The genomic landscape of pancreatic and periampullary adenocarcinoma

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

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

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

Abstract

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

Introduction

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

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

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

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

Results

1. Genomic aberrations in periampullary adenocarcinomas

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

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

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

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

2. Clinicopathological characteristics of periampullary adenocarcinomas

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

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

3. Copy number changes characterizing periampullary adenocarcinomas with different morphology

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

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

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

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

4. Driver genes in periampullary adenocarcinomas

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

5. Integrated analysis of copy number and gene expression data

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

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

6. Co-occurrence of chromosomal aberrations in periampullary adenocarcinomas

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

7. Gene Set Enrichment Analysis

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

8. Prognostic implications of copy number gain

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

Discussion

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

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

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

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

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

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

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

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

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

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

Material and Methods

DNA extraction

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

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

Tumor and matching normal samples

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

Affymetrix SNP 6.0 arrays

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

Data analysis

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

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

Frequency plots

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

mRNA expression analysis

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

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

Driver genes in amplified and deleted regions

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

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

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

Correlation analysis of copy number aberrations and gene expression data

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

Gene Set Enrichment Analysis

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

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

Survival analysis

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

20. Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A et al., 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801-1806.

21. Kao J, Pollack JR. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer. 2006; 45:761-9.

22. Kim TM, Yim SH, Lee JS, Kwon MS, Ryu JW, Kang HM, Fiegler H, Carter NP, Chung YJ. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. Clin Cancer Res. 2005;11:8235–824.

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

24. Maire, V., Nemati, F., Richardson, M., Vincent-Salomon, A., Tesson, B., Rigaill, G., Gravier, E., Marty-Prouvost, B., De Koning, L., Lang, G. et al., 2013. Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. Cancer research 73, 813-823.

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

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

27. Natrajan, R., Mackay, A., Wilkerson, P.M., Lambros, M.B., Wetterskog, D., Arnedos, M., Shiu, K.K., Geyer, F.C., Langerod, A., Kreike, B. et al., 2012. Functional characterization of the 19q12 amplicon in grade III breast cancers. Breast cancer research : BCR 14, R53.

28. Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M. et al., 2012. The life history of 21 breast cancers. Cell 149, 994-1007.

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

30. P. Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman & Michael R. Stratton (2004). A census of human cancer genes2Nature Reviews Cancer 4, 177-183 | doi:10.1038/nrc1299.

31. Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z. et al., 2009. Supervised risk predictor of breast cancer based

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

32. Prat, A., Adamo, B., Fan, C., Peg, V., Vidal, M., Galvan, P., Vivancos, A., Nuciforo, P., Palmer, H.G., Dawood, et al., 2013. Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. Scientific reports 3, 3544.

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

34. Sandhu, V., Bowitz Lothe, I.M., Labori, K.J., Lingjaerde, O.C., Buanes, T., Dalsgaard, A.M., Skrede, M.L., Hamfjord, J., Haaland, T., Eide, T.J. et al., 2015. Molecular signatures of mRNAs and miRNAs as prognostic biomarkers in pancreatobiliary and intestinal types of periampullary adenocarcinomas. Molecular oncology 9, 758-771.

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

36. Tang TC, Sham JS, Xie D, Fang Y, Huo KK, Wu QL, Guan XY. Identification of a candidate oncogene SEI-1 within a minimal amplified region at 19q13.1 in ovarian cancer cell lines. Cancer Res. 2002;62:7157–7161.

37. Tiseo, M., Bordi, P., Bortesi, B., Boni, L., Boni, C., Baldini, E., Grossi, F., Recchia, F., Zanelli, F., Fontanini, G. et al., 2013. ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. British journal of cancer 108, 1695-1703.

38. Waddell, N., Pajic, M., Patch, A.M., Chang, D.K., Kassahn, K.S., Bailey, P., Johns, A.L., Miller, D., Nones, K., Quek, K.,et al., 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518, 495-501.

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

40. Wennerstrom, A.B., Lothe, I.M., Sandhu, V., Kure, E.H., Myklebost, O., Munthe, E., 2014. Generation and characterisation of novel pancreatic adenocarcinoma xenograft models and corresponding primary cell lines. PloS one 9, e103873.

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

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

43. Zheng, H.Q., Zhou, Z., Huang, J., Chaudhury, L., Dong, J.T., Chen, C., 2009. Kruppel-like factor 5 promotes breast cell proliferation partially through upregulating the transcription of fibroblast growth factor binding protein 1. Oncogene 28, 3702-3713.

44. Zhou, Y., Han, C., Li, D., Yu, Z., Li, F., Li, F., An, Q., Bai, H., Zhang, X., Duan, Z., et al., 2015. Cyclin-dependent kinase 11(p110) (CDK11(p110)) is crucial for human breast cancer cell proliferation and growth. Scientific reports 5, 10433.

Figures and tables legends:

Figure 1: The frequency of copy number aberrations in the OUH and the TCGA cohort. The x-axis represents the genomic position while the y-axis represents the frequencies of amplification and deletion. The red and green bars in the figure show amplifications and deletions of the chromosomes. Note the overrepresentations in deletions in both cohorts.

Figure 2: Frequency of copy number alterations in periampullary adenocarcinomas.
2A) The frequency plot of the genomic aberration pattern for the pancreatobiliary and intestinal subtypes. The x-axis represents the genomic position and is divided into 22 facets for the 22 chromosomes. The y-axis represents the frequency of chromosomal gains and losses for the two subtypes based on morphology.
2B) The radial plot shows the relative percentage of chromosomal copy number gains and losses based on morphology. The cytobands are marked in red and green for gains and losses, respectively. The color bars refer to the morphology of the samples.
2C) The radial plot shows the percentage of copy number gains and losses based on site of origin. The cytobands are marked in red and green for gains and losses, respectively. The color bars refer to the site of origin of the sample.

Figure 3: The figure shows the average frequencies of amplifications and deletions of genes in both the OUH and TCGA cohorts. The x-axis shows the frequencies of amplifications and deletions of genes, and the color bars on the y-axis represents the chromosomes. The cytobands and genes are marked in red and green for amplifications and deletions, respectively.

Figure 4: Kaplan Meier OS and DFS for patients carrying PA tumors in the OUH cohort 4 A) Kaplan Meier DFS curves according to 18p11.22 chromosomal locus status. 4 B-C) Kaplan Meier DFS curves based on *RAB12* and *COLEC12* gene expression status. The genes are located in loci 18p11.22 and 18p11.32, respectively. 4 D) Kaplan Meier OS curves according to 19q13.2 chromosomal status. 4 E-F) Kaplan Meier OS curves based on *SERTAD3* and *ERCC1* gene expression status. The genes are located on 19q13.2 and 19q13.32, respectively. In Figure 4 D, tumors carrying deletions and tumors without

19q13 alterations (normal samples) were combined in the survival analysis due to limited number of tumors with deletion events (n = 4).

Table 1: Clinical features of the PAs from the OUH cohort (n = 55) and from the TCGA cohort (n = 127). Note: For comparative reasons, the overall and disease free survival in patients from the OUH cohort were only calculated for PDAC since the TCGA cohort is primarily composed of PDAC tumors.

Table 2: Pathway analysis using frequently amplified and deleted genes in the OUH and the TCGA cohorts. The table shows the enriched pathways, number of genes enriched in the pathway, *P*-value and FDR corrected *P*-value using the Benjamini and Hochberg correction.

Figure 1

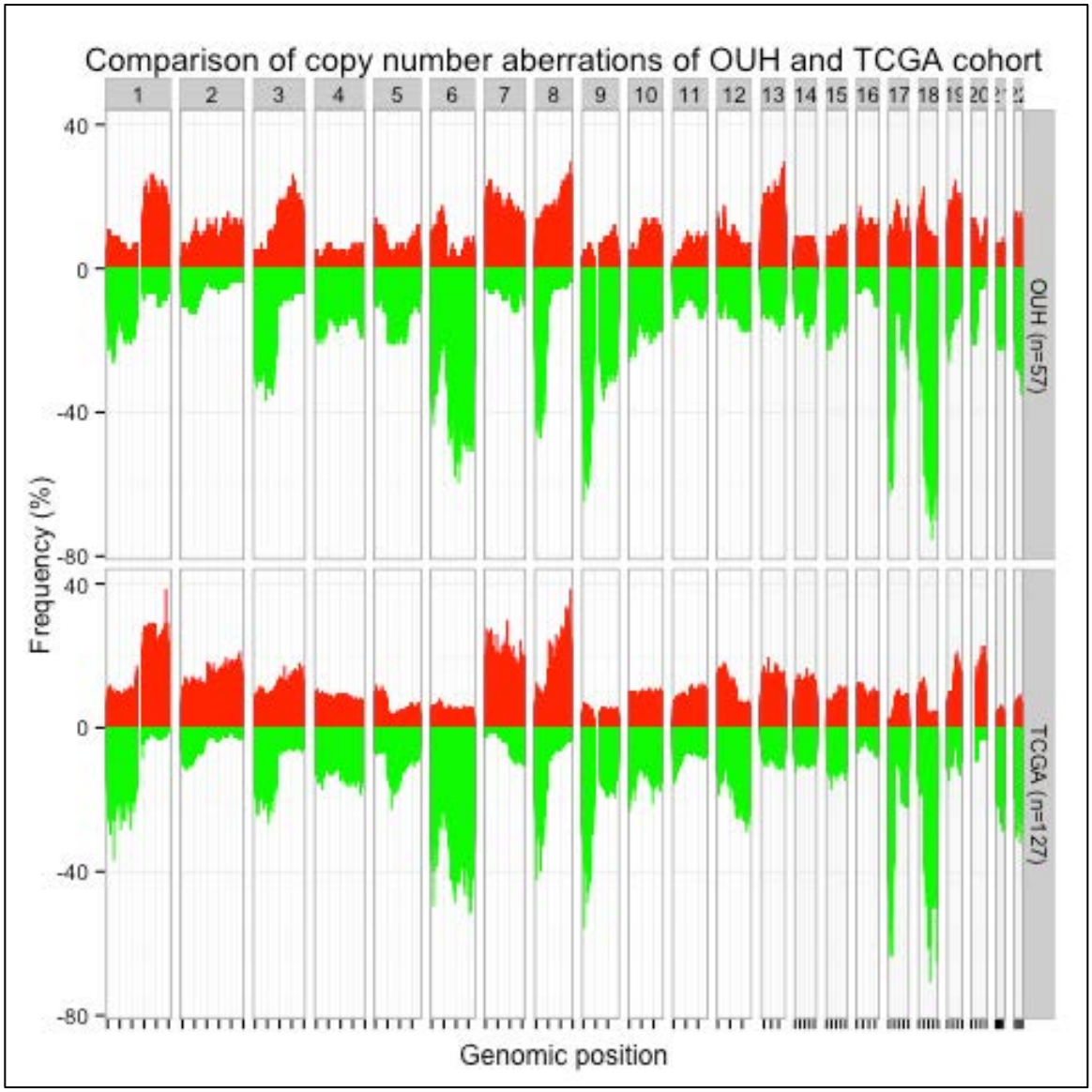
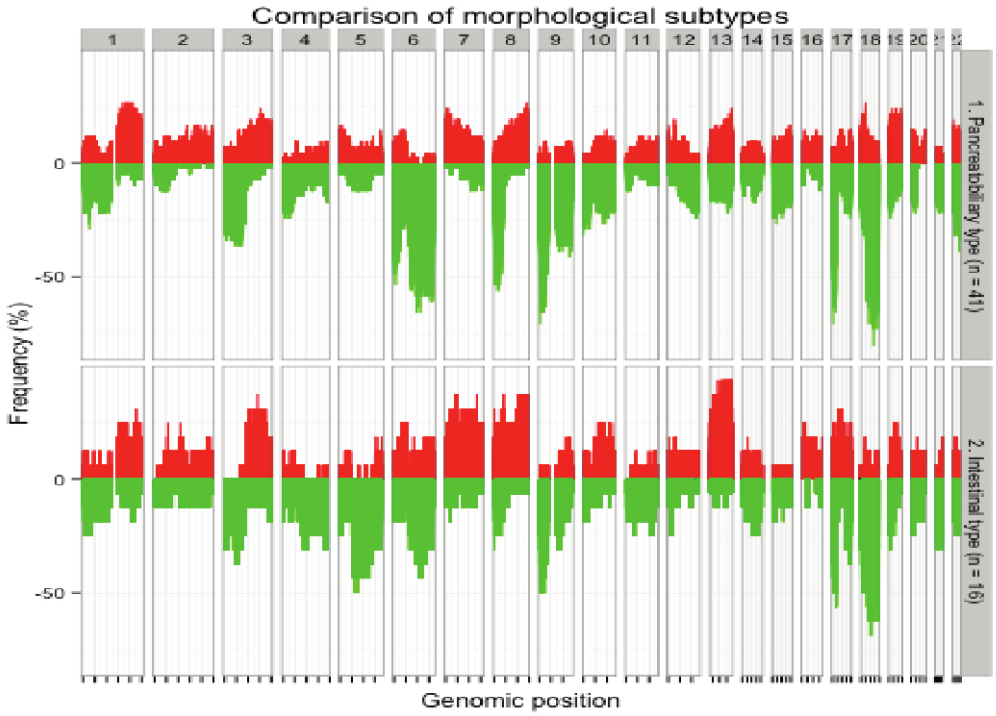
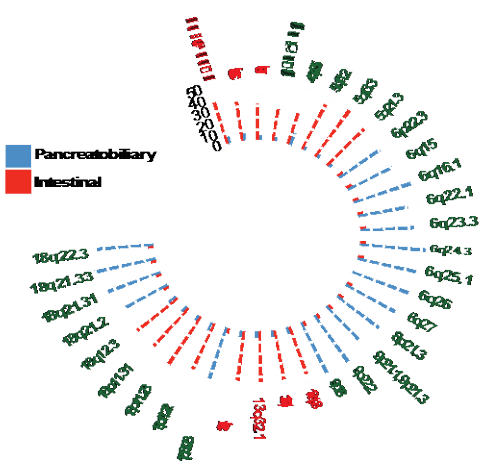

Comparison of copy number aberrations of OUH and TCGA cohort
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
40
0
-40
-80
OUH (n=57)
TCGA (n=127)
Frequency (%)
Genomic position

Figure 2

2A.



2B. Morphology based



2C. Site of origin based

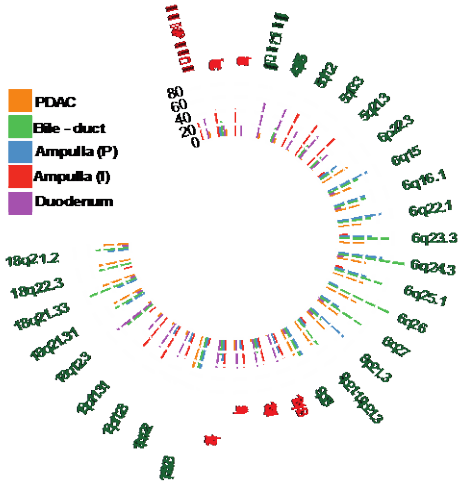


Figure 3

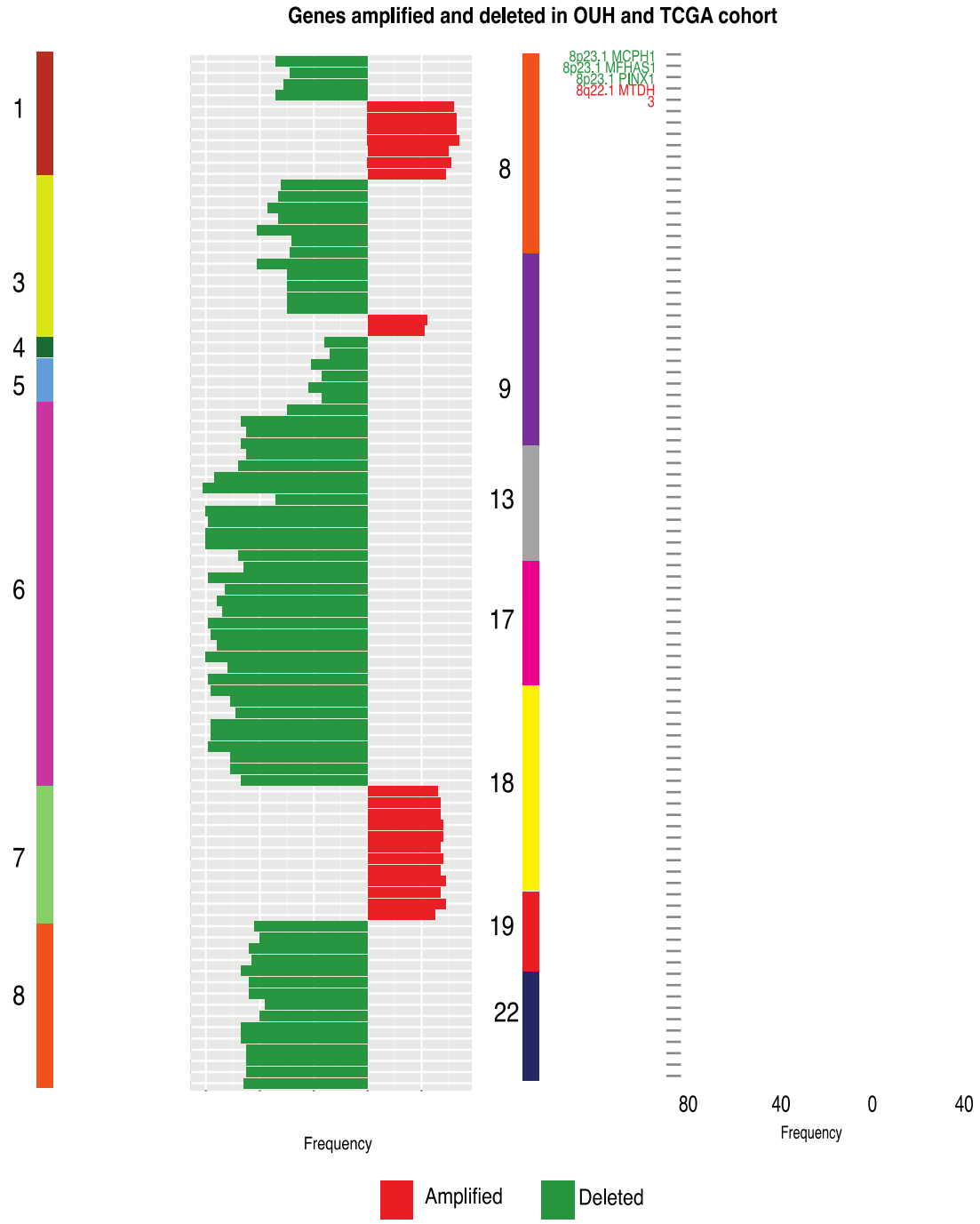

Genes amplified and deleted in OUH and TCGA cohort
1
3
4
5
6
7
8
8
9
13
17
18
19
22
8p23.1 MCPH1
8p23.1 MFHAS1
8p23.1 PINX1
8q22.1 MTDH
3
80
40
0
40
Frequency
Frequency
Amplified
Deleted

Figure 4

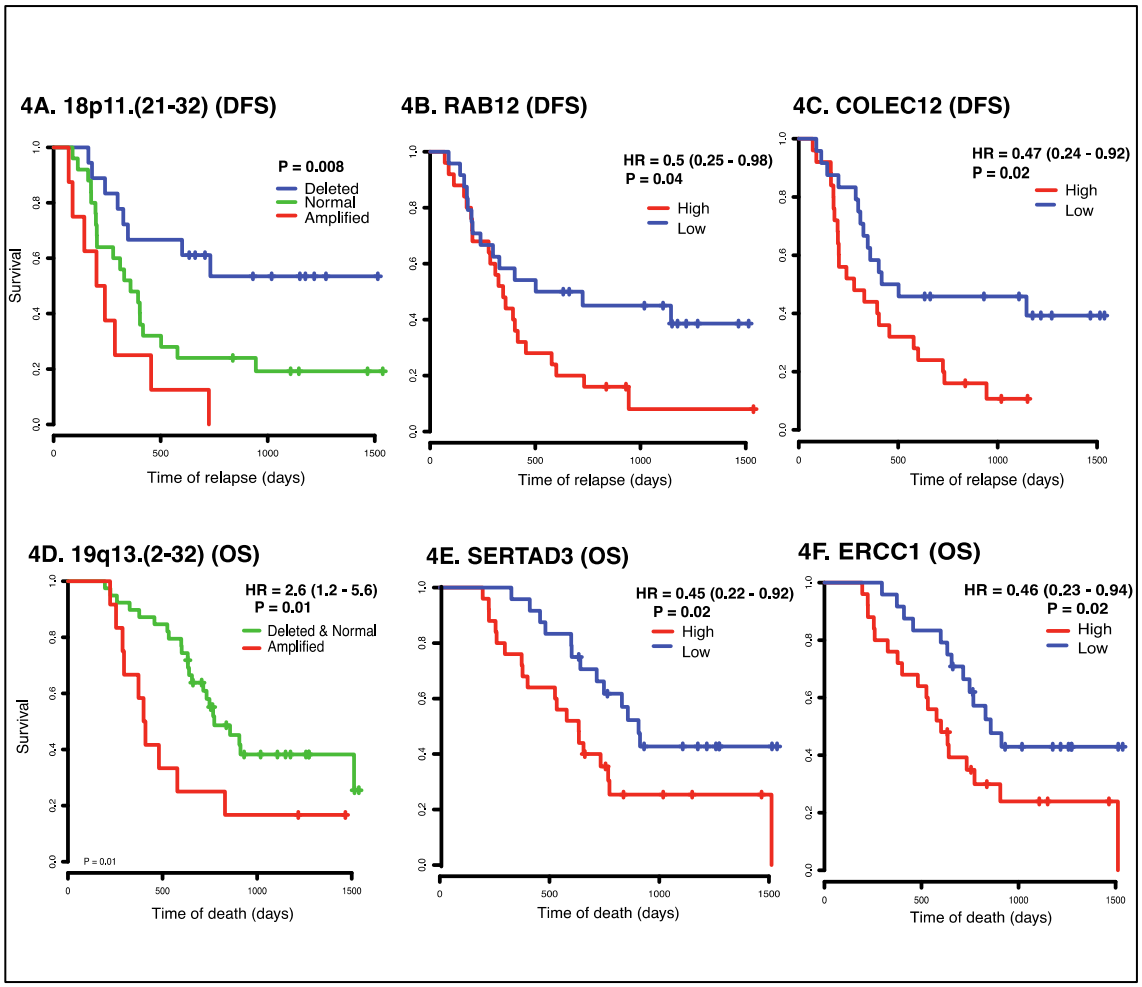

4A. 18p11.(21-32) (DFS)
P = 0.008
Deleted
Normal
Amplified
Survival
Time of relapse (days)
4B. RAB12 (DFS)
HR = 0.5 (0.25 - 0.98)
P = 0.04
High
Low
Time of relapse (days)
4C. COLEC12 (DFS)
HR = 0.47 (0.24 - 0.92)
P = 0.02
High
Low
Time of relapse (days)
4D. 19q13.(2-32) (OS)
HR = 2.6 (1.2 - 5.6)
P = 0.01
Deleted & Normal
Amplified
P = 0.01
Survival
Time of death (days)
4E. SERTAD3 (OS)
HR = 0.45 (0.22 - 0.92)
P = 0.02
High
Low
Time of death (days)
4F. ERCC1 (OS)
HR = 0.46 (0.23 - 0.94)
P = 0.02
High
Low
Time of death (days)

Table 1

Clinical Features		The OUH cohort (n=55) Frequency (percentage)	The TCGA cohort (n=127) Frequency (percentage)
Gender	Female	30 (55)	52 (41)
	Male	25 (45)	75 (59)
Type	PDAC	29 (52)	111(87)
	PDAC-other subtypes	-	16(13)
	Bile duct	4 (7)	-
	Ampulla pancreatobiliary type	6 (10)	-
	Ampulla intestinal type	7 (12)	-
	Duodenum	9 (16)	-
pT	T1	4 (7)	2(1)
	T2	9 (16)	11(9)
	T3	36 (65)	110(87)
	T4	6 (11)	3(2)
	TX	-	1(1)
N	N0	19 (35)	34(27)
	N1	35 (64)	91(72)
	N2	1 (2)	0(0)
	NX	-	2(1)
M	M0	52 (95)	59(47)
	M1	3 (5)	4(3)
	MX	-	64(50)
R	R0	37 (67)	68(54)
	R1	18 (33)	42(33)
	R2	-	4(3)
	RX	-	4(3)
	Not available	-	9(7)
Differentiation/ Grade	Well (G1)	18 (33)	14(11)
	Moderately (G2)	37 (67)	72(57)
	Poor (G3)	-	39(31)
	Undetermined (GX)	-	2(1)
Mean overall survival		578 days	247 days
Median disease free survival		259 days	-
Median age		65 years	66 years

Table 2

	OUH cohort			TCGA cohort		
Pathways	Number of genes	*P*-value	Adjusted *P*-value	Number of genes	*P*-value	Adjusted *P*-value
MAPK signaling	44	6.83E-07	5.18E-06	65	2.2E-13	3.1E-12
Jak-STAT signaling	29	3.86E-06	1.95E-05	44	6.0E-12	5.2E-11
Cell cycle	22	0.0001	0.0003	27	1.7E-05	3.8E-05
p53 signaling	13	0.0014	0.0026	20	1.8E-06	5.9E-06
Apoptosis	14	0.005	0.0081	25	1.5E-07	6.9E-07
Insulin signaling	19	0.0071	0.0104	22	0.007	0.008
TGF-beta signaling	12	0.0219	0.0269	16	0.003	0.004
Wnt signaling	18	0.0312	0.0374	34	5.8E-07	2.2E-06