Prognostic Relevance of Pancreatic Adenocarcinoma (PAC) Whole-tumor Transcriptomic Subtypes and Components

Shulin Zhao, Rémy Nicolle, Jérémy Augustin, Magali Svrcek, Louis de Mestier, Delphine Le Corre, Daniel Pietrasz, Olivier Caliez, Jérôme Cros, Pierre Laurent-Puig, et al.

Prognostic Relevance of Pancreatic Adenocarcinoma (PAC) Whole-tumor Transcriptomic Subtypes and Components

Authors: Shulin Zhao[1,2], Rémy Nicolle[3], Jérémy Augustin[4], Magali Svrcek[5,6], Louis de Mestier[7], Delphine Le Corre[1], Daniel Pietrasz[1,8], Olivier Caliez[1,5,9], Jérôme Cros[10], Pierre Laurent-Puig[1,11], Jean-Baptiste Bachet[1,5,9]

Affiliations:

[1]Centre de Recherche des Cordeliers, INSERM UMRS1138, CNRS, Sorbonne Université, USPC, Université de Paris, Equipe labellisée Ligue Nationale contre le cancer, CNRS SNC 5096, Paris, France; [2]Department of General Surgery, Pancreatic Disease Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China;[3] Programme Cartes d'Identité des Tumeurs (CIT), Ligue Nationale Contre le Cancer, 75014 Paris, France; [4]Sorbonne Université, Assistance Publique - Hôpitaux de Paris, Department of Pathology, Groupe Hospitalier Pitié Salpêtrière, Paris, France; [5]Sorbonne Université, UPMC Université, IUC, Paris, France; [6]Assistance Publique - Hôpitaux de Paris, Department of pathology, Hôpital Saint Antoine, Paris, France; [7]Assistance Publique - Hôpitaux de Paris, Department of Gastroenterology and Pancreatology , and INSERM U1149, Hôpital Beaujon, Clichy, France; [8]Paris-Saclay University, Assistance Publique - Hôpitaux de Paris, Department of digestive surgery, Paul Brousse Hospital, Villejuif, France; [9]Assistance Publique - Hôpitaux de Paris, Department of Hepato-gastroenterology, Groupe Hospitalier Pitié Salpêtrière, Paris, France; [10]Assistance Publique – Hôpitaux de Paris, INSERM, Université de Paris, Department of Pathology, Hôpital Beaujon, Clichy, France; [11]Assistance Publique - Hôpitaux de Paris, Department of biochemistry, Hôpital Européen Georges Pompidou, Paris, France.

Running title: Pancreatic Adenocarcinoma Transcriptomic Profiling

Corresponding author: Pr. Jean-Baptiste Bachet, Department of Gastroenterology and Digestive Oncology, Hôpital Pitié-Salpêtrière, Sorbonne University, 47-83 Boulevard de l'Hôpital, Paris 75013, France; Phone: 331-4216-1041; Fax: 331-4216-1238; E-mail: jean-baptiste.bachet@aphp.fr.

Conflict of interest statement:

J. Augustin has received personal fees from Servier outside the submitted work. M. Svrcek has received personal fees for consulting or advisory role from Bristol-Myers Squibb, MSD Oncology, Sanofi; travel, accommodations, expenses from Bristol-Myers Suibb, Ventana/Roche. L. de Mestier has received personal fees from AAA, Ipsen, Keocyt, and SIRTex, and grants from Keocyt, outside the submitted work. J. Cros has received personal fees from Novartis, IPSEN, outside the submitted work. P. Laurent-Puig has received grants from Fondation Roche, grants and non-financial support from Ligue Nationale de lutte contre le cancer, and grants from Siric Carpem during the conduct of the study; personal fees from Amgen,

Astra Zeneca, Biocartis, BMS, Lilly, Pfizer, Pierre Fabre, Roche, Sanofi, Servier, non-financial support from Biocartis and grants from Biorad, BMS, Servier outside the submitted work. J.B. Bachet has received personal fees from Amgen, AstraZeneca, Bayer, Merck Serono, Pierre Fabre, Roche, Sanofi, Servier, Shire, and non-financial support from Amgen, Merck Serono, and Roche.

S. Zhao, R. Nicolle, D. Le Corre, D. Pietrasz, O. Caliez declare no competing interests.

Statement of Translational Relevance

Previous studies of transcriptomic profile-based pancreatic adenocarcinoma (PAC) subtypes have focused on the qualitative classification. Using RNA micro-array technique, our team previously defined six quantitative tumor and stroma components and developed a qualitative subtyping schema with the help of these components. In this study, we demonstrated that the quantitative components are more robust than qualitative classifications from different transcriptomic profile techniques (RNA sequencing and RNA micro-array). Comparing with the qualitative subtypes, combining six quantitative components showed an advantage in the prognosis of disease-free survival (DFS) and overall survival (OS) in resected PAC. We created and validated a new DFS based multivariate Cox regression prognostic model, including six PAC transcriptomic component levels and pathological characteristics. We used a monocentric cohort for training of the model and validated the model by an independent multicentric cohort. The statistical significances were observed in both cohorts.

Abstract

Purpose:

Our team previously defined six quantitative transcriptomic components, and a classification in five subtypes by association of these components. In this study, we compared the robustness of quantitative components and qualitative classifications from different transcriptomic profiling techniques, investigated their clinical relevance and proposed a new prognostic model.

Experimental Design:

210 patients from a multicentric cohort and 149 patients from a monocentric cohort were included in this study. RNA microarray profiles were obtained from 165 patients of the multicentric cohort. RNA sequencing (RNA-seq) profiles were obtained from all the patients.

Results:

For the patients with both RNA micro-array and RNA-seq profiles, the concordance in subtype assignment was partial with an 82.4% coherence rate. The correlation between the two techniques projections of the six components ranged from 0.85 to 0.95, demonstrating an advantage of robustness. Based on the Akaike Information criterion, the RNA components showed more prognostic value in univariate or multivariate models than the subtypes. Using the monocentric cohort for training, we developed a multivariate Cox regression model using all six components and clinicopathological characteristics (node invasion and resection margins) on DFS. This prognostic model was highly associated with DFS ($p<0.001$). The evaluation of the model in the multicentric cohort showed significant association with DFS and OS ($p<0.001$).

Conclusions:

We described the advantage of the prognostic value and robustness of the whole-tumor transcriptomic components than subtypes. We created and validated a new DFS based multivariate Cox regression prognostic model, including six PAC transcriptomic component levels and pathological characteristics.

Introduction

Pancreatic adenocarcinoma (PAC) accounts for 2.5% of all cancers and 4.5% of cancer mortality worldwide[1], in 2018. At diagnosis, only 15-20% of patients may benefit from curative resection[2]. The 5-year overall survival (OS) rate is only 9%, which is the lowest of all the cancers[3].

In the past decade, transcriptomic profiling was used to identify PAC subtypes. Using gene expression micro-array analysis, Collison et al.[4] identified three subtypes: classical, quasi-mesenchymal (QM-PDA) and exocrine-like. Moffitt et al.[5] stratified two tumor tissue subtypes (basal-like tumor and classical like tumor) and two stromal tissue subtypes (normal stromal and activated stromal) in 2015. Using formalin-fixed paraffin embedded (FFPE) samples and transcriptomic profiles acquired using Affymetrix HG-U219 micro-arrays, our team published a new classification system with more detailed differentiation of tumor microenvironment in 2018[6]. We described the existence of two tumor-specific transcriptomic components (basal-like tumor component and classical tumor component) and identified four stromal-specific transcriptomic components (activated stroma component, inactive structural stroma component, inflammatory stroma component, and immune stroma component). By association of the tumoral and stromal transcriptomic components, we proposed a classification in 5 subtypes (pure classical, immune classical, desmoplastic, stroma activated and pure basal-like) with significant prognostic values. We demonstrated that the exocrine-like phenotype resulted from normal exocrine pancreatic tissue contamination. Comparing with other studies[7,8,9], the two tumor components, basal like or QM-PDA and the classical, were commonly identified in most classification systems[10]. The basal-like subtype carried the worse prognosis compared with classical subtype (median OS 17-19.2 months vs 19-43.1 months).[10]

Recent work based on single-cell analyzes has suggested that most PAC tumors are composed of a mixture of subtypes[11,12] , making discrete subtypes irrelevant resulting in disagreeing subtyping methods[13]. Given the complexity of PAC intra-tumor heterogeneity, the concept of a Molecular Gradient that grades tumors rather than assigning them to disjoint subtypes has been proposed as a clinically relevant alternative[14]. While the idea of scoring or estimating the proportion of each phenotype was proven to be effective for the epithelial compartment, it is unclear how the stromal component, which is a major constituent of PAC lesions and holds key clinical value[5], should be integrated.

In this study, we used an affordable RNA sequencing (RNA-seq) assay to investigate the clinical value of transcriptomic tumor and stroma components in resected PAC.

Patients and Methods

Sample collection and clinicopathologic data

We included two cohorts of patients in this study. The first cohort is a multicentric cohort (Cohort Multi). The first part of this cohort included 165 patients among the 309 patients who were included in our precedent study[6]. For these 165 patients, enough remaining RNA was available. The second part of this cohort included 45 patients who had curative intent resection for PAC and who received an adjuvant chemotherapy from May 2011 to May 2018 in the Pitié Salpêtrière Hospital (Paris, France). The Pitié Salpêtrière Hospital was one of the centers in the precedent study and these 45 patients were successive patients of the precedent study. The second cohort is a monocentric cohort (Cohort Mono). This cohort included 149 consecutive PAC patients who were operated in Beaujon Hospital (Paris, France) from April 1997 to April 2009. For these two cohorts, all types of pancreatic resections were eligible. Exclusion criteria were preoperative chemotherapy or radiotherapy, macroscopic incomplete resection (R2), ampulla of Vater adenocarcinoma, or pancreatic tumors other than adenocarcinoma. Patients who died of postoperative complications during the 30 days following the surgery were also excluded because they were not informative for translational study.

This study was conducted in accordance with principles of the International Conference of Harmonization Good Clinical Practices and Declaration of Helsinki and was approved by an independent ethics committee (CPP Ile-de-France 2014/58NICB and 2014/59NICB). The tumor samples included in this study came from routine care, and no additional samples were taken in the context of this study. An information note was given to the patient to inform them of the use of their samples for future molecular analysis. Patients had the possibility to sign an opposition note to such analysis. All the included patients accepted to participate. The FFPE samples of the first part of the Cohort Multi and the Cohort Mono were extracted as we previously published (two cores with diameters of 1.5mm)[6]. For the Cohort Mono, specialist pancreatic pathologist (J.C) confirmed the presence of neoplastic cells, selected a representative FFPE tumor block after examination of H&E-stained slides and gave a visual estimation of tumor cellularity. Two cores with diameters of 1.5mm in the zone of tumor were extracted. For the second part of the Cohort Multi, the pathologist specialized in pancreatic disease (J.A) confirmed pathology diagnosis, selected representative cores (1 core diameter 1.5 mm for RNA extraction) after the examination of H&E-stained slide.

The following data were collected in a prospective database: clinical and pathologic characteristics (gender, age, medical history, date of diagnosis, location of the primary tumor, primary tumor diameter, tumor differentiation grade, and TNM stage), follow-up data (date of primary resection, date and type of relapse, date of diagnosis of metastatic disease, date and type of chemotherapy regimen, date, and type of chemoradiotherapy, date of death or last follow-up). The TNM stage was redefined according to the AJCC 8th edition by the originally collected data.

Among the 359 patients included, 86 did not receive an adjuvant chemotherapy, 207 received a gemcitabine-based adjuvant chemotherapy and 56 other chemotherapy regimens. Only 3 patients (<1%) received an adjuvant mFOLFIRINOX.

RNA extraction and RNA-seq

In the first part of the Cohort Multi and the Cohort Mono, DNA/RNA was extracted as we previously published using the ALLPrep FFPE tissue kit (Qiagen®, Venlo, The Netherlands)[6]. In the second part of the Cohort Multi, total RNA was extracted from FFPE samples using RNeasy FFPE kit (Qiagen®, Germany). A total of 150 ng of RNA was used as the starting material for QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lexogen®, Austria), according to the manufacturer's instructions. RNA-seq reads were mapped using STAR[15] with the proposed ENCODE parameters on the human hg38 genomes and transcript annotation (Ensembl 75). Gene expression profiles were obtained using FeatureCount[16]. The counts per gene were normalized to CPM (counts per million). The CPM normalized data were then transformed with log2 using an offset of 1.

Determination of PAC whole-tumor RNA component levels and classification of subtypes previously published

We previously described six RNA components by independent component analysis (ICA)[17] based on Affymetrix result[6,18]. These components can quantitatively describe the composition of the tumor (basal-like tumor component and classical tumor component) and stroma (activated stroma component, inactive structural stroma component, inflammatory stroma component, and immune stroma component) component levels. Using the reference of the six components, we projected the component levels of the present study.

We also published[6] the centroid of each of the 5 subtypes (pure classical, immune classical, desmoplastic, stroma activated and pure basal-like). This subtype classification was originally clustered by non-supervised clustering and nominated by their expression of six component levels. The centroid comprises the average subtype expression value of 404 selected genes. The selected gene expression profile of the present study was then correlated to each of the 5 centroids using Spearman rank correlation, as previously published; the subtype centroid with the highest correlation defines the predicted class of the test samples in the present study.

The molecular characterization results in two types of sample phenotyping. Subtyping resulted in a stratification for which each patient is found to be a member of one of the 5 subtypes. On the other hand, the projection on the 6 components resulted, for each patient, in 6 scores measuring the relative level of each tumor and stromal phenotype encoded in the components. For representation purposes, the component projections were then scaled so that each component had a mean of 0 and standard deviation of 1. Ultimately, the classification into 5 subtypes results in one qualitative variable (with five modalities), while the component projections result in 6 quantitative variables (one for each component).

Deriving a prognostic model and subtypes from RNA component

A Cox proportional hazards regression model was trained on the disease-free survival (DFS) of the monocentric cohort using the 6 RNA components, resection margins (R) and N status. A conditional risk score was then obtained for any new patient using the trained coefficients on the patient's RNA components value as well as R and N status. Prognostic groups were then defined by the patient's predicted months of DFS.

Statistical analysis

Means were compared with the independent-samples *t* test, respectively. Continuous distributions were compared to binary or categorical variables using WilCoxon's signed ranks test or Kruskal-Wallis' tests, respectively. Associations between categorical variables was done using χ^2 test or fisher's exact test. χ^2 test was used to compare the variables in more than two groups and fisher's exact test was used to compare the variables in two groups. Associations between two continuous variables, for instance ICA projections, were done using Spearman's rank correlation. OS was defined as the date of surgery to death resulting from any cause. DFS for resected patients was calculated from the date of surgery until first recurrence or death (whatever the cause). Survival curves were estimated using the Kaplan–Meier technique and compared with the log-rank test. For each test, statistical significance was set at a 2-sided p-value of < 0.05. Univariate and multivariate Cox regression analysis and Kaplan-Meier curves were computed using the forest and survival package of the R statistical suite (R Core Development Team). The Cox regression model was used for analyzes and for estimating the hazard ratio with 95% confidence intervals (95% CI).

Results

Patient Population

A total of 359 resected patients were included for this study originating from one multicentric cohort (Cohort Multi n=210) and one monocentric cohort (Cohort Mono n=149). Clinicopathological characteristics of these two cohorts are summarized in Table 1. The patients in Cohort Mono were significantly slightly younger than patients in the Cohort Multi.

Evaluation of PAC whole-tumor subtypes

To evaluate the technical robustness of subtyping-defining RNA signatures, the samples from the first part of multicentric cohort were profiled using both a micro-array and RNA-seq FFPE-compatible approach. The concordance in subtype assignment was partial with an 82.4% coherence rate of all subtypes (Figure 1. A). The distribution of subtypes in each cohort was uniform in both cohorts (Figure 1. B). The prognostic value of whole-tumor subtypes was investigated and showed a significant association with DFS (Figure 1. C) and OS (Figure 1. D) in the pooled cohorts (n=359). For the DFS, pure basal-like subtype and the stroma activated subtype significantly showed the earliest relapse, with a median DFS of 10.23 months and 11.74 months, respectively. Desmoplastic subtype and pure classical subtype showed an equivalent outcome, with a median DFS of 17.18 months and 17.97 months, respectively. Immune classical showed the best outcome, with a median DFS of 36.10 months (Figure 1. C). Similar results were observed for OS (Figure 1. D).

Evaluation of tumor and stroma RNA components

Given the impaired technical robustness of subtypes and the recent results on tumor heterogeneity, we next thought to evaluate the use of components to quantify the two tumor phenotypes (basal-like and classical) and the four types of stroma (inactive structural stroma, activated stroma, inflammatory stroma, and immune stroma). The correlation between the micro-array and RNA-seq-based projections of the six components ranged from 0.85 to 0.95 (Figure 2. A-B), demonstrating a high robustness.

Univariate Cox regression between DFS and RNA components (Figure 3. A) or subtypes (Figure 3. B) showed significant associations in both cases. Based on the Akaike Information criterion (AIC), the RNA components showed more prognostic value in univariate or multivariate models than the subtypes (Supplementary Figure. 1).

Defining a novel prognostic model and subtypes based on RNA components

Given the technical robustness and high prognostic value of the RNA components, we then trained in Cohort Mono a multivariate Cox regression model using all six RNA signatures and common clinicopathological characteristics (node invasion and resection margins) on DFS (Figure 3. C). This prognostic model was highly associated with DFS ($p<0.001$). To improve the interpretability and usability of the prognostic model, three prognostic groups were derived from the regression model's prediction into groups of poor (less than 12 months predicted DFS), medium (between 12 and 36 months predicted DFS) and good prognosis (36 months or more of months predicted DFS) showing significant differences in DFS (Figure 3. D) and OS (Supplementary Figure 2. A). The evaluation of the model in the Cohort Multi showed significant association with DFS (Figure 3. E) and OS (Supplementary Figure 2. B).

No significant difference of the distribution of the prognostic groups in each cohort was observed (Supplementary Figure. 3). In all patients of two cohorts, we observed 262 patients with relapse in 5 years and 59 patients without relapse in 5 years. The prognostic model was significantly associated with the 5-year DFS rate (Supplementary Figure 4. A). Including the impact of adjuvant therapy as a variable, the multivariate Cox regression analysis showed the prognosis value of the model is independent for DFS (Supplementary Figure 4. B) and OS (Supplementary Figure 4. C).

Characterization of the proposed prognostic groups

The three prognostic groups showed specific RNA and clinic-pathological patterns (Figure 4. A). In particular, the poor prognostic group is frequently associated with N2 lymph node stage (p=0.001) and R1 resection margins (p=0.034). The prognostic groups showed differential distribution in the five RNA subtypes (Figure 4. B). Pure basal-like subtype was significantly more observed in the groups with poorer prognostic (Good: 0.00%, Medium: 9.52% and Poor: 31.00%; p<0.001) while the desmoplastic (Good: 42.97%, Medium: 26.98% and Poor: 13.00%; p<0.001) and immune classical (Good: 14.84%, Medium: 9.52% and Poor: 2.00%; p=0.004) were more frequent in groups of improved prognostic.

Specific enrichment in each RNA component further demonstrates the enrichment of each prognostic group in each RNA-defined tumor and stromal phenotypes (Figure 4. C-H).

For the two RNA tumor components, basal-like tumor component level was significantly higher expressed in the poorer prognostic groups (p<0.001) (Figure 4. D). Conversely, classical tumor (Figure 4. C) as well as inactive structural stroma (Figure 4. F) and immune stroma (Figure 4. H) component levels were significantly higher in improved prognostic group.

Discussion

In this study, we validated the existence of whole-tumor subtypes and components from an independent cohort (194 new patients) by RNA-seq data. The transcriptomic components and subtypes were originally developed on RNA micro-array data. We validated the prognostic values of the subtypes that we originally described. We also validated the prognostic value of the components. We compared the difference between the qualitative subtypes and the quantitative components (Figure. 5).

Since the first publication of transcriptomic subtypes based on RNA micro-array data of PAC published by Collinsson et al. [4], several studies overlapped these findings by micro-array based data[8,19]. RNA-seq is a technology that came into practice recently after the publication of these papers. RNA-seq is superior to micro-array analysis, with advantages in a broader dynamic range and detecting low abundance transcripts[10]. Recently studies based on RNA-seq came into people's sight[9,12,13]. To our knowledge, there is no study that had compared the results from RNA-seq and micro-array data for the classification and phenotyping of PAC. In the first cohort of this study, we compared the components values and subtypes classification from RNA-seq and RNA micro-array data (Figure. 5). A good concordance for both transcriptomic subtypes and components was confirmed.

Compared with the confirmed prognostic value of the subtypes that we previously described, combining the components in multivariate Cox regression showed a better performance to predict DFS and OS for the patients who had PAC curative intent resection. The concordance for components was also better than that of subtypes. This result indicates that PACs may be better characterized by quantitative components than the simple subtypes classification, potentially because of technical bias of the profiling methods and/or the effect of sampling. We developed and validated a Cox regression model using the RNA component's levels and common clinicopathological characteristics (node invasion and resection margins) in this article. This model could be prospectively evaluated in the future.

Classical and basal-like tumor associated signatures were described in the most of previous studies[10]. The transcriptomic result of the COMPASS trial showed that the basal-like subtype is chemo-resistant and can be distinguished from classical PAC by GATA6 expression[20]. This result was confirmed by RNA in situ hybridization (ISH) by the same team[21]. However, considering the heterogeneity of the tumor, these two tumor components can be existing in the same patient. Preclinical models showed that these two components could be modulated by different therapies demonstrating the plasticity of PDAC cells that contributes to the heterogeneity of PDAC tumors and their intrinsic resistance to a broad spectrum of therapies[22]. In our study, the classical tumor component level was associated with better prognostic while the basal-like tumor component level was associated with poor prognostic. The clinical usage of the tumor component level in the prediction of adjuvant therapy should be evaluated in future studies.

Stroma components are another important part of the PAC component, by shaping the intratumoral architecture of PAC and contributing to PAC heterogeneity[23]. To our knowledge, there is no consensus on the stroma signature compared with the tumor signature. Moffit R et al. [5] defined an activated stroma and a normal stroma. Neuzillet C et al.[24] concluded four subtypes of PAC associated fibroblasts. Using the laser capture microdissected materials, D. Birnbaum et al.[25] identified three subtypes of components. Like the tumor signatures, these signatures are classifications and are qualitative. In our study, we used four quantitative stroma components instead of the classification to describe the stroma more completely. The immune component level was independently associated with later relapse after

surgery. The prognostic value of the immune component was also observed in Mahajan U's study by IHC result[26].

This study is a retrospective study, and only the patients who had initial curative surgery were included. Further prospective studies should include borderline resectable, locally advanced, and metastatic patients.

Conclusion

In conclusion, using RNA-seq results, we confirmed the prognostic value of 5 PAC whole-tumor classification subtypes we previously published based on Affymetrix results. We described that the prognostic value of the whole-tumor transcriptomic components had a better prognostic value than the subtypes. We created and validated a new DFS based multivariate Cox regression prognostic model, including six PAC transcriptomic component levels and pathological characteristics.

Acknowledgements: This work was supported by Roche Institut, the Ministère de l'Enseignement Supérieur et de la Recherche, the Université de Paris, the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the SIRIC CARPEM, the ligue nationale contre le cancer (LNCC, Program "Equipe labelisée LIGUE"; no. EL2016.LNCC). S. Zhao was supported by a Chinese Scholarship Council (CSC) Fellowship. The authors would like to thank all patients and their families participating in the study, our colleagues who contributed to the tumor and data collections, and the Roche Institut for its support.

Authors' Contributions

Conception and design: S. Zhao, R. Nicolle, P. Laurent-Puig, J.B. Bachet.

Development and methodology: S. Zhao, R. Nicolle, J. Augustin, M. Svrcek, J. Cros, P. Laurent-Puig, J.B. Bachet.

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Zhao, L. de Mestier, D. Pietrasz, O. Caliez, J. Cros, P. Laurent-Puig, J.B. Bachet.

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Zhao, R. Nicolle, P. Laurent-Puig, J.B. Bachet.

Writing, review, and/or revision of the manuscript: S. Zhao, R. Nicolle, J. Augustin, M. Svrcek, L. De Mestier, D. Le Corre, D. Pietrasz, O. Caliez, J. Cros, P. Laurent-Puig, J.B. Bachet.

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Zhao, R. Nicolle, D. Pietrasz, O. Caliez, D. Le Corre, P. Laurent-Puig, J.B. Bachet.

Study supervision: P. Laurent-Puig, J.B. Bachet.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A.

   Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **2018**;68:394–424.

2. Li D, Xie K, Wolff R, Abbruzzese JL.

   Pancreatic cancer. *Lancet* **2004**;363:1049–57.

3. Siegel RL, Miller KD, Jemal A.

   Cancer statistics, 2020. *CA Cancer J Clin* **2020**;70:7–30.

4. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al.

   Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* **2011**;17:500–3.

5. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al.

   Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **2015**;47:1168–78.

6. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, et al.

   Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. *Gastroenterology* **2018**;155:1999-2013.e3.

7. Schlitter AM, Segler A, Steiger K, Michalski CW, Jäger C, Konukiewitz B, et al.

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

8. Namkung J, Kwon W, Choi Y, Yi SG, Han S, Kang MJ, et al.

   Molecular subtypes of pancreatic cancer based on miRNA expression profiles have independent prognostic value. *J Gastroenterol Hepatol* **2016**;31:1160–7.

9. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al.

   Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **2016**;531:47–52.

10. Martens S, Lefesvre P, Nicolle R, Biankin A V., Puleo F, Van Laethem JL, et al.

    Different shades of pancreatic ductal adenocarcinoma, different paths towards precision therapeutic applications. *Ann Oncol* **2019**;30:1428–36

11. Juiz N, Elkaoutari A, Bigonnet M, Gayet O, Roques J, Nicolle R, et al.

    Basal-like and classical cells coexist in pancreatic cancer revealed by single-cell analysis on biopsy-derived pancreatic cancer organoids from the classical subtype. *FASEB J* **2020**;34:12214–28.

12. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM, et al.

    Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor

evolution. *Nat. Genet* **2020**; 52:231–240

13. Topham JT, Karasinska JM, Lee MKC, Csizmok V, Williamson LM, Jang GH, et al.

    Subtype-discordant pancreatic ductal adenocarcinoma tumors show intermediate clinical and molecular characteristics. *Clin Cancer Res* **2021**;27:150–7.

14. Nicolle R, Blum Y, Duconseil P, Vanbrugghe C, Brandone N, Poizat F, et al.

    Establishment of a pancreatic adenocarcinoma molecular gradient (PAMG) that predicts the clinical outcome of pancreatic cancer. *EBioMedicine* **2020**;57.

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al.

    STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**;29:15–21.

16. Liao Y, Smyth GK, Shi W.

    FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**;30:923–30.

17. Cardoso JF, Souloumiac A.

    Blind beamforming for non-Gaussian signals. *IEE Proceedings, Part F Radar Signal Process* **1993**;140:362–70.

18. Hilmi M, Cros J, Puleo F, Augustin J, Emile JF, Svrcek M, et al.

    Tumour and stroma RNA signatures predict more accurately distant recurrence than clinicopathological factors in resected pancreatic adenocarcinoma. *Eur J Cancer* **2021**;148:171–80.

19. Kim S, Kang MJ, Lee S, Bae S, Han S, Jang JY, et al.

    Identifying molecular subtypes related to clinicopathologic factors in pancreatic cancer. *Biomed Eng Online* **2014**;13 Suppl 2(Suppl 2):S5.

20. Aung KL, Fischer SE, Denroche RE, Jang G, Dodd A, Creighton S, et al.

    Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. *Clin Cancer Res* **2018**;15;24(6):1344-1354.

21. O'Kane GM, Grunwald BT, Jang GH, Masoomian M, Picardo S, Grant RC, et al.

    GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. *Clin Cancer Res* **2020**;26:4901–10.

22. Porter RL, Magnus NKC, Thapar V, Morris R, Szabolcs A, Neyaz A, et al.

    Epithelial to mesenchymal plasticity and differential response to therapies in pancreatic ductal adenocarcinoma. *Proc Natl Acad Sci U S A* **2019**;116:26835–45.

23. Ligorio M, Sil S, Malagon-Lopez J, Nieman LT, Misale S, Di Pilato M, et al.

    Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. *Cell* **2019**;178:160-175.e27.

24. Neuzillet C, Tijeras-Raballand A, Ragulan C, Cros J, Patil Y, Martinet M, et al.

Inter- and intra-tumoural heterogeneity in cancer-associated fibroblasts of human pancreatic ductal adenocarcinoma. *J Pathol* **2019**;248:51–65.

25. Birnbaum DJ, Begg SKS, Finetti P, Vanderburg C, Kulkarni AS, Neyaz A, et al.

Transcriptomic analysis of laser capture microdissected tumors reveals cancer-and stromal-specific molecular subtypes of pancreatic ductal adenocarcinoma. *Clin Cancer Res* **2021**;27:2314–25.

26. Mahajan UM, Langhoff E, Goni E, Costello E, Greenhalf W, Halloran C, et al.

Immune Cell and Stromal Signature Associated With Progression-Free Survival of Patients With Resected Pancreatic Ductal Adenocarcinoma. *Gastroenterology* **2018**;155:1625-1639.e2.

Table 1. Clinicopathological characteristics of patients and tumors. [α]Mean [range], [β]Numbers of patients n (percent).

		Cohort Multi N=210	Cohort Mono N=149	p value
Age, in years[α]		64.2 (37.0-87.6) (sd.10.2)	61.4 (34.1-79.2) (sd.9.4)	0.009
Gender[β]	Female	87 (41.4%)	83 (55.7%)	
	Male	123 (58.6%)	66 (44.3)	1.000
Largest tumor diameter(mm)[α]		30.0 [7.0-130.0] (sd.15.7)	30.0 [10.0-150.0] (sd.16.5)	0.625
T stage[β]	T1	43 (20.5%)	41 (27.5%)	
	T2	131 (62.4%)	95 (63.8%)	
	T3	30 (14.3%)	12 (8.1%)	
	Unknown	6 (2.9%)	1 (0.7%)	0.095
N stage[β]	N0	56 (26.7%)	36 (24.2%)	
	N1	91 (43.3%)	62 (41.6%)	
	N2	63 (30.0%)	51 (34.2%)	0.682
Grade[β]	G1	80 (38.1%)	76 (51.0%)	
	G2	85 (40.5%)	47 (31.5%)	
	G3	39 (18.6%)	22 (14.8%)	
	Unknown	6 (2.9%)	4 (2.7%)	0.050
Resection Margin[β]	R0	160 (76.2%)	113 (75.8%)	
	R1	45 (21.4%)	36 (24.2%)	
	Unknown	5 (2.4%)		0.719
Adjuvant therapy[β]	Yes	150 (71.4%)	113 (75.8%)	
	No	50 (23.8%)	36 (24.2%)	
	Unknown	10 (4.8%)		0.901

Figure Legends

Figure 1. (A) Comparison of the PAC whole-tumor subtypes from Affymetrix and RNA-seq in the Cohort Multi first part (N=165). The number of patients showed in the bubble. The percentage below each label showed the subtype agreement rate of the current technique compared with another technique. (B) Numbers of PAC whole-tumor classification subtypes in each cohort (N=359). (C-D) Prognostic value of PAC whole-tumor classification subtypes in all patients. (N=359). (C) DFS. (D) OS.

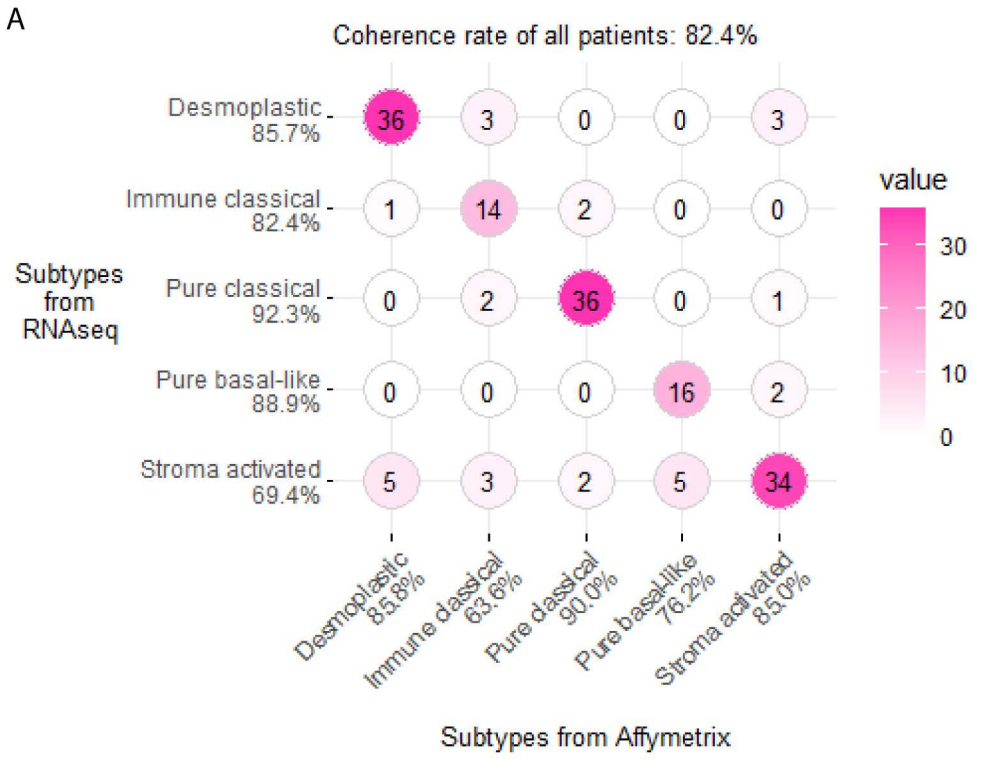
Figure 2. Correlation of six PAC transcriptomic component levels from Affymetrix and RNA-seq in the Cohort Multi first part (N=165). (A) Two tumor component levels. (B) Four stroma component levels.

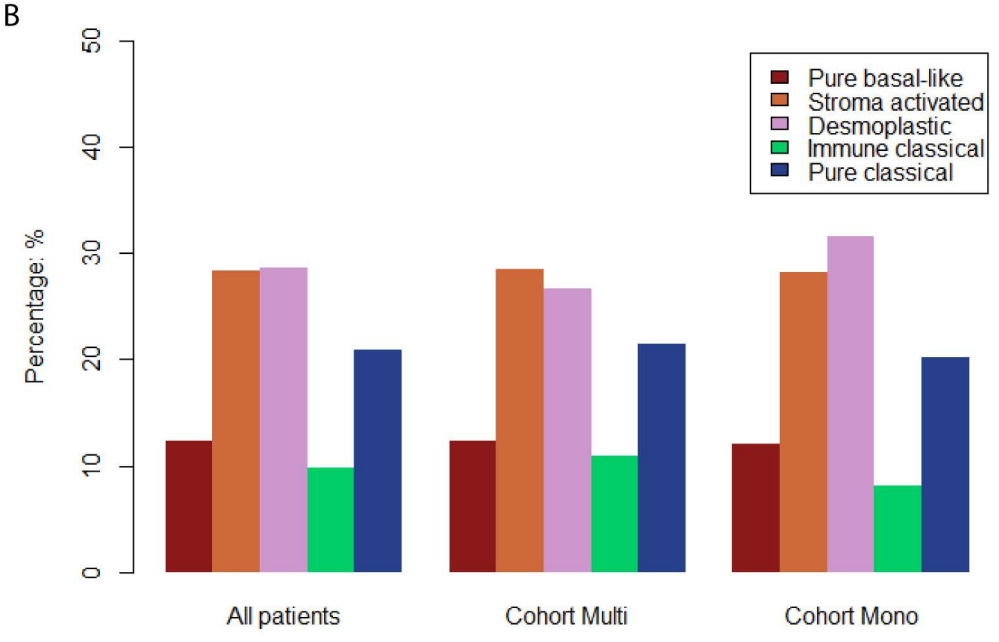
Figure 3. (A-B) Prognostic relevance of the component levels and subtypes on DFS (A) Prognostic value of six PAC transcriptomic component levels by univariate Cox regression in all patients (N=359). (B) Prognostic value of PAC whole-tumor subtypes by univariate Cox regression in all patients (N=359). (C) Prognostic values of the variables included in the training cohort by multivariate Cox regression (Cohort Mono N=149). (D-E) Disease-free survival curve of three prognostic groups in the training cohort (Cohort Mono N=149) and the validation cohort (Cohort Multi, N=205, 5 patients in the Cohort Multi without enough clinicopathological data were excluded). (D) Training cohort. (E) Validation cohort.

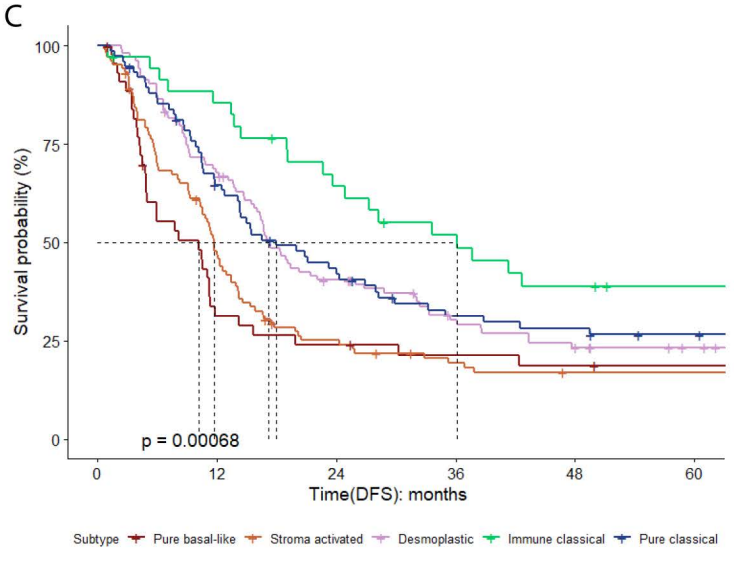
Figure 4. (A) Heatmap of the variable values in the multivariate Cox regression model of all patients (N=354, 5 patients in the Cohort Multi without enough clinicopathological data were excluded). (B) Numbers of PAC whole-tumor classification subtypes in each prognostic group. (C-H) Comparison of six PAC transcriptomic component levels in each prognostic group of all patients. (C) Classical tumor component. (D) Basal-like tumor component. (E) Activated stroma component. (F) Inactive structural stroma component. (G). Inflammatory stroma component. (H) Immune stroma component.

Figure 5. Graphical abstract. Comparison of RNA subtypes and components. The disease-free survival curve of 5 subtypes included all patients in both cohorts (N=359). The disease-free survival curve of the new clinical transcriptomic prognosis model included all patients in both cohorts, except 5 patients without enough clinicopathological data (N=354).

Figure 1


A
Coherence rate of all patients: 82.4%
Desmoplastic 85.7%
Immune classical 82.4%
Pure classical 92.3%
Pure basal-like 88.9%
Stroma activated 69.4%
Subtypes from RNAseq
value
30
20
10
0
Desmoplastic 85.8%
Immune classical 63.6%
Pure classical 90.0%
Pure basal-like 76.2%
Stroma activated 85.0%
Subtypes from Affymetrix


B
Percentage: %
Pure basal-like
Stroma activated
Desmoplastic
Immune classical
Pure classical
All patients
Cohort Multi
Cohort Mono


C
Survival probability (%)
p = 0.00068
Time(DFS): months
Subtype Pure basal-like Stroma activated Desmoplastic Immune classical Pure classical

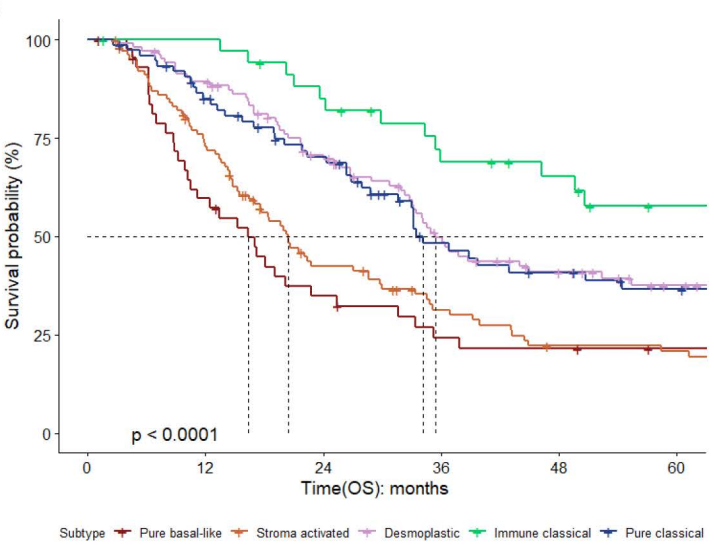

D
Survival probability (%)
p < 0.0001
Time(OS): months
Subtype Pure basal-like Stroma activated Desmoplastic Immune classical Pure classical

Figure 2

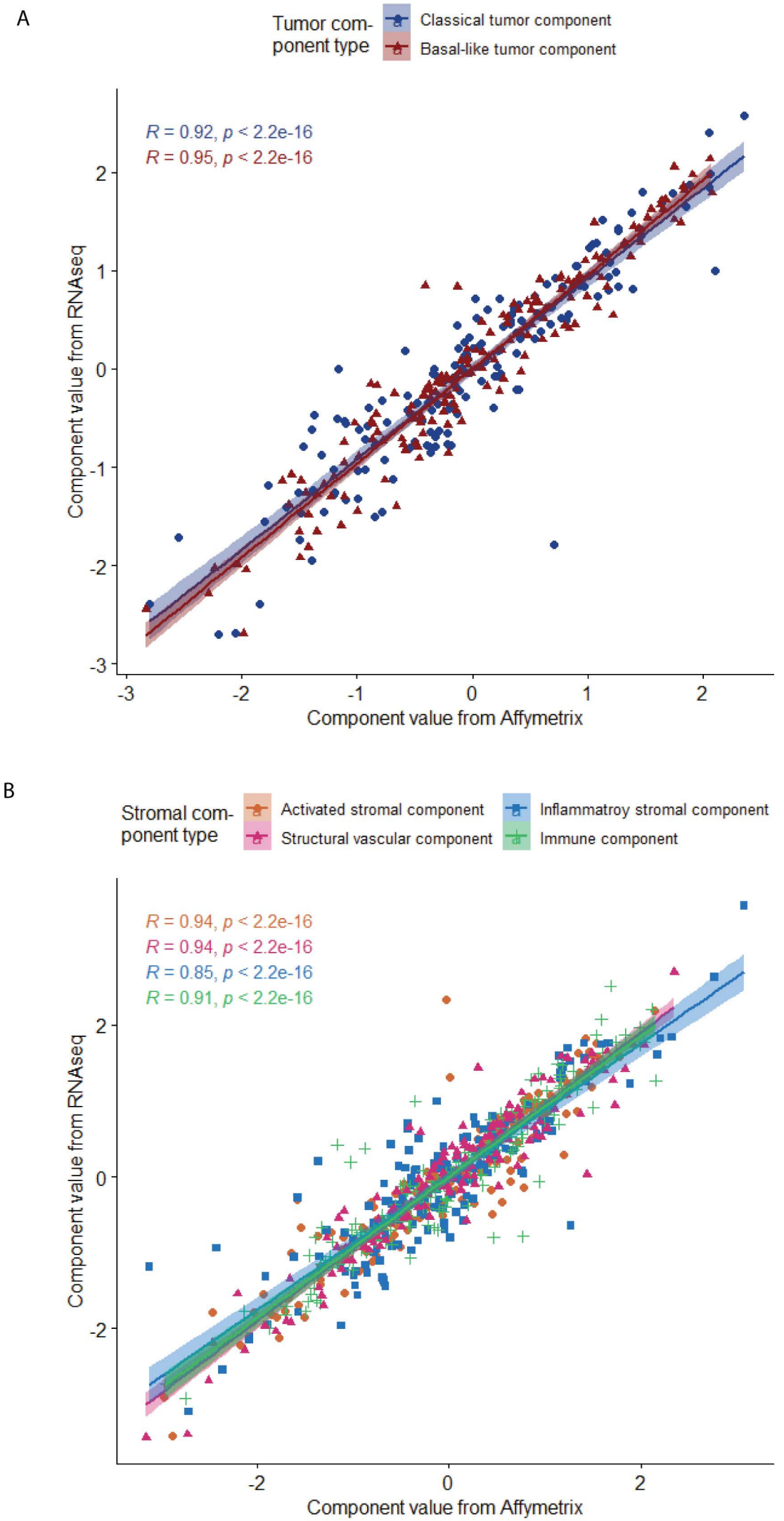

A
Tumor component type
Classical tumor component
Basal-like tumor component
R = 0.92, p < 2.2e-16
R = 0.95, p < 2.2e-16
Component value from RNAseq
Component value from Affymetrix
B
Stromal component type
Activated stromal component
Structural vascular component
Inflammatroy stromal component
Immune component
R = 0.94, p < 2.2e-16
R = 0.94, p < 2.2e-16
R = 0.85, p < 2.2e-16
R = 0.91, p < 2.2e-16
Component value from RNAseq
Component value from Affymetrix

Figure 3

A

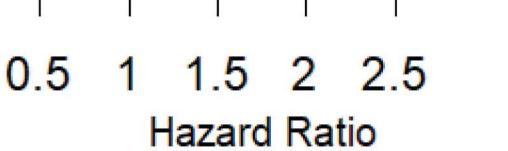
Variable	Coefficient	P value	Hazard Ratio(95% CI)
Classical tumor	-0.164	0.007	0.848(0.753-0.955)
Basal-like tumor	0.266	<0.001	1.304(1.139-1.494)
Actived stroma	0.064	0.299	1.066(0.945-1.203)
Inactive structural stroma	-0.172	0.002	0.842(0.754-0.941)
Inflammatory stroma	0.119	0.039	1.126(1.006-1.260)
Immune stroma	-0.156	0.014	0.855(0.755-0.969)

B

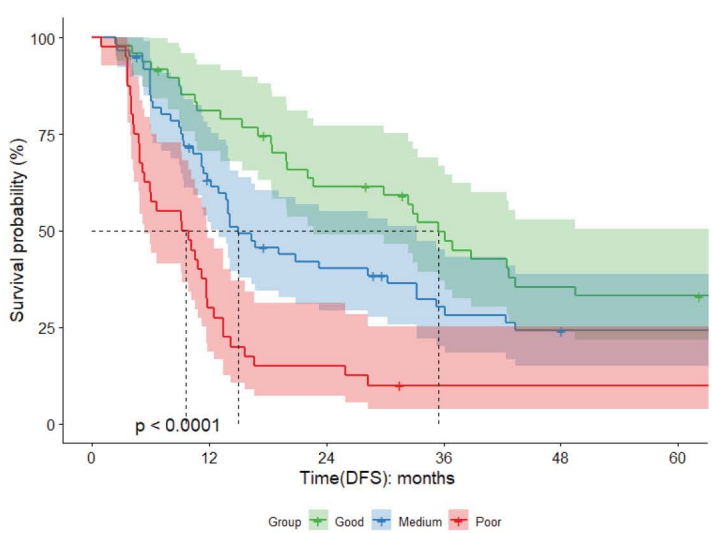
Subtype	Coefficient	P value	Hazard Ratio(95% CI)
Pure basal-like: Referrence			
Stroma activated	-0.096	0.638	0.908(0.609-1.356)
Desmoplastic	-0.490	0.017	0.612(0.409-0.918)
Immune classical	-0.918	0.001	0.399(0.233-0.684)
Pure classical	-0.477	0.029	0.621(0.405-0.952)

C

Variable	Coefficient	P value	Hazard Ratio(95% CI)
Classical tumor	-0.298	0.009	0.742(0.593-0.929)
Basal-like tumor	0.334	0.020	1.397(1.054-1.850)
Activated stroma	-0.268	0.057	0.765(0.581-1.008)
Inactive structural stroma	-0.046	0.679	0.955(0.770-1.185)
Inflammatory stroma	0.202	0.092	1.223(0.967-1.547)
Immune stroma	-0.287	0.027	0.750(0.582-0.968)
N stage:N2 vs N0&1	0.686	0.001	1.986(1.308-3.016)
Resection margin: R1 vs R0	0.481	0.034	1.618(1.038-2.524)



D



E

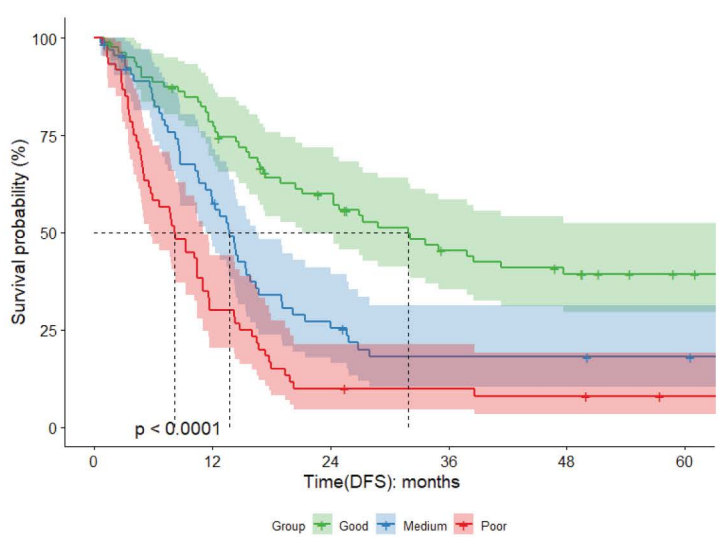


Figure 4

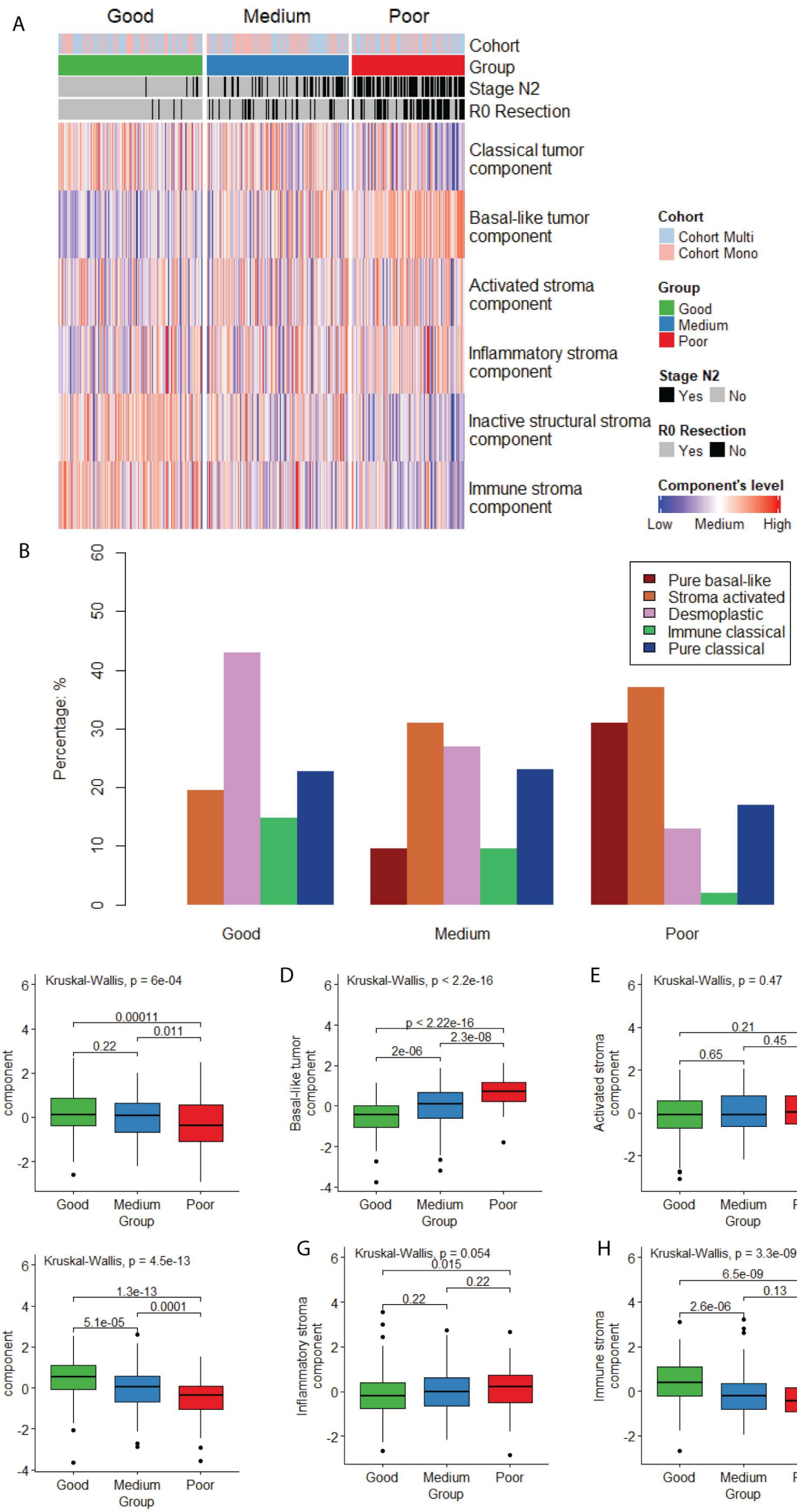
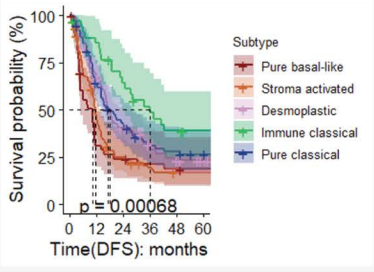

A
Good
Medium
Poor
Cohort
Group
Stage N2
R0 Resection
Classical tumor component
Basal-like tumor component
Activated stroma component
Inflammatory stroma component
Inactive structural stroma component
Immune stroma component
Cohort
Cohort Multi
Cohort Mono
Group
Good
Medium
Poor
Stage N2
Yes
No
R0 Resection
Yes
No
Component's level
Low
Medium
High
B
Percentage: %
0
10
20
30
40
50
60
Pure basal-like
Stroma activated
Desmoplastic
Immune classical
Pure classical
Good
Medium
Poor
C
Kruskal-Wallis, p = 6e-04
0.00011
0.011
0.22
Classical tumor component
Good
Medium
Poor
Group
D
Kruskal-Wallis, p < 2.2e-16
p < 2.22e-16
2.3e-08
2e-06
Basal-like tumor component
Good
Medium
Poor
Group
E
Kruskal-Wallis, p = 0.47
0.21
0.45
0.65
Activated stroma component
Good
Medium
Poor
Group
F
Kruskal-Wallis, p = 4.5e-13
1.3e-13
0.0001
5.1e-05
Inactive structural stroma component
Good
Medium
Poor
Group
G
Kruskal-Wallis, p = 0.054
0.015
0.22
0.22
Inflammatory stroma component
Good
Medium
Poor
Group
H
Kruskal-Wallis, p = 3.3e-09
6.5e-09
0.13
2.6e-06
Immune stroma component
Good
Medium
Poor
Group

Figure 5

	Subtypes	Components
Definition and property	Qualitative Example: Patient A: Pure basal-like Patient B: Pure classical Patient C: Stroma activated Patient E: Immune classical Patient F: Desmoplastic	Quantitative Example: Patient A Basal-like tumor Activated stroma Inactive structural stroma Immune stroma Inflammatory stroma Classical tumor Patient B Patient C Patient D
Robustness • Micro-array • RNAseq	Medium	High
Prognosis ability	5 Subtypes (**) 	New prognosis model based on 6 components, N stage and Resection margins (***)