

Предсказание метаданных эксперимента по экспрессии генов

Алексей Сергушичев



Cold
Spring
Harbor
Laboratory

bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT
| CHANNELS

New Results

Improving the value of public RNA-seq expression data by phenotype prediction

Shannon E Ellis, Leonardo Collado Torres, Jeffrey Leek

doi: <https://doi.org/10.1101/145656>

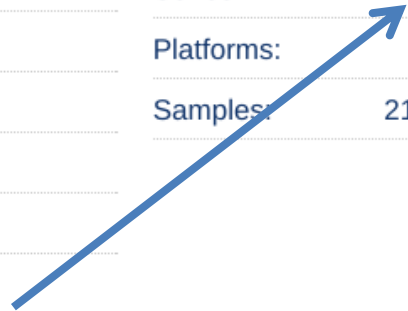
This article is a preprint and has not been peer-reviewed [what does this mean?].

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	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 4348
About GEO DataSets	Search GEO Documentation	Series:  87316
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 17534
About GEO2R Analysis	GEO BLAST	Samples: 2148664
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	



87316 экспериментов!

Не для всех образцов есть хорошая аннотация

Sample GSM789762		Sample GSM310012 Query DataSet	
Status	Public on Sep 01, 2011	Status	Public on Jul 30, 2009
Title	ctx_261_ko_f_2	Title	Global gene expression profiling of human pleural mesotheliomas; me
Sample type	RNA	Sample type	RNA
Source name	Mouse Brain Tissue	Source name	mesothelioma tissue
Organism	Mus musculus	Organism	Homo sapiens
Characteristics	genotype: ko region: ctx timepoint: 2 month gender: F replicate: 4	Characteristics	Mesothelioma tissue was obtained from those patients with a confi diagnosis and who had not received prior therapy. Intraoperative mali samples and nodules were dissected from associated fat and connec microdissection was performed. H&E staining was performed to verif neoplastic cells and to determine the histological subtype. Sample RNAlater (Ambion) following the manufacturer's protocol until RNA ex
Treatment protocol	None	Extracted molecule	total RNA
Growth protocol	Standard	Extraction protocol	Total RNA was extracted from each of the tumor and control samples Mini kit (Qiagen, Valencia, CA).
Extracted molecule	total RNA	Label	Biotin
Extraction protocol	Qiagen RNAeasy	Label protocol	Biotinylated cRNA target was produced starting from 3 µg of total RN Affymetrix (Santa Clara, CA) instructions and for each sample 15 µg to a length of 20-200 bp
Label	Biotin	Hybridization protocol	All the hybridization, washing, staining procedures were done u Affymetrix station (Fluidics station 450), GeneChip Scanner 3000) as manufacturer.
Label protocol	Standard Illumina protocol	Scan protocol	Scanning was done using a Genechip Affymetrix station (GeneChip : raccomended by manufacturer.
Hybridization protocol	Standard Illumina protocol	Description	9 mesothelioma mixed type tumors were compared to the normal pleu
Scan protocol	Standard Illumina protocol	Data processing	Microarray quality control and statistical validation were performed u: The presence of hybridization/construction artifacts was evaluatec function. The probe (PM) intensity distribution was evaluated us (Bioconductor package affy). Probe set intensities were obtained by r and normalization was done according to quantiles me OneChannelGUI, a graphical interface to Bioconductor packages, wa of the described analyses.
Description	Cortex, KO mice, 2 MO 4697095041_H		
Data processing	BeadStudio, quantile normalization		

Исходные данные

- Три типа предсказаний
 - Пол: мужской, женский
 - Ткань: легкие, мозг, печень, мышцы
 - Мышиная линия: C57BL/6, FVB, BALB/C
- 10 типов микрочипов
- По 100 образцов для каждого класса для обучения и 100 – для тестирования

Задачи

- Найти дополнительные датасеты для обучения
- Подобрать метод машинного обучения для выполнения предсказаний
- Попробовать применить биологические знания для улучшения предсказаний