Published in final edited form as:

Liver Int. 2014 April ; 34(4): 621–631. doi:10.1111/liv.12292.

CellMinerHCC: a microarray-based expression database for hepatocellular carcinoma cell lines

Frank Staib^{1,*}, Markus Krupp^{2,*}, Thorsten Maass³, Timo Itzel³, Arndt Weinmann¹, Ju-Seog Lee⁴, Bertil Schmidt², Martina Müller³, Snorri S. Thorgeirsson⁵, Peter R. Galle¹, Andreas Teufel³

¹Department of Medicine I, Johannes Gutenberg University, Mainz, Germany

²Institute of Computer Science, Johannes Gutenberg University, Mainz, Germany

³Department of Medicine I, University of Regensburg, Regensburg, Germany

⁴Kleberg Center for Molecular Markers, Department of Systems Biology, MD Anderson Cancer Center, Houston, Texas, USA

⁵Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, USA

Abstract

Background & Aims: Therapeutic options for hepatocellular carcinoma (HCC) still remain limited. Development of gene targeted therapies is a promising option. A better understanding of the underlying molecular biology is gained in *in vitro* experiments. However, even with targeted manipulation of gene expression varying treatment responses were observed in diverse HCC cell lines. Therefore, information on gene expression profiles of various HCC cell lines may be crucial to experimental designs. To generate a publicly available database containing microarray expression profiles of diverse HCC cell lines.

Methods: Microarray data were analyzed using an individually scripted R program package. Data were stored in a PostgreSQL database with a PHP written web interface. Evaluation and comparison of individual cell line expression profiles are supported via public web interface.

Results: This database allows evaluation of gene expression profiles of 18 HCC cell lines and comparison of differential gene expression between multiple cell lines. Analysis of commonly regulated genes for signaling pathway enrichment and interactions demonstrates a liver tumor phenotype with enrichment of major cancer related KEGG signatures like ‘cancer’ and ‘inflammatory response’. Further molecular associations of strong scientific interest, e.g. ‘lipid metabolism’, were also identified.

Conclusions: We have generated CellMinerHCC (<http://www.medicalgenomics.org/cellminerhcc>), a publicly available database containing gene expression data of 18 HCC cell lines.

Correspondence Andreas Teufel, MD, PhD, Department of Medicine I, University of Regensburg, Franz-Josef-Strauss Allee 11, Regensburg 93053, Germany, Tel: +49 941 9447006, Fax: +49 941 9447002, andreas.teufel@ukr.de, Received 9 October 2012
Accepted 24 July 2013, DOI:10.1111/liv.12292.

*Both authors contributed equally to this work.

This database will aid in the design of *in vitro* experiments in HCC research, because the genetic specificities of various HCC cell lines will be considered.

Keywords

bioinformatics; HCC; liver cancer; oncogenomics; systems biology

Hepatocellular carcinoma (HCC) is among the most common malignancies worldwide and its incidence is rising, especially in Asia and Sub-Saharan Africa, but also in Western countries. Simultaneously, the therapeutic options for this disease besides surgery still remain limited. Although both cellular changes that lead to HCC and the aetiological factors responsible for the majority of HCC cases have been recognized, the molecular pathogenesis of this disease still remains elusive (1). However, key to achieve further progress in the therapy of HCC will rely on a better understanding of the underlying biology of HCC development and growth allowing the development of subsequent targeted therapies against essential molecular mechanisms.

In vitro cell culture experiments provide the opportunity of modelling the complex mechanisms of HCC development. Thus, *in vitro* experiments may be used to easily test gene expression and corresponding biological behaviour of HCC cells in order to identify genetic signatures that are related to tumour growth and proliferation and may therefore be potential therapeutic targets. Furthermore, these cells are used in xenograft models of HCC development in mice in order to study tumour development and biological behaviour *in vivo*.

Although these *in vitro* and *in vivo* approaches using immortalized tumour cell lines have provided some insights in the molecular mechanisms underlying HCC development, it was simultaneously and repeatedly noted that diverse cell lines show diverse biological behaviour in response to drug treatment or genetic manipulation (2).

These diverse forms of behaviour and underlying genetic signatures made it necessary to take these differences in gene expression signatures into account, especially while planning targeted strategies to influence the biological behaviour. For example, a potential target of interest may be expressed very low in a particular cell line and therefore not be suitable as a therapeutic target whereas other cell lines may exhibit a strong expression, which may very well be blocked by therapeutic strategies in order to slow down tumour growth or migration.

Being aware of these differences may be crucial to the outcome of an experiment and therefore should be of importance in experimental design. Thus, information on the diverse gene expression profiles of the HCC cell lines may be crucial to experimental designs of modelling HCC *in vitro*.

The completely assembled human genome provides the opportunity to approach this issue with a large amount of genetic information and high-throughput genomic data (3). However, analysis of these data is not feasible for biologists or physicians not familiar with bioinformatics or at least microarray analysis. Even with a profound experience in microarray analysis, analysis of such data may not be done easily and would take a considerable amount of time. In conclusion, the available gene expression profiles of HCC

cell lines must therefore be considered as not accessible for the very most of the hepatological community. As a result, gene expression profiles of common HCC cell lines are still not considered in most publications on *in vitro* models studying molecular mechanisms of HCC development.

We have, therefore, established the CellMinerHCC database providing easy access to the microarray (raw) data on the 18 most commonly used HCC cell lines. Also, within this publicly accessible database, these data were made easily comparable in order to assist in future designing of *in vitro* HCC modelling.

Material and methods

Data source

Gene expression data were kindly provided by Lee and Thorgeirsson, Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, USA. The data include 18 HCC cell lines and normal liver samples. The HCC cell lines are the following: 7703, Focus, Hep3B, Hep3B-TR, Hep40, HepG2, HLE, HLF, HUH-1, HUH-6, HUH-7, PLC/PRF/5, SK-Hep1, SNU-182, SNU-387, SNU-389, SNU-449 and SNU-475. The HCC cell lines were handled as previously described (4). As control for dual channel microarrays, a pool of total RNA from 19 normal liver samples was used (5). Oligo microarrays were produced at the Advanced Technology Center at the National Cancer Institute, NIH, USA using 70-mer probes of 21 329 genes. Microarray experiments were performed as previously described (5), with the exception of dye swap experiments that were not included in this data set.

Data organization, webinterface

The genetic profile of the 18 HCC cell lines were stored in a PostgreSQL database (<http://www.postgresql.org>) and made publically accessible and searchable through a web interface implemented in PHP (<http://de.php.net>) (Fig. 1), as previously described (6). The data may also be downloaded as text files. Furthermore, CellMinerHCC data were linked to multiple well established databases such as NCBI Entrez database (7), HGNC (8), HPRD (9), OMIM (10), BioGPS (11), Nexbio (12) and Gent (13). Furthermore, the data were connected to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (14) and the Gene Ontology (GO) database (15) to enable an integrative comparison and querying for functional, molecular and signalling events. Moreover, CellMinerHCC is linked to the publically available, web-based tool DAVID ('Database for Annotation, Visualization and Integrated Discovery'; <http://david.abcc.ncifcrf.gov/home.jsp>), enabling further functional analysis (16, 17). Finally, the CellMinerHCC was cross-linked to our expression profiling database on normal tissues by next-generation sequencing database RNA-Seq Atlas (18) as well as our liver-specific databases on molecular associations LoMA (6).

Functional analysis

For annotation of commonly regulated genes across all 18 HCC cell lines with GO-terms, KEGG pathways, and SP-PIR Keywords, as defined by the SwissProt/Uniprot and PIR groups, DAVID has been used with default settings (16, 17).

Ingenuity Pathway Analysis (<http://www.ingenuity.com/>) was performed using 195 commonly regulated genes (M -value >1 , respectively <-1) across all 18 HCC cell lines applying default settings.

Results

CellMinerHCC database system

Diverse biological behaviour of HCC cell lines may result from diverse underlying genetic profiles and expression signatures, which may differ significantly among immortalized HCC tumour cell lines. The awareness of these differences made it necessary to take the diverse gene expression signatures into account, especially while planning targeted strategies to influence the biological behaviour.

However, the analysis of these data does not seem to be feasible for biologists or physicians not familiar with bioinformatics or at least microarray analysis. Even with profound experience in microarray analysis, analysis of such data is a complex and time-consuming task. We have, therefore, established the CellMinerHCC database providing easy access to the microarray (raw) data on the 18 most commonly used HCC cell lines. Currently, this database holds genome wide expression profiles of the 18 most commonly used HCC cell lines namely 7703, Focus, Hep3B, Hep3B-TR, Hep40, HepG2, HLE, HLF, HUH-1, HUH-6, HUH-7, PLC/PRF/5, SK-Hep1, SNU-182, SNU-387, SNU-389, SNU-449 and SNU-475. For all cell lines, our database contains the expression data of 21 329 genes. As some genes were represented by more than only one probe, expression data of all probes representing the same gene were averaged. In order to provide a rapid overview over the expression of individual genes in multiple cell lines, we furthermore colour coded the gene expression profiles. Mouse over the individual spots allows then to retrieve the detailed expression values for each gene and cell line.

The CellMinerHCC database provides multiple search options to support complex genetic analyses. Firstly, CellMinerHCC offers the option to search for individual genes and their genetic profile. This search may be performed by means of a search for individual gene names or cell lines or even a combination of both from the search page of the CellMinerHCC drop down menu. Also, more complex searches may be performed by a genetic pathway from KEGG (14), or a gene ontology from the 'explore genetic association' panel, providing a highly detailed search option. These search options together offer valuable complex analysis options (Fig. 1). For example, one can now, for the first time, easily select all genes associated with the Wnt signalling pathway and display their expression profile as an over view in the 18 HCC cell lines. Furthermore, one could search all gene products located to the nucleus by means of gene ontologies in order to identify transcription factors associated with the development of HCC and also display their expression profiles as a consolidated overview in all 18 HCC cell lines. In addition, the data search site offers the option of a free text search in order to provide a maximum flexibility in designing database queries (Fig. 2). After executing a search, the result page for these searches offers the genetic associations to individual diseases if present. Mainly, the results page provides a summary on gene name, corresponding NCBI Gene ID (7), Ensembl Gene ID (19), and most importantly the expression profiles of the gene in selected cell lines (Fig.

3). Additionally, the DAVID (16, 17) integration enables the user to further analyse their resulting gene set. Besides, more details on the specific gene such as gene alias names, chromosomal location, information on biological function, participation in biological processes and subcellular localization by supplying gene ontology information, and associated genetic pathways are accessible through the details button and respective site for each gene (Fig. 4). This information may be of significant value in designing complex and highly selective queries to the database.

Since local availability of the data in this repository may significantly speed up high-throughput searches for interested users, we also provide data in flat files for a complete download.

A key issue in developing this database was to provide the hepatological community with a powerful but simultaneously highly reliable and comprehensive database to perform systems biology-based high-throughput searches and comparison of gene expression. Our database was linked to multiple other bioinformatics resources, providing valuable connections and supporting advanced search and evaluation strategies. These links offer a strong backbone for bioinformatics research on chronic liver disease. In detail, CellMinerHCC has been linked to the most commonly used bioinformatics databases, such as PubMed (7), the European Bioinformatics Institute Website Ensembl (19), the Bioinformatics Resource of the National Center of Biotechnology Information Entrez Gene (7), and the Gene Ontology database, holding such as multiple sequence information, microarray expression data, conserved domains, as well as information on a gene's function (20).

Differential gene expression in HCC cell lines

Evaluating all 18 HCC cell lines for differential gene expression with a cut-off of at least two-fold changes in gene expression, we identified between 1.638 genes in HepG2 and 3.214 genes in SNU-398 as being differentially regulated. A box-plot providing the distribution of differentially regulated genes with differentiation into up- and downregulated genes is provided in Figure S1. The pattern of genes being upregulated vs. genes being downregulated was mostly evenly distributed. Solely SNU-182 cells, significantly more genes were found to be downregulated as compared to genes upregulated. Distribution of *M*-values as a measure of change in gene expression followed a 'normal distribution curve' in all cell lines (Figure S2). Of these, 195 genes were identified as consistently either up- or downregulated with 163 genes being downregulated and 32 genes being upregulated (Table S1).

Biological functions of commonly regulated genes in HCC cell lines

Having established such a rich data resource, we were curious about the biological functions of these 195 most commonly regulated genes among these 18 cell lines they may hold key functions related to liver cancer development. In order to obtain a comprehensive overview on the biological functions of these genes, we performed several advanced bioinformatics analyses. The list of 195 commonly regulated genes was imported into DAVID (16, 17). The majority of significantly enriched GO-terms, KEGG pathways and SP-PIR Keywords are highly specific for liver and a malignant phenotype. Among the top 20 enriched categories

were ‘liver’ ($P = 6.5 \times 10^{-14}$), ‘plasma’ ($P = 2.0 \times 10^{-25}$), ‘complement and coagulation cascades’ ($P = 1.2 \times 10^{-14}$), ‘metalloprotein’ ($P = 1.5 \times 10^{-7}$), ‘extracellular space’ ($P = 1.1 \times 10^{-9}$) and ‘disease mutation’ ($P = 8.3 \times 10^{-8}$) (Table 1).

Furthermore, all commonly regulated genes were analysed by means of Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) (Fig. 5). As a proof of principle, the highest ranking disease terms associated with these commonly regulated genes in HCC cell lines were ‘cancer’ (76 out of 195 genes; $P = 3.8 \times 10^{-10}$ – 1.2×10^{-2}) followed by ‘gastrointestinal disease’ (54 out of 195 genes; $P = 3.8 \times 10^{-10}$ – 1.2×10^{-2}), ‘hepatic system disease’ (31 out of 195 genes; $P = 3.8 \times 10^{-10}$ – 1.2×10^{-2}) and ‘inflammatory response’ (23 out of 195 genes; $P = 9.2 \times 10^{-7}$ – 1.2×10^{-2}). In addition, other networks also demonstrated a high association with HCC cell lines as well. In particular, the lipid metabolism ranked as third most significantly enriched category (36 out of 195 genes; $P = 6.2 \times 10^{-9}$ – 1.2×10^{-2}). In concordance with the DAVID analysis pointing out a significant role of many genes in metabolic processes, these data point towards an important role of gene/protein members of the lipid metabolism functional pathway (21). Furthermore, top biofunctions contained several categories related to immune functions. Among the top 20 biological functions, ‘immunological disease’ (32 out of 195 genes; $P = 3.0 \times 10^{-7}$ – 1.2×10^{-2}), and ‘immune cell trafficking’ (19 out of 195 genes; $P = 9.2 \times 10^{-7}$ – 1.2×10^{-2}) were identified. As further, ‘diseases and disorders’ term ‘inflammatory response’ (23 out of 195 genes; $P = 9.2 \times 10^{-7}$ – 1.2×10^{-2}) was highly significant enriched. Thus, among the 20 highest ranked biological functions were three categories related to immune system and inflammatory responses indicating that immune system and inflammation may play an important role in this disease.

We were also interested in expression of hepatocellular cancer-like stem cell genes among those commonly regulated genes in all 18 HCC cell lines. Lee *et al.* previously described a hepatoblast-like HCC subgroup, which included the expression of hepatic oval cell marker genes like KRT7, KRT19 and VIM (22). These genes were not found among the 195 commonly regulated genes. However, in a CellMinerHCC search, these genes have easily been identified as differentially regulated across the 18 HCC cell lines. All three marker genes are upregulated in SK-Hep1 and SNU-182, while they are downregulated in SNU-387. A recent study by Woo *et al.* described an eight gene stem cell-like signature associated with TP53 mutations and poor prognosis in patients with HCC. Again, these genes were not listed among the 195 commonly regulated genes. A CellMinerHCC search identified six out of these eight genes as minor regulated (within the two-fold change) except for ES1 (corresponds with HES1 in CellMinerHCC), which is significantly downregulated in most of the cell lines.

Altogether, these data indicate that these 195 commonly regulated genes across all 18 HCC cell lines not only reflect typical liver biofunctions and a liver cancer phenotype, there are also multiple additional enriched functions that may play an important role in HCC.

Discussion

Genetic mutations and a variable genetic background have been demonstrated to significantly influence the development and course of HCC as well as the efficiency of its treatment with diverse drugs.

Over the past decades, multiple molecular mechanisms and individual factors have been shown to be involved in the development of HCC and it has become clear that development and course of this disease under-ly complex genetic interactions. We have, in the past, published a large genetic database summarizing genes known to be involved in HCC development (6). However, these data do not represent a genome-wide screen as they were collected through a text mining approach of the known literature.

However, to investigate these complex molecular interactions, data resources providing a comprehensive collection of differential gene expression data involved in the development of HCC are urgently needed. We therefore present a novel database resource targeted to be of significant aid in the complex modelling of HCC development and evaluation of treatment options *in vitro*.

In vitro cell culture experiments provide the opportunity of modelling the complex mechanisms of HCC development. Thus, *in vitro* experiments may be used to easily test gene expression and corresponding biological behaviour of HCC cells in order to identify genetic signatures that are related to tumour growth and proliferation and may therefore be potential therapeutic targets. However, the diverse genetic backgrounds and differential gene expression profiles of these cell lines were not publicly available. We therefore analysed the 18 most commonly used liver cancer cell lines by means of microarray analysis in order to provide the basis for a more specific selection of liver cancer cell lines for future design of experiments in molecular hepatology research. To our knowledge, this is one of the first databases of its kind summarizing genome-wide gene expression profiles of cell lines of a specific tumour type. Thus, comparative transcriptomics analyses to similar databases were not possible. However, we thought it would be of definite interest to evaluate the genes being consistently over- or underexpressed in all cell lines as they may be crucial to the biological mechanisms of liver cancer development.

The analysis by DAVID revealed several GO-terms, KEGG pathways and SP-PIR keywords describing liver tissue and many of its physiological functions, which may at least partially be involved in HCC (16, 21). Among these are ‘complement and coagulation’, a KEGG pathway that has been previously described to be possibly involved in HCC (23, 24), an ‘acute inflammatory response’ may be involved in carcinogenesis if an imbalance and prolongation occur (25), ‘disease mutation’ plays multiple roles in carcinogenesis in general, and ‘metalloprotein’ involved in the generation of the extracellular matrix may also be involved in HCC, when unequal distributions occur (26) and even more when their decomposing metalloproteinases are disturbed (27). These findings are complemented and further differentiated by Ingenuity Pathway Analysis. Here, the disease terms describe a significant enrichment and association of 76 out of 195 commonly regulated genes in all 18 cell lines with ‘cancer’. With 54 and 31 out of 195 commonly regulated genes, the terms

‘gastrointestinal disease’ and a ‘hepatic system disease’ provide a more general heading for HCC. Thus, many of these 195 genes have already been associated with cancerous diseases in the gastrointestinal tract. The disease term ‘inflammatory response’ is complementary to the DAVID analysis revealing the GO term ‘acute inflammatory response’. The link between ongoing acute inflammation and HCC has become increasingly tight (28, 29). Identification of these categories in this gene set is providing additional evidence for this association. Categories like ‘immunological disease’ connect the inflammatory response with the immune system that is involved in HCC. Enrichment of additional categories like ‘humoral immune response’ and ‘immune cell trafficking’ point towards the important role the immune system is playing in development and progression of HCC (30). Finally, the category ‘lipid metabolism’ has been identified as third highest ranked biological function when analysing the 195 commonly regulated HCC cell line genes. For aberrant lipid metabolism, an association with HCC has already been described (31) and may also play a role in obesity and chronic inflammation related development of HCC (21, 29). Taken together, analysis of the commonly regulated genes among the 18 most often used HCC cell lines for enrichment of signalling pathways, proteins and interactions not only described a liver tumour phenotype, it also identified molecular associations and numerous categories currently under intense scientific development.

Over the past decade, it has become increasingly clear that stem cells are not only beneficial and that tumours of various origins also contain stem cells that help them to proliferate but also to escape conventional chemo-therapy (32, 33). Such cancer stem cells were identified in multiple tumours. The existence of cancer stem cells in hepatocellular carcinoma still remains elusive. However, there is accumulating evidence for the concept of cancer stem cells in HCC similar to other solid tumours, where cancer stem cells have already been conclusively identified (34, 35). We therefore analysed our differentially regulated genes for the appearance of stem cell markers. However, among the 195 genes consistently regulated in all 18 HCC cell lines, we did not find any of the conventional stem cell markers. In addition, searching the 195 commonly expressed genes in all 18 HCC cell lines for expression of a set of 3 classical markers for hepatic oval cells (KRT7, KRT19 and VIM) that have previously been identified among a group of genes defining a hepatoblast HCC subgroup associated with poor prognosis did not reveal any hit (36). However, running a CellMinerHCC search for these three target genes revealed a differential expression pattern across all cell lines. The graphical output of CellMinerHCC made identification of those cell lines with up- or downregulation of all three marker genes easy indicating its usefulness and easy-to-use properties.

Since our database is the first of its kind and such a novel tool in HCC research, we set a high value on a user friendly but at the same time usability for advanced bioinformatics analyses. To guarantee easy data access and connectivity, we transformed this database into a powerful web application. Since multiple systems biology queries and analyses require complex connections between information from diverse databases, the provided links offer a strong backbone for bioinformatics research on chronic liver disease.

We furthermore used our novel genetic resource in HCC research to identify key issues in transformed HCC cell lines such as the enrichment of genetic signalling pathways or

biological functions within these complex transcriptomics queries. We therefore realized a rich embedding of our database into the current scenery of bioinformatics repositories providing valuable connections, which may support advanced search and evaluation strategies. The provided links to further bioinformatics repositories were selected as they may in addition support automated correlation with additional genomic information such as multiple sequence information, microarray expression data, conserved domains, as well as information on a gene's function. Since many users of a preliminary version of our database were interested in genetic pathway analysis, we made this information instantly available for advanced queries through a drop down menu at the front page of the data search option.

Altogether, CellMinerHCC is the first database providing a comprehensive view and analysis options for microarray data of the most commonly used HCC cell lines and may be of significant use for *in vitro* modelling of HCC. CellMinerHCC is freely accessible at <http://www.medicalgenomics.org/cellminerhcc>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Maass T, Sfakianakis I, Staib F, et al. Microarray-based gene expression analysis of hepatocellular carcinoma. *Curr Genomics* 2010; 11: 261–8. [PubMed: 21119890]
2. Herzer K, Hofmann TG, Teufel A, et al. IFN-alpha-induced apoptosis in hepatocellular carcinoma involves promyelocytic leukemia protein and TRAIL independently of p53. *Cancer Res* 2009; 69: 855–62. [PubMed: 19141642]
3. Hocquette JF. Where are we in genomics? *J Physiol Pharmacol* 2005; 56(Suppl. 3): 37–70.
4. Lee J-S, Thorgerirsson SS. Functional and genomic implications of global gene expression profiles in cell lines from human hepatocellular cancer. *Hepatology* 2002; 35: 1134–43. [PubMed: 11981763]
5. Lee J-S, Chu I-S, Heo J, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 2004; 40: 667–76. [PubMed: 15349906]
6. Buchkremer S, Hendel J, Krupp M, et al. Library of molecular associations: curating the complex molecular basis of liver diseases. *BMC Genomics* 2010; 11: 189. [PubMed: 20302666]
7. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2011; 39(Database issue): D38–51. [PubMed: 21097890]
8. Seal RL, Gordon SM, Lush MJ, Wright MW, Bruford EA. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res* 2011; 39(Database issue): D514–9. [PubMed: 20929869]
9. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human protein reference database–2009 update. *Nucleic Acids Res* 2009; 37(Database issue): D767–72. [PubMed: 18988627]
10. McKusick VA. Mendelian inheritance in man and its online version. *AmJ Hum Genet* 2007; 80: 588–604. [PubMed: 17357067]
11. Wu C, Orozco C, Boyer J, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 2009; 10: R130. [PubMed: 19919682]
12. Kupersmidt I, Su QJ, Grewal A, et al. Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS ONE* 2010; 5: e13066. [PubMed: 20927376]
13. Shin G, Kang T-W, Yang S, et al. GENT: gene expression database of normal and tumor tissues. *Cancer Inform* 2011; 10: 149–57. [PubMed: 21695066]
14. Aoki KF, Kanehisa M. Using the KEGG database resource. *Curr Protoc Bioinformatics* 2005; 1: 112. [PubMed: 18428742]

15. 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. [PubMed: 10802651]
16. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37: 1–13. [PubMed: 19033363]
17. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44–57. [PubMed: 19131956]
18. Krupp M, Marquardt JU, Sahin U, et al. RNA-Seq Atlas—a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics* 2012; 28: 1184–5. [PubMed: 22345621]
19. Spudich G, Fernández-Suárez XM, Birney E. Genome browsing with Ensembl: a practical overview. *Brief Funct Genomic Proteomic* 2007; 6: 202–19. [PubMed: 17967807]
20. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res* 2001; 11: 1425–33. [PubMed: 11483584]
21. Becker D, Sfakianakis I, Krupp M, et al. Genetic signatures shared in embryonic liver development and liver cancer define prognostically relevant subgroups in HCC. *Mol Cancer* 2012; 11: 55. [PubMed: 22891627]
22. Lee J-S, Heo J, Libbrecht L, et al. A novel prognostic sub-type of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med* 2006; 12: 410–6. [PubMed: 16532004]
23. Li T, Wan B, Huang J, Zhang X. Comparison of gene expression in hepatocellular carcinoma, liver development, and liver regeneration. *Mol Genet Genomics* 2010; 283: 485–92. [PubMed: 20358383]
24. Wurmbach E, Chen Y, Khitrov G, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* 2007; 45: 938–47. [PubMed: 17393520]
25. Schetter AJ, Heegaard NHH, Harris CC. Inflammation and cancer: interweaving microRNA, free radical, cytokine and p53 pathways. *Carcinogenesis* 2010; 31: 37–49. [PubMed: 19955394]
26. Liu Y, Li L, Gao Y, et al. Distribution of metalloproteins in hepatocellular carcinoma and surrounding tissues. *Hepatogastroenterology* 2007; 54: 2291–6. [PubMed: 18265650]
27. Vihinen P, Kähäri V-M. Matrix metalloproteinases in cancer: prognostic markers and therapeutic targets. *Int J Cancer* 2002; 99: 157–66. [PubMed: 11979428]
28. Castello G, Scala S, Palmieri G, Curley SA, Izzo F. HCV-related hepatocellular carcinoma: from chronic inflammation to cancer. *Clin Immunol* 2010; 134: 237–50. [PubMed: 19910258]
29. Toffanin S, Friedman SL, Llovet JM. Obesity, inflammatory signaling, and hepatocellular carcinoma—an enlarging link. *Cancer Cell* 2010; 17: 115–7. [PubMed: 20159605]
30. Budhu A, Forgues M, Ye Q-H, et al. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell* 2006; 10: 99–111. [PubMed: 16904609]
31. Wu J-M, Skill NJ, Maluccio MA. Evidence of aberrant lipid metabolism in hepatitis C and hepatocellular carcinoma. *HPB (Oxford)* 2010; 12: 625–36. [PubMed: 20961371]
32. Lee TKW, Castilho A, Ma S, Ng IOL. Liver cancer stem cells: implications for a new therapeutic target. *Liver Int* 2009; 29: 955–65. [PubMed: 19490415]
33. Teufel A, Galle PR. Collecting evidence for a stem cell hypothesis in HCC. *Gut* 2010; 59: 870–1. [PubMed: 20581232]
34. Iovino F, Meraviglia S, Spina M, et al. Immunotherapy targeting colon cancer stem cells. *Immunotherapy* 2011; 3: 97–106. [PubMed: 21174560]
35. Papailiou J, Bramis KJ, Gazouli M, Theodoropoulos G. Stem cells in colon cancer. A new era in cancer theory begins. *Int J Colorectal Dis* 2011; 26: 1–11. [PubMed: 20680304]
36. Woo HG, Wang XW, Budhu A, et al. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology* 2011; 140: 1063–70. e8. [PubMed: 21094160]

ag bio-informatik
I. MEDIZINISCHE KLINIK UND POLIKLINIK MAINZ

Medicalgenomics

Genetic data and bioinformatics services

[About us](#)
[BlotBase](#)
[CellLineNavigator](#)
[CellMinerHCC](#)
[LoMA](#)
[RNA Seq Atlas](#)

[News](#)
[Search](#)
[Data](#)
[Download](#)

Quick Gene Search

Search Gene

Search

CellMinerHCC - Search

Fulltext CellMinerHCC Search

Search for one gene symbol resp. identifier.
(For multiple genes use 'Explore genetic profile' or separate genes by ",")

Submit

HCC cell line

- 7703
- Focus
- Hep3B
- Hep3B-TR
- Hep40
- HepG2
- HLE
- HLF
- HUH-1
- HUH-6
- HUH-7
- PLCPRF5
- SK-Hep1
- SNU-182
- SNU-387
- SNU-398
- SNU-449
- SNU-475

Explore genetic profile

HCC cell line	Gene Symbol	KEGG Pathway	Gene Ontology
7703	A1BG	1,4-Dichlorobenzene degr	'de novo' GDP-L-fucose bi
Focus	A1CF	1- and 2-Methylnaphthal	'de novo' IMP biosynthetic
Hep3B	A2LD1	3-Chloroacrylic acid degra	'de novo' posttranslational
Hep3B-TR	A2M	ABC transporters	'de novo' pyrimidine base
Hep40	A2ML1	Acute myeloid leukemia	(N-acetylneuraminy)-gali
HepG2	A4GALT	Adherens junction	(R)-3-amino-2-methylpro
HLE	A4GNT	Adipocytokine signalling p	(S)-2-hydroxy-acid oxidas
HLF	AAAS	Alanine and aspartate me	(S)-3-amino-2-methylpro
HUH-1	AACS	Alkaloid biosynthesis I	(S)-limonene 6-monooxy
HUH-6	AADAC	Alkaloid biosynthesis II	(S)-limonene 7-monooxy
HUH-7	AADACL1	Allograft rejection	(alpha-N-acetylneuramin
PLCPRF5	AADAT	Alzheimer's disease	1,4-alpha-glucan branchin
SK-Hep1	AAK1	Aminoacyl-tRNA biosynth	1-acylglycerol-3-phosphat
SNU-182	AAMP	Aminophosphonate metal	1-alkyl-2-acetylgllyceroph
SNU-387	AANAT	Aminosugars metabolism	1-alpha,25-dihydroxyvital

Make any combination of one ore more selections.
For multiple selections inside the same box hold Ctrl and click.

Query Reset

Fig. 1. CellMinerHCC offers multiple search options. Searches may be performed by means of individual gene names, NCBI Gene IDs, Ensembl Gene IDs, or disease names. Also, more complex searches may be performed by selecting disease, gene symbol, a genetic pathway from KEGG or a gene ontology from the 'explore genetic association' panel.

Fig. 5.

Top 20 significantly enriched biological functions identified by Ingenuity Pathway Analysis using the 195 commonly regulated genes across all 18 HCC cell lines. The bars represent the degree of significance ($-\log$ of the P -value) on the x -axis. The threshold line indicates the level above which the degree of enrichment has become statistically significant. The significant categories are provided along the y -axis.

Top 20 DAVID annotation chart for the 195 commonly regulated genes across all 18 HCC cell lines sorted by their degree of significance. Terms are identified by gene enrichment and functional annotation analysis. The categories on their left side provide the original databases and resources where the enriched terms originate from. The number of genes involved in each particular term out of the 195 gene list is provided graphically, in total numbers, and in percentage to the list analysed. Finally, the *P*-value of the gene enrichment in relation to the human genome and the Benjamini correction for multiple testing are provided on the right.

Table 1.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	SP_PIR_KEYWORDS	plasma	RT		25	13.4	5.1E-28	2.0E-25
☐	KEGG_PATHWAY	Complement and coagulation cascades	RT		19	10.2	6.1E-17	1.2E-14
☐	SP_PIR_KEYWORDS	liver	RT		15	8.0	3.1E-16	6.5E-14
☐	SP_PIR_KEYWORDS	Secreted	RT		52	27.8	7.1E-14	9.2E-12
☐	SP_PIR_KEYWORDS	acute phase	RT		11	5.9	9.1E-14	8.9E-12
☐	GOTERM_BP_FAT	acute inflammatory response	RT		16	8.6	4.2E-13	6.3E-10
☐	GOTERM_CC_FAT	extracellular space	RT		35	18.7	4.4E-12	1.1E-9
☐	SP_PIR_KEYWORDS	signal	RT		67	35.8	5.0E-10	3.9E-8
☐	UP_SEQ_FEATURE	signal peptide	RT		67	35.8	6.5E-10	4.3E-7
☐	SP_PIR_KEYWORDS	disease mutation	RT		43	23.0	1.3E-9	8.3E-8
☐	GOTERM_CC_FAT	extracellular region	RT		57	30.5	2.4E-9	3.0E-7
☐	GOTERM_BP_FAT	response to wounding	RT		26	13.9	2.4E-9	1.8E-6
☐	SP_PIR_KEYWORDS	metalloprotein	RT		13	7.0	2.7E-9	1.5E-7
☐	GOTERM_BP_FAT	protein maturation by peptide bond cleavage	RT		12	6.4	4.5E-9	2.3E-6
☐	UP_SEQ_FEATURE	domain:Peptidase S1	RT		12	6.4	6.5E-9	2.1E-6
☐	SP_PIR_KEYWORDS	blood coagulation	RT		9	4.8	8.3E-9	4.1E-7
☐	UP_SEQ_FEATURE	domain:Sushi 2	RT		9	4.8	8.7E-9	1.9E-6
☐	UP_SEQ_FEATURE	domain:Sushi 1	RT		9	4.8	8.7E-9	1.9E-6
☐	GOTERM_CC_FAT	extracellular region part	RT		36	19.3	9.0E-9	7.7E-7
☐	SMART	Tryp_SPC	RT		12	6.4	1.0E-8	8.4E-7