***httrpathway***

An R Package for Performing Pathway-Based Concentration Response Modeling from High Throughput Transcriptomics Data

Authors: Richard Judson, Thomas Sheffield, Joshua Harrill, Imran Shah, Logan Everett, R. Woodrow Setzer, …

US EPA, Center for Computational Toxicology and Exposure

Contact: Judson.richard@epa.gov

Contents

[Introduction 1](#_Toc74143221)

[Environment 1](#_Toc74143222)

[Code Repository 1](#_Toc74143223)

[Directory Structure 1](#_Toc74143224)

[Reference Files 1](#_Toc74143225)

[Input Transcriptomics Data 2](#_Toc74143226)

[Running Signature-based Concentration Response Analyses 2](#_Toc74143227)

[Running Gene-based Concentration Response Analyses 2](#_Toc74143228)

# Introduction

The R package httrpathway was developed at the US Environmental Protection Agency (EPA) Center for Computational Toxicology and Exposure (CCTE) to analyze concentration-response high-throughput transcriptomics data (HTTr) at the pathway (gene signature) level. It uses relatively standard computational approaches but is designed for analyzing large data sets (hundreds to thousands of chemicals). A “signature” is simply a set of genes, and the package uses a large number of such signature sets from open sources, including responses to chemical perturbagens, diseases, transcription factor activation, etc. The input to the package is data at the level of log-2 fold changes (l2fc) for a set of probes measured in a set of chemical samples as multiple concentrations. The end result are concentration-response parameters (potency, efficacy, overall activity or hitcalls) for each chemical sample for each signature. The package can also perform gene-by-gene concentration-response modeling.

Because httrpathway is designed to operate on large data sets, calculations can take many days on even a parallel server and produce data sets of several gigabytes in size. To manage this, there are multiple intermediate and final files produced that are placed in specific directories. Therefore, this package is designed around a specific directory structure, described below.

# Environment

# Code Repository

# Directory Structure

# Reference Files

# Input Transcriptomics Data

# Running Signature-based Concentration Response Analyses

# Running Gene-based Concentration Response Analyses