A classification of tasks for the systematic study of immune response using functional genomics data

Supplemental data file 1 – Classification of relevant information

C Hedeler[1,2*], N. W. Paton[1], J. M. Behnke[3], J. E. Bradley[3], M. G. Hamshere[3], K. J. Else[2]

[1] *School of Computer Science, The University of Manchester,Oxford Road, Manchester, M13 9PL, UK*
[2] *Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester, M13 9PT, UK*
[3] *School of Biology, Nottingham University, Nottingham, NG7 2RD, UK*

Email: C Hedeler - chedeler@cs.manchester.ac.uk;

*Corresponding author

1 Classification of relevant information

1.1 Genome

IID	Data	Data sources	Reference
G-SEQ	**Sequence**		
G-SEQ-DNA	DNA sequence	ENSEMBL	(Hubbard *et al.*, 2002)
G-SEQ-PP	Protein sequence	ENSEMBL	
G-SEQ-TRANS	Transcripts (alternative)	ENSEMBL	
G-SEQ-EXON	Exons, Introns	ENSEMBL	
G-SEQ-FEAT	Feature (UTR)		
G-SEQ-EST	EST		

Table 1. Genome data - Sequence

IID	Data	Data sources	Reference
G-LOC	**Location**		
G-LOC-CHR	Chromosomal location	ENSEMBL, MGD	(Blake *et al.*, 2003)
G-LOC-MARK	Marker	ENSEMBL, MGD, uniSTS	http://www.ncbi.nlm.nih.gov/entrez/
G-LOC-BP	bp position	ENSEMBL	
G-LOC-CM	cM position	MGD	

Table 2. Genome data - Location

IID	Data	Data sources	Reference
G-STR	**Strain**		
G-STR-SNP	SNP	ENSEMBL, dbSNP	(Wheeler *et al.*, 2003)
G-STR-HAP	Haplotype		(Wade *et al.*, 2002)
G-STR-PHENO	Phenotype classification	MGD	
G-STR-ALL	Alleles	MGD	

Table 3. Genome data - Strain

IID	Data	Data sources	Reference
G-FA	**Functional annotation**		
G-FA-GO	*Gene Ontology Annotation*	Gene Ontology Consortium, MGD	(The Gene Ontology Consortium, 2000)
G-FA-GO-MF	Molecular Function	GO, MGD	
G-FA-GO-BP	Biological Process	GO, MGD	
G-FA-GO-CC	Cellular Component	GO, MGD	
G-FA-GO-HS	Hierarchical structure	GO	
G-FA-GO-SS	Semantic similarity		(Lord *et al.*, 2003)
G-FA-SPKW	*SwissProt Keywords*	SwissProt	(Boeckmann *et al.*, 2003)
G-FA-EC	*Enzyme numbers*	LIGAND	(Kanehisa *et al.*, 2002)
G-FA-EC-HS	Hierarchical structure	LIGAND	
G-FA-PF	*Protein families*		
G-FA-PF-ENS	Ensembl protein family	ENSEMBL	
G-FA-PF-IP	InterPro	InterPro	(Mulder *et al.*, 2003)
G-FA-PF-PRO	Prosite	Prosite	(Hulo *et al.*, 2004)
G-FA-PF-PFAM	PFAM	PFAM	(Bateman *et al.*, 2002)
G-FA-PF-PRI	PRINTS	PRINTS	(Attwood *et al.*, 2003)
G-FA-PF-SM	Smart	Smart	(Letunic *et al.*, 2002)
G-FA-PF-TIGR	TIGRFAMs	TIGRFAMs	http://www.tigr.org/TIGRFAMs/
G-FA-PF-PIR	PIR	PIR	(Wu *et al.*, 2003)
G-FA-RIF	*GeneRIF: Gene References into Function*	LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/

Table 4. Genome data - Functional annotation

IID	Data	Data sources	Reference
G-SC	**Species comparison**		
G-SC-HM	Homology matches	ENSEMBL	

Table 5. Genome data - Species comparison

1.2 Transcriptome

IID	Data	Data sources	Reference
T-MD	**Microarray description**	SMD, GEO	(Gollub *et al.*, 2003), (Wheeler *et al.*, 2003)
T-MD-GL	Gene list		
T-MD-MT	Microarray type (Affy, glass)		
T-MD-ST	Spot type (cDNA, oligo)		
T-ECOND	**Experimental condition** Strain, tissue, timepoint, treatment, infection with pathogen, growth condition	SMD, GEO	
T-R	**Result**	SMD, GEO	
T-R-EL	Expression levels		
T-R-TC	Timecourse		

Table 6. Transcriptome data

1.3 Proteome

IID	Data	Data sources	Reference
P-ECOND	**Experimental condition** Strain, tissue, timepoint, treatment, infection with pathogen, growth condition	SWISS-2DPAGE	http://us.expasy.org/ch2d/
P-R	**Result**	SWISS-2DPAGE	
M-R-P	Protein list	SWISS-2DPAGE	

Table 7. Metabolome data

1.4 Metabolome

IID	Data	Data sources	Reference
M-MP	**Metabolic pathways**		
M-MP-CP	classification of pathways		
M-MP-LG	list of genes	KEGG	(Kanehisa *et al.*, 2002)
M-MP-OG	order of genes	KEGG	

Table 8. Metabolome data

1.5 Interactome

IID	Data	Data sources	Reference
I-UT	Used technique		
I-PPI	Observed protein-protein interaction		
I-COMPL	Observed complexes		

Table 9. Interactome data

1.6 Control

IID	Data	Data sources	Reference
C-CMRP	**Cellular, molecular, regulatory pathways**	BioCarta, KEGG	http://www.biocarta.com/
C-CMRP-CP	classification of pathways	BioCarta, KEGG	
C-CMRP-LP	list of proteins	BioCarta	
C-CMRP-OP	order of proteins	BioCarta	
C-CMRP-RP	interaction of proteins	BioCarta	

Table 10. Control data

References

Attwood, T., Bradley, P., Flower, D., Gaulton, A., Maudling, N., Mitchell, A., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A., and Zygouri, C. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research,* **31**, 400–402.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiler, L., Eddy, S., Griffiths-Jones, S., Howe, K., Marshall, M., and Sonnhammer, E. (2002). The Pfam Protein Families Database. *Nucleic Acids Research,* **30**, 276–280.

Blake, J. A., Richardson, J., Bult, C., Kadin, J., Eppig, J., *et al.* (2003). MGD: The Mouse Genome Database. *Nucleic Acids Research,* **31**, 193–195.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M., Estreicher, A., Gasteiger, E., Martin, M., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research,* **31**, 365–370.

Gollub, J., Ball, C. A., Binkley, G., Demeter, J., Finkelstein, D. B., Hebert, J. M., Hernandez-Boussard, T., Jin, H., Kaloper, M., Matese, J. C., *et al.* (2003). The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Research,* **31**, 94–96.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., *et al.* (2002). The Ensembl genome database project. *Nucleic Acids Research, 2002,* **30**, 38–41.

Hulo, N., Sigrist, C., Le Saux, V., Langendijk-Genevaux, P., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Research,* **32**, D134–D137.

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research,* **30**, 42–46.

Letunic, I., Goodstadt, L., Dickens, N., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R., Ponting, C., and Bork, P. (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Research,* **1** (242–244).

Lord, P. W., Stevens, R. D., Brass, A., and Goble, C. A. (2003). Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics,* **19**, 1275–1283.

Mulder, N., Apweiler, R., Attwood, T., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., *et al.* (2003). The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Research,* **31**, 315–318.

The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics,* **25**, 25–29.

Wade, C. M., III, E. J. K., Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., Lindblad-Toh, K., and Daly, M. J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature,* **420**, 574–578.

Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., and Wagner, L. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Research,* **31**, 28–33.

Wu, C. H., Yeh, L.-S. L., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.-Z., Ledley, R. S., Kourtesis, P., Suzek, B. E., Vinayaka, C. R., Zhang, J., and Barker, W. C. (2003). The Protein Information Resource. *Nucleic Acids Research,* **31**, 345–347.