

KOBAS server: a web-based platform for automated annotation and pathway identification

Jianmin Wu, Xizeng Mao, Tao Cai, Jingchu Luo* and Liping Wei*

Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering,
College of Life Sciences, Peking University, Beijing 100871, P. R. China

Received February 13, 2006; Revised February 23, 2006; Accepted March 20, 2006

ABSTRACT

There is an increasing need to automatically annotate a set of genes or proteins (from genome sequencing, DNA microarray analysis or protein 2D gel experiments) using controlled vocabularies and identify the pathways involved, especially the statistically enriched pathways. We have previously demonstrated the KEGG Orthology (KO) as an effective alternative controlled vocabulary and developed a standalone KO-Based Annotation System (KOBAS). Here we report a KOBAS server with a friendly web-based user interface and enhanced functionalities. The server can support input by nucleotide or amino acid sequences or by sequence identifiers in popular databases and can annotate the input with KO terms and KEGG pathways by BLAST sequence similarity or directly ID mapping to genes with known annotations. The server can then identify both frequent and statistically enriched pathways, offering the choices of four statistical tests and the option of multiple testing correction. The server also has a 'User Space' in which frequent users may store and manage their data and results online. We demonstrate the usability of the server by finding statistically enriched pathways in a set of upregulated genes in Alzheimer's Disease (AD) hippocampal cornu ammonis 1 (CA1). KOBAS server can be accessed at <http://kobas.cbi.pku.edu.cn>.

INTRODUCTION

Automated analysis of large sets of genes and proteins requires that they be annotated with a common controlled vocabulary. Gene Ontology (GO) (1) which comprises over 19 000 terms

in molecular function, biological process and cellular component, has been one of the most widely used controlled vocabularies. GO has been used to annotate whole genomes and find enriched functional categories in upregulated or clustered genes in microarray experiments. A variety of web-based tools have been developed for GO-based analysis, including Gotcha (2), GoFigure (3), FatiGO (4), GFINDER (5), GOstat (6), NetAffx (7), GOToolBox (8) and Onto-Tools (9). However, a weakness of GO is that its terms do not correspond directly to pathways.

Knowing the pathways involved in a set of genes or proteins, especially the statistically enriched pathways, could offer more biological insights and generate more directly testable hypotheses. Towards this goal, we have previously studied the KEGG Orthology (KO) (10,11), part of the KEGG suite of resources (12), as an alternative controlled vocabulary (13). We demonstrated that KO is effective in automated annotation of sets of sequences based on similarity to sequences with known KO annotations in the KEGG GENES database. Moreover, because KO is directly linked to KEGG pathways, it enables pathway identification. We developed and tested a KO-Based Annotation System (KOBAS) to find both the most frequent and most statistically significantly enriched pathways in a set of genes or proteins (13). KOBAS can be used to analyze whole genomes and results from DNA microarray and protein 2D gel experiments. For instance, Shi *et al.* (14) used KOBAS to find that ethylene biosynthesis is the most enriched pathway in a set of cotton fiber-specific genes from the cotton transcriptome profiling data; they then validated the finding by physiological and biochemical experiments.

The standalone KOBAS package has been downloaded and used by scientists worldwide. However, given that KOBAS comprises underlying SQLite relational databases, R statistics package, Python scripts and other necessary programs, we have received comments citing difficulties in installation and updating. Therefore, to benefit the entire scientific community, we here report a KOBAS web server with a user-friendly interface.

*To whom correspondence should be addressed. Tel: +86 10 6275 5206; Fax: +86 10 6275 9001; Email: weilp@mail.cbi.pku.edu.cn

*Correspondence may also be addressed to Jingchu Luo. Tel: +86 10 6275 7281; Fax: +86 10 6275 9001; Email: luojc@pku.edu.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

In addition to the existing functionalities of standalone KOBAS, the new web server has the following new features: (i) it allows the input of not only gene or protein sequences but also sequence identifiers (IDs); (ii) it offers three more statistical tests—binomial, χ^2 and Fisher's exact tests—in addition to the original hypergeometric distribution test to find significantly enriched pathways; (iii) the underlying databases are updated regularly, a process that is transparent to users; (iv) it has an improved HTML output format; and (v) it allows users to store and manage input data and output results online.

In addition to KOBAS, several tools and servers have recently been developed to identify enriched pathways in microarray data; these tools include ArrayXPath (15), PathwayExplorer (16), VAMPIRE (17) and Pathway-Express (9). However, a common feature of all these other tools and servers is that they can only take IDs as input to map directly to pathway databases. This greatly limits the usefulness of the tools because many important organisms have not been annotated or are poorly annotated in pathway databases (e.g. cotton). Direct ID mapping would fail for sequences in these organisms as well as for any newly sequenced genome, EST or cDNA sequences. Even for well-annotated genomes, usually not all transcripts of the same gene are annotated in the pathway databases.

During preparation of this manuscript, a KEGG Automatic Annotation Server (KAAS, <http://www.genome.jp/kegg/kaas/>) came online that can also annotate a set of sequences with KO terms. However, it does not have any other functionalities, such as the important statistical testing of significance of the identified pathways or input by ID. Assigning statistical significance to the pathways is one of the critical features of KOBAS and has been shown to lead to validated hypotheses. To date, KOBAS is the only server that has integrated all the aforementioned functionalities.

ANALYSIS TOOLS

The KOBAS server divides the analysis into two steps to provide more flexibility. The first step annotates a set of genes or proteins (as IDs or sequences) with KO terms. The second step identifies the frequent or statistically significant pathways. Users can also manage their data and results online.

KO annotation

If a user inputs a list of IDs from popular sequence databases, KOBAS will map these IDs directly to genes with known KO annotations using the cross-links we parsed from the KEGG GENES database. If there is a match, the KO terms of the KEGG gene are assigned to the query gene or protein. The list of acceptable sequence databases for each organism is available on the web site in the FAQ section of Documentation. For example, for human, GI numbers of GenBank and IDs of the NCBI Gene database, UniProt, GDB and OMIM are acceptable; for *Saccharomyces cerevisiae*, GI numbers of GenBank and IDs of the NCBI Gene database, UniProt, SGD and MIPS are acceptable. If there is a match, the output contains four columns: query sequence identifier, KO term, KO term definition and the KEGG gene to which the query is mapped.

If a user inputs a set of nucleotide or amino acid sequences in FASTA format (by uploading a file or pasting directly into

the input window), KOBAS assigns the KO terms for each sequence based on sequence similarity with entries in KEGG GENES using BLAST (18). We chose BLAST E -value $\leq 10^{-5}$ and rank ≤ 5 as the default cutoffs, meaning that a new sequence is assigned the KO term(s) of the first BLAST hit that (i) has BLAST E -value $\leq 10^{-5}$, (ii) has known KO assignments and (iii) has less than five other hits with a lower E -value that do not have KO assignments (13). Users have the option to adjust the cutoffs to increase sensitivity or specificity. A lower E -value or rank returns more reliable mapping results but may leave more sequences unannotated, whereas a higher E -value or rank annotates more sequences but may have some false positives. The choice of the cutoff criteria and the tradeoff between sensitivity and specificity was previously studied (13).

Figure 1 shows the output of KO annotation when the input is a set of FASTA sequences. Each row corresponds to a query DNA or protein and lists the sequence identifier extracted from the input, the assigned KO term (hyperlinked to detailed description in KEGG), definition of the KO term, the rank, E -value, score and percent identity of the BLAST hit, and the gene ID of the hit in the KEGG GENES database. If one sequence is annotated with multiple KO terms, each KO annotation is presented in a separate row.

| SEQUENCE IDENTIFIER | KO TERM | KO DEFINITION | RANK | E-VALUE | SCORE | IDENTITY (%) | BLAST HIT |
|-----------------------------------|------------------------|---|------|---------|--------|--------------|---------------------------|
| gi 307414 gb L10338.1 HUMSCN18A | K04845 | sodium channel, voltage-gated, type I, beta | 1 | 1e-101 | 370.0 | 93.91 | hsa:6324 |
| gi 27894305 ref NM_000576.2 | K04510 | interleukin 1, beta | 1 | 1e-145 | 516.0 | 95.17 | hsa:3553 |
| gi 3927886 dbj AB017365.1 | K02378 | frodoled 7 | 1 | 0.0 | 1113.0 | 93.38 | hsa:8324 |
| gi 1778716 gb J64197.1 HSU64197 | K05514 | chemokine (C-C motif) ligand, other | 2 | 3e-39 | 163.0 | 83.33 | hsa:6344 |
| gi 1908170 gb JAF015956.2 | K02308 | death-associated protein 6 | 1 | 0.0 | 1083.0 | 78.1 | hsa:1616 |
| gi 183988 gb M12036.1 HUM-ER2B | K05683 | v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 | 2 | 5e-31 | 137.0 | 64.41 | mo:2427 |
| gi 29903 emb V01512.1 HSCF05 | K04379 | v-fos FBJ murine osteosarcoma viral oncogene homolog | 2 | 5e-99 | 364.0 | 76.68 | hsa:2353 |
| gi 190871 gb M39258.1 HUMRARGA | K06500 | integrin beta 7 | 5 | 0.0 | 696.0 | 91.49 | mo:2511 |
| gi 4504492 ref NM_003086.1 | K02512 | activator of apoptosis harakon (neural death protein DPS), BCL2-interacting | 1 | 7e-29 | 128.0 | 68.13 | hsa:8730 |
| gi 3415122 gb JF081281.1 AF081281 | K06128 | lysophospholipase I | 1 | 1e-131 | 470.0 | 100 | hsa:10434 |
| gi 32027 emb Z11697.1 HSHB15PNA | K06510 | COB9 antigen | 1 | 1e-113 | 411.0 | 97.07 | hsa:9208 |
| gi 189177 gb M58603.1 HUMMF4B | K02580 | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 | 1 | 0.0 | 1836.0 | 95.46 | hsa:4790 |
| gi 34577121 ref NM_003998.2 | K02580 | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 | 1 | 0.0 | 1843.0 | 95.67 | hsa:4790 |
| gi 14730537 ref XM_051899.1 | K00500 | prostaglandin-endoperoxide synthase | 1 | 0.0 | 1184.0 | 97.12 | hsa:5743 |
| gi 338005 gb M28225.1 HJMSECP3 | K05509 | chemokine (C-C motif) ligand 2/7/8/11/21/3 | 1 | 2e-08 | 59.3 | 100 | hsa:6347 |
| gi 13570016 gb J47077.5 HSU47077 | K06642 | protein kinase, DNA-activated, catalytic polypeptide | 1 | 0.0 | 7941.0 | 97.09 | hsa:5591 |

Figure 1. Screenshot of the output of KO annotation when the input is FASTA sequences. The 21 of the 36 upregulated genes in AD CA1 were assigned KO terms based on sequence similarity. Each row corresponds to a query DNA or protein input by the user. The first column contains sequence identifiers extracted from the input. The second column contains the assigned KO terms hyperlinked to detailed descriptions in KEGG. The third column contains KO term definitions. The fourth to seventh columns show the rank, E -value, score and identity of the BLAST hit. The last column contains the gene ID of the hit hyperlinked to the KEGG GENES database. Users can choose to view the results in HTML or text format, edit the text format online and download results to local disks. Users can also select the program for further analysis using the annotation results as input directly.

Because BLAST is computationally intensive, it may not be possible to return results to users immediately if the input is large. The KOBAS server displays a URL if the job cannot be finished within one minute so that the user can access the results later. Alternatively, if the user supplies an e-mail address, the results will be emailed automatically upon completion of the job.

Pathway identification

After a set of genes or proteins are annotated with KO, the user can choose to identify the frequently occurring or the statistically enriched pathways in the set. The input is the output of the previous step, 'KO Annotation'. Since the third level in the KO hierarchy corresponds to KEGG pathways, we can trace the KO terms of a gene back through the KO hierarchy to its associated pathways. The frequently occurring pathways can be easily identified by tallying the number of genes or proteins associated with each pathway and ranking the numbers. The output lists the name of each of the pathways and the number and percentage of the query genes or proteins that are involved in each pathway.

However, as some pathways are naturally large and would involve more genes or proteins in any set just by chance, it is important to identify the statistically significantly enriched pathways compared with a background distribution. The user can use a whole genome as the background distribution by selecting from the list of genomes annotated with KO, or they can use any set of genes or proteins (e.g. the entire probe set on a microarray) annotated with KO as background by uploading or pasting the annotations. Next, the user may choose from four statistical tests—binomial, χ^2 , Fisher's exact and hypergeometric distribution tests—and the choice of whether to perform multiple testing correction using FDR. The output shows the statistically enriched pathways, listing the pathway name, the number and percentage of the query genes or proteins that are involved in each pathway, the number and percentage of the background genes or proteins that are involved in each pathway, right-tailed p -value and FDR-corrected q -value (if applicable) (Figure 2). The pathways are sorted by increasing p -value, or q -value (if applicable) from most significant to least. Each pathway name is linked to a page of detailed information including all the genes/proteins involved and hyperlinks to the KEGG pathway maps, with the relevant KO terms highlighted.

Online data management

For the convenience of frequent users of KOBAS, the server provides online data management functionalities. Registration is free and open to all. A registered user can save input files and output files in a private 'User Space' on the server. The User Space supports a tree-like structure to organize directories and files as shown in Figure 3. The saved input files and intermediate output files can be easily selected for repeated analysis using different parameters. All user input and output are strictly confidential. For guest users, the files are kept on the server for 7 days, whereas for registered users, the files are kept for 6 months.

The user can view the analysis history and monitor the job status online, including information on the program, input and output, start time and elapsed time and status of each job.

| PATHWAY | COUNT AND RATIO | P-VALUE | Q-VALUE |
|--|---------------------------|-------------------|------------------|
| Apoptosis | 5 / 23.81% 110 / 1.90% | 4.40921664574e-05 | 0.00119048849435 |
| MAPK signaling pathway | 6 / 28.57% 262 / 4.53% | 0.000275661785324 | 0.00372143410187 |
| Toll-like receptor signaling pathway | 4 / 19.05% 113 / 1.95% | 0.000718668370863 | 0.00646801533777 |
| Cytokines | 5 / 23.81% 223 / 3.85% | 0.00108457980742 | 0.00732091370006 |
| Cytokine-cytokine receptor interaction | 5 / 23.81% 270 / 4.67% | 0.00249861219818 | 0.0134925058702 |
| B cell receptor signaling pathway | 3 / 14.29% 100 / 1.73% | 0.00571424674331 | 0.0257141103449 |
| T cell receptor signaling pathway | 3 / 14.29% 110 / 1.90% | 0.00739151375494 | 0.0285101244834 |
| Type I diabetes mellitus | 2 / 9.52% 43 / 0.74% | 0.0112302439862 | 0.0379020734534 |
| Biotin metabolism | 1 / 4.76% 8 / 0.14% | 0.0321018759672 | 0.0787955137377 |
| Adipocytokine signaling pathway | 2 / 9.52% 72 / 1.24% | 0.0287743067227 | 0.0787955137377 |
| Calcium signaling pathway | 3 / 14.29% 190 / 3.28% | 0.0309354500873 | 0.0787955137377 |
| Hematopoietic cell lineage | 2 / 9.52% 87 / 1.50% | 0.0403893080778 | 0.090875943175 |
| Prostaglandin and leukotriene metabolism | 1 / 4.76% 42 / 0.73% | 0.144741052789 | 0.300616032717 |

Figure 2. Screenshot of the list of statistically enriched pathways identified in the upregulated genes in AD CA1, sorted by increasing q -value. The first column shows the name of the pathway. The second column lists the number and percentage of input genes or proteins involved in the pathway (top) and the number and percentage of background genes or proteins involved in the pathway. The third and fourth columns list the p -value and q -value of the statistical significance, respectively.

A submitted job has five possible statuses: submitted, running, in queue, finished and failed. A job is put in queue if 10 other jobs are already running on the server. When a job finishes, the output files are automatically saved in the User Space. If a job fails because of invalid input or other reasons, the detailed error message is logged.

IMPLEMENTATION

The KOBAS web server was developed using the platform-independent Java language. Apache Tomcat was used as a container for Java Servlet and JSP. User account information, uploaded input files and analysis results are stored in the MySQL database. The KOBAS server runs on a Linux box (4 Intel Xeon 2.20 GHz and 8 GB RAM). For very large jobs, the user can download the standalone version of KOBAS to run locally. The server web site includes a step-by-step tutorial (with screenshots) for general users as well as detailed technical documentation and an online browser of KOBAS source code for software developers.

EXAMPLE APPLICATION

One of our other research projects involves the genes and pathways involved in Alzheimer's Disease (AD). Colangelo



Figure 3. Screenshot of User Space. Users can organize their data and results in a tree-like structure. Users can upload files from their local disk to the KOBAS server and use them later as input. The output of any analysis will be automatically stored in the User Space for further analysis.

et al. (19) analyzed the expression profiles of 12 633 genes in the AD hippocampal cornu ammonis 1 (CA1) versus healthy controls using DNA microarrays. Using their dataset (http://www.medschool.lsuhsu.edu/neuroscience/faculty/Lukiw_researchspreadsheet.xls), we identified 36 genes as upregulated in AD CA1 with standard criteria ($P \leq 0.05$, 2-fold change). We then used 'KO Annotation' on the KOBAS server to annotate 21 of the 36 upregulated genes with KO terms using default cutoffs and 'Pathway Identification' to find statistically enriched pathways using the χ^2 test (Figure 1 and 2). A literature review showed that the top five pathways all have been associated with AD; these include apoptosis (caspase activation, a key step in apoptosis, leads to the proteolytic cleavage of *tau*) (20), mitogen-activated protein kinase (MAPK) signaling (implicated in the hyperphosphorylation of *tau*, a major component of the neurofibrillary tangles) (21), Toll-like receptor signaling (activating signal transduction pathways that stimulate immune function) (22), cytokines (promoting and sustaining inflammatory responses—a central feature of AD) (23) and cytokine–cytokine receptor interactions (associated with MAPK expression) (21).

DISCUSSION

Although there are several online servers for pathway analysis, KOBAS provides the most comprehensive set of functionalities including input by both IDs and sequences, finding both frequent and statistically enriched pathways, four choices of statistical tests, online management of data and analysis, both web-based and standalone versions of the program and both step-by-step tutorial for novice users and detailed technical documentation for bioinformaticians.

The power of KOBAS is limited by the number of input genes or proteins that can be assigned KO terms, which in turn is limited by the number of genes and proteins that have known KO annotations. Our previous experience indicates that typically 30–50% of gene or protein sequences in a newly sequenced genome can be assigned KO terms by BLAST similarity. This percentage is slightly lower than that for GO-based annotations, as more genes and proteins have known GO annotations than have KO annotations. However, this gap will decrease as more KO annotations become available.

The implementation of four statistical tests offers more flexibility to suit different analysis needs. The hypergeometric test requires that input annotations be a subset of the background annotations. For the chi-square test, when χ^2 becomes unreliable (expected frequencies <5), KOBAS will automatically switch to the Fisher's exact test. The binomial test is faster when the number of sequences is large.

Currently the KOBAS server allows 10 jobs to run concurrently and puts the other jobs in queue. KO annotation by sequence similarity is limited to 500 sequences per job. There is no limit for input by IDs or pathway identification. We are currently developing a distributed computing version of KOBAS on a cluster, which will enable the server to handle more jobs at a higher computational rate.

ACKNOWLEDGEMENTS

This work was supported by the China Human Liver Proteome Project and the China National High-tech Program (863). Funding to pay the Open Access publication charges for this article was provided by Ministry of Science and Technology of China.

Conflict of interest statement. None declared.

REFERENCES

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
- Martin, D.M., Berriman, M. and Barton, G.J. (2004) GOTcha: a new method for prediction of protein function assessed by the annotation of seven genomes. *BMC Bioinformatics*, **5**, 178.
- Khan, S., Situ, G., Decker, K. and Schmidt, C.J. (2003) GoFigure: automated Gene Ontology annotation. *Bioinformatics*, **19**, 2484–2485.
- Al-Shahrour, F., Diaz-Uriarte, R. and Dopazo, J. (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, **20**, 578–580.
- Masseroli, M., Martucci, D. and Pinciroli, F. (2004) GFINDER: Genome Function INtegrated Discoverer through dynamic annotation, statistical analysis, and mining. *Nucleic Acids Res.*, **32**, W293–W300.
- Beissbarth, T. and Speed, T.P. (2004) Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, **20**, 1464–1465.
- Cheng, J., Sun, S., Tracy, A., Hubbell, E., Morris, J., Valmeekam, V., Kimbrough, A., Cline, M.S., Liu, G., Shigeta, R. *et al.* (2004) NetAffx Gene Ontology Mining Tool: a visual approach for microarray data analysis. *Bioinformatics*, **20**, 1462–1463.
- Martin, D., Brun, C., Remy, E., Mouren, P., Thieffry, D. and Jacq, B. (2004) GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol.*, **5**, R101.
- Draghici, S., Khatri, P., Bhavsar, P., Shah, A., Krawetz, S.A. and Tainsky, M.A. (2003) Onto-Tools, the toolkit of the modern

- biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. *Nucleic Acids Res.*, **31**, 3775–3781.
10. Kanehisa, M. (1997) A database for post-genome analysis. *Trends Genet.*, **13**, 375–376.
 11. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
 12. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
 13. Mao, X., Cai, T., Olyarchuk, J.G. and Wei, L. (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, **21**, 3787–3793.
 14. Shi, Y.H., Zhu, S.W., Mao, X.Z., Feng, J.X., Qin, Y.M., Zhang, L., Cheng, J., Wei, L.P., Wang, Z.Y. and Zhu, Y.X. (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell*, **18**, 651–664.
 15. Chung, H.J., Kim, M., Park, C.H., Kim, J. and Kim, J.H. (2004) ArrayXPath: mapping and visualizing microarray gene-expression data with integrated biological pathway resources using Scalable Vector Graphics. *Nucleic Acids Res.*, **32**, W460–W464.
 16. Mlecnik, B., Scheideler, M., Hackl, H., Hartler, J., Sanchez-Cabo, F. and Trajanoski, Z. (2005) PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res.*, **33**, W633–W637.
 17. Hsiao, A., Ideker, T., Olefsky, J.M. and Subramaniam, S. (2005) VAMPIRE microarray suite: a web-based platform for the interpretation of gene expression data. *Nucleic Acids Res.*, **33**, W627–W632.
 18. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
 19. Colangelo, V., Schurr, J., Ball, M.J., Pelaez, R.P., Bazan, N.G. and Lukiw, W.J. (2002) Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *J. Neurosci. Res.*, **70**, 462–473.
 20. Cotman, C.W., Poon, W.W., Rissman, R.A. and Blurton-Jones, M. (2005) The role of caspase cleavage of tau in Alzheimer disease neuropathology. *J. Neuropathol. Exp. Neurol.*, **64**, 104–112.
 21. Ho, G.J., Drego, R., Hakimian, E. and Masliah, E. (2005) Mechanisms of cell signaling and inflammation in Alzheimer's disease. *Curr. Drug Targets Inflamm. Allergy*, **4**, 247–256.
 22. Aderem, A. and Ulevitch, R.J. (2000) Toll-like receptors in the induction of the innate immune response. *Nature*, **406**, 782–787.
 23. McGeer, P.L. and McGeer, E.G. (2001) Inflammation, autotoxicity and Alzheimer disease. *Neurobiol. Aging*, **22**, 799–809.