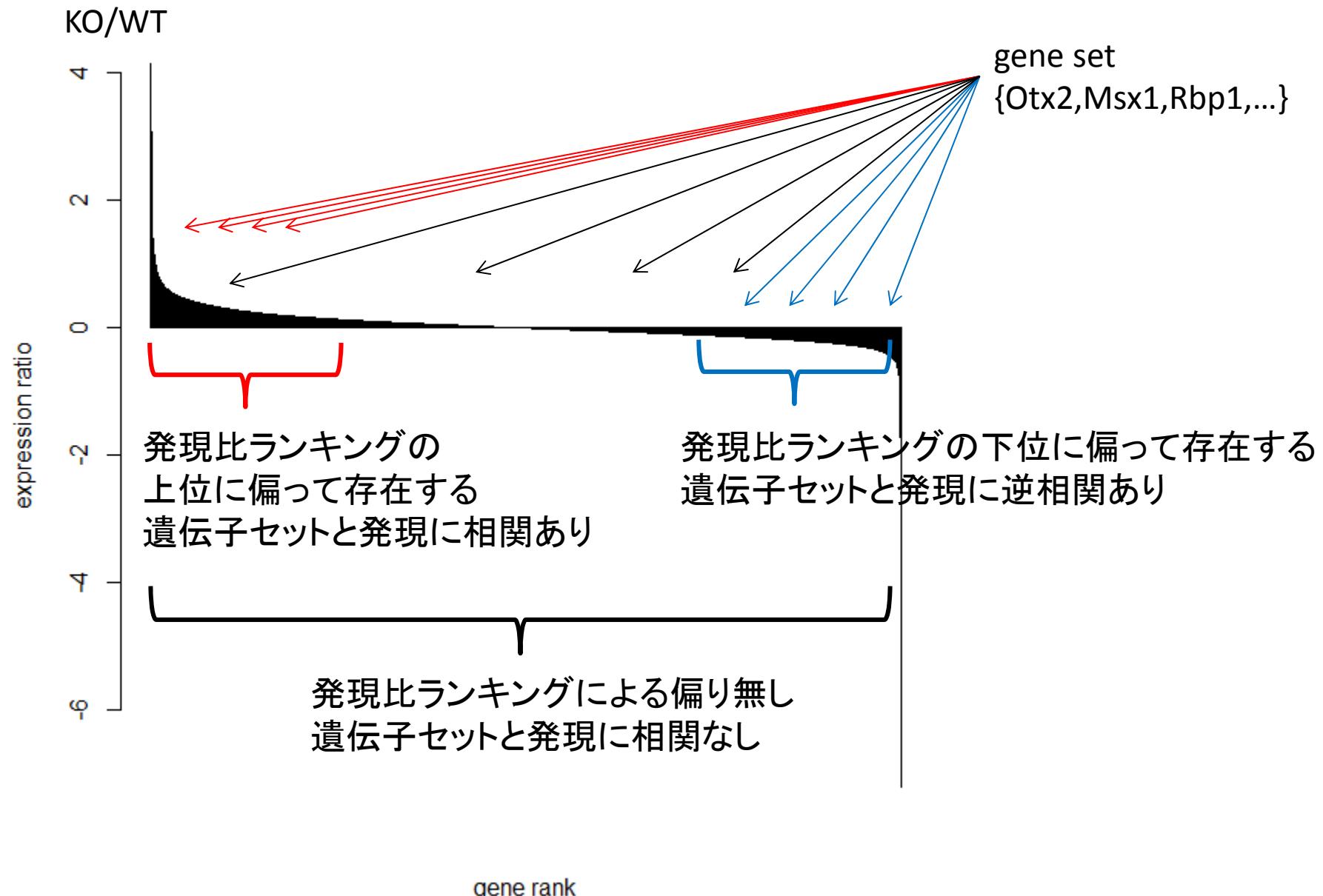




2016年度第5回 バイオインフォマティクス実習

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

受信トレイ - junakabayashi x MSN Japan x 中林 潤 - Outlook Web Access x GSEA x

www.broadinstitute.org/gsea/index.jsp

logged in as nakabayashi@yokohama-cu.ac.jp

BROAD INSTITUTE

GSEA

Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether *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.

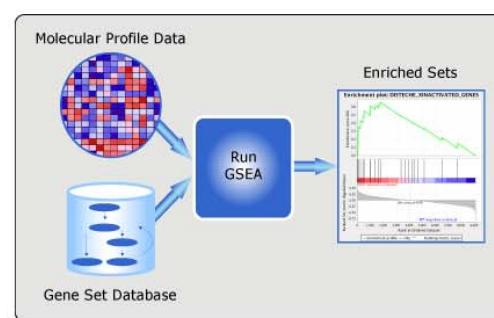
05-Jun-2013: Version 4.0 of the Molecular Signatures Database (MSigDB) is now available, which includes a new gene set collection (C7) of 1,910 immunologic signatures generated as part of the Human Immunology Project Consortium. We also released a newer version (2.0.13) of the GSEA desktop application.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

Molecular Profile Data

Gene Set Database

Run GSEA

Enriched Sets



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.

NATIONAL CANCER INSTITUTE

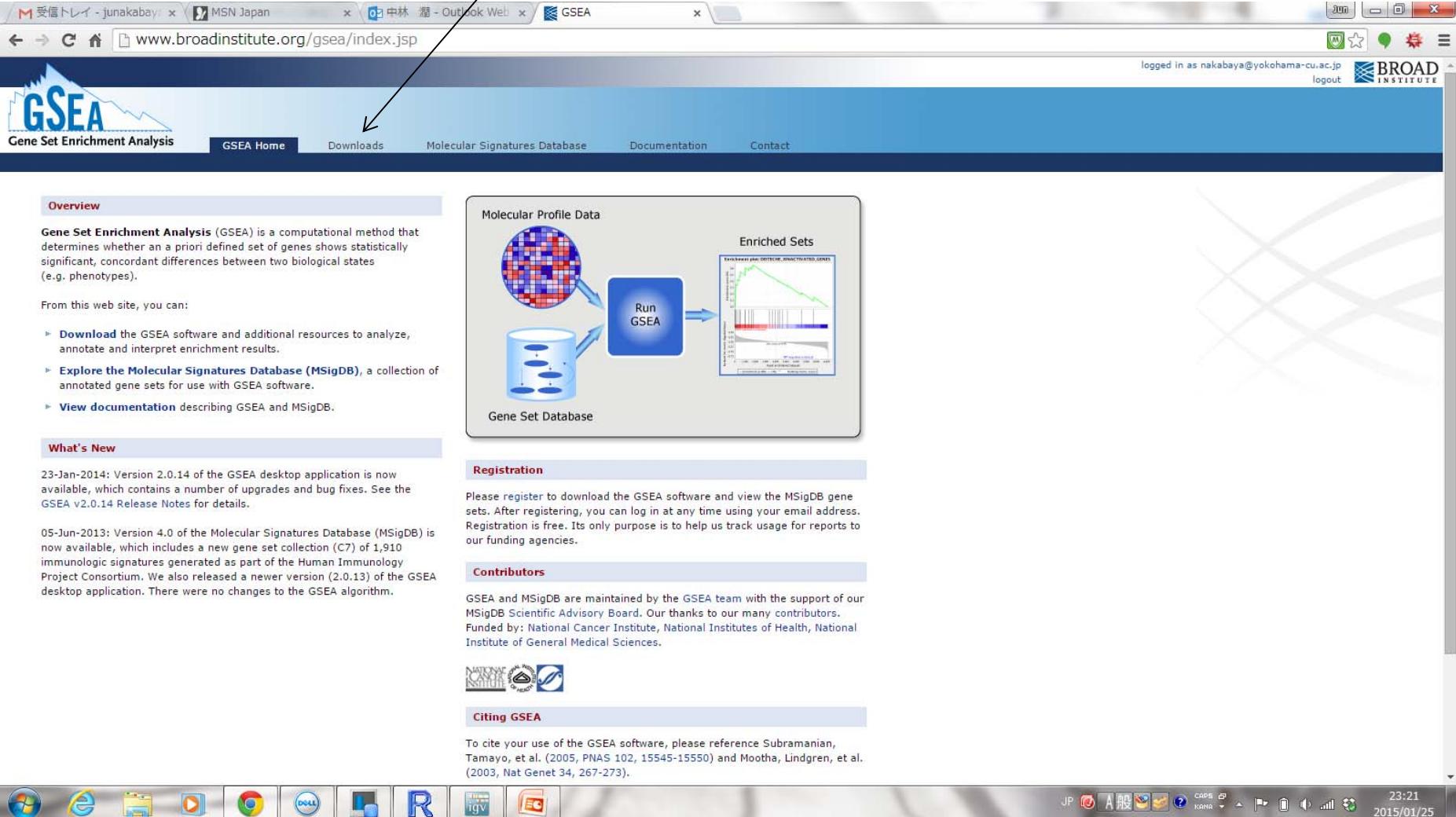
Citing GSEA

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).

JP A 航 CHIPS KIRI 23:21 2015/01/25

Windows Internet Explorer Google Chrome Dell R IGV EC

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



The screenshot shows the GSEA (Gene Set Enrichment Analysis) website at www.broadinstitute.org/gsea/index.jsp. A black arrow points from the text "DownloadセクションからGSEAを取得" to the "Downloads" menu item in the top navigation bar. The "Downloads" section contains a diagram illustrating the GSEA workflow: "Molecular Profile Data" (represented by a circular heatmap) and "Gene Set Database" (represented by a cylinder) feed into a central box labeled "Run GSEA", which then produces "Enriched Sets" (represented by a bar chart and a line graph). The top right of the page shows a user is logged in as "nakabaya@yokohama-cu.ac.jp" with a "logout" link and the "BROAD INSTITUTE" logo. The bottom of the page shows a Windows taskbar with various icons and the date "2015/01/25" and time "23:21".

Download

The screenshot shows the GSEA Downloads page on a computer screen. The page has a blue header with the GSEA logo and navigation links for Home, Downloads, Molecular Signatures Database, Documentation, and Contact. The main content is titled 'Downloads' and includes a note about the software being freely available for research purposes. It lists four 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' link. A dropdown menu for memory is open over the first option, showing '1GB (for 32 or 64-bit Java)'. A large Japanese annotation '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' (JNLP file, download and run the program each time) is overlaid on the page, with an arrow pointing to the 'Launch' button. Another Japanese annotation 'Java実行ファイル' (Java execution file) is overlaid on the 'download' 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.

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
 - Enrichment in phenotype: mouse vs. non-mouse
 - Enrichment in phenotype: liver vs. non-liver
4. Leading edge analysis
 - Leading edge finds genes driving enrichment results

Gene Set Tools

Chip2Chip mapping

- Convert gene sets between platforms

Chip2Chip mapping

Getting Help

GSEA web site:
www.broadinstitute.org/gsea

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

Email the GSEA team:
gsea@broadinstitute.org

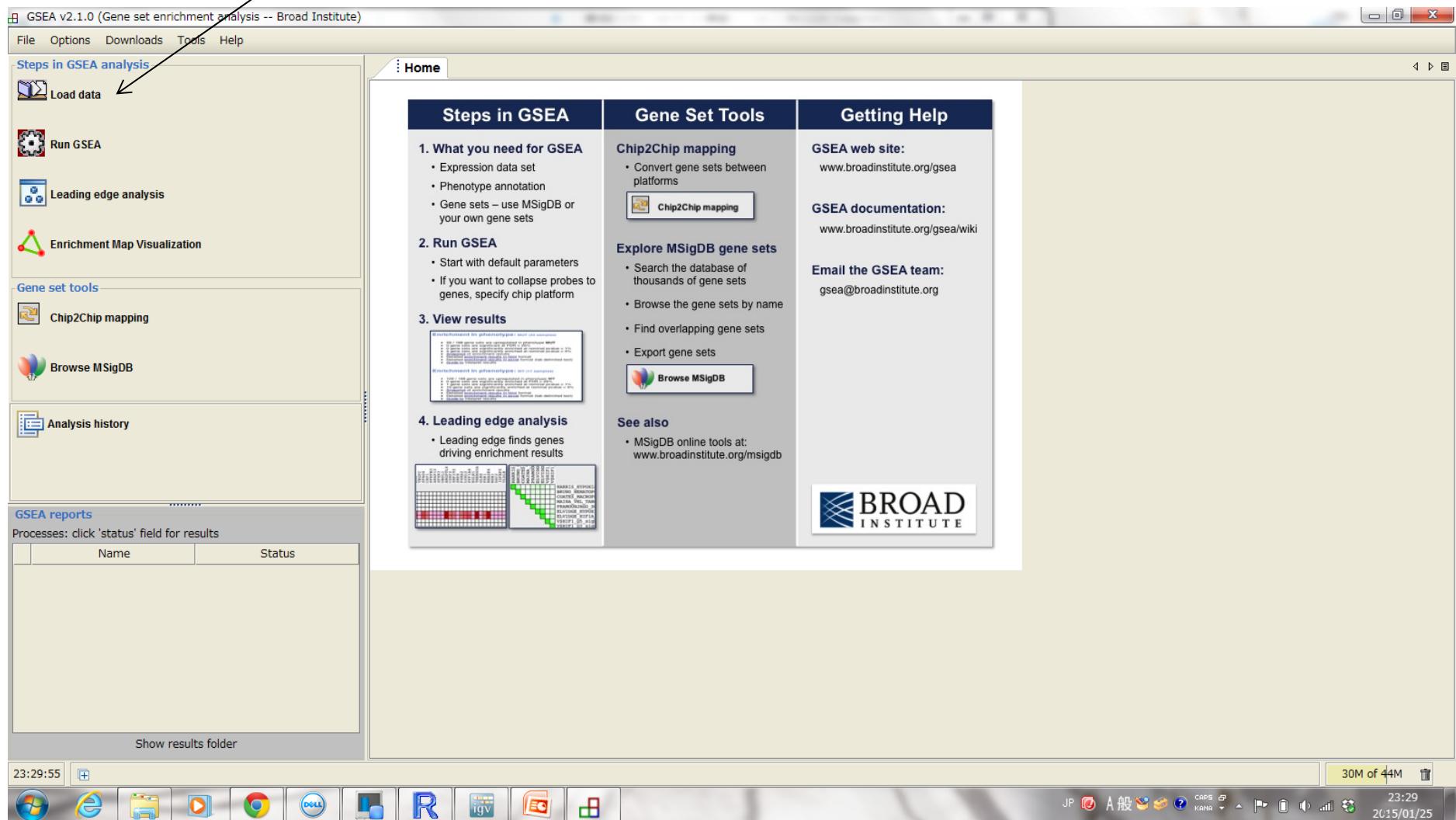
See also

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

BROAD INSTITUTE

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

データファイルをload



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

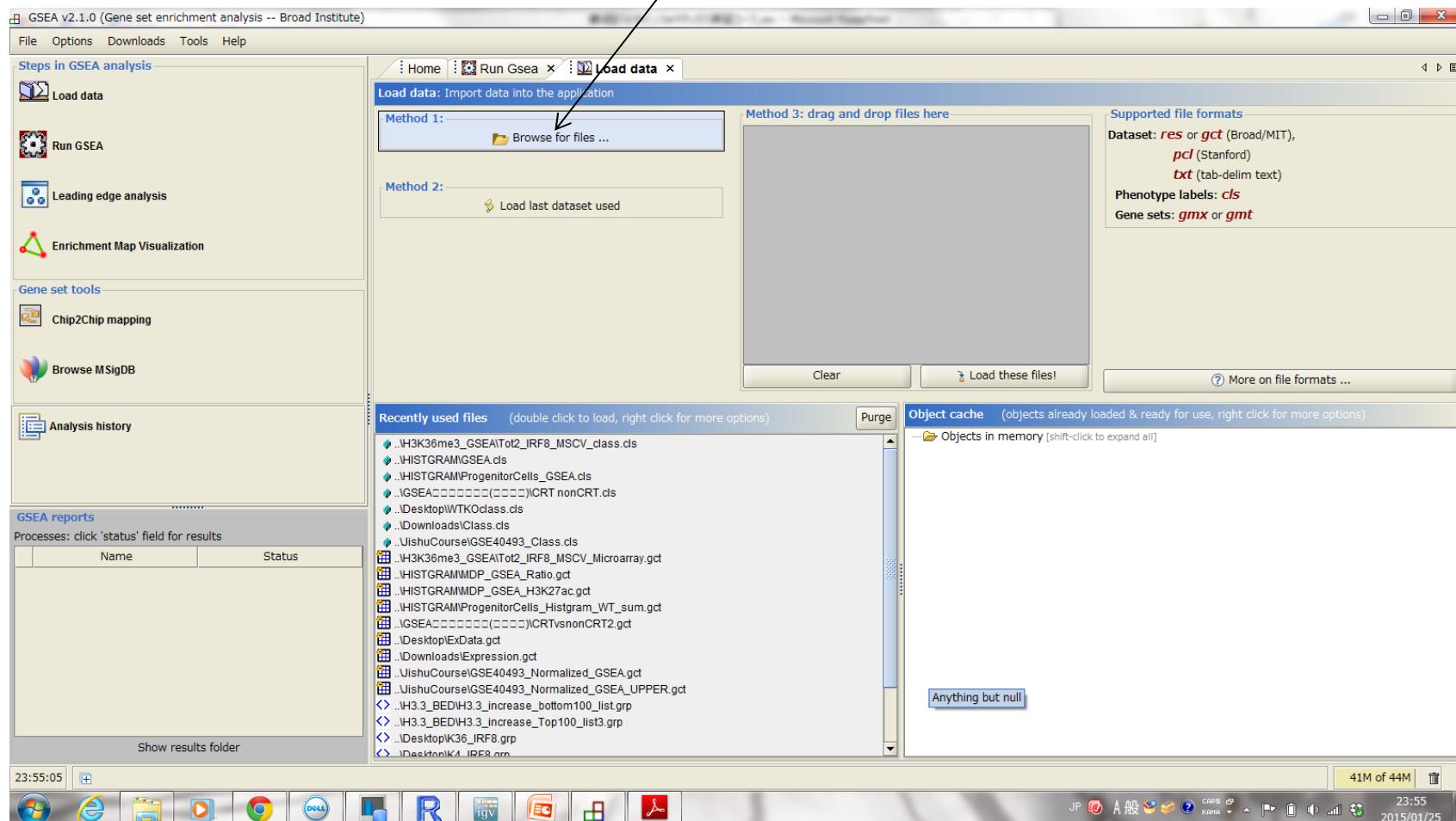
サンプル数
クラス数
常に必要
4 2 1
#KO WT
KO KO WT WT

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

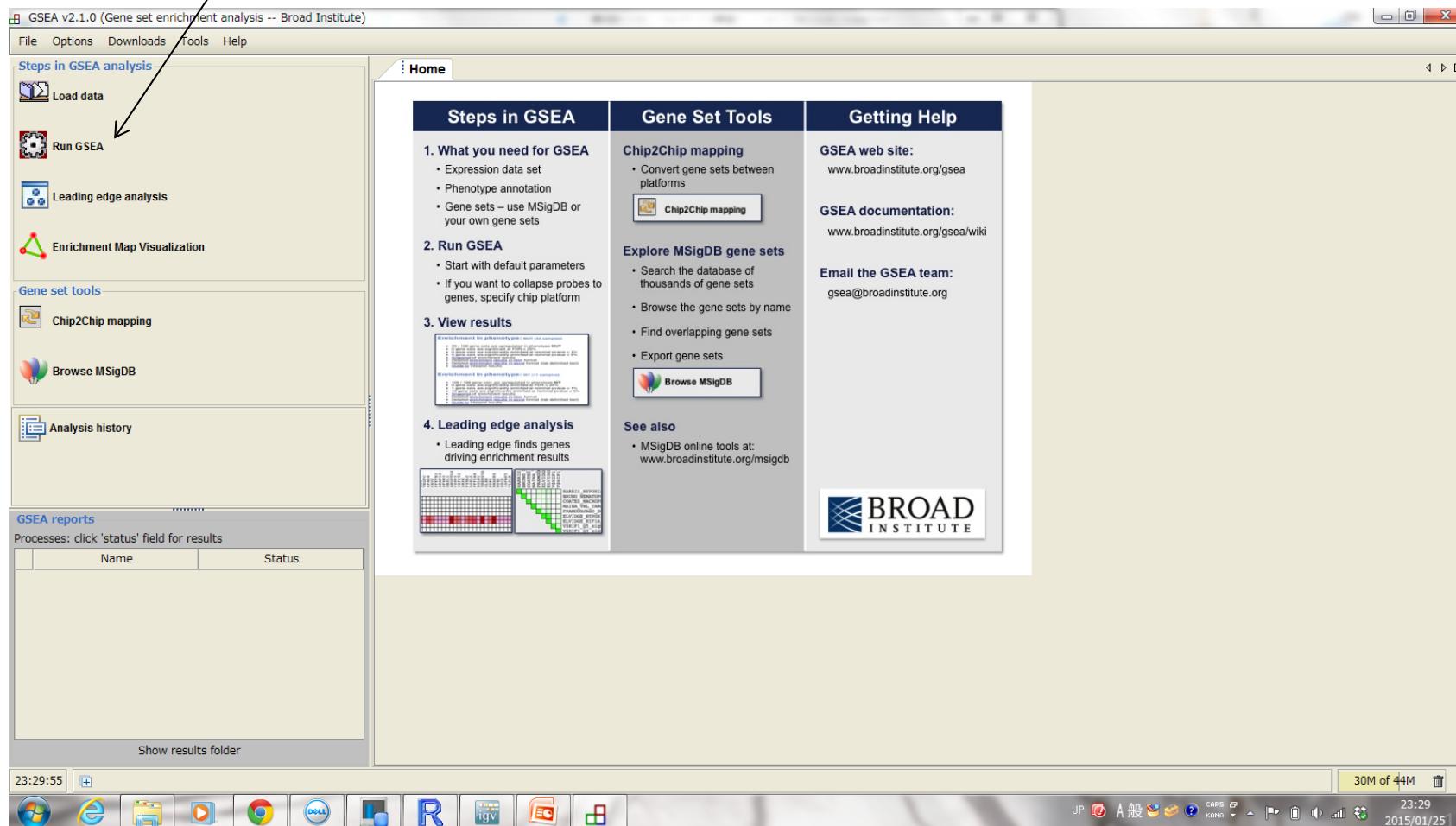
Load Data

Browse for filesをクリックしてファイルを選択

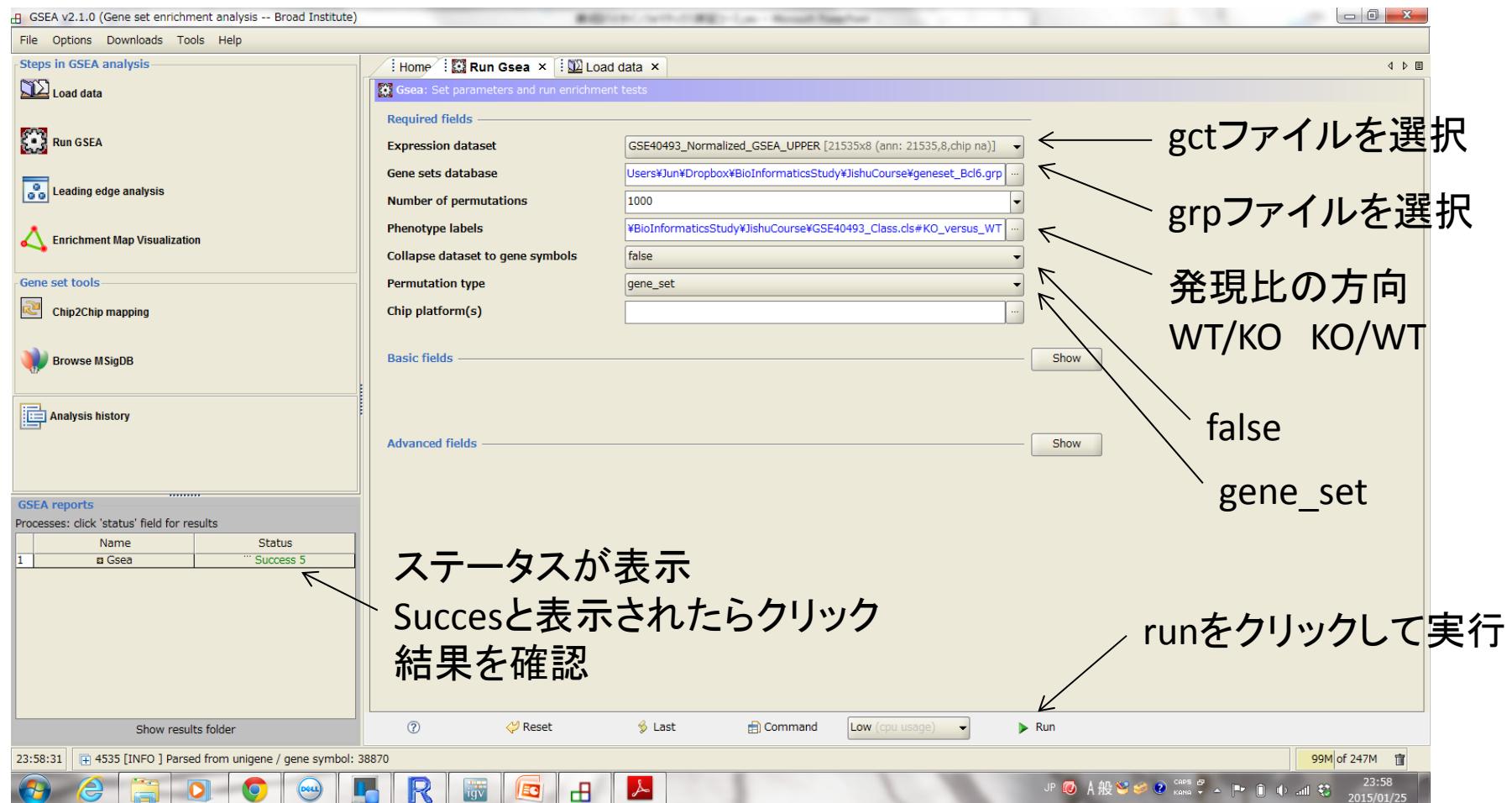


Run

Run GSEAをクリックして実行



Run



ブラウザ上で結果を表示

受信トレイ - junakabayi × MSN Japan × 中林 - Outlook Web × GSEA | Downloads × Index for xtools.gsea.G ×

file:///C:/Users/Jun/gsea_home/output/125/my_analysis.Gsea.1422197838982/index.html

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

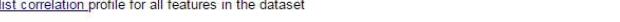
Gene set details

- 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



JP A股 ? CAPS 0:02
2015/01/26

ブラウザ上で結果を表示

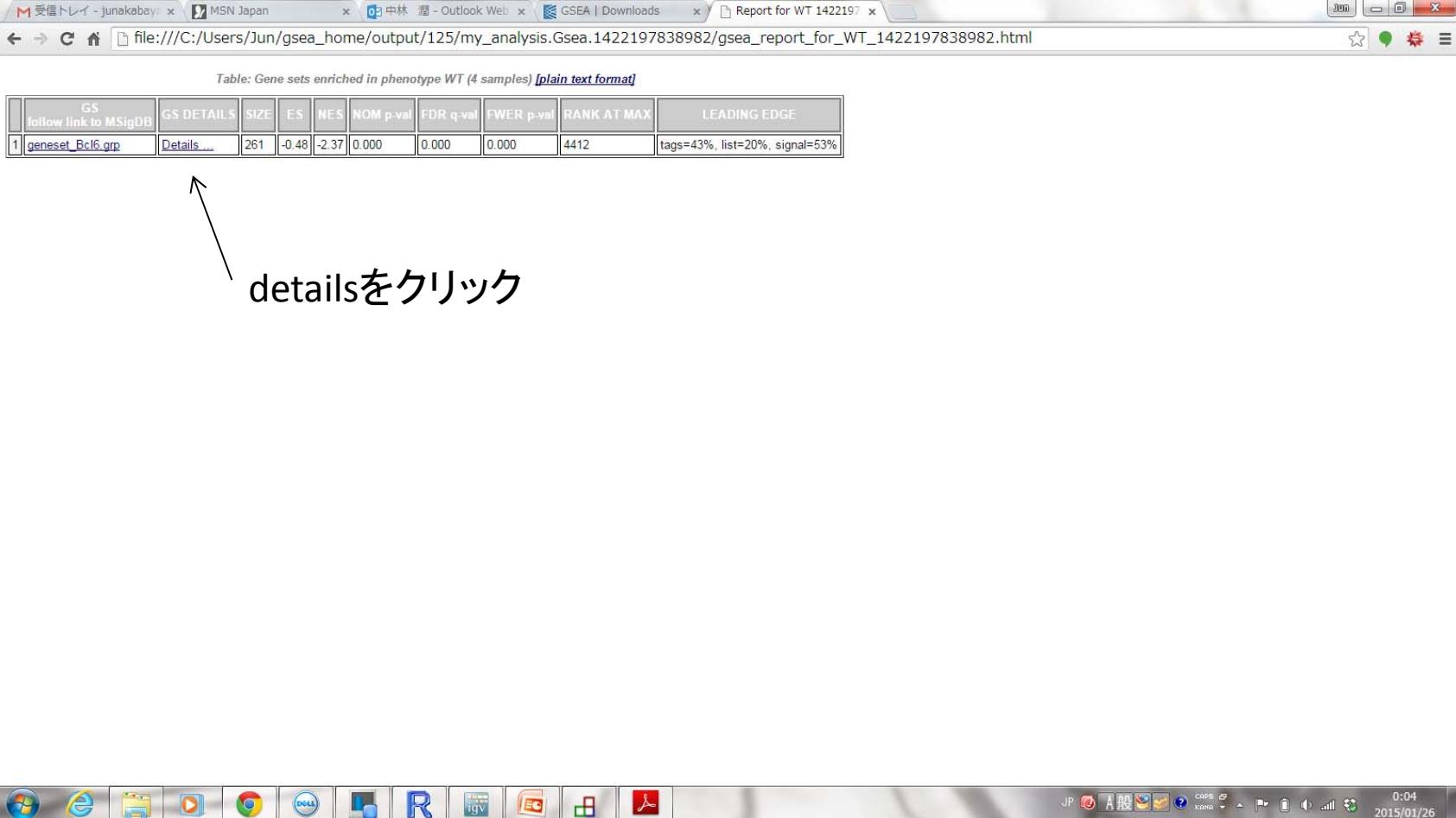


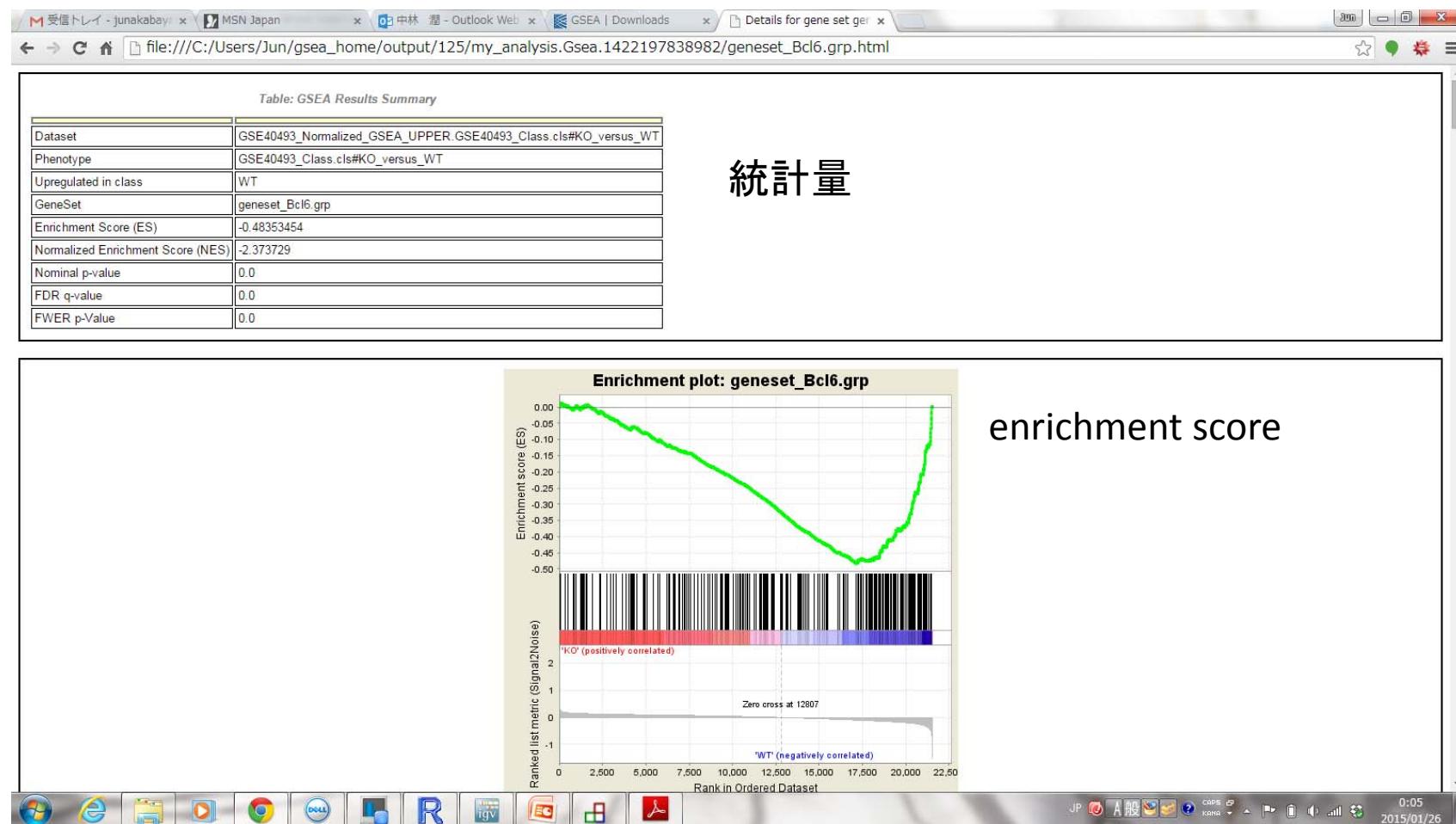
Table: Gene sets enriched in phenotype WT (4 samples) [plain text format]

	GS	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	geneset_Bcl6.grp	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>

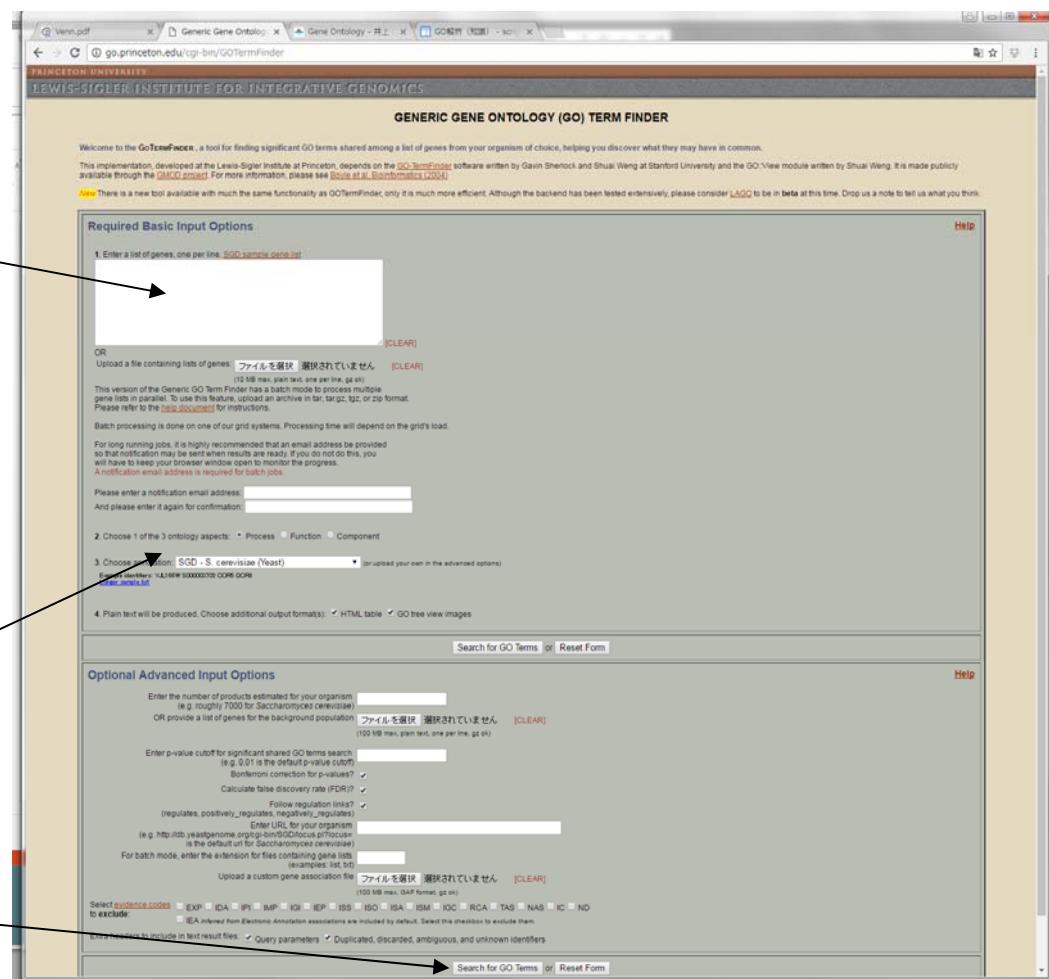
入力欄に遺伝子名のリストを
入力

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



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

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

Enter objects:

Examples: Select

Alternatively, enter the file name containing the data:

ファイルを選択 指定されていません

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

Include aliases

Display objects not found in the search

Search pathways containing all the objects (AND search)

Exec Clear

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