIN from gastric IPMN both at the transcriptional and methylation (figure 5E) level. Immunohistochemical analysis confirmed a more frequent MUCL3 expression in gastric IPMN than in PanIN (p=0.02) (online supplemental figure 5).

DISCUSSION

In this study, pancreatic cancer precursors with gastric and intestinal phenotype were analysed using genomic, epigenomic and transcriptomic approaches to address their molecular profile and assess their cell identity and possible similarity to adult pancreatic cell compartments. A major result of our integrated genome-based approach is the evidence of distinct genomic structural patterns and of an epigenetic regulation of mucin genes underlying the different phenotype of PDAC precursors.

Gastric IPMN and PanIN show a significant overlap in their gene expression and DNA methylation profile, with numerous commonly regulated pathways, including Notch signalling, which is supported by previous functional studies on Notch signalling in PanIN-driven pancreatic carcinogenesis.[32–34] So far, most functional data have focused on acinar-ductal metaplasia and PanIN formation and progression. While several mouse models have been reported to elicit different subtypes of IPMN by combining the KC or KPC mouse[35 36] with genetic targeting of additional pathways, such as the G-protein coupled receptor (GPCR) (*guanine nucleotide binding protein, alpha stimulating (GNAS)*), tranforming growth factor-alpha (TGFα), SWItch/Sucrose Non-Fermentable (SWI/SNF), Wingless/Integrated (WNT) Wand phosphoinosotide-3 kinase (PI3K) pathways,[37] the functional role of Notch signalling in the development of gastric and intestinal IPMN is not well-defined. Our finding suggests that Notch signalling is selectively involved in gastric but not in intestinal precursor development. It is tempting to speculate that PanIN and gastric IPMN have distinct precursor cells (compared with intestinal IPMN) being responsive to or requiring Notch signalling activity. In addition, they display very similar methylation profiles compared with ductal cells both from the main and the branch-duct compartment (figure 3B,D). This may appear in contrast with previous studies ascribing an essential role to Kras and Notch signalling for acinar-ductal reprogramming and development of PDAC precursors.[38] However, by comparing our results with those obtained from single cell RNA-expression analysis of normal pancreatic tissues,[31] we observed a retained expression of acinar markers in PanIN and gastric IPMN, which may still point towards a contribution of the acinar cell compartment to these two lesions, although minor contamination by acinar cells during the process of microdissection cannot be completely ruled out.

Interestingly, WGS and DNA methylation profiles revealed slightly more frequent (76% vs 61%) CNV in gastric IPMN than in PanIN lesions, with chromosomal regions affected by recurrent deletions only found in gastric IPMN, thus suggesting a higher impact of mutagenic factors in gastric IPMN, which potentially affect their progression.[39 40] The significant differential expression of *MUCL3* in gastric IPMN further consolidates the hypothesis of a higher potential for progression of these lesions compared with PanIN, since this molecule has been previously described to be overexpressed in PDAC and to promote its progression by affecting the nuclear factor-kappa B signalling pathway,[41] but more data are necessary to confirm these observations. In addition, the diagnostic value MUCL3 in distinguishing PanIN from gastric IPMN might be only relevant in case of lack of expression, which would then exclude gastric IPMN. The degree of dysplasia does not appear to be related to the presence of recurrent CNV, since the proportion of low-grade and high-grade lesions was not substantially different in the two groups. This is in contrast with previous studies, which showed infrequent and mostly non-recurrent CNV both in low-grade PanIN and IPMN[42] and identified a correlation between more frequent CNV and higher histological grade in IPMN.[43] On the other hand, CNV analysis in IPMN of mostly mixed phenotypes revealed recurrent gains in chromosome 3, 7, 8 and 12, in line with our results.[44]

To our knowledge, this study is the first to show that intestinal IPMN have profoundly divergent methylation profiles compared with both PanIN and gastric IPMN (figure 3A–C), suggesting a distinct cell identity. Indeed, single cell sequencing studies have revealed a certain degree of heterogeneity in the cell populations of the healthy adult pancreas.[31 45] A minor ductal cell population (defined as MUC5B-positive ductal cells), characterised by higher expression levels of genes related to mucin secretion, such as TFF3, has been described, which, according to our gene and protein expression data (figure 5A–B, online supplemental figure 5), could be related to intestinal IPMN. It is tempting to speculate that exposure of the ductal cell compartment to environmental carcinogens, for example, due to bile reflux as a consequence of an anatomic variation at the pancreato-biliary junction[46 47] or to an altered oral, gastric and intestinal microbiome,[48] could induce an intestinal phenotype switch as first adaptive response of a 'susceptible' cell type, followed by dysplasia and cancer. Since the ductal cell compartment is the only one that can achieve long-term expansion in organoids obtained from adult healthy mice,[49] such a 'susceptible' cell type might represent an adult progenitor-like cell residing in the pancreatic ducts. This model is supported by the clinical observation that intestinal IPMN are usually localised in the main pancreatic duct, where the contact with environmental carcinogens is more direct than in the periphery of the duct system. In addition, the intestinal differentiation-dysplasia-carcinoma model, possibly involving progenitor-like cells, has been already validated in other tumour types, such as Barrett adenocarcinoma of the oesophagus and gastric cancer.[50] Although the existence of progenitor cells in the human pancreas is still debated,[51] the differential methylation pattern of progenitor genes in intestinal IPMN compared with other lesion types found in this study argues for different adult duct cell types involved in these lesions and includes the activation of mucin-secretion signatures with intestinal reprogramming.

Notably, the comparison between the methylation profiles of low-grade and high-grade precursor lesions showed no apparent difference in methylated regions in the various comparison of lesion subtypes. This analysis was limited by the overall comparatively small numbers of high-grade lesions, which are rarely found in the clinical samples, and the unbalanced group sizes. However, despite these limitations, our results provide no clear evidence for a major role of differentially methylated regions between low-grade and high-grade pancreatic precursor lesions.

Based on these results and on recently published data, a model of development of gastric and intestinal pancreatic precursors can be proposed (figure 6). According to this model, *KRAS* mutations, which are very frequent genetic events both in PanIN and in primary and recurrent IPMN,[52 53] induce gastric reprogramming in pancreatic ductal cells independently

from their localisation. In the peripheral compartment, where PanIN and gastric IPMN are usually localised, progression and malignant transformation are rare events, probably related to Notch signalling activation, to the observed recurrent deletions in gastric IPMN and to the acquisition of additional, possibly subclonal, mutational events not detectable with 'whole lesion' approaches, like the one of the present study. Single cell sequencing has indeed revealed intralesional genetic heterogeneity in gastric IPMN, with subclonal mutations involving the *ARID1A* and *RNF43* genes.[54] These could provide a selective advantage of single cell groups within a definite lesion and explain the 'missing' genetic driver events in a subset of the precursor lesions investigated in the present study. Similarly, a single cell transcriptomics study performed on a small collective of IPMN of different subtypes identified cell clusters in low-grade IPMN with changes in gene expression similar to those found in high-grade lesions.[55] In the main duct compartment on the other hand, exposure to environmental carcinogens and to chronic inflammation induces intestinal differentiation. This can occur de novo or in a previously *KRAS*-mutated and gastric differentiated cell and is associated with higher frequency of recurrent amplifications, as shown in the CNV analysis, and with higher proliferating activity even in low-grade lesions[39 40] (figure 2, online supplemental figure 5). Accordingly, we found that the mucin O-glycan biosynthesis pathway, which has been shown to affect relevant processes of progression and metastasis in human cancer, including pancreatic cancer[42 56] is among the highest differentially regulated pathway between intestinal and gastric IPMN and between intestinal IPMN and PanIN (figure 3C). As shown in figure 6, the occurrence of mixed phenotypes, and also the evolution from gastric to intestinal IPMN, as suggested by some,[16] could be explained by this model. Overall, the observed distinct epigenomic patterns support further exploration of different adult cell compartments in the human pancreas, as well as aetiological and environmental factor analysis as an exciting research area.

This study has some limitations, mainly of methodological type and related to the difficulty of performing multiple, genome-wide analyses on archived paraffin material, which restricted the number of analysed samples on one side and influenced the choice of the type of analysis, for example, targeted versus whole-exome sequencing, on the other. In addition, for similar reasons, intralesional heterogeneity, which has been previously reported,[44] was not addressed in this study and some precursor lesions, such as pancreatobiliary IPMN, had to be excluded due to low case number.

Nevertheless, by applying multiple targeted and genome-wide analyses, we were able to provide the first comprehensive, large-scale molecular analysis of pancreatic cancer precursors with gastric and intestinal phenotype, showing their molecular heterogeneity, which is possibly related to a different cell identity and to a different aetiology. Furthermore, regardless of the above-mentioned technical limitations, we noted several overlaps in targets at the methylation and transcriptome level, as exemplified by *TFF3*, *MUC2* and *MUCL3*, even though these analyses were not all always performed in the same tissue specimens. We therefore strongly believe that studies concerning precursor lesions of PDAC should differentiate between the different entities and subtypes and not consider them as a group. Further studies are needed to better characterise susceptible cell types in the pancreatic ductal compartment as well as to identify potentially removable causes of intestinal reprogramming.

Author affiliations
[1]Bridge Institute of Experimental Tumor Therapy, West German Cancer Center, University Hospital Essen, Essen, Germany
[2]Division of Solid Tumor Translational Oncology, German Cancer Consortium (DKTK, partner site Essen) and German Cancer Research Center, DKFZ, Heidelberg, Germany
[3]Institute of Pathology, Heinrich-Heine University and University Hospital of Dusseldorf, Dusseldorf, Germany
[4]Department of General, Visceral and Pediatric Surgery, Heinrich-Heine-University and University Hospital of Dusseldorf, Dusseldorf, Germany
[5]Institute of Pathology, Technische Universitaet Muenchen, Munich, Germany
[6]HI-STEM—Heidelberg Institute for Stem Cell Technology and Experimental Medicine GmbH, Heidelberg, Germany
[7]Division of Stem Cells and Cancer, German Cancer Research Centre and DKFZ-ZMBH Alliance, Heidelberg, Germany
[8]German Cancer Consortium, (DKTK), Heidelberg, Germany

Correction notice This article has been corrected since it published Online First. The formatting of table 2 has been corrected and figure 3 updated. The open access licence has also been updated to CC BY.

Twitter Irene Esposito @IEspositoPATH

Acknowledgements We thank the microarray unit of the DKFZ Genomics and Proteomics Core Facility for providing the Illumina Human Methylation arrays and related services. PF Cheung and K Savvatakis are acknowledged for their help in preparing FACS-sorted cells from FFPE tissues. We would like to acknowledge the support of the Biobank of the University Hospital of Dusseldorf, Germany.

Contributors S-TL performed the EPIC-based CNV analysis and the analysis of the methylome and transcriptome data and wrote the manuscript. LG analysed the methylome data and the provided comparison with the single-cell data and wrote the manuscript. LF collected the samples, performed the low-pass WGS and the targeted NGS, she prepared the DNA/RNA for further analyses by manual or laser-assisted microdissection and performed part of the immunostaining. LH evaluated the cohort and the immunohistochemical results and wrote the manuscript. AY contributed to morphological and immunohistochemical analyses and wrote the paper. RV performed and evaluated immunohistochemistry. WG and FVO evaluated the panel sequencing and performed and evaluated the fusion transcript analyses. NS contributed to the study design and the planning and interpretation of the WGS experiments. WTK contributed to the study design, to the sample collection and to the data interpretation. AMS contributed to sample collection and analysis and to the study design. GK contributed to the study design and results interpretation and wrote the paper. EE and AT contributed to the methylome analysis and interpretation of the results. JTS planned and supervised the methylome and CNV analysis, contributed to the interpretation of the results and wrote the manuscript. IE conceived the study, planned all experiments, performed the morphological re-evaluation, supervised the WGS and NGS analysis, the sample preparation for further analysis and the immunohistochemistry, interpreted the results, wrote and submitted the manuscript. IE and JTS act as guarantors of the manuscript.

Funding This work was supported by grants of the German Research Foundation (DFG) to IE (grant N° ES285/6-1 and ES 285/8-1). IE is further supported by the European Commission (H2020, Grant Agreement N° 824946–SiMBiT). JTS is supported by the German Cancer Consortium (DKTK), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; #405344257/SI 1549/3-2 and SI1549/4-1), by the German Cancer Aid (#70112505/PIPAC, #70113834/PREDICT-PACA) and by the German Federal Ministry of Education and Research (BMBF; 01KD2206A/SATURN3). Part of this work represents the doctoral thesis of one of the authors (LF).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was approved by the ethic committee of the University Hospital of Dusseldorf, Germany (ethic no. 3821).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement RNAseq data are available in a public, open access repository. EPIC data are available upon reasonable request.

ORCID iDs
Aslihan Yavas http://orcid.org/0000-0002-8408-3063
Friederike V Opitz http://orcid.org/0000-0002-5025-3781
Anna Melissa Schlitter http://orcid.org/0000-0001-6431-2036
Elisa Espinet http://orcid.org/0000-0002-0690-9878
Irene Esposito http://orcid.org/0000-0002-0554-2402

REFERENCES

1 American Cancer Society. *Cancer facts & figures*. Atlanta, 2020.
2 Kleeff J, Korc M, Apte M, *et al*. Pancreatic cancer. *Nat Rev Dis Primers* 2016;2:16022.
3 Balachandran VP, Łuksza M, Zhao JN, *et al*. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* 2017;551:512–6.
4 Stark AP, Sacks GD, Rochefort MM, *et al*. Long-Term survival in patients with pancreatic ductal adenocarcinoma. *Surgery* 2016;159:1520–7.
5 Strobel O, Lorenz P, Hinz U, *et al*. Actual five-year survival after upfront resection for pancreatic ductal adenocarcinoma. *Ann Surg* 2022;275:962–71.
6 Wood LD, Canto MI, Jaffee EM, *et al*. Pancreatic cancer: pathogenesis, screening, diagnosis, and treatment. *Gastroenterology* 2022;163:386–402.
7 Park W, Chawla A, O'Reilly EM. Pancreatic cancer: a review. *JAMA* 2021;326:851–62.
8 Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med Overseas Ed* 2014;371:1039–49.
9 Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J Oncol* 2019;10:10–27.
10 Goggins M, Overbeek KA, Brand R, *et al*. Management of patients with increased risk for familial pancreatic cancer: updated recommendations from the International cancer of the pancreas screening (CAPS) Consortium. *Gut* 2020;69:7–17.
11 Tomasetti C, Li L, Vogelstein B. Cancer etiology, and cancer prevention. *Science* 2017;355:1330–4.
12 WHO Classification of Tumours Editorial Board (eds.). WHO Classification of Tumours. In: *Digestive system tumours*. 5th Edition, Volume 1. Lyon, 2019.
13 Klöppel G, Basturk O, Schlitter AM, *et al*. Intraductal neoplasms of the pancreas. *Semin Diagn Pathol* 2014;31:452–66.
14 Hruban RH, Adsay NV, Albores-Saavedra J. Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. *Am J Surg Pathol* 2001;25:579–86.
15 Peters MLB, Eckel A, Mueller PP, *et al*. Progression to pancreatic ductal adenocarcinoma from pancreatic intraepithelial neoplasia: results of a simulation model. *Pancreatology* 2018;18:928–34.
16 Omori Y, Ono Y, Kobayashi T, *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* 2020;477:21–31.
17 Roth S, Zamzow K, Gaida MM, *et al*. Evolution of the immune landscape during progression of pancreatic intraductal papillary mucinous neoplasms to invasive cancer. *EBioMedicine* 2020;54:102714.
18 Matthaei H, Wu J, Dal Molin M, *et al*. GNAS sequencing identifies IPMN-specific mutations in a subgroup of diminutive pancreatic cysts referred to as "incipient IPMNs". *Am J Surg Pathol* 2014;38:360–3.
19 Matthaei H, Hong S-M, Mayo SC, *et al*. Presence of pancreatic intraepithelial neoplasia in the pancreatic transection margin does not influence outcome in patients with R0 resected pancreatic cancer. *Ann Surg Oncol* 2011;18:3493–9.
20 Marchegiani G, Mino-Kenudson M, Ferrone CR, *et al*. Patterns of recurrence after resection of IPMN: who, when, and how? *Ann Surg* 2015;262:1108–14.
21 Yachida S, Jones S, Bozic I, *et al*. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 2010;467:1114–7.
22 Rhim AD, Mirek ET, Aiello NM, *et al*. Emt and dissemination precede pancreatic tumor formation. *Cell* 2012;148:349–61.
23 Notta F, Chan-Seng-Yue M, Lemire M, *et al*. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 2016;538:378–82.
24 Noë M, Niknafs N, Fischer CG, *et al*. Genomic characterization of malignant progression in neoplastic pancreatic cysts. *Nat Commun* 2020;11:4085.
25 Basturk O, Hong S-M, Wood LD, *et al*. A revised classification system and recommendations from the Baltimore consensus meeting for neoplastic precursor lesions in the pancreas. *Am J Surg Pathol* 2015;39:1730–41.
26 ZMc HV. Enhanced copy-number variation analysis using Illumina DNA methylation arrays. R package version 1.9.0. Available: http://bioconductor.org/packages/conumee/
27 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
28 Tian Y, Morris TJ, Webster AP, *et al*. Champ: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 2017;33:3982–4.
29 Espinet E, Gu Z, Imbusch CD, *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* 2021;11:638–59.
30 Moss J, Magenheim J, Neiman D, *et al*. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* 2018;9:5068.
31 Tosti L, Hang Y, Debnath O, *et al*. Single-nucleus and in situ RNA-sequencing reveal cell topographies in the human pancreas. *Gastroenterology* 2021;160:1330–44.
32 Miyamoto Y, Maitra A, Ghosh B, *et al*. Notch mediates TGF alpha-induced changes in epithelial differentiation during pancreatic tumorigenesis. *Cancer Cell* 2003;3:565–76.
33 Mazur PK, Einwächter H, Lee M, *et al*. Notch2 is required for progression of pancreatic intraepithelial neoplasia and development of pancreatic ductal adenocarcinoma. *Proc Natl Acad Sci U S A* 2010;107:13438–43.
34 Hidalgo-Sastre A, Brodylo RL, Lubeseder-Martellato C, *et al*. Hes1 controls exocrine cell plasticity and restricts development of pancreatic ductal adenocarcinoma in a mouse model. *Am J Pathol* 2016;186:2934–44.
35 Hingorani SR, Wang L, Multani AS, *et al*. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* 2005;7:469–83.
36 Hingorani SR, Petricoin EF, Maitra A, *et al*. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* 2003;4:437–50.
37 Li J, Wei T, Zhang J, *et al*. Intraductal papillary mucinous neoplasms of the pancreas: a review of their genetic characteristics and mouse models. *Cancers* 2021;13:5296.
38 De La O J-P, Emerson LL, Goodman JL, *et al*. Notch and KRAS reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. *Proc Natl Acad Sci U S A* 2008;105:18907–12.
39 Hovhannisyan G, Harutyunyan T, Aroutiounian R, *et al*. Dna copy number variations as markers of mutagenic impact. *Int J Mol Sci* 2019;20:4723.
40 Arlt MF, Wilson TE, Glover TW. Replication stress and mechanisms of CNV formation. *Curr Opin Genet Dev* 2012;22:204–10.
41 Yan J, Chen G, Zhao X, *et al*. High expression of diffuse panbronchiolitis critical region 1 gene promotes cell proliferation, migration and invasion in pancreatic ductal adenocarcinoma. *Biochem Biophys Res Commun* 2018;495:1908–14.
42 Gupta R, Leon F, Rauth S, *et al*. A systematic review on the implications of O-linked glycan branching and truncating enzymes on cancer progression and metastasis. *Cells* 2020;9:446.
43 Mateos RN, Nakagawa H, Hirono S, *et al*. Genomic analysis of pancreatic juice DNA assesses malignant risk of intraductal papillary mucinous neoplasm of pancreas. *Cancer Med* 2019;8:4565–73.
44 Fischer CG, Beleva Guthrie V, Braxton AM, *et al*. Intraductal papillary mucinous neoplasms arise from multiple independent clones, each with distinct mutations. *Gastroenterology* 2019;157:1123–37.
45 Muraro MJ, Dharmadhikari G, Grün D, *et al*. A single-cell transcriptome atlas of the human pancreas. *Cell Syst* 2016;3:385–94.
46 Adachi T, Tajima Y, Kuroki T, *et al*. Bile-reflux into the pancreatic ducts is associated with the development of intraductal papillary carcinoma in hamsters. *J Surg Res* 2006;136:106–11.
47 Muraki T, Reid MD, Pehlivanoglu B, *et al*. Variant anatomy of the biliary system as a cause of pancreatic and peri-ampullary cancers. *HPB* 2020;22:1675–85.
48 Kiss B, Mikó E, Sebő Éva, *et al*. Oncobiosis and microbial metabolite signaling in pancreatic adenocarcinoma. *Cancers* 2020;12:1068.
49 Huch M, Bonfanti P, Boj SF, *et al*. Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *Embo J* 2013;32:2708–21.
50 Hayakawa Y, Nakagawa H, Rustgi AK, *et al*. Stem cells and origins of cancer in the upper gastrointestinal tract. *Cell Stem Cell* 2021;28:1343–61.
51 Alvarez Fallas ME, Pedraza-Arevalo S, Cujba A-M, *et al*. Stem/Progenitor cells in normal physiology and disease of the pancreas. *Mol Cell Endocrinol* 2021;538:111459.
52 Hata T, Suenaga M, Marchionni L, *et al*. Genome-Wide somatic copy number alterations and mutations in high-grade pancreatic intraepithelial neoplasia. *Am J Pathol* 2018;188:1723–33.
53 Pea A, Yu J, Rezaee N, *et al*. Targeted DNA sequencing reveals patterns of local progression in the pancreatic remnant following resection of intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Ann Surg* 2017;266:133–41.
54 Kuboki Y, Fischer CG, Beleva Guthrie V, *et al*. Single-cell sequencing defines genetic heterogeneity in pancreatic cancer precursor lesions. *J Pathol* 2019;247:347–56.
55 Bernard V, Semaan A, Huang J, *et al*. Single-Cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. *Clin Cancer Res* 2019;25:2194–205.
56 Pan S, Brentnall TA, Chen R. Glycoproteins and glycoproteomics in pancreatic cancer. *World J Gastroenterol* 2016;22:9288–99.