MARCARviz: Interactive web-platform for exploratory analysis of toxicogenomics data for nongenotoxic hepatocarcinogenesis

Michael Römer[1], Heidrun Ellinger-Ziegelbauer[2], Bettina Grasl-Kraupp[3], Michael Schwarz[4], and Andreas Zell[1]

[1]**Department of Cognitive Systems, University of Tuebingen, Sand 1, 72076 Tuebingen, Germany**
[2]**Investigational Toxicology, Bayer Pharma AG, 42096 Wuppertal, Germany**
[3]**Department of Medicine, Medical University of Vienna, 1090 Vienna, Austria**
[4]**Department of Toxicology, University of Tuebingen, Wilhelmstraße 56, 72074 Tuebingen, Germany**

ABSTRACT

The late detection of non-genotoxic carcinogens in the drug development process can delay drug candidates for unmet medical needs from reaching the market despite considerable investments in their development. To enable faster, safer, and less expensive development of medications for patients, the MARCAR project generated a large set of transcriptomic data to investigate the underlying mechanisms of non-genotoxic hepatocarcinogenesis and to identify potential biomarkers for early detection of tumor formation in the rodent liver. The effective mining of these high-dimensional datasets is a non-trivial task that usually requires bioinformatics support to extract relevant mechanistic patterns and confirm toxicological hypotheses. Here, we present MARCARviz, a web-platform that enables biologists to (a) quickly address the most common questions associated with the MARCAR microarray data, to (b) identify relevant patterns in the data, and to (c) generate or confirm mechanistic hypotheses about non-genotoxic effects leading to cancer formation. The major advantage of MARCARviz is that there is no software or advanced technical knowledge required to perform powerful analyses and generate visualizations of the MARCAR data. MARCARviz greatly facilitates the confirmation of published MARCAR results and generation of new insights from the collected data by the greater public without the requirement for tedious pre-processing steps. MARCARviz is publicly available from https://tea.cs.uni-tuebingen.de/.

Keywords: Toxicology; Transcriptomics; Visualization; Nongenotoxic Carcinogenicity

INTRODUCTION

The detection of tumors during preclinical toxicology studies during drug development can result in the delay of drug candidates reaching the market depending on the mode of action, human relevance, and the intended therapeutic indication. Drug candidates associated with direct genotoxicity leading to DNA mutations are efficiently eliminated early during development with established short-term assays. However, a significant proportion of cancerogenic drug candidates induces tumors through mechanisms other than DNA mutations and are therefore called non-genotoxic carcinogens (NGC). To identify such compounds, time-consuming *in vivo* lifetime rodent assays are required. These 2-year rodent assays require a large number of animals, take at least three years to be completed and cost between $1 and $2 million US dollar (Johnson, 2012). The earlier detection of cancer development upon compound treatment in conjunction with an understanding of the effect of a drug at the molecular level could lead to significant savings in time, cost, and animal numbers, and could benefit patients regarding drug safety.

In this context, the MARCAR project (2010-2015, http://www.imi-marcar.eu/) generated a wealth of data to increase insights into mechanisms of NGC action and to detect carcinogenic effects of drug candidates earlier (see for example (Lempiainen et al., 2013; Unterberger et al., 2014; Thomson et al., 2014; Römer et al., 2014)). The MARCAR project explored primarily the liver, which is the major target organ of NGC induced tumors in rodents (Knight et al., 2006). A particular goal was

the identification of early biomarkers of NGC effects to provide the rationale for designing new short-term assays for earlier detection of potential NGCs. For differentiation of NGCs from genotoxic carcinogens and non-carcinogens, data were also generated from samples after treatment of rodents with such compound classes.

A fundamental concept of the MARCAR project was the integration of data collected at multiple levels of gene expression regulation. In addition to traditional profiling of mRNA abundance with microarrays, non-coding mRNAs were measured by array technology, protein abundance and modifications were assessed by reverse-phase protein arrays, and DNA methylation patterns were identified (Unterberger et al., 2014; Römer et al., 2014; Thomson et al., 2012). Potential biomarkers identified include, e.g., the *Dlk1-Dio3* imprinted gene cluster of noncoding RNAs (Lempiainen et al., 2013), epigenetic changes in the 5-hydroxymethylome (Thomson et al., 2012), and effects on the hepatic mesenchyme (Riegler et al., 2015). To effectively explore the interaction of multiple layers of gene regulation, the InCroMAP software was developed, which provides KEGG pathway maps with overlaid information on, e.g., mRNA expression, miRNA-gene interactions, and methylation changes (Wrzodek et al., 2013). Toxicogenomics approaches were applied to extract signatures, i.e., lists of biomarkers, to identify NGC action in both rats and mice (Kossler et al., 2015; Eichner et al., 2013). In total, the MARCAR project has generated 27 datasets across three species (*Homo sapiens*, *Rattus norvegicus*, *Mus musculus*) and four regulation levels (mRNA and miRNA expression, protein abundance, and methylation changes), which are available from the Gene Expression Omnibus (GSE68387).

The majority of MARCAR datasets were generated with traditional mRNA microarray profiling techniques. These arrays measure the expression of almost all known genes in the genome of a certain species at once, providing an enormous amount of raw data that requires preprocessing, normalization, and statistical or computational approaches to effectively identify patterns and specific markers and generate or confirm mechanistic hypotheses. Traditionally, bioinformaticians would process the data and provide tables and static visualizations to biologists, who would then try to interpret the data. To facilitate the process of generating insights from the wealth of data, we created MARCARviz, a set of software tools that allows an interactive visual analysis to quickly identify relevant patterns in large amounts of microarray data even by non-bioinformaticians. MARCARviz can be used to quickly answer common questions associated with the experiments, e.g., which genes are affected by the treatment with a specific compound, generate a mechanistic hypothesis, for example by finding enriched pathways and Gene Ontology (GO) terms, or compare the effects of several compounds. Also, MARCARviz provides cross-platform and cross-species analysis, which usually require the mapping of several types of identifiers (e.g., Affymetrix probe IDs to gene names) or identification of orthologous genes. The most relevant visualizations for microarray data, for example, heat maps, Venn diagrams, or volcano plots, are provided and enriched with interactive functionality for pattern identification. MARCARviz is a web-platform that requires no additional plugins or computational resources. All analyses are performed on a computation cluster to provide results as fast as possible. The major advantage of MARCARviz is that it allows biologists without advanced bioinformatics knowledge to quickly answer the most common questions that might be asked of the MARCAR mRNA expression data, extract the data and figures that support their hypothesis, and generate new insights and hypotheses about NGC mechanisms and biomarkers.

METHODS AND MATERIAL

Datasets

In total, 16 microarray datasets have been generated and processed for inclusion in MARCARviz (see Table S1). The animal experiments have been approved by the respective ethics committees and were performed according to established experimental guidelines. Study design and raw data generation have been previously described (Unterberger et al., 2014; Riegler et al., 2015; Braeuning et al., 2010; Luisier et al., 2014; Eichner et al., 2014; Lempiäinen et al., 2011; Ellinger-Ziegelbauer et al., 2008; Braeuning et al., 2016). These datasets cover three species: *Mus musculus* and *Rattus norvegicus* (*in vivo*), and *Homo sapiens* (*in vitro*). The datasets can be grouped into two broad categories: short-term effects of non-genotoxic carcinogens and mechanistic analysis of a model NGC (Phenobarbital). To study the common characteristics of non-genotoxic carcinogens, rats and mice were administered daily doses of several non-genotoxic, carcinogenic substances for up to four weeks. For comparison, genotoxic carcinogens and non-hepatocarcinogens have been included to identify effects that are specific for non-genotoxic carcinogenesis (see Table S2). For the mechanistic investigation of the model substance phenobarbital, several knockout studies have been performed in mice, along with transcriptomic profiling of tumor and normal tissue. In rats, mesenchymal cells and hepatocytes

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

have been isolated to identify effects that occur specifically in different cell types. All raw data were submitted to GEO and is available under the accession number GSE68387.

Data processing

Microarray quality control was performed using the R/Bioconductor package arrayQualityMetrics to remove outliers and low-quality samples (Gentleman et al., 2004; Kauffmann et al., 2009). Raw data processing was performed with the R/Bioconductor packages *affy* and *limma* (Gautier et al., 2004; Smyth, 2005). In short, Affymetrix 3' IVT expression array data was normalized with the Robust Multiarray Average (RMA) method and summarized to Entrez Gene IDs using custom Brainarray CDF files (Dai, 2005). Agilent microarray data was within-array background corrected, quantile normalized between arrays, and summarized to Entrez Gene IDs. For each dataset, we normalized all samples together. To eliminate batch effects, all studies use designs that ensure that treated and control samples are run in a single batch. A moderated t-value implemented in *limma* was used to compute p-values for significant deregulation of genes. We used the Benjamini-Hochberg method to correct for multiple hypothesis testing. Logarithmized (base 2) fold changes were calculated as the log2 of the observed mean intensity ratio between treated and control animals for each gene. The *biomaRt* package for R/Bioconductor was used to identify orthologous genes for cross-species data comparison (Durinck et al., 2009).

Pathway and gene ontology enrichment

Gene to gene set mapping files were obtained from the Molecular Signature Database (MSigDB, Liberzon et al. (2011)) for the KEGG, BioCarta, and Reactome pathway databases and the Gene Ontology database (Ashburner et al., 2000; Nishimura, 2001; Matthews et al., 2009; Kanehisa et al., 2012). For the enrichment analysis, all gene identifiers are mapped to orthologous HGNC gene symbols based on information available from the Entrez Gene database (Maglott, 2004). A hypergeometric test is used to calculate p-values for gene set enrichment. The formula for the test is

$$P(X \geq m) = \sum_{i=m}^{M} \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

where N is the total number of genes measured, M is the number of measured genes that are present in the gene set, n is the total number of differentially expressed genes, and m is the number of differentially expressed genes in the gene set. The p-values calculated by this hypergeometric tests are very close to those computed using Fisher's exact test. All calculated p-values were corrected for multiple hypothesis testing with the method developed by Benjamini-Hochberg.

Web-platform setup

The MARCARviz front end that is visible to the user is based entirely on the current web standards HTML5, JavaScript, and CSS. No Flash or Java applets are required for the interactive functionality. Only a browser with HTML5 support is needed. We use Highcharts.js for rendering interactive scatter, volcano, and bar plots, customized versions of InCHlib.js (Skuta et al., 2014) for interactive heat maps, jvenn.js (Bardou et al., 2014) for Venn diagrams, and Bootstrap.js for the general layout of the website. At the back end, MARCARviz uses a Node.js web server to handle communication between server and client. The Node.js server enables the scalable handling of simultaneous requests from multiple clients. All communication is secure and encrypted according to the HTTPs standard. Requests are distributed to an R backend for necessary data processing. A NoSQL based MongoDB is used for schema-free data storage and fast data retrieval. A schematic overview is shown in Fig. 1. Currently, the upload of user-provided datasets is not supported but this feature is planned for future releases.

RESULTS AND DISCUSSION

Main features and user interface

MARCARviz is a web-platform that is navigated through a menu available on all pages. We made a major distinction between "Analysis" and "Visualization" tools. Typically, an analysis will give a report in the form of a table, e.g., of differentially expressed genes or enriched pathways. A visualization will produce an interactive plot that can be used for visual exploration of the data or as a figure supporting a hypothesis in a manuscript or presentation. Currently, MARCARviz supports two analyses and five visualizations, which we will describe below along with possible use cases. Examples of the user interface and the interactive visualization of the data are shown in Fig. 2 and 3.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

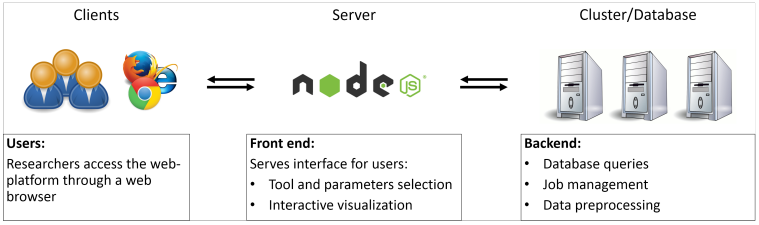

Figure 1. Architecture of the MARCARviz web platform. This is a simplified scheme of the MARCARviz platform architecture that demonstrates the interaction of the general components. Researchers access MARCARviz through the internet with the supported major browsers and use the web interface to select required analyses or visualizations. A node.js server handles the requests, serves static content, and manages the distribution of the requested analyses to the computation cluster. On the computation cluster, the necessary database queries, data preprocessing and integration (e.g., for data from multiple studies) are performed. The results of the requested analyses are then returned to the user by the server and rendered by the browser using Javascript, HTML5, and CSS.

The analysis tool "Differential expression" allows the identification of genes that are affected by a specific condition, e.g., treatment with a non-genotoxic carcinogen. Differential expression is established for example by comparing the gene expression in treated animals or tumor tissue with expression in untreated control animals or normal, non-tumor tissue. The strength of differential expression is measured as a fold change, i.e., the ratio of expression between treated and control animals. The significance of deregulation is given as the p-value of a moderated t-value. The user can define the fold change and p-value cutoffs used to identify affected genes. The deregulated genes are shown in a table that also gives a summary for each gene and provides links to external databases with additional information on the gene. The user can also choose to inspect the observed expression in individual samples to confirm the observed deregulation. The list of genes identified by this tool can be saved to be used as input for other tools, e.g., for a gene set enrichment analysis, or exported as a CSV, PDF, or Excel file.

The "Gene set enrichment" analysis tool identifies gene sets, e.g., KEGG or Reactome pathways or GO categories, for which a higher-than-expected number of genes is up- or downregulated in a condition. The test for enrichment is performed with a hypergeometric test that is equal to Fisher's exact test and calculates the probability of observing the proportion of deregulated genes in a gene set given the base proportion of deregulated genes in all measured genes. The significance of a gene set enrichment is reported by the q-value, i.e., the p-value obtained from the hypergeometric test after multiple-testing correction with the Benjamini-Hochberg method. The user can define the thresholds applied to filter differentially expressed genes and select the gene sets for which enrichment is tested. Currently, MARCARviz supports enrichment tests for GO categories and three pathway databases: KEGG, BioCarta, and Reactome. The results of the gene set enrichment analysis are shown in a table that reports the significance for each gene set along with the deregulated genes, the gene set statistics, and links to external databases with additional information on the gene sets. The table can also be exported as a CSV, PDF, or Excel file The visualization tool "Volcano plots" allows visual representation of the strength and significance of differential gene expression in a single condition. As for the "Differential expression" tool, the fold change and p-value are used to represent condition-dependent gene regulation. The interactive plot created by the "Volcano plots" tool shows the strength, i.e., the fold change, on the x-axis and the significance, i.e., the p-value, on the y-axis. Each point in the plot corresponds to a single gene. When the user hovers over a point, the corresponding gene symbol is shown, and if the point is clicked, the observed expression of the gene in the condition and control samples is displayed to confirm the regulation status. Again, the user can set the thresholds for fold change and p-value which define differential regulation and save the list of deregulated genes for further analyses. The volcano plot can be exported in standard image formats (PNG or PDF).

To compare the general effects on gene expression in two conditions, the "Scatter plots" visualization tool creates a fold change versus fold change plot. For each gene, the fold change in the first condition is plotted on the x-axis and the fold change in the second condition on the y-axis. In addition, a linear regression analysis is performed to provide statistical measures (for example the R2 value) of the concordance between the conditions. This tool supports cross-platform and cross-species

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

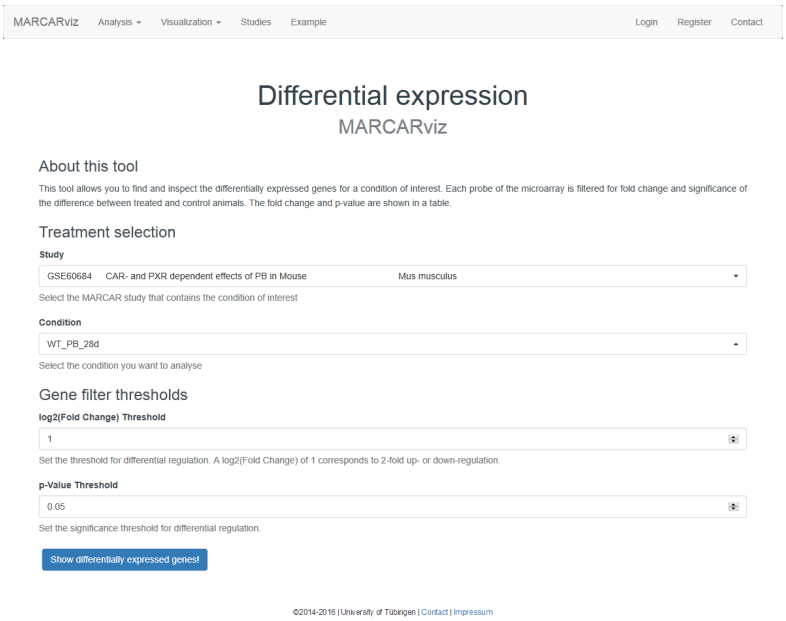

Figure 2. User interface of the MARCARviz tool for differential gene expression. This is a screenshot of the MARCARviz user interface for the tool "Differential expression". At the top of the page, the user can use the menu to navigate the page, e.g., to select the desired analysis or visualization tool. An interactive example of the use of MARCARviz is provided, as well as an overview of all studies that are available through MARCARviz. In the main body of the page, the user selects the conditions of interest and adjusts the settings to his needs. For all settings, default values are provided. When the user has finished the selection of condition and settings, a job is submitted by clicking on the button at the bottom of the page. This will redirect the user to a new page, on which the progress and the result (e.g., the interactive table or visualization) of the submitted job are shown.

comparisons, e.g., to compare the effects of the administration of a carcinogenic substance in rats and mice. The user can set a fold change threshold to exclude non-affected genes from the comparison and export the resulting plot in standard image formats.

The visualization tool "Venn diagrams" is the second tool for comparing multiple treatments. A Venn diagram shows the overlap of lists of deregulated genes for up to six different conditions, which enables the identification of genes that are deregulated in several conditions. For example, this can be used to find genes that are affected by several non-genotoxic substances or by one substance at multiple time points. As with the other tools, the user can set the fold change and p-value thresholds used to identify deregulated genes or choose to include only up- or downregulated genes. The "Venn diagrams" tool also supports cross-platform and cross-species comparisons by internally using data from orthologous genes. The lists of shared deregulated genes can be saved for other analyses or downloaded as standard text files.

The most powerful tool for visual data exploration is the "Heat maps" visualization tool, which creates heat maps, i.e., color-coded matrix representations of the strength of gene deregulation in multiple conditions. This enables an easy, visual identification of common expression patterns that are shared across several conditions. Control conditions, e.g., non-hepatocarcinogenic or genotoxic substances, can be included to visually identify genes that are deregulated specifically in the conditions of interest. The visual exploration is further facilitated by the interactivity of the heat map: the user can zoom in on genes or gene clusters that are interesting, search for specific genes, or hide conditions that are not of interest. Genes are clustered hierarchically by their expression pattern in the selected conditions to group co-expressed genes. The user can set thresholds for fold change and p-value to exclude genes that are not deregulated in any selected condition or provide a list of genes that should be shown in the heat map, e.g., from a previous analysis of deregulated genes in a specific condition. The heat map can be downloaded as a PNG file for inclusion in manuscripts or presentations. The genes that the user identified can be saved as a gene list for further analysis. A modified variant of the heat map tool, "Gene set heat maps", allows the same visual exploration for gene set enrichments, i.e., the significance of the hypergeometric test for gene set enrichment is color coded in the heat

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

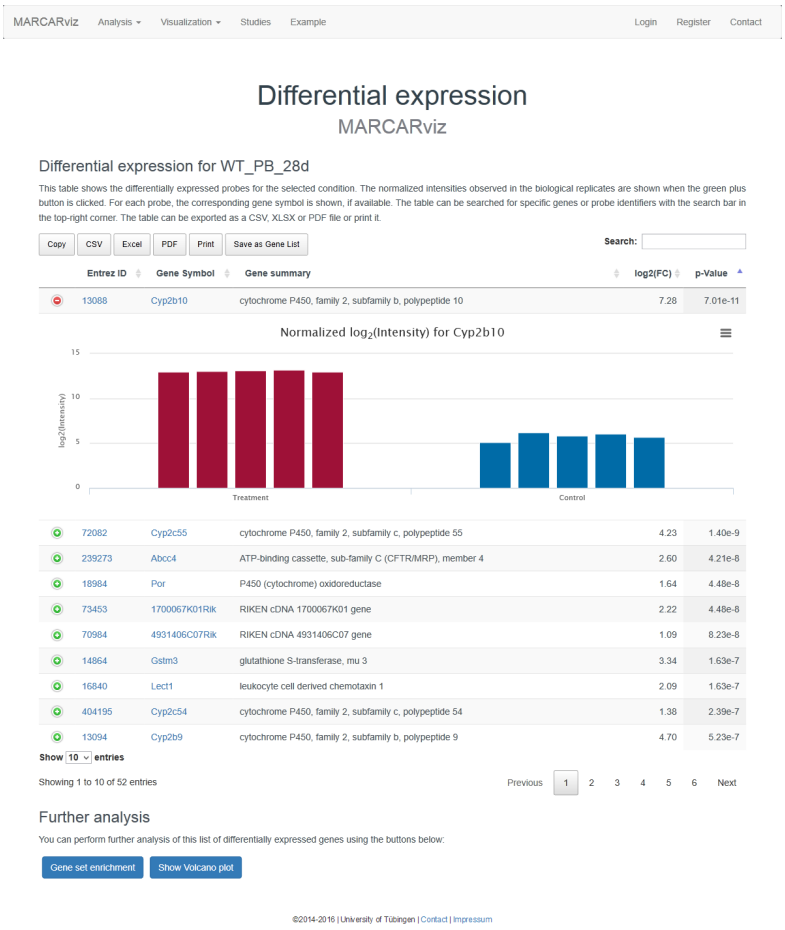

Figure 3. User interface of an interactive table of differential expression. This screenshot shows the result page of the tool differential expression. The differentially regulated genes that meet the inclusion criteria defined by the user are shown in a table. The table provides supplementary information, e.g., the official gene symbol and a short gene summary. By clicking on the table row, the expression data for the selected gene is shown in more detail. The table can be exported in common file formats (CSV, Excel, and PDF) or saved as a gene list for further analysis. Quick links to do gene set enrichment analysis or display a volcano plot for the treatment are also provided.

map. This exploratory analysis can help with the generation of new hypotheses on the mechanisms of non-genotoxic carcinogens, both on the level of single genes and gene sets.

Use case: Identification of phenobarbital target genes and pathways

Phenobarbital is an anticonvulsant drug that is used to treat many types of seizures in patients and has been in use for over a century. It is included in the WHO Model List of Essential Medicines. However, it has repeatedly been shown that phenobarbital acts as a non-genotoxic hepatocarcinogen in male and female mice (see, e.g., in the Carcinogenic Potency Database by Fitzpatrick (2008)) and is listed as a group 2B carcinogen by the IARC. Therefore, phenobarbital has been extensively studied as a model substance for non-genotoxic carcinogens. Here, we will use MARCARviz and a MARCAR dataset to identify potential target genes of phenobarbital, pathways that are affected by treatment with phenobarbital, and investigate the dependence of these effects on the constitutive androstane receptor (CAR) and the pregnane X receptor (PXR). The dataset is available from the Gene Expression Omnibus under the accession number GSE60684. An extensive analysis of this dataset was previously published by Luisier et al. (2014). This use case is also available as an interactive example online at the MARCARviz web platform. The dataset was generated with Affymetrix microarrays. We assessed the array and sample quality, normalized the raw data with the RMA method, and calculated fold changes and Benjamini-Hochberg multiple-testing corrected *p*-values as

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

described in the Methods section.

As the first step in our analysis, we use the "Differential expression" tool to identify potential target genes. We filtered genes that are at least two-fold deregulated (up- or downregulated) and show significant differences between treated and vehicle control animals (corrected *limma* p-value ≤ 0.05) for wild-type mice receiving phenobarbital each day for up to 13 weeks. Gene expression has been profiled with microarrays at five time points, after 1, 7, 14, 28, and 91 days. At all five time points, *Cyp2b10* and *Cyp2c55* are among the top deregulated genes. Using the "Venn diagram" tool, we identified 11 genes that are significantly up- or downregulated at all five time points, among them four *Cyp* genes (*Cyp2b10*, *Cyp2c55*, *Cyp2c37*, and *Cyp2c54*), the Wnt signaling inhibitor Wisp1, and other genes that have previously been linked with phenobarbital treatment, e.g., *Gstm3* (Lempiäinen et al., 2011; Lempiainen et al., 2013). Similar observations are made in mice in which CAR and PXR have been replaced with humanized CAR and PXR. Again, *Cyp2b10* and *Cyp2c55* are among the top deregulated genes at all five time points, and five other genes are deregulated at all five time points: *Abcc4*, *Akr1b7*, *Cbr3*, *Gstm3*, and *Por*. In contrast, in mice in which CAR and PXR have been knocked out, differential regulation is almost entirely eliminated at all time points. We also used the "Gene set enrichment" tool to find pathways that are deregulated after treatment with phenobarbital. For wild-type mice, the most affected KEGG pathways are the drug metabolism by cytochrome P450, glutathione metabolism, and retinol metabolism. These are consistently deregulated (q-value ¡ 0.001) at all five time points. Again, we find very similar results for the mice with humanized CAR and PXR. For the CAR and PXR knockout mice, no pathway deregulation is observed as expected due to the lack of deregulation of individual genes. These results obtained with MARCARviz are in concordance with the results of the analysis performed by Luisier et al. (2014). To visualize the major effects of phenobarbital on gene expression in wild type, humanized CAR/PXR, and CAR/PXR knockout mice, we used the "Heat maps" tool, using the same filtering (two-fold deregulation and corrected *limma* p-value ≤ 0.05) and clustering of both genes and conditions (see Fig. 4). The clustering of the conditions shows a large difference in gene expression between wild-type and humanized CAR/PXR mice on one hand and the CAR/PXR knockout mice on the other.

Discussion

MARCARviz allows researchers to perform the most commonly used analyses and visualizations for microarray data, i.e., detection of differentially expressed genes and enriched gene sets along with visualization of this data in heat maps, volcano and scatter plots, or Venn diagrams. Throughout the web-platform, MARCARviz provides additional information on genes and pathways or offers links to databases that contain additional information. This allows researchers to address a wide range of questions using the MARCAR data. In general, these can be classified into three overall tasks: identifying the differentially regulated genes in a condition, mechanistic analysis of these genes, and comparison of the observed effects in two or multiple conditions. By facilitating these analyses, MARCARviz provides toxicological researchers with the opportunity to generate and confirm mechanistic hypotheses supported by the MARCAR data, which is one of the largest resources for data representing, e.g., effects of shorter-term exposures of NGCs on gene expression in rodents, in comparison to a selection of genotoxic carcinogen and non-carcinogens or gene expression profiles in different types of mouse liver tumors characterized by either Ctnnb1 or Ha-ras oncogenic mutations. The interactive results can easily be shared with collaborators by sending links to the result web page. Alternatively, all tables and figures can be exported to standard table and figure formats. As the raw MARCAR gene expression data are provided in addition through GEO, bioinformatics analyses could also be performed starting with these raw data. However, MARCARviz provides preprocessed, ready-for-analysis data along with a user interface, and no necessity to install any additional software. Only a modern browser is required. The cross-species and cross-platform comparison of multiple datasets offered by MARCARviz is a particular advantage that would otherwise require mapping of the different manufacturer identifiers and identifying orthologous genes across species. All preprocessed data are available for download to allow researchers the use of their preferred tools to perform analyses that are not already offered by MARCARviz. In the future, we will respond to user feedback and continue to develop and improve the functionality and user interface to provide the main point-of-entry for analyzing MARCAR data.

CONCLUSION

MARCARviz is a web-platform for interactive visual exploration of the effects of non-genotoxic carcinogens on transcriptional regulation in rodents and human *in vitro* hepatocyte models, in comparison to some genotoxic carcinogens and non-carcinogen. We collected data from 16 datasets

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

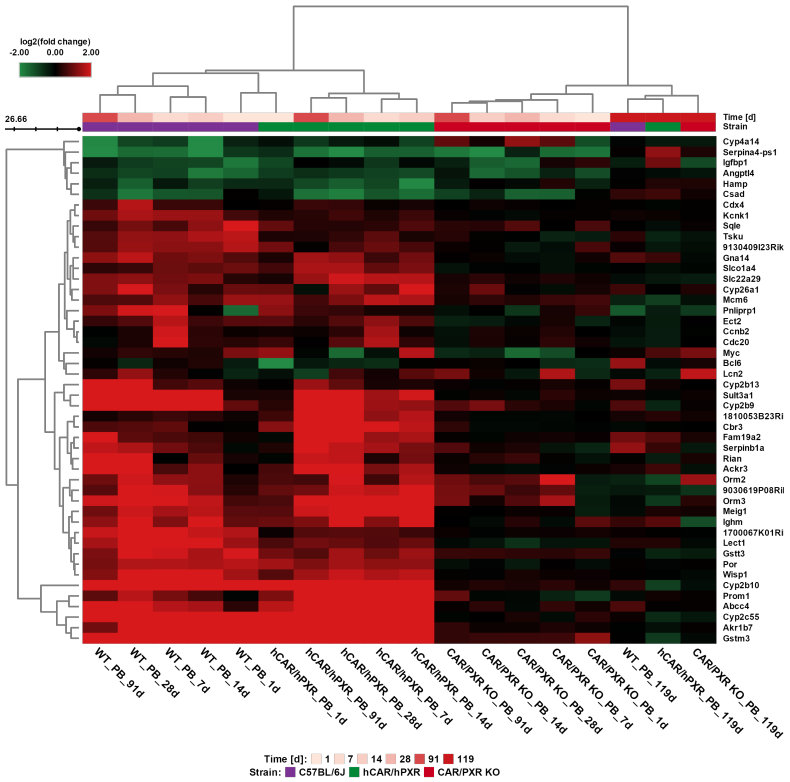

Figure 4. Heat map of deregulated genes after treatment with phenobarbital of wild-type mice, mice with humanized CAR/PXR and CAR/PXR knockout mice. Genes have been filtered for three-fold up- or downregulation and significant deregulation (corrected limma p-value ≤ 0.01) in at least one condition. Each row corresponds to a gene, each column to a condition (i.e., a combination of time point and mouse strain), each cell shows the log2 fold change between treated and vehicle control samples. The annotation bar above the heat map shows the time point and strain for each condition. The hierarchical clustering demonstrates the similarities of phenobarbital-mediated effects in the wild-type and humanized CAR/PXR mice. In contrast, gene deregulation is almost completely eliminated in CAR/PXR knockout mice. After recovery (conditions after 119 days), gene expression in wild-type and humanized CAR/PXR mice is similar to the knockout mice.

comprising 274 different conditions that were generated and analyzed over the course of the MARCAR project. The data included in MARCARviz cover the two most used rodent species in pre-clinical risk assessment: mouse (*Mus musculus*) and rat (*Rattus norvegicus*), as well as human *in vitro* data. We provide the most commonly used analyses and visualizations for microarray data in an interactive fashion that allows visual exploration to aid the generation and validation of new hypotheses about the mechanism of non-genotoxic carcinogenesis. This will facilitate the discovery of biomarkers and new methods for early detection of carcinogenic effects of drug candidates in preclinical risk assessment.

The datasets described in this article are available from the Gene Expression Omnibus (GSE68387, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68387). The web-platform MARCARviz is publicly available from https://tea.cs.uni-tuebingen.de/. There are no restrictions on its use by academic users.

ACKNOWLEDGMENTS

The research leading to this work was funded by the Innovative Medicine Initiative Joint Undertaking (IMI JU) under grant agreement number 115001 (MARCAR project). We acknowledge support by the German Research Foundation (DFG) and the Open Access Publishing Fund of the University of Tuebingen.

REFERENCES

Ashburner, M., Ball, C. A., and Blake, J. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC bioinformatics*, 15(1):293.

Braeuning, A., Gavrilov, A., Geissler, M., Wenz, C., Colnot, S., Templin, M. F., Metzger, U., Römer, M., Zell, A., and Schwarz, M. (2016). Tumor promotion and inhibition by phenobarbital in livers of conditional Apc-deficient mice. *Archives of Toxicology*, pages 1–14.

Braeuning, A., Singh, Y., Rignall, B., Buchmann, A., Hammad, S., Othman, A., Recklinghausen, I., Godoy, P., Hoehme, S., Drasdo, D., Hengstler, J. G., and Schwarz, M. (2010). Phenotype and growth behavior of residual β-catenin-positive hepatocytes in livers of β-catenin-deficient mice. *Histochemistry and Cell Biology*, 134(5):469–481.

Dai, M. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research*, 33(20):e175–e175.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8):1184–1191.

Eichner, J., Kossler, N., Wrzodek, C., Kalkuhl, A., Bach Toft, D., Ostenfeldt, N., Richard, V., and Zell, A. (2013). A Toxicogenomic Approach for the Prediction of Murine Hepatocarcinogenesis Using Ensemble Feature Selection. *PLoS ONE*, 8(9):e73938.

Eichner, J., Wrzodek, C., Römer, M., Ellinger-Ziegelbauer, H., and Zell, A. (2014). Evaluation of toxicogenomics approaches for assessing the risk of nongenotoxic carcinogenicity in rat liver. *PloS ONE*, 9(5):e97678.

Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H. J. (2008). Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutation Research*, 637(1-2):23–39.

Fitzpatrick, R. B. (2008). CPDB: Carcinogenic Potency Database. *Medical Reference Services Quarterly*, 27(3):303–311.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80.

Johnson, D. E. (2012). Estimating Human Cancer Risk from Rodent Carcinogenicity Studies: The Changing Paradigm for Pharmaceuticals. *Journal of Drug Metabolism & Toxicology*, 03(06).

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(D1):D109–D114.

Kauffmann, A., Gentleman, R., and Huber, W. (2009). arrayQualityMetrics–a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25(3):415–416.

Knight, A., Bailey, J., and Balcombe, J. (2006). Animal carcinogenicity studies: 2. Obstacles to extrapolation of data to humans. *ATLA Alternatives to Laboratory Animals*, 34(1):29–38.

Kossler, N., Matheis, K. A., Ostenfeldt, N., Bach Toft, D., Dhalluin, S., Deschl, U., and Kalkuhl, A. (2015). Identification of specific mRNA signatures as fingerprints for carcinogenesis in mice induced by genotoxic and nongenotoxic hepatocarcinogens. *Toxicological Sciences*, 143(2):277–95.

Lempiainen, H., Couttet, P., Bolognani, F., Muller, A., Dubost, V., Luisier, R., del Rio-Espinola, A., Vitry, V., Unterberger, E. B., Thomson, J. P., Treindl, F., Metzger, U., Wrzodek, C., Hahne, F., Zollinger, T., Brasa, S., Kalteis, M., Marcellin, M., Giudicelli, F., Braeuning, A., Morawiec, L., Zamurovic, N., Langle, U., Scheer, N., Schubeler, D., Goodman, J., Chibout, S.-D., Marlowe, J., Theil, D., Heard, D. J., Grenet, O., Zell, A., Templin, M. F., Meehan, R. R., Wolf, R. C., Elcombe, C. R., Schwarz, M., Moulin, P., Terranova, R., and Moggs, J. G. (2013). Identification of Dlk1-Dio3 Imprinted Gene Cluster Noncoding RNAs as Novel Candidate Biomarkers for Liver Tumor Promotion. *Toxicological Sciences*, 131(2):375–386.

Lempiäinen, H., Müller, A., Brasa, S., Teo, S.-S., Roloff, T.-C., Morawiec, L., Zamurovic, N., Vicart, A., Funhoff, E., Couttet, P., Schübeler, D., Grenet, O., Marlowe, J., Moggs, J., and Terranova, R. (2011). Phenobarbital Mediates an Epigenetic Switch at the Constitutive Androstane Receptor (CAR) Target Gene Cyp2b10 in the Liver of B6C3F1 Mice. *PLoS ONE*, 6(3):e18216.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12):1739–1740.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016

Luisier, R., Lempiainen, H., Scherbichler, N., Braeuning, A., Geissler, M., Dubost, V., Muller, A., Scheer, N., Chibout, S.-D., Hara, H., Picard, F., Theil, D., Couttet, P., Vitobello, A., Grenet, O., Grasl-Kraupp, B., Ellinger-Ziegelbauer, H., Thomson, J. P., Meehan, R. R., Elcombe, C. R., Henderson, C. J., Wolf, C. R., Schwarz, M., Moulin, P., Terranova, R., and Moggs, J. G. (2014). Phenobarbital Induces Cell Cycle Transcriptional Responses in Mouse Liver Humanized for Constitutive Androstane and Pregnane X Receptors. *Toxicological Sciences*, 139(2):501–511.

Maglott, D. (2004). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research*, 33(Database issue):D54–D58.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., Kanapin, A., Lewis, S., Mahajan, S., May, B., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L., and D'Eustachio, P. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research*, 37(Database):D619–D622.

Nishimura, D. (2001). BioCarta. *Biotech Software & Internet Report*, 2(3):117–120.

Riegler, T., Nejabat, M., Eichner, J., Stiebellehner, M., Subosits, S., Bilban, M., Zell, A., Huber, W. W., Schulte-Hermann, R., and Grasl-Kraupp, B. (2015). PRO-Inflammatory Mesenchymal Effects Of The Non-Genotoxic Hepatocarcinogen Phenobarbital: A Novel Mechanism Of Anti-Apoptosis And Tumor Promotion. *Carcinogenesis*, page bgv135.

Römer, M., Eichner, J., Metzger, U., Templin, M. F., Plummer, S., Ellinger-Ziegelbauer, H., and Zell, A. (2014). Cross-platform toxicogenomics for the prediction of non-genotoxic hepatocarcinogenesis in rat. *PLoS ONE*, 9(5):e97640.

Skuta, C., Bartůněk, P., and Svozil, D. (2014). InCHlib - interactive cluster heatmap for web applications. *Journal of cheminformatics*, 6(1):44.

Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A., and Dudoit, S., editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Statistics for Biology and Health, chapter V, pages 397–420. Springer-Verlag, New York.

Thomson, J. P., Lempiäinen, H., Hackett, J. A., Nestor, C. E., Müller, A., Bolognani, F., Oakeley, E. J., Schübeler, D., Terranova, R., Reinhardt, D., Moggs, J. G., and Meehan, R. R. (2012). Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome. *Genome Biology*, 13(10):R93.

Thomson, J. P., Moggs, J. G., Wolf, C. R., and Meehan, R. R. (2014). Epigenetic profiles as defined signatures of xenobiotic exposure. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 764:3–9.

Unterberger, E. B., Eichner, J., Wrzodek, C., Lempiäinen, H., Luisier, R., Terranova, R., Metzger, U., Plummer, S., Knorpp, T., Braeuning, A., Moggs, J., Templin, M. F., Honndorf, V., Piotto, M., Zell, A., and Schwarz, M. (2014). Ha-ras and β-catenin oncoproteins orchestrate metabolic programs in mouse liver tumors. *International Journal of Cancer*, 135(7):1574–1585.

Wrzodek, C., Eichner, J., Buchel, F., and Zell, A. (2013). InCroMAP: integrated analysis of cross-platform microarray and pathway data. *Bioinformatics*, 29(4):506–508.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2393v1 | CC BY 4.0 Open Access | rec: 26 Aug 2016, publ: 27 Aug 2016