

OPEN

Hepamine - A Liver Disease Microarray Database, Visualization Platform and Data-Mining Resource

Timo Itzel^{1,2}, Melanie Neubauer¹, Matthias Ebert³, Matthias Evert⁴ & Andreas Teufel^{1,2*}

Numerous gene expression profiling data on liver diseases were generated and stored in public databases. Only few were used for additional analyses by the hepatology research community. This may mostly be due to limited bioinformatics knowledge of most biomedical research personnel. In order to support an easy translation of bioinformatics data into translational hepatology research, we created Hepamine, a liver disease gene expression, visualization platform and data-mining resource. Microarray data were obtained from the NCBI GEO database. Pre-analysis of expression data was performed using R statistical software and the limma microarray analysis package from the Bioconductor repository. We generated Hepamine, a web-based repository of pre-analyzed microarray data for various liver diseases. At its initial release Hepamine contains 13 gene expression datasets, 20 microarray experiments and approximately 400 000 gene expression measurements. A self-explanatory website offers open and easy access to gene expression profiles. Results are furthermore visualized in simple three-color tables indicating differential expression. All data were linked to common functional and genetic databases particularly through the DAVID bioinformatics suite. Hepamine provides comprehensive data and easy access to hepatologic gene expression data even without in depth bioinformatics or microarray profiling experience. <http://www.hepamine.de>.

Chronic liver disease is a major health burden worldwide. In particular viral hepatitis B and C but also an increasing number of patients with non-alcoholic fatty liver disease/non-alcoholic steatohepatitis (NAFLD/NASH)^{1,2} in developed countries result in a steadily increasing number of patients with liver fibrosis, cirrhosis or hepatocellular carcinoma (HCC) as the common end stage of chronic liver disease³.

Recent advances in drug development for viral hepatitis will certainly be beneficial for many patients⁴. However, for patients with fatty liver disease, liver fibrosis, cirrhosis or liver cancer still only limited therapeutic options are available^{5,6}. Thus, investigating the molecular cause of many chronic liver diseases but also its common end stage fibrosis, cirrhosis and HCC are still urgently needed to provide better care for these patients³.

Over the past two decades, gene expression profiling of common liver diseases was extensively performed. Aim of this extensive survey of the liver transcriptome was the identification of “druggable” molecular targets and signaling networks to be targeted by specifically designed molecules. Even more, several groups aimed at using gene expression data particularly with respect to liver cancer in order to predict survival of patients or recurrence of this disease after resection. These efforts recently cumulated in the definition of consensus molecular subtypes (CMS) being helpful for prognosis prediction but also as an entry to precision medicine. As a result of these extensive efforts a vast amount of transcriptomic data was deposited in public databases and made publicly available. These data provide a rich source for further, comparative and integrative OMICS analysis to better understand the molecular basis of chronic liver disease. However, in order to utilize this information, knowledge of the available database resources but also bioinformatics tools are indispensable being a major hurdle in re-utilizing these data. Biomedical researchers are generally not capable of analyzing these data themselves due to a lack of programming skills. However, bioinformaticians are rather focused on the development of novel analysis algorithms.

In order to overcome this significant gap in knowledge transfer from molecular biology and bioinformatics to translational medical and ultimately clinical research, we provide Hepamine, a liver disease microarray database,

¹Division of Hepatology, Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ²Division of Bioinformatics, Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ³Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ⁴Department of Pathology, University of Regensburg, Regensburg, Germany. *email: andreas.teufel@medma.uni-heidelberg.de

Dataset	Array Platform		Label	Disease	Comparison	
	GEO	Annotation			Probe	Control
GSE11190	GPL570	hgu133plus2.db	E-GEOD-11190 HCV	HCV	Post-interferon liver biopsy	Pre-interferon liver biopsy
GSE14323	GPL571	hgu133a2.db	E-GEOD-14323 HCC	HCC	HCC	Normal
			E-GEOD-14323 cirrhosis	Fibrosis Cirrhosis	cirrhosis	Normal
GSE17548	GPL570	hgu133plus2.db	E-GEOD-17548 HCC	HCC	Ankara HCC tissue, Izmir HCC tissue	Ankara cirrhosis tissue, Izmir cirrhosis tissue
			E-GEOD-17548 HCC subset Ankara	HCC	Ankara HCC tissue	Ankara cirrhosis tissue
			E-GEOD-17548 HCC subset Izmir	HCC	Izmir HCC tissue	Izmir cirrhosis tissue
			E-GEOD-17967 HCC	HCC	HCV + cirrhosis with HCC	HCV + cirrhosis without HCC
GSE26566	GPL6104	hgu133a2.db	E-GEOD-26566 CCC	CCC	Cholangiocarcinoma	Normal intrahepatic bile duct
GSE32225	GPL8432		E-GEOD-32225 CCC	CCC	Human intrahepatic cholangiocarcinoma	Human normal biliary epithelial cells
GSE32958	GPL6244	hugene10sttranscriptcluster.db	E-GEOD-32958 CCC	CCC	Intrahepatic Cholangiocarcinoma (ICC)	Non-Tumor Tissue
	GPL6244	hugene10sttranscriptcluster.db	E-GEOD-32958 FNH	FNH	Focal Nodular Hyperplasia (FNH)	Non-Tumor Tissue
GSE33814	GPL6884		E-GEOD-33814 NASH	NAFLD NASH	steatohepatitis	normal
			E-GEOD-33814 Steatosis	NAFLD NASH	steatosis	normal
GSE38941	GPL570	hgu133plus2.db	E-GEOD-38941 HBV	HBV	HBV-associated acute liver failure	normal
GSE45001	GPL14550		E-GEOD-45001 CCC	CCC	Tumoral	Non-Tumoral
GSE46960	GPL6244	hugene10sttranscriptcluster.db	E-GEOD-46960 Biliary	Biliary Atresia	Biliary Atresia	Normal control
GSE49541	GPL570	hgu133plus2.db	E-GEOD-49541 Fibrosis	Fibrosis Cirrhosis	advanced (fibrosis stage 3–4)	mild (fibrosis stage 0–1)
GSE56140	GPL18461		E-GEOD-56140 HCC	HCC	hepatocellular carcinoma	cirrhosis
			E-GEOD-56140 HCC central	HCC	tumor central	cirrhosis
			E-GEOD-56140 HCC peripheral	HCC	tumor peripheral	cirrhosis

Table 1. Currently available pre-analyzed data sets.

visualization platform and data-mining resource, which provides easy but still detailed access to transcriptomic data on liver disease.

Material and Methods

Microarray data were obtained from the NCBI GEO Gene Expression Omnibus Archive of Functional Genomics Data (<https://www.ncbi.nlm.nih.gov/geo/>)⁷. Pre-analysis of expression data was performed using R statistical software and NCBI GEO's geo2r scripts provided for each data set (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)⁸. For reasons of comparing results between different experiments, all analysis scripts were adjusted to apply the newest array annotations using the hgu133plus2.db, hgu133a2.db, hgu133a.db, or hugene10sttranscriptcluster.db packages provided by the Bioconductor suite (<https://www.bioconductor.org/>)⁹ (Table 1). Next, all spots representing a specific gene were summarized by means of calculation an average expression in order to obtain a gene centered result. For further analysis, geo2r scripts were based on the Bioconductor limma package. Expression data were stored locally in a MySQL database. The business logic for visualization and filtering of the data was implemented on a python based Django 2.0 framework.

In order to support a functional characterization of the result tables, we provide predefined functionally related gene sets, e.g. signaling pathways (Suppl. Table 1). Those (KEGG) pathways were obtained from MSigDB provided by Broad Institute¹⁰.

WikiPathways¹¹ were obtained from the collaborative WikiPathways platform at https://www.wikipathways.org/index.php/Download_Pathways.

Included data sets were selected as follows: Firstly, all data sets related to chronic liver disease were selected from the GEO public functional genomics data repository (<https://www.ncbi.nlm.nih.gov/geo/>). Secondly, only data sets with GEO2R available scripts were selected for further processing. Thirdly, the analysis script utilized provided series matrix files. If data sets did not contain such a series matrix file, the respective data were not further processed or integrated in our database. Finally, dual-channel arrays were also excluded from these analyses as their analysis algorithm does not fit the used analysis scripts.

Web interface. After selecting the “Search Hepamine” option on the landing page, users will be guided to an initial data overview. This page contains all available data sets and lists all genes recognized in at least one of the available data sets. Primary data view displays gene expression as a three-way, traffic light visualization. Genes with a green arrow were upregulated, gene expression displayed as a yellow button was unchanged, and genes

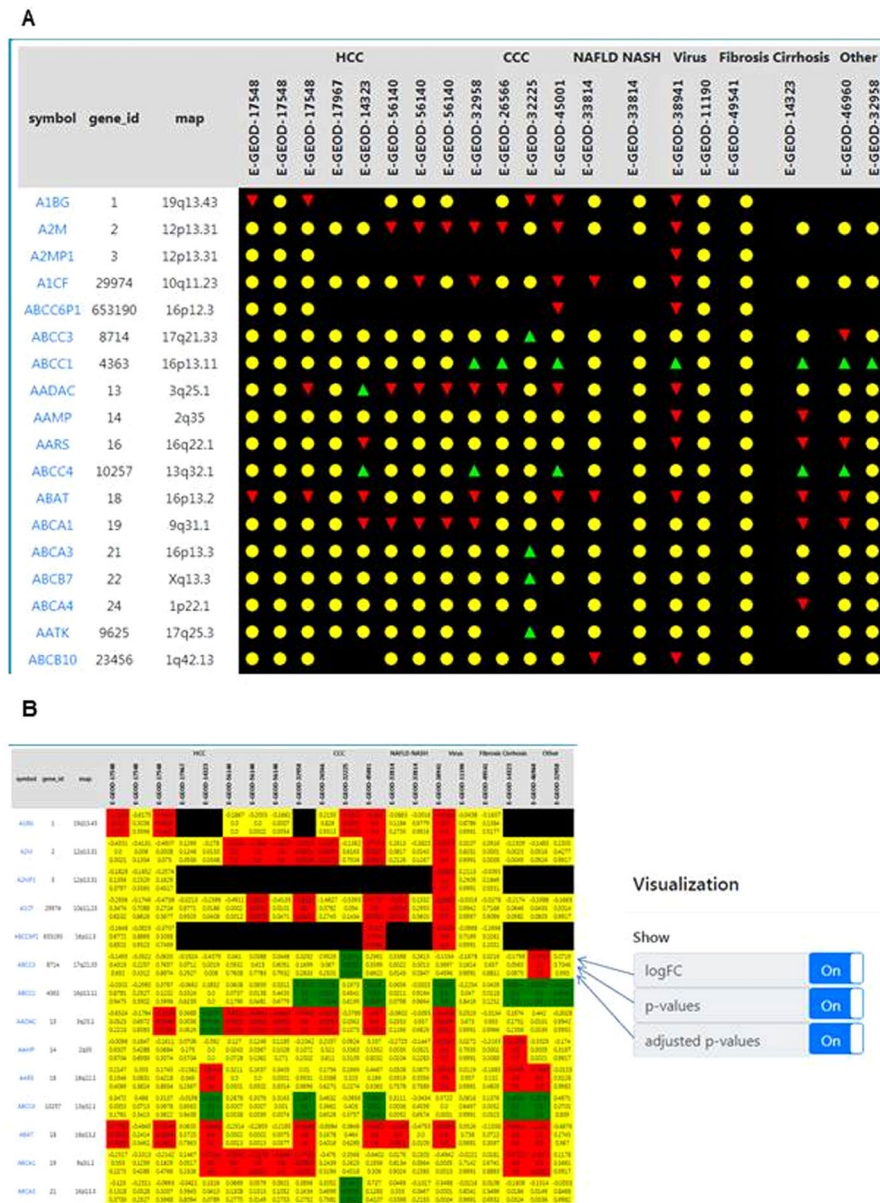


Figure 1. Hepamine visualization options of gene expression. (A) Simplified three-colour traffic light scheme: ▲ up-regulated, ● unchanged, ▼ down-regulated (B) Custom view by selection from fold change, p-value and adjusted p-value.

with a red arrow down were down regulated. This display offers an easy digestible overview of the excessive data accumulated (Fig. 1A).

However, as advanced users may be interested in more details about the differential gene expression in terms of fold-change, p-value, or adjusted p-value, Hepamine offers an alternative data view incorporating all this information for each gene in every data set. This alternative view may be selected by pressing the “Configure” button and selecting the respective features from the data filter panel on top of the page. Users may then select which features should be displayed selecting from fold change, p-value, and adjusted p-value. (Fig. 1B).

Data selection options. As users may be interested in a particular set of genes, we offer diverse options to limit Hepamine searches to genes, gene sets, or diseases of interest. Selecting the “Search” button, users may insert a list of genes in which they are interested in and have the output results limited to only these genes. Given a total of up to 25 185 genes per analysis results pages (searching the full genome annotated genes), this option performs focused searches of our liver disease specific data. If the user is interested in the data of only one of the displayed datasets, there is a link to the analysis view. There the data are shown as table or volcano plot (Fig. 2).

Alternatively users may upload a gene list of interest via the same site. Data need to be formatted as a.txt file and contain official gene symbol abbreviations as defined by the HUGO Gene Nomenclature Committee (HGNC)¹². Besides, we provide predefined gene sets (e.g. signaling pathways, gene ontologies) from the Kyoto

A

E-GEOD-14323 HCC					
symbol	gene_id	logFC	pval	padj	
JCHA2N	3512	3.02993408488842	8.69166018425036e-08	1.41961703242145e-06	
KCNJ16	3773	1.26970135257895	0.000626583631546926	0.00237060968722778	
KCNV2	3781	-1.90354254923665	3.8483800856457e-15	1.4955683836212e-13	
KCTD12	115207	1.26893907967104	1.05788128057861e-06	1.09449347797427e-05	
KSAAD000	9703	-1.0351514351579	1.46305834277679e-14	3.9342177629302e-12	
KLHL2	11275	-1.11980968971053	9.7679281994325e-12	6.88293383468868e-10	
KRT23	25984	2.52056739013158	8.09507229176925e-12	5.99764192492654e-10	
LACTB2	51110	-1.23326109660927	2.10093169993963e-06	1.9552540292994e-05	
LAPTM5	7805	1.26423926832884	5.45073024981242e-10	1.8186649005365e-08	
LEH	81606	1.3436174545	5.76174628622932e-13	6.74622481019125e-11	
LEP	3929	-1.43351307776316	3.47283834415186e-07	4.22779798488726e-06	
LDH8	3945	1.32607621280283	6.89180056541062e-09	1.51961758566394e-07	
LECT2	3950	1.48434099926315	5.01240242841256e-06	4.05039361647971e-05	
LGALS1	3956	1.34836037392104	3.54100838219036e-10	1.12245920231097e-08	
LGALS3	3958	3.09951617881578	7.16216054235854e-22	4.45343142523854e-18	
LGALS3BP	3959	1.77706824965789	3.38177171331427e-08	5.87115227180612e-07	
LGALS4	3960	2.38654606816421	8.12673846301308e-11	3.82818634566298e-09	
LHP	10186	1.5588944129473	1.107820020954532e-08	2.27340793541215e-07	
LIPA	3988	1.02922300631579	1.436757584878e-09	4.10747924725123e-08	
LIPG	9388	-1.0519513903138	1.37032617774713e-05	9.58457812287023e-05	
LOXL1	4016	1.00521344105263	2.64687608606279e-09	6.88830774189881e-08	
LPA	4018	-2.26261406234211	1.2186953840897e-12	1.23213414793935e-10	
LPAR6	10161	1.35471353489474	1.56438409013687e-10	5.06303515557417e-09	
LRP1	55791	-1.05390986147368	4.8252190482933e-06	3.9954535940662e-05	
LUM	4050	3.5552812456842	1.8968854652821e-21	7.8832254916391e-18	
LXN	56925	1.83897948723684	1.72888482054522e-08	3.34509924316204e-07	

Statistics

Number of features:
396

B

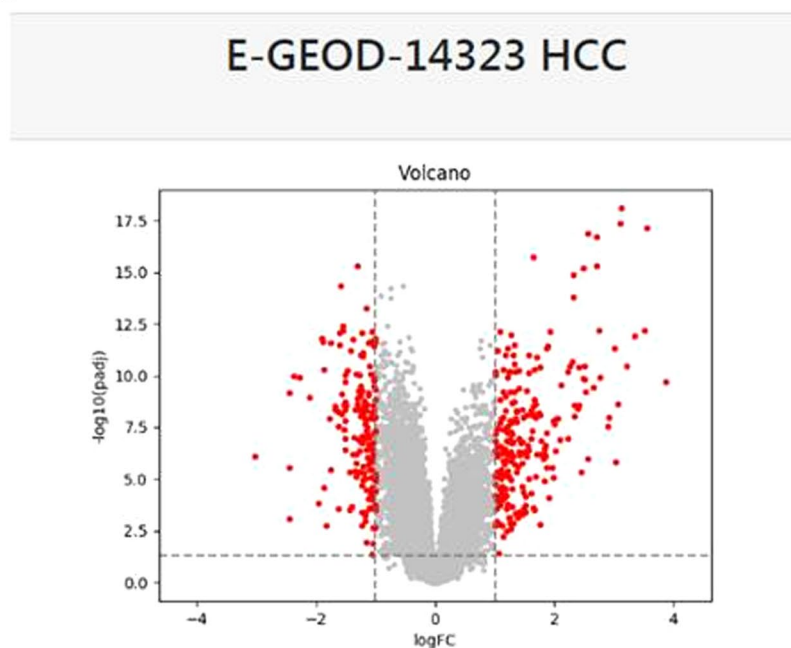


Figure 2. Additional data visualization options (analysis mode): (A) Result table, differential gene expression (B) Volcano plot.

Encyclopedia of Genes and Genomes (KEGG)¹³ or WikiPathways¹¹ to specifically view functional related genes of interest.

Finally, selecting the “Visualize Context” option, users may limit their searches to any disease of interest: HCC, CCA, NAFLD/NASH, viral hepatitis, fibrosis/cirrhosis, or other.

Data filtering options. Advanced users may change the selection criteria of significance within the data evaluation process. Again, the data filter bar on top of the page offers detailed selection options for fold change and adjusted p value of any preferred values.

Also, as some users may find it useful to perform more focused data searches, the top data filtering bar offers the option to select specific data sets such as HCC, CCA, viral hepatitis, NAFLD/NASH, fibrosis/cirrhosis, or other liver diseases such as biliary atresia.

Functional data analysis. Having retrieved a list of genes and their differential expression in one or even more liver disease, a user may likely be interested in further characterizing their selected group of genes, e.g. by means of pathway analysis or gene set enrichment. Enabling such access to deeper transcriptional information was achieved by linking Hepamine to the DAVID (database for annotation, visualization and integrated discovery) bioinformatics analysis platform^{14,15}. DAVID allows for detail information retrieval of gene annotation, visualization of results and integrated functional analysis. Upon selection of the “DAVID transfer” button, the gene list of interest/results are automatically transferred and DAVID results presented in a separate browser tab. Since the DAVID API only allows a fully automatic transfer of up to 400 genes, larger gene lists must be transferred in a semi-automated fashion. By activating the “DAVID transfer” button, users will have their gene list saved to the “clipboard”, obtain a detailed instruction on how to transfer their data, and be able to simply paste their selected gene list/content to the DAVID website. From thereon, users may continue using the DAVID website just as if they had used an automatic data transfer.

Use case. As a potential use case for Hepamine we compared differential gene expression among different data sets of liver cancer (HCC) development in order to identify potential drivers of hepatocarcinogenesis being showing robust differential expression among all data sets.

Initial step was selection of all HCC related data sets from the “Diseases” panel. This selection resulted in a subset table of 4 available HCC datasets with a total of 8 subsets (Fig. 3A). As a default setting, gene expression profiles were depicted color coded: green arrow up meaning up regulation, red arrow down meaning down regulation, yellow circles signaling no changes in differential gene expression (Fig. 3B). The available GSE numbers of the available datasets were then used to confirm the corresponding experimental description. Switching back to the results table significance level for differential gene expression were set to 4-fold regulation (\log_2 -fold 2) and a p-value of 0,05 (Fig. 3C). This search returned 223 genes being selected by these criteria.

For an even more stringed evaluation statistics adjusting for multiple testing was turned on selecting adjusted p-values “On”, at the right of the p-value field. Applying these parameters, we identified 175 genes being differentially expressed in at least one data set. By pressing the “Configure” button the visualization options become available. Selecting “logFC”, “p-values”, and “adjusted p-values” to be turned “On” resulted in a very detailed data table showing all important parameters fold-change, p-value, and adjusted p-value (Fig. 3D).

In order to get a functional assessment of these genes we automatically transferred these genes to DAVID via direct data transfer provided by Hepamine. Subsequently, these data were immediately available for further functional analysis to be selected from the DAVID Annotation summary results.

Discussion and Conclusions

As outlined above, gene expression profiling of common liver diseases was extensively performed over the past two decades. Aim of most studies on the liver transcriptome were the identification of “druggable” molecular targets, identification of prognostic subgroups, or predictive markers to accompany the entry into an era of precision medicine. Furthermore, multiple studies for liver cancer aimed at using gene expression data particularly to predict survival of patients or recurrence of the disease after resection¹⁶.

However, the analysis of the large amount of publicly available data on gene expression in various liver diseases seems not feasible for biologists or physicians not familiar with bioinformatics or at least microarray analysis. Even with profound experience in microarray experiments, analysis of such data is a complex and time-consuming task. We therefore implemented a novel database resource, Hepamine, to aid fast access to key hepatological expression data sets for a broader community.

We expect the Hepamine database to be highly useful in correlating *in vitro/ex vivo* models of disease to *in vivo* patient data. Furthermore, comparison of gene expression in diverse liver diseases and progressive liver disease, e.g. hepatitis, liver fibrosis, and liver cirrhosis will be possible with Hepamine. Finally, as demonstrated in the supplied use case, robustness of gene expression data may be increased by comparing several data sets with the same disease. If being up- or downregulated in several data sets the genes can be much more trusted to be a true driver of disease development rather than being just a biologically not meaningful by-stander¹⁷.

Since our database is the first of its kind and a novel tool in hepatological research, we set high value on a user friendly but also high performance usability for advanced bioinformatics analyses. To guarantee easy data access and connectivity we transformed this data base into a powerful web application. However, since requirements on such a web application may vary significantly depending on user’s needs, we offer two different data views. Traffic light visualization gives a fast and easy to digest overview on large gene expression data. It furthermore gives a quick impression on the consistency of expression data from different data sets reflecting gene expression changes of the same disease. In contrast, for users in need of a more detailed summary on the observed differential gene expressions, our data view displays all important numeric information for each gene and in every data set. This may be of interest in selecting genes with a high fold change or a strong p-value as it may be argued that expression changes of those genes have a higher probability of being robust in additional experiments to be potentially performed by the user. Those selections may furthermore be supported by means of selecting a higher stringency from the filter selection, increasing fold-change and/or p-value.

As advanced users may also want to further investigate a gene list consisting of genes being differentially expressed in a subset of liver diseases of interest we chose to automatically link our website to the DAVID bioinformatics resources consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists^{14,15}. The DAVID web resource is also publicly available

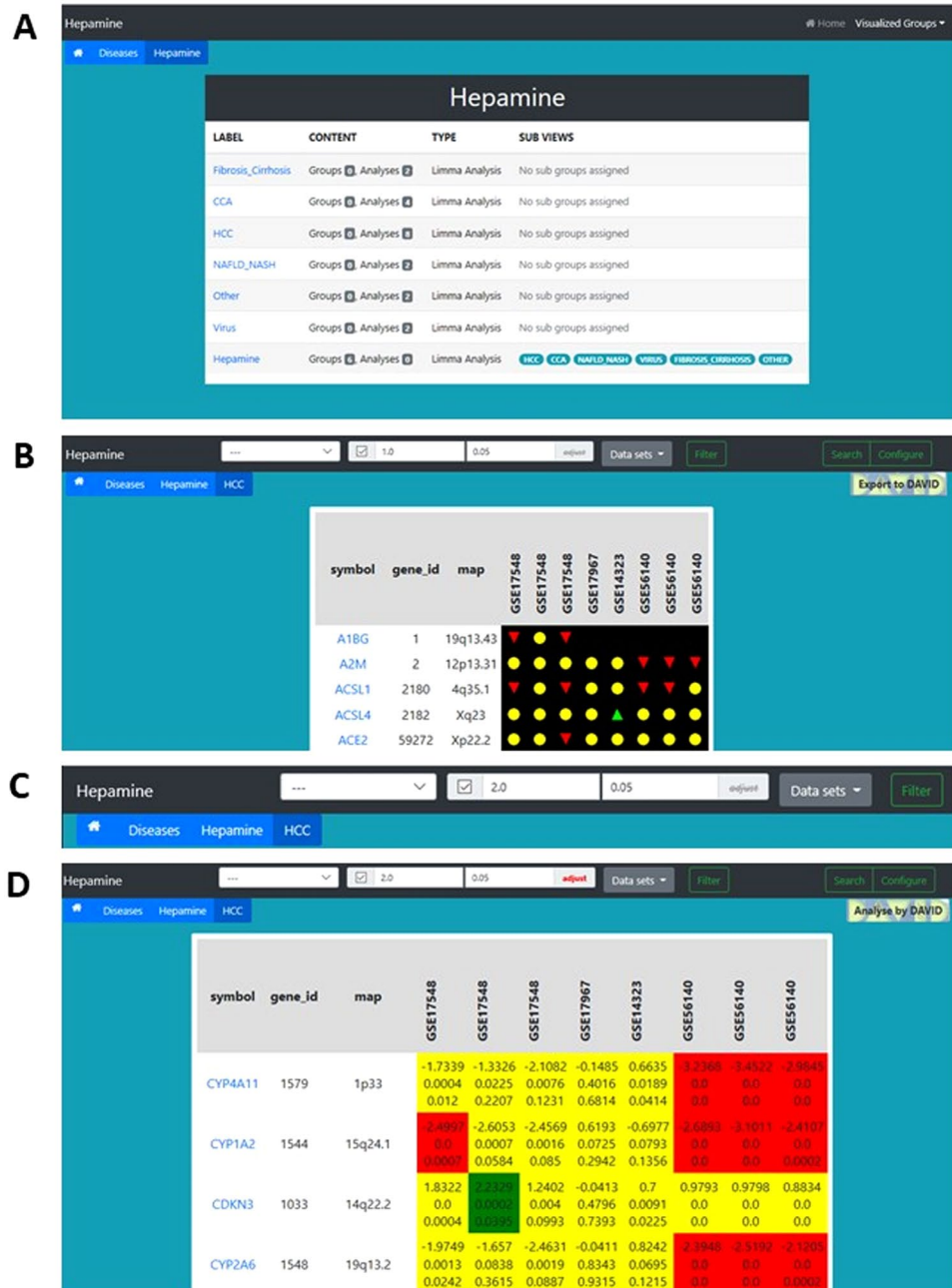


Figure 3. Screenshots of main analysis steps of the described use case HCC analysis in Hepamine. (A–D) Analysis in Hepamine.

and contains approximately 68 bioinformatics enrichment tools. Among those functions are gene functional classification tools, functional annotation charts or clustering and functional annotation tables^{14,15}.

DAVID supports automatic transfer of genes lists only up 400 genes. In order to at least ease transfer of larger genes lists our website supports a semi-automated fashion selecting a DAVID transfer button and subsequently copied into the working storage. Thus, users may easily paste their list into DAVID and continue using the website.

However, our web tool also has some short comings. Thus, it is obvious that microarray data are limited to genes known and annotated at the time of data production. In contrast, RNA sequencing data include additional unknown transcripts that could be later characterized and annotated. In order to ensure highest possible cooperativity among the diverse experiments performed under most likely highly different conditions, we chose at this stage only to incorporate data sets that could be evaluated using a Geo2R script provided by the GEO data repository⁸ and thus limited our selection to microarray data.

The fact that the selected experiments and respective data sets were not hybridized under the same or at least comparable conditions is also a major drawback of our data collection. However, at this point there is no large data set on liver diseases available (except for liver cancer, TCGA data) and our assembled gene expression array

provides the most comprehensive overview. Given the diverse experimental modalities, users must be aware of the limited comparability of the selected data sets. Given these limitations our database is well suited for hypothesis generation – as demonstrated in the supplementary use case - but definitely not for validation of otherwise obtained hypotheses.

In summary, we implemented the Hepamine gene expression web-resource, which is easy to use but yet contains comprehensive data to close the knowledge gap from molecular biology and bioinformatics to translational medical and ultimately clinical research on liver diseases.

Received: 28 May 2019; Accepted: 13 February 2020;

Published online: 16 March 2020

References

- Vernon, G., Baranova, A. & Younossi, Z. M. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment. Pharmacol. Ther.* **34**, 274–285, <https://doi.org/10.1111/j.1365-2036.2011.04724.x> (2011).
- Armstrong, M. J. *et al.* Presence and severity of non-alcoholic fatty liver disease in a large prospective primary care cohort. *J. Hepatol.* **56**, 234–240, <https://doi.org/10.1016/j.jhep.2011.03.020> (2012).
- Liou, I. W. Management of end-stage liver disease. *Med. Clin. North. Am.* **98**, 119–152, <https://doi.org/10.1016/j.mcna.2013.09.006> (2014).
- Manns, M. P. & von Hahn, T. Novel therapies for hepatitis C - one pill fits all? *Nat. Rev. Drug. Discov.* **12**, 595–610, <https://doi.org/10.1038/nrd4050> (2013).
- Schuppan, D. & Schattenberg, J. M. Non-alcoholic steatohepatitis: pathogenesis and novel therapeutic approaches. *J. Gastroenterol. Hepatol.* **28**(Suppl 1), 68–76, <https://doi.org/10.1111/jgh.12212> (2013).
- Worns, M. A. & Galle, P. R. HCC therapies—lessons learned. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 447–452, <https://doi.org/10.1038/nrgastro.2014.10> (2014).
- Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–995, <https://doi.org/10.1093/nar/gks1193> (2013).
- Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinforma.* **23**, 1846–1847, <https://doi.org/10.1093/bioinformatics/btm254> (2007).
- Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121, <https://doi.org/10.1038/nmeth.3252> (2015).
- Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425, <https://doi.org/10.1016/j.cels.2015.12.004> (2015).
- Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667, <https://doi.org/10.1093/nar/gkx1064> (2018).
- Yates, B. *et al.* Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.* **45**, D619–D625, <https://doi.org/10.1093/nar/gkw1033> (2017).
- Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, <https://doi.org/10.1093/nar/28.1.27> (2000).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
- Marquardt, J. U., Galle, P. R. & Teufel, A. Molecular diagnosis and therapy of hepatocellular carcinoma (HCC): an emerging field for advanced technologies. *J. Hepatol.* **56**, 267–275, <https://doi.org/10.1016/j.jhep.2011.07.007> (2012).
- Liang, L. *et al.* Validation of a multi-omics strategy for prioritizing personalized candidate driver genes. *Oncotarget* **7**, 38440–38450, <https://doi.org/10.18632/oncotarget.9540> (2016).

Author contributions

T.I. and A.T.: Data analysis, database set up, online presentation; M.N., M.E. and M.E.: critical data review, preparation of manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-61508-y>.

Correspondence and requests for materials should be addressed to A.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020