

Title: Dynamics and predicted drug response of a gene network linking dedifferentiation with β -catenin dysfunction in liver cancer

Authors: Claude Gérard¹, Mickaël Di-Luoffo¹, Léolo Gonay^{1,2}, Stefano Caruso^{3,4}, Gabrielle Couchy^{3,4}, Axelle Lorient¹, Darko Castven⁵, Junyan Tao⁶, Katarzyna Konobrocka¹, Sabine Cordi¹, Satdarshan P. Monga⁶, Emmanuel Hanert², Jens U. Marquardt⁵, Jessica Zucman-Rossi^{3,4}, Frédéric P. Lemaigre^{1,*}

¹ de Duve Institute, Université catholique de Louvain, Brussels, Belgium.

² Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

³ Centre de Recherche des Cordeliers, Sorbonne Universités, Inserm, UMRS-1138, F-75006 Paris, France.

⁴ Functional Genomics of Solid Tumors, USPC, Université Paris Descartes, Université Paris Diderot, Université Paris 13, Labex Immuno-Oncology, équipe labellisée Ligue Contre le Cancer, F-75000 Paris, France.

⁵ Department of Medicine I, Johannes Gutenberg University, Mainz, Germany.

⁶ Department of Pathology, Medicine and the Pittsburgh Liver Research Center, University of Pittsburgh, School of Medicine and University of Pittsburgh Medical Center, Pittsburgh, PA, USA.

Corresponding author:

Frédéric Lemaigre

de Duve Institute, Université catholique de Louvain,

Avenue Hippocrate 75 (B1-75.03)

1200 Brussels, Belgium

Email: frederic.lemaigre@uclouvain.be

Phone : +32-2-764-7583 (direct) / +32-2-764-7550 (secretariat)

Fax : +32-2-764-7573

Conflict of interest: The authors declare no conflict of interest.

Financial support: The work of FPL was supported by the Interuniversity Attraction Pole Programme (Belgian Science Policy, PVII-47), the D.G. Higher Education and Scientific Research of the French Community of Belgium (ARC 15/20-065), the F.R.S.-FNRS (Belgium: Grants T.007214 and J.0058.15), and the Belgian Foundation Against Cancer (grant 2014-125). J.Z-R's group is supported by INSERM, the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex Oncolmmunology (investissement d'avenir), Coup d'Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM and Fondation Mérieux. J.U.M. is supported by grants from the German Research Foundation (MA 4443/2-2) and the Volkswagen Foundation (Lichtenberg program)

Author contributions:

C.G. and F.P.L. conceived and designed the study. C.G. performed transcriptomic data analysis and mathematical modelling. M.D-L., K.K. and S. Cordi performed transfection experiments and quantified RNA and proteins. L.G. created the web-based platform under supervision of E.H. A.L. did the survival analysis and helped with transcriptomic data analysis. S. Caruso, G.C. and J.Z-R. provided transcriptomic data and human cell lines and contributed to the interpretation of the modelling results. D.C. and J.U.M. provided primary human HCC cells. J.T. and S.P.M. shared and analysed data from transgenic mice. C.G., M.D-L. and F.P.L. wrote the manuscript with input from all authors.

Abstract

Background and aims : Alterations of individual genes variably affect development of hepatocellular carcinoma (HCC), prompting the need to characterise the function of tumor-promoting genes in the context of gene regulatory networks (GRN).

Methods : Using data from The Cancer Genome Atlas, from the LIRI-JP (Liver Cancer – RIKEN, JP project), and from our transcriptomic, transfection and mouse transgenic experiments, we identify a GRN which functionally links LIN28B-dependent dedifferentiation with dysfunction of β -CATENIN (*CTNNB1*). We further generated and validated with human cell lines and *in vivo* expression data a quantitative mathematical model of the GRN.

Results : We found that LIN28B and CTNNB1 form a GRN with SMARCA4 (BRG1), Let-7b, SOX9, TP53 and MYC. GRN functionality is detected in HCC and gastrointestinal cancers, but not in other cancer types. GRN status negatively correlates with HCC prognosis, and positively correlates with hyperproliferation, dedifferentiation and HGF/MET pathway activation, suggesting that it contributes to a transcriptomic profile typical of the proliferative class of HCC. The mathematical model predicts how the expression of GRN components changes when the expression of another GRN member varies or is inhibited by a pharmacological drug. The dynamics of GRN component expression reveal distinct cell states that can switch reversibly in normal condition, and irreversibly in HCC. The mathematical model is available via a web-based tool which can evaluate GRN status of HCC samples and predict the impact of therapeutic agents on the GRN.

Conclusions : We conclude that identification and modelling of the GRN provides insight into prognosis and mechanisms of tumor-promoting genes in HCC.

Keywords: hepatocellular carcinoma ; mathematical model ; gene regulatory network ; HGF/MET pathway ; principal component analysis ; microRNA ; LIN28B

Lay summary

Hepatocellular carcinoma (HCC) is a heterogeneous disease driven by the concomitant deregulation of several genes functionally organised as networks. Here, we identified a gene regulatory network involved in a subset of HCCs. This subset is characterised by increased proliferation and poor prognosis. We developed a mathematical model which uncovers the dynamics of the network and allows us to predict the impact of a therapeutic agent, not only on its specific target but on all the genes belonging to the network.

Introduction

Various etiologies are associated with HCC, leading to heterogeneity in clinical outcome, histology, transcriptomic profile and mutational spectrum [1-5]. Such heterogeneity causes a variable response to therapeutic agents, as in mouse models with *Ctnnb1*-induced HCCs which show heterogeneous sensitivity to CTNNB1 inhibitors [6]. Thus, designing novel therapeutic strategies against HCC requires the identification of inhibitors of individual tumor-promoting genes and also the characterisation of the molecular networks in which those genes exert their functions.

Dedifferentiation of hepatic cells contributes to HCC progression [7-9]. In this context, poorly differentiated HCC develop as a result of forced induction of LIN28B, a RNA-binding protein which is repressed during normal hepatic cell differentiation. LIN28B is re-expressed in a subset of human HCC's characterised by high serum levels of α -fetoprotein [10, 11], thereby associating dedifferentiation, HCC progression and LIN28B expression [12]. In parallel, *CTNNB1* is one of the most frequently mutated genes in HCC [13]. Therefore, we here explore the possibility that HCC progression depends on a gene regulatory network (GRN) linking LIN28B-dependent dedifferentiation with CTNNB1 dysfunction.

Several approaches can be used to identify GRNs [14]. Here, we selected a method to capture the biological logic of the system whose size permits to identify its quantitative dynamics. The identification of the network is literature-based and its analysis implements mathematical modelling [15, 16]. We identified a HCC-promoting GRN comprising several members connecting LIN28B with CTNNB1 via Let-7b, MYC, SMARCA4 (BRG1), TP53 and SOX9. We further investigated the system-level dynamics of the GRN using a quantitative mathematical model, which was calibrated and validated using mRNA and protein expression data from HCC cell lines, primary HCC cells, patient databanks and mouse models.

Materials and methods

Data normalization and statistical analysis. Data normalization is described in Supplementary Information. Measured data are means $\pm$ SD. Significance was assessed by Student t-test (p value). For RNASeq data of the HCC cohort from The Cancer Genome Atlas (TCGA), we calculated the adjusted p value using a Benjamini-Hochberg correction (p* value) on the entire transcriptome.

RNASeq and miRNASeq. Data were from TCGA (<http://firebrowse.org/>), and from LIRI-JP (RIKEN) of the International Cancer Genome Consortium (ICGC; <https://dcc.icgc.org/projects/LIRI-JP>). For TCGA cohort, we converted the “scaled_estimate” in the “illuminahisec_rnaseqv2_unc_edu_Level_3_RSEM_genes” file into TPM by multiplying by 10^6 .

Plasmids and microRNAs. pCDNA3.1, pcDNA3-MYC, pCI-neo β -Catenin (CTNNB1^{S33Y}), pBABE-BRG1 were from Thermo Fisher Scientific (Waltham, MA, USA), Wafik El-Deiry (Addgene plasmid # 16011), Bert Vogelstein (Addgene plasmid # 16519) and Robert Kingston (Addgene plasmid # 1959) respectively. The miR mimic control (MIMAT0000039: UCACAACCUCCUAGAAAGAGUAGA) and miR mimic hsa-let-7b: hsa-let-7b-5p (MIMAT0000063: UGAGGUAGUAGGUUGUGUGGUU), hsa-let-7b-3p (MIMAT0004482: CUAUACAACCUACUGCCUUCCC) were from Dharmacon (Lafayette, CO, USA).

Cell culture. Human Huh7, HepG2, Hep3B cell lines and primary human HCC cells HCC31 [17] were grown in DMEM (Lonza, Leusden, Netherlands), 10% foetal bovine serum (5% for HCC31; Merck, Darmstadt Germany), L-Glutamine (2 mM) (Thermo Fisher Scientific), Penicillin-Streptomycin (50 U/mL and 50 μ g/mL) (Gibco™, Waltham, MA, USA) and Amphotericin B (Gibco™) (2.5 μ g/mL). Cells were grown in 60 mm dishes and transfected with 3 μ g plasmid and 120 nM miRmimic, using jetPRIME® (Polyplus-Transfection, Illkirch-Graffenstaden, France) for 48, 72 or 96h, in at least three independent experiments. DNA was transfected 24h after plating the cells. For LIN28 inhibition, Huh7 cells were grown in the presence of 120 μ M (N-

Methyl-*N*-[3-(3-methyl-1,2,4-triazolo[4,3-*b*]pyridazin-6-yl)phenyl]acetamide) for 6 days; the medium with inhibitor was changed every day.

Western blotting. Protocol and antibodies are described in Supplementary Information.

RNA extraction and analysis

Total RNA was isolated from Huh7 cells using Trizol (#1029602, Invitrogen, Life technologies). cDNA synthesis was performed with MMLV reverse transcriptase (#28025-13, Invitrogen, Life technologies) according to manufacturer's protocol. MicroRNA expression (Let-7a and Let-7b) was quantified by RT-qPCR using Kapa SYBR Fast 2X Universal Master Mix (#KK4601, Sopachem, Ochten, Netherlands). Specific stem-loop primers were used for reverse transcription, and RT-qPCR was performed using a specific forward primer and a common universal reverse primer.

Let-7a Fwd/Rev:

ACACTCCAGCTGGGTGAGGTAGTAGGTTG/CTCAACTGGTGTCGTGGAGTCGGC
AATTCAGTTGAGACTATACA; *Let-7b* Fwd/Rev:

ACACTCCAGCTGGGTGAGGTAGTAGGTTGT/
CTCAACTGGTGTCGTGGAGTCGGCAATTCAGTTGAGAACCACAC; *ACTB*

Fwd/Rev: TCCTGAGCGCAAGTACTCTGT/CTGATCCACATCTGCTGGAAG. Each ΔC_t between the measured transcripts and the housekeeping genes was normalised to their control conditions by using $2^{-\Delta\Delta C_t}$ method.

Results

Identification of a gene regulatory network driving hepatocellular carcinoma.

We selected an approach in which GRN members meet stringent functional criteria. First, their role in tumor promotion must be validated by animal experimentation and/or high-throughput sequencing data from patients. Second, GRN members must be connected by direct or indirect functional links characterised by protein-protein and protein-DNA interactions, or epistatic relationship identified in loss- and gain- of function analyses. By combining data from the literature, we first reconstituted a GRN

comprising 7 cross-regulating components: the miRNA *Let-7b*, the RNA-binding protein LIN28B, the ATP-dependent helicase SMARCA4, and the transcription factors SOX9, MYC, CTNNB1 and TP53 (Fig. 1A; Supplementary Information, Section 1). When interactions have been identified in non-hepatic cells, we verified whether they also occurred in cultured HCC cell lines (see below).

Data from TCGA (Fig. 1A) showed that expression of *Let-7b* is reduced in HCC as compared to adjacent non tumor tissue. This is consistent with the concomitant overexpression of LIN28B, a repressor of *Let-7b*. All other GRN components, except MYC, were increased in HCC. Further, a PCA analysis, which considers only the expression of the 7 GRN components in 50 non tumor controls and 368 HCCs revealed that non tumor tissue and HCC tumors clustered separately (Fig. 1B and Supplementary Fig. 1A). In addition, we clustered the HCC cohort in two groups of 100 samples based on the low versus high expression levels of each GRN component. This analysis showed that, except for *Let-7b*, high expression of a component is associated with high expression of the other GRN components. ~~Additional analysis also indicated that the expression of the GRN components is correlated~~ (Supplementary Figs. ~~1C-D and~~ 2; Supplementary Information, Section 2).

PCA investigating multiple targets of the GRN components (Supplementary Table 1) identified separate clusters for non tumor tissue and HCC tumors, indicating that the components of the GRN were not only dysregulated in HCC but also that they were actively controlling their targets (Fig. 1C and Supplementary Fig. 1B). Therefore, we considered that concomitant overexpression of the GRN components LIN28B, SMARCA4, SOX9, CTNNB1 and TP53, and downregulation of *Let-7b* is indicative of GRN activity.

Out of the 368 HCCs from the TCGA cohort whose GRN component expression had been analysed by PCA we selected the 150 samples with the highest Dimension 1 to which *CTNNB1*, *SMARCA4*, *SOX9*, *MYC* and *TP53* expression contribute the most, and out of the latter we selected the 100 samples with highest Dimension 2 to which *LIN28B* and *Let-7b* contribute the most (Supplementary Fig. 1A and red dots in Fig. 1B). These 100 samples were defined hereafter as displaying high GRN status and were compared with the other 268 tumors, which were defined as low GRN status,

and with non tumor liver tissue. The expression of the GRN components differed significantly between the tumors with low and high GRN status (Fig. 1D). Moreover, the samples with high GRN status displayed the highest expression of proliferation, embryonic and oncogenic markers, and the lowest levels of differentiation markers (Fig. 1D).

HCC's are divided in proliferative and nonproliferative classes [13]. GRN status, defined as above, correlated with increased MET signaling, a pathway whose activation characterises a subset of the proliferative HCC class [18]. Indeed, HCC's with low or high GRN status clustered separately from the non tumor samples in a PCA plot when considering the expression of 110 HGF/MET target genes (Fig. 2A). In addition, the 100 tumors with highest HGF/MET target gene expression displayed increased expression of *CTNNB1*, *SOX9*, *SMARCA4*, *MYC* and *TP53* as compared to the other 268 tumors (Fig. 2B).

Earlier transcriptomic analyses classified HCC patients in 6 subgroups, G1 to G6 ~~according to a predictor formula that considered the expression of 16 genes~~ [19]. PCA analysis using ~~those 16 genes, or~~ 216 randomly selected genes that are differentially expressed in G1 to G6, revealed that tumors from the TCGA cohort with high GRN status cluster the farthest from the non tumor tissue (Fig. 2C-D). Among the 216 genes differentially expressed between the G1 to G6 subgroups, the 50 genes that contribute the most to the Dimension 1 in the PCA plot were all upregulated in G1, G2 and G3 and were mainly involved in cell proliferation (Supplementary Table 2), supporting that GRN functionality may contribute to confer a transcriptomic profile typical of the HCC proliferation class.

We also investigated if GRN status correlated with prognosis. The survival probability was lower in the 100 HCC patients with high GRN status than in the 268 HCC patients with low GRN status (Fig. 1E). When performing the same analysis with individual components of the GRN, the expression of *Let-7b*, *MYC* and *TP53* did not correlate with survival, while expression of *CTNNB1*, *SOX9*, *SMARCA4* and *LIN28B* were inversely correlated with survival (Supplementary Fig. 3). Therefore, the status of the GRN is a good marker of prognosis.

To determine if the characteristics of HCC with high GRN status are valid for a distinct patient cohort, we analysed the GRN in the LIRI-JP (Riken) cohort from the ICGC. As in the TCGA cohort, the HCC with low or high GRN status clustered separately (Fig. 3A). The GRN is also functional in LIRI-JP since HCC with high GRN status overexpress the GRN target genes (Fig. 3A). HCC with high GRN status also overexpress proliferation markers and the HGF/MET target genes (Fig. 3B-C).

In conclusion, we have identified a GRN of functionally interacting partners which are misexpressed in HCC. The status of the GRN, characterised by consistent and concomitant misexpression of its components and targets, correlates with proliferation, dedifferentiation and prognosis.

Mathematical model of the GRN. To develop a tool that determines how variation of an individual GRN component affects the expression of all the others, we built a quantitative mathematical model, using a set of 20 kinetic equations describing the expression of each network component (mRNA and protein) as a function of time (Supplementary Information, Sections 3-4). We quantitatively calibrated the model using the mRNA levels of all GRN components available in TCGA and in a dataset from 34 human HCC cell lines [20] (Supplementary Fig. 4A-C). Two sets of parameter values and initial conditions were determined for normal tissues on one hand, and for HCC tumors and cell lines on the other hand (Supplementary Tables 3-7).

To quantitatively calibrate the model with protein expression values we transiently overexpressed GRN components in cultured HCC cell lines. These experiments validated in HCC the cross-regulating pairs of the GRN which had originally been identified in non-hepatic tumors. Indeed, overexpression of MYC in HepG2 or Huh7 cells stimulated TP53 and LIN28B expression, thereby validating in HCC the MYC→TP53 and MYC→LIN28B interactions (Fig. 4A and Supplementary Fig. 5A). Similarly, overexpression of constitutively active CTNNB1^{S33Y} and of SMARCA4 in HepG2 or Huh7 cells validated the CTNNB1→LIN28B and CTNNB1→MYC and SMARCA4→MYC interactions in HCC (Fig. 4B and Supplementary Fig. 5B-C).

Finally, transient transfection of *Let-7b-5p* mimic RNA repressed LIN28B, MYC and SOX9, confirming the regulatory links in the GRN (Fig. 4C).

The duration of the transient transfections in Fig. 4 (48h, 72h, 96h) reflected the minimal time-lapse required to monitor a significant change in TP53, LIN28B, MYC or SOX9 protein following overexpression of their respective stimulator or repressor. Therefore, both the level of protein induction and the timing required to observe a significant change in protein level were used to further calibrate the mathematical model, while maintaining the quantitative calibration of all mRNA expression levels. We then simulated the transient transfection conditions by increasing the transcription rate constants of MYC, CTNNB1 or *Let-7b*, starting at $t = 0h$ (Fig. 4A-C, right). The results showed that simulating a 10-fold increase in MYC -which mimics the observed 10-fold increase in transfected MYC protein- predicted a ~2-fold increase in LIN28B and TP53 after 72h; this predicted increase in LIN28B and TP53 matched closely the measured values (compare black and orange bars in Fig. 4A). Similarly, simulating an 8-fold increase in CTNNB1^{S33Y} predicted a ~3- and 11-fold induction of LIN28B and MYC after 48h; these inductions again matched closely the measured 4- and 9-fold induction of LIN28B and MYC, following an 8-fold increase in transfected CTNNB1^{S33Y} (Fig. 4B). Moreover, following a simulated 8-fold induction of *Let-7b* the model reproduced quantitatively the experimentally measured impact of *Let-7b* induction on LIN28B, MYC, and SOX9 (Fig. 4C). Finally, *Let-7b* induction in Huh7 cells did not affect CTNNB1 expression, which fitted with the simulation (Fig. 4C). We concluded that the mathematical model faithfully recapitulates the expression of GRN components.

Validation of the mathematical model in mouse and human. To validate the model, we challenged it by predicting the impact of a stabilising CTNNB1 mutation on LIN28B and *Let-7b* expression and compared the prediction with values from TCGA. To this end we first partitioned the HCC samples of the TCGA database in two cohorts characterised by the presence or absence of stabilising CTNNB1 mutation, the most frequent CTNNB1 mutations being missense mutations in exon 3. This revealed that CTNNB1 mRNA levels were higher in the presence of a mutation (Fig. 5A). Therefore, to simulate CTNNB1 stabilization in the model we increased the CTNNB1 mRNA synthesis rate, increased the activation rate of the inactive

(complexed) CTNNB1 form, and decreased the inactivation rate of the active (stabilised) form. Under these conditions, when *CTNNB1* mRNA values from the mutated and non-mutated samples were introduced in the model (Fig. 5A, orange bars), it faithfully predicted the reduction in *Let-7b* observed in non-mutated HCC samples as well as the slightly larger decrease of *Let-7b* seen in HCC samples with *CTNNB1* mutation. The model also predicted the increased *LIN28B* expression in HCC which was similar whether *CTNNB1* was mutated or not (compare orange and black bars in Fig. 5A).

A second validation was obtained by considering previous results from a mouse model of MYC-induced liver cancer [12]. We simulated a 4000-fold increase in *MYC* mRNA, which corresponds to the experimental induction of *Myc* in mice. Our simulation predicted a MYC-induced 7000-fold overexpression of *Lin28b* mRNA and a 0.4-fold reduction in *Let-7b*. This prediction fitted well with the experimentally-measured induction of *Lin28b* and reduction of *Let-7b* (Fig. 5B).

Third, as a proof of concept that the mathematical model can predict the impact of a pharmacological inhibitor of a GRN component, we simulated the inhibition of LIN28B in HCC: simulating 60 % inhibition of LIN28B protein predicted an increase in *Let-7b* and a decrease in SOX9 and MYC protein; CTNNB1 remained unaffected (Fig. 6). We then evaluated the results of the simulation in a HCC cell line, Huh7 (Fig. 6A), and in HCC31 primary liver cancer cells [17] (Fig. 6B), by growing these cells in the presence of LIN28 inhibitor (*N*-Methyl-*N*-[3-(3-methyl-1,2,4-triazolo[4,3-*b*]pyridazin-6-yl)phenyl]acetamide) [21]. The inhibitor reduced cell proliferation (not shown), and, as expected, induced the expression of *Let-7b* and *Let-7a*. Importantly, the changes of SOX9, MYC and CTNNB1 protein levels measured by western blot matched well with the simulations (Fig. 6), thereby validating the mathematical model as a tool to predict the impact of a pharmacological inhibitor on the GRN.

GRN dynamics are characterised by a bistable switch. To characterise the dynamical properties of the GRN and determine whether they identify distinct HCC cell states, we modeled the steady-state levels of *Let-7b*, selected as a representative variable, as a function of *CTNNB1* mRNA. In normal conditions, high *Let-7b* was associated with low *CTNNB1* mRNA, and *vice versa* (Fig. 5C, left). The

system exhibits a reversible bistable switch from high to low levels of *Let-7b*, at supraphysiological levels of *CTNNB1* mRNA. Green circles in Fig. 5C represent measured *Let-7b/CTNNB1* mRNA values in control samples from TCGA. However, simulating *Let-7b* as a function of *CTNNB1* mRNA in HCC conditions revealed an irreversible bistable switch occurring at low *CTNNB1* mRNA levels (Fig. 5D). Indeed, rising *CTNNB1* mRNA from low to supraphysiological levels in HCC would induce a switch from high to low *Let-7* expression, but reverting from supraphysiological *CTNNB1* to low *CTNNB1* levels would not allow to restore high *Let-7* levels. The red circles in Fig. 5D reflect tumor-to-tumor heterogeneity of the HCC's and are predominantly positioned around the upper branch of the bistable switch. Blue dots in Fig. 5D represent *Let-7b/CTNNB1* values in 34 human HCC cell lines and reflect the cell line-to-cell line heterogeneity. These dots are predominantly positioned around the lower branch of the bistable switch, suggesting that culture conditions of cell lines promote switching of the GRN to a state distinct from the mean cell state in patient tumors. In non tumor condition (Fig. 5C), the expression of the GRN components is well below that of the threshold at which the bistable switch occurs. Therefore, a bistable switch is unlikely to contribute to the initiation of the disease. In contrast, in HCC, the expression level of GRN components in tumors and cell lines is around the threshold of the bistable switch, suggesting that the bistable switch can contribute to HCC progression (Fig. 5D).

The distribution of measured *Let-7b/CTNNB1* values in Fig. 5C-D reflects inter-sample heterogeneity. When modelling a normal heterogeneous cell population with 30% of random variations around the basal value of each parameter, the *Let-7b/CTNNB1* values remained clustered in the upper stable steady state (Fig. 5C, right). Modelling a HCC cell population with the same % of parameter variation generated a distribution of *Let-7b/CTNNB1* values similar to that seen in HCC patients and cell lines (Fig. 5D, right), indicating that the model accounts for the sample's heterogeneity.

We analysed the robustness of the bistable switch dynamics in normal and HCC conditions towards variation in the value of different parameters (Supplementary Figs. 6-7). The results showed that the GRN dynamics rest on a robust reversible bistable switch that should not occur in normal condition, while it exhibits a sensitive

irreversible bistable switch in HCC condition that could cause heterogeneity among tumor samples.

GRN status in gastrointestinal cancers

We next verified if the GRN might be activated in cancers distinct from HCC. We looked at RNA expression in TCGA cohorts of cholangiocarcinoma (CHOL), stomach and esophageal carcinoma (STES), colorectal adenocarcinoma (COADREAD), thyroid carcinoma (THCA), kidney renal clear cell carcinoma (KIRC), breast carcinoma (BRCA), bladder urothelial carcinoma (BLCA), and lung adenocarcinoma (LUAD) (Fig. 7A and Supplementary Fig. 8). Significant and consistent induction of all tumor-promoting GRN components was detected in cholangiocarcinoma, stomach and esophageal carcinoma, and colorectal adenocarcinoma, *i.e.* in gastrointestinal cancers. *Let-7b* did not show the expected reduction in those tumors, yet depending on the tumor type, other *Let-7* family members were downregulated, like in cholangiocarcinoma where *Let-7c* is strongly repressed. The other cancer types did not show a consistent increase in GRN components, suggesting that the GRN does not display tumor-promoting activity in these tumors.

Tu and coworkers developed a mouse model of colorectal cancer with similar gene expression pattern as in human cancer [22], which offered the opportunity to validate the mathematical model in this cancer type. The published mouse data showed that induction of *Lin28b* for 6 months triggered cell proliferation, downregulation of *Let-7* and upregulation of *Sox9*, the latter being illustrated by immunostaining [22]. The subsequent removal of *Lin28b* for 2 months partially restored the expression of *Let-7* and proliferation markers. As in the experiments [22], the mathematical model predicted that a 6-month induction of *LIN28B* causes downregulation of *Let-7b* and upregulation of *SOX9*, and that those expression levels partially revert to near normal values 2 months after *LIN28* removal (Fig. 7B), thereby validating the GRN model in colorectal cancer. The mathematical model also showed that the expression of *MYC* (Fig. 7B) and *CTNNB1* mRNAs (not shown) is not affected by *LIN28B* induction, which fitted well with the observations [22].

Discussion

Here we identified a GRN which combines CTNNB1, SMARCA4, SOX9, LIN28B, Let-7b, TP53 and MYC. *MYC* expression is not strongly affected but it plays a role in mouse models of HCC [12]. Also, when comparing the HCC samples with low MYC expression with those expressing high levels of MYC, we found that MYC expression positively correlates with nearly all GRN components, namely CTNNB1, SOX9, SMARCA4 and TP53 (Supplementary Fig. 2).

Since the mathematical model resorts to two sets of parameters, one for normal condition and one for tumors and cell lines, it accounts for the GRN dynamics in each state, but not for the dynamical transition between normal and HCC states. Indeed, to properly calibrate the model on mRNA expression, the transcription rates of *LIN28B*, *SOX9*, mutant *TP53*, *CTNNB1* and *SMARCA4* had to be increased in the HCC conditions (*'tumor'* parameter in Supplementary Tables 4-5). This indicates that the interactions between GRN components are not sufficient to account for the normal-to-HCC transition. The parameters implicitly integrate the impact of external regulators of the GRN. Yet, the modelling strategy cannot integrate the full spectrum of regulations, and a number of regulators might not be known. Since modelling the transition from normal to HCC requires adaptation of the transcription rates of *LIN28B*, *SOX9*, mutant *TP53*, *CTNNB1* and *SMARCA4*, we suggest that the mechanisms controlling the expression of those genes warrant further investigation.

Classifications of HCC's identified proliferative *versus* non proliferative classes, or subdivided HCC patients in subgroups G1 to G6 [19]. Consistently, the GRN is most activated in the G1, G2 and G3 subgroups which are associated with poor differentiation, severe prognosis, and overexpression of genes regulating cell proliferation [19]. The targets of CTNNB1 are heterogeneous as CTNNB1 induces progenitor-type genes in tumor cells such as cyclin-D1 or VEGF-A, but also regulates expression of antioxidant, pro-survival and pro-hepatocyte differentiation genes in normal hepatocytes [23]. The G5 and G6 subgroups are associated with activation of

CTNNB1 targets typical for mature hepatocytes, whereas the G1, G2, and G3 subgroups show predominant activation of progenitor-type targets. Interestingly, high GRN status was associated with high expression of progenitor-type CTNNB1/TCF/LEF targets (Supplementary Fig. 9 and Supplementary Table 8). *TERT* promoter mutation are found in 60% of HCC's [13, 24]. However, out of the 368 HCC samples of the TCGA cohort, the 100 samples with highest *TERT* expression did not consistently misexpress GRN components as compared to the 268 HCC samples with lower *TERT* expression levels (Supplementary Fig. 9C).

The GRN comprises positive feedback loops which are at the origin of bistable switches. To our knowledge, our work identified the first potential irreversible bistable switch involved in a specific subset of HCC tumors. The identification of irreversible states in HCC provides evidence that targeting specific GRN components using drugs might be therapeutically ineffective when cancer cells are in a locked state with regard to the function of the drug target. Also, our comparison of patient tumors and cultured HCC lines indicates that cell lines may be in a state distinct from that of tumors with regard to the function of the GRN.

To facilitate the analysis of the GRN, we set up a web-platform allowing to check GRN status in new samples of HCC and gastrointestinal tumors. This platform presents a graphical user interface that integrates the expression levels of the GRN components of tumor samples, and which implements the mathematical model to test which component of the network is the best target to modulate the network dynamics (<http://biomodelling.eu/apps.html>).

In conclusion, we anticipate that identification and dynamical modelling of a set of GRNs composed of interacting components that are often mutated and/or dysregulated in HCC, as is the case here for CTNNB1 and Let-7/LIN28 [6, 10, 12, 25], will contribute to provide a global picture of tumor-promoting gene function in HCC. Our study presents the concept of tools that help in the design of bespoke therapies for treating HCC.

Acknowledgements

The authors thank P. Jacquemin, C. Pierreux, L. Desmet and the Lemaigre laboratory members for help, and L. Nguyen and H. Zhu for sharing information.

References

- [1] Calderaro J, Couchy G, Imbeaud S, Amaddeo G, Letouze E, Blanc JF, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J Hepatol* 2017;67:727-738.
- [2] Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science* 2017;357: pii: eaan2507.
- [3] Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015;47:505-511.
- [4] Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 2017;169:1327-1341 e1323.
- [5] Breuhahn K, Gores G, Schirmacher P. Strategies for hepatocellular carcinoma therapy and diagnostics: lessons learned from high throughput and profiling approaches. *Hepatology* 2011;53:2112-2121.
- [6] Tao J, Zhang R, Singh S, Poddar M, Xu E, Oertel M, et al. Targeting beta-catenin in hepatocellular cancers induced by coexpression of mutant beta-catenin and K-Ras in mice. *Hepatology* 2017;65:1581-1599.
- [7] Zeng X, Lin Y, Yin C, Zhang X, Ning BF, Zhang Q, et al. Recombinant adenovirus carrying the hepatocyte nuclear factor-1alpha gene inhibits hepatocellular carcinoma xenograft growth in mice. *Hepatology* 2011;54:2036-2047.
- [8] Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;31:339-346.
- [9] Jung KH, Zhang J, Zhou C, Shen H, Gagea M, Rodriguez-Aguayo C, et al. Differentiation therapy for hepatocellular carcinoma: Multifaceted effects of miR-148a on tumor growth and phenotype and liver fibrosis. *Hepatology* 2016;63:864-879.
- [10] Takashima Y, Terada M, Udon M, Miura S, Yamamoto J, Suzuki A. Suppression of lethal-7b and miR-125a/b Maturation by Lin28b Enables

- Maintenance of Stem Cell Properties in Hepatoblasts. *Hepatology* 2016;64:245-260.
- [11]Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, et al. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet* 2009;41:843-848.
- [12]Nguyen LH, Robinton DA, Seligson MT, Wu L, Li L, Rakheja D, et al. Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* 2014;26:248-261.
- [13]Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* 2015;149:1226-1239 e1224.
- [14]Wilkinson AC, Nakauchi H, Gottgens B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst* 2017;5:319-331.
- [15]Rata S, Suarez Peredo Rodriguez MF, Joseph S, Peter N, Echegaray Iturra F, Yang F, et al. Two Interlinked Bistable Switches Govern Mitotic Control in Mammalian Cells. *Curr Biol* 2018;28:3824-3832 e3826.
- [16]Ryu H, Chung M, Dobrzynski M, Fey D, Blum Y, Lee SS, et al. Frequency modulation of ERK activation dynamics rewires cell fate. *Mol Syst Biol* 2015;11:838.
- [17]Castven D, Becker D, Czauderna C, Wilhelm D, Andersen JB, Strand S, et al. Application of patient-derived liver cancer cells for phenotypic characterization and therapeutic target identification. *Int J Cancer* 2018; Epub ahead of print.
- [18]Kaposi-Novak P, Lee JS, Gomez-Quiroz L, Coulouarn C, Factor VM, Thorgeirsson SS. Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. *J Clin Invest* 2006;116:1582-1595.
- [19]Boyault S, Rickman DS, de Reynies A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007;45:42-52.
- [20]Rebouissou S, La Bella T, Rekik S, Imbeaud S, Calatayud AL, Rohr-Udilova N, et al. Proliferation Markers Are Associated with MET Expression in Hepatocellular Carcinoma and Predict Tivantinib Sensitivity In Vitro. *Clin Cancer Res* 2017;23:4364-4375.

- [21]Roos M, Pradere U, Ngondo RP, Behera A, Allegrini S, Civenni G, et al. A Small-Molecule Inhibitor of Lin28. *ACS Chem Biol* 2016;11:2773-2781.
- [22]Tu HC, Schwitalla S, Qian Z, LaPier GS, Yermalovich A, Ku YC, et al. LIN28 cooperates with WNT signaling to drive invasive intestinal and colorectal adenocarcinoma in mice and humans. *Genes Dev* 2015;29:1074-1086.
- [23]Nejak-Bowen KN, Monga SP. Beta-catenin signaling, liver regeneration and hepatocellular cancer: sorting the good from the bad. *Semin Cancer Biol* 2011;21:44-58.
- [24]Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun* 2013;4:2218.
- [25]Wu L, Nguyen LH, Zhou K, de Soysa TY, Li L, Miller JB, et al. Precise let-7 expression levels balance organ regeneration against tumor suppression. *Elife* 2015;4:e09431.

Figure legends

Figure 1. Identification of a GRN involved in HCC. (A) Structure of the GRN (left), and RNA levels of GRN components in non-tumor tissue (n=50) and HCC tumors (n=368) from the HCC cohort in TCGA (right). (B) PCA plots based on the expression of the GRN components, and (C) on selected GRN targets (Supplementary Table 1) in control and HCC samples. (D) Tumors were ranked according to low or high GRN status. Expression of the GRN components correlated well with high GRN status (top); and with expression of proliferation (middle), embryonic and oncogenic markers (bottom), and with lowest levels of the differentiation marker miR-122 (bottom). (E) Kaplan-Meier curves (overall survival) showed that patients with high GRN status (median time = 1372 days, 40 deceased patients amongst 100) exhibited a lower survival probability than patients with low GRN status (median time = 1852 days, 90 dead patients amongst 268) (Wilcoxon test, $p = 0.02$). p^* = p value with BH correction on the entire transcriptome.

Figure 2. The gene regulatory network promotes a proliferative HCC phenotype. (A) Non-tumor samples (blue; n=50) and HCC's (n=368) cluster separately in a PCA analysis based on the expression of 110 HGF/MET target genes (Table 1 in Ref. [18]). (B) Expression of GRN components in non tumor samples (blue, n=50) and in HCC with low (grey; n=268) or high HGF/MET target expression (red; n=100). 100 tumor samples with high HGF/MET signaling were selected by PCA analysis based on the expression levels 110 of HGF/MET target genes (panel A). (C, ~~D~~) Clustering of normal (blue) *versus* HCC with low (grey) and high GRN status (red) in a PCA analysis based on ~~(C)~~ the expression of ~~16 predictor genes and (D) of~~ 216 differentially expressed genes of the 6-group HCC classification. (B) Data are means $\pm$ SD.

Figure 3. Validation of the GRN in the LIRI-JP HCC cohort. (A, *left*) mRNA levels of GRN components in non-tumor tissue (n=202) and HCC tumors (n=243) from the HCC cohort (LIRI-JP) in ICGC. (A, *middle*) PCA plots based on the expression of the GRN components, and (A, *right*) on GRN targets (listed in Supplementary Table 1) in control (blue), HCC low (grey) or high GRN status (red). (B) Expression of the GRN

components correlated well with expression of proliferation markers. (C, *left*) Expression of GRN components in non tumor samples (blue, n=202) and in HCC with low (grey; n=142) or high HGF/MET target expression (red; n=100). (C, *right*) Non-tumor samples (blue; n=202), HCC low GRN (grey; n=142) and HCC high GRN status (red; n=100) cluster separately in a PCA analysis based on the expression of 110 HGF/MET target genes. Data are means $\pm$ SD. (D) Kaplan-Meier curves (overall survival) showed that patients with high GRN status (23 deceased patients amongst 86) exhibited a lower survival probability than patients with low GRN status (16 dead patients amongst 122) (Wilcoxon test, $p = 0.003$).

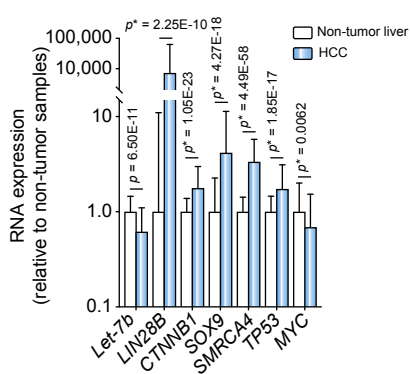
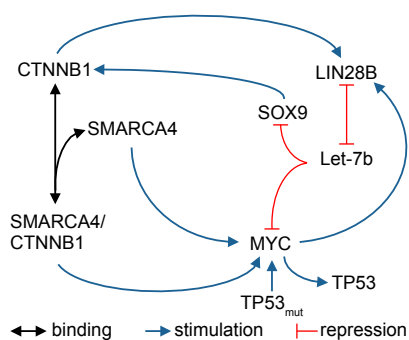
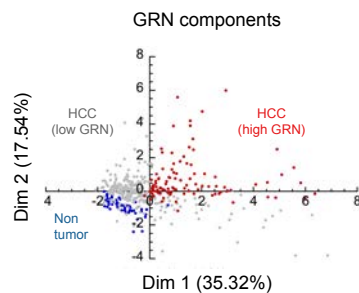
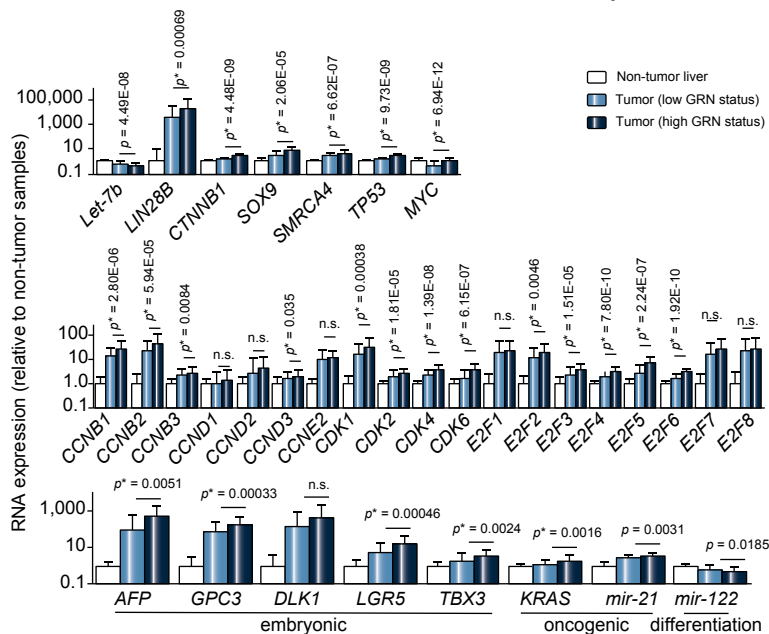
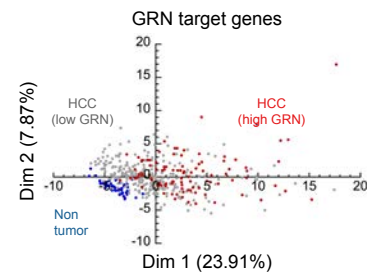
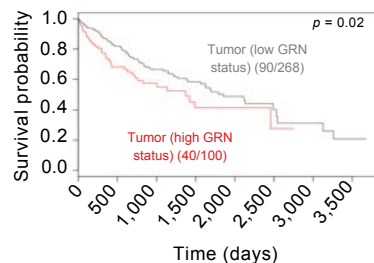
Figure 4. Calibration of the mathematical model on protein expression in human HCC cell lines. (A) Expression of MYC, LIN28B, TP53 and SOX9 protein following MYC overexpression in HepG2 cells. (B) Expression of CTNNB1, LIN28B and MYC protein following CTNNB1^{S33Y} overexpression in HepG2 cells. mCTNNB1 corresponds to endogenous mutant CTNNB1 resulting from inframe deletion of exon 3. (C) Expression of LIN28B, MYC, CTNNB1 and SOX9 in the presence of *Let-7b-5p* mimic RNA in Huh7 cells. Data in bar graphs are means $\pm$ SD, $n \geq 3$. *, $p < 0.05$; **, $p < 0.01$ and *** $p < 0.001$. Orange bars correspond to protein levels calculated in the mathematical model with (A) overexpression of MYC, (B) overexpression of CTNNB1^{S33Y} and (C) overexpression of *Let-7b*. The right panels in A-C show the simulated temporal evolution, with the appropriate time scale, of the GRN proteins following MYC, CTNNB1^{S33Y}, or *Let-7b* mimic transfection. See Supplementary Information for parameter values and initial conditions.

Figure 5. Validation of the model and GRN dynamics in HCC. (A) Expression of stabilised CTNNB1 mutant (n=98) and wild-type CTNNB1 (n=270) in HCC as compared to normal tissue (n=50). The predicted levels of *Let-7b* and *LIN28* following a simulation of increased expression and stabilisation of CTNNB1 fit with the experimental observations. (B) Measured and simulated impact of MYC induction on *LIN28B* mRNA and *Let-7b* expression levels, in a mouse model of MYC-induced HCC. (A, B) Results are means $\pm$ SD. (C) Modelling *Let-7b* levels as a function of CTNNB1 mRNA in normal conditions (left). Values in non tumor tissue (n=50; green circles) are superimposed on the model curve. Solid curves, stable steady states; dashed curves, unstable steady states. Modelling 30% random variations around the

basal value of each parameter in 200 cells resulted in a distribution of *Let-7b/CTNNB1* values in the upper stable state (right) which recapitulated the heterogeneity measured in normal samples. (D) Modelling *Let-7b/CTNNB1* mRNA values in HCC conditions (left). *Let-7b/CTNNB1* mRNA values in 368 HCC tumors (red circles) or 34 cell lines (blue dots) are superimposed on the model curve. Random variation of parameters like in (C) recapitulated the heterogeneity measured in HCC patients and cell lines (right). See Supplementary Information for parameter values and numerical simulations.

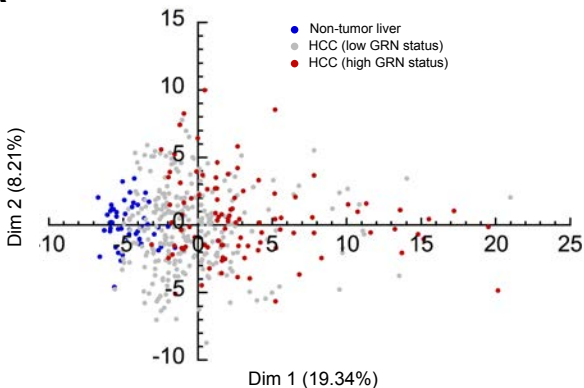
Figure 6. Mathematical model as a tool to predict the impact of LIN28 inhibition on the GRN dynamics. (A, B) Simulating the absence ($K_{\text{ILET7}} = 21$) or presence ($K_{\text{ILET7}} = 82$) of LIN28 inhibition for 6 days (144h) predicts expression levels of *Let-7b*, MYC, CTNNB1 and SOX9 which are close to the levels measured following (A) treatment of Huh7 cells and (B) HCC31 primary liver cancer cells with 120 μM of LIN28 inhibitor [21]. (B) *Let-7* expression levels were simulated and measured after 96h. (A, B) Results are means $\pm$ SD; $n \geq 3$.

Figure 7. Cancer-type specificity of GRN functionality and validation of the model in colorectal cancer. (A) Tumor-promoting components of the GRN are induced in cholangiocarcinoma (CHOL), stomach and esophageal carcinoma (STES), and colorectal adenocarcinoma (COADREAD). Thyroid carcinoma (THCA) had a distinct gene expression profile. Data (mean $\pm$ SD) are from TCGA; the number of normal and tumor samples is 9 and 36 (CHOL), 50 and 600 (STES), 51 and 624 (COADREAD), and 59 and 501 (THCA). (B) RNA levels of *Let-7b*, *Myc* and *Sox9* in experiment (black) and in the model (orange) in non tumor condition, after 6 months of *Lin28b* induction (LIN28B), and after 6 months of *Lin28b* induction followed by 2 months of partial *Lin28b* removal (remLIN28B) (see Supplementary Information for parameter values).

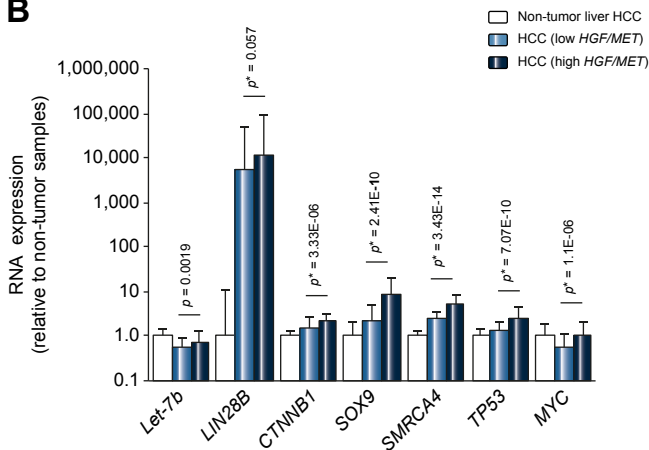
A**B****D****C****E**

A

HGF/MET target genes

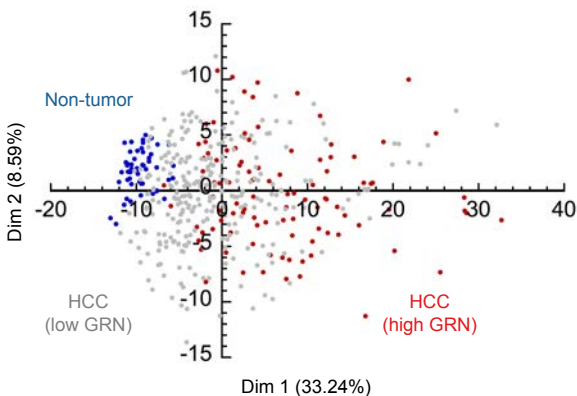


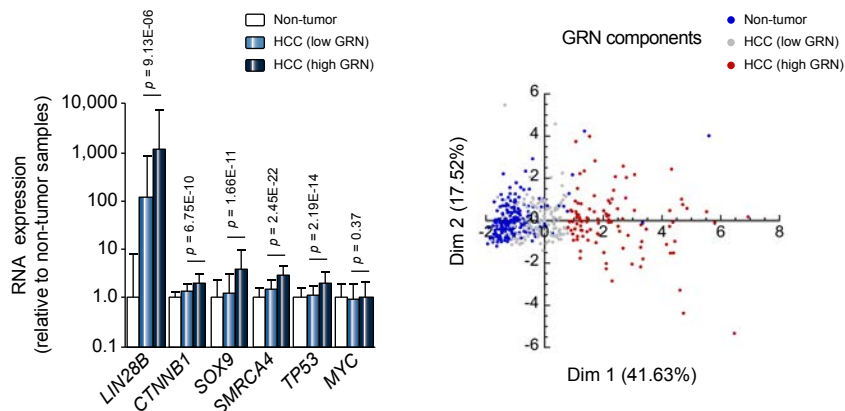
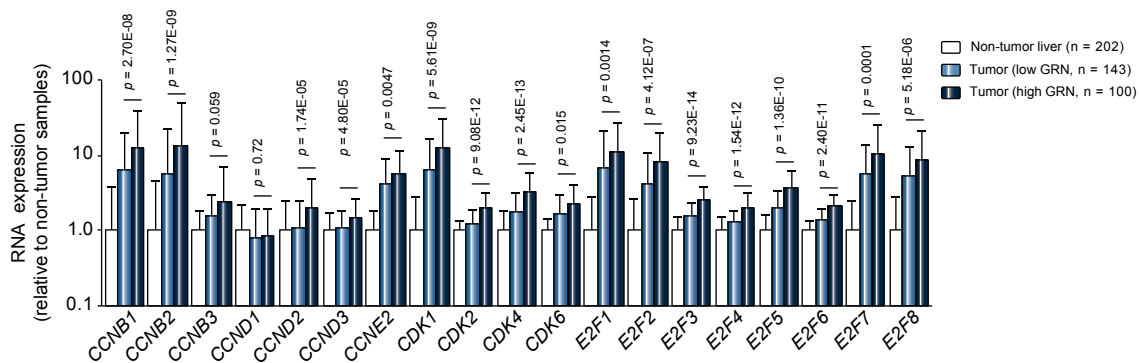
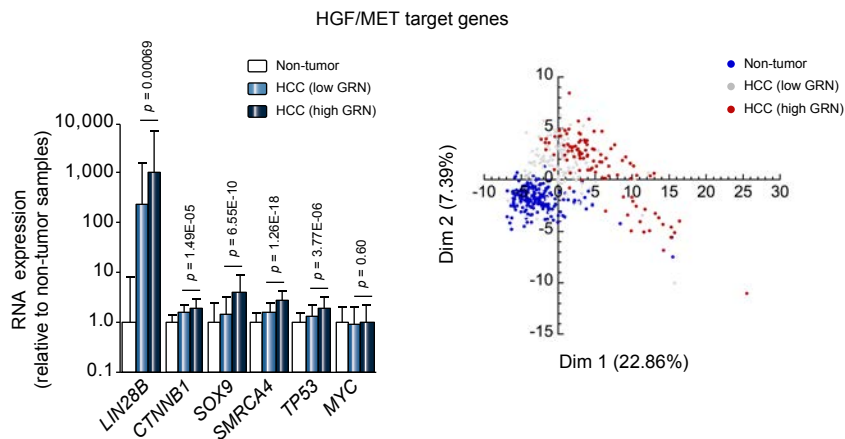
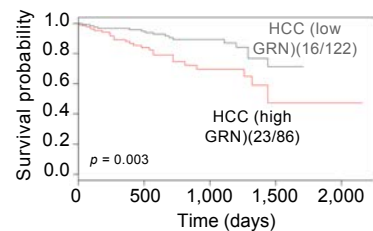
B

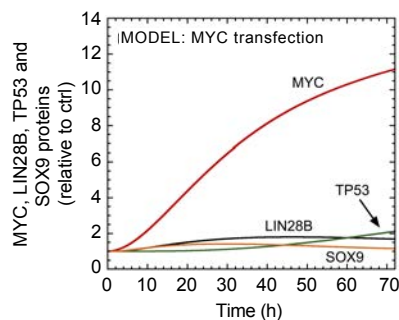
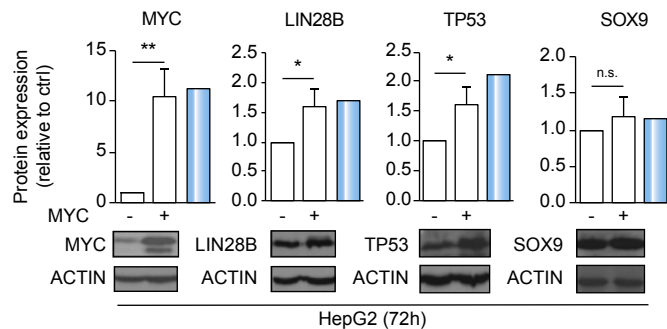
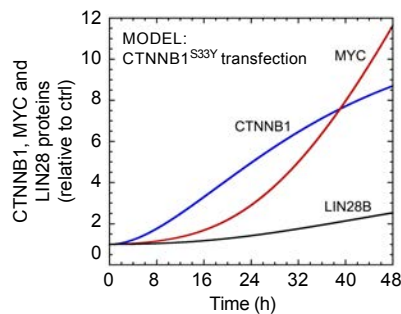
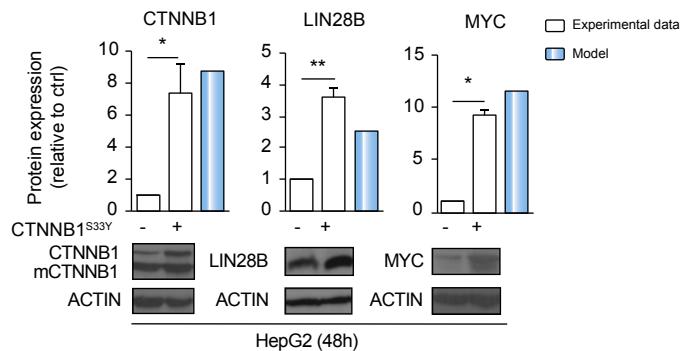
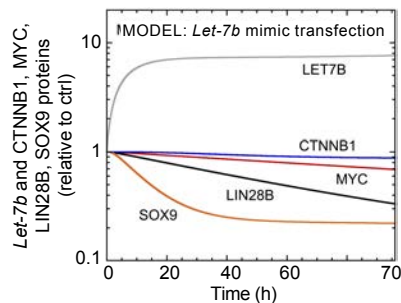
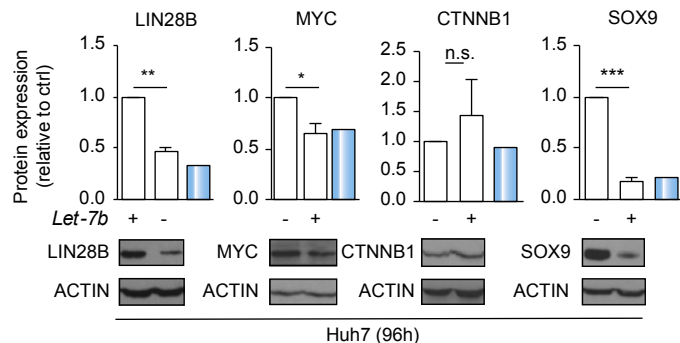


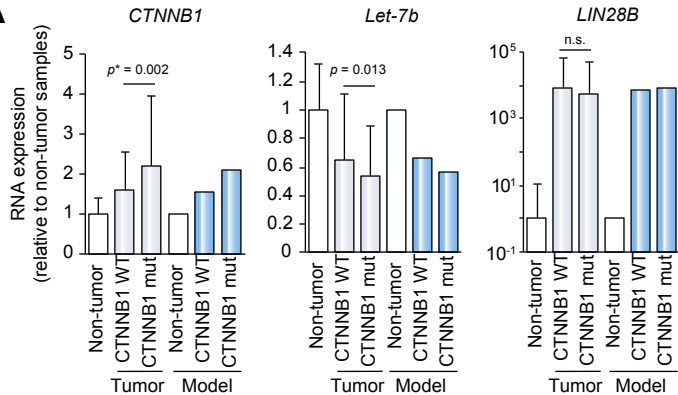
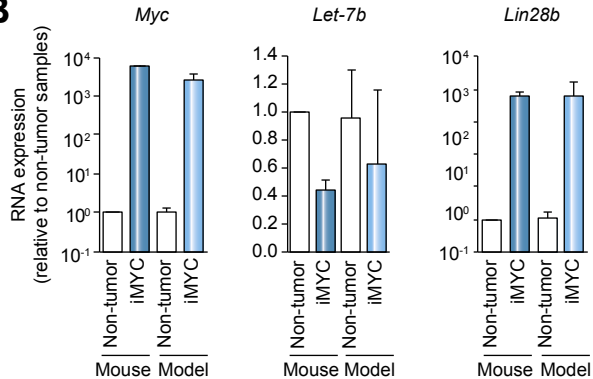
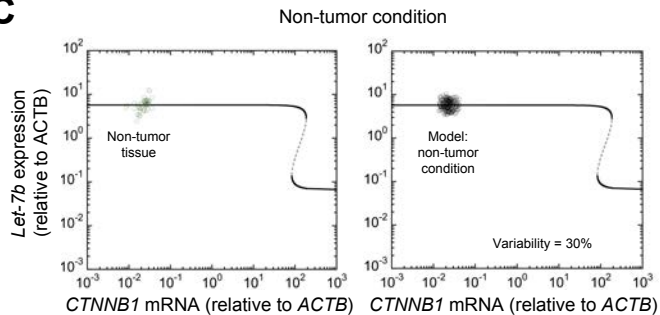
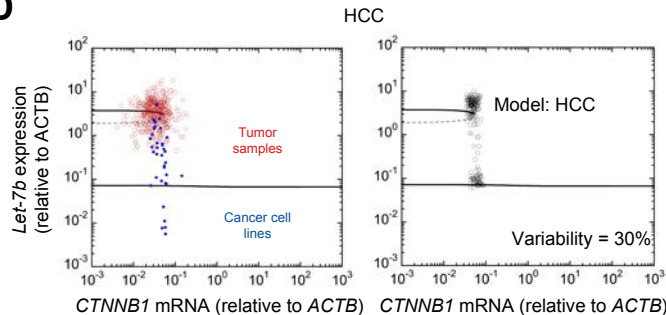
C

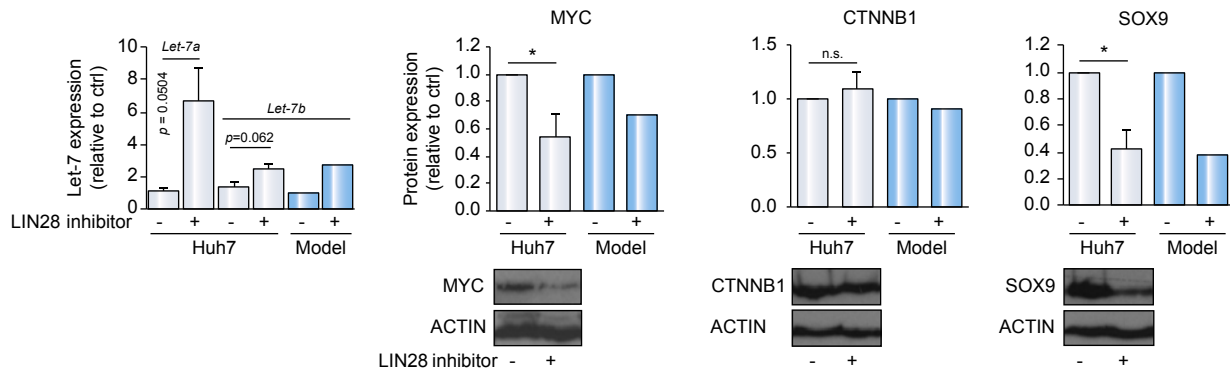
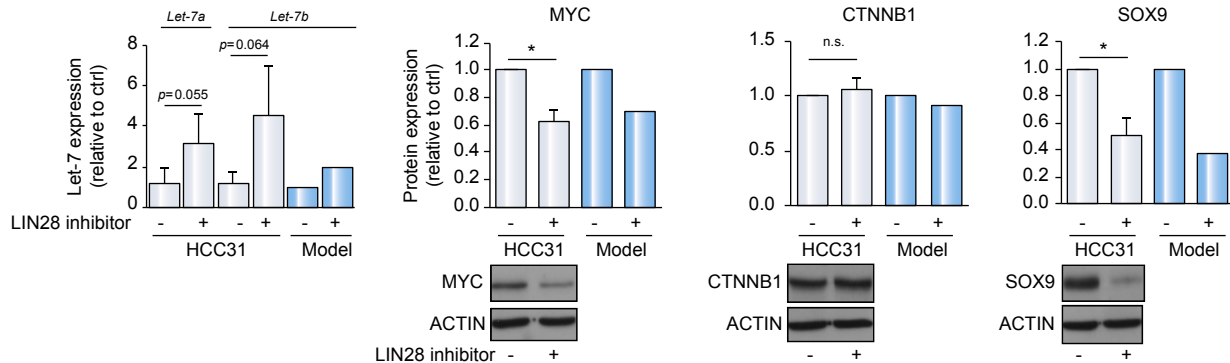
216 genes differentially expressed in G1-G6 subgroups

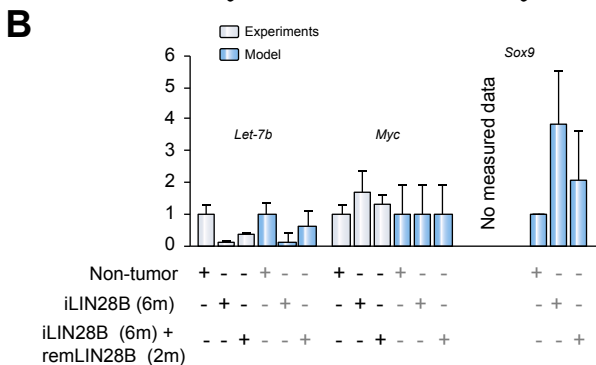
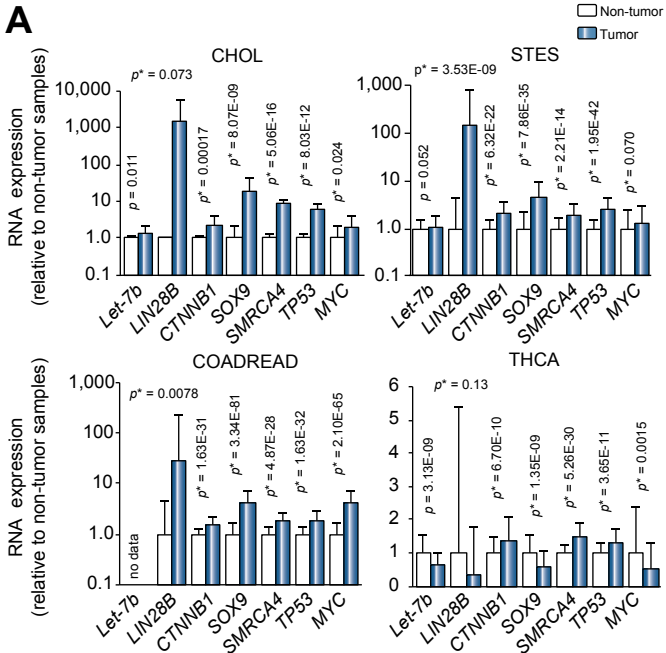


A**B****C****D**

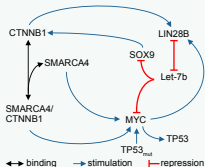
A**B****C**

A**B****C****D**

A**HCC cell line****B****Primary cancer cells**



Structure of the gene
regulatory network
(GRN) driving HCC



Mathematical model:
identification
of the **GRN**
dynamics

**Quantitative model
calibration :**
mRNA and
protein levels
of the GRN

Model predictions:
Impact of each
individual GRN
component on all
other components

Validated model
as a **tool** to predict
the effect of
**pharmacological
inhibitors**

**Challenging the
model for
validation**