

Supplemental Table 1: The features of the discussed bioinformatics tools. Y, Yes; N, No; **used tools.**

Tools	Only gene list	Gene Expression data	Graphical tool	Cytoscape	Many thousand of genes	Several replicates	Databases	Statistical methods	Gene correlation
Reactome FI(Wu et al., 2010)	Y	N	Y	Y	N	N	Reactome	-	Y
ClueGO (Bindea et al., 2009)	Y	N	Y	Y	N	N	GO, KEGG	-	-
WikiPathways (Kelder et al., 2012)	Y	N	Y	Y	<u>N</u>	<u>N</u>	Wikipathway	-	N
GSEA (Subramanian et al., 2005)	N	Y	Y	Y	Y	Y	MsigDB	Sample randomisation	N
PLAGE (Tomfohr et al., 2005)	N	Y	Y	N	<u>N</u>	<u>N</u>	KEGG Biocarta	Gene randomisation	N
Onto-Tools (Draghici et al., 2007)	N	Y	web	N	Y	Y	KEGG, GO	-	N
SigPathway (Tian et al., 2005)	N	Y	N	N	Y	Y	GO	Permutation	Y
Enrichment MAP (Merico et al., 2010)	N	Y	Y	Y	Y	Y	GO, KEGG,NCI, Biocarta	Sample randomisation	N
Disease /	N	Y	N	N	Y	Y	Own	-	N

Tools	Only gene list	Gene Expression data	Graphical tool	Cytoscape	Many thousand of genes	Several replicates	Databases	Statistical methods	Gene correlation
Phenotype <i>web-PAGE</i> (De et al., 2010)									
GAGE (Luo et al., 2009)	N	Y	N	N	<u>N</u>	<u>N</u>	KEGG, GO	Fold change	N
TopGO (Adrian Alexa, 2012)	N	Y	Y	Y	Y	Y	GO	-	N
NetBox (Research, 2011)	Y	N	N	N	N	N	PID, MSKCC	-	N

Tools	Only gene list	Gene Expression data	Graphical tool	Cytoscape	Many thousand of genes	Several replicates	Databases	Statistical methods	Gene correlation
GeneAnswers (Feng et al., 2010)	N	Y	Y	N	<u>N</u>	<u>N</u>	GO, KEGG, Biomart	-	N
GeneMANIA (Montejo et al., 2010)	Y	Y	Y	Y	<u>N</u>	<u>N</u>	GO	-	N
Gostat (Beissbarth and Speed, 2004)	Y	N	N	N	N	N	GO	-	N
GSEA (Castelo et al., 2011)	Y	Y	N	N	Y	N	GO,KEGG Biocarta	Sample randomisation	N
GEOGLE (Yu et al., 2009)	N	Y	N	N	Y	N	SigPathway	-	Y
T-Profiler (Boorsma et al., 2005)	N	Y	Y	Y	Y	Y	GO	permutation	N
DAVID (Huang et al., 2009a, 2009b)	Y	N	N	N	N	N	-	-	N
ReactomePA (Yu et al., 2012)	Y	Y	Y	N	<u>N</u>	<u>N</u>	KEGG, GO	hypergeometric	N
Cluster Profile (Yu et al., 2012)	Y	Y	Y	N	<u>N</u>	<u>N</u>	KEGG, GO	hypergeometric	N

Software/packages tried but not used in this work

Multiple Variable Correlation Analysis Package

Canonical Correlation Analysis (CCorA) (Jari Oksanen, et al. 2012)

CCA: An R Package to Extend Canonical Correlation Analysis (Ignacio González et al., 2008)

Cormotif ("Bioconductor - Cormotif," 2012).

WGCNA: an R package for weighted correlation network analysis (Langfelder and Horvath, 2008)

(<http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html>)

MPT corr: Multivariate Permutation test for Correlations (Yoder et al., 2004).

Canonical correlation analysis (CCA) ("CRAN - Package CCA," 2012).

Enrichment Network Package

GSEA: Gene Set Enrichment Analysis (Subramanian et al., 2005)

(<http://www.broadinstitute.org/gsea/index.jsp>)

PLAGE: Pathway Level Analysis of Gene Expression (Tomfohr et al., 2005)

Onto-Tools (Draghici et al., 2007) (<http://vortex.cs.wayne.edu/projects.htm>)

Enrichment MAP (Merico et al., 2010) (<http://baderlab.org/Software/EnrichmentMap>)

Disease / Phenotype *web-PAGE* (De et al., 2010)

GAGE: generally applicable gene set enrichment for pathway analysis (Luo et al., 2009)

(<http://www.bioconductor.org/packages/release/bioc/html/gage.html>)

TopGO (Adrian Alexa, 2012) (<http://topgo.bioinf.mpi-inf.mpg.de/>)

(<http://bioconductor.org/packages/release/bioc/html/topGO.html>)

GeneAnswers (Feng et al., 2010)

GeneMANIA (Montejo et al., 2010) (<http://www.genemania.org/>)

WikiPathways (Kelder et al., 2012) (<http://www.pathvisio.org/wiki/SuperPathways>)

NetBox (Research, 2011) (<http://cbio.mskcc.org/tools/netbox.html>)

The Ontologizer (Grossmann et al., 2007)

(<http://compbio.charite.de/contao/index.php/ontologizer2.html>)

Gostat (Beissbarth and Speed, 2004) (<http://gostat.wehi.edu.au/>)

(<http://www.bioconductor.org/packages/release/bioc/html/GOstats.html>)

GSVA: Gene Set Variation Analysis (Castelo et al., 2011)
(<http://www.bioconductor.org/packages/devel/bioc/html/GSVA.html>)

Gene-set analysis and reduction (Dinu et al., 2009)

SigPathway: Pathway Analysis with Microarray Data (Tian et al., 2005)
(<http://www.bioconductor.org/packages/2.9/bioc/html/sigPathway.html>)

GEOGLE (Yu et al., 2009)

GO_enrichment (http://davidang.org/wiki/tiki-index.php?page=go_enrichment.pl)

T-Profiler (Boorsma et al., 2005)

DAVID (Huang et al., 2009a, 2009b) (<http://david.abcc.ncifcrf.gov/home.jsp>)

ReactomePA (Yu et al., 2012)
(<http://www.bioconductor.org/packages/release/bioc/html/ReactomePA.html>)

ClusterProfiler (Yu et al., 2012)
(<http://www.bioconductor.org/packages/2.10/bioc/html/clusterProfiler.html>)

Clippda: A package for clinical proteomic profiling data analysis (Nyangoma, 2012)
(<http://www.bioconductor.org/packages/release/bioc/html/clippda.html>)

Supplemental References for Supplemental Table 1

- Adrian Alexa, 2012. Bioconductor - topGO [WWW Document]. URL <http://www.bioconductor.org/packages/release/bioc/html/topGO.html>
- Beissbarth, T., Speed, T.P., 2004. Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics (Oxford, England)* 20, 1464–5.
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)* 25, 1091–3.
- Bioconductor - Cormotif [WWW Document], 2012. . URL <http://www.bioconductor.org/packages/release/bioc/html/Cormotif.html>
- Boorsma, A., Foat, B.C., Vis, D., Klis, F., Bussemaker, H.J., 2005. T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic acids research* 33, W592–5.
- Castelo, R., Guinney, J., Sonja, H., 2011. BIOINFORMATICS GSVA : Gene Set Variation Analysis. Doctor 1–7.
- CRAN - Package CCA [WWW Document], 2012. . URL <http://cran.r-project.org/web/packages/CCA/>
- De, S., Zhang, Y., Garner, J.R., Wang, S.A., Becker, K.G., 2010. Disease and phenotype gene set analysis of disease-based gene expression in mouse and human. *Physiological genomics* 42A, 162–7.
- Dinu, I., Potter, J.D., Mueller, T., Liu, Q., Adewale, A.J., Jhangri, G.S., Einecke, G., Famulski, K.S., Halloran, P., Yasui, Y., 2009. Gene-set analysis and reduction. *Briefings in Bioinformatics* 10, 24–34.
- Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R., 2007. A systems biology approach for pathway level analysis. *Genome research* 17, 1537–45.
- Feng, G., Du, P., Krett, N.L., Tessel, M., Rosen, S., Kibbe, W.A., Lin, S.M., 2010. A collection of bioconductor methods to visualize gene-list annotations. *BMC research notes* 3, 10.
- Grossmann, S., Bauer, S., Robinson, P.N., Vingron, M., 2007. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics (Oxford, England)* 23, 3024–31.
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37, 1–13.
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44–57.
- Ignacio González, Sébastien Déjean, Pascal G. P. Martin, Alain Baccini, 2008. CCA: An R Package to Extend Canonical Correlation Analysis. *Journal of statistic software* 23.
- Jari Oksanen, F., Guillaume, Blanchet, Roeland, K., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Peter, S., Stevens, H.H., Wagner, H., n.d. Canonical Correlation Analysis (CCorA).
- Kelder, T., Van Iersel, M.P., Hanspers, K., Kutmon, M., Conklin, B.R., Evelo, C.T., Pico, A.R., 2012. WikiPathways: building research communities on biological pathways. *Nucleic acids research* 40, D1301–7.
- Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 9, 559.
- Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., Woolf, P.J., 2009. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics* 10, 161.

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D., 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS one* 5, e13984.

Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S.L., Morris, Q., Bader, G.D., 2010. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics (Oxford, England)* 26, 2927–8.

Nyangoma, S., 2012. Clippda: A package for clinical proteomic profiling data analysis. submitted.

Research, T.C.G.A., 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–15.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545–50.

Tian, L., Greenberg, S.A., Kong, S.W., Altschuler, J., Kohane, I.S., Park, P.J., 2005. Discovering statistically significant pathways in expression profiling studies. *Proc Natl Acad Sci U S A* 102, 13544–13549.

Tomfohr, J., Lu, J., Kepler, T.B., 2005. Pathway level analysis of gene expression using singular value decomposition. *BMC bioinformatics* 6, 225.

Wu, G., Feng, X., Stein, L., 2010. A human functional protein interaction network and its application to cancer data analysis. *Genome biology* 11, R53.

Yoder, P.J., Blackford, J.U., Waller, N.G., Kim, G., 2004. Enhancing power while controlling family-wise error: an illustration of the issues using electrocortical studies. *Journal of clinical and experimental neuropsychology* 26, 320–31.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics : a journal of integrative biology* 16, 284–7.

Yu, Y., Tu, K., Zheng, S., Li, Y., Ding, G., Ping, J., Hao, P., Li, Y., 2009. GEOGLE: context mining tool for the correlation between gene expression and the phenotypic distinction. *BMC bioinformatics* 10, 264.