



UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



# Tutorial - Diferença de expressão usando GEO2R

Leandro Martins de Freitas



## 1. Introdução

GEO2R é uma ferramenta online integrada ao NCBI que permite identificar genes diferencialmente expressos entre dois ou mais grupos.

## 2. Objetivo

Realizar a análise diferencial de expressão gênica utilizando a ferramenta GEO2R com o dataset GSE15471. Esse conjunto de dados contém 78 amostras: 39 amostras de tecido pancreático tumoral e 39 amostras de tecido pancreático não tumoral adjacente.

### Passo 1: Acessar o dataset GSE28735 e recuperar as informações de expressão

Acesse o link da base de dados GEO:

<https://www.ncbi.nlm.nih.gov/geo/>

Fazer a busca pelo GSE15471

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

**Getting Started**

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

**Tools**

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- Studies with Genome Data Viewer Tracks
- Programmatic Access
- FTP Site

**Browse Content**

Repository Browser

DataSets:	4348
Series:	238523
Platforms:	26625
Samples:	7454936

O estudo GSE15471 envolve a comparação entre tecido tumoral e não tumoral do pâncreas.

Há 78 amostras no total, divididas igualmente entre os dois tipos de tecido.



UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display**

Not logged in | **Login**

Scope: **Self** | Format: **HTML** | Amount: **Quick** | GEO accession: **GSE28735** | **GO**

**Series GSE28735** [Query DataSets for GSE28735](#)

Status	Public on Jul 09, 2012
Title	Microarray gene-expression profiles of 45 matching pairs of pancreatic tumor and adjacent non-tumor tissues from 45 patients with pancreatic ductal adenocarcinoma
Organism	<a href="#">Homo sapiens</a>
Experiment type	Expression profiling by array
Summary	<p>In order to identify biologically relevant tumor markers with prognostic significance, we set out to analyze gene expression profiling of tumor and adjacent non-tumor tissues from PDAC cases.</p> <p>We compared the microarray gene-expression profiles of 45 matching pairs of pancreatic tumor and adjacent non-tumor tissues. This data set were used to obtained genes that were differentially expressed and associated with survival. 51 genes were selected for further validation.</p>
Overall design	<p>90 total samples were analyzed. We compared gene expression profile of 45 pairs of pancreatic tumor and adjacent non-tumor tissues using Affymetrix GeneChip Human Gene 1.0 ST arrays. Tumor gene expression profiles were distinctly different from non-tumor profiles. Using ANOVA in Partek®, 7352 independent genes were found to be differentially expressed in tumors (<math>P &lt; 0.01</math>). We next performed Cox-regression analysis and identified 928 of the differentially expressed genes that were associated with survival (<math>P &lt; 0.1</math>). The list of genes was then subjected to pathway and biomarker analyses using Ingenuity Pathways Analysis (IPA). Based on literatures supporting their roles in cancer, 51 genes were selected for further analyses</p>

### GSE (GEO Series)

GSE é o identificador do conjunto de dados completo que agrupa diversas amostras e descreve o contexto do experimento como um todo. Uma série de experimentos que representa um conjunto de dados completo gerado em uma pesquisa específica. Cada GSE contém várias amostras (GSMs) que estão relacionadas a um experimento comum. O dataset GSE15471 é uma série que reúne amostras de tecido tumoral e não tumoral de pâncreas.

### GSM (GEO Sample)

GSM é o identificador de cada amostra individual no experimento. Refere-se a uma amostra específica dentro de uma série GSE. Cada GSM representa o resultado de uma análise de expressão em uma amostra biológica (por exemplo, um tecido tumoral ou controle). Uma amostra individual do GSE15471 pode ser GSM388076 (amostra de tecido tumoral).

### GPL (GEO Platform)

GPL é o identificador que descreve a plataforma tecnológica usada no experimento. Descreve a plataforma de microarray ou a tecnologia utilizada para medir a expressão gênica. Isso inclui o design dos "chips" ou a tecnologia de sequenciamento usada para gerar os dados.

O GPL possui informações sobre o tipo de chip, genes incluídos, e o fabricante da plataforma. A série GSE15471 foi realizada usando a plataforma GPL570, que corresponde ao chip de microarray HuGene-1\_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version].

### Passo 2: Acessar GEO2R

Na página do GSE15471, clique no botão “Analyze with GEO2R”.



UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > Accession Display [?](#) Not logged in | [Login](#) [?](#)

Scope: [Self](#) Format: [HTML](#) Amount: [Quick](#) GEO accession: [GSE28735](#) [GO](#)

Series [GSE28735](#) [Query DataSets for GSE28735](#)

Status

Public on Jul 09, 2012

Title

Microarray gene-expression profiles of 45 matching pairs of pancreatic tumor and adjacent non-tumor tissues from 45 patients with pancreatic ductal adenocarcinoma

Organism

[Homo sapiens](#)

Experiment type

Expression profiling by array

Summary

In order to identify biologically relevant tumor markers with prognostic significance, we set out to analyze gene expression profiling of tumor and adjacent non-tumor tissues from PDAC cases.  
We compared the microarray gene-expression profiles of 45 matching pairs of pancreatic tumor and adjacent non-tumor tissues. This data set were used to obtained genes that were differentially expressed and associated with survival. 51 genes were selected for further validation.

Overall design

90 total samples were analyzed. We compared gene expression profile of 45 pairs of pancreatic tumor and adjacent non-tumor tissues using Affymetrix GeneChip Human Gene 1.0 ST arrays. Tumor gene expression profiles were distinctly different from non-tumor profiles. Using ANOVA in Partek®, 7352 independent genes were found to be differentially expressed in tumors (P<0.01). We next performed Cox-regression analysis and identified 928 of the differentially expressed genes that were associated with survival (P<0.1). The list of genes was then subjected to pathway and biomarker analyses using Ingenuity Pathways Analysis (IPA). Based on literatures supporting their roles in cancer, 51 genes were selected for further analyses

Contributor(s)

[Hussain SP](#)

Citation(s)

Zhang G, Schetter A, He P, Funamizu N et al. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 2012;7(2):e31507. PMID: [22363658](#)  
Zhang G, He P, Tan H, Budhu A et al. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin Cancer Res* 2013 Sep 15;19(18):4983-93. PMID: [23918603](#)

Analyze with GEO2R

Submission date

Apr 20, 2011

Last update date

Jul 26, 2018

Contact name

Perwez Hussain

E-mail(s)

[hussainp@mail.nih.gov](mailto:hussainp@mail.nih.gov)

Organization name

NCI/NIH

Street address

37 Convent Drive

City

Bethesda

ZIP/Postal code

20892

A interface do GEO2R será aberta com todas as amostras já carregadas para análise.

NCBI > GEO > GEO2R > GSE28735

GEO Publications | FAQ | MIAME | Email GEO | Login

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full Instructions

GEO accession: [GSE28735](#) [Set](#) Microarray gene-expression profiles of 45 matching pairs of pancreatic tumor and adjacent non-tumor tissues from 45 patients with pancreatic ductal adenocarcinoma

• Samples

• Define groups

Selected 8 out of 90 samples

Group	Accession	Title	Source name	Tissue	Survival_months	Cancer_death
-	GSM711994	human pancreatic tumor tissue, patient sample 1	tumor tissue, patient 1	T	51	1
-	GSM711995	human pancreatic non-tumor tissue, patient sample 1	non-tumor tissue, patient 1	N	51	1
-	GSM711996	human pancreatic tumor tissue, patient sample 2	tumor tissue, patient 2	T	7	1
-	GSM711997	human pancreatic non-tumor tissue, patient sample 2	non-tumor tissue, patient 2	N	7	1
-	GSM711998	human pancreatic tumor tissue, patient sample 3	tumor tissue, patient 3	T	3	1
-	GSM711999	human pancreatic non-tumor tissue, patient sample 3	non-tumor tissue, patient 3	N	3	1
-	GSM711910	human pancreatic tumor tissue, patient sample 4	tumor tissue, patient 4	T	42	1
-	GSM711911	human pancreatic non-tumor tissue, patient sample 4	non-tumor tissue, patient 4	N	42	1
-	GSM711912	human pancreatic tumor tissue, patient sample 5	tumor tissue, patient 5	T	na	na
-	GSM711913	human pancreatic non-tumor tissue, patient sample 5	non-tumor tissue, patient 5	N	na	na
-	GSM711914	human pancreatic tumor tissue, patient sample 6	tumor tissue, patient 6	T	36	1
-	GSM711915	human pancreatic non-tumor tissue, patient sample 6	non-tumor tissue, patient 6	N	36	1
-	GSM711916	human pancreatic tumor tissue, patient sample 7	tumor tissue, patient 7	T	2	1
-	GSM711917	human pancreatic non-tumor tissue, patient sample 7	non-tumor tissue, patient 7	N	2	1
-	GSM711918	human pancreatic tumor tissue, patient sample 8	tumor tissue, patient 8	T	na	na

GEO2R

Options

Profile graph

R script

• Quick start

• Specify a GEO Series accession and a Platform if prompted.

• Click "Define group" and enter names for the groups of Samples you plan to compare, e.g., "tumor and control."

### Passo 3: Agrupar as Amostras

Identifique as amostras:

IMS/CAT-UFBA - Rua Rio de Contas, 58 – Quadra 17 – Lote 58 – Bairro Candeias  
Vitória da Conquista – BA - CEP 45.029-094/ Fone: (77) 3429 2709. E-mail: [leandromartins@ufba.br](mailto:leandromartins@ufba.br)



# UNIVERSIDADE FEDERAL DA BAHIA

## Instituto Multidisciplinar em Saúde

### Campus Anísio Teixeira



A lista de amostras aparece no canto direito, com códigos de acesso individuais (ex.: GSM711904).

Crie dois grupos de amostras com os seguintes nomes:

Group	Accession	Title	Source name	Tissue	Samplu_morth	Cancer_Stage
-	GSM711904	human pancreatic tumor tissue, patient sample 1	tumor tissue, patient 1	T	51	1
-	GSM711905	human pancreatic non-tumor tissue, patient sample 1	non-tumor tissue, patient 1	N	51	1
-	GSM711906	human pancreatic tumor tissue, patient sample 2	tumor tissue, patient 2	T	7	1
-	GSM711907	human pancreatic non-tumor tissue, patient sample 2	non-tumor tissue, patient 2	N	7	1
-	GSM711908	human pancreatic tumor tissue, patient sample 3	tumor tissue, patient 3	T	3	1
-	GSM711909	human pancreatic non-tumor tissue, patient sample 3	non-tumor tissue, patient 3	N	3	1
-	GSM711910	human pancreatic tumor tissue, patient sample 4	tumor tissue, patient 4	T	42	1
-	GSM711911	human pancreatic non-tumor tissue, patient sample 4	non-tumor tissue, patient 4	N	42	1
-	GSM711912	human pancreatic tumor tissue, patient sample 5	tumor tissue, patient 5	T	na	na
-	GSM711913	human pancreatic non-tumor tissue, patient sample 5	non-tumor tissue, patient 5	N	na	na
-	GSM711914	human pancreatic tumor tissue, patient sample 6	tumor tissue, patient 6	T	36	1
-	GSM711915	human pancreatic non-tumor tissue, patient sample 6	non-tumor tissue, patient 6	N	36	1
-	GSM711916	human pancreatic tumor tissue, patient sample 7	tumor tissue, patient 7	T	2	1
-	GSM711917	human pancreatic non-tumor tissue, patient sample 7	non-tumor tissue, patient 7	N	2	1

**TUMOR** - Grupo 1: Tecidos tumorais (45 amostras).

Use a descrição das amostras ou clique manualmente nas amostras que indicam tecido tumoral.

**CONTROLE** - Grupo 2: Tecidos não tumorais (45 amostras).

Group	Accession	Title	Source name	Tissue	Samplu_morth	Cancer_Stage
normal	GSM711981	pancreatic non-tumor tissue, patient sample 42	non-tumor tissue, patient 42	N	11	0
normal	GSM711983	pancreatic non-tumor tissue, patient sample 43	non-tumor tissue, patient 43	N	4	1
normal	GSM711985	pancreatic non-tumor tissue, patient sample 44	non-tumor tissue, patient 44	N	10	1
normal	GSM711987	human pancreatic non-tumor tissue, patient sample 45	non-tumor tissue, patient 45	N	10	0
normal	GSM711989	human pancreatic non-tumor tissue, patient sample 39	non-tumor tissue, patient 39	N	15	1
normal	GSM711991	human pancreatic non-tumor tissue, patient sample 40	non-tumor tissue, patient 40	N	5	1
normal	GSM711993	human pancreatic non-tumor tissue, patient sample 41	non-tumor tissue, patient 41	N	13	1
Tumor	GSM711904	human pancreatic tumor tissue, patient sample 1	tumor tissue, patient 1	T	51	1
Tumor	GSM711906	human pancreatic tumor tissue, patient sample 2	tumor tissue, patient 2	T	7	1
Tumor	GSM711908	human pancreatic tumor tissue, patient sample 3	tumor tissue, patient 3	T	3	1
Tumor	GSM711910	human pancreatic tumor tissue, patient sample 4	tumor tissue, patient 4	T	42	1
Tumor	GSM711912	human pancreatic tumor tissue, patient sample 5	tumor tissue, patient 5	T	na	na
Tumor	GSM711914	human pancreatic tumor tissue, patient sample 6	tumor tissue, patient 6	T	36	1
Tumor	GSM711916	human pancreatic tumor tissue, patient sample 7	tumor tissue, patient 7	T	2	1
Tumor	GSM711918	human pancreatic tumor tissue, patient sample 8	tumor tissue, patient 8	T	na	na
Tumor	GSM711920	human pancreatic tumor tissue, patient sample 9	tumor tissue, patient 9	T	na	na

## Passo 4: Configuração e Normalização

Escolha o método de normalização:

O GEO2R utiliza por padrão a normalização log2. Mantenha esta opção ativada, pois ela é importante para reduzir a variabilidade e facilitar a comparação entre as amostras.

Verifique se os grupos estão corretos:

Confirme se todas as 45 amostras de cada grupo foram alocadas corretamente.

## Passo 5: Realizar a Análise

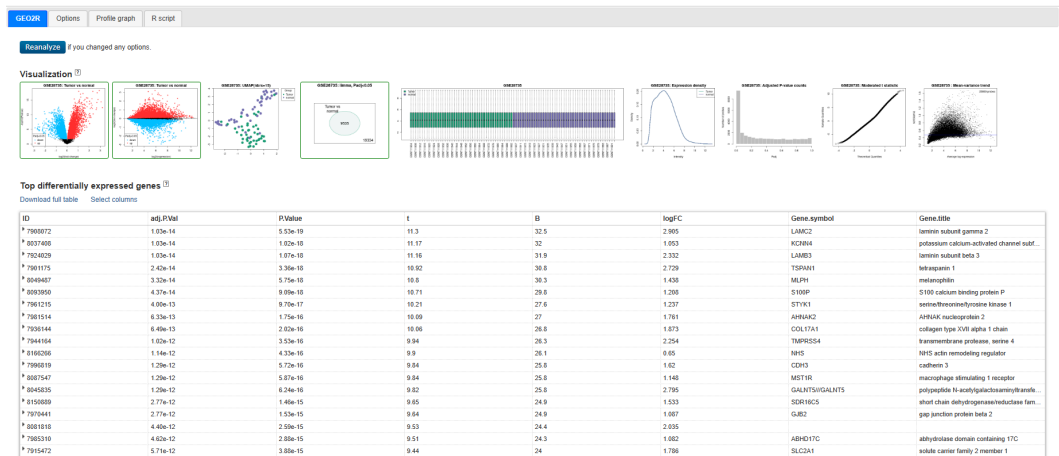
Clique em "Top 250" ou "Analyze" para rodar a análise.



# UNIVERSIDADE FEDERAL DA BAHIA

## Instituto Multidisciplinar em Saúde

### Campus Anísio Teixeira



## Interpretação dos Resultados:

O GEO2R gera uma tabela com os seguintes campos principais:

Gene symbol: Nome do gene.

LogFC (Log Fold Change): Variação na expressão entre os grupos.

P-value: Significância estatística.

Adjusted P-value (FDR): Correção de múltiplos testes (False Discovery Rate).

Top differentially expressed genes

Download full table Select columns

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
<sup>#</sup> 7908072	1.03e-14	5.53e-19	11.3	32.5	2.905	LAMC2	laminin subunit gamma 2
<sup>#</sup> 8037405	1.03e-14	1.02e-10	11.17	32	1.053	KCNH4	potassium calcium-activated channel subf...
<sup>#</sup> 7924029	1.03e-14	1.07e-10	11.16	31.9	2.332	LAMB3	laminin subunit beta 3
<sup>#</sup> 7901175	2.42e-14	3.36e-10	10.92	30.8	2.729	TSPAN1	tetraspanin 1
<sup>#</sup> 8049467	3.32e-14	5.75e-10	10.8	30.3	1.438	MLPH	melanophilin
<sup>#</sup> 8093950	4.37e-14	9.09e-10	10.71	29.8	1.208	S100P	S100 calcium binding protein P
<sup>#</sup> 7961215	4.00e-13	9.70e-17	10.21	27.6	1.237	STYK1	serine/threonine/tyrosine kinase 1
<sup>#</sup> 7981514	6.33e-13	1.75e-16	10.09	27	1.761	AHRK2	AHRK2 nucleoprotein 2
<sup>#</sup> 7936144	6.49e-13	2.02e-16	10.06	26.8	1.873	COL1A1	collagen type XVII alpha 1 chain
<sup>#</sup> 7944164	1.02e-12	3.53e-16	9.94	26.3	2.254	TMPS4	transmembrane protease, serine 4
<sup>#</sup> 8166266	1.14e-12	4.33e-16	9.9	26.1	0.65	NHS	NHS actin remodeling regulator
<sup>#</sup> 7996819	1.29e-12	5.72e-16	9.84	25.8	1.62	CDH3	cadherin 3
<sup>#</sup> 8087547	1.29e-12	5.87e-16	9.84	25.8	1.148	MST1R	macrophage stimulating 1 receptor

## Análise de sobrevivência

<https://kmplot.com/analysis/>

## LAMC2

