Genomic Landscape of Pancreatic Adenocarcinoma in Younger vs Older Patients: Does Age Matter?

Irit Ben-Aharon¹, Moshe Elkabets², Raphael Pelosi³, Kenneth H. Yu⁴,⁵,⁶, Christine A. Iacubuzio-Donahue⁵,⁷, Steven D. Leach⁸, Maeve A. Lowery⁹, Karyn A. Goodman¹⁰, Eileen M. O'Reilly⁴,⁵,⁶

1. Division of Oncology, Rambam Health Care Center, Haifa 31096, Israel
2. Shraga Segal Department of Microbiology, Immunology and Genetics, Faculty of Health Sciences Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel
3. Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA
4. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA
5. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA
6. Weill Cornell Medical College, New York, NY 10065, USA
7. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA
8. Geisel School of Medicine, Dartmouth, Hanover, New Hampshire 03755-1404, USA
9. Trinity St James Cancer Institute, Trinity College Dublin, Ireland
10. Department of Radiation Oncology, University of Colorado School of Medicine, Aurora, CO 80045, USA

Corresponding Author:
Eileen M. O’Reilly, MD
Memorial Sloan Kettering Cancer Center, New York, NY 10065
Telephone: 646 888-4182
Facsimile: 646 888-4542
Email: oreillye@mskcc.org
Abstract: 240
Tables: 1
Figures: 4

Running Title: Pancreatic adenocarcinoma genomics, transcriptome and age

Key words: Pancreatic adenocarcinoma, early age, TGF beta

Disclosure of Potential Conflicts of Interest

E.M. O’Reilly: Funding support to MSK: Genentech, Roche, BMS, Halozyme, Celgene, MabVax Therapeutics, ActaBiologica.
Consulting: Cytomx, BioLineRx, Targovax, Halozyme, Celgene
No potential related conflicts of interest were disclosed by the authors.
Irit Ben-Aharon¹, Moshe Elkabets², Raphael Pellossof³, Kenneth H. Yu⁴,⁵,⁶, Christine A. Iacubuzio-Donahue⁵,⁷, Steven D. Leach⁸, Maeve A. Lowery⁹, Karyn A. Goodman¹⁰
Purpose: State-of-the-art genomic analyses of pancreatic adenocarcinoma (PDAC) have yielded insight into signaling pathways underlying carcinogenesis. PDAC is characterized by substantial genomic heterogeneity. We aimed to determine if early-onset PDAC (EOPC; ≤ 55 years) displays a distinctive molecular landscape from average-age onset PDAC (AOPC; ≥ 70 years).

Experimental Design: Three distinct datasets for PDAC were analyzed. In the first, patients undergoing treatment at Memorial Sloan Kettering (MSK) were consented for MSK-IMPACT next generation sequencing. The second cohort analyzed was The Cancer Genome Atlas (TCGA) dataset for differences in somatic mutations, gene expression and protein expression. The third dataset was an Australian cohort of PDAC. Clinical data were correlated with genomic analyses.

Results: Two hundred and ninety-three samples were analyzed, yielding 90 patients (pts) aged ≤ 55 years and 203 pts aged ≥ 70 years. Among the genes known to be associated with carcinogenesis SMAD4 displayed higher mutation rates in younger patients. Comprehensive transcriptomic analysis of cellular pathways indicated that the TGFb pathway has increased activation and the expression levels of phospo-GSK3 were higher in EOPC. Survival outcomes revealed no differences between age groups.

Conclusions: These exploratory analyses suggest that there may be somatic gene alterations within the population of early onset PDAC patients that involve unique cellular pathways compared with average onset PDAC. Former studies imply these cellular pathways may play a role in smoking-related PDAC carcinogenesis. Larger genomic datasets are warranted for future evaluation to extend these observations.
Translational Relevance:
Pancreatic adenocarcinoma (PDAC) is characterized by substantial genomic heterogeneity. We aimed to determine if early-onset pancreatic cancer (EOPC) displays a distinctive molecular landscape from average-age onset pancreatic cancer (AOPC). Three distinct genomic datasets for PDAC were analyzed for differences in somatic mutations, gene expression and protein expression levels. Genomic data and analyses were correlated with clinical data and outcomes. We identified N= 56 patients aged ≤ 55 years and N= 127 pts aged ≥ 70 years. SMAD4 displayed higher mutation rates in EOPC as well as increased activation of TGFb pathway. Our exploratory analysis suggests that there may be somatic gene alterations within the population of EOPC that involve unique cellular pathways compared with AOPC. In an era of precision medicine, EOPC may represent a unique biological entity that should be further explored for potentially optimal treatment tailoring.
Introduction:

Pancreatic cancer (Pancreatic ductal adenocarcinoma; PDAC) is a highly lethal malignancy with a five-year mortality of 97–98%, usually due to metastatic disease (1,2). The challenging prognosis is attributed to the late stage at which most patients are diagnosed, along with the lack of effective systemic therapies to control the disease (2). PDAC is a disease of the elderly, with the median age of diagnosis of PDAC in the US being 71 years (1). Nearly 5% of patients diagnosed with PDAC are under the age of 50 years, referred as early-onset pancreatic cancer (EOPC; http://seer.cancer.gov/statfacts/html/pancreas.html; (2)). In the current Surveillance, Epidemiology, and End Results (SEER) data, there is a trend towards an increase in the rate of EOPC over the last two decades, as depicted in Figure 1, data obtained from the SEER website. There is a paucity of data regarding the clinicopathological features of EOPC. Several retrospective studies have described cohorts of EOPC, and indicated a trend for improved survival in the EOPC compared with patients with average onset pancreatic cancer (AOPC), although there are conflicting studies as well with smaller sample sizes (3-5). Newer studies have implied that EOPC is significantly more prevalent in males (ratio male/female >2 below the age of 50 years compared with <1.4 above the age of 60 years). Moreover, smoking has been found to be a key player in the pathogenesis of EOPC (6,7). The genomic characterization of PDAC has greatly evolved recently and revealed many of the genetic alterations that underline distorted signaling pathways in PDAC (8). The genomic landscape of PDAC is defined by four mutational pathways that are essential drivers of carcinogenesis, KRAS, TP53, CDKN2A and SMAD4, although additional candidate driver genes that may be attributed to either sporadic or familial PDAC occur in a minority of cases (such as BRCA, ATM and PALB2). The key mutation in PDAC is activated KRAS (>95%) which occurs early in carcinogenesis, while mutations in SMAD4 occurs later and are thought to facilitate the progression of PDAC (9). Loss of Smad4 protein correlates with worse clinical outcomes, although some data are conflicting (10,11). Molecular characterization may shed insight for newer therapeutic targets that may improve clinical care and outcomes. A better understanding of the molecular and genomic landscape of PDAC will facilitate allocation of
distinctive phenotypes and as a result to the definition of subgroups that will potentially benefit from specific therapies.

In light of putated differences in young patients who develop PDAC in comparison to a more typical age group, and with mounting genetic and pathologic evidence that had been gathered in the last decade, we sought to evaluate whether EOPC displays differing clinical characteristics, genomic landscape and survival outcomes compared to AOPC utilizing three distinct cohorts, which used exome sequencing and copy number analysis to define genomic aberrations.

Methods

Determining Old and Young PDAC Patient Cohorts

For statistical analysis (described below) we defined an “early-onset” (EOPC; < 55 years) and “average-onset” (AOPC; > 70 years) groups of patients at time of diagnosis. The intervening patients were classified as an “intermediate” group. We hypothesized that the biology of EOPC and AOPC might differ and that the intermediate group might represent a mixture of these biological entities. We have excluded the intermediate group from the analyses reported herein.

Memorial Sloan Kettering Cohort

Patients who were treated at Memorial Sloan Kettering (MSK) and had a confirmed diagnosis of PDAC who provided informed consent to participate in genetic profiling of their tumor under NCT 01775072 “Tumor Genomic Profiling in Patients Evaluated for Targeted Cancer Therapy” between March 2014 and April 2016 were enrolled into the study. Pancreatic tumor samples were profiled using the MSK-IMPACT custom assay, designed for targeted sequencing of all exons and selected introns of 410 oncogenes, tumor suppressor genes, and members of pathways deemed actionable by targeted therapies, as previously described (12). Clinical data collected including demographics, family history of malignancy, smoking history, pre-existing...
diabetes diagnosis and survival outcomes. Institutional Review and Privacy Board oversight was obtained for the MSK dataset.

**TCGA Cohort**

All analyses were performed on The Cancer Genome Atlas (TCGA) publicly available dataset. Eligible patients were those who were defined as having PDAC in the TCGA dataset and who had complete information on age at PDAC diagnosis, tumor histology, clinical demographics, survival, and had DNA analysis, gene expression and reverse protein phase array (RPPA) analyses available. Mutation and copy number data were obtained from the TCGA pan-cancer project (8). Analysis of the gene expression pathway was explored using the Gene Set Enrichment Analysis (GSEA) website (software.broadinstitute.org/gsea).

**Australian Cohort**

We included an Australian cohort in our analysis (13). Due to the small sample size of young-onset patients in this cohort, it was not included in the transcriptomic analysis dataset, but for clinical data only – patient clinicopathological demographics and survival analysis.

**Statistical Analyses**

Mutation type frequencies were determined for each patient, excluding indels and multiple-base mutations. Group frequencies were determined by averaging the individual patient frequencies: this prevented skewing of the data by patients with large numbers of mutations. The statistical significance of mutation and copy number frequencies was determined by Fisher's exact test. Statistical analysis for survival was performed using the statistical software SPSS version 17.0 for Windows (SPSS, Chicago, IL, USA). Survival was estimated using the Kaplan–Meier method and survival curves were compared using the log-rank test.

**Results**

**Patients**
From the three genomic cohorts, N= 293 patients were included in the analysis: N= 90 early-onset patients (EOPC) and N= 203 average-onset patients (AOPC). Thirty-seven EOPC patients had metastatic disease, all from the MSK cohort. There was a slight female predominance in the young cohort (53/90; 59%), though not statistically significant. The most common primary tumor location was the head of the pancreas in both groups. There was not a significant difference in the prevalence of smoking between the two groups (43/90; 48% never smokers in the EOPC group compared with 88/203; 43%, never-smokers in the AOPC). The analysis of family history was performed only for the MSK cohort, since data was not available for the Australian cohort, and there was a substantial number of cases with “unknown family history” in the TCGA cohort that may confound the results. For the MSK cohort, the EOPC group had a significantly higher rate of having a family history of cancer: 54% versus 39% (p= 0.05). There was no increased rates of pancreatitis or diabetes in the EOPC compared with the AOPC. The prevalence of diabetes was 13.2% in the EOPC and 20.3% in the AOPC groups. Patient characteristics of the three cohorts are presented in Table 1.

Genomic Landscape

Figure 2a illustrates the genomic landscape related to age for the TCGA cohort, and figure 2b for the MSK cohort. Consistent with prior molecular studies of PDAC, KRAS mutations were the most prevalent type in both cohorts, followed by TP53 and CDKN2A. In the MSK cohort, KRAS mutations were significantly less frequent in the EOPC cohort (83% versus 96%; p= 0.016); there was no significant difference in the TCGA dataset.

Mutations in the SMAD4 gene were significantly more prevalent in the EOPC group in both TCGA and MSK cohorts (p= 0.02). There was a trend of higher prevalence of mutations in PIK3CA gene in EOPC group in the TCGA cohort (p= 0.07) though did not differ across ages in the MSK cohort. Mutations in BRCA1 and BRCA2 were not more prevalent in EOPC.

TGFbeta Signaling is Associated with EOPC

Figure 3 describes the signaling pathways associated with age. Briefly, we utilized the RNAseq data available from the TCGA pancreatic tumors to further characterize the differences
between tumor gene expression signatures and genes. KEGG Gene enrichment set analysis was assessed per patient, and the enrichment score was calculated for each signature. Then a correlation between the enrichment score and the patient’s age was performed. Among the tested signatures, we identified four signaling pathways that were statistically associated with age (Figure 3a): Hedgehog, TGF-beta, Hypoxia and UV (ultraviolet) response. TGF-beta signaling was the second most significant signature, after hedgehog (Figure 3b-c).

To support our findings, we analyzed the reverse-protein-phase array (RPPA) data base available in the TCGA cohort. Comparing the TGF-beta pathway related protein, GSK3, we observed a higher expression of phosphorylated GSK3 in the two residues 9 and 21 in the EOPC cohort, as depicted in Figure 4.

**Survival**

Kaplan–Meier method and survival curves were performed only for the TCGA cohort and Australian cohort, due to lack of sufficient follow up interval for the MSK cohort. Survival analyses revealed no differences between the age groups. Survival curves are presented in Figure 5.
Discussion

State-of-the-art genomic analysis of PDAC has yielded new insights related to underlying carcinogenic signaling pathways (12-13). PDAC is characterized by substantial genomic homogeneity in that virtually all PDAC’s have a mutation in one of the four key driver oncogenes but also substantial genomic heterogeneity given the array of lower frequency other gene mutations with a frequency of <10% beyond driver mutations. Defining specific genomic patterns may underpin tailoring individual therapy to a targetable phenotype and may improve clinical outcomes. Waddell et al. in a pioneering study defined the genomic landscape of PDAC by performing deep whole-genome sequencing of a hundred PDACs and showed that structural variation is an important mechanism of DNA damage in pancreatic carcinogenesis (13). The analysis indicated that a combination of structural variation, mutational spectra and gene mutations creates a signature that may predict therapeutic responsiveness for platinum-based therapy, which are current therapeutic options for PDAC. Lowery at al. recently described a cohort of PDAC patients enrolled into the MSK study in which comprehensive next-generation sequencing was performed to determine the real-time feasibility of molecular profiling and in order to identify potential novel therapeutic targets (12). This report of 336 patients demonstrated frequent genetic alterations in key signaling pathways, including Notch signaling, chromatin remodeling, DNA repair, cell cycle, RNA processing, WNT and TGF-beta signaling, and KRAS activation, in concordance with former studies (9,13-16).

Due to former limited evidence pointing at different clinical and pathological features between EOPC and AOPC, we sought to evaluate whether EOPC displays a different genomic landscape compared to AOPC utilizing three distinct cohorts from MSK, the TCGA and an Australian cohort, which used genome sequencing and copy number analysis to define genomic aberrations. This is a first systematic review of contemporary genomic data on age groups of PDAC. We used the clinical definition of EOPC as patients ≤55 years at diagnosis, and AOPC as patients ≥70 years at diagnosis. Limited by small sample size, the analysis depicts a differential display in several genes across the two groups. We observed a significant differential mutation pattern in SMAD4 and PIK3CA. Using the gene set enrichment analysis tool (gsea; www.broadinstitute.org/gsea), we expanded the analysis and found that the RTK pathways...
displayed the most remarkable change across this cluster that reflected also in the protein expression (RPPA). Interestingly, we did not detect a higher *BRCA* mutation rate in the EOPC cases. This observation is in accordance with evidence indicating that while the median age at diagnosis for BRCA carriers is a few years earlier than that of the non-carrier cases (17,18), the age at onset does not meet the EOPC definition (e.g., the median age of familial cases enrolled into the Pancreatic Cancer Genetic Epidemiology Consortium (PacGENE); (19)).

Former retrospective studies suggested that EOPC has a better survival compared to their older counterparts (3,4,8,21,22). Notably, survival analyses did not show any significant difference between the EOPC and AOPC group in the TCGA cohort and the Australian cohorts although, it is notable that the TCGA cohort is comprised of mostly resected early stage patients. With regard to clinical characteristics, we looked for environmental and genetic factors that may contribute to the etiology of PDAC: cigarette smoking (20), which according to former studies is attributable in 20-25% of PDAC cases, family history, pancreatitis and diabetes mellitus (the latter two were available only for the TCGA and MSK cohorts) we noted that smoking was most prevalent in the younger patients in all three cohorts. However, there were no specific clinical, histopathological, genomic, or proteomic features which that can clearly distinguish EOPC over AOPC. Smoking has previously been correlated in multiple cohort and case-control studies to an increased risk for PDAC (23). The risk increases with the extent of smoking history. It has been shown that smoking itself is associated with a younger age at presentation (younger by 8 years) and furthermore, smoking was found to be a strong predisposing factor for PDAC patients with familial clustering or hereditary pancreatitis. Raimondi et al., indirectly suggested a relationship between heavy smoking and EOPC in European countries (7). Piciucchi et al. concluded that EOPC is related to active and early smoking but not to familial syndromes. We did not detect a higher rate of family history among EOPC compared with AOPC (24). Nevertheless, although *BRCA*-associated PDAC is diagnosed at a younger age of onset than sporadic pancreatic cancer, the onset is usually earlier by a decade, and by definition the EOPC cohort of this study is comprised of patients younger than 55 years. The lack of association between family history of cancer and EOPC has also been reported by others (6) and might indicate other etiologies such as certain environmental exposure that might predispose to EOPC.
The Catalogue of Somatic Mutations in Cancer (COSMIC) database depicts the five most frequent genetic aberrations in PDAC: KRAS, TP53, CDKN2A, SMAD4, and ARID1A ((25,26), COSMIC; http://www.sanger.ac.uk/cosmic). The challenge is therefore to translate these “actionable” mutations into tailoring therapeutic opportunities that may improve the prognosis of PDAC and to look for potential correlation between PDAC risk factors and genetic alterations that may enable individualizing therapy over the typical genomic landscape.

We found that SMAD4 and PI3KCA were more frequently mutated in EOPC compared with AOPC in the larger TCGA cohort. SMAD4 plays a role in gene transcription and tumor suppression through the TGF beta signaling pathway (27). SMAD4 mutations have been documented in 20% of PDAC and are associated with a poorer prognosis and increased tendency for metastases (28,29). Once SMAD4 is inactivated, the TGF-beta pathway is not suppressed, resulting in accelerated cell growth (30,31). In the EOPC cohort SMAD4 mutations were more prevalent, and moreover, upregulation of the TGF-beta pathway was observed. PIK3CA mutations were observed in 2-3% of PDAC patients. PI3K has been previously studied in PDAC. Asano et al. descried that PI3K and its downstream signaling pathway are constitutively activated in pancreatic tumors, whereas expression of PTEN, an inhibitor of this pathway, was reduced or lost probably due to promoter methylation (32). The reciprocal interaction between PI3K and MYC was documented whereas activation of PI3K/Akt leads to NF-κB activation and c-Myc stabilization and these two factors are responsible of an increase in cell proliferation.

In summary, key components of two different signaling pathways, SMAD4 (TGF beta) and PIK3CA and related proteins (PI3K pathway) depict a trend of differential expression in EOPC compared with AOPC. These pivotal effectors may interplay, as previously established in the literature, and may delineate a unique profile that is characteristic for younger patients. Though restricted by a small sample size and heterogeneity of the datasets, mainly by the stage (MSK cohort comprised of mostly metastatic patients whereas TCGA comprised mainly of non-metastatic patients) the analyses herein nonetheless shed light on potential genomic differences attributable to younger versus average age patients with PDAC. Future genomic studies using cutting-edge technologies, and prospective follow-up of the datasets included in
this analysis may elucidate the enigma of early onset PDAC that may eventually translate into personalized novel therapeutic strategies to improve clinical outcome.

**Authors' Contributions**

Conception and design: I. Ben-Aharon, M. Elkabets, E.M. O'Reilly


Acquisition of data: I. Ben-Aharon, M. Elkabets, R. Pellossof, E. M O'Reilly

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): I. Ben-Aharon, M. Elkabets, R. Pellossof, K.H Yu, C. A Iacubuzio-Donahuz, S.D Leach, M.A Lowery, K.A Goodman, E. M O'Reilly

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Ben-Aharon, M. Elkabets, R. Pellossof, K.A Goodman, E. M O'Reilly

Study supervision: E.M. O'Reilly

**Grant support:**

This work was funded in part by the Marie-Jos’ee and Henry R. Kravis Center for Molecular Oncology, the National Cancer Institute Cancer Center Core Grant No. P30-CA008748, the Suzanne Cohn Simon Pancreatic Cancer Research Fund, and the David M. Rubenstein Center for Pancreatic Cancer Research.

**Acknowledgements:**

This study was supported by the Institute of Oncology, Davidoff Center, Rabin Medical Center for the work of IBA during part of the study period.

Figure legends

Table 1: Patient characteristics
Figure 1: Pancreatic Cancer; Recent Trends in SEER Incidence Rates, 2000-2015

Figure 2a: Genetic alterations in TCGA cohort
Figure 2b: Genetic alterations in MSK cohort

Figure 3: Upregulated pathways EOPC Gene Expression. Analysis of the gene expression pathway was explored using the Gene Set Enrichment Analysis (GSEA) website (software.broadinstitute.org/gsea). (a) Upregulated pathways in TCGA cohort (b) Gene expression of Hedgehog and TGF-beta pathways across age. (c) Enrichment plot of TGF-beta signaling pathway and volcano plot of expressed genes in EOPC vs. AOPC.

Figure 4: Reverse Phase Protein Array (RPPA) of the TCGA dataset was analyzed by signaling pathway according to GSEA annotation, and Receptor Tyrosine Kinase (RTK) expression.

Figure 5: Kaplan-Meier plot of EOPC and AOPC in the TCGA and Australian cohorts.
References:


31. Levy L, Hill CS. Smad4 dependency defines two classes of transforming growth factor {beta} (TGF-{beta}) target genes and distinguishes TGF-{beta}-induced epithelial-

<table>
<thead>
<tr>
<th>Table 1: Patient Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Stage</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>FH</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pancreatitis</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Figure 1
Figure 3

A

<table>
<thead>
<tr>
<th>Signatures</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HALLMARK_HEDGEHOG_SIGNALING</td>
<td>0.00807</td>
</tr>
<tr>
<td>HALLMARK_TGF_BETA_SIGNALING</td>
<td>0.00942</td>
</tr>
<tr>
<td>HALLMARK_HYPOXIA</td>
<td>0.0138</td>
</tr>
<tr>
<td>HALLMARK_UV_RESPONSE_DN</td>
<td>0.016</td>
</tr>
<tr>
<td>HALLMARK_PI3K_AKT_MTOR_SIGNALING</td>
<td>0.0624</td>
</tr>
<tr>
<td>HALLMARK_ANDROGEN_RESPONSE</td>
<td>0.08</td>
</tr>
<tr>
<td>HALLMARK_PROTEIN_SECRETION</td>
<td>0.086</td>
</tr>
<tr>
<td>HALLMARK_KRAS_SIGNALING_UP</td>
<td>0.1</td>
</tr>
</tbody>
</table>

B

C

Downloaded from clincancerres.aacrjournals.org on January 7, 2019. © 2019 American Association for Cancer Research.
Figure 4

Signaling pathway

RTK
Figure 5

Cancer Specific Survival
TCGA

Survival (%)

0 2 4 6
Years

Young
Old

Cancer Specific Survival
Waddell

Survival (%)

0 2 4 6
Years

Young
Old