ELSEVIER

Journal of Hepatology 50 (2009) 1122–1131

Journal of Hepatology

www.elsevier.com/locate/jhep

A systems biology perspective on cholangiocellular carcinoma development: Focus on MAPK-signaling and the extracellular environment☆

Chunxia Wang[1,†], Thorsten Maass[2,†], Markus Krupp[2], Florian Thieringer[2], Susanne Strand[2], Marcus A. Wörns[2], Ana-Paula Barreiros[2], Peter R. Galle[2], Andreas Teufel[2,*]

[1]*Central Laboratory, Shandong Provincial Hospital, Jinan, PR China*
[2]*Department of Medicine I, Johannes Gutenberg University, Building 301, Langenbeckstr. 1, 55101 Mainz, Germany*

***Background/Aims*: Multiple genes have been implicated in cholangiocellular carcinoma (CCC) development. However, the overall neoplastic risk is likely associated with a much lower number of critical physiological pathways.**

***Methods*: To investigate this hypothesis, we extracted all published genetic associations for the development of CCC from PubMed (genetic association studies, but also studies associating genes and CCC in general, i.e. functional studies in cell lines, genetic studies in humans, knockout mice etc.) and integrated CCC microarray data.**

***Results*: We demonstrated the MAPK pathway was consistently enriched in CCC. Comparing our data to genetic associations in HCC often successfully treated by a multityrosine kinase inhibitor, sorafenib, we demonstrated a similar overrepresentation of MAPK. In contrast, most cancer-related genetic studies focusing on genes related to transcription and cell cycle control, we consistently found genes coding for products in the extracellular environment to be significantly enriched. Thus, CCC must be regarded as developing in the context of an altered extracellular environment.**

***Conclusions*: Our study suggests the liver microenvironment holds essential functions and structures key to CCC progression. Furthermore, we identified the MAPK signaling pathway consistently enriched, pointing towards a critical role in CCC development. These data may provide a rationale for treatment of CCC with sorafenib.**

***Keywords*: Liver cancer; Cholangiocarcinoma; Systems biology; Bioinformatics; Oncogenomics**

1. Introduction

Cholangiocellular carcinoma (CCC) is a comparatively rare cancer arising from the bile ducts. However, rates of cholangiocellular carcinoma have been rising worldwide over the past several decades, particularly rates of intrahepatic CCC [1,2].

Known risk factors account for only a few cases of CCC. Highest incidences of CCC were documented among patients with primary sclerosing cholangitis (PSC) [3]. Furthermore, particularly in Asia, infection with parasitic liver flukes causes a significant number of CCC [4].

Therapeutic options for CCC currently remain very limited. Besides surgery, the use of chemotherapy is still a matter of intense debate with many arguing for best supportive care as the standard of treatment [5]. Recently, targeted therapies have become novel therapeutic options in a variety of diseases. For example the multityrosine kinase inhibitor sorafenib was the first drug to significantly improve overall survival in hepatocellular carcinoma (HCC), previously regarded as resistant to conventional chemotherapeutic strategies [6].

Received 18 May 2008; received in revised form 5 January 2009; accepted 7 January 2009; available online 31 March 2009
Associate Editor: K. Koike

☆ The authors who have taken part in this study declared that they do not have anything to disclose regarding funding from industry or conflict of interest with respect to this manuscript.

* Corresponding author.
E-mail address: teufel@uni-mainz.de (A. Teufel).

† These authors contributed equally to this work.

0168-8278/$36.00 © 2009 Published by Elsevier B.V. on behalf of the European Association for the Study of the Liver.
doi:10.1016/j.jhep.2009.01.024

Numerous individual genes have been studied with respect to their level of expression in liver tissue. However, the overall picture is still undefined and general rules or factors regulating gene expression in the liver have not yet been established [7]. A vast number of genes have been reported to be involved in cancer development [9]. With this sheer number of genes the overall neoplastic risk has been suggested to be influenced rather by a much lower number of essential (physiological) pathways [8,10]. Thus, we performed a genome-wide analysis of genetic factors involved in CCC development.

2. Material and methods

2.1. Data aquisition and accessibility

In order to establish a comprehensive dataset on genetic associations of CCC, the complete PubMed database, currently containing approximately 17 million publications, has initially been searched by means of MeSH terms for abstracts containing the terms "CCC", "biliary tract cancer", "gallbladder cancer", "cholangiocellular carcinoma" and "cholangiocarcinoma". Subsequently, the extracted abstracts were searched for the official gene names of the Human Genome Organiszation (HUGO, http://www.hugo-international.org/) including all alias gene names obtained from the NCBI Gene Website as well as the corresponding murine gene names also from NCBI [11]. This strategy revealed 13710 abstracts assumed to potentially describe genetic associations in CCC development. Subsequently, these publications were individually and manually validated for genetic associations, leaving 236 confirmed potential genetic associations for CCC development. Publications included not only genetic association studies (case-control studies), but studies associating genes and cholangiocarcinoma in general, i.e. functional studies in cell lines, genetic studies in humans, knockout mice etc. These genes as well as the genes regulated in the microarray experiment by Obama et al. [12] were designated "CCC associated" in this paper. Finally, validated information was stored in our publicly available database, Library of Genetic Associations (http://www.medicalgenomics.org/databases/loga/news, Fig. 1).

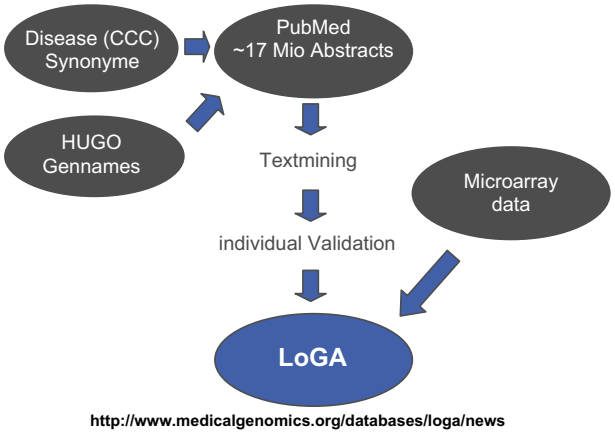

Fig. 1. Schematic drawing of the data acquisition process. Identified potential genetic associations to CCC development were made publicly available through the LoGA data base (www.medicalgenomics.org/databases/loga/news). [This figure appears in colour on the web.]

2.2. Microarray data

Microarray data were obtained from the microarray experiment by Obama et al. as published (http://www.interscience.wiley.com/jpages/0270-9139/suppmat/index.html, [12]). Of these, 51 upregulated and 324 downregulated genes could be assigned an individual gene ID and thus were then available for automated analysis. Genes/EST not being assigned an Entrez Gene ID were mostly EST of unknown functional relevance.

2.3. Automated analysis

As for many genes multiple alias names exist, we had to address each gene with the gene-ID provided by the NCBI Entrez website [11]. For some genes, generally uncharacterized genes, unique IDs have not yet been assigned. Thus, after matching individual Entrez-Gene-IDs to the individual genes we obtained a gene set of 601 genes for further automated analysis, 592 of which were available for WebGestalt (see [13]) analysis as they had an individual gene name assigned and the associaton with CCC was found in human tissue. As for those genes not having an Entrez-Gene-ID, no functional information was available, we did not expect the removal of a few genes from the gene set to be critical for our further functional analysis.

2.4. Gene set analysis

Gene set analysis of the PubMed isolated gene sets and microarray gene sets as well as the complete data set were performed using the WebGestalt Toolkit [13]. For comparison to reference collectives, we selected the human reference gene set "WEBGESTALT_HUMAN". For statistical testing choice was left on "no preference". Due to the strictly human reference gene set, we discarded genes that had only been demonstrated in mouse or rat to be associated with CCC from this analysis. As these were only four genes confirmed in rat 3 in mouse, and 1 in hamster we did not expect this to be critical to the analysis.

2.5. Analysis of genetic association with hepatocellular carcinoma (HCC)

In order to be able to investigate genetic associations for hepatocellular carcinoma, we repeated the same selection strategy as for CCC. MeSH terms were "Hepatocellular" [MeSH] OR "hepatocellular carcinoma" OR "HCC" OR "hepatoma" OR "liver cancer" OR "primary liver cancer" OR "liver tumor" OR "liver carcinoma" OR "primary liver cancer" OR "hepatic tumor". We identified 608 confirmed human potential genetic associations for HCC. The validated associations were stored in our publicly available database (www.medicalgenomics.org/databases/loga/news).

Microarray data were provided by Lee et al. and had previously been published [14,15].

3. Results

3.1. Chromosomal distribution

To identify genetic hotspots we analyzed the chromosomal distribution of the genes associated with CCC. Looking at the complete list of genes, these were found to be evenly distributed. An exception to this was the Y-chromosome which contained only two of these genes. Also at Chromosome 17 a regional accumulation of CCC associated genes was observed. However, none of these genes were neighbouring one another. This general distribution pattern was observed in all three subgroups of the complete list, the PubMed extracted genes and the array of up/downregulated genes (Fig. 2).