

Chris Bauer<sup>1</sup>, Karol Stec<sup>1</sup>, Alexander Glintschert<sup>1</sup>, Kristina Gruden<sup>2</sup>, Christian Schichor<sup>3</sup>, Michal Or-Guil<sup>4,5</sup>, Joachim Selbig<sup>6</sup> and Johannes Schuchhardt<sup>1</sup>

<sup>1</sup>Research and Development, MicroDiscovery GmbH, Berlin, Germany. <sup>2</sup>Department for Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia. <sup>3</sup>Department of Neurosurgery, Klinikum der Ludwig-Maximilian-Universität München, Munich, Germany. <sup>4</sup>Systems Immunology Lab, Department of Biology, Humboldt University, Berlin, Germany. <sup>5</sup>Research Center ImmunoSciences, Charité University of Medicine Berlin, Berlin, Germany. <sup>6</sup>Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany.

**ABSTRACT:** Personalized medicine is promising a revolution for medicine and human biology in the 21st century. The scientific foundation for this revolution is accomplished by analyzing biological high-throughput data sets from genomics, transcriptomics, proteomics, and metabolomics. Currently, access to these data has been limited to either rather simple Web-based tools, which do not grant much insight or analysis by trained specialists, without firsthand involvement of the physician. Here, we present the novel Web-based tool “BioMiner,” which was developed within the scope of an international and interdisciplinary project (SYSTHER<sup>†</sup>) and gives access to a variety of high-throughput data sets. It provides the user with convenient tools to analyze complex cross-omics data sets and grants enhanced visualization abilities. BioMiner incorporates transcriptomic and cross-omics high-throughput data sets, with a focus on cancer. A public instance of BioMiner along with the database is available at <http://systherDB.microdiscovery.de/>, login and password: “systher”; a tutorial detailing the usage of BioMiner can be found in the Supplementary File.

**KEYWORDS:** data mining, multiomics data integration, biomarker detection, pathway visualization, personalized medicine, cancer

**CITATION:** Bauer et al. BioMiner: Paving the Way for Personalized Medicine. *Cancer Informatics* 2015;14:55–63 doi: 10.4137/CIN.S20910.

**RECEIVED:** October 09, 2014. **RESUBMITTED:** December 17, 2014. **ACCEPTED FOR PUBLICATION:** December 25, 2014.

**ACADEMIC EDITOR:** J.T. Efrid, Editor in Chief

**TYPE:** Technical Advance

**FUNDING:** This work was supported by the grants INREMOS/SYSTHER and ColoNET from the German Federal Ministry of Education and Research (BMBF) (0315000A and 0315417H), and by the Slovenian Ministry of Health. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** MicroDiscovery GmbH has a trademark on the name “BioMiner” since 2001 until at least 2021. This information can be accessed from the German patent office and at <https://register.dpma.de/DPMARegister/marke/register/301488312/DE?lang=en>. BioMiner has no connection to other tools with the same or a similar name. J. Schuchhardt discloses financial support for the company MicroDiscovery GmbH from the University of Potsdam, during the conduct of this

study. The financial support originated from a BMBF grant. Other authors disclose no competing interests.

**CORRESPONDENCE:** johannes.schuchhardt@microdiscovery.com

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review by minimum of two reviewers. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

## Introduction

On the way to the medicine of the 21st century, personalized (also called individualized) medicine is a great and crucial step. It allows physicians to tailor therapy plans to the patient’s distinctiveness, providing better treatment outcomes for patients and overall less drug use. This equates to faster recovery, less side effects, and lower expenses for the health care system.<sup>1</sup>

Personalized medicine is especially promising with regard to the treatment of neoplastic disorders. They are among the most complex diseases known. Individualized medicine helps in developing pharmaceuticals and treatments that can efficiently eliminate cancer cells and at the same time cause the least amount of damage to the surrounding tissues and the patient overall. For personalized medicine and the fight against cancer, disease-centered high-throughput data sets are ideal resources. They can be used to generate or verify hypotheses, to compare own findings with public data, and to find novel biomarkers and identify their functional contexts. To achieve this,

a complex data processing pipeline is required, comprising (at least) the following steps<sup>2,3</sup>:

1. Searching and downloading a suitable public data set
2. Importing data into a statistical environment
3. Performing adequate data normalization
4. Performing quality control of data and normalization
5. Formulating and implementing a statistical analysis and
6. Interpreting the results

The realization of this workflow is very time consuming and requires special know-how in bioinformatics and statistics.<sup>4</sup> Physicians and biologists not experienced in high-throughput data analysis can profit immensely from a tool that allows easy and intuitive access to cross-omics high-throughput data sets, granting them firsthand access to this vast resource of knowledge.

The majority of biological high-throughput data sets are stored in public data repositories such as Gene Expression Omnibus (GEO)<sup>5</sup> and ArrayExpress<sup>6</sup> for microarray data, PRoteomics IDentifications (PRIDE)<sup>7</sup> for proteomics data, or Sequence Read Archive (SRA)<sup>8</sup> of the National Center for

<sup>†</sup>Systems Biology Tools Development For Cell Therapy And Drug Development.



Biotechnology Information for next-generation sequencing data. These repositories provide data storage for a dedicated set of single “omic” data type but do not support linking of different types of omics data (cross-omics). GEO and PRIDE also provide mining tools to inspect their data sets and to perform some limited data analysis.<sup>9,10</sup> Workflow management systems such as KNIME,<sup>11</sup> TAVERNA,<sup>12</sup> Orange4WS,<sup>13</sup> and Galaxy<sup>14</sup> use graphical user interfaces to build complex multistep data-mining workflows. These tools are designed to grant some limited analysis capabilities but do not provide unified (cross-omics) access to data sets. Performing data analysis based on cross-omics data sets proves to be a challenging task, even for those experienced with data analysis. Table 1 presents a comparison of BioMiner with other systems.

In a joint effort, the partners of the binational SYSTHER (Systems Biology Tools Development for Cell Therapy and Drug Development – [www.systher.eu](http://www.systher.eu)) project successfully investigated the molecular mechanisms of different cancer types. To analyze cancer from multiple points of view, complementary high-throughput data sets from transcriptomics, metabolomics, and proteomics were collected.<sup>15–18</sup> To store these different “omics” data sets together with complementary data from public resources, we created SystherDB, resorting to previous work.<sup>19,20</sup>

A database can only be beneficial to the user if the stored data are well presented and easily accessible. The complex nature of biological high-throughput data makes this a challenging task. Rising to this challenge, we developed BioMiner, a Web-based toolbox for mining and visualizing data stored in SystherDB. The main objective is to give all partners in the SYSTHER project a tool that they can use to obtain their own results from high-throughput data. Regardless of their expertise, the partners can use BioMiner to interactively explore public and proprietary multiomics data sets.

A public instance of BioMiner is freely available online. It currently contains 18 different studies (Table 2), with almost 4,000 microarrays and more than 187 Mio measured values of genes, proteins, or metabolites. Because BioMiner was developed in the SYSTHER project, most of the studies are related to the focus of the project, glioblastoma multiforme (GBM). Most of them have been imported from different public resources. The large studies in particular highlight the potential of BioMiner to easily inspect and analyze huge data sets. We demonstrate the application by presenting typical use cases for the identification of cancer biomarkers. A graphical representation of the workflow can be found in Figure 1.

All analyses presented in the methods section can readily be reproduced and expanded using the public Web server <http://systherDB.microdiscovery.de/>, with login and password “systher.”

## Methods

BioMiner is designed as a Web application and uses Google Web Toolkit (GWT) for the graphical user interface. GWT functionality is enhanced by customized plot-applets tailored

**Table 1.** Comparison of features between BioMiner and equivalent tools.

FEATURES	BioMiner	GEO	KNIME	GALAXY	R
Interactive Plotting	✓	×	×	×	×
Easy access to omics data	✓	×	×	×	×
Contains data-base with relevant information	✓	×	×	×	×
Shared data / remote access	✓	✓	×	×	×
Import own/public data	×	✓	✓	✓	✓
Programmable	×	×	✓	✓	✓
ANALYTICAL CAPABILITIES	BioMiner	GEO	KNIME	GALAXY	R
Differential analysis	✓	✓	✓	✓	✓
Correlation analysis	✓	×	✓	✓	✓
Cross-omics mapping	✓	×	×	×	×
Enrichment analysis	✓	×	✓	✓	✓
ANOVA, advanced modelling	×	×	✓	✓	✓
Clustering	×	×	✓	✓	✓

**Notes:** A “✓” indicates that the tool supports the feature. A “×” indicates that the feature is missing in standard installations. Please note that in KNIME, Galaxy, and R, advanced users may extend standard capabilities.

for interactive visualization of large data sets. Results are presented in two parallel views composed of a table and a plot. Both views are interactive and user-defined selections can be synchronized. Pathway visualization is achieved by extending the PathVisio library.<sup>21</sup>

Experimental data from genomics, proteomics and metabolomics are stored in a large manually curated MySQL<sup>‡</sup> database (SystherDB). To ensure data consistency, data import has to be performed by a dedicated specialist. Specific indexing methods have been implemented to achieve good response times. Even when querying large studies comprising several millions of measurements, response times are typically within just a few seconds. Metabolite data are annotated using three different identifier systems: Golm Metabolome Database,<sup>22</sup> Human Metabolome Database (HMDB),<sup>23</sup> and Kyoto Encyclopedia of Genes and Genomes (KEGG).<sup>24</sup> Integrating the different data types into a single outcome requires a predefined cross-omics relationship, eg, a mapping of metabolites onto genes or vice versa. Integration of semantic information is achieved by singular enrichment analysis.<sup>25</sup> Especially, pathway and functional information from Reactome,<sup>26</sup> KEGG,<sup>24</sup> Wiki-Pathways,<sup>27</sup> and GeneOntology<sup>28</sup> are supported. Enrichment

<sup>‡</sup><http://www.mysql.com>

**Table 2.** Studies in SystherDB: overview of the studies currently available in the SystherDB.

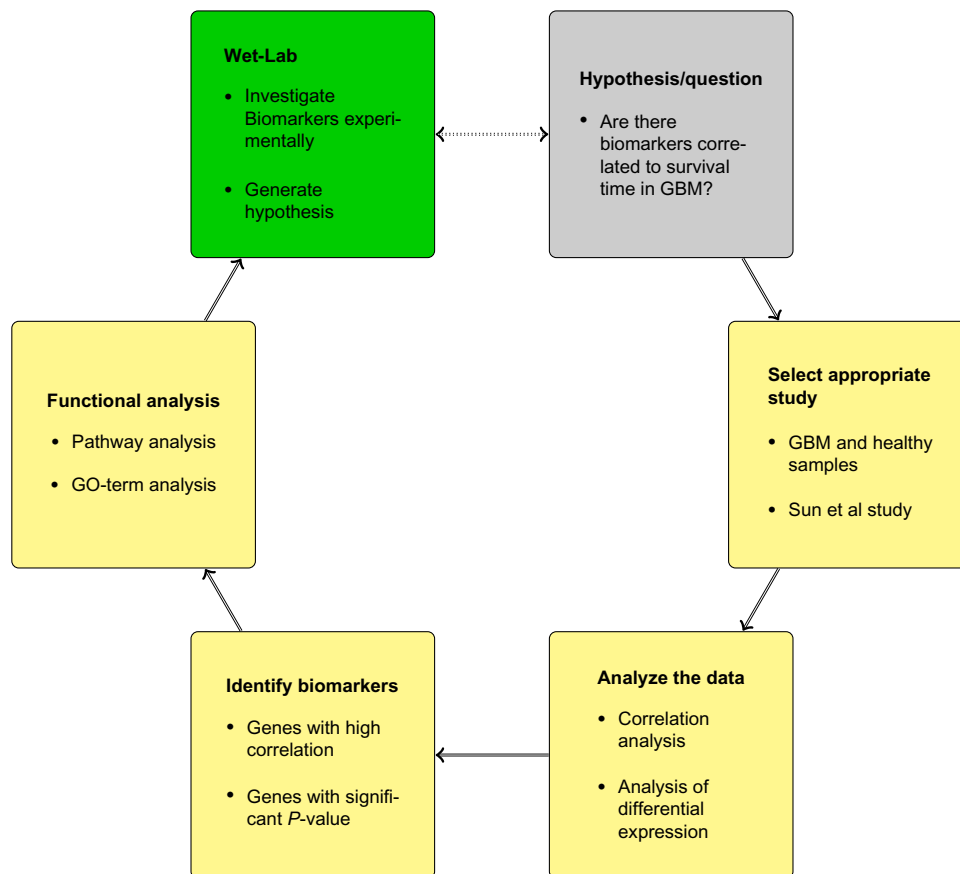
TITLE	TYPE	FACTORS	GEO-ID	SAMPLES	PM-ID	BIOMOLECULES
Glioma-derived stem cell factor effect on angiogenesis in the brain	Glioma	Cell Type; Tumor Grade	GSE4290	180	16616334 <sup>36</sup>	Genes
High-grade gliomas (HG-U133B)	Glioma	Cell Type; Tumor Grade; Necrosis; Survival Time	GSE4271	100	16530701 <sup>46</sup>	Genes
Gliomas of grades III and IV	Glioma	Cell Type; Tumor Grade	GSE4412	85	15374961 <sup>41</sup>	Genes
Expression profiles of human glioblastoma frozen tumors and cell lines	Glioma	Cell Type; Cell Number	GSE9171	30	18394558 <sup>47</sup>	Genes
Glioblastoma from a homogenous cohort of patients treated within clinical trial	Glioma	Disease State; Survival Time	GSE7696	84	18565887	Genes
Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development	Glioma	Cell Type	GSE9171	30	18394558	Genes
Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age (HG-U133A)	Glioma	Tumor Type; Survival Time	GSE13041	191	18940004	Genes
Transcriptome profile of human colorectal adenomas	CRC	Disease State	GSE8671	64	18171984 <sup>43</sup>	Genes
Clinical significance of osteoprotegerin expression in human colorectal cancer	CRC	Disease State; Method	GSE21510	148	21270110 <sup>48</sup>	Genes
Human colorectal cancer cell lines treated with several inhibitors of PI3Kinase–AKT signaling pathway	CRC	Cell Type; Treatment	GSE18005	15	20546605	Genes
RAS signaling in colon carcinoma: target gene deregulation and growth control through Y-box-binding protein 1	CRC	Cell Type; Treatment	GSE18232	18	21170361	Genes
Expression data from 290 primary colorectal cancers	CRC	Tumor Grade; Survival Time	GSE14333	226	19996206	Genes
NCI60 expression profiling using the Agilent Whole Human Genome Oligo Microarray	Cancer	Cell Line; Cell Name	GSE22821	249		Genes, Proteins, Metabolites
Expression data from the Cancer Cell Line Encyclopedia (CCLE)	Cancer	Tumor Location; Tumor Histology	GSE36133	917	22460905	Genes
Gene expression profile of peripheral blood lymphocytes: comparison between melanoma patients and healthy controls	Melanoma	Disease; Cell Type	GSE6887	46	17488182 <sup>49</sup>	Genes
Global control of cell cycle transcription by coupled CDK and network oscillators	Cell Cycle	Group; Time	GSE8799	60	18463633 <sup>50</sup>	Genes
Human body index – transcriptional profiling		Tissue; Disease; State	GSE7307	677		Genes
A genomic storm in critically injured humans		Age; Sex	GSE36809	812	22110166	Genes

statistics with multiple testing-corrected  $P$ -values are calculated as described in Bluthgen et al.<sup>29</sup> Correlation analyses are based on Pearson correlation coefficients. To filter noise and improve response times, correlations are calculated for high-variance genes (by default, the top 500 genes).

BioMiner complies with public data management standards such as Minimum Information About a Microarray Experiment (MIAME),<sup>30</sup> Minimum Information About a Proteomics Experiment (MIAPE),<sup>31</sup> and Minimum Information About a Metabolomics Experiment (MIAMET).<sup>32</sup>

**Cross-omics mapping.** For the mapping between genes and proteins, we use the relations provided by ENSEMBL<sup>#33</sup> database accessible via BioMart.<sup>34</sup> This mapping is updated on a regular basis (coupled with new versions of ENSEMBL). For establishing relations between genes and metabolites, we use the combined information of ConsensusPathDB<sup>35</sup> and HMDB.<sup>23</sup> These databases provide information about metabolites (compounds) and associated metabolic enzymes.

<sup>#</sup>www.ensembl.org



**Figure 1.** Typical workflow in which BioMiner can be integrated. The yellow boxes represent steps within the field of bioinformatics (or in this case, done in BioMiner), the green box represents the wet-lab work, and the gray box represents the theoretical part (eg, study design) of the workflow.

Genes coding for the metabolic enzymes are mapped to the corresponding metabolite. This mapping is updated with new versions of HMDB.

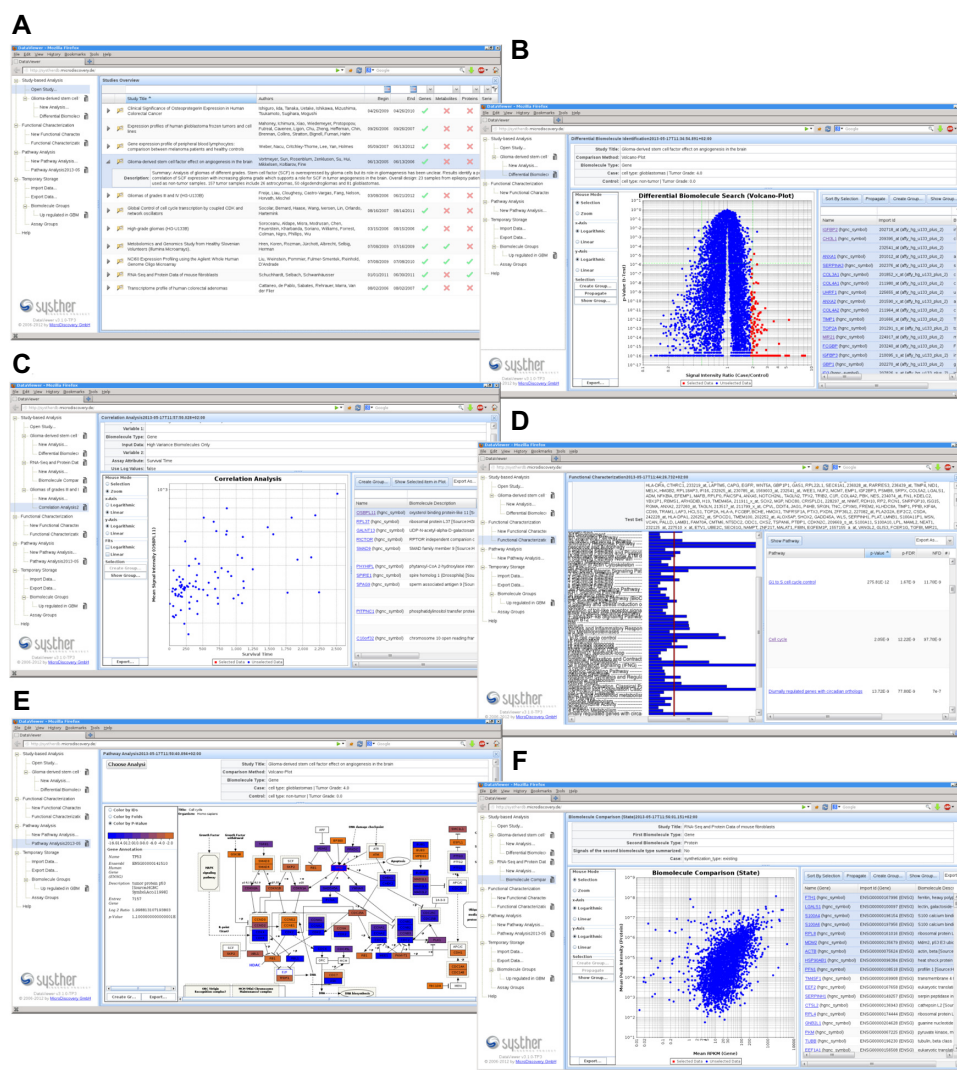
## Results

The main goal of BioMiner (beyond easily accessible storage of high-throughput cross-omics data) is to provide a set of tools that allow the performing of a number of statistical investigations. To demonstrate the intuitive usability of the application, we present typical workflows used for the identification of genes and pathways associated with GBM. The presented use cases are mainly based on data of the GBM Study by Sun et al.<sup>36</sup>, containing data from 180 microarrays derived from tissues of patients with different brain tumor types and grades, as well as from tissues of nontumor controls. This study has a relatively simple experimental design with a large number of replications. The major analysis steps of BioMiner are shown in Figure 2, including study overview, detection of differentially expressed genes, identification of pathways, visual pathway inspection, comparison of gene expression with protein expression, and correlation of genes with survival times. These steps are demonstrated in more detail.

**Study selection and inspection.** In order to access data, an experiment (study) is selected from an overview presenting

key properties of all studies in the database (Fig. 2A). To find a study of interest, the table can be filtered using keywords such as “glioma” or biomolecule types such as “genes.” Once a study is selected, different biological and technical properties can be inspected, such as number and grouping of samples, experimental or clinical parameters, or the experimental technologies used.

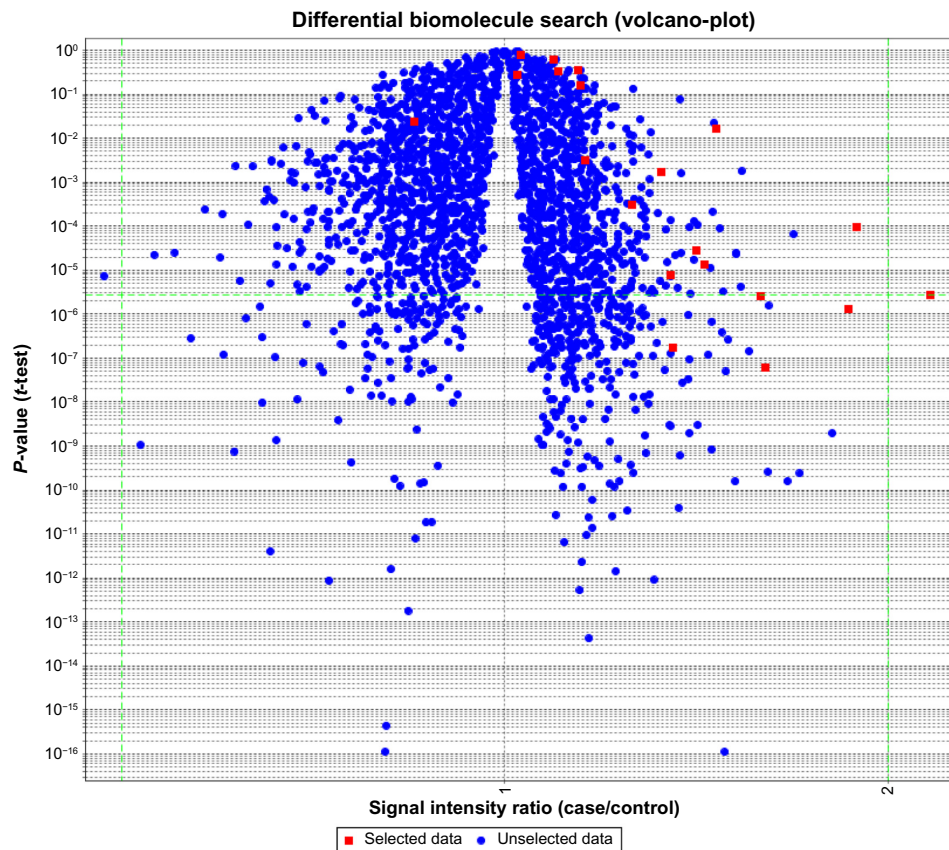
**Differential analysis.** The simplest and most common approach for the detection of tumor-associated biomarkers is a differential analysis based on statistical tests such as the *t*-test. Differential biomolecule identification can be configured using a drag-and-drop mechanism of the relevant samples. As an example, we compare GBM WHO grade IV samples versus nontumor control samples in the study of Sun et al.<sup>36</sup> The comparison is performed on the fly, and the resulting fold values and *P*-values are presented in a tabular view and can be displayed in interactive plots (eg, volcano plot: Fig. 2B). The two top upregulated genes between glioma WHO grade IV and nontumor samples are *IGFBP2* and *CHI3L1*, both of which are described as important glioma-related markers in the literature.<sup>37–40</sup> The table with the differential results can be filtered for a specific gene name or description. Searching for the ratio of curvise reveals a 2.6-fold upregulation in GBM WHO grade IV (Log<sub>2</sub> ratio of 1.4).



**Figure 2.** Data mining with BioMiner. Screenshots of different results from data mining with BioMiner including the following: (A) study overview, (B) detection of differentially expressed genes, (C) correlation of gene expression and survival time, (D) identification of significantly enriched pathways, (E) visual pathway inspection based on predefined layouts, and (F) biomolecule comparison of gene and protein expression. Results are typically presented in synchronized, parallel views composed of a table and a plot. The pathway inspection is shown in more detail in Figure 4.

**Correlation analysis.** Correlation is a general and versatile tool to identify associations and predictive relationships between biomolecules or experimental parameters. For example, if a study is annotated with survival times, the analysis of correlated events allows the identification of markers associated with survival. This association may be positive (high gene expression indicates long survival) or negative (high gene expression indicates short survival). BioMiner implements different types of correlation analyses to search for an association between different biomolecules or between biomolecules and clinical parameters. As an example, Figure 2C shows the results of a correlation analysis between genes and survival time based on the study of Freije et al.<sup>41</sup> A ranked table of highly correlated genes is complemented by a scatter plot of the top gene *OSBPL11*. We included several GBM- and colorectal cancer (CRC)-related studies with annotated survival times (Table 2).

**Working with gene groups.** The user can define and store a specific group of biomolecules, eg, by selecting a set of genes with significant *P*-values or high fold changes. Defining a biomolecule group offers the possibility for cross-study analyses, addressing, for instance, the question: How do the top upregulated genes from the data set of Sun et al.<sup>36</sup> behave in other data sets? To this end, a group of biomolecules can be loaded into existing plots. With this approach, the user can easily compare two studies. An example comparing two similar studies is given in Figure 3. This volcano plot visualizes the differential biomolecule comparison for astrocytoma grade III versus GBM grade IV using the data set from Freije et al.<sup>41</sup> Top upregulated genes identified with the data set from Sun et al.<sup>36</sup> are highlighted in red (refer “Differential analysis” section earlier). The majority of the highlighted genes show a very good agreement between both GBM studies. This agreement even increases if the highlighted gene set



**Figure 3.** Volcano plot for cross-study comparison. Volcano plot visualizing the differential comparison of astrocytoma grade III versus GBM grade IV using the data set from Freije et al.<sup>41</sup> Top upregulated genes identified with the data set from Sun et al.<sup>36</sup> (GBM grade IV vs control) are highlighted to investigate the relation between different studies. The majority of the highlighted genes show a good agreement between both brain tumor experiments.

is created from the comparison of tumor grade III with tumor grade IV in the study by Sun et al.<sup>36</sup> (data not shown). The top upregulated gene in both studies is *MIR21* (rightmost red point), which has been identified recently as a useful biomarker for GBM.<sup>42</sup> But there are also some disagreements between both data sets. Especially, the gene *MALAT1* (leftmost point in Fig. 3) is downregulated in the data based on Freije et al.<sup>41</sup> while showing an upregulation in the data set from Sun et al.<sup>36</sup>

Beyond comparing closely related studies, this approach can also be used to investigate relations between more heterogeneous studies. By highlighting the top genes from Sun et al.<sup>36</sup> in the differential comparison of adenoma versus normal mucosa,<sup>43</sup> we detect general and more specific cancer markers. *IGFBP2* (already reported above) and *TGFBI* are upregulated in both tumor types, while the gene *PLOD2* shows upregulation only in GBM but not in CRC.<sup>44</sup>

**Functional characterization and interactive pathway plot.** Typically, the result of a differential analysis is not a single gene but rather a set of up- or downregulated genes that may be related in a systemic context (such as a biochemical pathway). The contextual interpretation of a given set of genes can be challenging, especially considering the multitude of different gene functions and involvements in biochemical

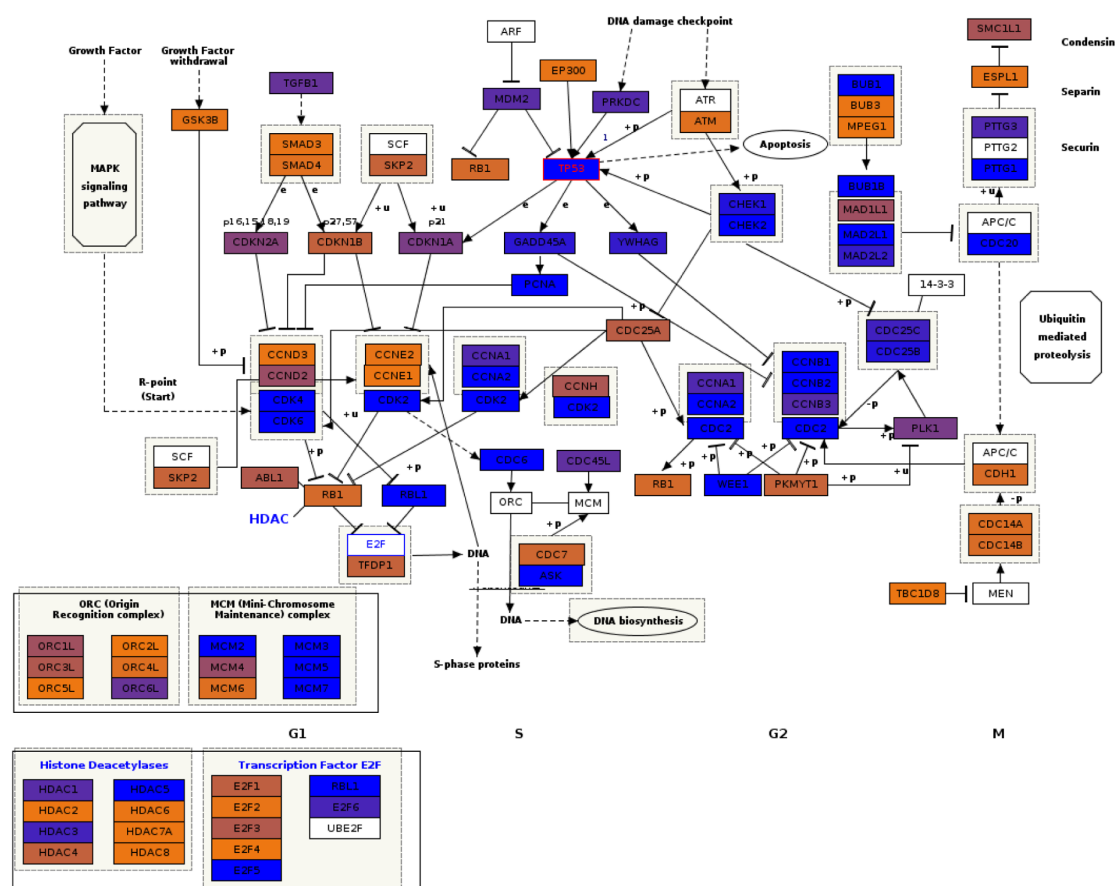
pathways. BioMiner provides a straightforward way to identify biochemical pathways, Gene Ontology terms, or chromosomal regions significantly related to a given set of genes (Fig. 2E). Pathways from WikiPathways and KEGG can be displayed and interactively inspected based on predefined layouts. Integrating results from differential analyses in the pathway graph allows the investigation of genes and metabolites in their systemic contexts. This is illustrated for the “cell cycle” pathway in Figure 4 using the differential biomolecule identification of GBM tumor grade 4 versus control (refer “Differential analysis” section).

**Cross-omics analysis.** The system internally contains cross-omics mappings between genes and proteins or between genes and metabolites. The complexity of the different cross-omics mappings is hidden from the user. For studies containing cross-omics data, the user can easily correlate different omics data types, eg, by creating a scatter plot comparing differential gene expression with differential protein expression (Fig. 2F).

## Discussion

In this article, we demonstrate that the Web-based application BioMiner is well suited to perform statistical analyses in cancer-related high-throughput experiments. The studies selected for illustration are related to GBM. All presented

Title: Cell cycle  
Organism: Homo sapiens



**Figure 4.** Pathway visualization. Interactive pathway visualization of the cell cycle pathway from WikiPathways repository.

**Notes:** The pathway graph can be scaled and exported as a png image. Additional information on genes and metabolites is available via selection. Color code refers to *P*-values of differential expression comparing GBM grade IV vs control using the Sun et al.<sup>36</sup> study (blue = significant differential expression).

analyses can be reproduced using the publicly available instance of the application.

In order to create a tool useful for a broad spectrum of life science users, our major goal was to keep BioMiner's usage as simple as possible while offering substantial flexibility in answering biological questions. Simplicity is achieved by following several different guiding principles. First, technical details are hidden from the user: ready-to-use normalized data are provided together with information on experimental design and clinical parameters. Second, experimental data are complemented by public knowledge on pathways and gene functions. Third, instead of providing an extensive repertoire of data analysis algorithms, we rather focus on a limited set of proven methods. To perform analyses beyond this set, it may be necessary to download result tables or complete study data and run analyses in more advanced statistical frameworks such as R/BioConductor.<sup>45</sup> Fourth, we model key aspects of the experimental design so that the main questions can be addressed rapidly. For instance, the case-control branches of a study are explicitly represented so that differential analysis can be achieved straightforwardly. More sophisticated analyses

deviating from the default design are possible but may result in slower response times.

Of course, easy accessibility comes at a price: performance-optimized queries with response times of a few seconds are available only for predefined groups. All preprocessing steps, notably data normalization, have to be performed before data import and cannot be changed during an analysis. At the current state of development, there is no option for users to import own data sets. The predefined cross-omics mappings cannot be changed or individualized by the user. They have to be updated by the system administrator on a regular basis. Given the benefit of a fluent and straightforward usability, we think these limitations are acceptable.

The true potential of BioMiner is revealed when looking at the huge studies with several hundreds of microarrays. It takes only seconds and a few mouse clicks to compare 165 lung cancer samples with 56 breast cancer samples from the Cancer Cell Line Encyclopedia (CCLE) study comprising >900 microarrays. The interactive result presentation makes it very easy to inspect the results, search for additional information, or identify functional relations.



The system is able to represent personalized data such as high-throughput data (genomics, epigenomics, proteomics, and metabolomics) in combination with clinical parameters. If used in a clinical context, BioMiner is well suited to address the needs of personalized medicine. Interfacing with tumor biobanks or clinical data management systems, the system will enable clinicians to directly evaluate patient-related high-throughput data and draw therapeutic conclusions.

## Conclusion and Outlook

BioMiner is a resource that enables rapid identification and assessment of relevant features in complex biological experiments. It provides clinicians and physicians a platform integrating high-throughput data together with clinical parameters. Straight design and ease of use make BioMiner perfectly suited for those who are not experts in bioinformatics and statistics. The application enables experimentalists to directly assess data, validate hypotheses, draw biological conclusions, and plan further experimental procedures. With an increasing number of experiments available for data analysis, the usefulness of BioMiner will still expand.

Given its intuitive usability, BioMiner may also serve as a platform for education, with instructors using the application for hands-on demonstration of high-throughput data analysis.

In the future, additional analytical capabilities of BioMiner will augment the use of this platform in the context of personalized medicine even further. Researchers will be able to submit their patients' molecular profiles to BioMiner and receive information about the patients' disease status and optimal course of treatment.

## Acknowledgments

We thank Sergio Grimbs and Matjaz Hren for helpful discussions.

## Author Contributions

CB, AG, and JSch developed and implemented the application. CB, KS, AG, KG, CS, MO-G, JS, and JSch drafted the manuscript. All authors read and approved the final manuscript.

## Supplementary Materials

**Supplementary File 1.** A tutorial detailing the usage of BioMiner.

**Supplementary File 2.** BioMiner methods for log-expression values, means and standard deviations, differential analysis with two groups, differential analysis with four groups, multiple testing adjustment and correlation analysis.

## REFERENCES

- Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med.* 2009;1(1):2.
- Godman B, Finlayson AE, Cheema PK, et al. Personalizing health care: feasibility and future implications. *BMC Med.* 2013;11:179.
- Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med.* 2010;363:4.
- Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008;5:e184.
- Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* 2011;39:D1005–10.
- Parkinson H, Sarkans U, Kolesnikov N, et al. ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.* 2011;39:D1002–4.
- Vizcaíno JA, Côté R, Reisinger F, et al. A guide to the proteomics identifications database proteomics data repository. *Proteomics.* 2009;9:4276–83.
- Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res.* 2011;39:19–21.
- Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res.* 2007;35:D760–5.
- Wang R, Fabregat A, Rios D, et al. PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat Biotechnol.* 2012;30:135–7.
- Berthold Michael R, Cebron N, Dill F, et al. KNIME: the Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Freiburg, Germany: Springer; 2007.
- Hull D, Wolstencroft K, Stevens R, et al. Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.* 2006;34:W729–32.
- Podpečan V, Lavrač N, Mozetič I, et al. SegMine workflows for semantic microarray data analysis in Orange4 WMS. *BMC Bioinformatics.* 2011;12:416.
- Goecks J, Nekrutenko A, Taylor J, Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010;11:R86.
- Gruden K, Hren M, Herman A, et al. A “crossomics” study analysing variability of different components in peripheral blood of healthy caucasoid individuals. *PLoS One.* 2012;7:e28761.
- Guo KT, Juerchott K, Fu P, et al. Isolation and characterization of bone marrow-derived progenitor cells from malignant gliomas. *Anticancer Res.* 2012;32:4971–82.
- Schichor C, Albrecht V, Korte B, et al. Mesenchymal stem cells and glioma cells form a structural as well as a functional syncytium in vitro. *Exp Neurol.* 2012;234:208–19.
- Jürchott K, Guo KT, Catchpole G, et al. Comparison of metabolite profiles in U87 glioma cells and mesenchymal stem cells. *Biosystems.* 2011;105:130–9.
- Schuchhardt J, Glintschert A, Hartl D, et al. BrainProfileDB – a platform for integration of functional genomics data. *Proteomics.* 2008;8:1162–4.
- Bauer C, Glintschert A, Schuchhardt J. ProfileDB: A resource for proteomics and cross-omics biomarker discovery. *Biochim Biophys Acta.* 2014;1844(5):960–6.
- van Iersel MP, Kelder T, Pico AR, et al. Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics.* 2008;9:399.
- Kopka J, Schauer N, Krueger S, et al. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics.* 2005;21:1635–8.
- Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 2009;37:D603–10.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 2010;38:D355–60.
- Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1–13.
- Matthews L, Gopinath G, Gillespie M, et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 2009;37:D619–22.
- Pico AR, Kelder T, Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol.* 2008;6:e184.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25–9.
- Bluthgen N, Brand K, Cajavec B, Swat M, Herzog H, Beule D. Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform.* 2005;16:106–15.
- Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet.* 2001;29:365–71.
- Taylor CF, Paton NW, Lilley KS, et al. The minimum information about a proteomics experiment (MIAPe). *Nat Biotechnol.* 2007;25:887–93.
- Bino RJ, Hall RD, Fiehn O, et al. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 2004;9:418–25.
- Flicek P, Aken BL, Ballester B, et al. Ensembl's 10th year. *Nucleic Acids Res.* 2010;38:D557–62.
- Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart Central Portal – unified access to biological data. *Nucleic Acids Res.* 2009;37:W23–7.
- Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* 2011;39:D712–7.
- Sun L, Hui AM, Su Q, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell.* 2006;9:287–300.



37. Li Y, Jiang T, Zhang J, et al. Elevated serum antibodies against insulin-like growth factor-binding protein-2 allow detecting early-stage cancers: evidences from glioma and colorectal carcinoma studies. *Ann Oncol.* 2012;23:2415–22.
38. Hsieh D, Hsieh A, Stea B, Ellsworth R. IGFBP2 promotes glioma tumor stem cell expansion and survival. *Biochem Biophys Res Commun.* 2010;397:367–72.
39. Seroo NV, Delfino KR, Southey BR, Beever JE, Rodriguez-Zas SL. Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med Genomics.* 2011;4:49.
40. Ku BM, Lee YK, Ryu J, et al. CHI3 L1 (YKL-40) is expressed in human gliomas and regulates the invasion, growth and survival of glioma cells. *Int J Cancer.* 2011;128:1316–26.
41. Freije WA, Castro-Vargas FE, Fang Z, et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* 2004;64:6503–10.
42. Ilhan-Mutlu A, Wagner L, Wöhrer A, et al. Plasma MicroRNA-21 concentration may be a useful biomarker in glioblastoma patients. *Cancer Invest.* 2012;30:615–21.
43. Sabates-Bellver J, Flier LG, Palo M, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res.* 2007;5:1263–75.
44. Dong S, Nutt CL, Betensky RA, et al. Histology-based expression profiling yields novel prognostic markers in human glioblastoma. *J Neuropathol Exp Neurol.* 2005;64:948–55.
45. R Development Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2011.
46. Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell.* 2006;9:157–73.
47. Wiedemeyer R, Brennan C, Heffernan TP, et al. Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. *Cancer Cell.* 2008;13:355–64.
48. Tsukamoto S, Ishikawa T, Iida S, et al. Clinical significance of osteoprotegerin expression in human colorectal cancer. *Clin Cancer Res.* 2011;17:2444–50.
49. Critchley-Thorne RJ, Yan N, Nacu S, Weber J, Holmes SP, Lee PP. Down-regulation of the interferon signaling pathway in T lymphocytes from patients with metastatic melanoma. *PLoS Med.* 2007;4:e176.
50. Orlando DA, Lin CY, Bernard A, et al. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature.* 2008;453:944–7.