

CCancer: a bird's eye view on gene lists reported in cancer-related studies

Sabine Dietmann¹, Wanseon Lee², Philip Wong², Igor Rodchenkov³ and Alexey V. Antonov^{2,*}

¹Freiburg Center for Data Analysis and Modeling, University of Freiburg, Eckerstr. 1, D-79104 Freiburg,

²Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Institute for Bioinformatics and Systems Biology, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany and ³Faculty of Medicine, University of Toronto, BBDMR, CCBR, 160 College Street, Toronto, ON, Canada

Received January 30, 2010; Revised April 27, 2010; Accepted May 21, 2010

ABSTRACT

CCancer is an automatically collected database of gene lists, which were reported mostly by experimental studies in various biological and clinical contexts. At the moment, the database covers 3369 gene lists extracted from 2644 papers published in ~80 peer-reviewed journals. As input, CCancer accepts a gene list. An enrichment analyses is implemented to generate, as output, a highly informative survey over recently published studies that report gene lists, which significantly intersect with the query gene list. A report on gene pairs from the input list which were frequently reported together by other biological studies is also provided. CCancer is freely available at <http://mips.helmholtz-muenchen.de/proj/ccancer>.

INTRODUCTION

At the moment, various high-throughput experimental platforms are employed intensively to provide new insights into the molecular mechanisms underlying a variety of biological phenomena (1,2). An increasing number of biological or clinical studies report differentially expressed genes, epigenetically silenced genes, frequently mutated genes, genes with copy number variations or other gene lists involved in common biological processes. Although being publicly available, this type of information, at the same time, is dissolved in hundreds of papers. The only way to collect this valuable data is to use automatic text mining systems.

Text-mining systems are employed by biomedical researchers to automatically extract relevant information from the literature [see ref. (3) for a review]. For example, PolySearch (4) is a generic text mining system

for extracting relationships between genes and diseases. Several other databases, which are based on text mining, focus on specialized research areas: PubMeth (5) and MeInfoText (6) collect information on gene methylation in cancer. DDOC (7) and DDEC (8) collect heterogeneous information about genes differentially expressed in ovarian and esophageal cancer, such as manually curated information about the promoter regions and associated transcription factors, as well as text-mined reports.

Recently, we have developed the PLIPS database, a collection of protein lists extracted from proteomics studies by text-mining (9). PLIPS also provides a statistical framework for the interpretation of a protein list. To generate the PLIPS database, relatively few 'text mining' efforts were required, since a majority of proteomics studies are published in a few highly specialized proteomics journals. PLIPS covers only five major proteomics journals (Proteomics, Journal of Proteome Research, Molecular and Cellular Proteomics, Proteomics—Clinical Applications) and ~1000 different protein lists extracted from 800 independent studies.

In contrast to proteomics, high-throughput genomic technologies were more frequently used and their results are published in a much wider spectrum of journals. Gene lists, which were characterized to play key roles in molecular mechanisms for a variety of biological phenomena, are regularly reported in general biological journals, as well as in highly specific medical journals. Thus, automatic extraction of this information requires a lot of additional efforts.

Here, we present a database, termed CCancer, which provides a collection of 3369 gene lists automatically extracted from tables in 2644 studies covering ~80 peer-reviewed journals. Cancer is a major focus of biomedical research. According to our estimates, more than a half of the gene lists stored in CCancer are extracted

*To whom correspondence should be addressed. Tel: +49 (0) 89 3187 2788; Fax: +49 (0) 89 3187 3585; Email: a.antonov@helmholtz-muenchen.de

from cancer related studies. This fact pre-defines the name of the database.

CCancer is not only a database but a web-based analysis platform, which employs an enrichment analyses framework (10–14) to interpret a given user-defined gene list. As input, CCancer accepts gene/protein list. As output, a catalogue of previously published studies that report a table of genes/proteins, which significantly intersects with a query list, is provided. Thus, CCancer supports the interpretation of the functional context for an experimentally derived gene lists. To illustrate the valuable and often unprecedented information that the user can get by using the CCancer database, we present several examples of data analyses.

MATERIALS AND METHODS

Text mining

We collected all articles (~150 000) published in 80 peer-reviewed journals for the last 5–7 years. The articles were screened for tables which report gene identifiers. The search algorithm was implemented to recognize a table with gene/protein identifiers within the paper text. If the table reports at least 10 unique gene identifiers of the same type [i.e. ‘Entrez Gene IDs’, ‘Gene Symbols’, ‘RefSeq’, ‘UNIGENE’, ‘ENSEMBL’, ‘Affymetrix Probes’, ‘IPISYN (International Protein Identifier)’, ‘Uniprot, SwissProt’] then the paper was selected. In total 3369 gene/protein lists were identified from 2644 papers. All gene list were mapped to ‘Entrez Gene IDs’. The data in CCancer covers ~20 000 unique ‘Entrez Gene IDs’.

The top journals, in terms of the number of extracted gene lists, were highly specific journals in cancer research: ‘Cancer Research’ (327 papers and 411 gene lists), ‘Oncogene’ (214 papers and 278 gene lists), ‘Clinical Cancer Research’ (149 papers and 178 gene list), ‘International Journal of Cancer’ (109 papers and 143 gene lists). The full list of journals is accessible on the web server (<http://mips.helmholtz-muenchen.de/proj/ccancer/journals>).

We would like to point out that the data were collected automatically. A gene list in the database may be incomplete (in comparison to the originally reported list in the paper) and might have false positive gene identifiers. In our estimate based on 100 randomly selected records, ~60% of records are of high quality (the original table in the paper and the CCancer record have <10% of false negative and false positive genes), ~20% of records are of good quality (containing 10–25% false negatives and not more than 10% of false positives) and ~15% of records contains ~35–75% of genes actually reported in the paper table. About 3–5% of the records may represent artefacts, i.e. a result from a wrongly recognized table which does not actually reports a gene list.

Automatic annotation of gene lists with human disease terms

A comprehensive hierarchical controlled vocabulary for human disease (http://do-wiki.nubic.northwestern.edu/index.php/Main_Page) was used to link articles to

human disease terms. First we computed the background distribution for each disease term using all available articles (~150 000). For each disease term ‘A’ we select the subset of articles where the term was present at least once and compute for each article from this subset the number of times the term ‘A’ was present in the paper. The average number of times the term ‘A’ was mentioned per paper across this subset was computed. If a term ‘A’ was mentioned twice as many times as computed average value, then the paper was annotated with term ‘A’.

The gene lists from CCancer database were annotated with ‘human disease terms’ based on the annotation of the paper from which it was extracted. We would like to point out that the biological context of genes reported in the table may not correspond to the context of the terms overrepresented in the paper text. In each case, manual analysis is required.

Statistical analysis: intersection of gene lists

To statistically link a given gene list (the query list) to the lists from the database we implement standard enrichment analyses. For each gene/protein list L in the database, the number of genes I common between the query list and the gene list L are counted. The null hypothesis H_0 ‘Genes from the query list (size NQ) and from the list L (size N_L) have at least I common genes by chance’ is tested. The hypergeometric test, adjusted for multiple testing by a Monte–Carlo simulation procedure, is employed to assess the significance of the intersection I . The estimated P -value corresponds exactly to the definition of an experiment-wise Westfall and Young P -value (11,12,15–17).

Statistical analysis: significantly associated gene pairs

Based on the data from the CCancer database we identified gene pairs which were significantly associated (frequently reported together). Let us denote N as the total number of tables in the CCancer database. Let us denote C_j to be the number of tables from CCancer database where gene j was reported. For each pair of genes (j and k) we compute C_{jk} the number of tables where both gene j and k were reported. The number (intersection) C_{jk} follows a hypergeometric distribution with parameters N , C_j and C_k (C_j balls were drawn without replacement from an urn containing ‘ N ’ balls in total, ‘ C_k ’ of which are white). A Monte–Carlo simulation procedure was employed to adjust the P -value for multiple testing (for each gene we tested K hypotheses where K is a total number of genes). At the significance level ($P < 0.05$) each gene from the CCancer database was associated on average with ~20 other genes.

Cross-linking of gene lists from the database

Extracted gene/protein lists were mapped to NCBI Entrez Gene IDs. Each gene/protein list L in the database was considered as ‘query’ list to identify the other lists from the database which have significant number of common genes (with ‘query’ list). Thus, we cross-linked all gene list pairs if they share a significant number ($P < 0.01$) of common genes. This information can be browsed online

(<http://mips.helmholtz-muenchen.de/proj/ccancer/journals/>).

Linking gene lists to gene ontology terms

Each gene/protein list from CCancer database was linked to gene ontology (GO) terms which were significantly ($P < 0.01$) enriched in the list (13). In analogy to the calculation of the intersection P -value, a hypergeometric test, adjusted for multiple testing using a Monte–Carlo simulation approach, was employed to estimate the statistical significance of a GO category.

Monte–Carlo simulation to adjust P -values for multiple testing

All three considered cases (intersection between gene list and CCancer records, GO enrichment of gene list and significantly associated gene pairs) can be modelled by the same statistical model ‘sampling without replacement’. In this model, k balls were drawn without replacement from an urn containing ‘ N ’ balls in total, n of which are white (all others are black). In this case, the number k_1 of white balls drawn from the urn follows a hypergeometric distribution with parameters N , k and n . However, in our cases the balls are multicolored and we actually test multiple hypotheses at the same time: ‘white balls were drawn randomly’, ‘blue balls were drawn randomly’, ‘red balls were drawn randomly’ and so on. As we select the most enriched whatever color (let say red for this case), the estimated P -value based on the hypergeometric distribution does not reflect the actual probability to get k_1 of whatever color balls; it reflects the probability to get k_1 red balls. To get the actual probability to get k_1 whatever color balls we need to adjust P -value for multiple testing. One way to do this is to use Monte–Carlo simulation to directly measure this probability based on, let say 1000 simulations.

In this case we simulate a random drawing of k balls 1000 times and each time we estimate the P -value based on hypergeometric distribution for the best (whatever) color. Thus, we got a distribution of size 1000 of the best P -values for a random drawn of k balls and compare it to the P -value for the best (whatever) color balls related to our original drawn of k balls. The estimate of the adjusted P -value is given by the share of random simulations where the best P -value was equal or superior (less) than the P -value for the best (whatever) color balls related to our original drawing of k balls.

RESULTS

CCancer: interpretation of gene list

The user can query his/her list of gene/protein identifiers to find statistically significant links to previously published studies as well as to identify gene pairs from the submitted list which were frequently reported together ($P < 0.05$). As input, CCancer accepts several types of gene identifiers. CCancer supports most gene and protein identifiers such as ‘Gene Symbol’, ‘Entrez Gene Id’, ‘UniProt/Swiss-Prot’, ‘UniGene’, ‘Ensembl’, ‘RefSeq Protein ID’, ‘RefSeq

Transcript ID’ and ‘Affymetrix probe codes’. As output, a catalogue of previously published studies that report a table of genes that significantly intersect with a query list is provided.

After a list of potential studies is generated, one needs to check manually all interesting hits. First, a list of gene IDs common between query and database list need to be checked (a link ‘Mapping protocol’ is provided on the resulting page). Additionally, by looking into the corresponding study (a link is provided on the resulting page) one can understand better the functional context of the ‘database’ gene list. As been already mentioned in ‘text mining’ section, the database was collected automatically and, thus, some hits may represent artefacts.

CCancer: browsing for gene lists with common biological properties

CCancer also provides an interface to browse gene lists from the database with a common property. At the moment, the user can select gene lists which are statistically linked to either a particular GO biological process (http://mips.helmholtz-muenchen.de/proj/ccancer/go_bp), molecular function (http://mips.helmholtz-muenchen.de/proj/ccancer/go_mf) or cellular component (http://mips.helmholtz-muenchen.de/proj/ccancer/go_cc). The possibility to browse gene lists based on their local properties is going to be extended in the future.

CCancer: examples of possible applications

Next we present examples of analyses of experimental data by CCancer to illustrate the potential utility of our database. In principle, interpretation of gene list using CCancer is based on the widely accepted guilt-by-association principle: significant similarities between protein lists can be indicative of similarity in molecular mechanisms between corresponding phenomena. The next example aims to illustrate this application of CCancer

Example 1: Relationship between the senescence phenotype and cancer. A study by Young *et al.* (18) provided evidence that autophagy-related genes mediate the acquisition of the senescence phenotype. The authors studied 53 autophagy- and senescence-related genes, which were up- or down-regulated after Ras induction. A screen in CCancer for studies, which report gene sets that significantly intersect with the genes reported in ref. (18), identified several related papers ($P < 0.01$, Table 1). For example, a study is detected, where Staber *et al.* (19) report genes associated with recurrent acute myeloid leukemia after high-dose chemotherapy. The genes, which were differentially expressed in patients with acute myeloid leukemia prior to high-dose chemotherapy and after relapse, significantly intersect with the senescence-related genes reported by ref. (18).

Two other studies detected by CCancer covered related topics. A study (21) described genes related to macrophage activation and TH1 immune response, which were induced by low-dose radiation therapy in follicular lymphoma. A second study of ref. (22) identified genes differentially

Table 1. An example of a CCancer output: a query list of autophagy- and senescence-related genes

P-value	l_r	Size of query list	Size of database list	Mapping protocol	Study 'Table number', the number and type of the protein/gene recognized, the title of the article	Human disease ontology terms over-represented in the article text
0.01	8	54	27	»Map	Table 3, 25 UNIGENE_1 in epithelial and stromal Cathepsin K and CXCL14 expression in breast tumour progression	Carcinoma <i>in situ</i> , breast cancer, radial scar, breast tumor, ductal carcinoma, cancer, carcinoma, scar
0.01	8	54	33	»Map	Table 3, 33 REFSEQ_RNA_NAV in effects of imatinib on monocyte-derived dendritic cells are mediated by inhibition of nuclear factor- κ B and Akt signalling pathways	Chronic myeloid leukemia, myeloid leukemia, leukemia
0.01	5	54	30	»Map	Table 1, 32 synonyms in identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors	Tear, disease
0.01	5	54	45	»Map	Table 2, 40 synonyms in <i>in vivo</i> p53 response and immune reaction underlie highly effective low-dose radiotherapy in follicular lymphoma	Follicular lymphoma, disease, lymphoma
0.01	7	54	99	»Map	Table 1, 86 synonyms in gene expression analysis of prostate hyperplasia in mice overexpressing the prolactin gene specifically in the prostate	Benign prostatic hyperplasia, intraepithelial neoplasia, dna fragmentation, hyperplastic, hyperplasia, cancer
0.01	4	54	41	»Map	Table 1, 41 REFSEQ_RNA_NAV in cytoglobin, the newest member of the globin family, functions as a tumor suppressor gene	Breast cancer, cancer
0.015	4	54	44	»Map	Table 2, 44 REFSEQ_RNA_NAV in common alterations in gene expression and increased proliferation in recurrent acute myeloid leukemia	Acute myeloid leukemia, myeloid leukemia, leukemia, disease
0.015	5	54	75	»Map	Table 1, 72 synonyms in identification of epigenetically silenced genes in tumor endothelial cells	Cancer
0.015	5	54	75	»Map	Table 2, 60 synonyms in molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells <i>in vitro</i>	Nephroblastoma
0.015	6	54	114	»Map	Table 1, 95 synonyms in knowledge-based gene expression classification via matrix factorization	Disease
0.015	5	54	84	»Map	Table 3, 73 synonyms in genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma	Squamous cell carcinoma, neck cancer, oral cancer, squamous carcinoma, cancer, disease, carcinoma
0.015	4	54	74	»Map	Table 1, 74 REFSEQ_RNA_NAV in dlk1/FAI regulates the function of human bone marrow mesenchymal stem cells by modulating gene expression of pro-inflammatory cytokines and immune response-related factors	
0.015	4	54	77	»Map	Table 1, 67 SYNONYMS in Comparative analysis of genes regulated by PML/RAR (20) and PLZF/RAR (20) in response to retinoic acid using oligonucleotide arrays	Leukemia
0.015	4	54	78	»Map	Table 2, 67 UNIGENE_1 in genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma	Squamous cell carcinoma, neck cancer, oral cancer, squamous carcinoma, cancer, disease, carcinoma
0.015	4	54	85	»Map	Table 1, 76 synonyms in genomic and proteomic analysis of the myeloid differentiation program	Leukemia
0.015	4	54	93	»Map	Table 3, 83 synonyms in differentiation of lung adenocarcinoma, pleural mesothelioma, and nonmalignant pulmonary tissues using DNA methylation profiles	Lung adenocarcinoma, malignant mesothelioma, pleural mesothelioma, cancer, adenocarcinoma, adenocarcinomas, mesothelioma, confusion
0.015	6	54	227	»Map	Table 3, 190 synonyms in inflammatory manifestations of experimental lymphatic insufficiency	Retinoblastoma, disease, lymphedema

(continued)

Table 1. Continued

P-value	I_i	Size of query list	Size of database list	Mapping protocol	Study 'Table number', the number and type of the protein/gene recognized, the title of the article	Human disease ontology terms over-represented in the article text
0.015	4	54	106	»Map	Table 1, 92 synonyms in overexpression of the Axl tyrosine kinase receptor in cutaneous SCC-derived cell lines and tumors	Skin cancer, cancer, carcinoma
0.025	7	54	360	»Map	Table W2, 433 AFFY in chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance	Squamous cell carcinoma, neoplastic transformation, leukemia, cancer, lipoma
0.025	5	54	206	»Map	Table 2, 179 synonyms in sequential gene expression changes in cancer cell lines after treatment with the demethylation agent 5-Aza-2-deoxycytidine	Cancer, carcinoma, hepatoma
0.025	4	54	137	»Map	Table 1, 119 synonyms in transcriptional response of lymphoblastoid cells to ionizing radiation—genome research	DNA damage

I_i : number of genes in the intersection.

Table 2. An example of a CCancer output: a query list of autophagy- and senescence-related genes

Gene ID	Associated gene	The number of times reported together	Disease terms, associated with gene pair
CTSL1	CTSB	12 (78,49)	Leukemia (3) cancer (7) carcinoma (3)
CTSL1	CTSS	5 (49,29)	Cancer (3) carcinoma (2)
CTSL1	CTSD	9 (66,49)	Leukemia (4) cancer (3) disease (3) myeloid leukemia (2)
CTSL1	CTSA	5 (49,20)	Leukemia (2) cancer (3)

Gene pairs frequently reported together (a part of the table).

expressed in response to ionizing radiation in lymphoblastoid cells.

A relationship between cancer- and senescence-related genes or pathways would have been certainly expected. Interestingly, some topics, which emerged from the comparison of senescence-related genes to the studies in the CCancer database, were related to cancer therapy or prognosis. For instance, the studies commonly reported *Cathepsin B and D* (CTSB/D) and *Cathepsin L1* (CTSL1), which participate in protein degradation and turnover (23,24).

Another output from Ccancer (Table 2) reports pairs of genes which were frequently reported together. For example, the already mentioned genes *Cathepsin L1* (CTSL1) and *Cathepsin B and D* (CTSB/D) were reported together much more frequently than it would be expected by chance. CTSL1 and CTSB were reported together by 12 Ccancer records while each gene was reported 78 and 49 times, respectively ($P < 0.05$). Interesting, 7 out of 12 papers were related to 'cancer' and three to 'LEUKEMIA'.

Example 2: Searching for novel potential clinical applications of drugs. CCancer can further be exploited to identify hints for novel clinical applications of known drugs (25) or drugs under development in the case, where the list of drug molecular targets is known. For a variety of cell disorders (including most cancer subtypes) CCancer stores lists of genes identified to be at differential states (comparison between normal versus disordered cells). This type of information can be mined to identify new potential therapeutic implications. Significant number of common genes between drug targets and gene lists from the database related to some cell disorder can be indicative of probable usability of the drug for the corresponding disease.

For example, Bosutinib is a novel promiscuous kinase inhibitor. We extracted a reported list of direct interactors of Bosutinib identified by chemical proteomics (26). We used CCancer to identify previously reported gene lists that have a significantly share of common genes/proteins (<http://mips.helmholtz-muenchen.de/proj/ccancer/example.html>) and, thus, to identify potential physiological conditions, where an application of Bosutinib might be effective. Among already known applications, like different types of 'leukemia' (27,28), CCancer suggests other specific cancer types, like 'oral squamous cell carcinoma'(28).

DISCUSSION

Here, we have generated a comprehensive collection of gene list reported by the papers in 80 peer-reviewed journals. Tables in articles usually present, to some extent, pre-processed gene lists, which are selected for significance by experts. To our knowledge, no existing text-mining system provides a similarly accessible and comparable collection of experimentally derived gene lists for analysis.

The CCancer database provides a computational interface, which generates a highly informative survey over recently published cancer-related studies, which report similar and significantly intersecting gene lists. As we have demonstrated, by applying this automatic analysis, the user may obtain sometimes unexpected links to previous studies. It would be a tedious, if not impossible task for experimental researchers to gain these insights by a manual analysis of the literature. Articles, which contain significantly intersecting gene lists, are not necessarily listed as ‘related articles’ in PubMed. In addition, CCancer implements a robust statistical treatment of the intersection between a query and a database gene list, and provides a valid estimate of the *P*-values by a Monte-Carlo simulation procedure. The *P*-values actually reflect the probability of getting an intersection of the same size, in terms of the number of genes, for any random query gene list.

FUNDING

This work was supported by the Helmholtz Alliance on Systems Biology (project ‘CoReNe’). Funding for open access charge: Helmholtz Center Munich – German Research Center for Environmental Health (GmbH).

Conflict of interest statement. None declared.

REFERENCES

- Fernandes,T.G., Diogo,M.M., Clark,D.S., Dordick,J.S. and Cabral,J.M. (2009) High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. *Trends Biotechnol.*, **27**, 342–349.
- Powell,A.K., Zhi,Z.L. and Turnbull,J.E. (2009) Saccharide microarrays for high-throughput interrogation of glycan-protein binding interactions. *Methods Mol. Biol.*, **534**, 313–329.
- Krallinger,M., Valencia,A. and Hirschman,L. (2008) Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome Biol.*, **9**(Suppl. 2), S8.
- Cheng,D., Knox,C., Young,N., Stothard,P., Damaraju,S. and Wishart,D.S. (2008) PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Res.*, **36**, W399–W405.
- Ongenaert,M., Van,N.L., De,M.T., Menschaert,G., Bekaert,S. and Van,C.W. (2008) PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res.*, **36**, D842–D846.
- Fang,Y.C., Huang,H.C. and Juan,H.F. (2008) MeInfoText: associated gene methylation and cancer information from text mining. *BMC Bioinformatics*, **9**, 22.
- Kaur,M., Radovanovic,A., Essack,M., Schaefer,U., Maqungo,M., Kibler,T., Schmeier,S., Christoffels,A., Narasimhan,K., Choolani,M. *et al.* (2009) Database for exploration of functional context of genes implicated in ovarian cancer. *Nucleic Acids Res.*, **37**, D820–D823.
- Essack,M., Radovanovic,A., Schaefer,U., Schmeier,S., Seshadri,S.V., Christoffels,A., Kaur,M. and Bajic,V.B. (2009) DDEC: Dragon database of genes implicated in esophageal cancer 2. *BMC Cancer*, **9**, 219.
- Antonov,A.V., Dietmann,S., Wong,P., Igor,R. and Mewes,H.W. (2009) PLIPS, an automatically collected database of protein lists reported by proteomics studies. *J. Proteome Res.*, **8**, 1193–1197.
- Huang,d.W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
- Dietmann,S., Georgii,E., Antonov,A., Tsuda,K. and Mewes,H.W. (2009) The DICS repository: module-assisted analysis of disease-related gene lists. *Bioinformatics*, **25**, 830–831.
- Berriz,G.F., King,O.D., Bryant,B., Sander,C. and Roth,F.P. (2003) Characterizing gene sets with FuncAssociate. *Bioinformatics*, **19**, 2502–2504.
- Antonov,A.V., Schmidt,T., Wang,Y. and Mewes,H.W. (2008) ProfCom: a web tool for profiling the complex functionality of gene groups identified from high-throughput data. *Nucleic Acids Res.*, **36**, W347–W351.
- Antonov,A.V., Dietmann,S., Wong,P., Lutter,D. and Mewes,H.W. (2009) GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res.*, **37**, W323–W328.
- Westfall,P.N. and Young,S.S. (1993) *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons Inc, New York.
- Antonov,A.V., Dietmann,S. and Mewes,H.W. (2008) KEGG spider: interpretation of genomics data in the context of the global gene metabolic network. *Genome Biol.*, **9**, R179.
- Antonov,A.V., Dietmann,S., Rodchenkov,I. and Mewes,H.W. (2009) PPI spider: a tool for the interpretation of proteomics data in the context of protein-protein interaction networks. *Proteomics*, **9**, 2740–2749.
- Young,A.R., Narita,M., Ferreira,M., Kirschner,K., Sadaie,M., Darot,J.F., Tavare,S., Arakawa,S., Shimizu,S., Watt,F.M. *et al.* (2009) Autophagy mediates the mitotic senescence transition. *Genes Dev.*, **23**, 798–803.
- Staber,P.B., Linkesch,W., Zauner,D., Beham-Schmid,C., Guelly,C., Schauer,S., Sill,H. and Hoefler,G. (2004) Common alterations in gene expression and increased proliferation in recurrent acute myeloid leukemia. *Oncogene*, **23**, 894–904.
- Gronborg,M., Bunkenborg,J., Kristiansen,T.Z., Jensen,O.N., Yeo,C.J., Hruban,R.H., Maitra,A., Goggins,M.G. and Pandey,A. (2004) Comprehensive proteomic analysis of human pancreatic juice. *J. Proteome Res.*, **3**, 1042–1055.
- Knoops,L., Haas,R., de Kemp,S., Majoor,D., Broeks,A., Eldering,E., de Boer,J.P., Verheij,M., van Ostrom,C., de Vries,A. *et al.* (2007) In vivo p53 response and immune reaction underlie highly effective low-dose radiotherapy in follicular lymphoma 1. *Blood*, **110**, 1116–1122.
- Jen,K.Y. and Cheung,V.G. (2005) Identification of novel p53 target genes in ionizing radiation response. *Cancer Res.*, **65**, 7666–7673.
- Benes,P., Vetvicka,V. and Fusek,M. (2008) Cathepsin D—many functions of one aspartic protease. *Crit Rev. Oncol. Hematol.*, **68**, 12–28.
- Cordes,C., Bartling,B., Simm,A., Afar,D., Lautenschlager,C., Hansen,G., Silber,R.E., Burdach,S. and Hofmann,H.S. (2009) Simultaneous expression of Cathepsins B and K in pulmonary adenocarcinomas and squamous cell carcinomas predicts poor recurrence-free and overall survival. *Lung Cancer*, **64**, 79–85.
- Antonov,A.V. (2010) Mining protein lists from proteomics studies: applications for drug discovery. *Expert Opin. Drug Discovery*, **5**, 322–331.
- Fernbach,N.V., Planyavsky,M., Muller,A., Breitwieser,F.P., Colinge,J., Rix,U. and Bennett,K.L. (2009) Acid elution and one-dimensional shotgun analysis on an Orbitrap mass spectrometer: an application to drug affinity chromatography. *J. Proteome Res.*, **8**, 4753–4765.
- Walters,D.K., Goss,V.L., Stoffregen,E.P., Gu,T.L., Lee,K., Nardone,J., McGreevey,L., Heinrich,M.C., Deininger,M.W., Polakiewicz,R. *et al.* (2006) Phosphoproteomic analysis of AML cell lines identifies leukemic oncogenes. *Leuk. Res.*, **30**, 1097–2004.
- Sticht,C., Freier,K., Knopfle,K., Flechtenmacher,C., Pungs,S., Hofele,C., Hahn,M., Joos,S. and Lichter,P. (2008) Activation of MAP kinase signaling through ERK5 but not ERK1 expression is associated with lymph node metastases in oral squamous cell carcinoma (OSCC). *Neoplasia*, **10**, 462–470.