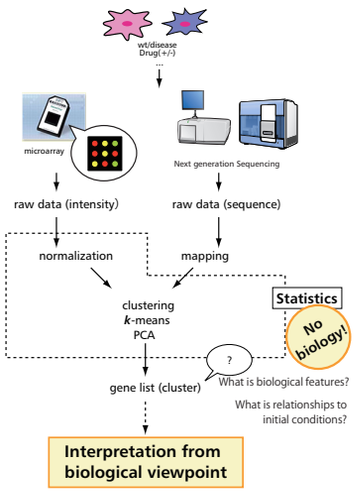
Database Center for Life Science (DBCLS)
Research Organization of Information and Systems (ROIS)

Functional profiling of gene and disease features using MeSH controlled vocabulary

Takeru Nakazato*, Hidemasa Bono
Database Center for Life Science (DBCLS), Research Organization of Information and Systems (ROIS), JAPAN
*: nakazato@dbcls.rois.ac.jp

Backgrounds and motivations



Microarray and next generation sequencer (NGS) is widely used for transcriptomics. Many methods to analyze these data are reported, but most of all are statistical one. So, "functional annotation" from biological viewpoint is required to interpret retrieved gene groups. Also, the relationships between gene groups and initial conditions of experiment such as wt/disease and drug(+/−) should be taken into account.
Researchers often use Gene Ontology (GO) but GO does not contain terms of diseases, drugs, and anatomy fields. We therefore generated feature profiles with Medical Subject Headings (MeSH).

Gene Ontology

Example) BRCA1 (Gene ID: 672)

Molecular Function
DNA binding
androgen receptor binding
enzyme binding
transcription coactivator activity
tubulin binding
ubiquitin protein ligase binding
...

Cellular Component
BRCA1-A complex
BRCA1-BARD1 complex
gamma-tubulin ring complex
nucleus
protein complex
ribonucleoprotein c
ubiquitin ligase complex

Biological Process
DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator
G2 DNA damage checkpoint
androgen receptor signaling pathway
apoptotic process
cellular response to indole-3-methanol
chromosome segregation
double-strand break repair via homologous recombination
positive regulation of DNA repair
positive regulation of protein ubiquitination
postreplication repair
regulation of cell proliferation
transcription from RNA polymerase II promoter
transcription from RNA polymerase III promoter
response to DNA damage stimulus
response to estrogen stimulus
response to ionizing radiation
...

No information on diseases and anatomy

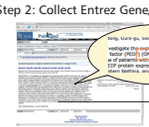
Feature profiles of genes and diseases using MeSH vocabulary

Methods

Step 1: Extract PMIDs cited in Entrez Gene/OMIM reference sections

Reference section → OMIM ID, PMID

Step 2: Collect Entrez Gene/OMIM IDs described in MEDLINE articles



OMIM ID, PMID → Gene/OMIM-PMID Pairs → Gene/OMIM-MeSH Pairs → Scoring (*p*-value)

Step 3: Collect PMIDs assigned by MeSH terms corresponding to genes and diseases

Corresponging MeSH terms → OMIM ID, PMID

PubMed Search

Results and Discussions

Example of feature profile using MeSH

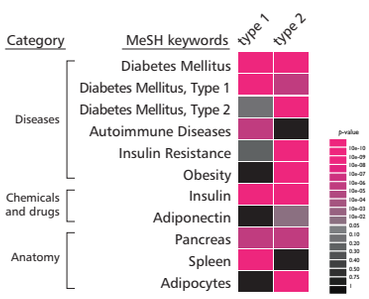
Example) BRCA1 (Gene ID: 672)

MeSH Terms	Category	*p*-value
Breast Neoplasms	Disease	0
Ovarian Neoplasms		0
BRCA1 Protein	Chemicals and Drugs	0
Tumor Suppressor Proteins		5.91×10^{-136}
Rad51 Recombinase		1.66×10^{-54}
Breast	Anatomy	1.52×10^{-43}
Chromosomes, Human, Pair 17		9.03×10^{-23}

BRCA1 is a tumor suppressor gene involved in breast cancer, and located at chr. 17. Phospholyrate BRCA1 repairs double strand damage of DNA with Rad51 protein. This profile illustrates these features.

Difference and similarity between disease features

Example)
%222100 Diabetes Mellitus, Insulin-dependent; IDDM
#125853 Diabetes Mellitus, Noninsulin-dependent; NIDDM



MeSH Terms	*p*-value
Type 1	
Autoimmune disease	4.55×10^{-5}
Spleen	5.53×10^{-7}
→ Autoimmune disease	
Type 2	
Obesity	1.18×10^{-15}
Adipocyte	5.17×10^{-5}
→ Metabolic disorder	

Data Visualization

FREE!

Feature profiles of genes and diseases

Gendoo
(gene, disease features ontology-based overview system)

http://gendoo.dbcls.jp/


Feature profiles


tree view

NGS data relevant to diseases

DBCLS SRA http://sra.dbcls.jp/


Frequency list

The NGS data is archived in public database, the sequence read archive (SRA). We downloaded these NGS data, and constructed the disease list by extracting MeSH disease terms from reference publications. Users can search public NGS data from diseases of their interests.

SRA ID	SRA Title	Disease	疾患名	PMID
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18987736
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18987736
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA009897	In-depth characterization of the microRNA transcriptome in a leukemia progression model	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18849523
SRA029797	Exome Sequencing Identifies Somatic Mutations in Acute Monocytic Leukemia	Leukemia, Myeloid, Acute	白血病-急性骨髄性	21399634

Public NGS data | Disease | Corresponding publication

NGS entry list

Reference:
Gendoo: Functional profiling of gene and disease features using MeSH vocabulary
Nakazato T., Bono H., Matsuda H., Takagi T.,
Nucleic Acids Research, 37 (Suppl. 2) (Web Server issue), 2009
doi:10.1093/nar/gkp483

Biocuration2013
Churchill College, Cambridge, UK
Apr. 7-10, 2013