W206–W211 *Nucleic Acids Research, 2007, Vol. 35, Web Server issue*
doi:10.1093/nar/gkm327

Onto-Tools: new additions and improvements in 2006

Purvesh Khatri[1], Calin Voichita[1], Khalid Kattan[1], Nadeem Ansari[1], Avani Khatri[1], Constantin Georgescu[1], Adi L. Tarca[2] and Sorin Draghici[1,]*

[1]Department of Computer Science, Wayne State University, 431 State Hall, Detroit, MI, 48202 and [2]Perinatology Research Branch-NIH/NICHD, 4 Brush, 3990 John R, Detroit, MI 48201, USA

Received January 31, 2007; Revised April 10, 2007; Accepted April 17, 2007

ABSTRACT

Onto-Tools is a freely available web-accessible software suite, composed of an annotation database and nine complementary data-mining tools. This article describes a new tool, Onto-Express-to-go (OE2GO), as well as some new features implemented in Pathway-Express and Onto-Miner over the past year. Pathway-Express (PE) has been enhanced to identify significantly perturbed pathways in a given condition using the differentially expressed genes in the input. OE2GO is a tool for functional profiling using custom annotations. The development of this tool was aimed at the researchers working with organisms for which annotations are not yet available in the public domain. OE2GO allows researchers to use either annotation data from the Onto-Tools database, or their own custom annotations. By removing the necessity to use any specific database, OE2GO makes the functional profiling available for all organisms, with annotations using any ontology. The Onto-Tools are freely available at http://vortex.cs.wayne.edu/projects.htm.

INTRODUCTION

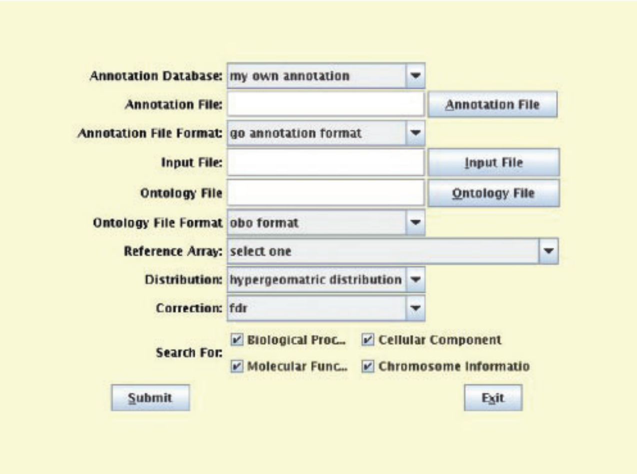
Together with the ability of generating a large amount of data per experiment, high-throughput technologies also brought the challenge of translating such data into a better understanding of the underlying biological phenomena. First released in 2001, Onto-Tools is a freely available web-accessible software suite that addresses some of these challenges. The Onto-Tools suite includes: (i) Onto-Express—used to translate lists of differentially regulated genes into a better understanding of the underlying biological phenomena (1–5); (ii) Onto-Design—used to select the best set of genes to be included on a custom microarray designed for the study of a given biological phenomenon (2,4); (iii) Onto-Compare—used to analyze the functional bias of various focused commercial microarrays and select the one that is most appropriate for a given biological hypothesis (2,6); (iv) Onto-Translate—used to translate lists of genes from one reference system to another (e.g. from GenBank accession numbers to UniGene cluster IDs to Affymetrix probe IDs, etc.) (2,5,7,8); (v) Onto-Miner—provides a unified access point and an application programming interface (API) allowing queries for various information such as the gene name, official symbol, reference accession number, coded protein, etc. (4); (vi) Promoter-Express—which allows the users to find condition-specific transcription factor binding sites (TFBSs) (7,9) and (vii) nsSNPCounter—which allows analysis of synonymous and non-synonymous codon substitutions in protein coding genes (7). Previous publications have described in detail the motivation, implementation and validation of these tools. The logical workflow between the Onto-Tools applications has also been previously explained (2,4). This article describes two new tools added to the ensemble and discusses other enhancements recently made to the existing tools.

OE2GO

Onto-Express (OE) is a web-based tool in the Onto-Tools suite that performs automated function profiling for a list of differentially expressed genes. However, Onto-Express does not support functional profiling for the organisms that do not have annotations in public domain, or use of custom (i.e. user-defined) ontologies. This limitation is also true for most of the other existing tools for functional profiling (10), which means that researchers working with uncommon organisms and/or new annotations or ontologies may be forced to construct such profiles manually.

Onto-Express-to-go (OE2GO) is a new tool added to the Onto-Tools ensemble to address these issues. OE2GO is built on top of OE to leverage its existing functionality. In OE2GO, the users now have an option to use either the Onto-Tools database as a source of functional annotations or provide their own annotations in a separate file (Figure 1). Currently, OE2GO supports annotation file in the Gene Ontology format. A GO-formatted annotation file has 15 tab-delimited columns that contain a database

*To whom correspondence should be addressed. Tel: +1 313 577 5484; Fax: +1 313 577 6868; Email: sorin@wayne.edu



Annotation Database: my own annotation

Annotation File:

Annotation File Format: go annotation format

Input File:

Ontology File

Ontology File Format obo format

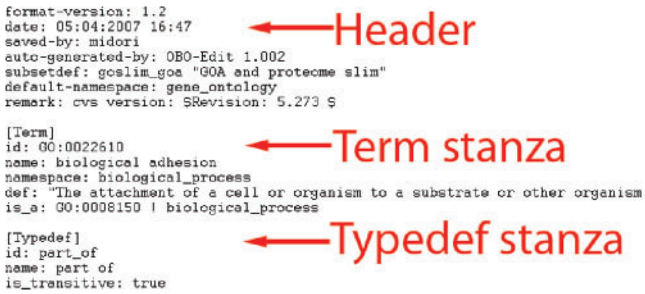
Reference Array: select one

Distribution: hypergeomatric distribution

Correction: fdr

Search For: ☑ Biological Proc... ☑ Cellular Component ☑ Molecular Func... ☑ Chromosome Informatio

Figure 1. OE2GO input interface for using custom annotations. When the user selects 'my own annotations', OE2GO allows the user to specify the name of the annotation file in GO format and the ontology file in OBO format. When using custom annotations, OE2GO assumes that the identifiers in the input file are of the same type as those used in the annotations file.



```
format-version: 1.2
date: 05:04:2007 16:47
saved-by: midori
auto-generated-by: OBO-Edit 1.002
subsetdef: goslim_goa "GOA and proteome slim"
default-namespace: gene_ontology
remark: cvs version: $Revision: 5.273 $

[Term]
id: GO:0022610
name: biological adhesion
namespace: biological_process
def: "The attachment of a cell or organism to a substrate or other organism."
is_a: GO:0008150 ! biological_process

[Typedef]
id: part_of
name: part of
is_transitive: true
```

Figure 2. An example OBO formatted ontology file that contains a header, a term stanza and a typedef stanza. Each stanza starts with either '[Term]' or '[Typedef]'. The header and the stanzas contain a set of tag-value pairs in <tag>:<value> format. The complete list of allowed tags for each type of stanza and the header are described on GO web site.

name, a unique ID for an entity being annotated, its corresponding symbol, one or more references supporting an annotation, type of evidence, date of annotation, type of entity (e.g. a gene or a protein), etc. The detailed description of the format and each column is available at www.geneontology.org/GO.annotation.shtml. The annotation files for approximately 40 organisms in GO format are available for download from GO ftp site that can be directly used with OE2GO (ftp://ftp.geneontology.org/pub/go/gene-associations/).

As shown in Figure 1, when using custom annotations, the user must also specify the ontology file in OBO format. Figure 2 shows an example ontology file in OBO format. An ontology file in OBO format contains a header section and a number of stanzas consisting of a set of tag-value pairs. A tag-value pair consists of a tag name, a colon and a tag value. The header section must be before any stanzas, and contains meta-information about the ontology such as its creation date, default namespace, remarks, format version, etc. (Figure 2). A stanza can be of type 'Term' or 'Typedef', where a term stanza describes an ontology term, and a typedef stanza describes a type of relationship between two terms in the ontology. A complete detailed description of the format is available at www.geneontology.org/GO.format.shtml. OE2GO uses Java code from an open-source tool OBO-edit in order to parse the ontology file in OBO format. The Gene Ontology Consortium provides an ontology file in OBO format that can be directly used with OE2GO (www.geneontology.org/ontology/gene_ontology_edit.obo).

For each gene in the input file, OE2GO searches the annotation file specified by the user. Hence, the functional profiles created by OE2GO depend on the information present in the annotation file. If the file does not contain the annotations that are otherwise well known, they are not provided in the OE2GO results. The strength of OE2GO is that it enables the researchers to use the functional annotations that are not yet in the public domain, by allowing them to be included in the annotation file. Another advantage of OE2GO is that it enables the researchers to avoid annotation bias (10) by allowing them to remove the biological processes that are more studied than the others from the annotation file.

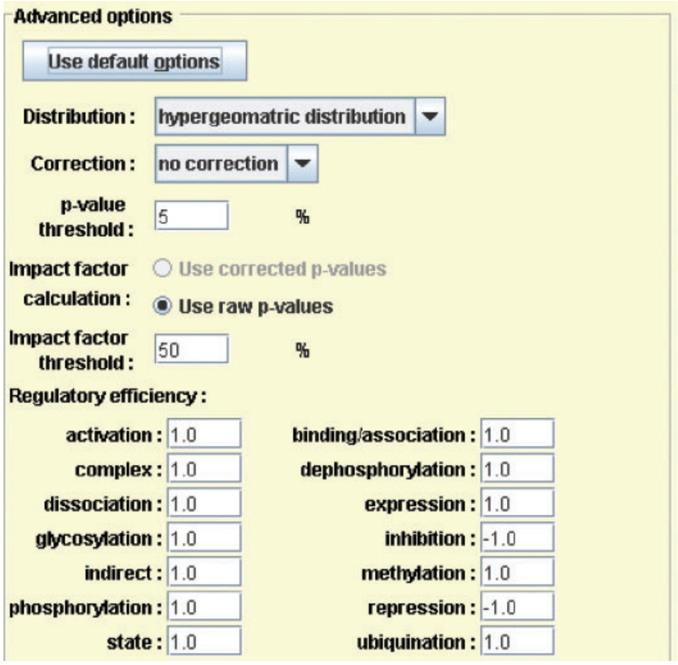
PATHWAY-EXPRESS

The automated functional profiling approach, first proposed by Onto-Express in 2001, has now become the *de facto* standard in the second stage analysis of gene expression data (10). A large number of tools performing similar ontological analysis are available today. Although this approach is widely adopted, it considers each biological process independent of the others, and ignores dependencies and interactions among them (10).

At the same time, a number of pathway databases available in public domain describe how genes interact with each other in metabolic and signaling pathways (e.g. KEGG (11), BioCarta (www.biocarta.com), Reactome (12), etc.). Several tools that allow researchers to reveal the pathways associated with a given set of differentially expressed genes already exist (13–23).

Pathway-Express (PE) is a tool in the Onto-Tools ensemble that is designed to perform a pathway analysis. When a user submits a list of genes, PE searches the Onto-Tools database and builds a list of all associated pathways. The Onto-Tools database currently contains signaling pathways from KEGG. However, PE can analyze any collection of pathways described in SBML (24). PE performs a classical enrichment analysis based on a hypergeometric distribution in order to identify those pathways that contain a proportion of differentially expressed genes that is significantly different from what is expected just by chance. This analysis produces a set of P-values that characterize the significance of the pathway from this statistical perspective (a lower P-value corresponds to a higher significance).

PE also calculates a perturbation factor $\mathrm{PF}(g)$ for each gene on each pathway. This perturbation factor takes into consideration the (i) normalized fold change of the gene



Distribution : hypergeomatric distribution

Correction : no correction

p-value threshold : 5 %

Impact factor calculation : ○ Use corrected p-values ● Use raw p-values

Impact factor threshold : 50 %

Regulatory efficiency :

activation :	1.0	binding/association :	1.0
complex :	1.0	dephosphorylation :	1.0
dissociation :	1.0	expression :	1.0
glycosylation :	1.0	inhibition :	-1.0
indirect :	1.0	methylation :	1.0
phosphorylation :	1.0	repression :	-1.0
state :	1.0	ubiquination :	1.0

Figure 3. Pathway-Express advanced options. Pathway-Express allows the users to choose advanced options such as the distribution to use for calculating significance of pathways, the type of correction to apply for multiple hypotheses, whether to use *P*-values or corrected *P*-values for calculating impact factor, etc. It also allows a user to specify different weights for different types of interactions (e.g. activation, binding, repression, phosphorylation, etc.) between genes on the pathways.

and (ii) the number and amount of perturbation of genes upstream of it (i.e. its position on a pathway). The users can use 'Advanced Options' button to specify different weights for different types of interactions between genes on the pathways (Figure 3). As shown in Figure 3, PE uses negative weights for 'inhibition' and 'repression'. This gene perturbation factor reflects the relative importance of each differentially expressed gene on the pathway. The impact factor of the entire pathway is calculated using a probabilistic term that takes into consideration the proportion of differentially expressed genes on the pathway and gene perturbation factors of all genes in the pathway. More details about the gene perturbation factors and pathway impact factors, a comparison with the existing methods, and a full discussion of the advantages and disadvantages of these methods are described elsewhere.

When a user submits a list of input IDs, PE converts the list into a list of gene IDs using the Onto-Tools database. The Onto-Tools database integrates a number of public databases including GenBank, dbEST, UniGene, Entrez Gene, RefSeq, KEGG, etc. After creating a list of gene IDs, PE searches the KEGG pathways in the Onto-Tools database for each input gene, and builds a list of pathways containing at least one input gene. Note that the pathways returned by PE depend on the annotations available in KEGG. If a gene is known to be involved in a pathway, but is not annotated as such in KEGG, PE does not return the pathway in its output.

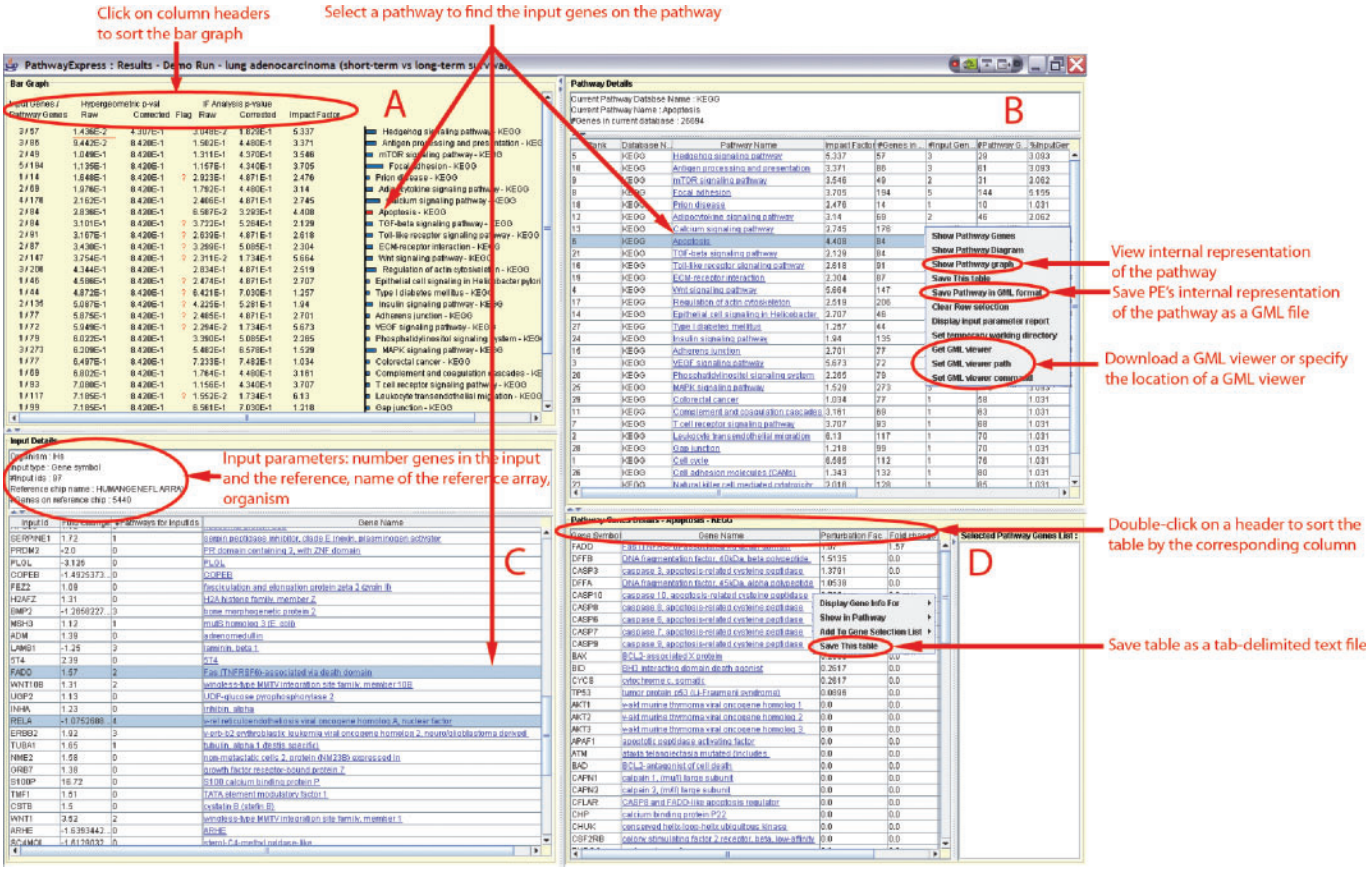
The output of PE is shown in Figure 4. The top left panel displays detailed results for each pathway including: number of input genes and total number of genes on a pathway, probability of obtaining the same number of genes on a given pathway by random chance, impact factor and probability of obtaining the impact factor by random chance for a given pathway, etc. The bar graph in panel A can be sorted in increasing or decreasing order of any column by clicking on the corresponding column header. The bar graph in panel A, pathway details in panel B and input details in panel C are synchronized with each other. For instance, in Figure 4, selecting the apoptosis pathway in panel B, highlights the corresponding bar in red color in panel A and also selects the input genes in panel C (i.e. genes FADD and RELA).

Right-clicking a mouse in the PE output window brings up a context-sensitive popup menu (Figure 4). For instance, the menu displayed by clicking a mouse on a pathway name allows the user to view the corresponding KEGG pathway diagram, download and specify a GML viewer to use with PE, etc. The KEGG pathway diagram that corresponds to the selected pathway can also be viewed by double-clicking the pathway name in either panel A or panel B. PE highlights the input genes in red (up-regulated genes) or blue (down-regulated) in the KEGG diagram (Figure 5). The popup menu also allows to save a pathway in GML format that can be viewed in any GML viewer. This can be done by selecting 'Save in GML format' from the popup menu. If a program able to read GML files (e.g. yEd (www.yed.com) or Cytoscape (21)) is already installed on the local machine, the user can specify its location by selecting 'Set GML viewer command' from the popup menu. After specifying the GML viewer, a user can select 'Show pathway graph' to access GML representation of the pathway from within PE. All tables in the PE output can be saved as a tab-delimited text file by selecting 'Save this table' from the popup menu.

Figure 5 also shows the internal representation of the apoptosis pathway (read from the GML file) with the two input genes represented by elliptic nodes. PE's internal representation also shows how perturbation introduced by the genes FADD and RELA is propagated throughout the pathway in the given condition. The perturbed genes on the apoptosis pathway are shown with colored background, whereas the unperturbed genes are shown with white background, and the direction of the propagation is indicated by an arrow between two genes. In Figure 5, notice that only the area of the pathway downstream of the input genes is perturbed, while the rest of the pathway is unperturbed.

ENHANCEMENTS

Over the past year, Onto-Miner (OM) has been reimplemented as a Java tool to make its interface user friendly and consistent with the Onto-Tools suite. The previous HTML input interface of OM is replaced by a new Java interface that is easier to use (Figure 6). Unlike previous version of OM that required the user to


Click on column headers
to sort the bar graph
Select a pathway to find the input genes on the pathway
A
B
C
D
View internal representation
of the pathway
Save PE's internal representation
of the pathway as a GML file
Download a GML viewer or specify
the location of a GML viewer
Input parameters: number genes in the input
and the reference, name of the reference array,
organism
Double-click on a header to sort the
table by the corresponding column
Save table as a tab-delimited text file
Show Pathway Genes
Show Pathway Diagram
Show Pathway graph
Save This table
Save Pathway in GML format
Clear Row selection
Display input parameter report
Set temporary working directory
Get GML viewer
Set GML viewer path
Set GML viewer command
Display Gene Info For
Show in Pathway
Add To Gene Selection List
Save This table

PathwayExpress : Results - Demo Run - lung adenocarcinoma (short-term vs long-term survival)

Bar Graph

Input Genes /	Hypergeometric p-val			IF Analysis p-value			
Pathway Genes	Raw	Corrected	Flag	Raw	Corrected	Impact Factor	
3/57	1.436E-2	4.307E-1		3.048E-2	1.829E-1	5.337	Hedgehog signaling pathway - KEGG
3/86	9.442E-2	8.420E-1		1.502E-1	4.480E-1	3.371	Antigen processing and presentation - KEGG
2/49	1.049E-1	8.420E-1		1.311E-1	4.370E-1	3.546	mTOR signaling pathway - KEGG
5/184	1.135E-1	8.420E-1		1.157E-1	4.340E-1	3.705	Focal adhesion - KEGG
1/14	1.648E-1	8.420E-1	?	2.923E-1	4.871E-1	2.476	Prion disease - KEGG
2/69	1.976E-1	8.420E-1		1.792E-1	4.480E-1	3.14	Adipocytokine signaling pathway - KEGG
4/178	2.162E-1	8.420E-1		2.406E-1	4.871E-1	2.745	Calcium signaling pathway - KEGG
2/84	2.836E-1	8.420E-1		6.587E-2	3.293E-1	4.408	Apoptosis - KEGG
2/84	3.101E-1	8.420E-1	?	3.722E-1	5.264E-1	2.129	TGF-beta signaling pathway - KEGG
2/91	3.167E-1	8.420E-1	?	2.639E-1	4.871E-1	2.618	Toll-like receptor signaling pathway - KEGG
2/87	3.430E-1	8.420E-1	?	3.299E-1	5.085E-1	2.304	ECM-receptor interaction - KEGG
2/147	3.754E-1	8.420E-1	?	2.311E-2	1.734E-1	5.664	Wnt signaling pathway - KEGG
3/208	4.344E-1	8.420E-1		2.834E-1	4.871E-1	2.519	Regulation of actin cytoskeleton - KEGG
1/46	4.586E-1	8.420E-1	?	2.474E-1	4.871E-1	2.707	Epithelial cell signaling in Helicobacter pylori
1/44	4.872E-1	8.420E-1	?	6.421E-1	7.030E-1	1.257	Type I diabetes mellitus - KEGG
2/135	5.087E-1	8.420E-1	?	4.225E-1	5.281E-1	1.94	Insulin signaling pathway - KEGG
1/77	5.875E-1	8.420E-1	?	2.485E-1	4.871E-1	2.701	Adherens junction - KEGG
1/72	5.949E-1	8.420E-1	?	2.294E-2	1.734E-1	5.673	VEGF signaling pathway - KEGG
1/79	6.022E-1	8.420E-1		3.390E-1	5.085E-1	2.265	Phosphatidylinositol signaling system - KEGG
3/273	6.209E-1	8.420E-1		5.482E-1	6.578E-1	1.529	MAPK signaling pathway - KEGG
1/77	6.497E-1	8.420E-1		7.233E-1	7.482E-1	1.034	Colorectal cancer - KEGG
1/69	6.802E-1	8.420E-1		1.764E-1	4.480E-1	3.161	Complement and coagulation cascades - KE
1/83	7.080E-1	8.420E-1		1.156E-1	4.340E-1	3.707	T cell receptor signaling pathway - KEGG
1/117	7.185E-1	8.420E-1	?	1.552E-2	1.734E-1	6.13	Leukocyte transendothelial migration - KEGG
1/99	7.185E-1	8.420E-1		6.561E-1	7.030E-1	1.218	Gap junction - KEGG

Input Details

Organism : Hs
Input type : Gene symbol
#Input ids : 97
Reference chip name : HUMANGENEFL ARRAY
#Genes on reference chip : 5440

Input Id	Fold change	#Pathways for InputIds	Gene Name
SERPINE1	1.72	1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator
PRDM2	-2.0	0	PR domain containing 2, with ZNF domain
PLOL	-3.126	0	PLOL
COPEB	-1.4925373...	0	COPEB
FEZ2	1.08	0	fasciculation and elongation protein zeta 2 (zygin II)
H2AFZ	1.31	0	H2A histone family, member Z
BMP2	-1.2658227...	3	bone morphogenetic protein 2
MSH3	1.12	1	mutS homolog 3 (E. coli)
ADM	1.39	0	adrenomedullin
LAMB1	-1.25	3	laminin, beta 1
5T4	2.39	0	5T4
FADD	1.57	2	Fas (TNFRSF6)-associated via death domain
WNT10B	1.31	2	wingless-type MMTV integration site family, member 10B
UGP2	1.13	0	UDP-glucose pyrophosphorylase 2
INHA	1.23	0	inhibin, alpha
RELA	-1.0752688...	4	v-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor
ERBB2	1.92	3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived
TUBA1	1.65	1	tubulin, alpha 1 (testis specific)
NME2	1.58	0	non-metastatic cells 2, protein (NM23B) expressed in
GRB7	1.38	0	growth factor receptor-bound protein 7
S100P	16.72	0	S100 calcium binding protein P
TMF1	1.51	0	TATA element modulatory factor 1
CSTB	1.5	0	cystatin B (stefin B)
WNT1	3.62	2	wingless-type MMTV integration site family, member 1
ARHE	-1.6393442...	0	ARHE
SC4MOL	-1.6129032...	0	sterol-C4-methyl oxidase-like

Pathway Details

Current Pathway Databse Name : KEGG
Current Pathway Name : Apoptosis
#Genes in current database : 26694

Rank	Database N...	Pathway Name	Impact Factor	#Genes in ...	#Input Gen...	#Pathway G...	%InputGen
5	KEGG	Hedgehog signaling pathway	5.337	57	3	29	3.093
10	KEGG	Antigen processing and presentation	3.371	86	3	61	3.093
9	KEGG	mTOR signaling pathway	3.546	49	2	31	2.062
8	KEGG	Focal adhesion	3.705	194	5	144	5.155
18	KEGG	Prion disease	2.476	14	1	10	1.031
12	KEGG	Adipocytokine signaling pathway	3.14	69	2	46	2.062
13	KEGG	Calcium signaling pathway	2.745	178			
6	KEGG	Apoptosis	4.408	84			
21	KEGG	TGF-beta signaling pathway	2.129	84			
16	KEGG	Toll-like receptor signaling pathway	2.618	91			
19	KEGG	ECM-receptor interaction	2.304	87			
4	KEGG	Wnt signaling pathway	5.664	147			
17	KEGG	Regulation of actin cytoskeleton	2.519	206			
14	KEGG	Epithelial cell signaling in Helicobacter	2.707	46			
27	KEGG	Type I diabetes mellitus	1.257	44			
24	KEGG	Insulin signaling pathway	1.94	135			
15	KEGG	Adherens junction	2.701	77			
3	KEGG	VEGF signaling pathway	5.673	72			
20	KEGG	Phosphatidylinositol signaling system	2.265	79			
25	KEGG	MAPK signaling pathway	1.529	273	3		3.093
29	KEGG	Colorectal cancer	1.034	77	1	58	1.031
11	KEGG	Complement and coagulation cascades	3.161	69	1	63	1.031
7	KEGG	T cell receptor signaling pathway	3.707	93	1	68	1.031
2	KEGG	Leukocyte transendothelial migration	6.13	117	1	70	1.031
28	KEGG	Gap junction	1.218	99	1	70	1.031
1	KEGG	Cell cycle	6.585	112	1	76	1.031
26	KEGG	Cell adhesion molecules (CAMs)	1.343	132	1	80	1.031
22	KEGG	Natural killer cell mediated cytotoxicity	2.018	128	1	85	1.031

Pathway Genes Details - Apoptosis - KEGG

Gene Symbol	Gene Name	Perturbation Fac	Fold change
FADD	Fas (TNFRSF6)-associated via death domain	1.57	1.57
DFFB	DNA fragmentation factor, 40kDa, beta polypeptide	1.5135	0.0
CASP3	caspase 3, apoptosis-related cysteine peptidase	1.3791	0.0
DFFA	DNA fragmentation factor, 45kDa, alpha polypeptide	1.0538	0.0
CASP10	caspase 10, apoptosis-related cysteine peptidase		
CASP8	caspase 8, apoptosis-related cysteine peptidase		
CASP6	caspase 6, apoptosis-related cysteine peptidase		
CASP7	caspase 7, apoptosis-related cysteine peptidase		
CASP9	caspase 9, apoptosis-related cysteine peptidase		
BAX	BCL2-associated X protein		0.0
BID	BH3 interacting domain death agonist	0.2617	0.0
CYCS	cytochrome c, somatic	0.2617	0.0
TP53	tumor protein p53 (Li-Fraumeni syndrome)	0.0898	0.0
AKT1	v-akt murine thymoma viral oncogene homolog 1	0.0	0.0
AKT2	v-akt murine thymoma viral oncogene homolog 2	0.0	0.0
AKT3	v-akt murine thymoma viral oncogene homolog 3	0.0	0.0
APAF1	apoptotic peptidase activating factor	0.0	0.0
ATM	ataxia telangiectasia mutated (includes	0.0	0.0
BAD	BCL2-antagonist of cell death	0.0	0.0
CAPN1	calpain 1, (mu/I) large subunit	0.0	0.0
CAPN2	calpain 2, (m/II) large subunit	0.0	0.0
CFLAR	CASP8 and FADD-like apoptosis regulator	0.0	0.0
CHP	calcium binding protein P22	0.0	0.0
CHUK	conserved helix-loop-helix ubiquitous kinase	0.0	0.0
CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity	0.0	0.0

Selected Pathway Genes List :

Figure 4. Pathway-Express output. PE output is organized in four panels. Panel **A** summarizes the results as a bar graph. Panel **B** provides details about each pathway including: (i) number of genes on the pathway, (ii) number of genes found in the reference array that are on the pathway, (iii) number of input genes found on the pathway, (iv) its impact factor, (v) database identifier of the pathway, etc. Panel **C** provides input details including the input ID, its corresponding gene name and symbol, number of pathways the gene is found on, and the fold change as provided by the user. Double-clicking a pathway name in panel B provides a list of genes on that pathway in panel D along with their perturbation factor.

manually download the results from the server, OM now automatically downloads the result file from the Onto-Tools server and saves it in the location specified by the user.

ACKNOWLEDGMENTS

This material is based upon work supported by the following grants: NSF DBI-0234806, CCF-0438970, 1R01HG003491-01A1, 1U01CA117478-01, 1R21CA100740-01, 1R01NS045207-01, 5R21EB000990-03 and 2P30 CA022453-24. Onto-Tools currently runs on equipment provided by Sun Microsystems EDU 7824-02344-U and by NIH(NCRR) 1S10 RR017857-01. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, NIH, DOD or any other of the funding agencies. Funding to pay the Open Access publication charges for this article was provided by NSF DBI-0234806.

Conflict of interest statement. None declared.

REFERENCES

1. Khatri,P., Draghici,S., Ostermeier,G.C. and Krawetz,S.A. (2002) Profiling gene expression using onto-express. *Genomics*, **79**, 266–270.
2. 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.
3. Draghici,S., Khatri,P., Martins,R.P., Ostermeier,G.C. and Krawetz,G.A. (2003) Global functional profiling of gene expression. *Genomics*, **81**, 98–104.
4. Khatri,P., Bhavsar,P., Bawa,G. and Draghici,S. (2004) Onto-tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res.*, **32**, W449–W456.
5. Khatri,P., Sellamuthu,S., Malhotra,P., Amin,K., Done,A. and Draghici,S. (2005) Recent additions and improvements to the onto-tools. *Nucleic Acids Res.*, **33**, W762–W765.
6. Draghici,S., Khatri,P., Shah,A. and Tainsky,M. (2003) Assessing the functional bias of commercial microarrays using the onto-compare database. *BioTechniques*, Microarrays and Cancer: Research and Applications, 55–61.
7. Khatri,P., Desai,V., Tarca,A.L., Sellamuthu,S., Wildman,D.E., Romero,R. and Draghici,S. (2006) New onto-tools: Promoter-express, nsSNPCounter and onto-translate. *Nucleic Acids Res.*, **34**, W626–W631.

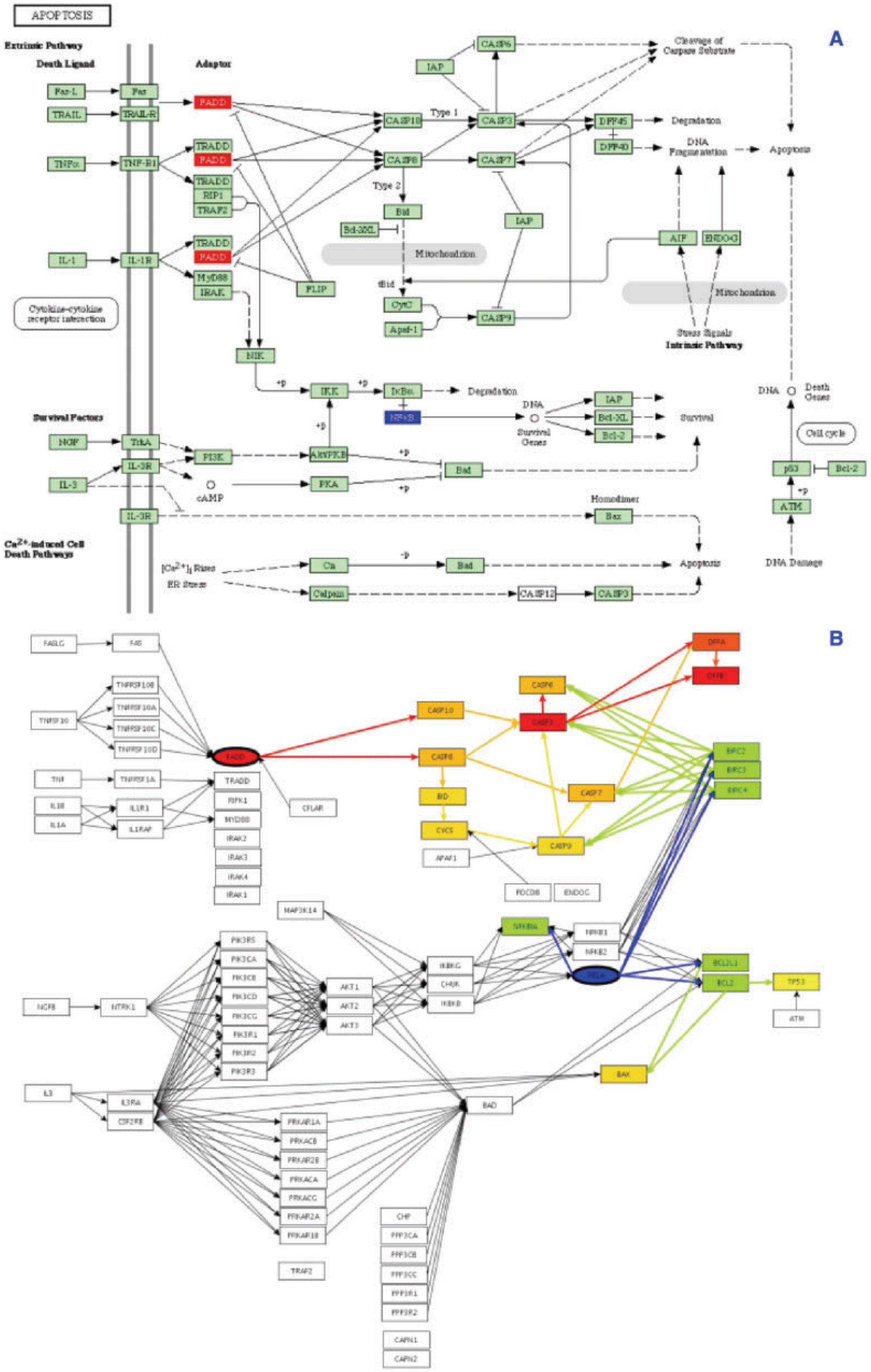
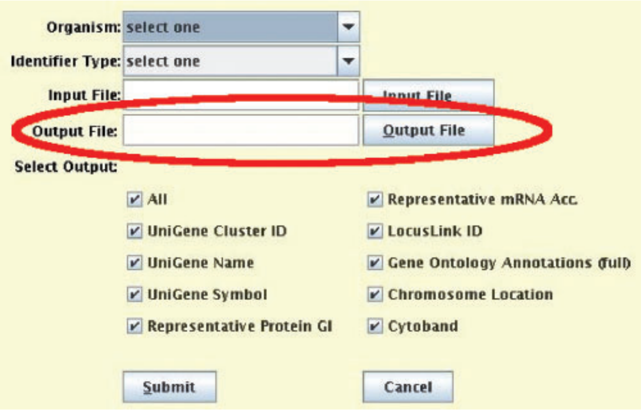


Figure 5. The apoptosis pathway as described by KEGG (panel **A**) and its internal representation, as constructed by Pathway-Express (panel **B**). PE can save this internal representation in a GML file that can be exported to any other visualization application. In this internal representation, each node corresponds to a gene, and each edge represents an interaction between two genes. In this example, the up-regulated input gene FADD is highlighted with red color and down-regulated gene RELA is highlighted with blue color in both panels. The internal graph also shows how PE propagates the perturbation through the pathway, where the direction of an arrow represents the direction of propagation. Note that in some cases, a gene in KEGG diagram may represent a group of genes. For instance, gene PI3K in the KEGG diagram in fact corresponds to a group of eight nodes in PE's internal representation: PIK3R5, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2 and PIK3R3.



Organism: select one

Identifier Type: select one

Input File:

Output File:

Select Output:

☑ All
☑ UniGene Cluster ID
☑ UniGene Name
☑ UniGene Symbol
☑ Representative Protein GI
☑ Representative mRNA Acc.
☑ LocusLink ID
☑ Gene Ontology Annotations (full)
☑ Chromosome Location
☑ Cytoband

Figure 6. New Onto-Miner input interface. OM automatically retrieves the results from the Onto-Tools server and saves them into the output file specified by the user.

8. Draghici,S., Sellamuthu,S. and Khatri,P. (2006) Babel's tower revisited: a universal resource for cross-referencing across annotation databases. *Bioinformatics*, **22**, 2934–2939.
9. Desai,V., Khatri,P., Done,A., Friedman,A., Tainsky,M. and Draghici,S. (2005) A novel bioinformatics technique for predicting condition-specific transcription factor binding sites. In *Proceedings of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*. San Diego, USA.
10. Khatri,P. and Draghici,S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**, 3587–3595.
11. Ogata,H., Goto,S., Sato,K., Fujibuchi,W., Bono,H. and Kanehisa,M. (1999) Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **27**, 29–34.
12. Joshi-Tope,G., Gillespie,M., Vasrik,I., D'Eustachio,P., Schmidt,E., de Bone,B., Jassal,B., Gopinath,G.R., Wu,G.R. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
13. 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.
14. Dahlquist,K.D., Salomonis,N., Vranizan,K., Lawlor,S.C. and Conklin,B.R. (2002) Genmapp, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **31**, 19–20.
15. Doniger,S.W., Salomonis,N., Dahlquist,K.D., Vranizan,K., Lawlor,S.C. and Conklin,B.R. (2003) Mappfinder: using gene ontology and genmapp to create a global gene expression profile from microarray data. *Genome biol.*, **4**, R7.
16. Grosu,P., Townsend,J.P., Hartl,D.L. and Cavalieri,D. (2002) Pathway processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res.*, **12**, 1121–1126.
17. Holford,M., Li,N., Nadkarni,P. and Zhao,H. (2004) Vitapad: visualization tools for the analysis of pathway data. *Bioinformatics*, **21**, 1596–1602.
18. Nikitin,A., Egorov,S., Daraselia,N. and Mazo,I. (2003) Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics*, **19**, 2155–2157.
19. Pan,D., Sun,N., Cheung,K.-H., Guan,Z., Ma,L., Holford,M., Deng,X. and Zhao,H. (2003) Pathmapa: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for arbidopsis. *BMC Bioinformatics*, **4**, 56.
20. Pandey,R., Guru,R.K. and Mount,D.W. (2004) Pathway miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics*, **20**, 2156–2158.
21. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
22. Hosack,D.A., Dennis,G.Jr, Sherman,B.T., Lane,H.C. and Lempicki,R.A. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol.*, **4**, P4.
23. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
24. Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D. *et al.* (2003) The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.