Causal modeling of stroma-cancer cell communication

Julia C Engelmann[1], Claus Hellerbrand[2], Rainer Spang[1]
1 *Statistical Bioinformatics, University of Regensburg, Germany*
2 *Internal Medicine I, University Hospital Regensburg, Germany*
julia.engelmann@ur.de

Abstract: Molecular communication between stroma and cancer cells is well recognized to have a crucial role in carcinogenesis, tumor growth and cancer cell migration. From a systems biology perspective it links the molecular networks of both cell types. We have described a systems level analysis combining experimental and computational approaches for studying inter-cellular communication via secreted gene products. Our approach builds on statistical methodology for causal analysis based on a combination of reverse network engineering and causal effect estimation using genome scale observational data. In the context of molecular interactions between hepatic stellate cells (HSC) and hepatocellular carcinoma (HCC) cells, we predicted causal effects of HSC secreted gene products on tumor (HCC) gene expression. The many cause-effect pairs were then condensed to a small set of stromal factors which together cause the majority of gene expression changes observed in HCC cells with a Bayesian approach borrowed from Model-based Gene Set Analysis (MGSA). This resulted in a set of 10 secreted HSC gene products which together cause the majority of gene expression changes observed in HCC cells. The set of secreted stromal factors contained both known and unknown cancer promoting factors, including Placental Growth Factor (PGF) and Periostin (POSTN) as representatives of the former, and Pregnancy-Associated Plasma Protein A (PAPPA) as an example of the latter. We could show that PAPPA contributes to the activation of NFkB signaling. In clinical data, higher levels of PAPPA are linked to advanced stage HCC.

1 Background

Cancer is a heterogeneous assembly of different cell types characterized, among others, by its composition and interactions of different cells. The basic building blocks of a cancer entity are epithelial cells, fibroblasts, vascular and inflammatory cells plus the extracellular matrix. Their interactions via (1) cell-cell contacts, (2) secreted factors like chemo- and cytokines, and (3) the modulation of the extracellular matrix are dynamic and influence cell proliferation, movement and differentiation [TC06].

Hepatocellular carcinoma (HCC) is one of the most prevalent and lethal malignant tumors worldwide. The major risk factor predisposing to HCC is hepatic cirrhosis. It arises through the activation of hepatic stellate cells (HSC), myofibroblast-like cells that are responsible for the excessive hepatic matrix deposition seen in chronically damaged livers. Moreover, HSCs infiltrate the stroma of liver tumors localizing around tumor sinusoids, fibrous septa, and capsules [WF14]. Conditioned medium collected from activated HSCs induces growth, migration and invasion of HCC cells in vitro. Furthermore, HSCs promote aggressive growth of HCC cells in experimental in vivo models [ZZY+11] and their presence predicts poor clinical outcome in HCC patients [JQF+09]. These data indicate that HSCs affect HCCs. Yet, the molecular mechanisms of this crosstalk are largely unknown.

Intra- and inter-cellular molecular mechanisms are typically studied using functional assays that involve the perturbation of the cellular systems. Unlike statistical associations in observational data, functional assays can directly distinguish between cause and effect. Their disadvantage is that they can be difficult to perform in high throughput. Recently, Maathuis and colleagues introduced a novel method to extract causal information from mere observational gene expression data [MKB09]. Their IDA ('Intervention calculus when the DAG is absent') algorithm combines local reverse network engineering using the PC-algorithm [SGS00] with causal effect estimation [Pea00, Pea03]. These virtual functional assays predict lists of genes that will change expression if the expression of a query gene was perturbed experimentally.

In [EAOR+15] we showed how functionally relevant secreted agents of stroma-tumor communication can be successfully predicted through a combination of novel experimental designs, causal network modeling, and data integration: Stromal hepatic stellate cells (HSC) from a set of human donors were

cultivated and the conditioned media were used to stimulate hepatocellular carcinoma cells (HCC). Gene expression was measured on the paired HSC and HCC cells before and after stimulation. With information on gene expression levels in both 'sender' and 'receiver' cells, we were able to infer which genes might play a role in communication of these two cell types. We used the IDA framework to predict the effects of virtual targeted interventions in HSCs on the expression of individual genes in stimulated HCCs. Finally, we integrated the large set of predicted pairs of causally interacting gene products to select the most important HSC secreted agents influencing cancer cell gene expression.

2 Results

Causal modeling approach

In our paper [EAOR$^+$15], we used virtual targeted interventions by means of the IDA algorithm [MKB09] to identify gene products that mediate the communication of stroma and cancer cells. IDA consists of two steps. First, a partially directed network of regulatory interactions is constructed using the PC algorithm [SGS00]. Second, causal effects are estimated using Pearl's Do-calculus [Pea00]. To infer a potential causal effect of a stromal gene x on a cancer gene y, IDA needs the expression of y, x, and all genes x' that directly influence the expression of x in the regulatory network. Since stromal cells were in no contact to cancer cells in our experimental setting, the genes x' must be stromal genes as well. Hence it is sufficient to confine the reconstruction of a regulatory network to stromal genes only. For each of the cancer genes that changed expression upon conditioned media stimulation (False Discovery Rate < 0.001), we used IDA to screen for potential stromal genes that when perturbed in expression would have a strong effect on the respective cancer gene. Therefore we focused on secreted gene products as candidate stromal regulators. However, these genes are most likely regulated by non-secreted gene products which hence also need to be included into the network reconstruction. To limit the computational burden, we included the most highly and variably expressed genes across the stromal samples into the analysis, assuming that they would translate into abundant and variable amounts of protein. Since we were interested in cellular communication via secreted gene products, we confined the list of potential activators of cancer genes to only secreted stroma genes.

For each of the target cancer genes, secreted stroma genes were ranked by the effect size estimated by IDA. This procedure corresponds to ranking by the predicted causal effect in a virtual perturbation experiment: Gene-by-gene, all secreted stroma genes were virtually repressed by one standard unit and the expected change of the cancer gene was calculated. Performing the analysis on standardized data allows comparing effects across genes, and thus, the stromal gene with the strongest expected effect was ranked first, and so on. We applied IDA modeling in a sub-sampling approach, reporting causal effects only when they were insensitive to small perturbations of the data. The experimental and computational model set-up is depicted in Figure 1.

A small set of stroma-secreted proteins can activate cancer gene expression in concert.

Although all secreted HSC proteins have the potential to affect the expression of HCC genes, we postulate that a much smaller set of proteins is sufficient to activate HCCs. Thus in [EAOR$^+$15] we aimed at identifying a small set of HSC genes that jointly account for the wide spectrum of expression changes in HCC cells observed in response to stimulation with HSC-CMs. We arranged the cause-effect pairs such that we obtained a list of potential HCC targets for each HSC cause. Since several HSC genes were predicted to affect multiple HCC genes, these lists overlapped. Model based Gene Set Analysis (MGSA) [BRG11] is an algorithm that aims at partially covering an input list of genes with as little Gene Ontology categories as possible. It balances the coverage with the number of categories needed.

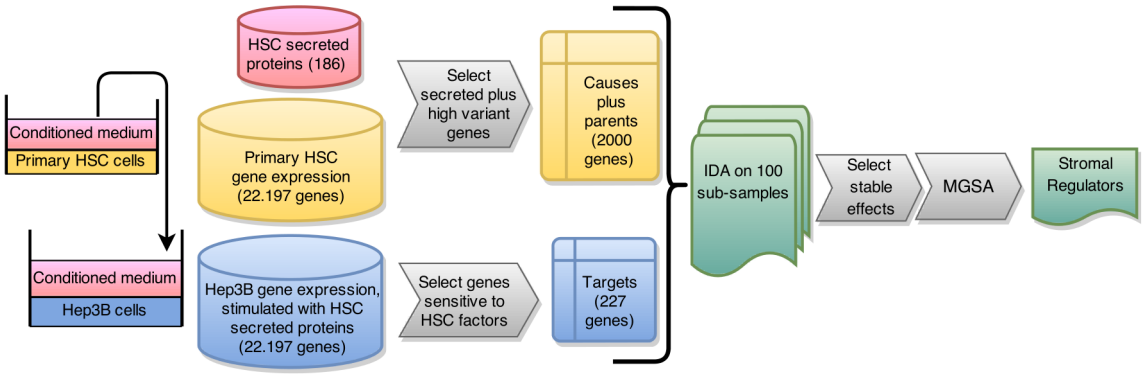

Figure 1: **Overview of the experimental and computational approach to identify secreted stromal (HSC) factors which influence tumor (HCC) gene expression.** Conditioned medium of primary human HSC (n=15) was transferred onto human Hep3B HCC cells. Gene expression data of HSC and HCC cells were filtered to reduce the dimensionality of the data and to build cause-and-effect (target) matrices. The matrices served as input for the IDA algorithm which estimates causal effects for each cause on each target gene. Causal effects that were stable across sub-sampling runs (i.e., that were stable with respect to small perturbations of the data) were retained and subjected to Model-based Gene Set Analysis (MGSA) to extract a sparse set of HSC genes influencing HCC cell gene expression.

We modified this algorithm in such a way that it covered the list of cancer genes responsive to stromal factors with the sets of HSC targets predicted by IDA. Thus HSC genes were in competition to each other: an analysis based on frequencies (how many HCC genes does each HSC gene affect) discovers redundant HSC genes that target the same HCC genes. Our approach strove for a maximum coverage of the target genes with a minimum number of HSC secreted genes. We identified 10 HSC secreted proteins which covered the majority of gene expression changes observed in HCC cells. The list consisted of PGF, CXCL1, PAPPA, IGF2, IGFBP2, POSTN, NPC2, CTSB, HGF, and CSF1. Notably, the set of the most influential HSC regulators included several well-known tumor-promoting genes such as placental growth factor (PGF) [PWB+05], and the chemokine CXCL1, which promotes HCC angiogenesis and growth [TML+12]. Periostin (POSTN) is a secreted cell adhesion protein whose expression levels are directly related to metastatic potential and poor prognosis of HCC [LWJ+13]. High expression levels of the macrophage colony-stimulating factor 1 (CSF1) are another indicator of tumor progression and poor survival in HCC patients [BFY+06]. Over-expression of cathepsin B (CTSB), on the other hand, promotes HCC cell migration and invasion [CCJ+12].

PAPPA is a novel stromal factor which activates NFkB signaling in cancer cells

In our paper [EAOR+15], we identified PAPPA as a novel stromal regulator of HCC cell gene expression. As many of the cancer genes that changed gene expression levels upon incubation with stroma-conditioned medium are NFkB pathway members or targets of the transcription factor NFkB, we experimentally tested whether PAPPA could induce NFkB activity. Indeed, recombinant PAPPA protein together with conditioned medium induced a stronger induction of NFkB signaling than conditioned medium alone. We could also show that PAPPA is solely secreted by HSCs and not by HCC cells, and its protein levels correlate with fibroblast markers in patient samples, indicating that stromal cells are the major source of PAPPA also in HCC tissue. Finally, we could show that increased levels of PAPPA indicate advanced stage HCC in clinical samples.

3 Presentation outline

The presentation will first motivate and introduce the importance of cell communication of different cell types in tumor tissue. Then the experimental setting to produce systems-wide unidirectional cell communication data will be presented. The main part of the presentation will focus on the computational model to derive the most important stromal factors which influence tumor cell gene expression. The last part will briefly highlight biological findings with this approach and their clinical implications.

References

[BFY+06] Anuradha Budhu, Marshonna Forgues, Qing-Hai Ye, Hu-Liang Jia, Ping He, Krista A. Zanetti, Udai S. Kammula, Yidong Chen, Lun-Xiu Qin, Zhao-You Tang, and Xin Wei Wang. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell*, 10(2):99–111, Aug 2006.

[BRG11] Sebastian Bauer, Peter N Robinson, and Julien Gagneur. Model-based gene set analysis for Bioconductor. *Bioinformatics*, 27(13):1882–1883, Jul 2011.

[CCJ+12] Wan-Nan Chen, Jin-Yan Chen, Bo-Yan Jiao, Wan-Song Lin, Yun-Li Wu, Ling-Ling Liu, and Xu Lin. Interaction of the hepatitis B spliced protein with cathepsin B promotes hepatoma cell migration and invasion. *J Virol*, 86(24):13533–13541, Dec 2012.

[EAOR+15] Julia C. Engelmann, Thomas Amann, Birgitta Ott-Rötzer, Margit Nützel, Yvonne Reinders, Jörg Reinders, Wolfgang E. Thasler, Theresa Kristl, Andreas Teufel, Christian G. Huber, Peter J. Oefner, Rainer Spang, and Claus Hellerbrand. Causal Modeling of Cancer-Stromal Communication Identifies PAPPA as a Novel Stroma-Secreted Factor Activating NFkB Signaling in Hepatocellular Carcinoma. *PLoS Comput Biol*, 11(5):e1004293, May 2015.

[JQF+09] Min-Jie Ju, Shuang-Jian Qiu, Jia Fan, Yong-Sheng Xiao, Qiang Gao, Jian Zhou, Yi-Wei Li, and Zhao-You Tang. Peritumoral activated hepatic stellate cells predict poor clinical outcome in hepatocellular carcinoma after curative resection. *Am J Clin Pathol*, 131(4):498–510, Apr 2009.

[LWJ+13] Yang Lv, Wei Wang, Wei-Dong Jia, Qi-Kai Sun, Jian-Sheng Li, Jin-Liang Ma, Wen-Bin Liu, Hang-Cheng Zhou, Yong-Sheng Ge, Ji-Hai Yu, Hong-Hai Xia, and Ge-Liang Xu. High-level expression of periostin is closely related to metastatic potential and poor prognosis of hepatocellular carcinoma. *Med Oncol*, 30(1):385, Mar 2013.

[MKB09] Marloes H Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133–3164, 2009.

[Pea00] Judea Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge Univ. Press, Cambridge, Cambridge, 2000.

[Pea03] Judea Pearl. Statistics and causal inference: A review. *Test*, 12:281–318, 2003.

[PWB+05] Christian Parr, Gareth Watkins, Mike Boulton, Jun Cai, and Wen G. Jiang. Placenta growth factor is over-expressed and has prognostic value in human breast cancer. *Eur J Cancer*, 41(18):2819–2827, Dec 2005.

[SGS00] Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search.* MIT Press, Cambridge, MA, Cambridge, 2nd edition, 2000. ISBN 0-262-19440-6.

[TC06] Thea D. Tlsty and Lisa M. Coussens. Tumor stroma and regulation of cancer development. *Annu Rev Pathol*, 1:119–150, 2006.

[TML+12] Kwan Ho Tang, Stephanie Ma, Terence K. Lee, Yuen Piu Chan, Pak Shing Kwan, Carol M. Tong, Irene O. Ng, Kwan Man, Ka-Fai To, Paul B. Lai, Chung-Mau Lo, Xin-Yuan Guan, and Kwok Wah Chan. CD133+ liver tumor-initiating cells promote tumor angiogenesis, growth, and self-renewal through neurotensin/interleukin-8/CXCL1 signaling. *Hepatology*, 55(3):807–820, Mar 2012.

[WF14] Michael C. Wallace and Scott L. Friedman. Hepatic fibrosis and the microenvironment: fertile soil for hepatocellular carcinoma development. *Gene Expr*, 16(2):77–84, 2014.

[ZZY+11] Wenxiu Zhao, Lei Zhang, Zhenyu Yin, Weixue Su, Guangli Ren, Changsheng Zhou, Junyong You, Jia Fan, and Xiaomin Wang. Activated hepatic stellate cells promote hepatocellular carcinoma development in immunocompetent mice. *Int J Cancer*, 129(11):2651–2661, Dec 2011.