

RESEARCH ARTICLE

Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers

Antonio Irigoyen¹✉, Cristina Jimenez-Luna²✉, Manuel Benavides³, Octavio Caba⁴*, Javier Gallego⁵, Francisco Manuel Ortuño⁶, Carmen Guillen-Ponce⁷, Ignacio Rojas⁶, Enrique Aranda⁸, Carolina Torres⁹, Jose Prados²

1 Department of Medical Oncology, Virgen de la Salud Hospital, Toledo, Spain, **2** Institute of Biopathology and Regenerative Medicine (IBIMER), Center of Biomedical Research (CIBM), University of Granada, Granada, Spain, **3** Department of Medical Oncology, Virgen de la Victoria Hospital, Malaga, Spain, **4** Department of Health Sciences, University of Jaen, Jaen, Spain, **5** Department of Medical Oncology, University General Hospital of Elche, Alicante, Spain, **6** Department of Computer Architecture and Computer Technology, Research Center for Information and Communications Technologies, University of Granada, Granada, Spain, **7** Department of Medical Oncology, Hospital Ramón y Cajal Hospital, Madrid, Spain, **8** Maimonides Institute of Biomedical Research (IMIBIC), Reina Sofia Hospital, University of Cordoba, Cordoba, Spain, **9** Department of Biochemistry and Molecular Biology I, Faculty of Sciences, University of Granada, Granada, Spain

✉ These authors contributed equally to this work.

* ocaba@ujaen.es



OPEN ACCESS

Citation: Irigoyen A, Jimenez-Luna C, Benavides M, Caba O, Gallego J, Ortuño FM, et al. (2018) Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers. PLoS ONE 13(4): e0194844. <https://doi.org/10.1371/journal.pone.0194844>

Editor: Surinder K. Batra, University of Nebraska Medical Center, UNITED STATES

Received: November 10, 2017

Accepted: March 9, 2018

Published: April 4, 2018

Copyright: © 2018 Irigoyen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data from both microarrays reported in this paper were deposited in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) with accession numbers GSE49641 and GSE74629 for the Affymetrix and Illumina platforms, respectively.

Funding: This work was supported by the Instituto de Salud Carlos III (grant number DTS15/00201 to OC), Ministerio de Economía y

Abstract

Applying differentially expressed genes (DEGs) to identify feasible biomarkers in diseases can be a hard task when working with heterogeneous datasets. Expression data are strongly influenced by technology, sample preparation processes, and/or labeling methods. The proliferation of different microarray platforms for measuring gene expression increases the need to develop models able to compare their results, especially when different technologies can lead to signal values that vary greatly. Integrative meta-analysis can significantly improve the reliability and robustness of DEG detection. The objective of this work was to develop an integrative approach for identifying potential cancer biomarkers by integrating gene expression data from two different platforms. Pancreatic ductal adenocarcinoma (PDAC), where there is an urgent need to find new biomarkers due its late diagnosis, is an ideal candidate for testing this technology. Expression data from two different datasets, namely Affymetrix and Illumina (18 and 36 PDAC patients, respectively), as well as from 18 healthy controls, was used for this study. A meta-analysis based on an empirical Bayesian methodology (ComBat) was then proposed to integrate these datasets. DEGs were finally identified from the integrated data by using the statistical programming language R. After our integrative meta-analysis, 5 genes were commonly identified within the individual analyses of the independent datasets. Also, 28 novel genes that were not reported by the individual analyses ('gained' genes) were also discovered. Several of these gained genes have been already related to other gastroenterological tumors. The proposed integrative meta-

Competitividad (the Spanish Ministry of Economy and Competitiveness) (grant number TIN2015-71873-R to IR), Consejería de Salud, Junta de Andalucía (PIN-0474-2016 to JP), Consejería de Economía, Innovación, Ciencia y Empleo, Junta de Andalucía (P12-TIC-2082 to IR) and the University of Granada (grant number 15/13 to OC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Manuel Benavides has received honoraria from Roche and honoraria for advisory role from Roche. Carmen Guillen-Ponce has received honoraria from Roche; honoraria for advisory role from Roche Pharma, Bayer, Merck Serono, Sanofi Aventis, Celgene y Novocure; other remuneration from Celgene, Roche Pharma, Sanofi Aventis, Merck Serono. Enrique Aranda has received honoraria for advisory role from Amgen, Bayer, Celgene, Merck, Roche, Sanofi. The other authors have stated that they have no conflicts of interest. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

analysis has revealed novel DEGs that may play an important role in PDAC and could be potential biomarkers for diagnosing the disease.

Introduction

Pancreatic ductal adenocarcinoma (PDAC), the most common type of pancreatic cancer (PC), is the fourth leading cause of cancer death in Western countries, with a 5-year survival rate of about 4% and a median survival rate of less than 6 months [1]. At the time of diagnosis, 80% of patients with PDAC are found to have unresectable locally advanced or metastatic disease [2]. The absence of reliable biomarkers for population screening is one of the most important limitations in the management of this malignancy [3].

Currently, the only biomarker in routine clinical use for PDAC is the carbohydrate antigen 19–9 (CA19–9) [4]. However, recent studies found this biomarker to be an unreliable diagnostic tool due to its limited sensitivity (~80%) and specificity (80–90%) [5]. Furthermore, elevated levels of CA19–9 may also appear in pancreatitis [6], benign diseases of the hepatobiliary system [7] and other malignancies of the gastrointestinal tract [8].

Microarray techniques have become a useful tool for determining gene expression profiles in cancer, allowing the discovery of possible tumor biomarkers [9]. However, sometimes biopsy from tumoral tissues can be complex and present complications. In this context, peripheral blood mononuclear cells (PBMCs) constitute an alternative, non-invasive source for finding tumor biomarkers [10,11]. These cells suffer modifications in their gene expression profile when in contact with the tumor microenvironment [12], and may therefore be used as an accessible source of cancer biomarkers.

Additionally, the so-called meta-analysis techniques have been increasingly employed to integrate data from different microarray platforms, making this technology more consistent and powerful. These meta-analyses are especially useful for combining several datasets related to the same disease when they are limited in size, therefore improving their statistical power [13]. Meta-analyses have recently been applied to identify DEGs in several tumor studies, including in breast [14,15], ovarian [16], prostate [17] and pancreatic cancers [18]. One of the main challenges in a meta-analysis is to adequately integrate datasets obtained using different platforms in order to make them comparable. Various methods have been developed to normalize datasets and provide reliable integration, removing batch effects and making cross-platform corrections, such as Distance Weighted Discrimination (DWD) [19], empirical Bayes methods (ComBat) [20], and cross-platform normalization (XPN) [21]. In this sense, ComBat and XPN have been proven to outperform DWD in term of minimizing inter-platform variance [13].

In this study, an integrated meta-analysis of two gene expression datasets from PDAC data was proposed for identifying DEGs in patients. The datasets were collected from two different microarray platforms, namely Affymetrix and Illumina. The expression data was integrated using an empirical Bayes method (ComBat) to avoid bias between the platforms.

Materials and methods

Study population

All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki. All participants gave written informed consent to participate before their enrolment in the study. The study was approved by the respective Ethics Committee at the

Hospital Universitario Puerta del Mar, Hospital Germans Trias i Pujol, Complejo Hospitalario de Navarra, Hospital Reina Sofia, Hospital General de Valencia, Hospital Sant Pau, Hospital Virgen de la Salud, Hospital Parc Taulí, Hospital Universitario Ramón y Cajal, Hospital Carlos Haya, Hospital Universitario Marques de Valdecilla, Hospital General de Elche, Hospital Son Llatzer, Hospital Universitario de Donostia, and Hospital Virgen de las Nieves.

The 54 patients with unresectable PDAC recruited in this study were divided into two independent cohorts. Samples from cohort 1, selected from our previous study [22], include 18 patients with PDAC recruited from January 2009 to July 2012 at the Virgen de las Nieves University Hospital in Granada. Cohort 2 was also independent and included 36 new patients with PDAC, from a phase 2 randomized trial, recruited from March 2012 to February 2013 from 15 different hospitals mediated by the Spanish cooperative group for gastrointestinal tumor therapy (TTD). The diagnosis of PDAC was based on clinical evaluation and imaging studies, which were histologically confirmed by surgery or imaging-guided biopsy. The same enrolment criteria were applied to both cohorts. Finally, 18 gender-, age-, and habit- matched healthy controls were included. The study was approved by the Ethics Committee of the different hospitals, and all clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was obtained from all patients and controls before their enrolment in the study.

Blood collection and isolation of total RNA from PBMCs

Prior to any chemotherapy regimen, peripheral blood samples (12 ml) from all patients and healthy controls were collected in PAXgene Blood RNA Tubes (PreAnalytix) and stored at room temperature for 24 hours, to achieve complete lysis of the blood cells and immediate and persistent RNA stabilization. The RNA from PBMCs was isolated using the PAXgene Blood RNA Kit (PreAnalytix) according to the manufacturer's instructions. The final concentration of purified RNA was quantified by absorbance at 260 nm in a NanoDrop 2000c spectrophotometer (Thermo Scientific). The quality was determined using the 2100 Bioanalyzer (Agilent Technologies). All samples presented an RNA integrity number (RIN) >7.0 and a 28S:18S rRNA ratio >1.0.

cDNA microarray analysis

Whole genome cDNA microarray hybridization of samples was performed using two different platforms to identify potential PDAC markers. Affymetrix microarray-based gene expression profiling was carried out on the samples from the patients included in Cohort 1 and 18 healthy controls, using GeneChip® Human Gene ST 1.0 Arrays (Affymetrix Inc.) according to the recommended protocol. Briefly, 1 µg of high-quality total RNA was used to synthesize double-stranded cDNA, and biotin-tagged cRNA was produced. This cRNA was recovered, purified and then hybridized to the chips overnight at 45°C. After being washed and stained, the arrays were scanned with a GeneChip Scanner 3000 7G (Affymetrix Inc.) following the manufacturer's protocol.

The gene expression levels of multi-plat were measured using the HumanHT-12 v4 Expression BeadChip (Illumina Inc.). In addition, the expression data for the same 18 healthy controls were recalculated using Illumina technology. Both Affymetrix and Illumina expression values from healthy controls were considered for the integrative meta-analysis. Briefly, 1 µg of high-quality total RNA isolated using the Illumina TotalPrep RNA Amplification Kit (Ambion) was amplified. Then, it was reverse transcribed into first and second strand cDNA, and biotin labeled cRNA were generated following the manufacturer's instructions. This labeled cRNA was hybridized overnight to the arrays. The beadchips were washed, stained

with dye-labeled streptavidin, and scanned with an Illumina IScan to measure the intensity. The raw data images were analyzed with Illumina Genome Studio software, which generated an average probe intensity for each sample.

Data deposition: the data from both microarrays reported in this paper were deposited in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) with accession numbers GSE49641 and GSE74629 for the Affymetrix and Illumina platforms, respectively.

Microarray data processing and integrative meta-analysis

All data processing and integration procedures were performed using the R statistical programming language. The data were integrated by adapting the scheme from Turbull et al. [13] with particular stages for the analysis of PDAC data in Affymetrix and Illumina. The workflow of the proposed meta-analysis is shown in Fig 1. More specifically, hybridization data from Affymetrix (Cohort 1) were first normalized using Robust Multi-array Average (RMA) analysis from the Bio-conductor R package *oligo* [23]. In the same way, Illumina expression data (Cohort 2) was pre-processed by applying Quantile Normalization (QN) from the R package *lumi* [24]. In both cases, genes with low variability expression values were discarded to reduce false-positive rates.

Data from both platforms were integrated with the *virtualArray* software R package [25]. This software allows data from different microarray platforms to be merged by considering several batch effect removal and cross-platform correction methods. Specifically, the data were integrated using the empirical Bayes method (ComBat) [19]. The ComBat method merges the information from several genes with similar expression distributions in each dataset to estimate the average and variance in each of those genes [26]. From the integrated data, those genes most likely to be differentially expressed in PDAC patients *versus* controls were selected by analyzing the gene expression microarray data with the linear models for microarray data (*limma*) software package [27]. The R script for the integrative meta-analysis is included as S1 File.

To validate the selected genes as PDAC biomarkers, a leave-one-out cross-validation (LOOCV) was performed with them. In this validation, one sample is consecutively discarded from the initial dataset, leaving a temporary training set and one left-out sample (test sample). This validation procedure is extensively used to assess a prediction model when no validation dataset is available.

Finally, we performed a GO enrichment analysis over the set of newly discovered genes after meta-analysis. For this purpose, an enrichment test using the Kolmogorov-Smirnov (KS) statistical test was carried out from *topGO* Bioconductor-R package. This analysis identified those biological functions and process that are shared by the differentially expressed genes.

Results

Patient characteristics

Our study included two independent cohorts of patients. The first group (1) included 18 PDAC patients and the independent cohort (2) comprised 36 PDAC patients. Table 1 shows the most relevant clinical characteristics of the patients from each cohort.

Cohort 1, selected from our previous study [22], consisted of 9 men (50%) and 9 women (50%) with a mean age of 61.4 (range 37–76). None of the patients had a history of chronic pancreatitis, but 7 (38.9%) had a history of type II diabetes mellitus prior to being diagnosed with PDAC. At the time of diagnosis, 12 patients (66.6%) had stage IV tumors and 6 (33.4%) presented stage III tumors.

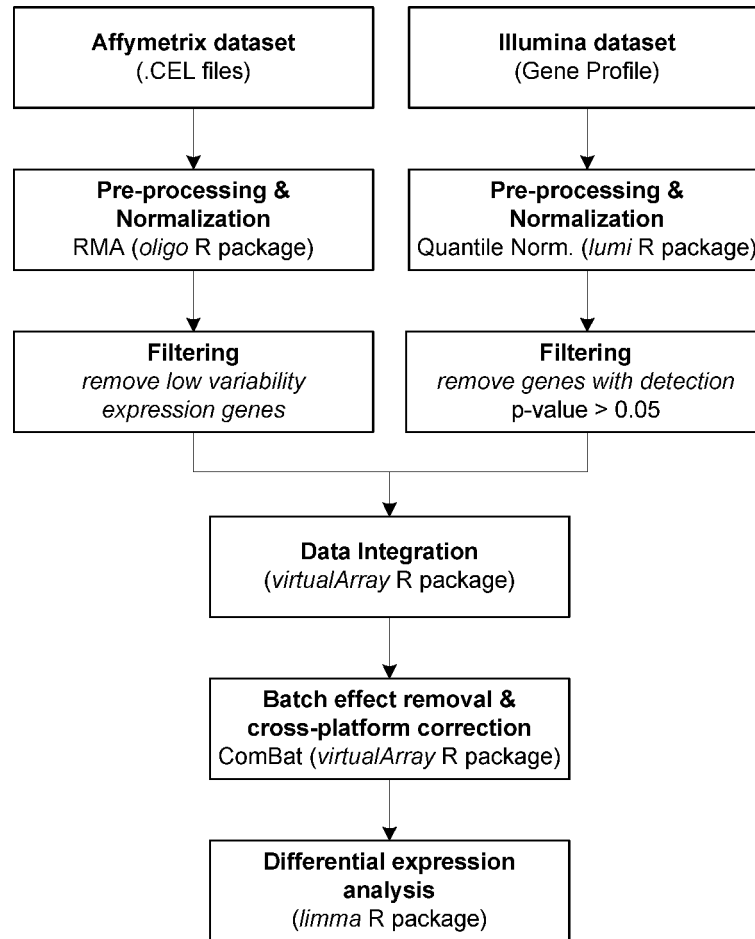


Fig 1. Workflow of the whole integrated meta-analysis for integration of Affymetrix/Illumina expression data from PDAC datasets.

<https://doi.org/10.1371/journal.pone.0194844.g001>

The PDAC patients from Cohort 2 comprised 24 men (67%) and 12 women (33%) with a mean age of 60.0 (range 42–73). Only 2 patients (5.6%) had a history of chronic pancreatitis, however, 14 patients (38.9%) had a history of type II diabetes mellitus. At the time of diagnosis, all patients had stage IV tumors.

Also, 18 healthy subjects were also included in the study. The control group consisted of 10 men (55.6%) and 8 women (44.4%) with a mean age of 60.4 (age range 35–74 years); none of these subjects had a history of either chronic pancreatitis or type II diabetes mellitus.

Differential gene expression profiling of PBMCs from PDAC patients

After normalization and integration using *virtual Array*, the statistical differences in gene expression between the PDAC patients and healthy controls were analyzed with *limma* software. The data were integrated following the ComBat approach in order to reduce the batch effect produced amongst arrays. The effectiveness of the ComBat method in our integration for batch removal can be confirmed according to the comparative boxplots and density plot at [S1 Fig](#). From this meta-analysis, 72 genes were consistently identified as being differentially expressed ($p < 0.01$) with at least a 1.5-fold differential expression between the groups. Of these 72 genes, 39 were overexpressed and 33 repressed ([Table 2](#) and [S1–S3 Tables](#)).

Table 1. Characteristics of both Cohort 1 and Cohort 2 groups of PDAC patients.

Characteristic	Cohort 1 (n = 18)	Cohort 2 (n = 36)
	N°. case (%)	N°. case (%)
<i>Sex</i>		
Male	9 (50%)	24 (67%)
Female	9 (50%)	12 (33%)
<i>Age Mean±SD</i>		
Maximum	76	73
Minimum	37	42
<i>Pancreatitis</i>		
Yes	0 (0%)	2 (5.6%)
No	18 (100%)	34 (94.4%)
<i>Diabetes</i>		
Yes	7 (38.9%)	14 (38.9%)
No	11 (61.1%)	22 (61.1%)
<i>Stage</i>		
I	0 (0%)	0 (0%)
II	0 (0%)	0 (0%)
III	6 (33.3%)	0 (0%)
IV	12 (66.7%)	36 (100%)

<https://doi.org/10.1371/journal.pone.0194844.t001>

The meta-analysis findings were also compared with those obtained by individual analyses in both datasets to evaluate bias and reproducibility across the microarray studies. As a result, 14 genes already identified in the Affymetrix study were also highlighted by our meta-analysis, whereas 35 genes were shared with the Illumina study (Fig 2). Five of these genes were consistently identified by the three studies (Affymetrix, Illumina and integrated meta-analysis) (Table 2). Also ROC curves and areas under the curve (AUC) metrics were calculated for those 5 genes (Fig 3). Finally, a leave-one-out cross-validation was performed over these genes to demonstrate their predictive power. The accuracy values (sensitivity/specificity) obtained from this cross-validation is shown in Table 3.

Additionally, 28 gained genes were found. Gained genes are those identified as differentially expressed in the meta-analysis but not in the individual studies. These genes may be only weakly relevant individually but provide more consistent expression patterns when several datasets are integrated [28]. In order to determine the predictive power of these gained genes, their individual ROC curves were also studied (S3 Fig). Therefore, each individual gene was able to discriminate between PDAC patients and healthy controls with an average sensitivity and specificity of 74.8% and 73.3%, respectively. The same prediction analysis was performed combining the 5 commonly expressed genes as well as the 28 gained genes (S4 Fig). In this case, the sensitivity and specificity results reached the 100% and 94% for the 5 common genes and 91% and 87% for the 28 gained genes.

Finally, we applied a GO enrichment analysis for the 28 gained genes. A total of 12 biological processes were found to be significant across these genes (Table 4).

Discussion

Affymetrix GeneChips and Illumina BeadChips are the main platforms used for gene expression microarrays. However, non-trivial systematic bias (batch effects) can occur in both making it necessary to use appropriate correction methods when integrating the datasets from the two technologies [13]. Also, differences in sequences and the number of probes make it even

Table 2. Coincident genes in the three analyzes: Affymetrix, Illumina and integrated meta-analysis.

Gene	Gene description	ENTREZ ^a	FC ^b	adj.P.Val
<i>FAIM3</i>	Fas apoptotic inhibitory molecule 3	9214	- 2.17	4.59E-11
<i>IRAK3</i>	interleukin-1 receptor-associated kinase 3	11213	1.84	4.59E-11
<i>DENND2D</i>	DENN/MADD domain containing 2D	79961	- 1.67	1.08E-09
<i>PLBD1</i>	phospholipase B domain containing 1	79887	1.67	1.50E-09
<i>AGPAT9</i>	1-acylglycerol-3-phosphate O-acyltransferase 9	84803	1.58	1.47E-08

^aEntrez Gene Name.

^bFold change.

<https://doi.org/10.1371/journal.pone.0194844.t002>

more difficult to integrate their datasets. Consequently, a complex integration method is mandatory in order to successfully perform consistent meta-analyses.

Several batch correction and cross-platform normalization approaches have been proposed for this purpose, including mean-centering (MC), DWD, and empirical Bayesian (ComBat) method. Even though the three proposals were compared for this integrated meta-analysis, the ComBat approach was finally selected. ComBat has been highly recommended in the literature due to its reduced computational cost and the fact it is independent of sample size [26]. It has also proven to be useful in reducing inter-platform variance, outperforming other similar approaches such as DWD or MC (see S2 Fig) [13]. Nevertheless, it is important to highlight

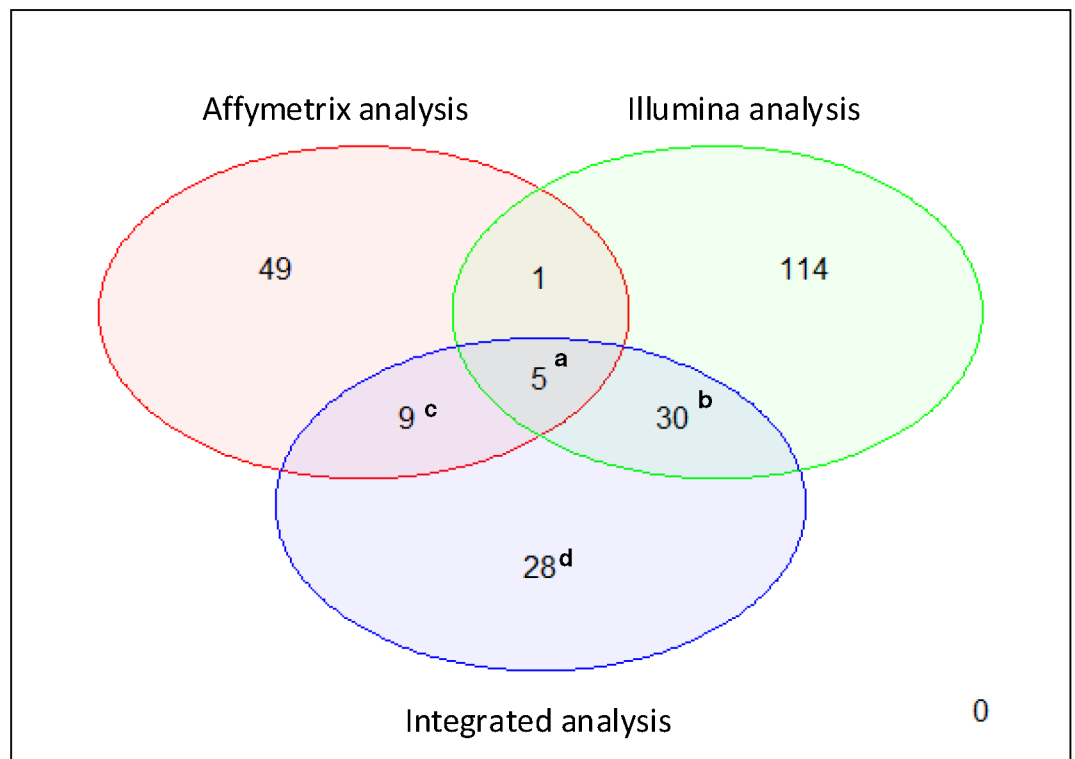


Fig 2. Comparison of individual analysis by technology with integrated analysis. **a** Coincident genes in the three analyzes: Affymetrix, Illumina and integrated meta-analysis (Table 2). **b** Remaining differentially expressed genes in individual Illumina and the integrative meta-analysis (S1 Table). **c** Remaining differentially expressed genes in individual Affymetrix and the integrative meta-analysis (S2 Table). **d** Differentially expressed genes in the integrative meta-analysis but not in individual analysis (gained genes) (S3 Table).

<https://doi.org/10.1371/journal.pone.0194844.g002>

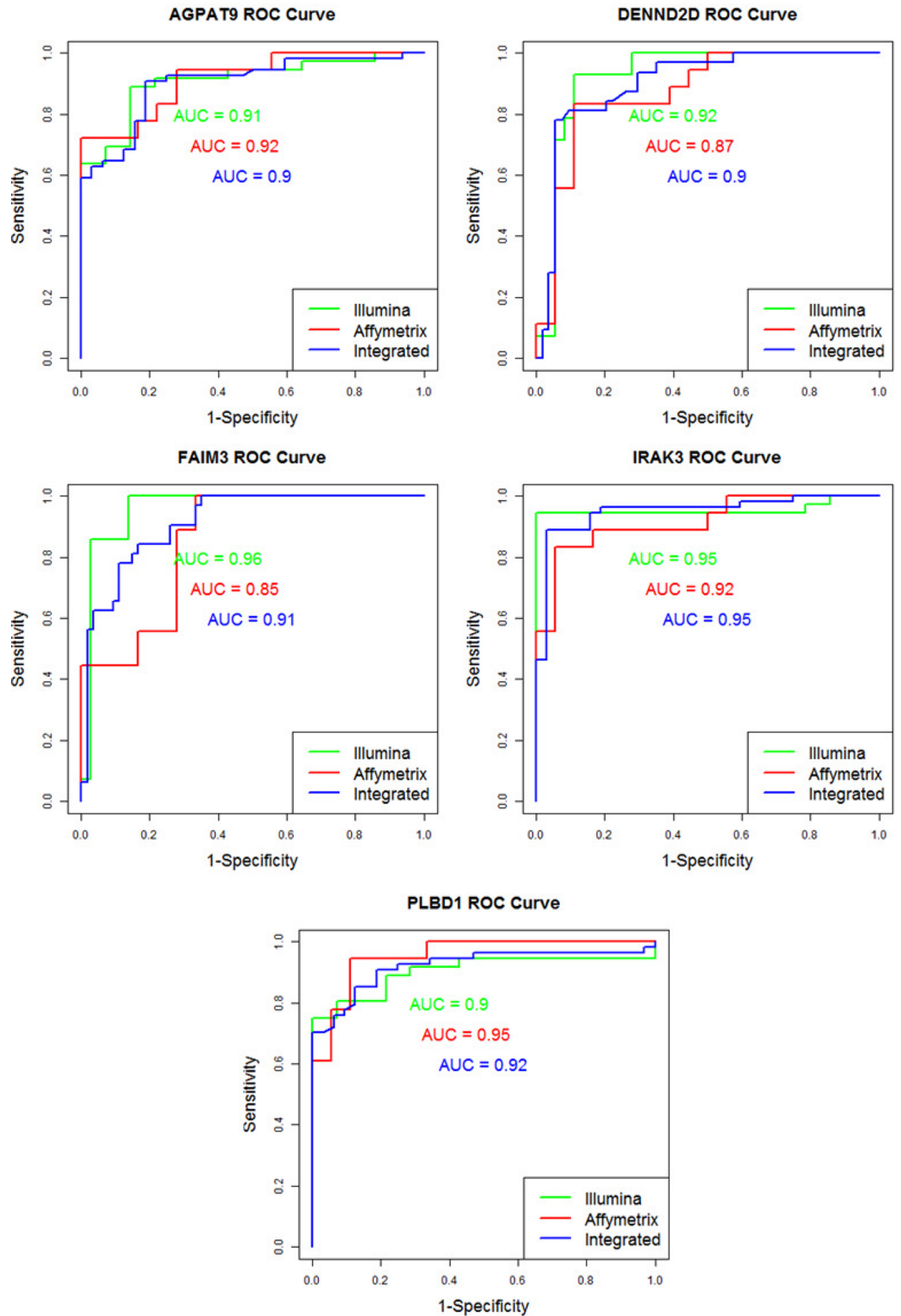


Fig 3. ROC Curves for the 5 genes commonly expressed: *FAIM3*, *IRAK3*, *DENND2D*, *PLBD1* and *AGPAT9*. Curves are provided for both Illumina and Affymetrix individual analyses as well as our integrative meta-analysis. The Area Under the Curve (AUC) metrics are also provided for each curve.

<https://doi.org/10.1371/journal.pone.0194844.g003>

that this methodology is still being carefully revised. Thus, novel alternatives are continually being proposed in the literature trying to correct bias more efficiently among disease samples, for instance, using co-normalization of control samples [29] or combining with other normalization approach like LOESS, SVN or QN [30,31].

Also, given that the multi-platform integration with virtualArray is based on ExpressionSet format, the proposed meta-analysis could be easily extended to integrate other data sources like RNA-Seq from next-generation sequencing technologies [25]. In fact, the widely used RNA-Seq expression analysis with the R package DESeq [32] already applies variance-stabilizing transformation to convert and normalize raw count values to ExpressionSet format.

Other similar meta-analyses have already been carried out to identify biomarkers in pancreatic cancers from several microarray datasets [18]. Nevertheless, these solutions provide DEGs merely by statistically determining the intersection between datasets. In contrast, a more thorough integrative approach including batch correction and cross-platform normalization is proposed in this work.

After this integration, 5 genes, namely Fas apoptotic inhibitory molecule 3 (*FAIM3* or *TOSO*), IL-1 Receptor-Associated Kinase 3 (*IRAK3*), DENN/MADD Domain Containing 2D gene (*DENND2D*), Phospholipase B Domain Containing 1 (*PLBD1*) and 1-Acylglycerol-3-Phosphate O-Acyltransferase 9 (*AGPAT9* or *MAG-1*), were identified as being commonly differentially expressed by the individual analyses in Affymetrix and Illumina as well as by the integrated meta-analysis. These genes were shown to be potential predictors for PDAC diagnosis given they showed areas under the curve (AUC) metrics higher than 0.9 for their corresponding ROC curves (Fig 3). Therefore, these genes were considered reliable targets since they showed consistent differential expression in the integrated analysis and higher predictive metrics. In fact, *IRAK-3* has already been studied and validated by RT-qPCR in our previous study using Affymetrix [22]. Also, the other three genes validated in Affymetrix, namely *ANKRD22*, *CLEC4D* and *VNN1* were similarly identified in the proposed meta-analysis.

More specifically, our results showed downregulation of the gene *FAIM3*, which plays an important role in the immune system as it encodes an Fc receptor for immunoglobulins (Ig), M. Fc receptors specifically bind to the Fc region of Igs to mediate the unique functions of each class [33,34]. The expression of *FAIM3* is reported in peripheral blood leukocytes and detected in high levels in chronic lymphocytic leukemia cells [35]. It has been demonstrated that a decrease in *FAIM3* expression results in increased apoptosis, however, increased *FAIM3* expression resulting from CD25 antibody treatment protects T cells from IL-2-mediated activation-induced cell death (AICD) [36] underlining an involvement in the immune process.

Table 3. Sensitivity and specificity values for the selected genes after a leave-one-out cross-validation (LOOCV) process.

Gene	Sensitivity	Specificity
<i>FAIM3</i>	0.889	0.75
<i>IRAK3</i>	0.87	0.969
<i>DENND2D</i>	0.944	0.75
<i>PLBD1</i>	0.852	0.813
<i>AGPAT9</i>	0.889	0.813

<https://doi.org/10.1371/journal.pone.0194844.t003>

Table 4. Shared Gene Ontology (GO) terms after the gene enrichment analysis applied over the 28 gained genes. The Kolmogorov-Smirnov statistical test was performed to determine their significance (p -value < 0.05).

GO ID	GO Term	Ontology	# Genes	p -value	Genes
GO:0044237	cellular metabolic process	BP	12	0.011	<i>ANXA3; HP; ITGB3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; SH2D1B; ST6GAL1; TXK; VAMP2</i>
GO:0044763	single-organism cellular process	BP	23	0.015	<i>ANXA3; BPI; CD177; CLEC2D; CLEC4E; DYSF; GPR141; HP; ITGB3; LCN2; MS4A1; MYL9; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; SH2D1B; SLC38A1; SORT1; ST6GAL1; TXK; VAMP2</i>
GO:0050776	regulation of immune response	BP	5	0.023	<i>NLRC4; PLSCR1; RASGRP1; SH2D1B; TXK</i>
GO:0044710	single-organism metabolic process	BP	6	0.037	<i>HP; PLSCR1; RASGRP1; S100A12; ST6GAL1; VAMP2</i>
GO:0006139	nucleobase-containing compound metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>
GO:0006725	cellular aromatic compound metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>
GO:0006807	nitrogen compound metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>
GO:0034641	cellular nitrogen compound metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>
GO:0034645	cellular macromolecule biosynthetic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RPS28; S100A12; ST6GAL1; TXK</i>
GO:0044249	cellular biosynthetic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RPS28; S100A12; ST6GAL1; TXK</i>
GO:0046483	heterocycle metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>
GO:1901360	organic cyclic compound metabolic process	BP	7	0.043	<i>ANXA3; NLRC4; PLSCR1; RASGRP1; RPS28; S100A12; TXK</i>

<https://doi.org/10.1371/journal.pone.0194844.t004>

The upregulation of the gene *IRAK3* may provide a clue about the mechanisms leading to immune evasion by tumor cells. This gene is expressed in monocytes and macrophages [37] and can be triggered by Toll-like receptors (TLRs) [38,39], which are expressed in various types of cancer [40]. Overall, *IRAK3* activation leads to immunosuppression [41] and allows the communication between tumor cells and macrophages facilitating cancer progression and a favorable microenvironment for the tumor [42,43]. In fact, monocytes from chronic myeloid leukemia and metastatic cancer patients present *IRAK3* upregulation, leading to tumor formation and growth [44]. In this sense, a study with mouse models carried out by Rothschild et al. [45] demonstrated the connection between *IRAK3* expression and both inflammation and colorectal cancer.

The *DENND2D* gene, another modified gene related to the immune system, has been suggested as a tumor suppressor gene [46]. DENN-domain proteins are differentially expressed in normal and neoplastic cells and regulate Rab GTPases, which play important roles in differentiation, proliferation processes, and regulation of cancer cells, among other things [47,48]. *DENND2D* has been proposed to suppress the tumorigenicity and proliferation of lung cancer cells [49,50]. In addition, the *DENND2D* mRNA expression level has been found to be significantly lower in esophageal squamous cell carcinoma tissues, hepatocellular carcinoma [51], lung cancers, immortalized bronchial epithelial cell lines and other precancerous lesions [46,50].

In our study, the *PLBD1* gene expression level coding was shown to be elevated in PDAC patients. This gene is highly expressed in neutrophils and monocytes [52] and members of this family have been related to antibacterial defense [53].

Metastatic ability is one of the major problems associated with pancreatic cancer. In this regard, our study reveals the overexpression of the *AGPAT9* gene, which has been associated

with the metastatic process in lung cancer [54,55]. Various important functions of *AGPAT9* have been described in this metastatic process. First, *AGPAT9* is involved in the adaptation to the microenvironment, regulating the metabolism and hypoxia, and contributing to vascular development increasing the expression of VEGF. Furthermore, *AGPAT9* is involved in mTOR pathway activation which is key in the metastatic process [56].

Additionally, 28 novel gained genes were found to have more robust patterns in the meta-analysis than in individual studies, making them statistically more significant as possible biomarkers (S3 Table).

Upregulated Annexin A3 (*ANXA3*) and downregulated Membrane-Spanning 4-Domains Subfamily A Member 1 (*MS4A1*) were novel gained genes discovered using this technique. These results are supported by Baine et al., who also included both genes as part of a predictor set of biomarkers in the PBMC of PC patients [57]. Also, Haptoglobin (*HP*) and Lipocalin 2 (*LCN2*) appeared upregulated in this new set of genes. The presence of fucosylated *HP* in serum has been associated to many cancers including hepatocellular, gastric and colon cancers, but the highest incidence has been observed in PC, mainly at an advanced stage [58]. Increased *LCN2* levels have been related to the epithelial to mesenchymal transition [59] and proposed as a serum marker for familial PC [60]. Moreover, we observed the upregulation of other genes like CD177 Molecule (*CD177*), Phospholipid Scramblase 1 (*PLSCR1*), Secretory Leukocyte Peptidase Inhibitor (*SLPI*), S100 Calcium Binding Protein A12 (*S100A12*) and Integrin Beta 3 (*ITGB3*), all of them related to the development of different gastrointestinal tumors [61–66].

It is also noteworthy that all the novel genes that appeared downregulated are associated with the immune response: Granulysin (*GNLY*) functions as a chemotactic for T-lymphocytes, monocytes and other inflammatory cells [67]; Natural Killer Cell Granule Protein 7 (*NKG7*) is expressed in several cell types, including NK and T-cells [68]; C-type Lectin Domain Family 2, Member D (*CLEC2D*) is a receptor present in NK cells [69]; TXK Tyrosine Kinase (*TKX*) takes part in the Th1 cytokine production and is implicated in the adaptive immune response [70]; and RAS Guanyl Releasing Protein 1 (*RASGRP1*) has been found to play an important role in T-cell development [71].

Conclusions

An innovative meta-analysis has been performed to combine two gene expression datasets containing PDAC data and identify robust DEGs in these patients. Integrative meta-analyses have been shown to be powerful tools for identifying more robust DEGs when working with different data sources. Thus, an empirical Bayes approach (ComBat) has been employed in this study to integrate data from two different microarray technologies, namely Affymetrix GeneChip® Human Gene ST 1.0 Arrays and Illumina HumanHT-12 v4 Expression BeadChip, removing the batch effect between technologies and increasing the statistical significance of the subsequent analysis. The integrative analysis has confirmed the DEGs previously published for the Affymetrix data but has also located a set of gained genes that were not robust enough to be identified in the individual analyses. Thus, most of the genes identified have already been annotated as biomarkers in PDAC whereas other gained genes observed in this meta-analysis have also been related to several gastroenterological cancers. The proposed method has therefore been proven useful for more in-depth analysis of heterogeneous expression datasets, improving the identification of DEGs and discovering novel potential biomarkers for diagnosing PDAC. Future RT-qPCR studies will be performed to validate the gained genes that are considered interesting for this purpose. The proposed meta-analysis is also planned to be extended using RNA-Seq data from additional PDAC samples.

Supporting information

S1 File. R Script including code used to obtain results showed in this paper for the integrative meta-analysis.

(ZIP)

S1 Fig. Graphical analysis of the batch analysis removal. (A) Boxplots for the gene expression distributions in Cohort 1 (Affymetrix), Cohort 2 (Illumina) and healthy controls before applying ComBat batch removal. (B) Same boxplots after ComBat batch removal. The distributions show the normalization and reduction of technical differences between cohorts. (C) Density plot and standard deviation of expression across arrays after integration. The red dotted line indicates the median of the standard deviation. An approximately horizontal red line indicates an effective removal of bias and batch effects among arrays.

(PNG)

S2 Fig. Comparison of batch removal method. (A) Boxplots and standard deviation of expression after applying the mean-centering (MC) method. (B) Boxplots and standard deviation after applying the distance discretization method. Although differences cannot be appreciated in boxplots, the median of the standard deviation (red dotted line) indicated a slightly better linearity in ComBat method (see [S1 Fig](#)). Additionally, the median standard deviation is also clearly lower for ComBat batch removal.

(PNG)

S3 Fig. Individual ROC curve for the 28 gained genes. ROC curves for the gained genes. The area under the curve (AUC) is performed to estimate the predictive power of each gene. A cut-off is determined to optimize the discrimination between PDAC patients and healthy controls. The corresponding specificity and sensitivity values are calculated accordingly.

(PDF)

S4 Fig. ROC curves for combined genes. (A) The ROC curve and its corresponding AUC, sensitivity and specificity are obtained for the combination of the 5 genes shared by the three studies (Illumina, Affymetrix and meta-analysis). (B) The ROC curve as well as AUC, sensitivity and specificity values is also obtained for the combination of the 28 gained genes.

(PNG)

S1 Table. Remaining differentially expressed genes in individual Illumina and the integrative meta-analysis.

(PDF)

S2 Table. Remaining differentially expressed genes in individual Affymetrix and the integrative meta-analysis.

(PDF)

S3 Table. Differentially expressed genes in the integrative meta-analysis but not in individual analysis (*gained genes*).

(PDF)

Author Contributions

Conceptualization: Antonio Irigoyen, Octavio Caba, Jose Prados.

Data curation: Francisco Manuel Ortuño, Ignacio Rojas.

Formal analysis: Francisco Manuel Ortuño.

Funding acquisition: Octavio Caba.

Investigation: Cristina Jimenez-Luna, Manuel Benavides, Octavio Caba, Javier Gallego, Carmen Guillen-Ponce, Ignacio Rojas, Enrique Aranda, Carolina Torres.

Methodology: Cristina Jimenez-Luna, Manuel Benavides, Carmen Guillen-Ponce, Enrique Aranda.

Resources: Antonio Irigoyen.

Writing – original draft: Octavio Caba, Jose Prados.

Writing – review & editing: Octavio Caba, Jose Prados.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin.* 2015; 65(1):5–29. <https://doi.org/10.3322/caac.21254> PMID: 25559415
2. Hidalgo M. Pancreatic cancer. *N Engl J Med.* 2010; 362(17):1605–17. <https://doi.org/10.1056/NEJMra0901557> PMID: 20427809
3. Chan A, Diamandis EP, Blasutig IM. Strategies for discovering novel pancreatic cancer biomarkers. *J Proteomics.* 2013; 81:126–34. <https://doi.org/10.1016/j.jprot.2012.09.025> PMID: 23026552
4. Ducreux M, Cuhna AS, Caramella C, Hollebecque A, Burtin P, Goere D, et al. Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2015; 26 Suppl 5:v56–68. <https://doi.org/10.1093/annonc/mdv295> PMID: 26314780
5. Goonetilke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19–9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol.* 2007; 33(3):266–70. <https://doi.org/10.1016/j.ejso.2006.10.004> PMID: 17097848
6. Steinberg W. The clinical utility of the CA 19–9 tumor-associated antigen. *Am J Gastroenterol.* 1990; 85(4):350–5. PMID: 2183589
7. O'Brien DP, Sandanayake NS, Jenkinson C, Gentry-Maharaj A, Apostolidou S, Fourkala EO, et al. Serum CA19-9 is significantly upregulated up to 2 years before diagnosis with pancreatic cancer: implications for early disease detection. *Clin Cancer Res.* 2015; 21(3):622–31. <https://doi.org/10.1158/1078-0432.CCR-14-0365> PMID: 24938522
8. Duffy MJ, Sturgeon C, Lamerz R, Haglund C, Holubec VL, Klapdor R, et al. Tumor markers in pancreatic cancer: a European Group on Tumor Markers (EGTM) status report. *Ann Oncol.* 2010; 21(3):441–7. <https://doi.org/10.1093/annonc/mdp332> PMID: 19690057
9. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer.* 2007; 7(7):545–53. <https://doi.org/10.1038/nrc2173> PMID: 17585334
10. Huang H, Dong X, Kang MX, Xu B, Chen Y, Zhang B, et al. Novel blood biomarkers of pancreatic cancer-associated diabetes mellitus identified by peripheral blood-based gene expression profiles. *Am J Gastroenterol.* 2010; 105(7):1661–9. <https://doi.org/10.1038/ajg.2010.32> PMID: 20571492
11. Tanday S. Biomarkers in blood could help to detect pancreatic cancer. *Lancet Oncol.* 2014; 15(3):e108.
12. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. *Nature.* 2003; 425(6961):905. <https://doi.org/10.1038/425905a> PMID: 14586448
13. Turnbull AK, Kitchen RR, Larionov AA, Renshaw L, Dixon JM, Sims AH. Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. *BMC Med Genomics.* 2012; 5:35. <https://doi.org/10.1186/1755-8794-5-35> PMID: 22909195
14. Pavlou MP, Dimitromanolakis A, Martinez-Morillo E, Smid M, Foekens JA, Diamandis EP. Integrating meta-analysis of microarray data and targeted proteomics for biomarker identification: application in breast cancer. *J Proteome Res.* 2014; 13(6):2897–909. <https://doi.org/10.1021/pr500352e> PMID: 24799281
15. Yasrebi H. Comparative study of joint analysis of microarray gene expression data in survival prediction and risk assessment of breast cancer patients. *Brief Bioinform.* 2016; 17(5):771–85. <https://doi.org/10.1093/bib/bbv092> PMID: 26504096
16. Waldron L, Haibe-Kains B, Culhane AC, Riester M, Ding J, Wang XV, et al. Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. *J Natl Cancer Inst.* 2014; 106(5). <https://doi.org/10.1093/jnci/dju049> PMID: 24700801

17. Pashaei E, Ahmady M, Ozen M, Aydin N. Meta-analysis of miRNA expression profiles for prostate cancer recurrence following radical prostatectomy. *PLoS One*. 2017; 12(6):e0179543. <https://doi.org/10.1371/journal.pone.0179543> PMID: 28651018
18. Goonsekere NC, Wang X, Ludwig L, Guda C. A meta analysis of pancreatic microarray datasets yields new targets as cancer genes and biomarkers. *PLoS One*. 2014; 9(4):e93046. <https://doi.org/10.1371/journal.pone.0093046> PMID: 24740004
19. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. *Bioinformatics*. 2004; 20(1):105–14. PMID: 14693816
20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8(1):118–27. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515
21. Shabalín AA, Tjelmeland H, Fan C, Perou CM, Nobel AB. Merging two gene-expression studies via cross-platform normalization. *Bioinformatics*. 2008; 24(9):1154–60. <https://doi.org/10.1093/bioinformatics/btn083> PMID: 18325927
22. Caba O, Prados J, Ortiz R, Jimenez-Luna C, Melguizo C, Alvarez PJ, et al. Transcriptional profiling of peripheral blood in pancreatic adenocarcinoma patients identifies diagnostic biomarkers. *Dig Dis Sci*. 2014; 59(11):2714–20. <https://doi.org/10.1007/s10620-014-3291-3> PMID: 25069573
23. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. 2010; 26(19):2363–7. <https://doi.org/10.1093/bioinformatics/btq431> PMID: 20688976
24. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008; 24(13):1547–8. <https://doi.org/10.1093/bioinformatics/btn224> PMID: 18467348
25. Heider A, Alt R. virtualArray: a R/bioconductor package to merge raw data from different microarray platforms. *BMC Bioinformatics*. 2013; 14:75. <https://doi.org/10.1186/1471-2105-14-75> PMID: 23452776
26. Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, et al. Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform*. 2013; 14(4):469–90. <https://doi.org/10.1093/bib/bbs037> PMID: 22851511
27. Smyth G. Limma: linear models for microarray data. In: Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S, editors. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer; 2005. p. 397–420.
28. Toro-Dominguez D, Carmona-Saez P, Alarcon-Riquelme ME. Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjogren's syndrome uncovered through gene expression meta-analysis. *Arthritis Res Ther*. 2014; 16(6):489. <https://doi.org/10.1186/s13075-014-0489-x> PMID: 25466291
29. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med*. 2016; 8(346):346ra91. <https://doi.org/10.1126/scitranslmed.aaf7165> PMID: 27384347
30. Muller C, Schillert A, Rothemeier C, Tregouet DA, Proust C, Binder H, et al. Removing Batch Effects from Longitudinal Gene Expression—Quantile Normalization Plus ComBat as Best Approach for Microarray Transcriptome Data. *PLoS One*. 2016; 11(6):e0156594. <https://doi.org/10.1371/journal.pone.0156594> PMID: 27272489
31. Delfani P, Dexlin Mellby L, Nordstrom M, Holmer A, Ohlsson M, Borrebaeck CA, et al. Technical Advances of the Recombinant Antibody Microarray Technology Platform for Clinical Immunoproteomics. *PLoS One*. 2016; 11(7):e0159138. <https://doi.org/10.1371/journal.pone.0159138> PMID: 27414037
32. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11(10):R106. <https://doi.org/10.1186/gb-2010-11-10-r106> PMID: 20979621
33. Kubagawa H, Carroll MC, Jacob CO, Lang KS, Lee KH, Mak T, et al. Nomenclature of Toso, Fas apoptosis inhibitory molecule 3, and IgM FcR. *J Immunol*. 2015; 194(9):4055–7. <https://doi.org/10.4049/jimmunol.1500222> PMID: 25888699
34. Lang KS, Lang PA, Meryk A, Pandya AA, Boucher LM, Pozdeev VI, et al. Involvement of Toso in activation of monocytes, macrophages, and granulocytes. *Proc Natl Acad Sci U S A*. 2013; 110(7):2593–8. <https://doi.org/10.1073/pnas.1222264110> PMID: 23359703
35. Vire B, David A, Wiestner A. TOSO, the Fcmicro receptor, is highly expressed on chronic lymphocytic leukemia B cells, internalizes upon IgM binding, shuttles to the lysosome, and is downregulated in response to TLR activation. *J Immunol*. 2011; 187(8):4040–50. <https://doi.org/10.4049/jimmunol.1100532> PMID: 21908732

36. Richter GH, Mollweide A, Hanewinkel K, Zobywalski C, Burdach S. CD25 blockade protects T cells from activation-induced cell death (AICD) via maintenance of TOSO expression. *Scand J Immunol*. 2009; 70(3):206–15. <https://doi.org/10.1111/j.1365-3083.2009.02281.x> PMID: 19703010
37. Kumar Pachathundikandi S, Brandt S, Madassery J, Backert S. Induction of TLR-2 and TLR-5 expression by *Helicobacter pylori* switches cagPAI-dependent signalling leading to the secretion of IL-8 and TNF-alpha. *PLoS One*. 2011; 6(5):e19614. <https://doi.org/10.1371/journal.pone.0019614> PMID: 21573018
38. O'Neill LA. The interleukin-1 receptor/Toll-like receptor superfamily: 10 years of progress. *Immunol Rev*. 2008; 226:10–8. <https://doi.org/10.1111/j.1600-065X.2008.00701.x> PMID: 19161412
39. Park JS, Gamboni-Robertson F, He Q, Svetkauskaite D, Kim JY, Strassheim D, et al. High mobility group box 1 protein interacts with multiple Toll-like receptors. *Am J Physiol Cell Physiol*. 2006; 290(3):C917–24. <https://doi.org/10.1152/ajpcell.00401.2005> PMID: 16267105
40. Jain A, Kaczanowska S, Davila E. IL-1 Receptor-Associated Kinase Signaling and Its Role in Inflammation, Cancer Progression, and Therapy Resistance. *Front Immunol*. 2014; 5:553. <https://doi.org/10.3389/fimmu.2014.00553> PMID: 25452754
41. Zhou H, Yu M, Fukuda K, Im J, Yao P, Cui W, et al. IRAK-M mediates Toll-like receptor/IL-1R-induced NFkappaB activation and cytokine production. *EMBO J*. 2013; 32(4):583–96. <https://doi.org/10.1038/emboj.2013.2> PMID: 23376919
42. del Fresno C, Otero K, Gomez-Garcia L, Gonzalez-Leon MC, Soler-Ranger L, Fuentes-Prior P, et al. Tumor cells deactivate human monocytes by up-regulating IL-1 receptor associated kinase-M expression via CD44 and TLR4. *J Immunol*. 2005; 174(5):3032–40. PMID: 15728517
43. Soares-Schanoski A, Jurado T, Cordoba R, Siliceo M, Fresno CD, Gomez-Pina V, et al. Impaired antigen presentation and potent phagocytic activity identifying tumor-tolerant human monocytes. *Biochem Biophys Res Commun*. 2012; 423(2):331–7. <https://doi.org/10.1016/j.bbrc.2012.05.124> PMID: 22659741
44. Kuo CC, Shih YL, Su HY, Yan MD, Hsieh CB, Liu CY, et al. Methylation of IRAK3 is a novel prognostic marker in hepatocellular carcinoma. *World J Gastroenterol*. 2015; 21(13):3960–9. <https://doi.org/10.3748/wjg.v21.i13.3960> PMID: 25852282
45. Rothschild DE, Zhang Y, Diao N, Lee CK, Chen K, Caswell CC, et al. Enhanced Mucosal Defense and Reduced Tumor Burden in Mice with the Compromised Negative Regulator IRAK-M. *EBioMedicine*. 2017; 15:36–47. <https://doi.org/10.1016/j.ebiom.2016.11.039> PMID: 27939424
46. Hibino S, Kanda M, Oya H, Takami H, Shimizu D, Nomoto S, et al. Reduced expression of DENND2D through promoter hypermethylation is an adverse prognostic factor in squamous cell carcinoma of the esophagus. *Oncol Rep*. 2014; 31(2):693–700. <https://doi.org/10.3892/or.2013.2901> PMID: 24317529
47. Subramani D, Alahari SK. Integrin-mediated function of Rab GTPases in cancer progression. *Mol Cancer*. 2010; 9:312. <https://doi.org/10.1186/1476-4598-9-312> PMID: 21143914
48. Cheng KW, Lahad JP, Kuo WL, Lapuk A, Yamada K, Auersperg N, et al. The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. *Nat Med*. 2004; 10(11):1251–6. <https://doi.org/10.1038/nm1125> PMID: 15502842
49. Marat AL, Dokainish H, McPherson PS. DENN domain proteins: regulators of Rab GTPases. *J Biol Chem*. 2011; 286(16):13791–800. <https://doi.org/10.1074/jbc.R110.217067> PMID: 21330364
50. Ling B, Zheng H, Fu G, Yuan J, Shi T, Chen S, et al. Suppression of non-small cell lung cancer proliferation and tumorigenicity by DENND2D. *Lung Cancer*. 2013; 79(2):104–10. <https://doi.org/10.1016/j.lungcan.2012.10.012> PMID: 23182661
51. Kanda M, Nomoto S, Oya H, Takami H, Hibino S, Hishida M, et al. Downregulation of DENND2D by promoter hypermethylation is associated with early recurrence of hepatocellular carcinoma. *Int J Oncol*. 2014; 44(1):44–52. <https://doi.org/10.3892/ijo.2013.2165> PMID: 24189587
52. Xu S, Zhao L, Larsson A, Venge P. The identification of a phospholipase B precursor in human neutrophils. *FEBS J*. 2009; 276(1):175–86. <https://doi.org/10.1111/j.1742-4658.2008.06771.x> PMID: 19019078
53. Koduri RS, Gronroos JO, Laine VJ, Le Calvez C, Lambeau G, Nevalainen TJ, et al. Bactericidal properties of human and murine groups I, II, V, X, and XII secreted phospholipases A(2). *J Biol Chem*. 2002; 277(8):5849–57. <https://doi.org/10.1074/jbc.M109699200> PMID: 11694541
54. Jiang Y, Zhang W, Kondo K, Klco JM, St Martin TB, Dufault MR, et al. Gene expression profiling in a renal cell carcinoma cell line: dissecting VHL and hypoxia-dependent pathways. *Mol Cancer Res*. 2003; 1(6):453–62. PMID: 12692265
55. Zhang J, Meng Y, Lin H, Liu G, Wang Y, Chen H, et al. [Promotion of MAG-1 on Metastasis of Lung Cancer Cells in vitro and Its Expression in Lung Cancer Tissue of 24 Cases.]. *Zhongguo Fei Ai Za Zhi*. 2009; 12(2):93–9. <https://doi.org/10.3779/j.issn.1009-3419.2009.02.01> PMID: 20716399

56. Wang Y, Jia H, Lin H, Tan X, Du Z, Chen H, et al. Metastasis-associated gene, mag-1 improves tumour microenvironmental adaptation and potentiates tumour metastasis. *J Cell Mol Med.* 2012; 16(12):3037–51. <https://doi.org/10.1111/j.1582-4934.2012.01633.x> PMID: 22985252
57. Baine MJ, Chakraborty S, Smith LM, Mallya K, Sasson AR, Brand RE, et al. Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. *PLoS One.* 2011; 6(2):e17014. <https://doi.org/10.1371/journal.pone.0017014> PMID: 21347333
58. Okuyama N, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, et al. Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer.* 2006; 118(11):2803–8. <https://doi.org/10.1002/ijc.21728> PMID: 16385567
59. Rodvold JJ, Mahadevan NR, Zanetti M. Lipocalin 2 in cancer: when good immunity goes bad. *Cancer Lett.* 2012; 316(2):132–8. <https://doi.org/10.1016/j.canlet.2011.11.002> PMID: 22075378
60. Slater EP, Fendrich V, Strauch K, Rospleszcz S, Ramaswamy A, Matthai E, et al. LCN2 and TIMP1 as Potential Serum Markers for the Early Detection of Familial Pancreatic Cancer. *Transl Oncol.* 2013; 6(2):99–103. PMID: 23544163
61. Toyoda T, Tsukamoto T, Yamamoto M, Ban H, Saito N, Takasu S, et al. Gene expression analysis of a *Helicobacter pylori*-infected and high-salt diet-treated mouse gastric tumor model: identification of CD177 as a novel prognostic factor in patients with gastric cancer. *BMC Gastroenterol.* 2013; 13:122. <https://doi.org/10.1186/1471-230X-13-122> PMID: 23899160
62. Fan CW, Chen CY, Chen KT, Shen CR, Kuo YB, Chen YS, et al. Blockade of phospholipid scramblase 1 with its N-terminal domain antibody reduces tumorigenesis of colorectal carcinomas in vitro and in vivo. *J Transl Med.* 2012; 10:254. <https://doi.org/10.1186/1479-5876-10-254> PMID: 23259795
63. Guo J, Li G, Zhuang J, Ji C, Liu F, Tao G, et al. [Expression and clinical significance of secretory leucocyte protease inhibitor in colon carcinoma]. *Nan Fang Yi Ke Da Xue Xue Bao.* 2013; 33(6):898–901. PMID: 23803207
64. Thierolf M, Hagmann ML, Pfeffer M, Berntenis N, Wild N, Roessler M, et al. Towards a comprehensive proteome of normal and malignant human colon tissue by 2-D-LC-ESI-MS and 2-DE proteomics and identification of S100A12 as potential cancer biomarker. *Proteomics Clin Appl.* 2008; 2(1):11–22. <https://doi.org/10.1002/prca.200780046> PMID: 21136775
65. Lei Y, Huang K, Gao C, Lau QC, Pan H, Xie K, et al. Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells. *Mol Cell Proteomics.* 2011; 10(10):M110 005397. <https://doi.org/10.1074/mcp.M110.005397> PMID: 21622897
66. Du XY, Liu X, Wang ZJ, Wang YY. SLPI promotes the gastric cancer growth and metastasis by regulating the expression of P53, Bcl-2 and Caspase-8. *Eur Rev Med Pharmacol Sci.* 2017; 21(7):1495–501. PMID: 28429358
67. Krensky AM, Clayberger C. Biology and clinical relevance of granulysin. *Tissue Antigens.* 2009; 73(3):193–8. <https://doi.org/10.1111/j.1399-0039.2008.01218.x> PMID: 19254247
68. Turman MA, Yabe T, McSherry C, Bach FH, Houchins JP. Characterization of a novel gene (NKG7) on human chromosome 19 that is expressed in natural killer cells and T cells. *Hum Immunol.* 1993; 36(1):34–40. PMID: 8458737
69. Germain C, Meier A, Jensen T, Knapnougel P, Poupon G, Lazzari A, et al. Induction of lectin-like transcript 1 (LLT1) protein cell surface expression by pathogens and interferon-gamma contributes to modulate immune responses. *J Biol Chem.* 2011; 286(44):37964–75. <https://doi.org/10.1074/jbc.M111.285312> PMID: 21930700
70. Tavares TS, Nanus D, Yang XJ, Gudas LJ. Gene microarray analysis of human renal cell carcinoma: the effects of HDAC inhibition and retinoid treatment. *Cancer Biol Ther.* 2008; 7(10):1607–18. PMID: 18769122
71. Ksionda O, Limnander A, Roose JP. RasGRP Ras guanine nucleotide exchange factors in cancer. *Front Biol (Beijing).* 2013; 8(5):508–32. <https://doi.org/10.1007/s11515-013-1276-9> PMID: 24744772