Week 1 Notes/Slides/Information for the Purdue Big Data For Biologists Workshop

2016
Welcome!

DAY 1
Session 1:
WELCOME

Encouraging Big Data Thinking in Biomedical research
Introductions

Who are you? Where are you from?

What is your research interest?

Why are you interested in “big data”?
Instructors and Teaching Assistants

Main Instructors:
James C. Fleet, PhD (Nutrition Science)
Wanqing Liu, PhD (Medicinal Chemistry and Molecular Pharmacology)
Pete Pascuzzi, PhD (Libraries)
Min Zhang, PhD (Statistics)

Teaching Assistants:
Chen Chen (Statistics)
Min Ren (Statistics)
Guest Lecturers

Week 1:
Doug Crabill (Purdue University)
Sean Davis (National Cancer Institute)
Nadia Atallah (Purdue University)
Ingenuity Systems Staff

Week 2:
Jennifer Neville (Purdue University)
Shi Li Lin (Ohio State University)
Martin Lundquist (John Hopkins University)
Strengths/Weaknesses Of Hypothesis-Driven Science
Disruptive Technologies Propel Scientific Advancement

Sequenced genomes helped scientists see that data generation need not be a barrier to scientific progress.

What can we do with all this information?

- Genomic: GEO ENCODE
- Genetic: WebQTL/JAX 1000 genomes
- Metabolomic: HMDB
- Proteomic: ProteomicsDB
- Phenotypes: JAX TCGA

2001 Breakthrough of the Year

2001 Sequenced Genomes

Fleet 2016
New technologies bring challenges to using and integrating data

12 Brain-imaging modalities

Requires Multidimensional Thinking and Analytical Tools

What is Biological Big Data Science (BBDS)?

An approach that integrates multi-dimensional, high-density, variously formatted data types to gain insight in the function or regulation of biological systems.
Big Data in Biomedicine Has Two “Flavors”

“Omics” Driven
- Genotyping
- Gene expression
- NGS data

Payer-provider Driven
- Electronic Medical Records
- Pharmacy information
- Insurance Records

Basic Biological Understanding and Treatment Discovery

Optimize Healthcare Delivery and Economics

Fleet 2016
Where do traditional biomedical researchers fit?

- Life Sciences
- Levels of Scale
  - Gene
  - Cell
  - Organism
  - Community
  - Global

Focus

Biological Big Data Science (BBDS)

- Analysis
- Characterization

- Computational Disciplines
- Analytical Disciplines

Fleet 2016
Central Dogma of Biology Has Been Reshaped by Big Data

Modified from Doerge (2002) NRG 3:43
Is Traditional Science Outdated?

“faced with massive data, this approach to science - hypothesize, model, test — is becoming obsolete.”
Chris Anderson, Editor, Wired Magazine, 2008

Hypothesis-driven vs Data-driven

- Needs ‘falsifiable’ hypothesis
- Needs appropriate data
- Statistical significance
- Small effects in lots of data
- Determines Causation

- No hypothesis needed
  - Exploratory?
- Full data not needed
- Post-hoc explanation
- Statistical significance?
- “Thin” data?
The Wrong Way to Do Bioinformatics

Well Designed Experiment 1

Experimental Scientist

Interpretation

Results

And then a miracle happens....
The Challenge of BBDS

Core Facility

Sample Analysis

Well Designed Experiment 1

Experimental Scientist

Interpretation

Results

Visualisation

Raw Data

Processed Data 1

Functional Analysis

Processed Data “n”

Computing Algorithms Statistics

Computing Algorithms Statistics

Algorithms Computing Statistics

Fleet 2016
Big Data Challenges in Biomedicine

From the NIH “Big Data to Knowledge” (BD2K) program

1) Locating data
2) Getting access to data
3) Organizing, Managing, and Processing data
4) Computational issues
   • Infrastructure
   • Speed (computer and algorithms)
5) Analytical methods
6) Training researchers to use BBD effectively
Stages of Big Data Education

**NAIVE**
- Appreciate Value
- Vocabulary to Talk with Experts to Define Goals

**FAMILIARITY**
- Use Pre-existing Software

**LOW LEVEL ABILITY**
- Code, Develop Tools

**HIGH LEVEL ABILITY**
- Creative Work, Novel Approaches

Skill + Independence
What We’ll cover during the workshop

Week 1
Unit 1: Microarray
Unit 2: Next Generation Sequencing

Week 2
Unit 3: Biomarker Discovery
Unit 4: Genetic Variation

Biological Goals:

• Regulation of gene expression
• Phenotype prediction
• Genotype-phenotype relationships

Fleet 2016
What We will cover during the workshop

**Week 1**
- **Unit 1**: Microarray
- **Unit 2**: Next Generation Sequencing

**Week 2**
- **Unit 3**: Biomarker Discovery
- **Unit 4**: Genetic Variation

**Technical Goals:**
- Analysis pipelines
- Statistical issues
- Visualization
- Functional annotation
- Databases
- Project management
- Computation and programming

*Fleet 2016*
Day 1
Session 2:
Working with the Purdue Computer Infrastructure

Doug Crabill
Department of Statistics
Purdue University
Sites to Understand Computing

- UNIX operating system
  - Learn UNIX
- Linux operating system
  - [http://www.tutorialspoint.com//operating_system/os_linux.htm](http://www.tutorialspoint.com//operating_system/os_linux.htm)
- R coding
  - [https://www.r-project.org/about.html](https://www.r-project.org/about.html)

Fleet 2016
Monday

BREAK #1
Day 1
Session 3: Data Repositories and Pre-processed Data Sites

James C. Fleet, PhD
Distinguished Professor
Department of Nutrition Science

Pete Pascuzzi, PhD
Assistant Professor
Purdue Libraries
<table>
<thead>
<tr>
<th>Data Archives</th>
<th>Web link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH Data Sharing Repositories</td>
<td><a href="https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html">https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html</a></td>
<td>Trans-NIH BioMedical Informatics Coordinating Committee (BMIC) sites</td>
</tr>
<tr>
<td><strong>Gene Expression Omnibus (GEO)</strong></td>
<td><a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a></td>
<td>NCBI; transcriptome and ChIP-seq datasets</td>
</tr>
<tr>
<td>Array Express</td>
<td><a href="http://www.ebi.ac.uk/arrayexpress/">http://www.ebi.ac.uk/arrayexpress/</a></td>
<td>EMBL-EBI repository to archive functional genomics data</td>
</tr>
<tr>
<td>European Nucleotide Archive (ENA)</td>
<td><a href="http://www.ebi.ac.uk/ena">http://www.ebi.ac.uk/ena</a></td>
<td>Comprehensive record of worlds nucleotide sequencing information</td>
</tr>
<tr>
<td>The Cancer Genome Atlas (TCGA)</td>
<td><a href="http://cancergenome.nih.gov/">http://cancergenome.nih.gov/</a></td>
<td>Multi &quot;omic&quot; phenotype characterization of tumors</td>
</tr>
<tr>
<td>Proteomics IDEntifications (PRIDE)</td>
<td><a href="http://www.ebi.ac.uk/pride/archive/">http://www.ebi.ac.uk/pride/archive/</a></td>
<td>European proteomics datasets</td>
</tr>
<tr>
<td>Metabolomics Workbench</td>
<td><a href="http://metabolomicsworkbench.org/standards/nominatecompounds.php">http://metabolomicsworkbench.org/standards/nominatecompounds.php</a></td>
<td>Metabolomic datasets</td>
</tr>
</tbody>
</table>
Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.
GEO Datasets Week 1

- **GSE15947**: Time course of 1,25(OH)2 D treated RWPE1 cells
- **GSE54783**: The Osteoblast to Osteocyte Transition: Epigenetic changes and response to the vitamin D3 hormone

<table>
<thead>
<tr>
<th>GSE #:</th>
<th>Accession number for a <strong>complete dataset</strong> that is submitted to GEO</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM #:</td>
<td>Accession number for a <strong>specific sample</strong> within a dataset</td>
</tr>
<tr>
<td>GPL #:</td>
<td>The <strong>platform</strong> used to generate a dataset</td>
</tr>
<tr>
<td>SRX #:</td>
<td>Accession number for a <strong>sample generated by NGS</strong> that is deposited in the Short Read Archive (SRA)</td>
</tr>
<tr>
<td>Data Archives</td>
<td>Web link</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Oncomine</td>
<td><a href="https://www.oncomine.org/resource/login.html">https://www.oncomine.org/resource/login.html</a></td>
</tr>
<tr>
<td>Gene Expression across Normal and Tumor tissue</td>
<td><a href="http://medical-genome.kribb.re.kr/GENT/">http://medical-genome.kribb.re.kr/GENT/</a></td>
</tr>
<tr>
<td>(GENT)</td>
<td></td>
</tr>
<tr>
<td>BioGPS</td>
<td><a href="http://biogps.org/#goto=welcome">http://biogps.org/#goto=welcome</a></td>
</tr>
<tr>
<td>cBioPrortal</td>
<td><a href="http://www.cbioportal.org/">http://www.cbioportal.org/</a></td>
</tr>
<tr>
<td>Genotype-Tissue Expression project (Gtex)</td>
<td><a href="http://www.gtexportal.org/home/">http://www.gtexportal.org/home/</a></td>
</tr>
<tr>
<td>Immunological Genome Project (Immgen)</td>
<td><a href="http://www.immgen.org/">http://www.immgen.org/</a></td>
</tr>
<tr>
<td>Kidney Systems Biology Project</td>
<td><a href="https://hpcwebapps.cit.nih.gov/ESBL/Database/">https://hpcwebapps.cit.nih.gov/ESBL/Database/</a></td>
</tr>
<tr>
<td>Saccharomyces Genome Database</td>
<td><a href="http://www.yeastgenome.org/trancriptome-data-in-yeastmine">http://www.yeastgenome.org/trancriptome-data-in-yeastmine</a></td>
</tr>
<tr>
<td>miRBase</td>
<td><a href="http://mirbase.org/">http://mirbase.org/</a></td>
</tr>
</tbody>
</table>
Day 1
Session 4: Understanding R and Bioconductor

James C. Fleet, PhD
Distinguished Professor
Department of Nutrition Science

Pete Pascuzzi, PhD
Assistant Professor
Purdue Libraries
The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.3.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The R Logo is available for download in high-resolution PNG or SVG formats.
- **userR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.
Using R: a guide for complete beginners

This tutorial is intended to introduce users quickly to the basics of R, focusing on a few common tasks that biologists need to perform: some basic analysis, load a table, plot some graphs, and perform some basic statistics. More extensive tutorials can be found on the project website and via Bioconductor (not covered here).

R-language: http://www.r-project.org

BioConductor: http://www.bioconductor.org

Advantages of R

- Free!
- Powerful, many libraries have been created to perform application specific tasks. e.g. analysis of microarray experiments and Next-Gen sequencing (Bioconductor: including Bioseq group).
- Presentation quality graphics
  - Save as a png, pdf or svg
- History
  - What you do can be saved for the next time you use R.
  - Ability to turn it into an automated script to perform again and again on different data

Disadvantages

- Lack of a comprehensive graphical user interface, but two do exist: However some do exist: R commander: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/ and Limma-gui (microarrays) : http://bioinf.wehi.edu.au

Citation
About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, 1211 software packages, and an active user community. Bioconductor is also available as an AMI (Amazon Machine Image) and a series of Docker images.

Install »
Get started with Bioconductor
- Install Bioconductor
- Explore packages
- Get support
- Latest newsletter
- Follow us on twitter
- Install R

Learn »
Master Bioconductor tools
- Courses
- Support site
- Package vignettes
- Literature citations
- Common work flows
- FAQ
- Community resources
- Videos

Use »
Create bioinformatic solutions with Bioconductor
- Software, Annotation, and Experiment

Develop »
Contribute to Bioconductor
- Developer resources
- Use Bio 'dev'
Rstudio is a Graphical User Interface (GUI) that let’s you use R more conveniently.
Day 1
Session 5-6:
Microarray Overview
Data Processing and QC

James C. Fleet, PhD
Distinguished Professor
Department of Nutrition Science

Pete Pascuzzi, PhD
Assistant Professor
Purdue Libraries
1,25(OH)₂ D Induces Gene Transcription through the Vitamin D Receptor

Classical Role: Regulate whole body calcium metabolism

NGS project: Osteoblast and osteocytes

Novel Role: Prevent and treat cancer

Microarray project: Normal prostate epithelial cell

Fleet et al. (2012) Biochem J 441:61
VDR Deletion Accelerates Cancer Development in Mice

**APC\textsuperscript{min} Mice**

- **Colon Cancer**
  - Tumor size
  - WT, HT, KO (5 months)

**LPB-Tag Mice**

- **Prostate Cancer**
  - Tumor size

**MMTV-neu Mice**

- **Breast Cancer**
  - Tumor Incidence

---

**Colon Cancer**

<table>
<thead>
<tr>
<th>Age (mo)</th>
<th>Tumor size (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
</tr>
</tbody>
</table>

**Prostate Cancer**

<table>
<thead>
<tr>
<th>Age (wks)</th>
<th>Tumor weight (mgs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

**Breast Cancer**

<table>
<thead>
<tr>
<th>Age (mo)</th>
<th>Tumor-Free Mice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>18</td>
<td>30</td>
</tr>
</tbody>
</table>

---

Larriba et al., 2011, PLoS One 6:e23524
Mordan-Mcombs et al., 2010, JSBMB 121:368
Zinser and Welsh 2004, Carcinogenesis 25:2361

Fleet 2016
Strategies of Cancer Prevention

Cancer Prevention

Cancer Treatment

Time (y)

Survival without Cancer

Survival with Cancer

Diagnosis

Research Paradigm:
Study Gene expression with DNA microarray

Kovalenko et al. (2010) 1,25 dihydroxyvitamin D-mediated orchestration of anticancer, transcript-level effects in the immortalized, non-transformed prostate epithelial cell line, RWPE1. BMC Genomics 11:26

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2820456/
**Study Design**

- **RWPE1 cells**
  - Human prostate epithelial cell line (ATCC CRL-11609)
  - 54 y old white male
    - Peripheral zone of normal health prostate
  - HPV18 immortalized
  - Not tumorigenic in nude mice

**2 x 3 factorial design, n = 4/group (24 total samples)**

<table>
<thead>
<tr>
<th>Media change</th>
<th>100 nM 1,25(OH)2 D or EtOH control</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 h</td>
<td></td>
</tr>
<tr>
<td>24 h</td>
<td></td>
</tr>
<tr>
<td>48 h</td>
<td></td>
</tr>
</tbody>
</table>

GEID = GSE15947

*Kovalenko et al. (2010) BMC Genomics 11:26*
Affymetrix Microarray Analysis Workflow

Well Designed Study
- High Quality RNA
- Affymetrix analysis
- Affymetrix QC analysis
- Affymetrix Raw Data

Normalized Vetted Data
- Process Data
  - RMA
  - Normalize
  - MAS5 (P/A)
- Additional QC analysis
- Data Filters
  - P/A call
  - Fold
- Statistical Analysis
- Differentially Expressed Gene List
- Clustering and visualization
- Pathway and Geneset Analysis
- Network building
- Interpret Experiment

Interpret Experiment
- Fleet 2016
Affymetrix Tiling Arrays

11-25 probes or probe pairs per gene

Perfect Match = 25 bp
Mismatch in one base

Used to define “background”

Algorithms from Affymetrix and others define “expression”
Affymetrix GeneChip Microarrays

**Probe Array**

**Pre-hybridization**

- Probe Cell
- 540,000 locations on each GeneChip array
- Actual strand = 23 base pairs
- U133 Plus 2 GeneChip: 54,000 probe sets

**Probe Array**

**Post-hybridization**

- 11 probe pairs/gene
- U133 Plus 2 GeneChip: 54,000 probe sets

**cDNA library and Biotin Label**

- Total RNA
- Reverse Transcription
- In Vitro Transcription
- Biotin-labeled cRNA

**Hybridization to Array**

- 25 base long probes
- 40 x 10^7 copies/cell

---

Fleet 2016
Residual Plots of Scanned GeneChips

- High quality (mostly white)
- Too Variable (intense red/blue)
- Scratched
- Uneven (edge drying)
- Uneven

Residuals
Red = (+)
Blue = (-)
White = 0

http://plmimagegallery.bmbolstad.com/

Fleet 2016
Relative Log Expression (RLE) provides a measure of reproducibility of gene expression data that can be compared across batches, experiments or trials.

Large spread and non-zero = bad

Outside control limits = bad

Control limits

IQR limits

Fleet 2016
Normalized Unscaled Standard Error (NUSE)

Values have no units. Used to assess relative quality of arrays within an analysis set.

- Large spread and non-zero = bad
- Outside control limits = bad

Fleet 2016
Day 1
Session 7:
Differential Gene Expression

James C. Fleet, PhD
Distinguished Professor
Department of Nutrition Science

Pete Pascuzzi, PhD
Assistant Professor
Purdue Libraries
Statistics for Biologists

Nature Publishing:

Collections of short articles on basic statistical concepts for biologists

http://www.nature.com/collections/qghhqm
What is a p-Value?

The probability of obtaining an effect at least as extreme as the one you observed.

Control Mean = 260
Treatment Mean = 330.6
### What is type I error?

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Null Hypothesis is True</th>
<th>Null Hypothesis is False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject Null Hypothesis</td>
<td><strong>Type I error</strong> (False positive, ( \alpha ))</td>
<td>Correct Inference (True positive)</td>
</tr>
<tr>
<td>Fail to Reject Null Hypothesis</td>
<td>Correct Inference (True negative)</td>
<td><strong>Type II error</strong> (False negative, ( \beta ))</td>
</tr>
</tbody>
</table>

Rejecting the null hypothesis when it is in fact true...a false positive.
Why is type I error a problem for “omics” studies?

\[ P(\text{at least one Type I error among } m \text{ tests}) = 1 - (1 - \alpha)^m \]

e.g. When \( \alpha = 0.05 \) and \( m = 10 \), \( P = 0.401 \)

“large p, small n problem”

100 repeats of same experiment: 5 potential false positives when \( \alpha = 0.05 \)

25,000 transcripts on an array = 1,250 false positives at \( \alpha = 0.05 \)
How do we control type I error in “omics” studies?

Familywise Error Rate Correction (FWER): the probability of making even one false discovery in a set of comparisons.

* e.g. Bonferroni test

\[
\alpha/\text{(\# comparisons)} = 0.05/25,000 = 0.000002
\]

Very conservative but useful if the goal is to only find the changes that are most reliable (and likely the largest)
How do we control type I error in “omics” studies?

False Detection Rate (FDR):

e.g. Benjamini and Hochberg procedure

\[
i/m(Q) = \frac{\text{rank}}{\text{# comparisons}} \times \text{false discovery rate}
\]

<table>
<thead>
<tr>
<th>Gene</th>
<th>P value</th>
<th>Rank</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td>0.0001</td>
<td>1</td>
<td>0.000002</td>
</tr>
<tr>
<td>Gene 2</td>
<td>0.0004</td>
<td>2</td>
<td>0.000004</td>
</tr>
<tr>
<td>Gene 1500</td>
<td>0.049</td>
<td>1500</td>
<td>0.003</td>
</tr>
<tr>
<td>Gene 25,000</td>
<td>0.988</td>
<td>25000</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Accepts a set rate of false positives within a number of comparisons (e.g. 5% FDR means 5/100 “significant” comparisons are likely false positives)
Data Reduction as a Strategy to Minimize Type I Error Problem

Filter out genes with:
  • Low expression
  • High variation

<table>
<thead>
<tr>
<th>MAS5 (use M=P)</th>
<th># “Present”</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 4/24 = P</td>
<td>28,883</td>
</tr>
<tr>
<td>75% total = P</td>
<td>21,448</td>
</tr>
<tr>
<td>75% P in at least 1 group</td>
<td>25,985</td>
</tr>
<tr>
<td>50% total = P</td>
<td>23,793</td>
</tr>
<tr>
<td>50% P in at least 1 group</td>
<td>29,260</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drop Bottom 25%</th>
<th># “Present”</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% total = P</td>
<td>40,597</td>
</tr>
<tr>
<td>75% P in at least 1 group</td>
<td>45,689</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drop genes SD/mean &gt;0.25</th>
<th># “Present”</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% total = P</td>
<td>18,661</td>
</tr>
</tbody>
</table>

54,677 genes on the U133 Plus 2.0 array

Hackstadt and Hess (2009) BMC Informatics 10:11

Fleet 2016
## Processing and Statistics Influence the DEG List

<table>
<thead>
<tr>
<th>Processing</th>
<th>Statistic</th>
<th>DEG at 48 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>gcRMA</td>
<td>SAM</td>
<td>3566</td>
</tr>
<tr>
<td>RMA</td>
<td>SAM</td>
<td>2249</td>
</tr>
<tr>
<td>RMA</td>
<td>limma</td>
<td>1021</td>
</tr>
</tbody>
</table>