

# OMIC TOOLS 1/12

Yusuke Kijima

# **SAFE** Significant Analysis of Function and Expression

- What can we do with SAFE?
  - Resampling-based method for testing functional categories in gene expression experiments.
- Good
  - Statistically detailed and multi-core processing

# SAFE Significant Analysis of Function and Expression

- What can we do with SAFE?
  - Resampling-based method for testing functional categories in gene expression experiments.
- Good
  - Whole-gene-expression-based functional analysis
  - Statistically detailed and multi-core processing

```
> results1 <- safe(e3.upp, p3.upp$er, C.mat2, print.it = FALSE)
> proc.time()-t
 ユーザ システム 経過
 130.232    52.006   183.860 ← Default
> t <- proc.time()
> results2 <- safe(e3.upp, p3.upp$er, C.mat2, method = "bootstrap.t", print.it = FALSE)
> proc.time()-t
 ユーザ システム 経過
  25.080    9.962   35.104 ← Bootstrap
>
> t <- proc.time()
> results2p <- safe(e3.upp, p3.upp$er, C.mat2, method = "bootstrap.t", print.it = FALSE, parallel = T)
> proc.time()-t
 ユーザ システム 経過
  32.148   16.048   23.873 ← 2 core and bootstrap
```

# SAFE Significant Analysis of Function and Expression

## How to use?

Samples

	samplename	p53
1	UPP_103B41	0
2	UPP_104B91	1
3	UPP_112B53	0
4	UPP_114B66	0
5	UPP_130B92	0
6	UPP_131B79	0
7	UPP_135B40	0
8	UPP_138B34	0
9	UPP_139B03	0
10	UPP_147B19	1
11	UPP_148B49	0
12	UPP_148B48	0
13	UPP_150B81	0
14	UPP_154B42	0

Sample - phenotype table

	UPP_103B41	UPP_104B91	UPP_131B79	UPP_139B03	UPP_147B19	UPP_162B98	UPP_194B60	UPP_238C87	UPP
1007_s_at	8.786	9.151	8.989	9.497	9.180	9.357	9.014	9.193	
1053_at	6.335	6.431	6.049	6.179	6.177	5.863	5.969	5.966	
117_at	6.246	6.371	6.088	6.002	7.290	6.702	6.365	6.477	
121_at	7.951	8.134	8.029	8.035	7.842	8.201	8.070	7.931	
1255_g_at	4.599	3.049	4.734	4.679	4.984	5.250	5.152	4.267	
1294_at	7.382	7.554	7.336	7.235	7.615	7.547	7.557	7.100	
1316_at	6.179	5.819	5.408	5.785	6.232	5.946	5.995	6.398	
1320_at	4.223	5.444	5.566	4.229	5.188	5.489	5.467	5.164	
1405_i_at	9.182	7.985	7.554	5.818	8.357	5.357	7.823	7.820	
1431_at	5.001	5.463	4.921	5.809	5.285	5.777	4.941	5.359	
1438_at	6.419	6.642	6.438	6.926	4.643	6.003	6.811	6.935	
1487_at	7.226	7.355	7.104	7.142	7.615	7.395	7.137	7.209	
1494_f_at	6.791	8.697	6.872	7.016	6.718	7.066	6.714	6.903	
1598_g_at	8.936	9.155	9.324	9.776	8.611	9.552	8.829	8.883	

Input matrix : gene expression × sample

Genes

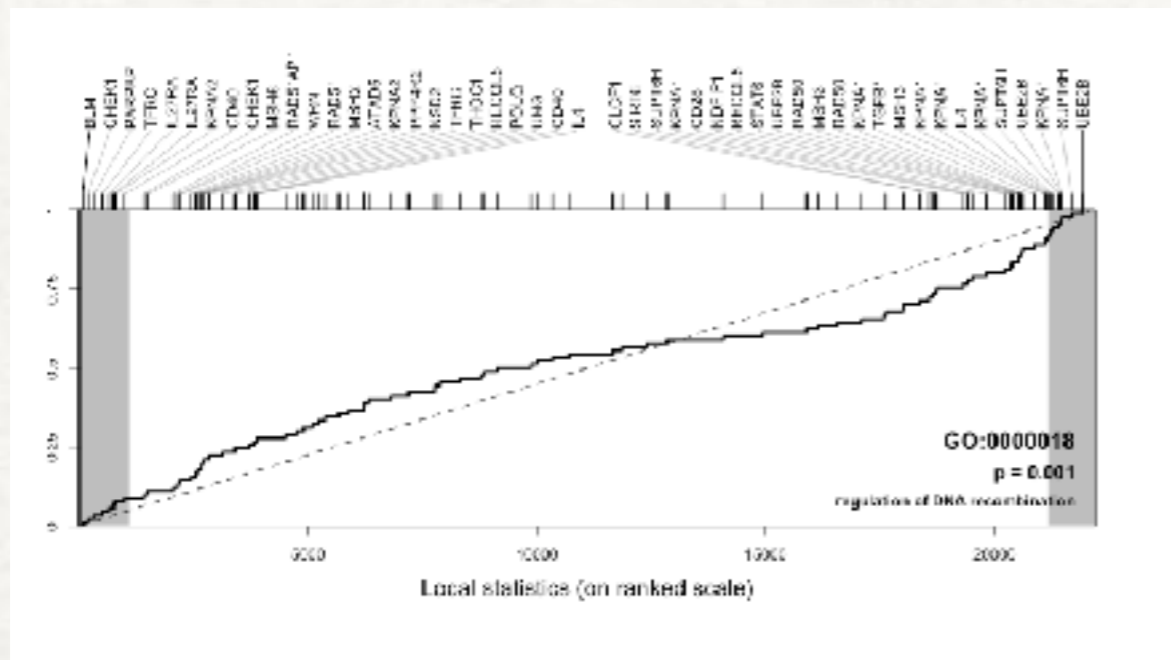
"Safe" function

GO database in R

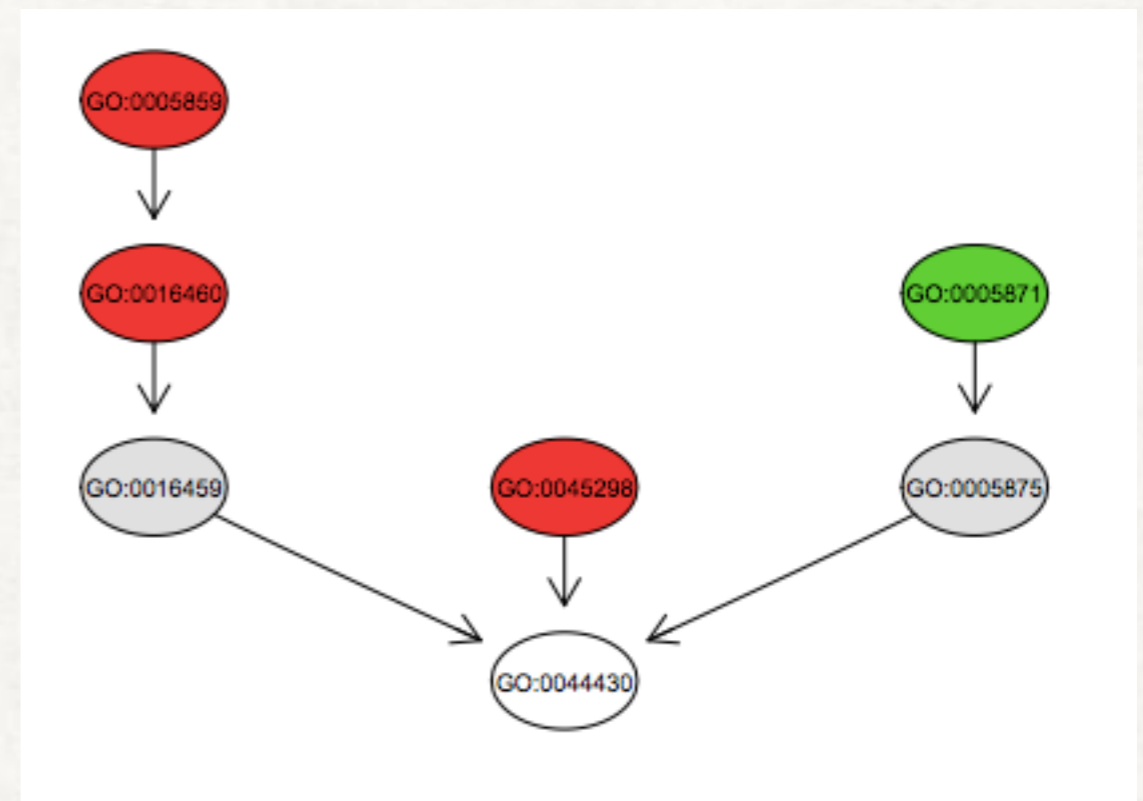
```
> safe.topTable(results1, number = 10, description = FALSE)
  GenesetID Size Statistic P.value Adj.p.value
1 GO:0008150 18393 206899228 0.001 0.3307
2 GO:0009987 17211 193728382 0.001 0.3307
3 GO:0048519 6616 75859417 0.001 0.3307
4 GO:0048523 6001 68863885 0.001 0.3307
5 GO:0051641 3608 41658518 0.001 0.3307
6 GO:0051348 509 6153705 0.001 0.3307
7 GO:0033673 379 4728481 0.001 0.3307
8 GO:0006469 351 4425780 0.001 0.3307
9 GO:0051224 298 3737798 0.001 0.3307
10 GO:0045778 156 2887595 0.001 0.3307
```

Result  
Top 10

# SAFE Significant Analysis of Function and Expression



Significance visualization



Overview of GO category statistics

# **SAFE** Significant Analysis of Function and Expression

- **Bad**

- Many other packages are needed.
- Detailed but complicated