Multi-pathway network analysis of mammalian epithelial cell responses in inflammatory environments

David C. Clarke and Douglas A. Lauffenburger*

Department of Biological Engineering and the Center for Cellular Decision Processes, Massachusetts Institute of Technology, 77 Mass Ave., Cambridge, MA 02139, U.S.A.

*Corresponding author

Phone: (617) 252-1629 Fax: (617) 258-0204 Email: lauffen@mit.edu

Running title: Cell signalling networks in inflammation

Keywords: inflammation, liver, hepatocellular carcinoma, mathematical model, logic, cell signalling.

Abstract

Inflammation is a key physiological response to infection and injury and while usually beneficial it can also be damaging to the host. The liver is a prototypical example in this regard because inflammation helps resolve liver injury but it also underlies the etiology of pathologies such as fibrosis and hepatocellular carcinoma. Cells sense their environment, including the inflammatory environment, through the activities of receptor-mediated signal transduction pathways. These pathways are organized in a complex interconnected network, and it is becoming increasingly recognized that cellular adaptations result from the quantitative integration of multi-pathway network activities, rather than isolated pathways causing particular phenotypes. Therefore, comprehending liver cell signalling in inflammation requires a scientific approach that is appropriate for studying complex networks. Here we review our application of systems analyses of liver cell signalling in response to inflammatory environments. Our studies feature broad measurements of cell signalling and phenotypes in response to numerous experimental perturbations reflective of inflammatory environments, the data from which are analyzed using Boolean and fuzzy logic models and regression-based methods in order to quantitatively relate the phenotypic responses to cell signalling network states. Our principal biological insight from these studies is that hepatocellular carcinoma cells feature uncoupled inflammatory and growth factor signalling, which may underlie their immune evasion and hyperproliferative properties.

1 Introduction

2 Inflammation is a mechanism for preserving homeostasis in response to noxious stimuli such as infection 3 and injury. Cells of the innate immune system drive the inflammatory response, which typically begins 4 when proinflammatory stimuli activate macrophages residing in the affected tissue to produce 5 chemoattractants that recruit neutrophils to the site of inflammation. Neutrophils are white blood cells that 6 sense and eliminate pathogens. They carry out their functions in part by release of oxygen radicals and 7 degradative enzymes whose leakage can cause collateral tissue damage [1]. Once the inflammatory 8 stimulus is cleared, the inflammatory response is damped by a process called resolution, which is marked 9 by the transition from neutrophil to monocyte recruitment. Monocytes are blood-borne macrophages that 10 differentiate into macrophages once inside tissue. They serve to clear debris and promote tissue repair. 11 The inflammatory response is therefore characterized by processes that cause both damage and repair. 12 The damaging processes must be sufficiently strong to eliminate the inflammatory stimulus but not too 13 strong so as to cause excess tissue damage. A maladapted inflammatory response can lead to chronic 14 inflammation, which is a hallmark of many complex diseases such as cancer, fibrosis, cardiovascular 15 disease and diabetes. A better understanding of inflammation could therefore improve therapeutic 16 approaches to acute and chronic diseases.

17

18 The liver plays a central role in maintaining homeostasis through its functions in metabolism,

19 detoxification and inflammation. The liver is a key participant in the initial systemic response to

20 inflammation, called the acute phase response, because it synthesizes acute phase proteins such as C-

reactive protein, serum amyloid A and fibrinogen [2]. Conversely, components of the inflammatory

response are important in liver physiology and pathophysiology. The cytokine interleukin-6, for example,

23 serves to protect the liver when it is injured and promotes liver regeneration [3]. Inflammation can

24 contribute to pathological states of the liver, perhaps best exemplified by chronic inflammation due to viral

infection (e.g., hepatitis B and C), toxic substance exposure (e.g., aflatoxin-B1), or steatosis serving as a

26 precursor to hepatocellular carcinoma [4]. Hepatocellular carcinoma is the fifth most prevalent cancer

27 worldwide and is notoriously difficult to treat, which underlies its status as the third most lethal type of

28 cancer [5]. The considerable burdens of liver disease and diseases linked to chronic inflammation

29 emphasize the need for investigating the interplay between the liver and inflammation.

30

Cells adapt to their environments by the activities of receptor-mediated signalling pathways. The biochemical activities of these signalling pathways regulate gene expression, metabolism and/or cell structure in order to modify cell physiology. For example, in the case of hepatocytes during the acute phase response, IL-6 released by macrophages and stromal fibroblasts at the site of inflammation acts hormonally on hepatocytes by binding and activating a receptor complex leading to phosphorylation and dimerization of the signal transducers and activators of transcription 3 (STAT3) [6, 7]. STAT3 is a transcription factor that regulates the transcription of a number of genes including many involved in the 38 acute-phase response. The communication between cells acting at the site of inflammation and those of

39 the liver exemplify how the cells of multi-cellular organisms communicate in order to maintain

40 homeostasis.

41

42 The activities of individual pathways such as IL-6-STAT3 have been well studied in liver cells. However, 43 studies of isolated pathways have limited applicability to the in vivo situation in which cells are continually 44 exposed to multiple extracellular molecules leading to the simultaneous activity of many signalling 45 pathways. In inflammation, for example, cells are exposed to multiple pro- and anti-inflammatory 46 cytokines (e.g., tumour necrosis factor- α (TNF- α), interleukin-1 (IL-1), IL-4, IL-10, etc.), growth factors, 47 hormones and other molecules. Intracellular signalling pathways crosstalk with each other, effectively 48 forming a network, but it is poorly understood how intracellular signalling networks processes the 49 combinatorial action of multiple environmental cues. Making progress in this area will improve our 50 understanding of complex physiological responses such as inflammation and help guide the development 51 of better therapeutics. Here we review our studies of the intracellular signalling systems of hepatocytes 52 and hepatoma cells in response to inflammatory environments using a systems approach. We begin by 53 explaining what a systems approach to biology means followed by reviewing three studies in which we 54 used this approach to investigate liver cell signalling and physiology in inflammatory contexts.

55

56 Scientific approach: Cue-signal-response experiments and mathematical modelling

A hallmark of engineering practice is applying a systems approach to the design process. In this context, a "systems approach" denotes studying a system by applying diverse inputs to the system and measuring the outputs. Mathematical models are then used to model the relationship between input and output. The parameters of the mathematical model are tuned such that the desired outputs are obtained from the inputs expected under operating conditions. The engineer then modifies the design to reflect these optimal parameter values.

63

64 We adopt a similar engineering approach in investigating biological systems, with a difference being that 65 we seek to "reverse engineer" the system by using input-output relationships and selected measurements 66 of the system to constrain a model of the intracellular signalling network. We implement an experimental 67 paradigm called "cue-signal-response" that reflects this input-system-output relationship [8]. The input 68 consists of molecules in the cell's environment, the system is the cell signalling network and the output is 69 the behaviour that the cell executes to adapt to the input. In practice, a cue-signal-response paradigm 70 applied to hepatocyte physiology during inflammation involves applying inflammatory cytokines and 71 growth factors (cues) to cultured hepatocytes or hepatoma cells, perturbing intracellular signalling by 72 inhibiting kinases with small-molecule inhibitors, assessing intracellular signalling by multiplexed 73 measurement of phospho-protein levels (signals), and measuring the secretion of cytokines (responses). 74 Specific cues include inflammatory cytokines such as IL-6 and IL-1 and inhibitors target kinases such as

75 mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) kinase (MEK),

phosphoinositide 3-kinase (PI3K) and p38 MAPK. The cells are lysed at specific time points and

77 multiplexed bead-based flow cytometric immunoassays based on Luminex xMAP technology are used to

78 measure the levels of ~15 phosphorylated proteins including Akt, ERK and signal transducer and

79 activator of transcription 3 (STAT3). Similar assays are also used to measure the levels of ~50 secreted

80 proteins in the cell culture media.

81

82 The resulting dataset features thousands of data points, which makes it challenging to interpret.

83 Mathematical tools are therefore used for downstream analyses. Classically, ordinary-differential-

84 equations have been the method of choice for analyzing cell signalling systems because they make direct

use of biochemical rate equations that describe the kinetics of enzyme-catalyzed reactions, protein-

86 protein interactions and transport processes. Disadvantages of ODE models include their critical

87 requirement for firm specification of network topological interactions, and their reliance on adjustable

88 parameters that must be robustly estimated in order to effectively represent the system. As the size of the

89 model grows, so does the uncertainty in the topology and the corresponding number of parameters,

90 which in turn increases the demand for more comprehensive biological knowledge and intensive

91 experimental data (as well as computational power, although that is a lesser challenge at this point).

92 While studying signalling from one or two pathways with ODEs is feasible, the networks that we study are

too uncertain and large for ODEs to be practically useful. We therefore use modelling techniques that

94 represent the system in a coarser grained fashion. In doing so, we require less data to obtain quantitative

95 insight into the system, albeit less than could be obtained with ODEs.

96

97 Our studies feature two types of mathematical frameworks, regression-based methods and logic-based 98 methods. Our implementation of these techniques has been reviewed in detail elsewhere [9-11]. Briefly, 99 regression models, such as multiple linear regression and partial-least-squares regression (PLSR), are 100 useful for guantifying the correlation between variables in context of one another. Regression models do 101 not incorporate information beyond the data itself, except that the variables included in the model are 102 specified by the modeller. This prior specification makes the models supervised but they are also strictly 103 empirical. Logic modelling, by contrast, involves translating prior knowledge or hypotheses about the 104 system structure or function into computable language. In this way, logic models are capable of bringing a 105 network diagram to operational function. We investigate cell signalling networks using logic modelling by 106 first constructing a diagram of the network based on published data (which we call a "prior knowledge 107 network", or PKN), then collecting a cue-signal-response dataset devoted to perturbing and measuring 108 aspects of the network, followed by using optimization algorithms to identify and quantify the connections 109 in the hypothetical network that are most important for explaining the data (Figure). The resulting fitted 110 models can then be used for simulation or analysis purposes. Irrespective of the modelling approach, 111 model predictions are experimentally validated. We have used this workflow to obtain considerable

insights into the systems-level operation of cell signalling networks in effecting phenotypic responses indiverse contexts, which we discuss in the following section.

114

115 Liver cell information processing during inflammation

116 We applied our interdisciplinary approach to the issue of epithelial cell signalling in inflammation by 117 devising logical and statistical modelling methods and applying them to data from cultured liver cells 118 exposed to inflammatory conditions. In the first paper from these studies, Saez-Rodriguez et al. extended 119 Boolean logic methods previously used to study biological networks from a theoretical standpoint to allow 120 the model to directly interface with experimental data [12]. Specifically, their algorithm translates a prior 121 knowledge network, in this specific example a database-derived PSN map, into a Boolean logic model 122 and optimizes the model topology to best fit experimental data. In this example, the data comprised 123 phospho-protein levels of intracellular signalling intermediates in HepG2 cells exposed to inflammatory 124 cytokines and inhibitors of several kinases. Interestingly, they found the resulting optimal models 125 consisted of substantially fewer connections than found in the PKN. Remarkably, an empty model, that 126 contained nodes but no edges connecting them, fit the data better than the PKN [12]. This result implies 127 that the comprehensive protein-protein interaction network maps and PSNs commonly used to depict 128 biological networks do not necessarily reflect networks operating in a specific cell type under specific 129 conditions. This lack of predictivity stems from two apparently paradoxical sources: 1) The networks 130 include too many interactions, presumably because they are typically curated from multiple sources, cell 131 types, time points and experimental conditions and 2) The networks lack interactions that are present and 132 functional in the network under study due to imperfect databases or incomplete understanding of the 133 biology. Indeed, by examining the data points that the original optimal model failed to adequately fit, 134 Saez-Rodriguez et al. tested new interactions to see which ones best improved the fit [12]. The existence 135 of two candidate interactions, one linking TNF-receptor associated factor 6 and MEK and another linking 136 ERK and insulin receptor substrate-1, was supported by published evidence.

137

138 An alternative modelling approach was used by Alexopoulos et al., who performed a comparative 139 analysis of the intracellular signalling networks of healthy and cancerous liver cells [13]. A cue-signal-140 response dataset was generated in which inflammatory cytokines and growth factors were applied to 141 primary human hepatocytes and HepG2 hepatocellular carcinoma cells in concert with small-molecule 142 inhibitors targeting seven kinases from different signalling pathways. Multiple linear regression analysis 143 was used to estimate the strength of relationships between the cytokines and signals, the inhibitors and 144 signals, and the signals and secreted cytokines. The networks were defined by the relationships featuring 145 the highest regression coefficient magnitudes. The networks for the primary hepatocytes and the HepG2 146 cells were then compared, revealing that HepG2 cells displayed reduced responsiveness to inflammatory 147 stimuli but increased responsiveness to progrowth stimuli, relative to the primary hepatocytes. In 148 particular, alterations of NF-kB signalling in HCC cells had profound phenotypic consequences because

- 149 primary hepatocytes secreted a number of cytokines that the HCC cells did not. These cytokines are
- thought to be involved in recruiting cells of the innate immune system, which implies that HCC cells
- 151 modify the secretion patterns in order to avoid detection and elimination by the immune system.
- 152

153 The above studies clearly demonstrate the utility of Boolean logic and regression methods in exploring 154 how normal and healthy liver cells differentially process environmental information. However, both 155 modelling methods have important limitations. Specifically, Boolean logic describes the activity of each 156 node as either "on" or "off", which ignores potentially important graded activity, and regression models do 157 not incorporate potentially valuable prior knowledge. Morris et al. addressed these limitations by 158 developing a novel logic-based modelling method called "constrained fuzzy logic" (cFL), which 159 incorporates prior knowledge in the same manner as Boolean logic but also models quantitative 160 behaviour [14]. Specifically, this method converts an input value from an upstream node into a continuous 161 value between 0 and 1 for the downstream node through a sigmoidal transfer function [14]. This capability 162 facilitates the ability of the resulting trained models to fit weak responses. The quantitative relationships 163 between proteins are also estimated, allowing for modelling of dose-response data, which could prove 164 valuable for pharmacological applications.

165

166 CFL was applied to the same dataset to which the BL approach had been previously applied [12]. CFL 167 was able to capture interactions that were missed by BL [14]. This included the moderate levels of 168 phosphorylation of c-Jun N-terminal kinase (JNK) and c-Jun by TGF- α stimulation. This interaction was 169 the only instance of growth factor pathway crosstalk with inflammatory pathways observed in 170 measurements of HepG2 cells, which the previous studies failed to detect [12, 13]. Furthermore, 171 stimulating HepG2 cells with IL-6 led to moderately increased phosphorylation of several species, 172 including Akt, MEK and p70 S6 kinase, in addition to strong phosphorylation of its canonical downstream 173 STAT3 pathway [14]. The PKN did not include links from the IL-6 receptor that allowed for the observed 174 moderate phosphorylation levels [14]. In the case of BL, the resulting lack of fit of these data points did 175 not adversely affect the overall fit, presumably because a similar absolute deviation resulted between the 176 intermediate levels of the measured phosphorylations and the model outputs of 0 or 1. In contrast, the 177 cFL model was sufficiently sensitive to this error that the model was deemed to inadequately fit this data 178 [14]. Morris et al. followed up this result by seeking to distinguish the pathway that most likely caused the 179 phosphorylations. To do so, they tested PKNs with new interactions either between the IL-6 receptor 180 (IL6R) and PI3K or IL6R and Ras. Most of the resulting fitted models contained the IL6R-Ras link, thus 181 indicating that the Ras-Raf-MEK pathway and not a PI3K-downstream pathway likely mediated the 182 phosphorylations [14]. This result was validated with dedicated experiments. 183

184 **Conclusions and future directions**

7

185 We have performed a series of studies in which mathematical models of proteomic data revealed 186 important insights into the signal transduction networks of healthy and cancerous liver cells in 187 inflammatory environments. Our principal biological insight is that hepatocellular carcinoma cells feature 188 both decreased responsiveness to inflammatory stimuli and increased responsiveness to growth factors 189 relative to normal hepatocytes, which could promote immune evasion and increased proliferation. Our 190 principal mathematical advances include devising methods for formally fitting Boolean logic models to 191 data and creating a fuzzy logic method useful for making quantitative models. For relatively small 192 networks such as the one studied here (i.e., downstream of ~5-7 receptors), cFL is a powerful approach. 193 Given the higher computational burden of cFL, Boolean logic will still be needed to model larger networks 194 until more efficient algorithms are developed. We note, however, that larger networks do not necessarily 195 provide additional predictive power because maximal predictivity was observed with models featuring 196 substantially reduced numbers of edges from the initial prior knowledge network. This finding supports the 197 use of our functional biochemistry approach because it provides data on components of the system that 198 actually carry out the cellular response. A distinctive feature of our approach is that it requires broad 199 sampling of network states, which is achieved by applying diverse experimental conditions (in our case, 200 cytokines and inhibitors, but could also include other treatments such as small interfering RNAs). 201 Attempting to process samples from hundreds of independent experiments with other proteomic 202 techniques such as mass spectrometry is currently unfeasible owing to technical limitations and time and 203 fiscal costs. Our approach therefore represents a rational, efficient and informative means to elucidating 204 epithelial cellular signalling and physiology in inflammatory contexts.

205

206 Going forward, we emphasize that systems-level approaches can be effectively used in vivo. A recent 207 study from our lab successfully extended previous systems-level analyses of data collected from colon 208 cancer cells in vitro [15-17] by applying PLSR modelling to signalling measurements taken from the 209 intestines of mice treated systemically with TNF- α [18]. Determining the biological effects of TNF- α is not 210 straightforward because stimulation of TNF-a receptors increases the activity of multiple downstream 211 signalling pathways, the quantitative integration of which determines the ultimate biological outcome. In 212 the case of mouse intestinal epithelial cells in vivo, for example, TNF-α was found to promote apoptosis in 213 cells of the proximal part of the small intestine, but not in the distal part, with the timing of apoptosis being 214 dose-dependent [18]. TNF- α administration also affected cell proliferation in a region-specific manner. 215 PLSR modelling of signalling and phenotype data revealed that the differential sensitivity of apoptosis 216 was due to guantitative differences in MAPK signalling kinetics between the two intestinal regions and 217 that growth arrest was related to c-Jun and activating transcription factor activation as well as MAPK 218 signalling kinetics [18]. Subsequent experiments validated the hypotheses generated from the original 219 dataset and model. This study demonstrates that our systems-level approach can be successfully applied 220 to in vivo contexts, despite their added complexity compared to in-vitro-cell-culture-based experiments. 221

8

- 222 We contend that our approach works irrespective of the experimental system because cells integrate
- 223 complex contextual information into biochemical activities of signalling pathways that form the basis for
- 224 phenotypic decisions. The cell signalling network is complex but manageable such that by measuring
- selected nodes across this network and using mathematical models to infer the network output, we are
- able to predict the ultimate biological outcome. We therefore anticipate systems-level approaches
- 227 becoming broadly applicable to the study of cellular signalling.
- 228

229 Acknowledgements

- 230 We thank Melody Morris, Julio Saez-Rodriguez and Scott Dixon for critically reading the manuscript. Our
- work is supported by the National Institutes of Health (NIH) [R24-DK090963], the Center for Cellular
- 232 Decision Processes [NIH P50-GM68762] and the Institute for Collaborative Biotechnologies through
- 233 contract no. W911NF-09-D-0001 from the U.S. Army Research Office.
- 234

235 References

236 Gordon, A. H. and Koj, A. (1985) The Acute-phase response to injury and infection : the roles of 1 237 interleukin I and other mediators. Elsevier. Amsterdam 238 Baumann, H. and Gauldie, J. (1994) The acute phase response. Immunol Today. 15, 74-80 2 3 239 Taub, R. (2004) Liver regeneration: from myth to mechanism. Nat Rev Mol Cell Biol. 5, 836-847 240 4 Farazi, P. A. and DePinho, R. A. (2006) Hepatocellular carcinoma pathogenesis: from genes to environment. Nat Rev Cancer. 6, 674-687 241 242 5 El-Serag, H. B. and Rudolph, K. L. (2007) Hepatocellular carcinoma: epidemiology and molecular 243 carcinogenesis. Gastroenterology. 132, 2557-2576 Singh, A., Jayaraman, A. and Hahn, J. (2006) Modeling regulatory mechanisms in IL-6 signal 244 6 245 transduction in hepatocytes. Biotechnol Bioeng. 95, 850-862 246 Heinrich, P. C., Behrmann, I., Muller-Newen, G., Schaper, F. and Graeve, L. (1998) Interleukin-6-7 247 type cytokine signalling through the gp130/Jak/STAT pathway. Biochem J. 334 (Pt 2), 297-314 Gaudet, S., Janes, K. A., Albeck, J. G., Pace, E. A., Lauffenburger, D. A. and Sorger, P. K. (2005) 248 8 249 A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Mol Cell 250 Proteomics. 4, 1569-1590 251 Kumar, N., Hendriks, B. S., Janes, K. A., de Graaf, D. and Lauffenburger, D. A. (2006) Applying 9 252 computational modeling to drug discovery and development. Drug Discov Today. 11, 806-811 253 Lazzara, M. J. and Lauffenburger, D. A. (2009) Quantitative modeling perspectives on the ErbB 10 254 system of cell regulatory processes. Exp Cell Res. 315, 717-725 Morris, M. K., Saez-Rodriguez, J., Sorger, P. K. and Lauffenburger, D. A. (2010) Logic-based 255 11 models for the analysis of cell signaling networks. Biochemistry. 49, 3216-3224 256 257 12 Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt,

- Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt,
 S. and Sorger, P. K. (2009) Discrete logic modelling as a means to link protein signalling networks with
 functional analysis of mammalian signal transduction. Mol Syst Biol. 5, 331
- Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A. and Sorger, P. K.
 (2010) Networks inferred from biochemical data reveal profound differences in toll-like receptor and
- inflammatory signaling between normal and transformed hepatocytes. Mol Cell Proteomics. 9, 1849-1865
 Morris, M. K., Saez-Rodriguez, J., Clarke, D. C., Sorger, P. K. and Lauffenburger, D. A. (2011)
 Training signaling pathway maps to biochemical data with constrained fuzzy logic: guantitative analysis of
- 265 liver cell responses to inflammatory stimuli. PLoS Comput Biol. **7**, e1001099
- 266 15 Janes, K. A., Albeck, J. G., Gaudet, S., Sorger, P. K., Lauffenburger, D. A. and Yaffe, M. B.
- 267 (2005) A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis.
 268 Science. **310**, 1646-1653

Miller-Jensen, K., Janes, K. A., Brugge, J. S. and Lauffenburger, D. A. (2007) Common effector
 processing mediates cell-specific responses to stimuli. Nature. 448, 604-608

17 Kreeger, P. K., Mandhana, R., Alford, S. K., Haigis, K. M. and Lauffenburger, D. A. (2009) RAS
mutations affect tumor necrosis factor-induced apoptosis in colon carcinoma cells via ERK-modulatory
negative and positive feedback circuits along with non-ERK pathway effects. Cancer Res. 69, 8191-8199
18 Lau, K. S., Juchheim, A. M., Cavaliere, K. R., Philips, S. R., Lauffenburger, D. A. and Haigis, K.
M. (2011) In vivo systems analysis identifies spatial and temporal aspects of the modulation of TNFalpha-induced apoptosis and proliferation by MAPKs. Sci Signal. 4, ra16

277 278

279 Figure legend

- 280 Logic modelling workflow. A prior knowledge network is derived from literature, databases and/or existing
- data. Experiments are conducted to systematically perturb and/or measure nodes distributed throughout
- the network. The data is then used by a model optimization (i.e., data fitting) algorithm to tune the model
- topology and if applicable, the model parameters, to minimize the discrepancy between the model output
- and data. The resulting model is then analyzed to derive insight into the biology of the system. The figure
- is adapted from figures contained in reference [14].

