



Programming for Biomedical Informatics

Lecture 10 “Differential Gene Expression”

<https://github.com/tisimpson/pbi>

Ian Simpson
ian.simpson@ed.ac.uk

Differential Gene Expression Analysis Using NCBI-GEO & PyDESeq2

NCBI

Resources

How To

Sign in to NCBI

GEO Home

Documentation

Query & Browse

Email GEO

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.

Keyword or GEO Accession

Search

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

ENCODE Data Listings and Tracks

Browse Content

Repository Browser

DataSets:

4348

Series:

238523

Platforms:

26626

Samples:

7454950

Information for Submitters

Login to Submit

Submission Guidelines

Update Guidelines

MIAME Standards

Citing and Linking to GEO

Guidelines for Reviewers

GEO Publications

README

Code of conduct

MIT license

PyDESeq2

downloads 176k

downloads 12k

license MIT

PyDESeq2 is a python implementation of the DESeq2 method [1] for differential expression analysis (DEA) with bulk RNA-seq data, originally in R. It aims to facilitate DEA experiments for python users.

As PyDESeq2 is a re-implementation of DESeq2 from scratch, you may experience some differences in terms of retrieved values or available features.

Currently, available features broadly correspond to the default settings of DESeq2 (v1.34.0) for single-factor and multi-factor analysis (with categorical or continuous factors) using Wald tests. We plan to implement more in the future. In case there is a feature you would particularly like to be implemented, feel free to open an issue.

Table of Contents

PyDESeq2

Table of Contents

Installation

Requirements

Getting started

Documentation

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

<https://github.com/owkin/PyDESeq2>

Retrieving Data from NCBI GEO

Programmatic Access to NCBI GEO

https://www.ncbi.nlm.nih.gov/geo/info/geo_paccess.html

Download Options

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

Workflow

- Use the qualifier fields in Entrez GEO DataSets to fine-tune a search
- Construct the appropriate eSearch query in your script/program
- Run the query, retrieve the results in the form of UIDs or history parameters (query_key and WebEnv) as needed
- Run eSummary or eFetch and/or eLink depending on your needs to retrieve the final metadata or accessions.
- If you need to download full records or supplementary files, use the accession information to construct an FTP URL and download the data.



Programming for Biomedical Informatics

Next Lecture - Tuesday 29th October
“Biological Networks”

Ask Questions on the EdStem Discussion Board

Background

<https://github.com/tisimpson/pbi>



THE UNIVERSITY of EDINBURGH
informatics