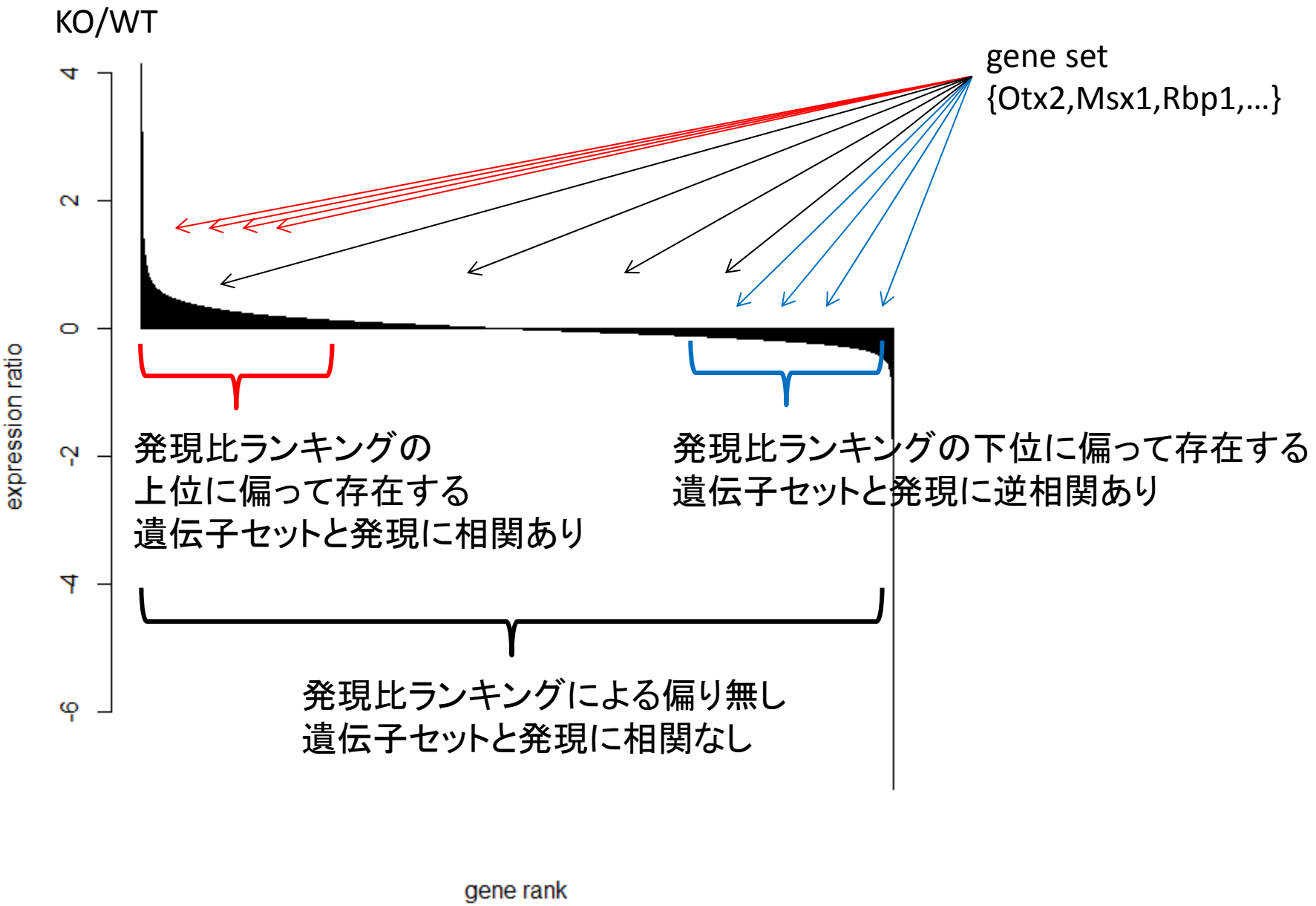
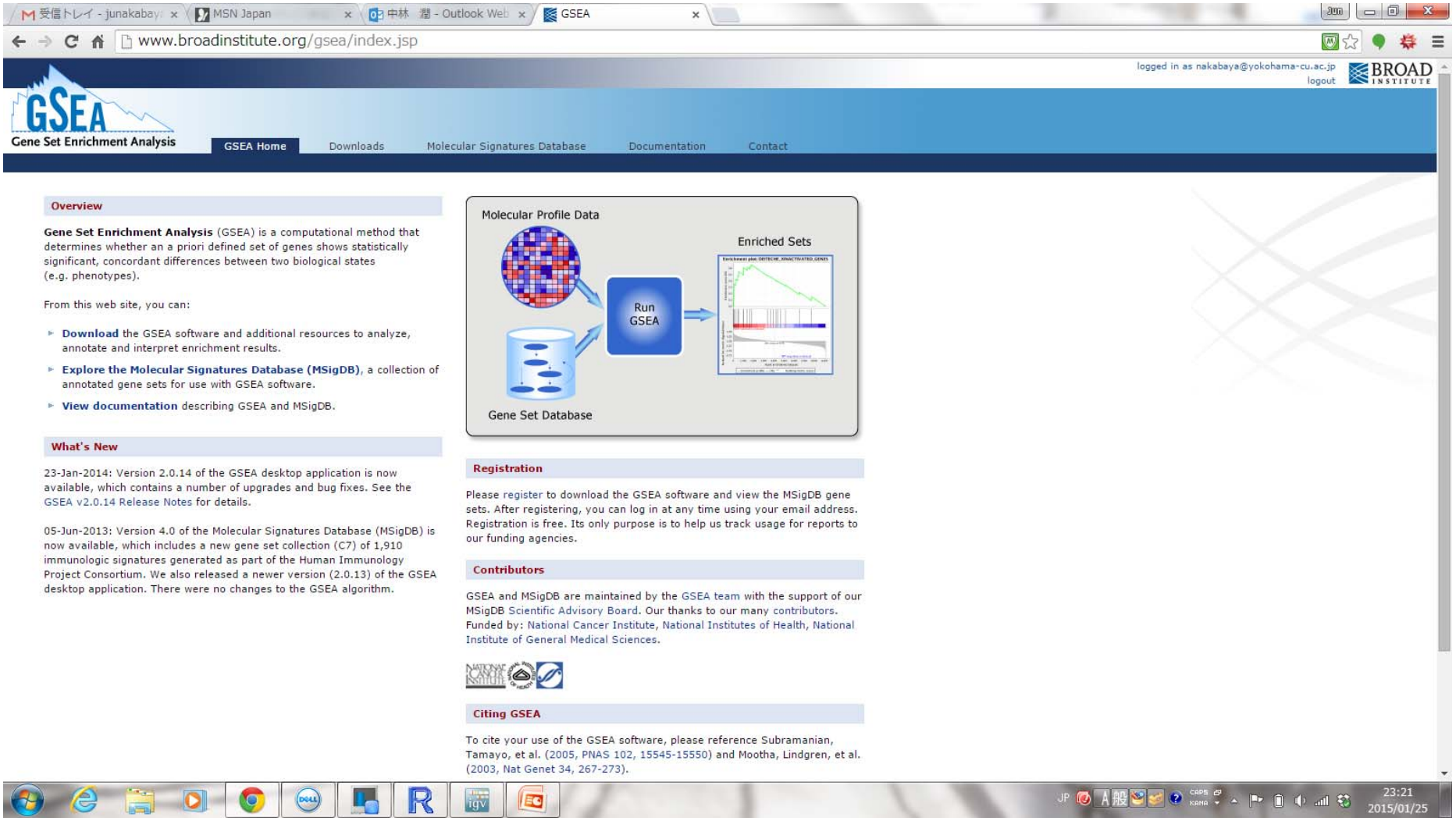
2016年度第5回
バイオインフォマティクス実習

発現変動遺伝子の機能解析
GSEA, pathway解析,GO解析

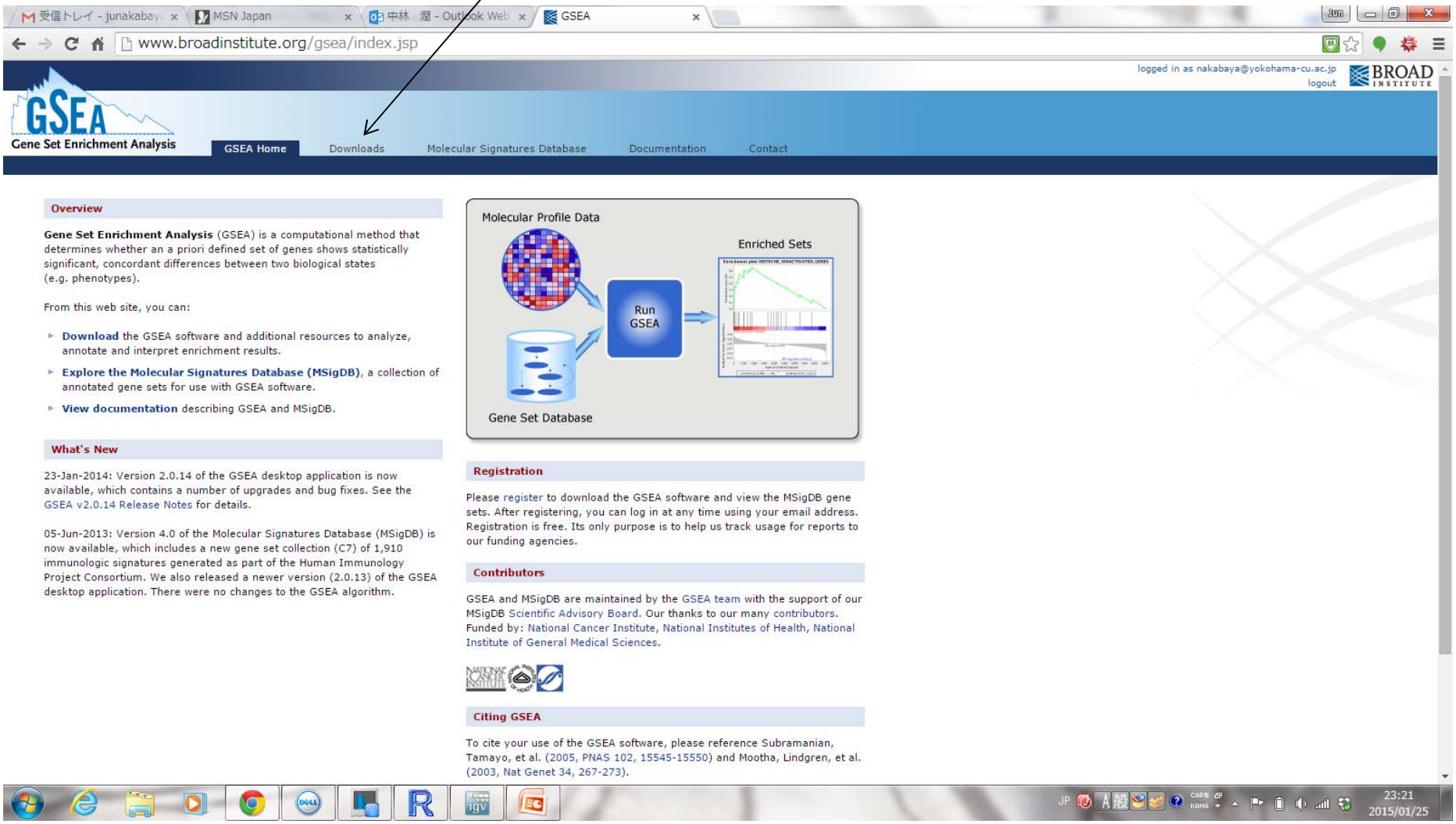
- Gene Set Enrichment Analysis (GSEA)
- 特定の遺伝子セットと発現比の間に相関があるか調べる


KO/WT
gene set
{Otx2,Msx1,Rbp1,…}
4
2
0
-2
-4
-6
expression ratio
発現比ランキングの
上位に偏って存在する
遺伝子セットと発現に相関あり
発現比ランキングの下位に偏って存在する
遺伝子セットと発現に逆相関あり
発現比ランキングによる偏り無し
遺伝子セットと発現に相関なし
gene rank

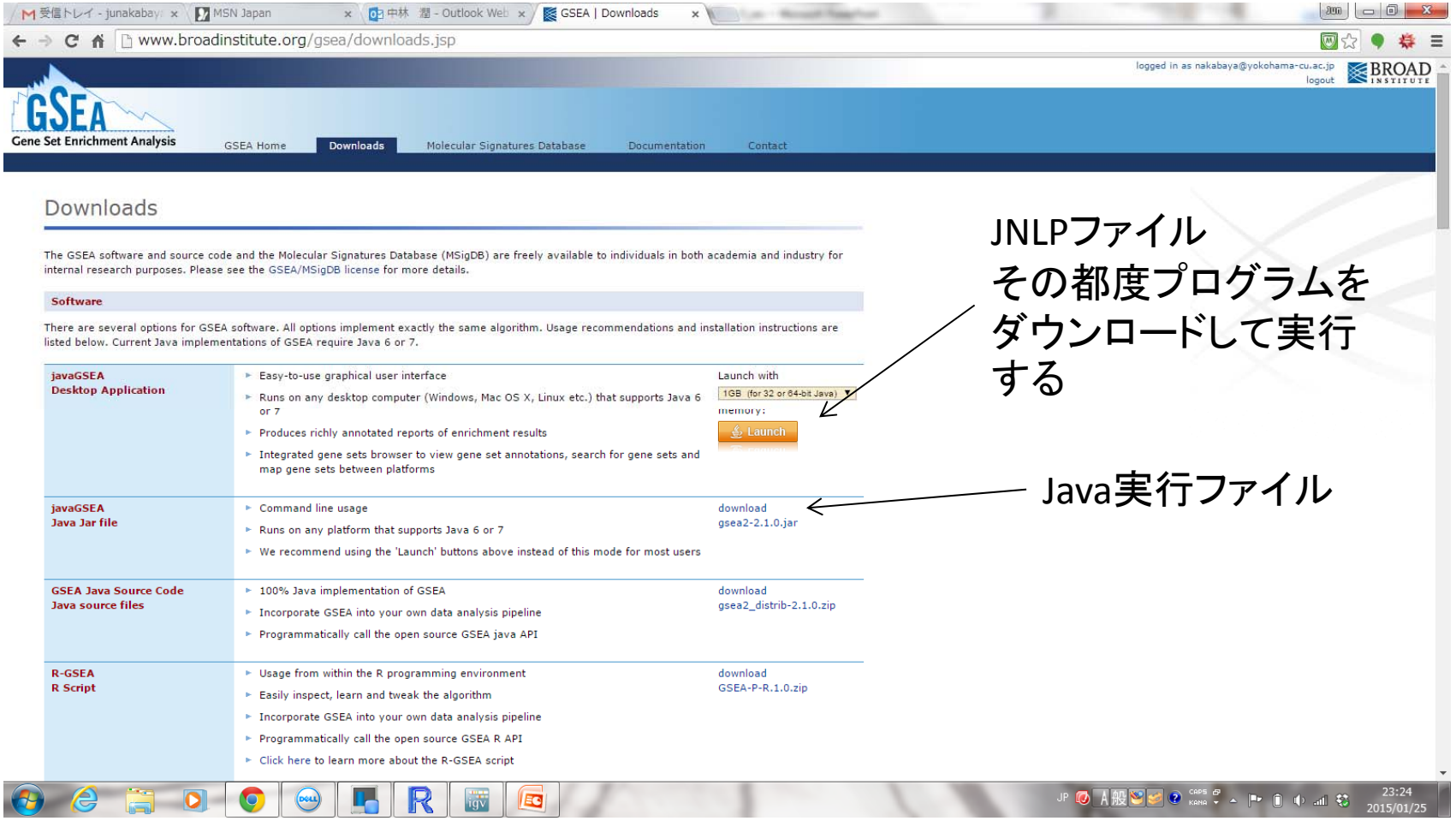
http://www.broadinstitute.org/gsea/index.jsp



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



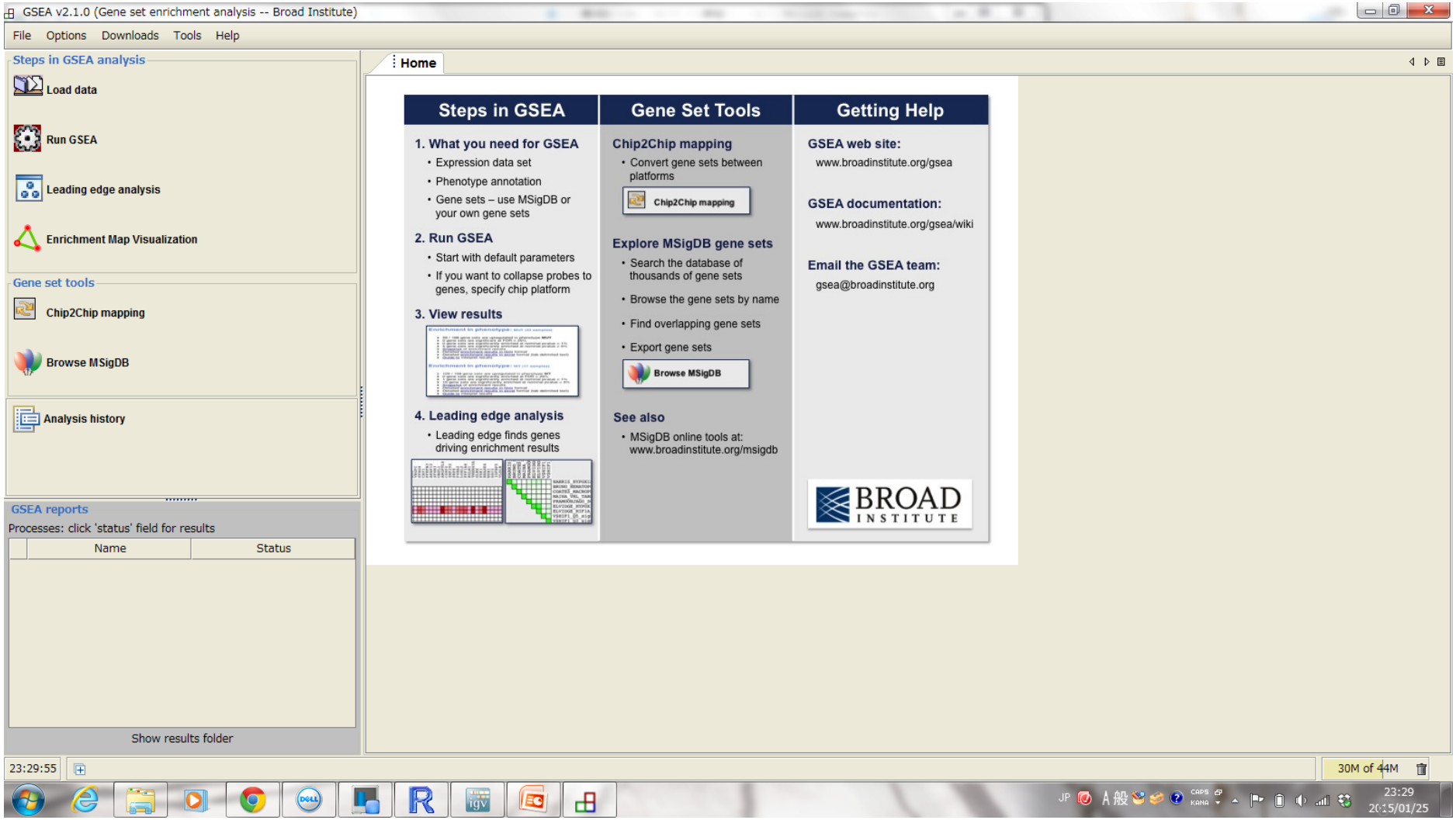
Download


GSEA | Downloads
www.broadinstitute.org/gsea/downloads.jsp
logged in as nakabaya@yokohama-cu.ac.jp
logout
BROAD INSTITUTE
GSEA
Gene Set Enrichment Analysis
GSEA Home
Downloads
Molecular Signatures Database
Documentation
Contact
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
Easy-to-use graphical user interface
Runs on any desktop computer (Windows, Mac OS X, Linux etc.) that supports Java 6 or 7
Produces richly annotated reports of enrichment results
Integrated 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
Command line usage
Runs on any platform that supports Java 6 or 7
We 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
100% Java implementation of GSEA
Incorporate GSEA into your own data analysis pipeline
Programmatically call the open source GSEA java API
download
gsea2_distrib-2.1.0.zip
R-GSEA
R Script
Usage from within the R programming environment
Easily inspect, learn and tweak the algorithm
Incorporate GSEA into your own data analysis pipeline
Programmatically call the open source GSEA R API
Click 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
23:24
2015/01/25

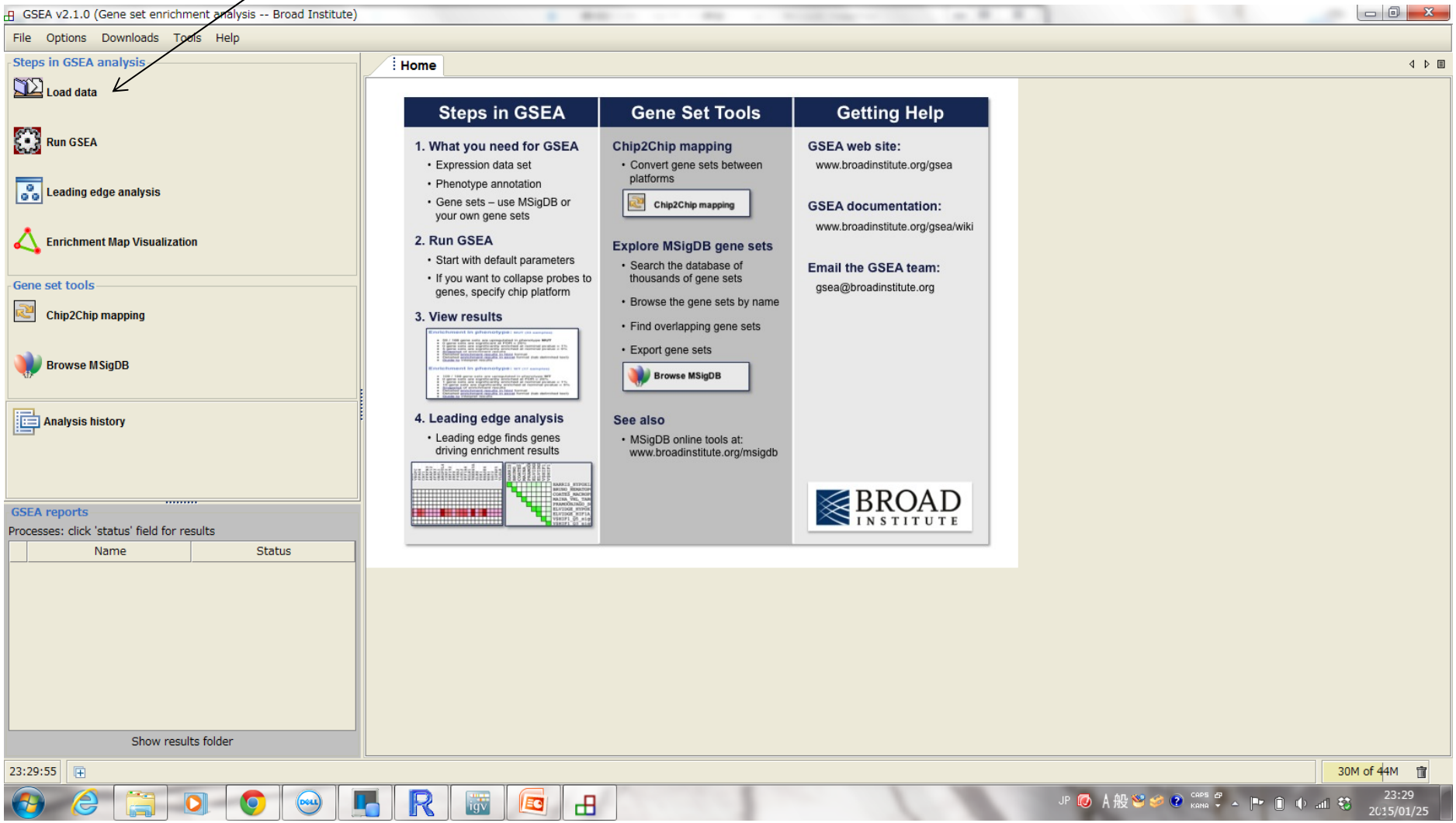
- 課題配布フォルダからgsea2-2.1.0を各自のデスクトップにコピー

- 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

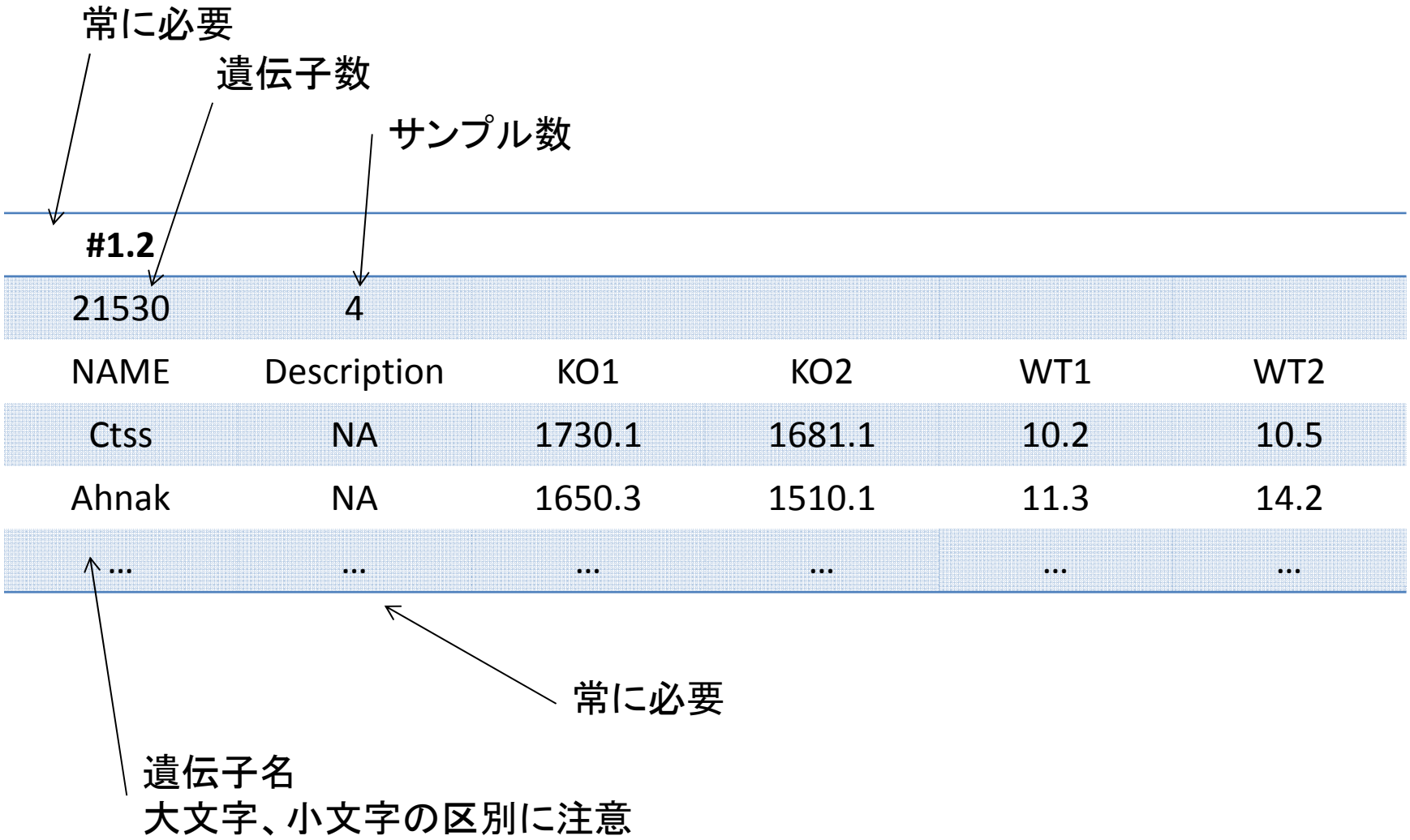
データファイルをload



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



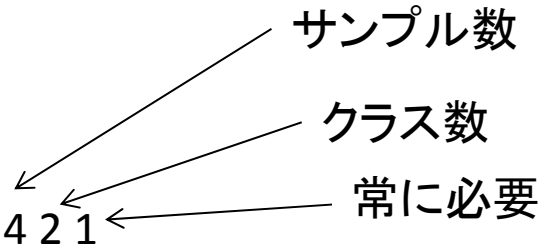
ファイル名の拡張子はgct

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

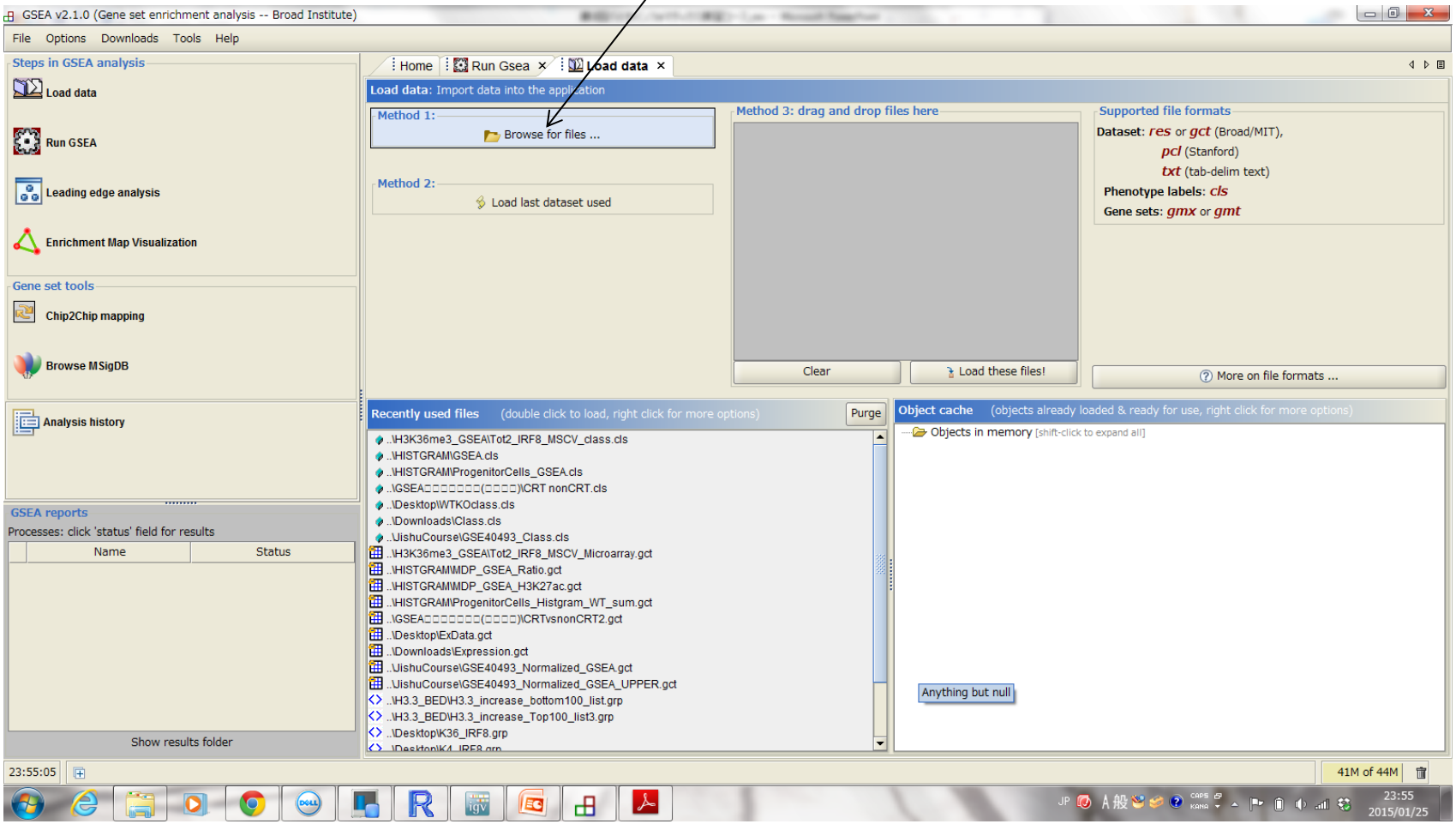


clsファイルはスペース区切りのテキストファイル
拡張子はcls

- 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

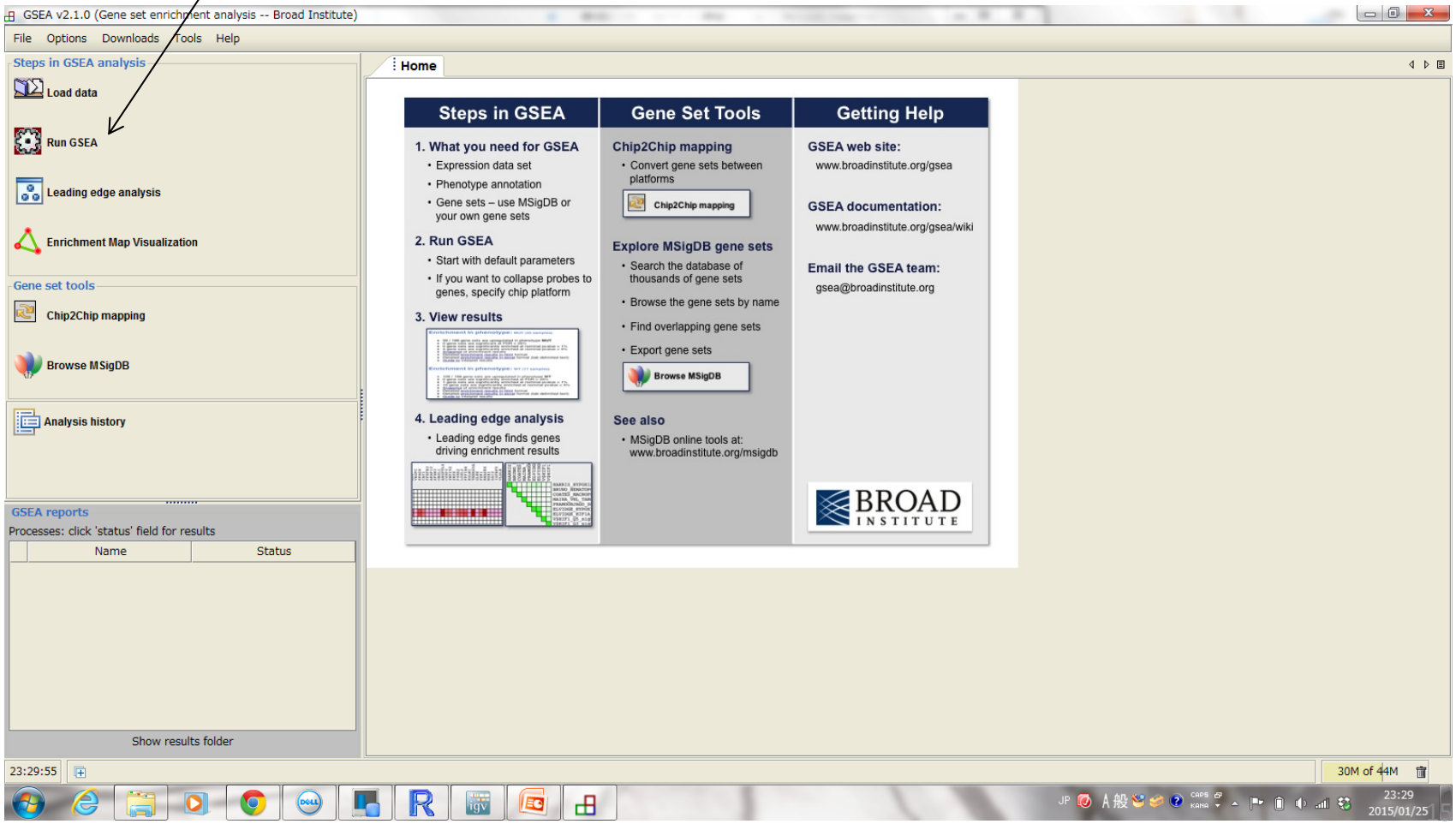
Load Data

Browse for filesをクリックしてファイルを選択

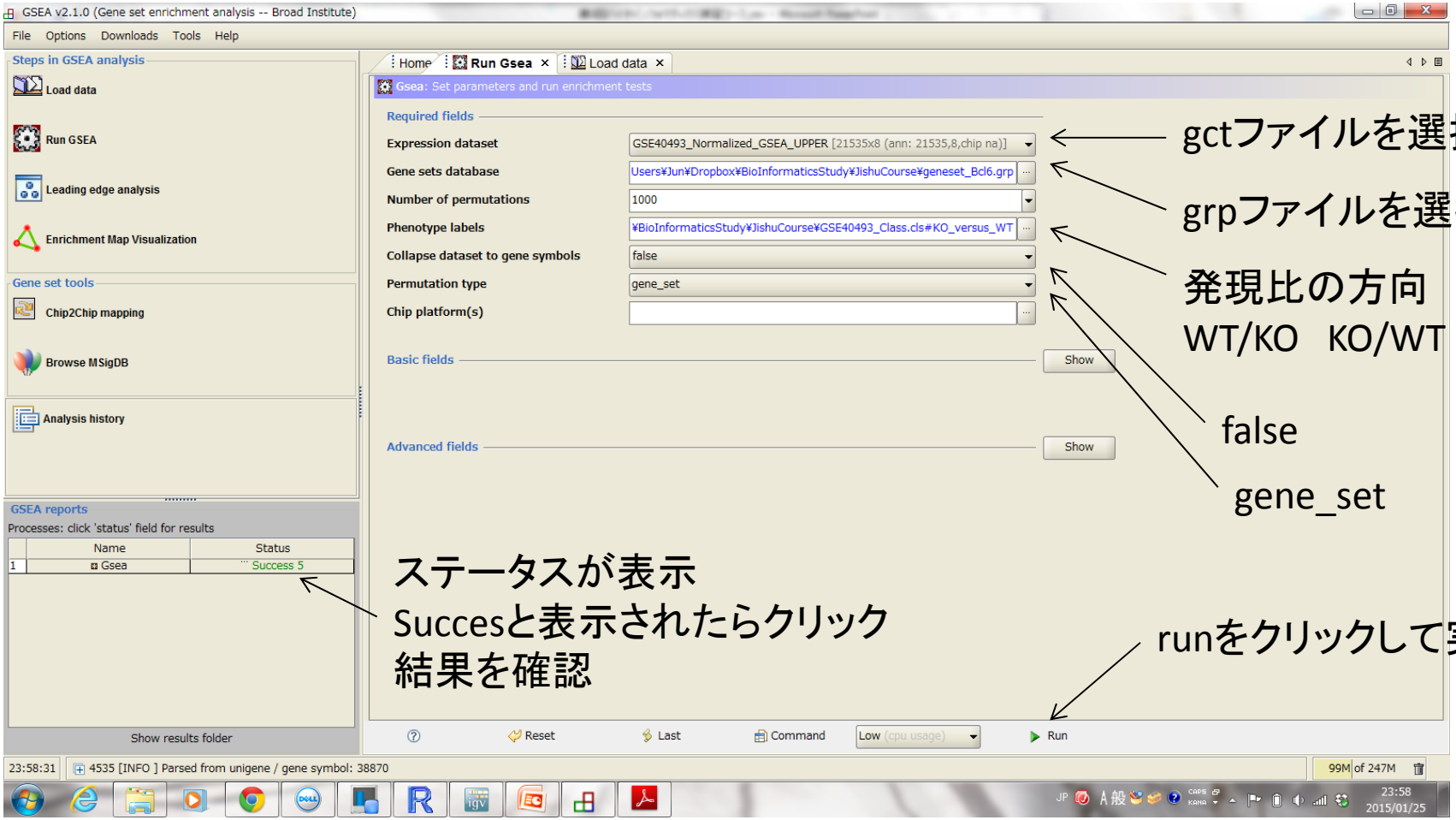


Run

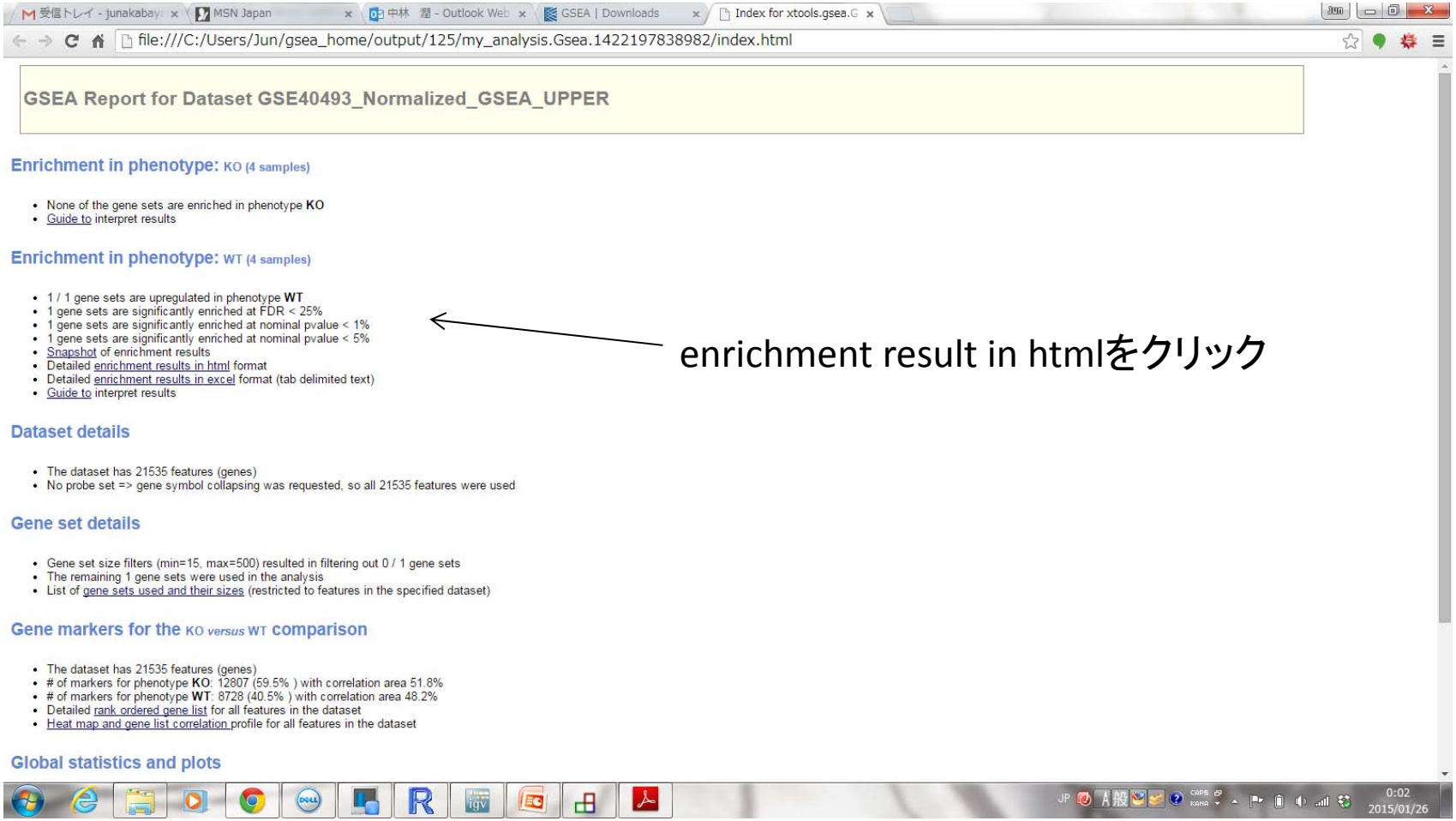
Run GSEAをクリックして実行



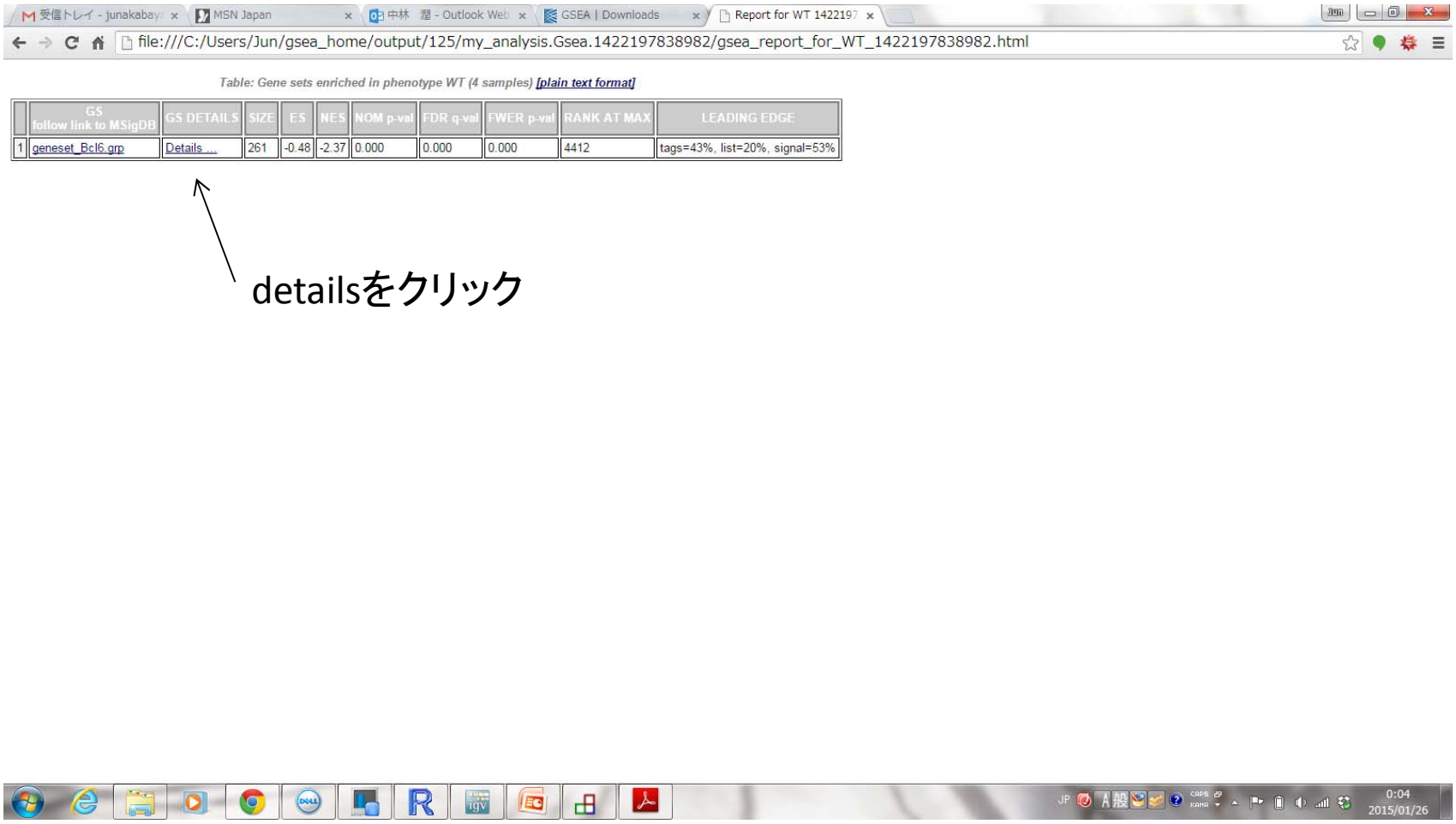
Run


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
Gsea
Success 5
Show results folder
Home
Run Gsea
Load data
Gsea: Set parameters and run enrichment tests
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
Reset
Last
Command
Low (cpu usage)
Run
23:58:31
4535 [INFO] Parsed from unigene / gene symbol: 38870
99M of 247M
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
ステータスが表示
Succesと表示されたらクリック
結果を確認
runをクリックして実行

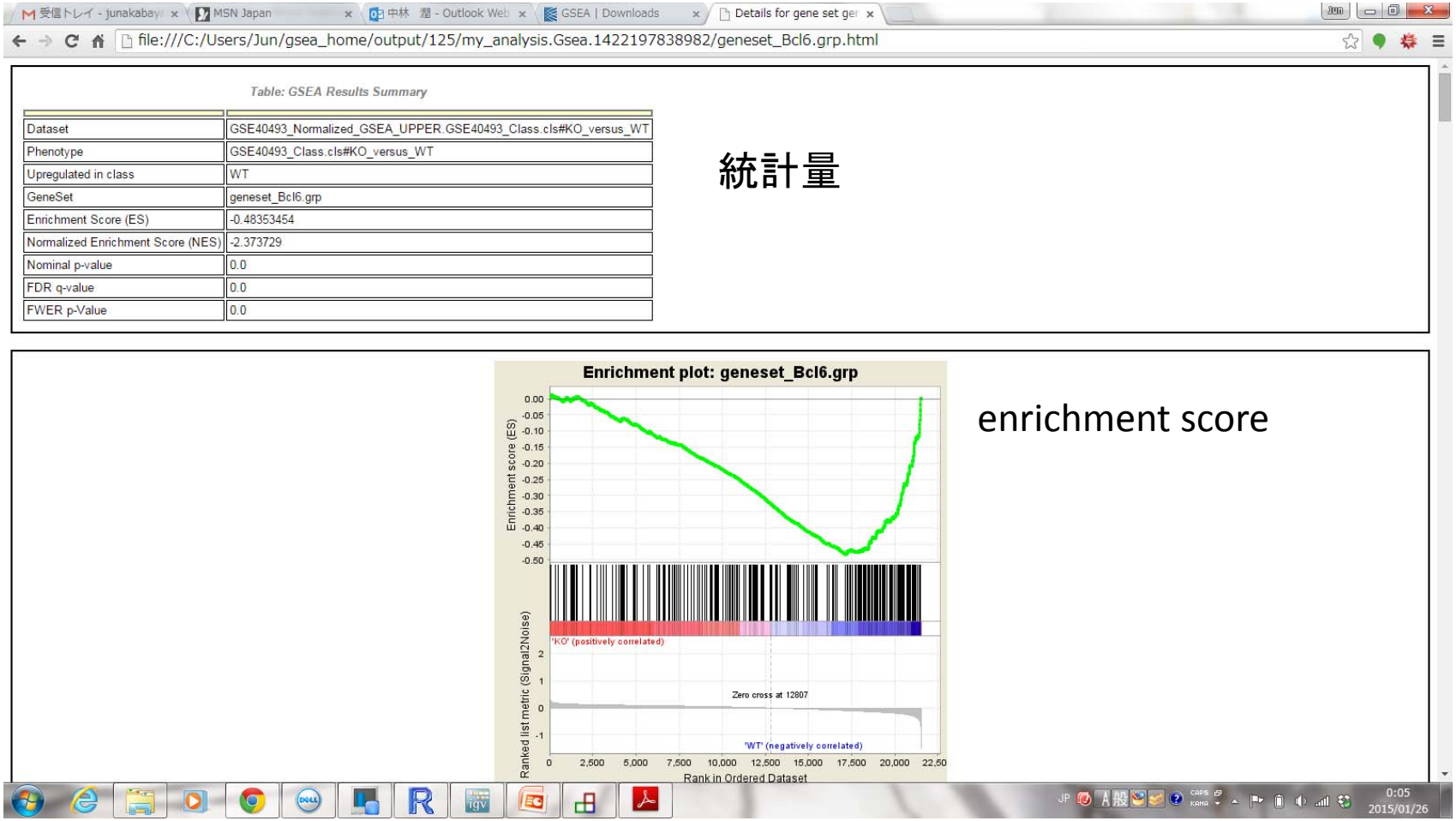
ブラウザ上で結果を表示



ブラウザ上で結果を表示



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

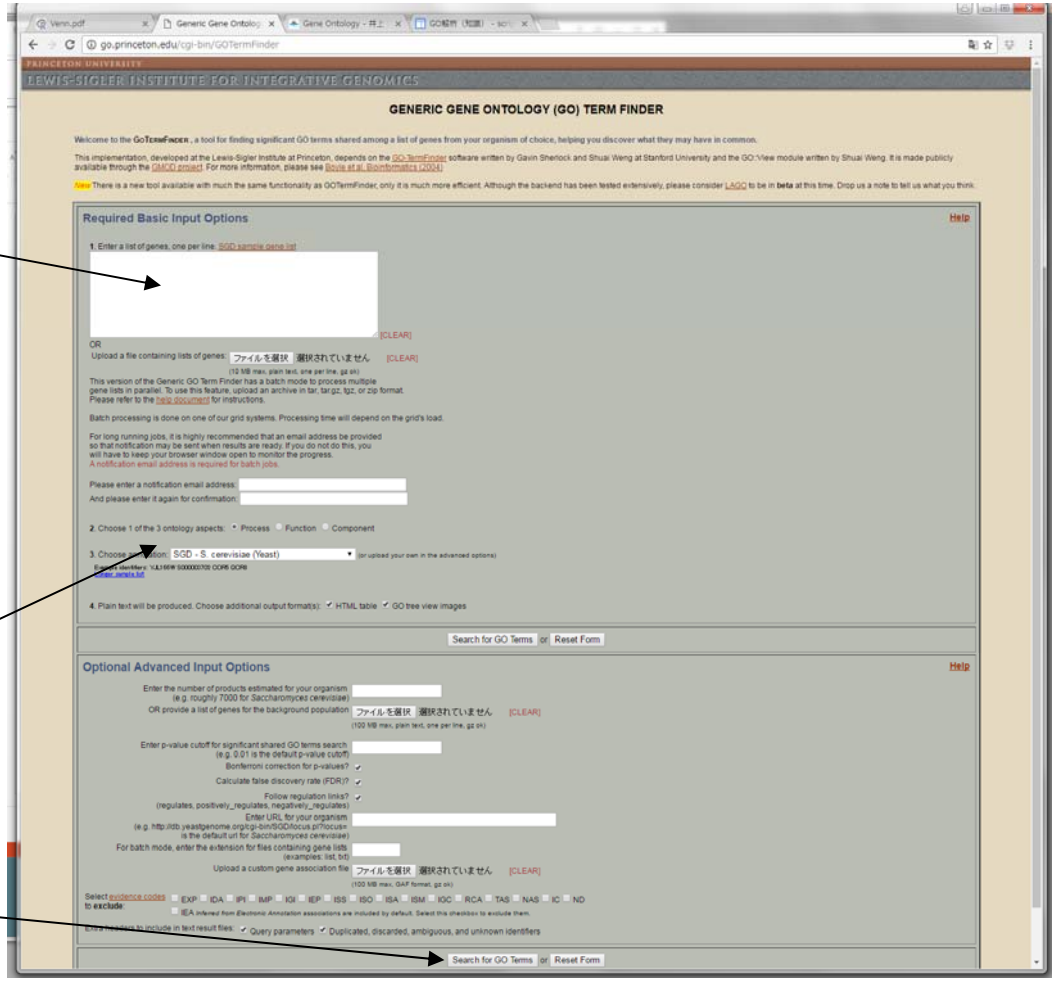
generic gene ontology term finder

- 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