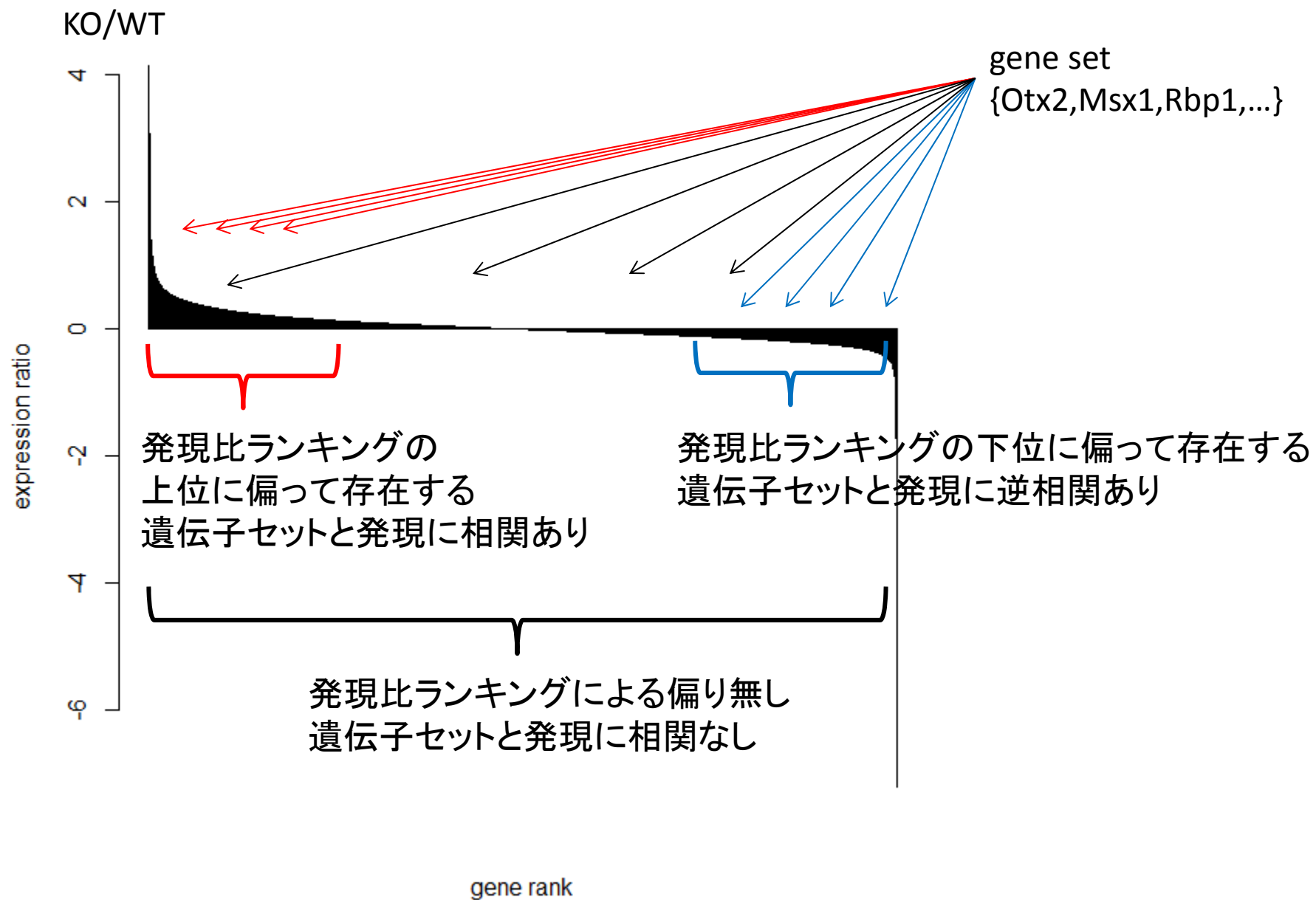




2016年度第5回 バイオインフォマティクス実習

発現変動遺伝子の機能解析
GSEA, pathway解析, GO解析

- Gene Set Enrichment Analysis (GSEA)
- 特定の遺伝子セットと発現比の間に相関があるか調べる



http://www.broadinstitute.org/gsea/index.jsp

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).

From this web site, you can:

- ▶ **Download** the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- ▶ **Explore the Molecular Signatures Database (MSigDB)**, a collection of annotated gene sets for use with GSEA software.
- ▶ **View documentation** describing GSEA and MSigDB.

What's New

23-Jan-2014: Version 2.0.14 of the GSEA desktop application is now available, which contains a number of upgrades and bug fixes. See the [GSEA v2.0.14 Release Notes](#) for details.

05-Jun-2013: Version 4.0 of the Molecular Signatures Database (MSigDB) is now available, which includes a new gene set collection (C7) of 1,910 immunologic signatures generated as part of the Human Immunology Project Consortium. We also released a newer version (2.0.13) of the GSEA desktop application. There were no changes to the GSEA algorithm.

Registration

Please register to download the GSEA software and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Contributors

GSEA and MSigDB are maintained by the GSEA team with the support of our MSigDB Scientific Advisory Board. Our thanks to our many contributors. Funded by: National Cancer Institute, National Institutes of Health, National Institute of General Medical Sciences.

Citing GSEA

To cite your use of the GSEA software, please reference Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550) and Mootha, Lindgren, et al. (2003, Nat Genet 34, 267-273).

Diagram: Molecular Profile Data and Gene Set Database feed into 'Run GSEA', which produces Enriched Sets (visualized as a plot).

DownloadセクションからGSEAを取得
Javaプログラム(OSに依存しない)
メールアドレスを登録する必要あり

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).

From this web site, you can:

- ▶ **Download** the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- ▶ **Explore the Molecular Signatures Database (MSigDB)**, a collection of annotated gene sets for use with GSEA software.
- ▶ **View documentation** describing GSEA and MSigDB.

What's New

23-Jan-2014: Version 2.0.14 of the GSEA desktop application is now available, which contains a number of upgrades and bug fixes. See the GSEA v2.0.14 Release Notes for details.

05-Jun-2013: Version 4.0 of the Molecular Signatures Database (MSigDB) is now available, which includes a new gene set collection (C7) of 1,910 immunologic signatures generated as part of the Human Immunology Project Consortium. We also released a newer version (2.0.13) of the GSEA desktop application. There were no changes to the GSEA algorithm.

Registration

Please register to download the GSEA software and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Contributors

GSEA and MSigDB are maintained by the GSEA team with the support of our MSigDB Scientific Advisory Board. Our thanks to our many contributors. Funded by: National Cancer Institute, National Institutes of Health, National Institute of General Medical Sciences.

Citing GSEA

To cite your use of the GSEA software, please reference Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550) and Mootha, Lindgren, et al. (2003, Nat Genet 34, 267-273).

Download

The screenshot shows the GSEA Downloads page on the Broad Institute website. The page lists four software options: javaGSEA Desktop Application, javaGSEA Java Jar file, GSEA Java Source Code Java source files, and R-GSEA R Script. Each option has a list of features and a download or launch link. Two Japanese annotations with arrows point to specific links: 'JNLPファイル その都度プログラムをダウンロードして実行する' points to the 'Launch' button for the Desktop Application, and 'Java実行ファイル' points to the 'download gsea2-2.1.0.jar' link for the Java Jar file.

Downloads

The GSEA software and source code and the Molecular Signatures Database (MSigDB) are freely available to individuals in both academia and industry for internal research purposes. Please see the GSEA/MSigDB license for more details.

Software

There are several options for GSEA software. All options implement exactly the same algorithm. Usage recommendations and installation instructions are listed below. Current Java implementations of GSEA require Java 6 or 7.

Software	Features	Download/Action
javaGSEA Desktop Application	<ul style="list-style-type: none">Easy-to-use graphical user interfaceRuns on any desktop computer (Windows, Mac OS X, Linux etc.) that supports Java 6 or 7Produces richly annotated reports of enrichment resultsIntegrated gene sets browser to view gene set annotations, search for gene sets and map gene sets between platforms	Launch with 1GB (for 32 or 64-bit Java) memory: Launch
javaGSEA Java Jar file	<ul style="list-style-type: none">Command line usageRuns on any platform that supports Java 6 or 7We recommend using the 'Launch' buttons above instead of this mode for most users	download gsea2-2.1.0.jar
GSEA Java Source Code Java source files	<ul style="list-style-type: none">100% Java implementation of GSEAIncorporate GSEA into your own data analysis pipelineProgrammatically call the open source GSEA java API	download gsea2_distrib-2.1.0.zip
R-GSEA R Script	<ul style="list-style-type: none">Usage from within the R programming environmentEasily inspect, learn and tweak the algorithmIncorporate GSEA into your own data analysis pipelineProgrammatically call the open source GSEA R APIClick here to learn more about the R-GSEA script	download GSEA-P-R-1.0.zip

JNLPファイル
その都度プログラムをダウンロードして実行する

Java実行ファイル

- 課題配布フォルダからgsea2-2.1.0を各自のデスクトップにコピー
- gsea2-2.1.0をダブルクリック

GSEA

GSEA v2.1.0 (Gene set enrichment analysis -- Broad Institute)

File Options Downloads Tools Help

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Gene set tools

- Chip2Chip mapping
- Browse MSigDB

Analysis history

GSEA reports


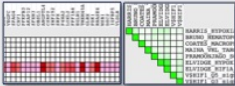
Processes: click 'status' field for results

Name	Status

Show results folder

Home

Steps in GSEA

- 1. What you need for GSEA**
 - Expression data set
 - Phenotype annotation
 - Gene sets – use MSigDB or your own gene sets
- 2. Run GSEA**
 - Start with default parameters
 - If you want to collapse probes to genes, specify chip platform
- 3. View results**
- 4. Leading edge analysis**
 - Leading edge finds genes driving enrichment results

Gene Set Tools

Chip2Chip mapping

- Convert gene sets between platforms

Chip2Chip mapping

Explore MSigDB gene sets

- Search the database of thousands of gene sets
- Browse the gene sets by name
- Find overlapping gene sets
- Export gene sets

Browse MSigDB

See also

- MSigDB online tools at: www.broadinstitute.org/msigdb

Getting Help

GSEA web site:

www.broadinstitute.org/gsea

GSEA documentation:

www.broadinstitute.org/gsea/wiki

Email the GSEA team:

gsea@broadinstitute.org

BROAD INSTITUTE

23:29:55 30M of 44M 23:29 2015/01/25

データファイルをload

GSEA v2.1.0 (Gene set enrichment analysis -- Broad Institute)

File Options Downloads Tools Help

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Gene set tools

- Chip2Chip mapping
- Browse MSigDB

Analysis history

GSEA reports


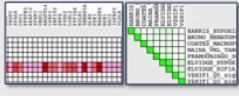
Processes: click 'status' field for results

Name	Status
------	--------

Show results folder

Home

Steps in GSEA

- 1. What you need for GSEA**
 - Expression data set
 - Phenotype annotation
 - Gene sets – use MSigDB or your own gene sets
- 2. Run GSEA**
 - Start with default parameters
 - If you want to collapse probes to genes, specify chip platform
- 3. View results**
- 4. Leading edge analysis**
 - Leading edge finds genes driving enrichment results

Gene Set Tools

Chip2Chip mapping

- Convert gene sets between platforms

Chip2Chip mapping

Explore MSigDB gene sets

- Search the database of thousands of gene sets
- Browse the gene sets by name
- Find overlapping gene sets
- Export gene sets

Browse MSigDB

See also

- MSigDB online tools at: www.broadinstitute.org/msigdb

Getting Help

GSEA web site:

www.broadinstitute.org/gsea

GSEA documentation:

www.broadinstitute.org/gsea/wiki

Email the GSEA team:

gsea@broadinstitute.org

BROAD INSTITUTE

23:29:55 30M of 44M 2015/01/25

データファイルのload

- 必要なファイルは3つ
- 発現プロファイル gctファイル
- 遺伝子セット grpファイル
- カテゴリー clsファイル

gctファイル

常に必要

遺伝子数

サンプル数

#1.2

21530	4				
NAME	Description	KO1	KO2	WT1	WT2
Ctss	NA	1730.1	1681.1	10.2	10.5
Ahnak	NA	1650.3	1510.1	11.3	14.2
...

常に必要

遺伝子名
大文字、小文字の区別に注意

ファイル名の拡張子はgct

grpファイル

#gene symbol
Evi1
Myct1
...

遺伝子名の羅列

gctファイルと大文字、小文字を一致させる
ファイル名の拡張子はgrp

clsファイル

サンプル数
クラス数
常に必要

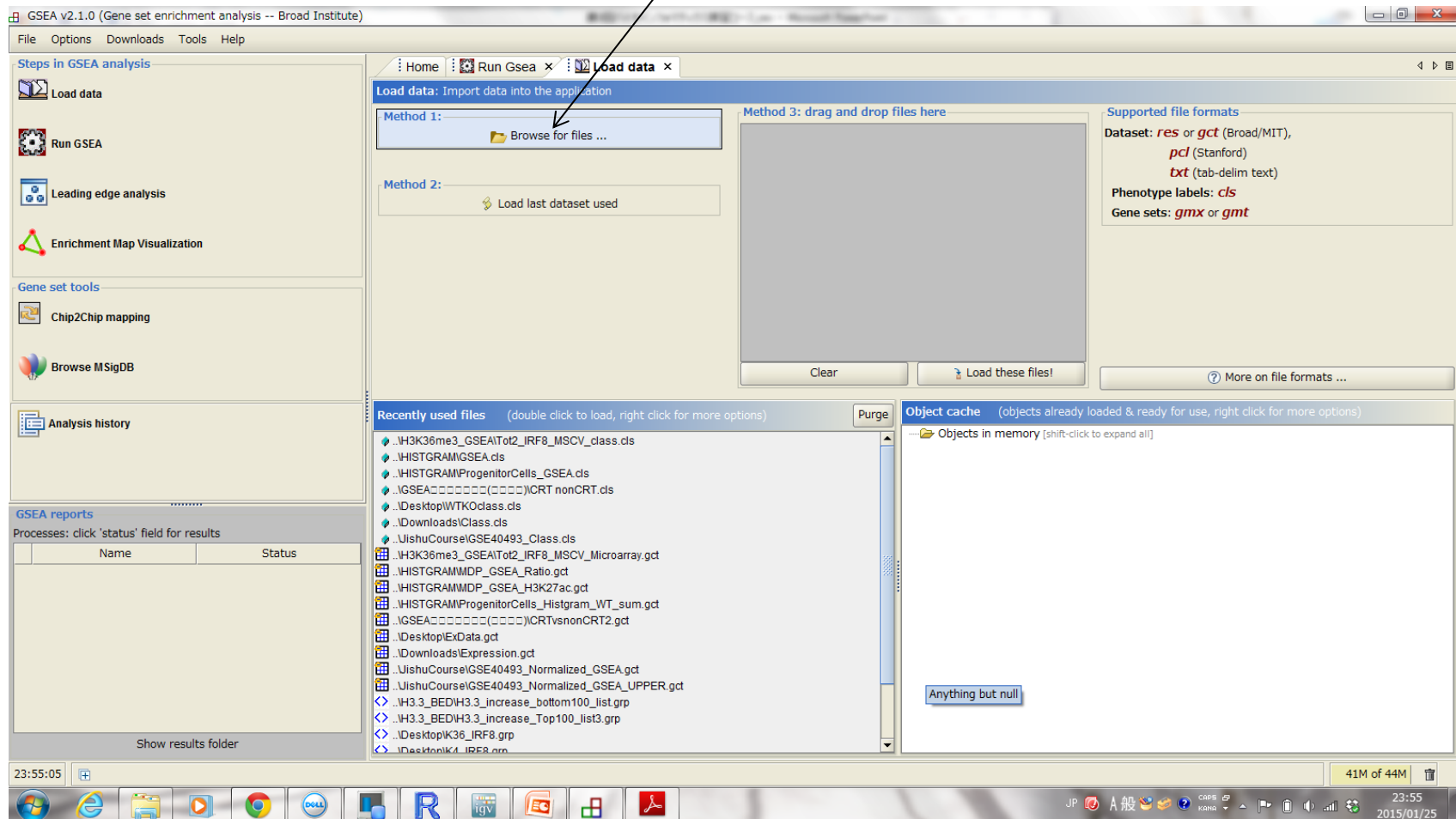
4 2 1
#KO WT
KO KO WT WT

clsファイルはスペース区切りのテキストファイル
拡張子はcls

- 課題配布フォルダから
- GSE40493_Normalized_GSEA_UPPER.gct
- geneset_Bcl6.grp,geneset_BRAIN.grp
- GSE40493_Class.cls
- 各ファイルを各自のデスクトップフォルダへコピー

Load Data

Browse for filesをクリックしてファイルを選択



Run

Run GSEAをクリックして実行

GSEA v2.1.0 (Gene set enrichment analysis -- Broad Institute)

File Options Downloads Tools Help

Steps in GSEA analysis

- Load data
- Run GSEA**
- Leading edge analysis
- Enrichment Map Visualization

Gene set tools

- Chip2Chip mapping
- Browse MSigDB

Analysis history

GSEA reports

Processes: click 'status' field for results

Name	Status
------	--------

Show results folder

Home

Steps in GSEA

- 1. What you need for GSEA**
 - Expression data set
 - Phenotype annotation
 - Gene sets – use MSigDB or your own gene sets
- 2. Run GSEA**
 - Start with default parameters
 - If you want to collapse probes to genes, specify chip platform
- 3. View results**
 - Enrichment in phenotype: [view the results](#)
 - Enrichment in phenotype: [view the results](#)
- 4. Leading edge analysis**
 - Leading edge finds genes driving enrichment results

Gene Set Tools

Chip2Chip mapping

- Convert gene sets between platforms

[Chip2Chip mapping](#)

Explore MSigDB gene sets

- Search the database of thousands of gene sets
- Browse the gene sets by name
- Find overlapping gene sets
- Export gene sets

[Browse MSigDB](#)

See also

- MSigDB online tools at: www.broadinstitute.org/msigdb

Getting Help

GSEA web site:
www.broadinstitute.org/gsea

GSEA documentation:
www.broadinstitute.org/gsea/wiki

Email the GSEA team:
gsea@broadinstitute.org

BROAD INSTITUTE

23:29:55 30M of 44M

23:29 2015/01/25

Run

The screenshot shows the GSEA v2.1.0 application window. The left sidebar contains a 'Steps in GSEA analysis' section with icons for 'Load data', 'Run GSEA', 'Leading edge analysis', and 'Enrichment Map Visualization'. Below this is a 'Gene set tools' section with 'Chip2Chip mapping' and 'Browse MSigDB'. At the bottom of the sidebar is an 'Analysis history' section. The main panel is titled 'Gsea: Set parameters and run enrichment tests' and contains several sections: 'Required fields' with dropdowns for 'Expression dataset' (GSE40493_Normalized_GSEA_UPPER), 'Gene sets database' (Users\Jun\Dropbox\BioinformaticsStudy\JishuCourse\geneset_Bcl6.grp), 'Number of permutations' (1000), 'Phenotype labels' (BioinformaticsStudy\JishuCourse\GSE40493_Class.cls#KO_versus_WT), 'Collapse dataset to gene symbols' (false), 'Permutation type' (gene_set), and 'Chip platform(s)'. Below these are 'Basic fields' and 'Advanced fields' sections, each with a 'Show' button. At the bottom of the main panel is a 'Run' button. The bottom status bar shows '23:58:31', '4535 [INFO] Parsed from unigene / gene symbol: 38870', and '99M of 247M'. Japanese annotations with arrows point to various elements: 'gctファイルを選択' points to the 'Expression dataset' dropdown; 'grpファイルを選択' points to the 'Gene sets database' dropdown; '発現比の方向 WT/KO KO/WT' points to the 'Phenotype labels' dropdown; 'false' points to the 'Collapse dataset to gene symbols' dropdown; 'gene_set' points to the 'Permutation type' dropdown; 'runをクリックして実行' points to the 'Run' button; and 'ステータスが表示 Successと表示されたらクリック 結果を確認' points to the 'Status' column in the 'GSEA reports' table.

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Gene set tools

- Chip2Chip mapping
- Browse MSigDB

Analysis history

GSEA reports

Processes: click 'status' field for results

	Name	Status
1	Gsea	Success 5

Required fields

Expression dataset: GSE40493_Normalized_GSEA_UPPER [21535x8 (ann: 21535,8,chip na)]

Gene sets database: Users\Jun\Dropbox\BioinformaticsStudy\JishuCourse\geneset_Bcl6.grp

Number of permutations: 1000

Phenotype labels: BioinformaticsStudy\JishuCourse\GSE40493_Class.cls#KO_versus_WT

Collapse dataset to gene symbols: false

Permutation type: gene_set

Chip platform(s):

Basic fields: Show

Advanced fields: Show

Run

Reset Last Command Low (cpu usage)

23:58:31 4535 [INFO] Parsed from unigene / gene symbol: 38870 99M of 247M

JP A 般 CAPS KANA

23:58 2015/01/25

gctファイルを選択

grpファイルを選択

発現比の方向
WT/KO KO/WT

false

gene_set

runをクリックして実行

ステータスが表示
Successと表示されたらクリック
結果を確認

ブラウザ上で結果を表示

GSEA Report for Dataset GSE40493_Normalized_GSEA_UPPER

Enrichment in phenotype: KO (4 samples)

- None of the gene sets are enriched in phenotype KO
- [Guide to interpret results](#)

Enrichment in phenotype: WT (4 samples)

- 1 / 1 gene sets are upregulated in phenotype WT
- 1 gene sets are significantly enriched at FDR < 25%
- 1 gene sets are significantly enriched at nominal pvalue < 1%
- 1 gene sets are significantly enriched at nominal pvalue < 5%
- [Snapshot of enrichment results](#)
- Detailed [enrichment results in html](#) format
- Detailed [enrichment results in excel](#) format (tab delimited text)
- [Guide to interpret results](#)

Dataset details

- The dataset has 21535 features (genes)
- No probe set => gene symbol collapsing was requested, so all 21535 features were used

Gene set details

- Gene set size filters (min=15, max=500) resulted in filtering out 0 / 1 gene sets
- The remaining 1 gene sets were used in the analysis
- List of [gene sets used and their sizes](#) (restricted to features in the specified dataset)

Gene markers for the KO versus WT comparison

- The dataset has 21535 features (genes)
- # of markers for phenotype KO: 12807 (59.5%) with correlation area 51.8%
- # of markers for phenotype WT: 8728 (40.5%) with correlation area 48.2%
- Detailed [rank ordered gene list](#) for all features in the dataset
- [Heat map and gene list correlation](#) profile for all features in the dataset

Global statistics and plots

enrichment result in htmlをクリック

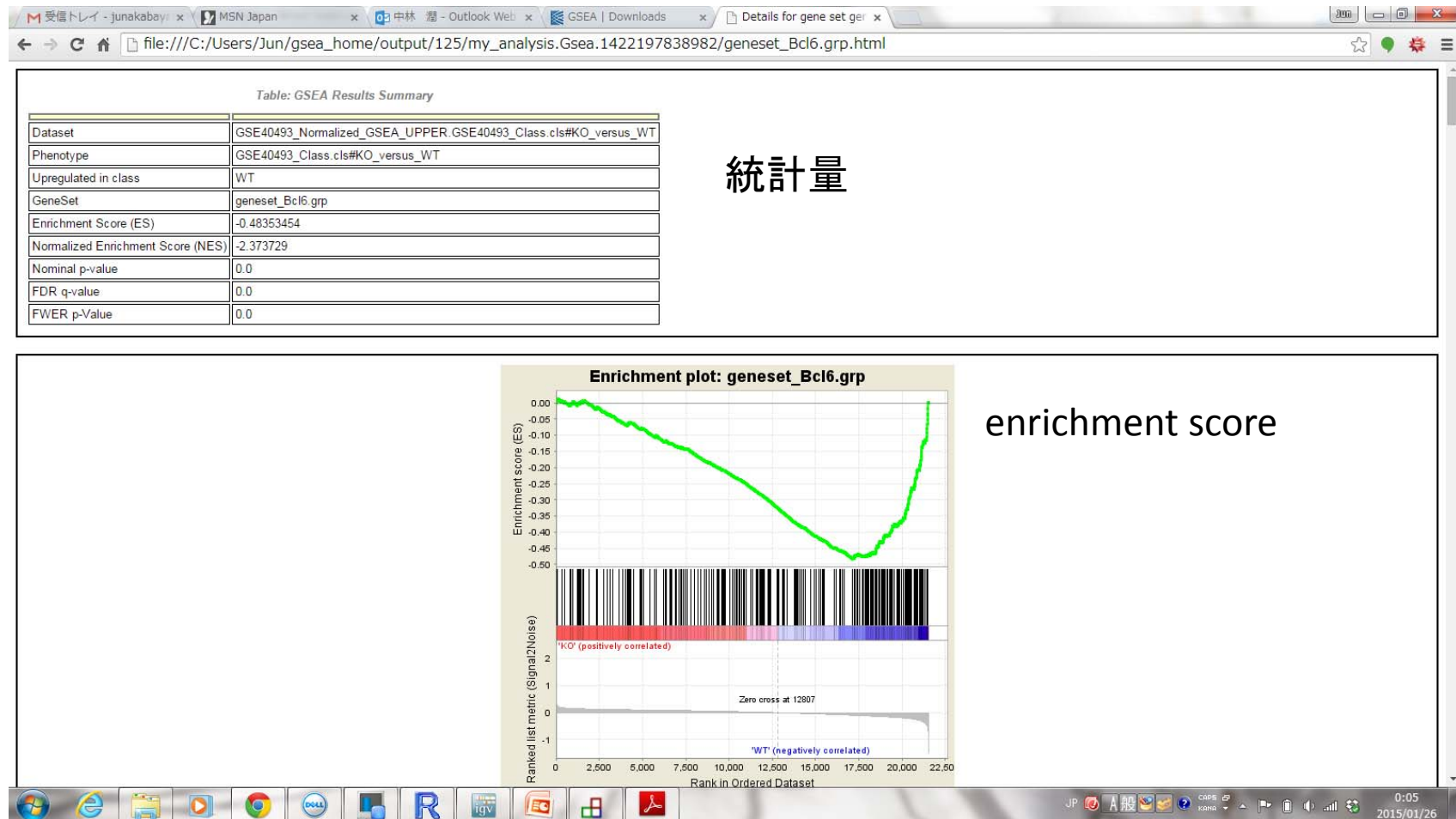
ブラウザ上で結果を表示

Table: Gene sets enriched in phenotype WT (4 samples) [\[plain text format\]](#)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	geneset_Bcl6.gpr	Details...	261	-0.48	-2.37	0.000	0.000	0.000	4412	tags=43%, list=20%, signal=53%

detailsをクリック

ブラウザ上で結果を表示



結果のファイル

- 結果はgsea_homeフォルダに自動的に保存されます。

Gene Ontology解析

- GO : 人が定義した遺伝子のアノテーション情報で、3つの階層を持つ。

Molecular Function MF

Cellular Component CC

Biological Process BP

generic gene ontology term finder

- <http://go.Princeton.edu/cgi-bin/GOTermFinder>

入力欄に遺伝子名のリストを
入力

or

遺伝子リストのテキストファイ
ルを選択

生物種を選択

クリックして検索

GENERIC GENE ONTOLOGY (GO) TERM FINDER

Welcome to the **GoTermFinder**, a tool for finding significant GO terms shared among a list of genes from your organism of choice, helping you discover what they may have in common.

This implementation, developed at the Lewis-Sigler Institute at Princeton, depends on the [GO_TermFinder](#) software written by Gavin Sherlock and Shuai Wang. It is made publicly available through the [JMKO project](#). For more information, please see [Lewis et al. Bioinformatics 2004](#).

NEW There is a new tool available with much the same functionality as GoTermFinder, only it is much more efficient. Although the backend has been tested extensively, please consider [LUGO](#) to be in **beta** at this time. Drop us a note to tell us what you think.

Required Basic Input Options

1. Enter a list of genes, one per line. [GO sample gene list](#)

OR

Upload a file containing lists of genes: [ファイルを選択](#) [選択されていません](#) [\[CLEAR\]](#)

This version of the Generic GO Term Finder has a batch mode to process multiple gene lists in parallel. To use this feature, upload an archive in tar, tar.gz, tgz, or zip format. Please refer to the [help document](#) for instructions.

Batch processing is done on one of our grid systems. Processing time will depend on the grid's load.

For long running jobs, it is highly recommended that an email address be provided so that notification may be sent when results are ready. If you do not do this, you will have to keep your browser window open to monitor the progress.

A notification email address is required for batch jobs.

Please enter a notification email address:

And please enter it again for confirmation:

2. Choose 1 of the 3 ontology aspects: ☐ Process ☐ Function ☐ Component

3. Choose [organism](#): [\[Go to update your own in the advanced options\]](#)

4. Plain text will be produced. Choose additional output format(s): ☒ HTML table ☒ GO tree view images

[Search for GO Terms](#) [Reset Form](#)

Optional Advanced Input Options

Enter the number of products estimated for your organism (e.g. roughly 7000 for *Saccharomyces cerevisiae*):

OR provide a list of genes for the background population: [ファイルを選択](#) [選択されていません](#) [\[CLEAR\]](#)

Enter p-value cutoff for significant shared GO terms search (e.g. 0.01 is the default p-value cutoff):

☒ Bonferroni correction for p-values?

☒ Calculate false discovery rate (FDR)?

☒ Follow regulation links? (regulates, positively_regulates, negatively_regulates)

Enter URL for your organism (e.g. <http://db.yeastgenome.org/cgi-bin/GO/GOcous.pl?house> is the default url for *Saccharomyces cerevisiae*):

For batch mode, enter the extension for files containing gene lists (examples: list.txt, list.xls)

Upload a custom gene association file: [ファイルを選択](#) [選択されていません](#) [\[CLEAR\]](#)

Select evidence codes: ☐ EXP ☐ IDA ☐ IMP ☐ IGI ☐ IEA ☐ ISS ☐ ISO ☐ ISA ☐ ISM ☐ IOC ☐ RCA ☐ TAS ☐ NAS ☐ IND

To exclude: ☐ IEA inferred from Electronic Annotation associations are included by default. Select this checkbox to exclude them.

Enter a comma to include in text result files: ☒ Query parameters ☒ Duplicated, discarded, ambiguous, and unknown identifiers

[Search for GO Terms](#) [Reset Form](#)

KEGG Pathway

- <http://www.genome.jp/kegg/>

生命システム情報統合データベース

分子レベルから細胞、個体、生態までの情報を取り揃えている

- KEGG mapper – Search pathway

http://www.genome.jp/kegg/tool/map_pathway1.html

生物種を選択

ヒト: hsa

マウス: mmu

入力欄に遺伝子名のリストを
入力

or

遺伝子リストのテキストファイル
を選択

クリックして検索

KEGG Mapper - Search Pathway

About KEGG Mapper

- Search Pathway
- Search&Color Pathway
- Color Pathway
- Color Pathway WebGL
- Search Brite
- Search&Color Brite
- Join Brite
- Join Brite Table
- Search Module
- Search&Color Module
- Search Disease
- Reconstruct Pathway
- Reconstruct Brite
- Reconstruct Module
- Map Taxonomy
- Convert ID
- Annotate Sequence BlastKOALA
- KEGG Atlas
- KEGG

Search against: Enter: map, ko, ec, rn, hsadd, or

Enter objects:

Examples:

Alternatively, enter the file name containing the data:

選択されていません

Filter1 Filter2 (to extract K/C/G/D/R/RP/RC numbers)

☒ Include aliases

☒ Display objects not found in the search

☐ Search pathways containing all the objects (AND search)

Search Pathway is the basic pathway mapping tool, where given objects (genes, proteins, compounds, glycans, reactions, drugs, etc.) are searched against KEGG pathway maps and found objects are marked in red. The objects in different types of **pathway maps** are specified by the following **KEGG identifiers** and aliases.

Prefix	Type	KEGG identifier	Alias
map	Reference pathway - metabolic	K/R/EC numbers C/G/D numbers	KO alias
map	Reference pathway - non-metabolic	K number C/G/D numbers	KO alias
ko	Reference pathway (KO)	K number C/G/D numbers	KO alias EC numbers
ec	Reference pathway (EC)	EC number C/G/D numbers	
rn	Reference pathway (Reaction)	R number C/G/D numbers	RP/RC numbers
org	Organism-specific pathway	gene identifier C/G/D numbers	gene alias (gene name) K/EC numbers

Last updated: June 10, 2014