Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of The Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Reverse engineering of gene regulatory networks governing cell-cell communication in the microenvironment of pancreatic cancer

Presented by
Zbigniew Rogon-Lamparski, M.Res. M.Sci.
Born in Torun, Poland

Oral-examination:……..

Heidelberg, 2011

Referees:

Prof. Dr. Roland Eils

Prof. Dr. Peter Angel

Table of Contents

Acknowledgements

This thesis represents more than just my work at the computer; it is a milestone and a compilation of experiences, which I have had the privilege to gather at the German Cancer Research Center. Throughout the years I have met here amazing individuals, scientists, and researchers who shaped this work in many ways, I thank them all.

I thank my advisor and boss Prof. Roland Eils, his support, helpful discussions and ideas have been invaluable throughout the progress of this work.

I wish to thank Dr. Nathalia Giese, a collaborator and advisor throughout my thesis, she has the unique ability to inspire excellence.

I thank the remaining members of my thesis committee Prof. Peter Angel for his valuable guidance and help, and Prof. Jan Schmidt for his support.

I thank Dr. Hauke Busch for his support, guidance, knowledge, and invaluable help throughout this process, especially in the beginning when it was needed the most.

I also thank an organization: the DKFZ – German Cancer Research Center, my employer for the past years, I have cherished every moment spent here. The Helmholtz Alliance on Systems Biology, which has supported this work through the network on Systems Biology of Signaling in Cancer (SBCancer).

Finally, to my family: Thank You. Without your continuous, unconditional, and loving support none of this would have been possible.

Abbreviations

AP1 – Activator protein 1 transcriptional complex;

BC – Bayesian Clustering;

CTRNN – Continuous Time Recurrent Neural Network;

ECM - Extracellular Matrix;

EGF – epidermal growth factor;

EGR – Early growth response factor;

EMT – Epithelial-Mesenchymal Transition;

FGF – fibroblast growth factors;

GA – Genetic Algorithm;

GO – Gene Ontology;

GRN – Gene Regulatory Network;

HGF – hepatocyte growth factor;

HGU133Plus2 – Human Genome-wide Affymetrix Microarray Chip;

IGF-1 – Insulin-like growth factor 1;

IQR – interquartile range;

JAK – Janus kinase;

LLE – Largest Lyapunov Exponent;

MAP kinases – mitogen-activated protein kinases;

NFκB – Nuclear factor κ B;

NN – Neural Network;

ODE - Ordinary Differential Equation;

PDAC – Pancreatic Ductal Adenocarcinoma;

PM – Perfect Match;

PSC – Pancreatic Stellate Cells;

RMA – Robust Multi-array Average;

STAT – Signal transducer and activator of transcription;

TC – Tumor Cells;

TF – transcription factor;

TFBSA – transcription factor binding site analysis;

TGFβ – transforming growth factor beta;

1. Abstract

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the leading causes of cancer death, with a five-year survival rate of <5% and a median survival of 6 months. Extensive desmoplastic reaction is a characteristic feature and a prognostic factor of PDAC, which conveys its resistance. Desmoplastic stroma accounts for approx. 90% of tumor volume and consists predominantly of non-malignant fibroblasts (pancreatic stellate cells, PSC). Previous studies have revealed the PSC mesenchymal origins, capacity to switch between quiescent and activated states, proinflammatory features, expression of soluble factors, ability to migrate, and phagocytize.

State of the art: Abundance of stroma has sparked previous attempts to dissect the interactions between PSC and tumor cells (TC) producing a common picture of a microenvironment supporting PDAC development. Unfortunately, focus on snapshot-like analysis has proven difficult to translate into therapeutical advances, as it discards the dynamic interactions in the microenvironment, as well as the temporal dynamics of gene expression itself. Gene regulatory networks (GRN) adapt to environmental cues by rewiring connections between genes, those induced modulations effectively lead to state-transitions e.g. PSC activation, or produce mutually exclusive cell-fate decisions e.g. differentiation, senescence, or death. We recognize that cell-specific assignment of stimuli, identification of genes forming the GRNs, as well as the identification of cellular state-changes remain undiscovered. We hypothesize that at an early stage, the quiescent → activated PSC transition yields a steady state PSC gene regulatory network (GRN), but the subsequent succession of impulse responses along TC→PSC→TC interaction axis drives both cell types into unstable states maintained only for the duration of the direct TC-PSC contact.

Aims: Through the application of a high-throughput complexity reduction approach and *in silico* modeling I aim to reconstruct the GRNs underlying the cell-cell communication, and identify key soluble factors shaping the double-paracrine interactions. I aim to use the models to gain a mechanistic and functional insight into how the cues are integrated and how they affect GRN maintenance. I hope to capture cell-fate decisions and identify key dynamic changes with the ultimate goal of finding genetic markers to aid development of novel therapeutic options for this deadly malignancy.

Results: We have individually stimulated PSC and TC with conditioned supernatant from the respective other cell type and recorded a time-series (1-24h) from which genome-wide microarray expression data has been generated. In this dissertation I used the time-resolved expression profiles to identify significant gene kinetics through an approach-involving gene ranking, filtering, and clustering

followed by gene ontology and pathway analysis. I identified key gene interactions using a genetic algorithm embedded in a continuous time recurrent neural network (CTRNN) modeling scheme. Then I used the derived GRN's to produce a picture of unique intercellular interactions. Through *in silico* simulations with the created models, and subsequent data analysis and interpretation I delivered targets for experimental testing on the inter- as well as intra-cellular levels.

Experimental validation of the selected gene targets using gene silencing and qRT-PCR confirmed the *in silico* predicted TC network behavior; validation of the intercellular connections confirmed their dependence on the identified networks.

2. Zusammenfassung

Hintergrund: Das duktale Adenokarzinom der Bauchspeicheldrüse (PDAC) ist eine der führenden Ursachen für Todesfälle durch Krebs, mit einer 5-Jahresüberlebensrate von <5% und einer mittleren Überlebenszeit von 6 Monaten. Umfangreiche desmoplastische Reaktion ist ein charakteristisches Merkmal und ein prognostischer Faktor für PDAC, welches auch dessen therapeutischen Widerstand vermittelt. Desmoplastisches Stroma bildet ca. 90% des Tumorvolumens und besteht überwiegend aus nicht-malignen Fibroblasten (Pankreas Sternzellen, PSC). Frühere Studien haben den mesenchymalen Ursprung dieser Zellen, ihre Fähigkeit zur Umschaltung zwischen ruhendem und aktiviertem Zustand, ihre proinflammatorischen Eigenschaften, die Expression von löslichen Faktoren, und ihre Fähigkeit zu wandern und zu phagozytieren enthüllt.

State of the art: Erhebliche Auswirkungen des Stromas auf die Tumorprogression wurden durch frühere Versuche gezeigt. Leider konzentrierten sich diese Studien nur auf eine Momentaufnahme und ignorierten dabei die dynamischen Wechselwirkungen in der Mikroumgebung, sowie die zeitaufgelöste Dynamik der Genexpression selbst was die Umsetzung in Therapieansätze behindert hat. Gen-regulatorische Netzwerke (GRN) passen sich dynamisch an Umweltreize an und Beeinflussen den zellulären Zustand. Diese induzierte Modulation führt effektiv zu Zell Veränderungen (e.g. PSC Aktivierung), oder sogar zu sich gegenseitig ausschließenden Zell-Schicksal Entscheidungen (e.g. Differenzierung, Seneszenz oder Tod). In diesem Zusammenhang blieben die Zell-spezifische Zuordnung von Reizen, die Identifizierung der entsprechenden Gene welche die zugrunde liegenden GRN bilden, sowie die Identifizierung von zellulären Zuständen bislang unerforscht. Wir vermuten, dass in einem frühen Stadium der Tumorentwicklung durch die Aktivierung von PSC ein stabiles GRN Netzwerk hergestellt wird, welches dann aber durch die anschließende Sequenz von Impulsantworten entlang der Tumorzellen (TC)→PSC→TC Interaktions Achse destabilisiert wird.

Ziele: Mein Ziel ist es, durch eine Reduktion von Komplexität und den Einsatz von *in silico* Modellierung die zugrunde liegenden Gen-regulatorische Netzwerke zu rekonstruieren und die Faktoren zu identifizieren, welche für die Doppel-Parakrine Stimulierung der Zellen in dieser Mikroumgebung verantwortlich sind. Ein weiteres Ziel ist es einen mechanistischen und funktionellen Einblick in die Signal Integration und dessen Wartung auf der Gen ebene zu gewinnen. Ich hoffe, die Zell-Schicksal Entscheidungen erfassen zu können und wichtige

dynamische Veränderungen zu identifizieren. Dadurch können genetische Marker für die Entwicklung neuer therapeutischer Optionen für diese tödliche Malignität gefunden werden.

Ergebnisse: Durch die Stimulierung von PSC und TC durch konditioniertes Medium von dem jeweils anderen Zelltyp wurden zwei Microarray-Expressionsdaten Zeitreihen von 1-24h Experimental aufgenommen. In dieser Arbeit durch den Ansatz von Gen Einstufung, Filtrierung und Clustering identifiziere ich die zeitaufgelösten Genexpressionsprofile welche dann mit Gen-Ontologie und Pathway-Analyse bearbeitet werden. Mit Hilfe eines genetischen Algorithmus in einem zeitverzögerten rekurrenten neuronalen Netz (CTRNN) konnte ich die Schlüssel-Gen Interaktionen erkennen. Die abgeleiteten Netzwerke habe ich dann zur Erzeugung eines Bildes der einzigartigen interzellulären Wechselwirkungen benutzt. Durch *in silico* Simulationen mit den erstellten Modellen, die anschließende Datenanalyse und Interpretation konnte ich Ziele für die experimentelle Prüfung auf der inter- und intra-zellulären Ebene identifizieren.

Experimentelle Validierung der ausgewählten Gen-Targets mit Gen-Stilllegung und qRT-PCR bestätigte das *in silico* vorhergesagtes Verhalten des TC Netzwerks, zusätzlich konnten wir auch die Abhängigkeit der ausgewählten interzellulären Verbindungen von den identifizierten Netzwerken bestätigen.

3. Introduction

3.1 Cancer

Cancer is a common descriptor of many heterogeneous and multifactorial diseases, which share certain features such as unconstrained growth (proliferation), invasiveness and metastasis of the tumor cells. Approximately two hundred distinct types of cancer have been recognized thus far and divided based on the histology into six major categories: carcinomas, sarcomas, lymphomas, myelomas, leukemias, and teratomas. These major categories, while significantly different from each other, are even more intricate as tumors can form with features intertwining multiple types and thus resulting in the eighth category of mixed types. Carcinomas comprise around 85% of all human cancers including pancreatic, lung, cervical, breast, skin and brain tumors, and are known to be invasive into the surrounding tissues and organs, metastasizing to lymph nodes and other sites.

3.2 Pancreatic cancer

Carcinoma of the exocrine pancreas, also known as the Pancreatic Ductal Adenocarcinoma (PDAC) is the main object of interest in the work presented here. It is a gastrointestinal malignancy and one of the leading causes of cancer death that for a very long time evaded the progress in cancer research, mostly due to the lack of distinctive early symptoms and resulting diagnostic difficulties (Li et al. 2004, Dickman et al. 2006). Patients with carcinoma of the exocrine pancreas have a dismal prognosis with a five-year survival rate of <5% and a median survival of 4-6 months depending on the staging of the disease upon diagnosis (Jemal et al. 2010, Gudjonsson 1987, Warshaw et al. 1992, Cartmel et al. 1997). Exact etiology of PDAC is not known, but alcohol abuse, poor dietary habits often resulting in overweight and obesity, diabetes, smoking, and family history are most often cited as the main risk factors (Lowenfels et al. 1993 and 2004, Weisburger et al. 1995, Klein et al. 2001, Luo 2010).

3.2.1 Pancreatic Ductal Adenocarcinoma (PDAC)

Pancreatic ductal adenocarcinoma constitutes approx. 90% of all primary malignant tumors of the pancreatic gland. The sources of the tumors are either pancreatic ducts (99%) or acinar cells (1%) (Anand et al. 2010). This form of pancreatic cancer is highly aggressive, causing organism-wide devastation with its rapid progression and resistance to all forms of treatment

(DiMagno et al. 1999, Sener et al. 1999, Kornmann et al. 2003). Statistics over the last 10 years show that in only around 7% of all diagnosed cases the tumors were still confined to the primary site, while over 90% of patients were diagnosed in the late stages of the disease. After the tumor has spread to surrounding tissues and lymph nodes treatment is impossible (Xu et al. 2010; WHO Data and Statistics 2010). Only early diagnosis combined with tumor resectability and staging (size, spread) determines the survival prognosis and outcome (Table 1), and only primary, non-metastatic tumors offer a chance for treatment. The only successful treatment option is a pancreaticoduedenectomy, surgical removal of the distal stomach, duodenum, common duct, and head of the pancreas, containing the pancreatic neoplasm, first reported by Whipple et al. in 1935. Unfortunately, resection alone is rarely curative. Out of around 15-20% patients who undergo this procedure, only one in five survive at least 5 years. All other cancer treatment options have been used in pancreatic cancer and include chemotherapy and radiation, however success rate is limited even with the most efficient chemotherapeutic drugs e.g. gemcitabine. One-year survival of pancreatic cancer patients treated with gemcitabine is around 18% (Burris et al. 1997, Rothenberg et al. 1996, Li et al. 2003), gemcitabine enhanced by Erlotinib yields an increase to 23% (Moore et al. 2007), while 5-fluorouracil (5-FU) is as effective as gemcitabine alone (Mulcahy, 2009). More recently combinatorial therapies have been introduced into clinical trials such as the adjuvant Chemo-Radio-Immunotherapy, however their impact on long-term survival is yet to be determined, other therapies have limited effect on the survival due to the intense resistance of pancreatic adenocarcinoma to all extant treatments. Once the cancer has spread, palliative treatment is used to improve the patient's quality of life by controlling the symptoms and complications and limiting the cachectic reaction, unfortunately disease at an advanced stage cannot be significantly slowed down.

Stage	Survival
IA	37% *
IB	21% *
IIA	12% *
IIB	6%
III	2%
IV	1%

*** tumor is resectable**

Table 1 5-year survival rates of patients depend on tumor stage at the time of diagnosis, and tumor resectability. Data based on a publication by the American Cancer Society, October 2009.

3.2.2 Tumor microenvironment

One of the explanations for the high resistance of PDAC to treatment may be the highly complex microenvironment and the poor vasculature of these tumors. The vast majority of cells in pancreatic ductal neoplasm are non-malignant stroma cells (Mahadevan et al. 2007) and the extensive desmoplastic reaction (a process in which fibrous tissue infiltrates and surrounds the neoplasm as observed also in breast, prostate and lung tumors) is one of the characteristic features of PDAC. The desmoplastic stroma in PDAC contains myofibroblasts (stellate cells), immune cells, nerve fibers and marrow-derived stem cells, all forming a unique environment, which not only harbors and nourishes tumor cells, but also likely drives their progression and metastasis, at the same time conveying therapy resistance (Korc 2007). Cancer microenvironment has previously been shown to affect many tumors in a manner that enhances proliferation, invasiveness, tumorigenicity and metastatic potential (Bhowmick et al. 2004, Hu et al. 2008, Nelson et al. 2006, Orimo et al. 2006, Tlsty et al. 2006). However, very little is known about specific interactions between cells comprising it, and even less about the molecular mechanisms underlying them. It has been shown recently that desmoplasia in PDAC may be promoted through the paracrine signaling of tumor cells and associated fibroblasts via the sonic hedgehog (SHH) (Bailey et al. 2008 and 2009, Yauch et al. 2008). SHH is a gene family forming a signaling pathway with a key role in regulating vertebrate organogenesis, and control of the adult stem cell division. Inhibition of SHH in a mouse model has been shown to increase the efficiency of chemotherapeutic delivery in pancreatic cancer (Olive et al. 2009).

Since the vast majority of cells in PDAC stroma are myofibroblasts, they are the natural candidates for tumor-stroma interaction investigation. Fibroblasts are however not the only cell type of interest as the presence of immune cells in the neoplasm combined with the quick progression of PDAC suggests that tumor cells are capable of creating favorable microenvironmental conditions with immunosuppressive features, as was proposed by Dunn et al. (2004) in the hypothesis of “cancer immunoediting”. According to this the immune system not only protects the host against tumor development, but can also promote tumor growth by selecting for tumor escape variants with reduced immunogenicity and through e.g. secretion of immunoregulatory proteins, or direct alteration of immune cell populations including T cells, NKT cells and dendritic cells.

The normal tissue-specific microenvironment arising as a result of intercellular interactions has been shown to be capable of inhibiting malignant cell growth in breast cancer, where fibroblasts (not tumor-associated) inhibit the growth of transformed mammary epithelial cell line *in vitro* (Sadlonova et al. 2005). Changes in normal tissue organization and homeostasis coming from e.g. chronic inflammation (pancreatitis) and wound healing may increase the chance of tumor initiation (Dvorak 1986, Schafer et al. 2008). When altered during a malignant transformation, the microenvironment provides a significant synergistic contribution to tumor development, usually by supporting tumor cell proliferation (Surowiak et al. 2006, Yazhou et al. 2004). In fact, breast cancer myofibroblasts (similar to stellate cells found in the PDAC stroma) correlate with a high proliferative rate of breast tumor cells and a poorer prognosis (Polyak et al. 2009); the latter aspect was recently also investigated by Farmer et al. (2009). Their results show that a stroma-related signature may be used in breast cancer to predict its resistance to neoadjuvant therapies. This suggests that perturbations of the normal mesenchymal-epithelial interactions can lead to unregulated growth (Farrow et al. 2008). It has been shown in prostate cancer (Cunha et al. 2002) that normal epithelial cells are affected by the tumor-stroma interactions differently than the tumor-associated fibroblasts indicating that abnormalities must be present both in epithelial and stromal cell compartments (Cunha et al. 2002, Polyak et al. 2009). These findings are in contrast with available therapies, which specifically target tumor cells, avoiding the associated stroma. Furthermore, some of the existing therapies may actually be suboptimal for stroma-rich cancers as demonstrated in prostate (Cunha et al. 2002), breast (Barcellos-Hoff in 1998), and pancreatic cancer (Ohuchida et al. 2004), where fibroblasts sub-lethally irradiated before injection have been shown to more potently enhance tumor growth than unirradiated cells, owing this to the acquisition of an activated phenotype (Polyak et al. 2009).

Investigation of the molecular crosstalk between malignant cells and tumor stroma in this context may be essential to unfold the pancreatic cancer ability to progress in such a rapid manner, as well as understand its resistance to therapy. A therapy targeting the non-malignant stroma cells may have the additional benefit of pursuing genetically stable cells rather than the inherently unstable tumor cells.

The potential of altering the tumor-stroma interactions recently became even clearer after a publication by Beatty et al. (2011) who have shown the true potential for the development of therapies. They have used an agonist of CD40 receptor in tumor associated macrophages and showed that the activation of this receptor causes a significant activation of immune reactivity leading to tumoricidal effects and depletion of stroma.

Given the importance of crosstalk between cells the work presented here focuses on the identification of the paracrine signaling and the underlying molecular regulation in tumor cells and pancreatic stellate cells, which are the main constituent of the PDAC stroma. While many connections between the various cell types are already known in both inflammation and carcinogenesis of pancreas (Figure 1), interactions leading to the development of the aberrant environment as well as further tumor cell progression towards metastasis remain largely undiscovered. Understanding the gene regulation underlying intercellular interactions resolved via the cytokine patterns taken up and released by each cell type may provide more accessible and more effective drug targets.

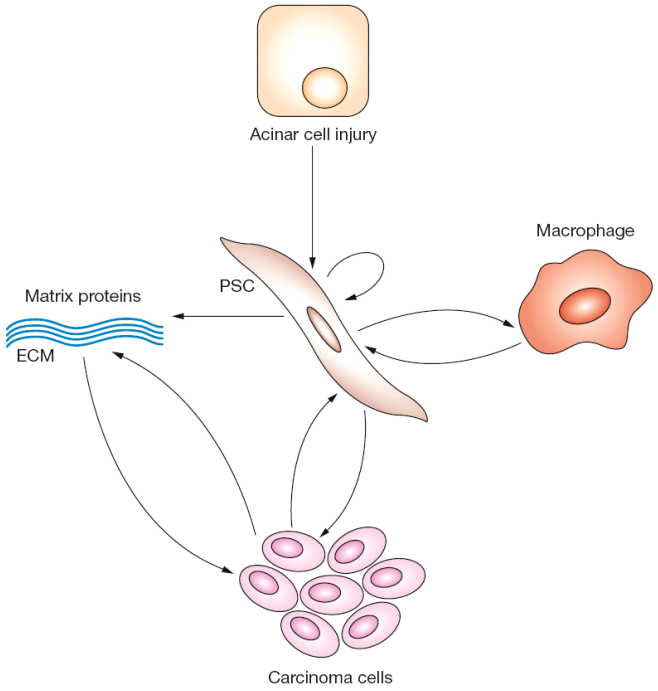


Figure 1 Interactions in inflammation and cancerogenesis of pancreatic cancer, modified after Algül et al. 2007

3.2.3 Pancreatic stellate cells

Pancreatic Stellate Cells (PSC) are the main component of the PDAC microenvironment. PSC are organ specific cells of the connective-tissue family located in the exocrine areas of the pancreas. PSC were first observed by Watari et al. in 1982 and Ikejiri 1990 using fluorescence and electron microscopy, and isolated by Apte et al. and Bachem et al. in 1998.

Stellate cells are of mesenchymal origin and have the capacity to switch between a quiescent and activated phenotype, which makes them in that manner similar to the hepatic stellate cells (HSC), a fact that awakened interest in them soon after their discovery. In a healthy pancreas, stellate cells are quiescent, and can be identified by vitamin A containing lipid droplets in their cytoplasm (Watari et al. 1982). Their activation *in vivo* may occur in response to various circumstances such as tissue damage e.g. through alcohol, prolonged inflammation, or infiltration of the tissue by immune cells. This activation is accompanied by a non-malignant transition from the quiescent into a myofibroblast-like phenotype, in which the vitamin A droplets are degraded and PSC begin to express α-smooth muscle actin (α-SMA), produce extracellular matrix components such as type I collagen, fibronectin and laminin, and actively proliferate (Yen et al. 2002, Masamune et al. 2009).

In vitro stellate cells can be activated by alcohol (ethanol), cytokines (TGFβ, PDGF, TNFα, IL1, IL6) (Bachem et al. 1998, Apte et al. 1999, Mews et al. 2002), growth factors (PDGF) (Luttenberger et al. 2000), and oxidative stress. Upon activation PSC acquire a variety of functions e.g. due to their proinflammatory phenotype they express intercellular adhesion molecule ICAM-1, as well as cytokines and chemokines e.g. IL6, IL8, and monocyte chemoattractant protein MCP-1. They are well recognized as key mediators of pancreatic fibrosis (Wells et al. 1998, Haber et al. 1999), which is a characteristic feature of chronic pancreatitis. In addition, it has been shown that PSC possess (*in vitro*) the ability to migrate in a chemotactic direction (Phillips et al. 2003), remove necrotic debris and aged polymorphonuclear cells by means of phagocytosis, inhibit apoptosis (Hwang et al. 2008, Vonlaufen 2008) and enhance the migration and invasion of pancreatic cancer cells (Bachem et al. 2008).

Stellate cells seem to be significant players in the neoplastic development and progression, and the microenvironment in which they exist has already been shown to influence the growth, differentiation, survival, and motility of cells (Crnogorac-Jurcevic et al. 2001, Koenig et al. 2006, Jaskiewicz et al. 2003, Lohr et al. 2001). However, the molecular mechanisms and the understanding of the underlying cellular regulation driving the progression towards invasion and metastasis remain scarce.

Abundant crosstalk between cells in PDAC, which occurs via autocrine, paracrine and juxtacrine connections between cells in the complex microenvironment influences the progression of tumor cells. Many of the signals driving proliferation and invasion may be attributed to the stromal component of tumors (Bissell et al. 2001), but the exact formation of the paracrine interactions, signal assignment to specific cells in the microenvironment, as well as the underlying gene regulation responsible for the formation of those interactions remain open. Pancreatic stellate cells are the major cellular component of the PDAC stroma and are known to produce a wide range of factors affecting the ECM composition e.g. collagen, fibronectin, proteoglycans, and proteinases (Mollenhauer et al. 1987, Seymour et al. 1994, Bachem et al. 2005). Production of those proteins is stimulated by various signaling pathways such as transforming growth factor β (TGFβ), hepatocyte growth factor (HGF), fibroblast growth factors (FGF), insulin-like growth factor 1 (IGF-1) and epidermal growth factor (EGF) (Ide et al. 2006, Mahadevan et al. 2007). The precise configuration of the ECM is regulated by various mechanisms in tumor and stellate cells. e.g. matrix metalloproteinases involved in degradation and remodeling of the ECM (Jones et al. 1999), and their inactivators – tissue inhibitors of metalloproteinases (TIMPs) (Gress et al. 1998, Bramhall et al. 1996, Neesse et al. 2010). Considering the impact of stroma on tumor progression previous studies by Apte et al. (1999), Luttenberger et al. (2000), Schneider et al. (2001), Mews et al. (2002), and Bachem et al. (2005) have attempted to dissect the interactions between PSC and TC, and produced a common picture of a microenvironment supporting the development of pancreatic cancer. None of the aforementioned studies have investigated in depth the dynamic nature of gene regulation responsible for the formation of the paracrine interactions, and most were performed using tissue samples, which contain a complex mixture of stroma cells, discarding the different impact each cell may have on the formation of intercellular connections. A recent study by Xu et al. (2010) has shown that cancer associated fibroblasts (PSC) may be more important to the progression of cancer than previously believed. Aside of having the ability to (de)construct the ECM they also have the capacity to accompany TC to distant metastatic sites, an important since the presence of "source" stoma likely significantly increases the chances of tumor cells to form neoplasms in remote locations.

PDAC cells are some of the most resilient to therapy we know, and the host response to pancreatic cancer shares many parallels with an infection or a wound, as many tumors arise in areas of infection or chronic inflammation. Existing evidence suggests an increasing role for inflammation as a critical link with tumor progression (Grivennikov et al. 2010). In a recent paper Feng et al. (2010) have modeled early tumor initiation and found that the immune system both

attacks and helps early development of tumors. Apparently tumor cells produce chemoattractant signals similar to those in wounded tissue, leading to infiltration by leukocytes, which are however usually incapable of destroying the TC, leading to a chronic inflammatory state that supports tumor growth (Feng et al. 2010). The term "wounds that never heal" best describes the current view of tumors. In light of our findings presented in Busch et al. 2008 regarding the keratinocyte migration in the process of wound healing the presence of cancer associated fibroblasts (PSC) in the stroma of PDAC suggests a strong potential impact of those cells on tumor cell progression.

3.2.4 Progression towards invasion and metastasis

Unconstrained growth, invasion and metastasis are the three main features of malignant tumors such as PDAC. In normal tissue, environment is the natural barrier for any motile cell, however TC are capable of modifying this environment in such a way that its constraints are loosened. There are a few significant rate-limiting steps in tumor progression including (after Albini, 1998):

- Proliferation – oncogene activation, suppressor inactivation
- Loss of cell adhesion e.g.loss of e-cadherin
- Matrix degradation – MMPs, uPa/plasmin, cathepsin activity
- Extravasation and reaching distant sites – resistance to immune system
- Adhesion to endothelial cells – E-selectin, mucins, V-CAM, I-CAM, integrins (β2, β4)
- Attachment to extracellular matrix – integrins (β1/β3/β5)
- Basement membrane degradation – gelatinases uPA/plasmin
- Migration – chemotactic growth factors, gradients of matrix proteins
- Proliferation at distant sites – organ specific growth factors
- Angiogenesis – VEGF, bFGF, HGF

A general separation of those steps into two phases has been recently proposed by Chaffer et al. (2011), and Shibue et al. (2010) who divide it into two general phases: physical translocation of a TC to a distant tissue, and colonization, and suggest that clonal evolution may be insufficient to explain the unique properties of metastasizing cell populations, offering cancer stem cells as a solution to the transdifferentiation events.

Pancreatic tumors show high motility resulting in a very fast progression towards metastasis, but at the same time slow proliferation. It has been shown that **PSC can stimulate the migration of cancer cells *in vitro*** (Vonlaufen et al. 2008) and that they may **accompany tumor cells during the process of migration to metastatic sites** (as cell clusters)(Xu et al. 2010), which would explain how the tumors are capable of forming metastasis so rapidly. In addition to that, secondary pancreatic tumors at metastatic sites are often genetically different from the primary tumors, which would suggests that the metastasis sets on at a very early stage of the disease, contrary to what we see in many other solid tumors. This however is in contrast with the most recent findings by Yachida et al. 2010, and Campbell et al. 2010, who show that the notion of rapid progression towards metastasis in pancreatic cancer may be coming from poor diagnostics and not actual molecular developments. Using next generation sequencing Campbell

et al. compared mutations found in the original tumors with metastatic sites. They confirmed that all mutations found in the metastases were also present in the original tumor. Yachida et al. 2010 have used estimates of proliferation and mutation rates to mathematically model and calculate the number of cell divisions between discrete events in tumor evolution. They estimate that the average time necessary for the formation of a non-metastatic neoplasm is on average 11.8 years, and it takes another 6.8 years for the emergence of clones leaving a large window of opportunity for early detection (Figure 3, Yachida et al. 2010).

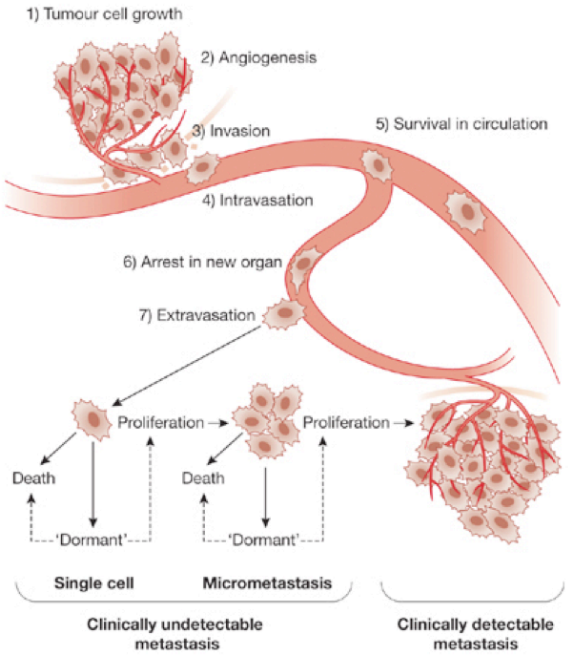


Figure 2 Progression to metastasis (Mammary gland biology and breast cancer. Conference on Common Molecular Mechanisms of Mammary Gland Development and Breast Cancer Progression; Sharon F. McGee (Adapted from Chambers et al. 2002; Chambers & Matrisian, 1997))

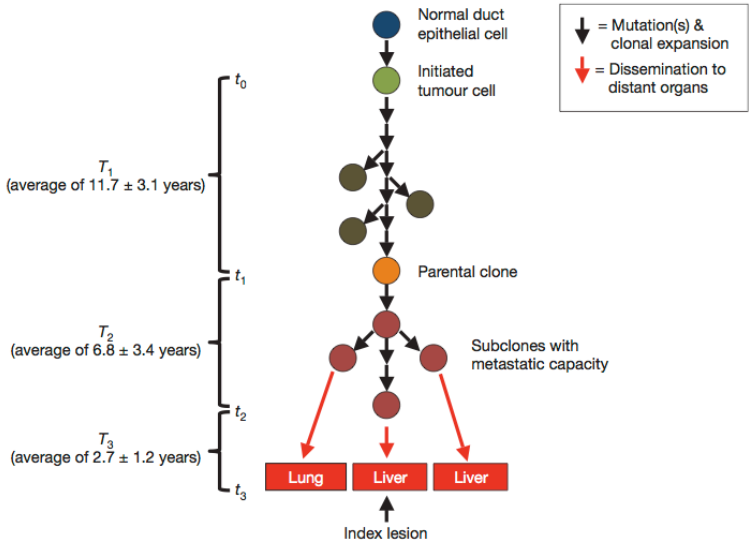


Figure 3 Schema of the genetic evolution of pancreatic cancer by Yachida et al. 2010

The high metastatic capacity of pancreatic tumors is a great challenge in the fight against them. Metastasis is one of the hallmarks of PDAC progression and is revealed with the activation of blood and lymph vessel angiogenesis together with inflammatory and immune-suppressive responses, which further promote migration and invasion (Kopfstein et al. 2006). The high complexity of this process becomes clear when we consider all the steps that are necessary for a cell to metastasize (Figure 2). First the malignant primary cell must compromise its basal structures, by breaking down the extracellular matrix (ECM), it must detach itself and migrate into the surrounding tissue where it can invade small blood or lymph vessels. Subsequently it has to survive the journey through the circulation, invade the endothelium and base membrane, prepare the new site to grow a colony, proliferate, and attract new blood vessels to nourish its growth and survival. Even if all those steps are successful, the cell is still not predestined for metastatic success as it may either proliferate to form a clinically detectable metastasis, remain dormant as a single cell, or run through a limited series of divisions to form a micro-metastasis, which is usually successfully destroyed by the organisms own immune defenses (Nguyen 2004). This complexity is consistent with the notion of the metastatic cell being a rare variant of the tumor cell, which arises at a very late stage of tumor progression, but it is clearly contradictory to the inescapable progression towards metastasis by practically all untreated carcinomas (few exceptions include glioma and basal cell carcinoma).

Genetics' view of metastatic progression defines metastasis as a series of molecular alterations, resulting from mutations. Cell biology however focuses on the regulator/effector

mechanisms that implement the metastatic phenotype especially the Epithelial Mesenchymal Transition (EMT). EMT is considered the single most important pro-metastatic transdifferentiation event after which epithelial cells acquire mesenchymal (embryonic) features. This, like any differentiation process, causes a change in the expression of various genes and proteins scattered across the entire genome. An interdisciplinary approach to the problem offers an alternative solution with a cell attractor model, the underpinnings of which have been provided by Delbrück as early as 1948 (Differentiated states correspond to the stable states in a bi-stable system) and Kauffman in 1969 (High-dimensional attractors of genomic networks represent cell types). Rather than looking at single events such as mutations, this approach looks at a cell in the form of states in a high-dimensional space. Attractors in this space are low energy sinks (valleys in the attractor landscape), which correspond to the specific mutually exclusive phenotypic events such as proliferation, differentiation, migration and apoptosis, and can undergo a switch-like transition from one to another (Figure 4). Each of those programs is represented in each cell type by a very specific gene expression profile, which corresponds to only one of those exclusive cell decisions (Huang and Ingber 2006).

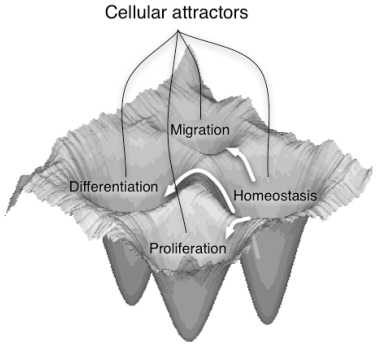


Figure 4 From a lecture by Dr. Hauke Busch (2008)

Keeping in mind the attractor model, we notice that our approach in this project has the capacity to recognize gene expression profiles corresponding to switch-like events associated with those decisions, consequently suggesting that the observation of the dynamic changes may provide new insights into how the cell behavior is established and affected.

3.2.5 Gene regulation

In light of the aforementioned complexity of tumor formation and associated changes in the microenvironment we recognize that the understanding of gene regulation in pancreatic cancer is limited, and while certain genes have been previously implicated in the process of tumor progression (e.g. Sato et al. 2004, Crnogorac-Jurcevic et al. 2001, Campagna et al. 2008, López-Casas et al. 2010), the connections between the regulatory networks and the system-wide changes still elude us. Among the goals of this work is to provide a deeper understanding of gene regulation in both cell types of interest (stellate and tumor cells) in the context of the tumor microenvironment, with a focus on tumor cell response to stellate cells.

It is well established, that gene expression is not constant, but varies with time and location, and is dependent on cell cycle, and genetic code, as well as additional modifications e.g. epigenetic DNA methylation, histone modifications. This multi-level regulation allows the cell to respond flexibly to its environment (Jaenisch and Bird 2003). Control over expression is exerted on every level starting with storage of information, conformation of chromatin, its arrangement in the nucleus, modifications to chromatin structure, and its accessibility, complex processes of replication, transcription, translation and post-translational modifications, and ending with proteins, which may also exert regulatory effects on the process of transcription (transcription factors), performing replication and transcription (various polymerases), cleaving the nucleic acids (nucleases), and histones (chromosome packaging). Each stage can be modulated and the phenotypic outcome is what can be described as cellular decisions e.g. proliferation, differentiation leading to the formation of tissues and organs, and cell death. The separation of time scales between the various processes in cells, as well as the high complexity of each stage of this modulation are the main reasons why gene regulation still eludes our understanding. In addition, accumulating evidence suggests that the structural order in which genes are stored in the eukaryotic genome is not as random as was once believed. Initially it was thought that the only gene clustering in DNA was the result of evolutionary events such as duplications, while everything else was the result of random arrangement of coding and non-coding sequences. Recently multiple genome-wide expression studies in organisms such as Drosophila (e.g. Spellman et al. 2002, Boutanaev et al. 2002, Kalmykova et al. 2005), mouse (e.g. Williams et al. 2002, Nelander et al. 2005, Singer et al. 2005, Sémon et al. 2006, and Purmann et al. 2007), as well as human (Caron et al. 2001, Lercher et al. 2002, Vogel et al. 2005) showed that genes with similar expression levels are non-randomly distributed within genomes and tend to cluster within genomic neighborhoods. Eisen et al. (1998) suggested that genes sharing similar time-resolved

patterns of expression (co-expressed) are likely to have similar function, an assumption which may not be entirely accurate as ectopic expression may be explained by expression leakage, caused by spreading of chromatin modifications or the transcription apparatus into neighboring genes (Yanai, 2006). Nonetheless, following the logic of the guilt-by-association principle we are tempted to extrapolate it stating that co-expressed genes are likely to be co-regulated (Walker et al. 1999, Quackenbush 2003, Joshi et al. 2004, Zhou et al. 2005, Yanai 2006), a finding which has turned out to be very useful in a complexity reduction approach where clustering is used to identify common gene expression profiles.

Transcriptional programs may be represented as gene networks, where products of expressed genes activate or repress secondary downstream targets. Since TFs bind selectively to *cis*-regulatory elements (binding sites) in promoters of their downstream targets, it is reasonable to assume that genes regulated by the same TF should all contain the corresponding binding sites in their regulatory regions and exhibit similar expression profiles as measured in e.g. microarrays. Pilpel et al. (2001) found that genes sharing pairs of binding sites are significantly more likely to be co-expressed than genes with only single binding sites in common. This is in agreement with the hypothesis that a limited number of transcription factors combine in various ways in order to respond to a much larger number of environmental conditions or stress factors. Segal et al. (2003) and Beer and Tavazoie (2004) further developed this idea to find combinations of regulatory mechanisms that best explain expression data.

In addition to the aforementioned intrinsic complexity of gene regulation, we also recognize a further complication coming from the extracellular effects. The model of extracellular-induced (cytokines, growth factors etc.) gene expression has two major components: the initial induction of primary response (immediate early genes) within seconds to minutes, followed by a compulsory delay allowing translation of TFs, which then induce the secondary response genes. The initial response does not require *de novo* protein synthesis and is therefore mediated by pre-existing transcription factors. Secondary response genes require *de novo* protein synthesis. The third class of genes, delayed primary response genes remains outside the scope of this work, as it is indistinguishable from delayed primary and secondary responses without additional experimentation (e.g. involving cycloheximide to block *de novo* protein synthesis) (Tullai et al. 2007). Taking into account the three main signaling patterns established between cells via soluble factors i.e. autocrine, paracrine, and mixed; we investigated the time ordered

sequence of events on the transcript level to differentiate between gene clusters stimulated by either signal (Figure 5).

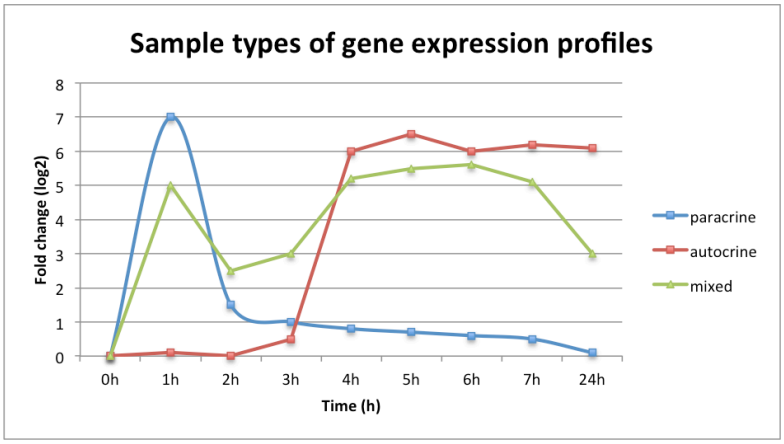


Figure 5 Signaling can be identified by its effects on time-resolved gene expression profiles

The initial stimulation of cells induced through extracellular paracrine signaling results in an upregulation of immediate early genes on a time scale of minutes, and at the latest within 2h. Genes that show an upregulation after 4h or later, are either delayed primary genes, or constitute a secondary wave most likely via the involvement of an autocrine set of signals already secreted by the cell at this stage, additionally the secondary genes can be induced without the extracellular signal involvement through feed-forward type of intrinsic signaling where initially induced TF act on their downstream targets. GRN modeling presented in this work operates on precisely this set of paradigms accepting both: internal and external stimuli to drive the internal network dynamics.

3.2.6 Investigating gene regulation in PDAC

The fast development of high throughput technologies is driving recent genome-wide expression profiling in pancreatic cancer. Twelve molecular pathways involved in pancreatic cancer were described recently using those techniques (Jones et al. 2008). One of the earliest attempts of gene expression evaluation was made using the serial analysis of gene expression (SAGE), which aims to find RNA molecules that are present in one total RNA preparation, but absent in another (Argani et al. 2001). Since then, the bulk of gene expression analysis in PDAC has been performed using hybridization-based methods i.e. microarrays e.g. two-channel cDNA spotted arrays (Han et al. 2002), with support of PCR-based techniques that aim to visualize the transcripts by amplification and quantification (at least relative) between amounts of a particular product in cell extracts from two experimental states e.g. RT-PCR, SIP-PCR, *in situ* qRT-PCR. Most of the datasets generated using those methods are located in the Pancreatic Expression Database (Chelala et al. 2007 and 2009, Cutts et al. 2011), and include, among others, expression profiling of laser microdissected tissues provided by Crnogorac-Jurcevic et al. 2002, Logsdon et al. 2003, Grutzmann et al. 2004, and Buchholz et al. 2005.

Unfortunately, while the benefit of using hybridization-based technologies such as microarrays is clear as we measure the expression levels of thousands of genes simultaneously and identify hundreds of significant genes differentially expressed in pancreatic cancer, building in essence expression profiles of the cells of interest, a few significant problems arise. The validation of the derived targets remains an issue, first and foremost because it is hampered by the very high cost in terms of work hours necessary to process the long lists of genes of interests at the involved laboratories; second, because the understanding of the complexity of gene regulation is incomplete, with new potential sources of variation in PDAC such as recent microRNA discoveries (López-Casas et al. 2010). Nonetheless, as our understanding of the molecular regulation in PDAC improves, it is becoming easier to combine various data sources to arrive at meaningful conclusions. Jones et al. (2008) combined the available genomic (sequencing, amplifications, deletions) and transcriptomic data (SAGE) in 24 PDAC patients to show twelve core pathways altered in pancreatic cancer, out of which six were described as shared among all investigated tumors including apoptosis, regulation of G_1/S phase transition, hedgehog signaling, KRAS, TGF-β, and Wnt/Notch signaling. In the work presented here, while employing only one transcriptomic data source, we use a wide range of supplementary techniques for complexity reduction, combined with Systems Biology-driven modeling, and additional literature-driven sources for the analysis of gene expression in PDAC.

3.2.7 Issues

Differential gene expression profiling of pancreatic cancer, while adding significantly to our understanding of molecular changes in PDAC, in nearly all cases focuses on snapshot-like identification of changes in tumor cells, disregarding not only the dynamic impact of the microenvironment, but even more importantly, the dynamic nature of gene expression itself, which is known to vary depending on the state of the environment. Although one may argue that the snapshot approach is valid when applied to well defined precursor lesions such as pancreatic intraepithelial neoplasia (PanIN), intraductal papillary mucinous neoplasms (IPMNs), and mucinous cystic neoplasms (Hruban et al. 2004) as they are the earliest form of neoplastic transformation, it cannot be forgotten that all of them arise in the rich microenvironment of pancreas and are no doubt heavily affected by it. From this picture we draw a clear conclusion that the expansion of our understanding of the underlying PDAC biology does not translate well to therapeutical advances because of the dynamical environment-dependent genome-wide changes in tumor cells. Therefore it seems hardly feasible that such approaches (i.e. gene expression profiling) will be able to produce sufficiently specific biomarkers from the given cellular context, and a different solution is needed.

The identification of biomarkers for diagnosis, prognosis and treatment is by far the most important task currently underway, it is therefore in our best interest to employ genome-wide technologies to evaluate the dynamical changes in gene expression regulation (e.g. time-series microarray experiments), and combine them with modern analytical methodologies, which allow not only the processing of genes *in silico*, but also modeling of the potential interactions between those genes prior to their experimental evaluation. Emerging techniques such as those provided by Systems Biology offer an integrative approach and show great promise as they allow us to reduce the complexity of the problem and focus on only a limited number of genes of interest. This seems to be one of the greatest challenges in the current data analysis as can be witnessed in the supplemental information of Harsha et al. 2009 compendium on potential biomarkers of pancreatic cancer, which provides a list of targets, the validation of which seems daunting.

3.3 Systems Biology in cancer research

The fast development of Biotechnology and vast amounts of biological data produced by high-throughput techniques over the last two decades have brought the life sciences into the age of information technology, at first tentatively with the introduction of Computational Biology and Bioinformatics, and more recently with the emergence of Systems Biology. Systems Biology is where biomedical techniques meet Bioinformatics, Physics and Mathematics to create a unique perspective on life at the molecular level. Today we are capable of recreating intracellular signaling *in silico*, and are presented with a long-term goal of modeling intercellular interactions in tissues, and at some point whole organisms, as complete systems. Disciplines such as Molecular Biology, Functional Genomics and Biochemistry, while unique on their own, draw significant benefits from the interdisciplinary and integrative systemic approach, which provides not only analysis of experimental results, but more importantly directs the attention towards new experimental designs. Successful integration and combination of all smaller parts creates a fully functional framework for Systems Biology, and gives a strong argument for its scientific validity. Traditional approaches combined with new technologies still seek to understand the function of genes and their products, and how they determine phenotypes, but on a greater scale.

Traditional biomedical research in cancer has focused for a long time on a 'single gene/protein/molecule of interest' approach, such as the search for an oncogene. We have long believed that such a single altered gene might be responsible for a disease, however more than 70 oncogenes have been found, and while they provide us with understanding of key changes in cells, only very few malignancies have ever been shown to be the result of the altered activity of a single oncogene (e.g. Burkitt's lymphoma, CHL Classical Hodgkin Lymphoma). Instead of looking at cancer on a per-gene basis, in Systems Biology we treat it as a system-wide problem, and this system dysfunction is what allows the cell to escape the normal growth control in a multicellular microenvironment. Systems Biology of cancer aims not only at improving our understanding of the cell as a complete unit in which all processes are intertwined, crosstalk is abundant and the behavior is often cumulative, but also to provide us with tools to establish means of restoring such disturbed system back to its original state. The mathematical approaches applied here aim to create a strict representation (preferably quantitative) of biological processes (e.g. pathways), units (e.g. cells) and systems (e.g. tissues, organs, organisms) with the goal of understanding their behavior, predicting, and altering them for the benefit of fighting this disease. The ability of Mathematics to describe a system in engineering terms provides us with a quantitative system-level understanding of living matter at a level previously unachievable by

experimentalists alone. All this is possible due to the significant technological advances in computer science, and biological high-throughput methods e.g. microarrays. The ability to measure the expression of tens of thousands of genes simultaneously and to study adaptive responses of cells to different stimuli on the gene level has turned out to be a major milestone.

The approach undertaken in this project is concept-based in a way that it provides the combination of systems theory with molecular biology in an iterative loop between the experiment and the model. The experiments are driven by the needs of the model and aim to provide a contextual understanding of the interactions between tumor cells and pancreatic stellate cells.

Ludwig von Bertalanffy proposed the underlying theory for Systems Biology, Systems Theory, between the 1940's and 1970's. It is based on principles from Physics, Biology and Engineering. A system, in general, is a set of elements, which interact with each other (and with their environment, if the system is *open*), and form an "entity" (e.g. a pattern), which is different from any of the separate parts. Therefore a system consists of four underlying features:

- Objects, which are its parts or variables;
- Attributes, which are qualities or properties of the system and its objects;
- Relationships between the objects, and the
- Context of a system, which exists in an environment.

Depending on the organization of the input/throughput/output a system can be either *open* or *closed* (a closed system does not interact with its environment). Several system characteristics are: wholeness and interdependence (the whole is more than the sum of all parts), correlations, perceiving causes, chain of influence, hierarchy, suprasystems and subsystems, self-regulation and control, goal-orientation, interchangeability with the environment, inputs/outputs, the need for balance/homeostasis, change and adaptability (morphogenesis) and equifinality (there are various ways to achieve similar goals) (Theory Clusters 2010). In the context of biological systems such as gene regulatory networks we speak of dynamical systems, which evolve over time, and mathematically have two parts: a state vector $\boldsymbol{x} \in \boldsymbol{R^n}$ (a list of numbers which may change as time progresses), which describes exactly the state of some real or hypothetical system, and a function, $f : R^n \rightarrow R^n$, which tells us, given the current state, what the state of the system will be in the next instant of time (R – real numbers, does not apply to complex dynamical systems where complex values take over).

There are two kinds of dynamical systems: discrete time and continuous time. Gene regulatory networks in this project are viewed as continuous-time dynamical systems, which change smoothly over time. Another important distinction in dynamical systems is between chaotic and deterministic dynamics, that is, between systems, which exhibit randomness and unpredictability versus those, which do not. A deterministic model will always produce the same output from a given starting condition. This seemingly unpredictable behavior is called chaos and most natural systems are chaotic (Rasband 1990, Ott 1993, Strogatz 1994). A gene regulatory network is considered a nonlinear, dynamical, and stochastic (random i.e. non-deterministic) system. Nonlinear dynamics deals with the long-term quantitative behavior of dynamical systems (solved numerically or approximated). However, often the focus is not on finding precise solutions to the equations defining the dynamical system (which is often hopeless), but rather to answer questions like "Will the system settle down to a steady state in the long term, and if so, what are the possible attractors?" or "Does the long-term behavior of the system depend on its initial condition?" (Boros, 2009).

In order to understand the behavior of a complex time-continuous dynamical stochastic system, such as a gene regulatory network, derived from time-resolved microarray data, we use Systems Biology to build and validate models of the gene interactions. In this project we apply an ODE-driven neural network (NN) approach to model the genes as nodes in a network. The model successfully provides a platform to investigate *in silico*, without experimental cost, whether the system has a steady state (or multiple), whether the initial conditions define the long-term behavior of the system, and what type of global behavior will result in response to external as well as internal perturbations of such system. This in the context of switch-like mechanisms leading to cell-fate decisions is of profound importance and interest.

3.4 Modeling of gene regulatory networks (GRN)

Gene regulatory networks are the control system of cells; they consist of genes interacting with each other and their products on multiple levels – transcriptomic, proteomic and metabolomic. When speaking of gene regulation, transcription factors are in the spotlight, as they form the underpinnings of this regulation, however the process is multi-level and has biochemical (signal transduction cascades) and mechanistic components (accessibility, location, translocation of products between cell compartments), which should be kept in mind.

In order to improve our comprehension of the regulatory processes and provide means of analyzing highly complex high-throughput data, without employing complex and expensive experimental setups, computer assisted procedures have been developed, which allow us to build entire GRN *in silico*. The goal of modeling GRN is to unravel the dynamical interactions between genes of interest in the context of cellular behavior and transitions under the assumption that causality of transcriptional regulation can be inferred from changes in mRNA expression profiles, where additional regulation levels are neglected or included as hidden factors in diverse gene regulatory models. This is what we call **reverse engineering**, and what provides a framework of describing a cell in terms defined by Physics and Mathematics (Hecker 2009, Fu et al. 2009). The objective of *in silico* modeling is to understand the cell as a whole with not only genome-wide models, but holistic models encompassing all levels of cellular regulation including protein, metabolite, and mechanistic. Unfortunately due to computational and mathematical complexity, and more importantly limited data availability (concurrent measurements on multiple levels of regulation are usually outside of the financial scope of a single lab) most of the modeling approaches, including the one presented in this work, use only gene expression as data sources. Of course, generalized multi-scale models of inter- and intra- cellular interactions exist, but the results they provide are often too complex to follow, understand, and more importantly validate (e.g. complex genome-wide Bayesian networks), some approaches for the sake of feasibility forego entirely, or at least simplify, the low level gene regulation often focusing on the phenotypic effects of experimental perturbations. The model presented in this work attempts to create a good approximation of the regulatory networks of interest, simplifying the problem by building networks of not only direct, but also indirect connections between genes.

Systems Biology of network inference is an iterative process between experiment and model, which aims to provide insights into the processes of interest without the need of building expensive experimental models (Figure 6).

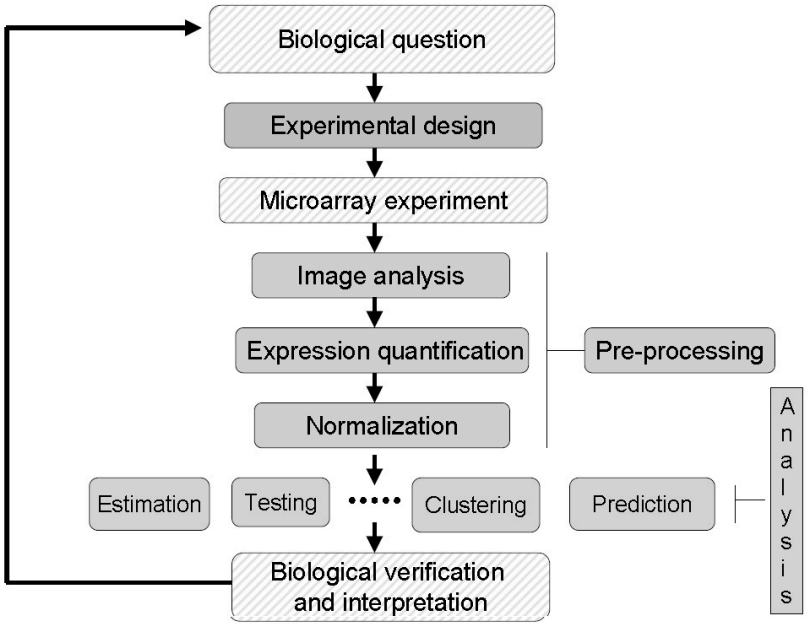


Figure 6 Iterative process of modeling and experimentation in Systems Biology. Modeling precedes biological verification and interpretation in the section of analysis. (from a lecture by David Edwards, courtesy of the Bioconductor project)

Experimental design, selection of data preprocessing and analysis, as well as the modeling methods are tailored to answer a specific biological question. There is no single 'best' modeling technique applicable to everything, but rather different formalisms exist that allow us to address different questions. Since aspects of structure and dynamics can be compared with principles governing man-made systems, mathematical modeling approaches apply many traditional engineering methods to address such questions as the modularity of a regulatory network, the response of the system to perturbations arising from its environment, and the robustness of its behavior in the presence of noise (Ropers et al. 2008).

In this context GRN modeling is, due to the high dimensionality of the genome-wide microarray data, a non-trivial task, further complicated by the dynamic nature of time-course experiments used to discover the dynamical, time-continuous changes among genes.

3.4.1 Artificial neural networks

In this project we apply an artificial neural network (ANN) approach (continuous time recurrent neural network CTRNN). The applicability and ability to reverse engineer GRNs from time-resolved microarray data was previously demonstrated in our work concerning wound healing (Busch et al. 2008) – a project that forms the base reference for the neural network methodology described throughout this thesis. Artificial neural networks are part of machine learning approaches, which allow computers to evolve behaviors based on empirical data. In fact other methods of machine learning are used throughout the presented work including genetic programming, Bayesian learning and clustering, but it is the neural networks, that form the core of reverse engineering.

With ANNs we attempt to mimic the brain, by representing neurons and their connections in mathematical terms. Just like the learning capacity of a brain, ANNs aim to teach computers to solve highly complex questions, without the use of standard computational, algorithmic approaches. Instead of defining each step of data processing, the neural network is set to discover the solution by itself in an iterative manner. Just like in the brain, where each neuron has up to 10.000 connections with other neurons, each node in an ANN is connected to other nodes, each capable of receiving signals, and if strong enough, activating and transmitting it onward to other connected nodes.

3.4.1.1 Continuous Time-Recurrent Neural Network (CTRNN)

Continuous time-recurrent neural networks are a wide class of recurrent neural networks in which inputs and outputs are functions of a continuous time variable and neurons have a temporal response (relating state to inputs) that is described by a differential equation in time (Pineda, 1987, Forcada 2011). Connections between neurons in this network type form a directed cycle creating an internal state of the network (Figure 7)(Websters 2011). Recurrence of the network defines the presence of feedback loops, which allow it to exhibit dynamic temporal behavior. Continuous time is in line with our desire to model a natural biological system, which evolves in an inherently continuous manner.

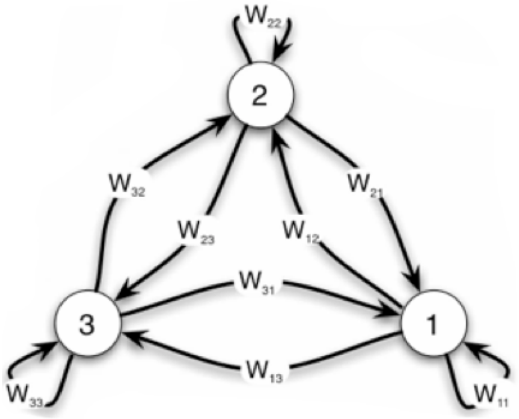


Figure 7 Continuous time recurrent neural network (CTRNN) directed cycle with auto-regulatory feedback loops (based on a presentation by Busch et al. 2008)

In a 1995 paper Randall D. Beer "On the dynamics of small continuous-time recurrent neural networks" suggested, that small dynamic neural networks have the potential of being "powerful building blocks for the modular construction of larger networks with desired dynamics". Based on the principle ideas of Beer's CTRNN approach, the method used in this project was proposed by David Camacho-Trujillo in his PhD dissertation "Reverse engineering of genetic networks with time delayed recurrent neural networks and clustering techniques" (University of Heidelberg, 2008), and used to build a working framework applicable to high throughput data. Performance of the CTRNN approach for reverse engineering GRNs was successfully evaluated by David Camacho-Trujillo, as well as confirmed in our aforementioned paper (Busch et al. 2008). The focus of the presented work is on the application of the method to draw biologically relevant conclusions in the context of gene regulation in pancreatic cancer.

3.4.2 Alternative methods of reverse engineering GRNs from time-resolved data

Depending on the type of data and the type of question multiple different methods of reverse engineering GRN exist (Hache et al. 2009). They may be static or dynamic, with a continuous or discrete time, linear or nonlinear, deterministic or stochastic and include: Boolean networks (Liang et al. 1998), linear models (D'haeseleer et al. 1999), differential equations (Chen et al. 1999), association networks (Basso et al. 2005, Schäfer et al. 2005), static Bayesian networks (Friedman et al. 2000), neural networks (Hache et al. 2007), state space models (Rangel et al. 2005), and dynamic Bayesian networks (Friedman et al. 1998, Yu et al. 2004, Werhli et al. 2006). Due to the nature of time-series expression data, only few are capable of modeling the dynamical behavior of genes (Table 2), including: systems of differential equations, dynamic Bayesian networks (DBN), relevance networks, Gaussian models, and neural networks. Direct comparisons between those methods may be found in an in-depth review offered by Hache et al. (2009), who summarized the results of comparisons in one sentence "Averaged over all results, the neural network approach shows the best performance." Unfortunately, currently no reverse engineering approaches are capable of dealing with large networks of tens to hundreds of genes. As Hache et al. (2009) have shown "Sensitivity, specificity, and precision are always low. Some methods predict only few gene interactions, such as DBN, indicated by a low sensitivity and, in contrast to that, other methods identify many false regulations, such as the correlation measures."

Author	Organism	Data	Time points	Model scheme	Learning algorithm
D'haeseleer et al. (1999)	Rat	RTQ-PCR	28	Linear difference equations	Least squares
Nariai et al. (2004)	Yeast	cDNA ma.	69	Bayesian network	Stepwise hill climbing
Bernard and Hartemink (2005)	Yeast	cDNA ma.	69	Dynamic Bayesian network	Stepwise simulated annealing
Guthke et al. (2005)	Human	cDNA ma	5	Linear differential equation	Stepwise
Kimura et al. (2005)	T.thermophilus	cDNA ma.	14	S-system model	Evolutionary algorithm
Van Someren et al. (2006)	Mouse	cDNA ma.	5	Linear difference equations	LASSO

Table 2 Examples of GRN inference approaches for time-series based on Hecker, et al. 2008.

3.5 Data sources

The source data underlying the expression analysis throughout this project was produced by hybridization of the total cell mRNA (reverse transcribed to cDNA) to an Affymetrix microarray. The choice of microarrays, a hybridization-based method, was determined by our need to visualize transcripts by labeling and quantification on a global, cellular scale, which is especially useful if it is not only the differences between the cells that are interesting but also the specific patterns of regulation. Microarray analysis was previously broadly applied in functional genomics and Systems Biology for gene discovery, disease diagnosis and prognosis, drug discovery (pharmacogenomics), and toxicological research (toxicogenomics). Microarrays enable us not only to look for a transcript among thousands of different genes simultaneously, but also to simultaneously measure the activity and interactions of thousands of genes. Typical scientific questions addressed by microarray experiments include the identification of expression patterns, which can answer questions concerning functional pathways and how cellular components work together to regulate and carry out cellular processes, in addition to the identification of differentially expressed genes, which can help us understand the changes happening in the cell under certain conditions and which genes, groups of genes, classes of genes are up or down regulated under those conditions.

Due to the dynamic nature of gene expression the most interesting data is gathered not by focusing on a single-snapshot differences between samples, but rather on following the dynamic changes in the expression levels of all genes over a long period of time. Such detailed information enables us to visualize not only differences between the control and the treatment at a specific time point, but also to understand how the cell gene levels change over time in response to a specific treatment. Time-course expression analysis is more complex than a snapshot approach and requires a stringent set of rules for every stage including the experimental design. In order to create experiments viable for subsequent analysis one has to clearly define the nature of the experiments, type and number of treatments, and number of time points (short time series of up to 6 time points, or long time series of over 6 time points). In addition, standard rules apply such as a clear definition of treatment, controls, and the overall setup making each experiment clearly distinct from the other.

4. Hypothesis and Objective

Pancreatic ductal adenocarcinoma (PDAC) with its high resistance to known therapies shows behavior distinct from more common types of solid tumors: slow proliferation, quick progression towards metastasis, and desmoplasia. The microenvironment in which pancreatic cancer arises is quite unique with multiple cell types forming the tissue and entering the neoplasm e.g. myofibroblasts (pancreatic stellate cells, PSC), immune cells, and nerve fibers. It is known that this microenvironment undergoes significant changes in abnormal conditions such as inflammation developing into pancreatitis (chronic inflammation) and fibrosis. Previous studies by other groups have revealed PSC ability to harbor and nourish cancer cells, as well as affect their progression, and convey radio- and chemo-resistance. Unfortunately those studies focus on search for the most abundant signals and snapshot-like static analyses of cultured cells and tissue samples ignored the most crucial factors characterizing response to perturbations in complex systems that is: the temporal dynamics in gene expression itself and that of intercellular interactions, and the redundancy of molecular signals underlying mutual relationships, meaning that an altered expression of a soluble factor in one cell type remains without consequences if matched by preexisting signal from the other cell type (or another factor with similar downstream signaling). It therefore remains unresolved whether the rapid progression of pancreatic cancer may be directly attributed to the PSC, which type of cell initiates the paracrine signaling, which soluble factors and which receptors are specifically involved in each cell, and what the dynamic nature is of this intercellular signaling and its underlying gene regulation.

We hypothesize that at an early stage, the quiescent→activated PSC transition yields a steady state PSC gene regulatory network (GRN), but the subsequent succession of impulse responses along TC→PSC→TC interaction axis drives both cells' GRNs it into unstable states maintained for the duration of TC-PSC contact. The double-paracrine connections between cells are established transiently as a result of the GRN formation and can be identified via those underlying networks.

Through the application of high-throughput complexity reduction approach and *in silico* modeling we aim to reconstruct the GRNs underlying the formation of the cell-cell interactions, and to identify key soluble factors shaping the double-paracrine communication. We aim to use the models to gain a mechanistic and functional insight into how the signals are integrated and how they affect GRN maintenance. We hope to capture cell-fate decisions and identify key dynamic changes with the ultimate goal of finding genetic markers to aid development of novel therapeutic options for this deadly malignancy.

5. Materials and Methods

Experimental work throughout this project was performed by collaboration partners at the European Pancreas Center (Dr. Nathalia Giese), the Department of Immunology at the University of Heidelberg (PD Dr. Thomas Giese), and at the Division of Functional Genome Analysis of the German Cancer Research Center DKFZ (Dr. Andrea Bauer).

5.1 Workflow

To uncover the intercellular connections and intracellular gene interactions in tumor and pancreatic stellate cells we designed three experiments, in which we individually stimulated cells with conditioned supernatant from the respective other cell type. This approach allowed us to record time-resolved, genome-wide microarray expression data (up to 24h after stimulation), and thus unravel the time-ordered sequence of events underlying the regulation and the paracrine signaling between the tumor and stellate cells. We recently showed that the long-term cellular behavior is captured and reflected in the cells' gene expression kinetics (Busch et al. 2008). Therefore, we set up a high throughput data analysis and modeling pipeline to describe the dynamics of the pancreatic tumor microenvironment *in silico* (Figure 8). Processing was optimized for data quality and reproducibility *in silico*.

The bioinformatics of microarrays begins with data pre-processing also referred to as **low-level analysis** (image processing, quantification, normalization), which prepares the data for the **high-level analysis** that answers the real biological questions. Once the final format for the data is achieved, data exploration may commence, and this part of the process is referred to as **exploratory analysis**. Typical exploratory tasks include classification (or class prediction), clustering (or automatic classification), correlation, association, and pathway analysis.

Low-level microarray data processing was performed using the Bioconductor framework of the R programming environment, a set of open source software packages for mathematical and statistical analysis of biological data. For high-level analysis as well as exploratory analysis we used a wide range of tools, most of which are open source in the form of websites, databases or standalone software packages. The only proprietary tools used in this work include the Transfac Professional database (a free version exists but is of limited use), and Ariadne Pathway Studio with its ResNet Mammalian manually curated database.

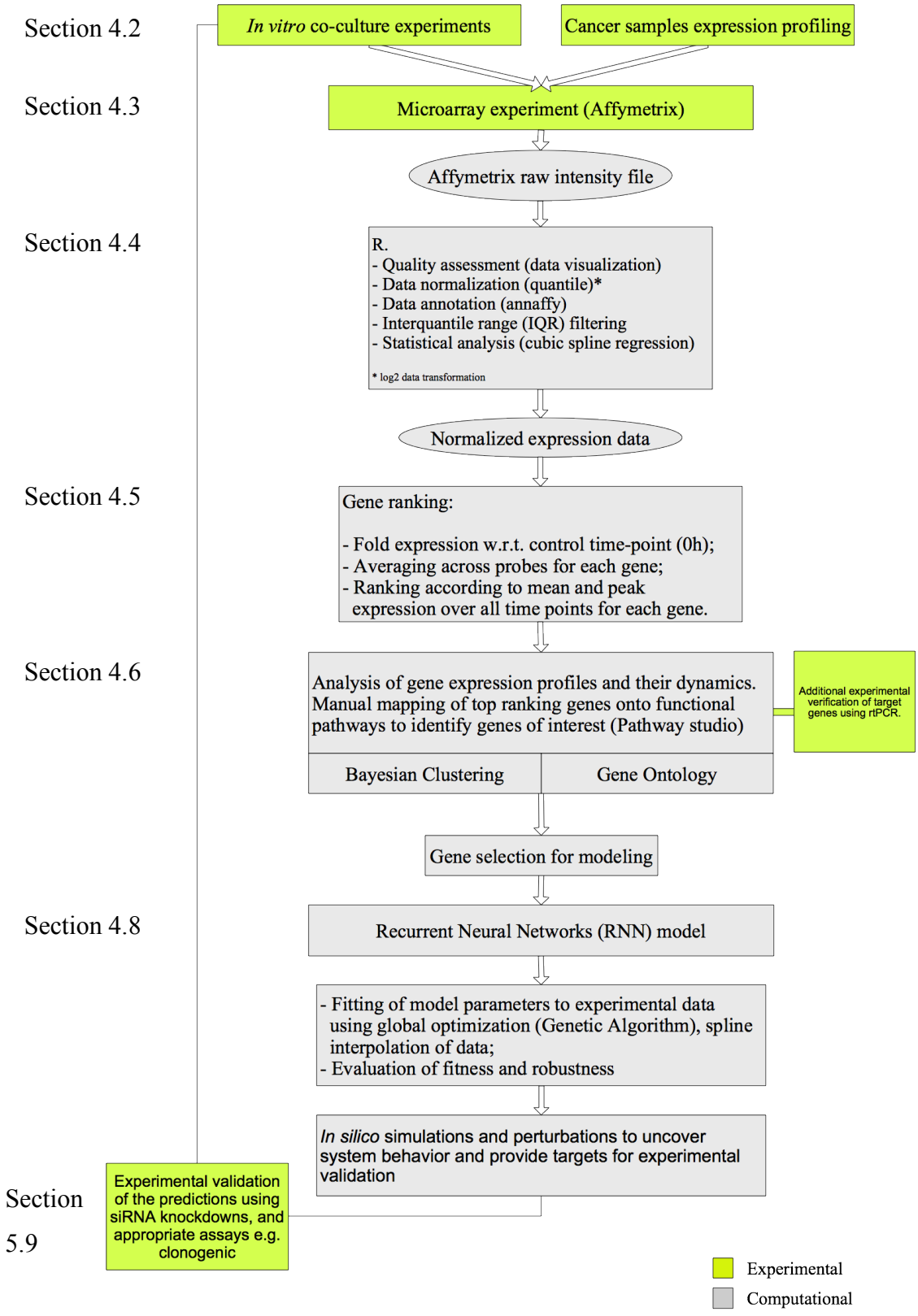


Figure 8 Workflow intertwining experiment with computational data analysis

5.2 Experiments for genome-wide expression profiling

To unravel the time-ordered sequence of events underlying the gene regulation and the paracrine signaling between tumor and stellate cells, three time-course experiments were performed (Figure 9). In all cases we individually stimulated cells with conditioned medium from the respective other cell type, exchanged medium, and collected samples at 8 intervals, once every hour between 0h and 7h with an additional measurement at 24 hours (by definition, it is a long time series). Untreated controls were collected at 24h to account for basal gene expression changes unrelated to experimental conditions. Experimental validations of the microarray results were performed using qRT-PCR to verify transcript levels and time-resolved behavior of genes of interest.

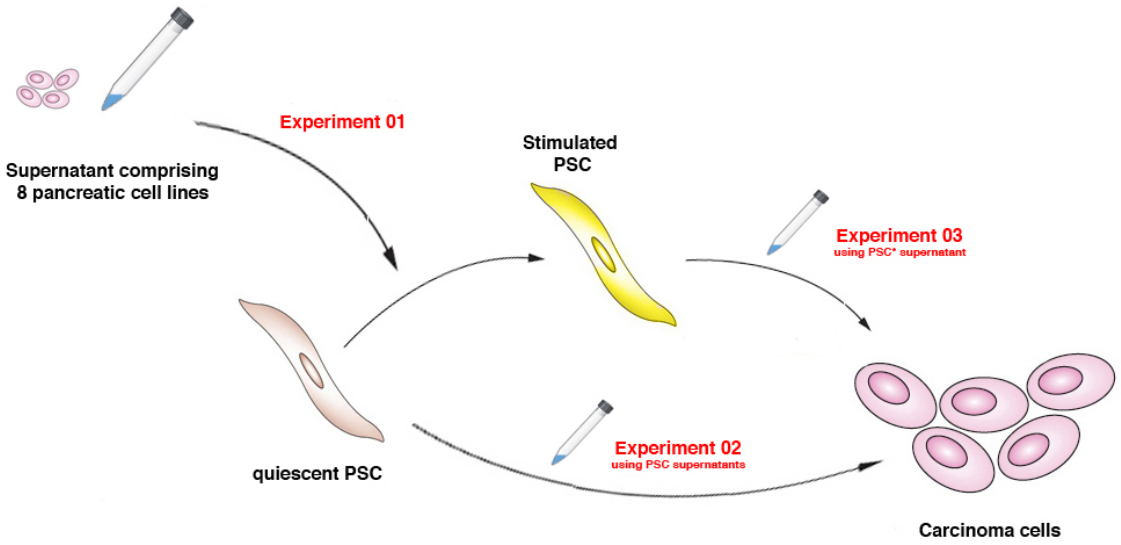


Figure 9 Experimental setup to unravel cell-cell communication in PDAC microenvironment

5.2.1 Cell lines

Eight cell lines were used, in combination, for PSC preconditioning. MiaPaCa2 cell line was used for the remaining two TC experiments, and 2 cell lines: Panc1 and MiaPaCa2 were used for experimental validation procedures. Additional validation of experimental results was expanded for all remaining TC lines using the Wagner CellLines dataset in Oncomine (www.oncomine.org).

MiaPaCa2	65-years-old Caucasian male (Yunis et al. 1977)
Panc1	56-years-old Caucasian male, epithelioid carcinoma (Lieber et al. 1975)
BxPC3	61-years-old female with a primary pancreatic adenocarcinoma
T3M4	Pancreatic adenocarcinoma
Colo357	Human Lymph node metastasis (Morgan et al. 1980)
SU8686	Pancreatic adenocarcinoma
Capan1	40-years-old Caucasian male, pancreatic adenocarcinoma from metastatic site in the liver
Aspc1	62-years-old Caucasian female, pancreatic adenocarcinoma derived from the patients ascites (fluid from the peritoneal cavity)

All cell lines were acquired from ATCC Bioresource Center.

5.2.2 Stellate cells (Experiment 1)

Supernatant from 8 human adenocarcinoma cell lines (Σ) was gathered and used for stimulation of primary human PSC derived from patient tissue (Figure 10). Tissue origin implies pre-exposure to TC, therefore to ensure GRN stabilization PSC were passaged and allowed to grow in a cell culture prior to the experiments.

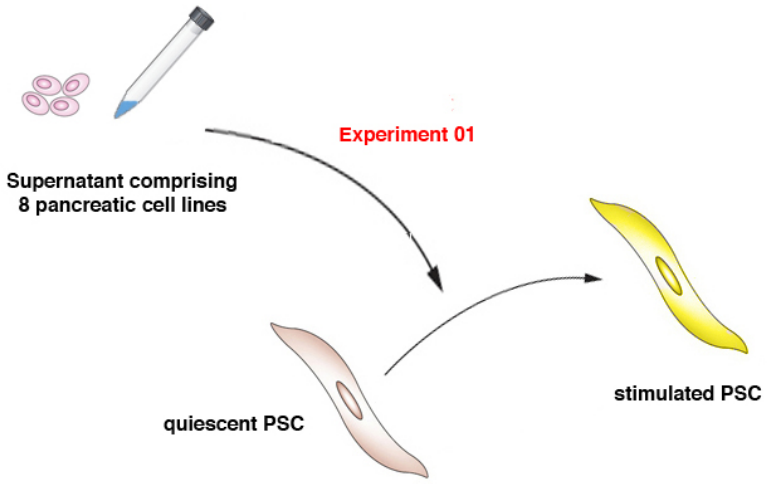


Figure 10 Stimulation of quiescent stellate cells with conditioned medium (8 human pancreatic adenocarcinoma cell lines) (Experiment 1)

5.2.3 Tumor cells (experiments #2 and #3)

Both TC experiments were performed on the MiaPaCa2 tumor cell line, and the cells were treated with supernatant from either quiescent (#2), or stimulated (#3) stellate cells (Figure 11). Serum free supernatant was used to treat TC.

Experiment #2:	Stimulation with the supernatant of quiescent stellate cells (PSC)
Experiment #3:	Exposure to the supernatant of stimulated stellate cells (PSC*)

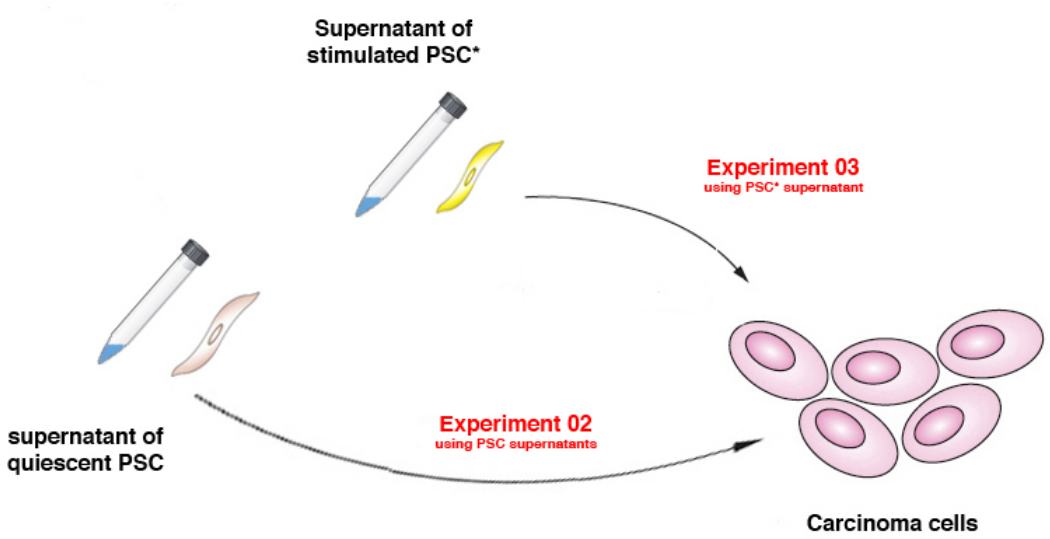


Figure 11 Stimulation of tumor cells with quiescent (Experiment 2), and stimulated (Experiment 3) stellate cells (PSC)

Each experiment was performed in triplicate. Total RNA was gathered for each time point using Qiagen kits, an aliquote for mRNA re-isolate with MagnaPur, samples were collected for later study, as well as reverse transcribed to cDNA for microarray hybridization. Confirmation experiments were performed with qRT-PCR (see section 6.5).

5.3 Microarrays

mRNA samples from each time point of each experiment were reverse-transcribed (cDNA) and hybridized to the Affymetrix Human Genome U133 Plus 2.0 arrays (HGU133Plus2 GeneChip) for further analysis (Figure 12). Altogether 27 Affymetrix chips were used for data analysis of the three main experiments. Quantification of the microarrays with the corresponding scanning software was performed at the Division of Functional Genome Analysis (DKFZ), and the output in the form of raw data files was used for the computational data preprocessing, analysis, modeling and interpretation of experiments in the presented work.

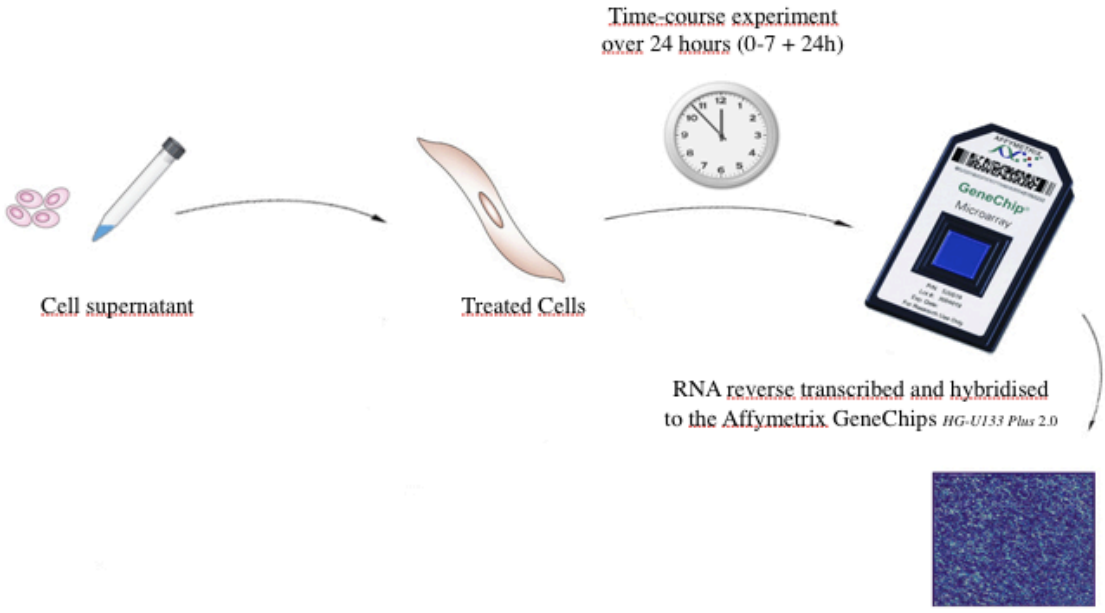


Figure 12 Experimental procedure from cell treatment to image acquisition

5.4 Low-level data analysis

Image acquisition and quantification is automated by the software provided with the Affymetrix scanners, and was executed by Dr. Andrea Bauer. Raw data along with microarray images form the starting point of the presented work.

5.4.1 Normalization of microarrays

Normalization of microarrays was performed using the Robust Multi-chip Average normalization (RMA) by Irizarry et al. as included in the function *JustRMA* of the Bioconductor '***affy***' library (Irizarry et al. 2003, Bolstad et al. 2003), which includes a 3-step procedure of: pre-normalization (PM perfect match only), background correction; normalization (quantile); post-normalization (median polish).

Quantile normalization was used to unify the distribution of the data on every chip (making distributions identical in statistical properties) through first ordering the genes by placing the strongest gene on top across all conditions (Figure 13), calculating mean for the row, and reshuffling them back into original positions. In the end the entire matrix contains not the original data but the mean values. All conditions have the same values but usually in different positions, data distribution is unified, and biological information is preserved.

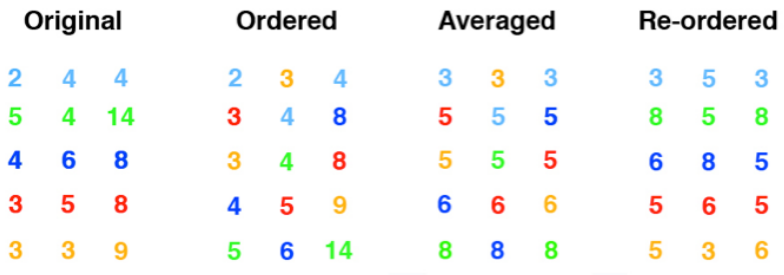


Figure 13 Quantile normalization steps from original data stored on an array, up to the re-ordered averaged values.

Data quality was evaluated with quality control (QC) plots using the ***overview()*** function of ***made4*** R library for multivariate analysis of microarrays (Culhane et al. 2005) (Figure 14).

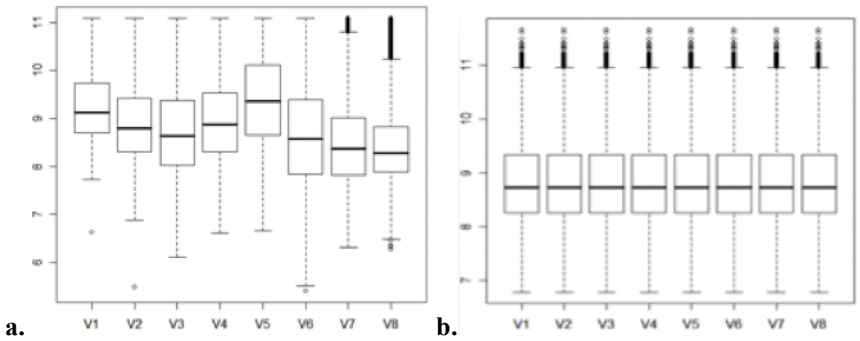


Figure 14 Sample data distribution prior (left), and after (right) quantile normalization using RMA

5.4.2 Post-normalization data filtering

We applied an adaptive spatial filtering approach based on the IQR (inter-quartile range) to remove genes that, due to their low overall intensity or variability, are unlikely to carry information about the phenotypes under investigation. IQR provides a measure of the statistical spread of the middle 50% of the intensity scores for each gene (between 75th and 25th percentile). Genes characterized by a relative large signal distribution were preserved and genes with either consistently low intensity values or low variance across all time points were removed (Figure 15) (von Heydebreck et al. 2004, Bossotti et al. 2007, Spugnini et al. 2006, Lo Iacono et al. 2006, and Olivero et al. 2006.).

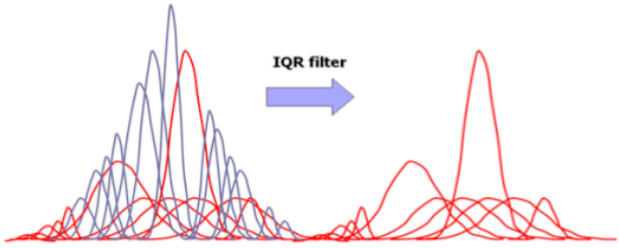


Figure 15 IQR filter removes genes that show little changes within the experimental points. The distributions of the various probe sets shown in red are retained by the filter, blue are rejected. (Figure by Calogero et al. 2010 OneChannelGUI package vignette, Bioconductor project)

Filter was implemented in R using ***genefilter*** package (Bioconductor) with code by Morten Mattingsdal of the Bioinformatic Core Facility at the Rikshospitalet-Radiumhospitalet HF, Oslo, Norway.

```
iqrFilter<-function(x, threshold){
            require(affy)
            require(genefilter)
            iqr <- function(x){IQR(x)>threshold}
            ff <- filterfun(iqr)
            which <- genefilter(x, ff)
            par(mfrow=c(1,2))
            text<-paste("Unfiltered probes= ", dim(exprs(x))[1], sep="")
            hist(exprs(x), main=text, breaks=100)
            text<-paste("Filtered probes= ", sum(which), sep="")
            hist(exprs(x[which,]), main=text, breaks=100)
            return(x[which,])
}
```

Annotation of the microarray data was performed using the ***annotate*** R package (Gentleman 2010) and the corresponding ***hgu133plus2.db*** annotation file for the Affymetrix HGU133Plus2 microarray.

5.5 High-level data analysis

High-level analysis of microarrays allows us to identify differentially expressed genes as well as perform a biological interpretation of the data using tools for gene clustering, correlation analysis, Gene Ontology (GO) and pathway analysis. We separate the latter into a section on exploratory data analysis (5.6).

5.5.1 Gene ranking

Normalized $\log_2$ transformed microarray data were used to calculate the gene **fold change** value (i.e. how many times the intensity of gene expression changed compared to the control). FC value has no measure of statistical significance but is biologically relevant and specifically usable for our data, which other than the time series has no other replication. Genes were ranked according to the **Euclidian distance** between the **maximum expression within a probe set** and the **mean expression of a probe set over all time points** (as we showed in Busch et al. 2008), both scaled (normalized) to the maximum value measured in each dataset.

The ranking procedure was implemented in an IDL script *rank.pro* as originally proposed by Hauke Busch, and further developed for the needs of this project.

5.5.2 Statistical analysis of the time-series

To identify a potential overlap between FC results and a recognized statistical approach we apply a cubic spline regression method appropriate for a long time series of over 6 time points (Leek et al. 2006) as implemented in the *Edge* statistical suite (Storey et al. 2005). We have shown previously (MaSigPro in Busch et al. 2008), that the statistical analysis can only approximate the ranking performed with FC values, which produce biologically meaningful data. Any statistical method and defined parameters such as p-values, or false discovery rates have a direct impact on the gene expression correlation structure, and therefore on which genes are identified as the top ranking.

An additional evaluation of the available statistical methods was performed using:

- Significance analysis of microarrays (SAM) (Tusher et al. 2001)
- *timecourse* – an R package with MB-statistics and/or T^2 statistic (Tai and Speed (2006) and Tai (2005)) derived using multivariate empirical Bayes approaches;
- BATS (Bayesian analysis of time series) by Angelini et al. 2008

Edge was selected as the optimal solution providing the most reproducible results when compared to FC ranking. The advantage of *Edge* is the ability to identify genes that are differentially expressed between two or more different biological conditions (e.g., healthy versus diseased, treated versus untreated) in a time course experiment (Storey et al. 2005 and 2007; Storey, Dai and Leek 2007). *Edge* also allows two types of time course significance analysis: the first one tests for genes whose expression changes over time, and the second identifies genes, which show different expression over time between two or more biological conditions.

5.6 Exploratory data analysis

Our ability to gather genome-wide expression data outstrips our ability to process it in a similar, high-throughput, manner. Large databases containing millions of gene measurements exist, but the applications of methods to draw biologically meaningful conclusions remain scarce. In this project we apply multiple steps of analysis to draw biologically relevant conclusions and to prepare data for the modeling stage. These steps include the clustering of gene expression profiles (using Bayesian clustering), Gene ontology analysis (using *David*, *WebGestalt*, and *eGOn*), Transcription Factor Binding Site Analysis *TFBSA* (using *TransFac* and *Paint*), and pathway analysis (using *Ariadne Pathway Studio, KEGG, WikiPathways, and Pathway Commons*).

5.6.1 Clustering of gene expression profiles

To identify patterns of regulation and potential directions for cell fate decisions we apply clustering as a complexity reduction approach to find genes sharing similar expression profiles (Figure 16). To eliminate bias we apply an unsupervised Bayesian clustering (BC), which clusters the data from scratch, without *a priori* knowledge of existing gene expression profiles. BC provides a distinct improvement over the alternative solutions with its ability to reliably cluster the data without a significant number of outliers, and more importantly to perform on-the-fly modifications to the cluster numbers without recalculation. BC is based on a representation of clusters through a stochastic population model. Clusters are formed based on a 'typical' expression profile and each gene assigned to the cluster follows the same typical kinetic and differs from this profile only due to individual variability (random effects) (Magni et al. 2008). The assignment occurs by a random walk process (Ferrazzi et al., 2005). BC identifies the optimal number of clusters automatically, and the quality is easily assessed by visually inspecting the result.

For BC clustering we used the *timeclust* package (Magni et al. 2008, Ferrazzi et al., 2005). The default gene clustering in this project encompasses top 500 ranked genes, both up- and down-regulated. BC was self-sufficient for sets of up to 250 genes, and supported by:

–Heuristics (for sets of 250-500 genes), and

–Self-organizing maps (> 500 genes).

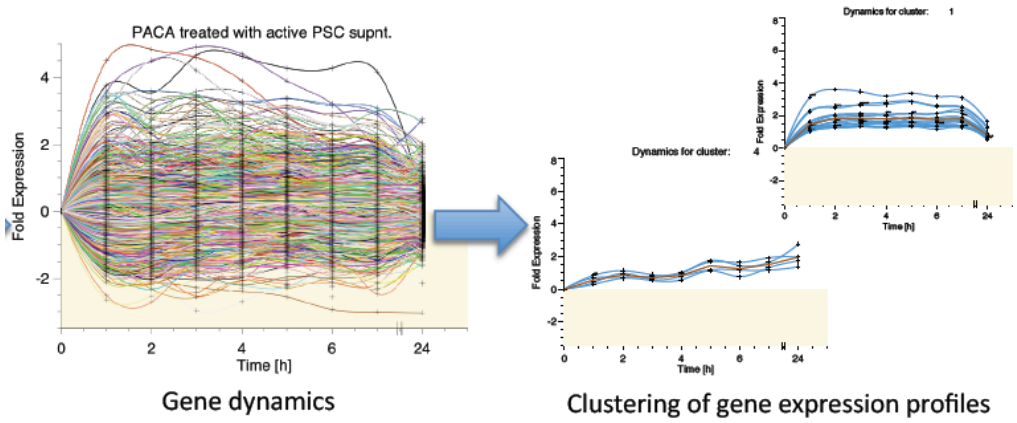


Figure 16 Microarray data set is clustered using a Bayesian algorithm

5.6.2 Manual clustering approach

Initial clustering of PSC data, which was performed manually by Dr. Axel Szabowski, was based on the visual inspection, and identification of gene families sharing similar expression profiles and function within each cluster. First a preselection of distinct profiles was performed to create general clusters containing TFs and known regulators. In the next step clusters were augmented with all available family members of those representative genes, along with genes of interest sharing the same expression profile. Since many of the selected genes have shown low expression levels, a summarized profile for each module was used instead of the standard computation of a mean value (avoids signal attenuation in each cluster). This knowledge-driven clustering, while in many respects successful and biologically strong, has proven to be suboptimal as the time necessary to perform it outweighed the quality of the achieved results when compared to Bayesian clustering, therefore an automated Bayesian clustering algorithm was applied to all datasets.

5.6.3 Gene ontology analysis

Gene Ontology (GO) analysis was performed using:

- ***David*** (Nat. Inst. of Allergy and Infectious Diseases (NIAID), NIH) (Huang et al. 2009),
- ***WebGestalt*** (Zhang et al. 2005), 2nd generation (Gene Set Analysis Toolkit)(Duncan 2010)
- ***eGOn***, part of ***GeneTools*** (Beisvag et al. 2006)

GO is a functional and statistical analysis, which based on the list of differentially expressed genes returns a list of significantly (FDR adjusted) enriched GO terms. Those terms are functionally grouped, and fall into three major categories – biological process, molecular function, cellular component, and more specific classes such as intracellular metabolism, transcriptional regulation, membrane, organelle, secreted factors etc. The results of this enrichment analysis give us an insight into how the cell responds to the given conditions in each of the experiments, but also provide detailed functional overview of each of the clusters coming from the BC. While systematically enhancing the biological interpretation of large lists of genes, GO has the disadvantage of disregarding correlation between the genes coming from their expression levels, and their behavior over time. Also both inhibitors and activators may be located within a single term, which defines a process that can be regulated both positively and negatively, while at the same time not defining the direction of this interaction, making it difficult to draw immediate conclusions.

David – a comprehensive collection of GO analysis tools, providing both functional annotation and functional classification of genes binding multiple data sources to draw from e.g. pathway (Biocarta, Kegg, Reactome), disease (e.g. OMIM_Disease), literature (PubMed). However, due to very limited visualization capabilities, including them all usually obscures the analysis.

WebGestalt – combines GO enrichment analysis with visualization tools and methods for pathway analysis (KEGG, WikiPathways, Pathway Commons). Additionally contains GO slim, which is a cut-down version of GO that provides a broad overview of the ontology content without the fine-grained details, which is particularly useful for summarizing results.

eGOn – allows for comparisons between multiple gene lists, which are analyzed simultaneously to compare the distribution of the annotated genes over the GO hierarchy. It applies a generalized linear model and generalized estimation equations (Leisering et al. 2000) in the form of a statistical test.

5.6.4 Correlation analysis

To identify similarities and differences between datasets we apply correlation analysis. Distance-measuring techniques apply various distance metrics to calculate the similarity between points, the most commonly used are: Euclidian distance, standard correlation, Pearson, and Spearman (ranked) correlation, and Manhattan distance.

Here we use the Pearson correlation, a nonparametric measure, which does not make any assumptions for the probability distribution of the data.

To compute the Pearson correlation coefficient of two vectors or of a correlation matrix of an $m \times n$ array, we apply the ***correlate*** function in the IDL programing environment. The results are plotted in a graphical form for direct comparisons of each gene in the given datasets.

5.6.5 Pathway analysis

Pathway analysis was performed using WebGestalt and Ariadne Pathway studio. *WebGestalt* provides the basic facility of Gene Set analysis using KEGG (Kyoto Encyclopedia of Genes and Genomes), Pathway Commons (pathwaycommons.org) combining e.g. IntAct, HumanCyc, BioGrid, Reactome, and Wikipathways (wikipathways.org). For the statistical analysis in WebGestalt we used the hsapiens_affy_hg_u133_plus_2 reference set, multiple test adjustment was performed using the Benjamini & Hochberg method with p-value cut-off of 0.05 (<5%).

Ariadne Pathway Studio was used with the default parameters and each dataset was analyzed to identify the known biological relationships, associations, interactions and facts that were extracted from the biomedical literature and are stored in the manually curated ResNet Mammalian Database (*www.ariadnegenomics.com*). With Pathway studio we can interpret gene expression data, build, expand and analyze pathways as well as find relationships among genes, proteins, cell processes and diseases. It is literature based and the identified interactions are not tissue specific (although it is possible to focus on a specific organ), therefore for each analysis step with this tool we add overlap checks against the microarray experiments in this project to ensure cell-type specificity.

Pathway analysis combined with the transcription factor binding site analysis TFBSA (*transfac* database) forms the underpinnings of our method of identifying specific regulatory patterns for a cell type of interest using the microarray data as a validation.

The most common applications for the pathway analysis included the search for the:

- Shortest path between genes of interest;
- Shortest path with common regulators;
- Shortest path with transcription factors;
- Location of proteins within cellular compartments (plasma membrane, mitochondria, ER, Golgi and nucleus) and extracellular matrix.

5.6.6 Transcription factor binding site analysis (TFBSA)

Transcription factor binding site analysis was performed using the ***Transfac Professional*** database as well as ***Paint*** (Promoter Analysis and Interaction Network Toolset version 4.0-pre) from Daniel Baugh Institute for Functional Genomics and Computational Biology (Vadigepalli et al. 2003).

Efficient identification of TFBS is a crucial step in the study of gene regulation as it provides us with a comprehensive resolution of transcriptional regulation even from filtered microarray data. This approach tackles the problem by modeling TF-binding sites using position weight matrices and searching for these sites in the DNA sequences of genes of interest (*transfac* database).

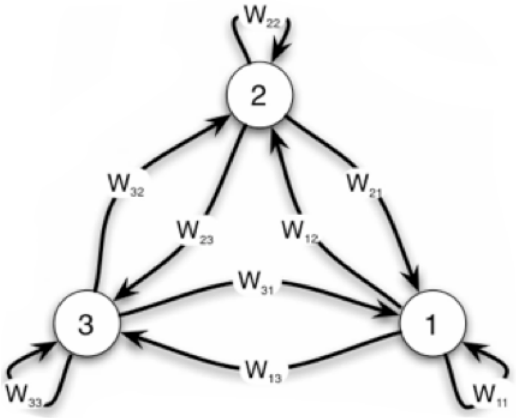
Following parameters were used for all TFBSA analyses:

Upstreamer	TF Retriever
Desired upstream length: 5000	Match (TRANSFAC Pro v.2010.01)
Gene Identifier type: Entrez ID Human – Affymetrix HG-U133 Plus 2	MATCH filter: Minimize False Positives MATCH Vertebrates non redundant
Plain file format	Core similarity threshold: 1.0

Table 3 TFBSA analyses parameter choices

5.7 Modeling of gene regulatory networks with CTRNNs

Modeling of the GRN from time-course data in this project was performed using a modified version of the continuous time-recurrent neural network (CTRNN) by Beer et al. (1995) as described in Busch et al. (2008). A network is defined here as a set of N-coupled ordinary differential equations, each one describing the kinetics and interactions of an individual gene.



$$\frac{\Delta g_i(t)}{\Delta t} = f_i(g_1, \ldots, g_n) = \frac{1}{\tau_i}\left[-g_i + \sum_{j=1}^{n} . W_{ij} \, . \sigma\left(g_j(t - \Delta t_j) - \theta_j\right) + I_j^0 \exp(-\lambda t)\right]$$

$$\sigma(x) = \frac{1}{1 + \exp\left(-\alpha x\right)}$$

Figure 17 Ordinary differential equation describing the kinetics and interactions of a gene in a CTRNN network (in red are parameters, which are being evolved in the network) and the corresponding sigmoidal activation function (Busch et al.2008)

This CTRNN approach proposed by David Camacho-Trujillo in his PhD dissertation "Reverse engineering of genetic networks with time delayed recurrent neural networks and clustering techniques" (University of Heidelberg 2008) based on concepts developed by Prof. Randal D. Beer (Dept. of Computer Engineering and Science, Case Western Reserve University, Cleveland, OH 44106).

We have previously shown that it is applicable to time-resolved microarray data (Busch et al. 2008). The initial implementation by David Camacho was modified into a working framework applicable to high throughput microarray data by Dr. Hauke Busch (University of Freiburg), and subsequently modified into its current iteration by me. The direct modifications of the approach included the construction of a workflow as shown in Figure 8, redefinition of model parameter boundaries, addition of custom input functions, ability to evolve additional parameters in the

input functions (i.e. input signal decay λ), ability to integrate external expression profiles as inputs, and modifications to the evolutionary algorithms and fitness evaluation procedures of the CTRNN to achieve optimal fitting results.

The performance of the CTRNN approach for reverse engineering GRNs was previously successfully evaluated by Dr. Camacho-Trujillo, as well as confirmed in the aforementioned publication, a theoretical evaluation was not the objective of the work presented here. The focus was the application and achieving biologically relevant results in the context of pancreatic cancer. Only an overview of the methodology will be provided in the subsequent sections wherever applicable, detailed background information may be located in Busch et al. 2008.

5.7.1 Gene selection strategy

The initial model construction was performed using a *'metagene'* approach, where a theoretical gene kinetic, averaged from all genes in a cluster, was taken. This stage provided us with a draft network for a quick evaluation of the dynamic behavior of the system. Subsequent modeling was driven by gene selection in search of the most stable and fit solutions. We have followed two approaches to gene selection. The first one attempted to select genes in an unbiased, semi-automatic way by taking e.g. the single most upregulated probe set in each cluster, mean of the top 10 most upregulated genes in each cluster, summarized expression of genes from a specific family of genes, a single representative (mean) of each cluster (same as the metagene approach), mean of all transcription factors in a cluster.

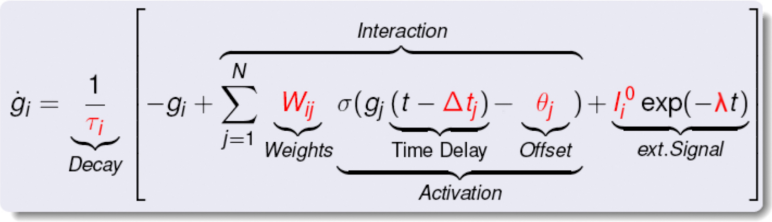
The second approach was knowledge-driven and followed the researchers goals and interests. This selection was supported by detailed analysis of the top ranking genes in all experiments (w.r.t. the existing literature, GO and pathway analysis), and detailed analysis of intra- and intercellular signaling within the top 500 differentially expressed genes. While introducing a bias into the selection, this method provided biologically sound gene selection, and has been used in the construction of the final models.

Final gene selection for the model was performed by collaborators:

- Dr. Nathalia Giese (European Pancreas Center) – TC model;
- Dr. Axel Szabowski (DKFZ) – PSC model.

5.7.2 Construction of a continuous recurrent neural network

Armed with gene selection we enter the modeling stage, where each gene cluster, which corresponds to a node in the neural network, is represented with an ODE (Figure 18).



g_i – fold expression of the respective gene;

τ_i – time constant defining transcript decay;

$\Delta\tau_j$ – is a time-delay term, a generalization of the original CTRNN definition (Hu *et al.*, 2005; Kim *et al.*, 2007), which accounts for the time delay between gene induction, transcription, translation and final effect of a gene j on any other gene;

W_{ij} – square matrix describing connection weights from gene j → i;

θ_j – offset term accounts for basal gene expression (noise in the system);

σ – non-linear sigmoid activation function incorporating interaction between genes by weighting the input of gene j → i, as defined by an exponentially decaying function, factor *a* controls the steepness of σ(*x*) to adjust the transition width between the off and the on state of a gene;

I_j^0 – external signal modeled here as an exponential decay function. The input in the form of an external cellular stimulus and its impact on the network is not known a priori, hence multiple approaches are possible - constant, pulse-like, decaying/increasing, or periodic;

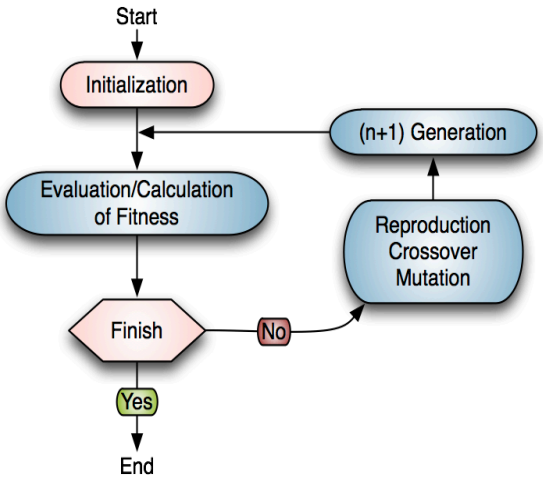
λ – input signal decay

Figure 18 ODE of a neural network type defining the interactions and behavior of each node in the network

Updating the state of a CTRNN is accomplished by numerically integrating the defining differential equations. The method of choice for the ODE integration is a forward Euler update, which is the default procedure. This first order solver, due to the nature of the neural network and multiple iterative processes involved in the evolutionary programming implemented in the algorithm, provides an optimal trade-off between computational burden and quality of the results. An adaptive method of a 4th-order Runge-Kutta update was implemented by me (inefficiently at the moment) for testing purposes, but has not yielded any significant fitness improvements, which can be attributed to successful fitness evaluation with mean square error function. A one step Midpoint method will be implemented in the future to improve overall efficiency. More detailed parameter ranges are presented in the corresponding result sections for each of the models.

5.7.3 Modeling procedure

A typical CTRNN simulation proceeds as follows: first, a CTRNN is created and initialized with a random set of parameters created within predefined (literature) biological boundaries, then a loop is entered in which the appropriate CTRNN state update function is called repeatedly. Model parameters, as defined in the ODE, are fitted to the experimental time series with a genetic algorithm (GA) as the global optimization method using mutation, crossing-over, selection, and an elitist strategy.



1. Choose initial population
2. Evaluate each individual's fitness
3. Repeat
4. Select individuals to reproduce
5. Mate pairs at random
6. Apply crossover operator (exchange whole sets of parameters)
7. Apply mutation operator (randomly modify a parameter)
8. Evaluate each individual's fitness
9. Until terminating condition (e.g. predefined no. of generations)

Figure 19 The canonical genetic algorithm procedure

The learning loop terminates when a satisfactory solution is found (exact fit to the data), or the number of generations has passed a preset limit (in our case 1500-2000). The subsequent stages are arranged in a manner shown in Figure 20 beginning with long-term network simulations, which are an attempt to establish the *in silico* system behavior. Solutions are selected among the resulting systems based on a combination of fitness (evaluated with a Mean Square Error function (MSE)) to experimental data and robustness to perturbations (established with the Largest Lyapunov Exponent (LLE) analysis).

Since the number of measured time-points is far lower than the number of evaluated parameters a dimensionality problem arises and is solved by including interpolation of data (cubic spline function with a sampling interval of Δt=0.05h (3 minutes)), using pre-defined, biologically reasonable parameter ranges, time delay on gene activity, search for a balance between network topology robustness and systems fitness to the experimental parameters (the fittest solution is not

always the true solution, because such solution must be also robust towards parameter variability, accounting for the inherent noise in a biological system). Additionally we evaluate a zero input response as an augmentation of the fitness function, under the assumption that the fold expression of genes should not change in any significant way in the absence of an external signal. Hence, systems, which homeostatic state becomes unstable over time in the absence of an external input receive a fitness penalty. See Busch et al. 2008 Supplementary materials for a more detailed description of the NN implementation. Model parameters and evolved parameter ranges are provided in the corresponding model result sections (see for PSC 6.1.7.2 and TC 6.2.8.3).

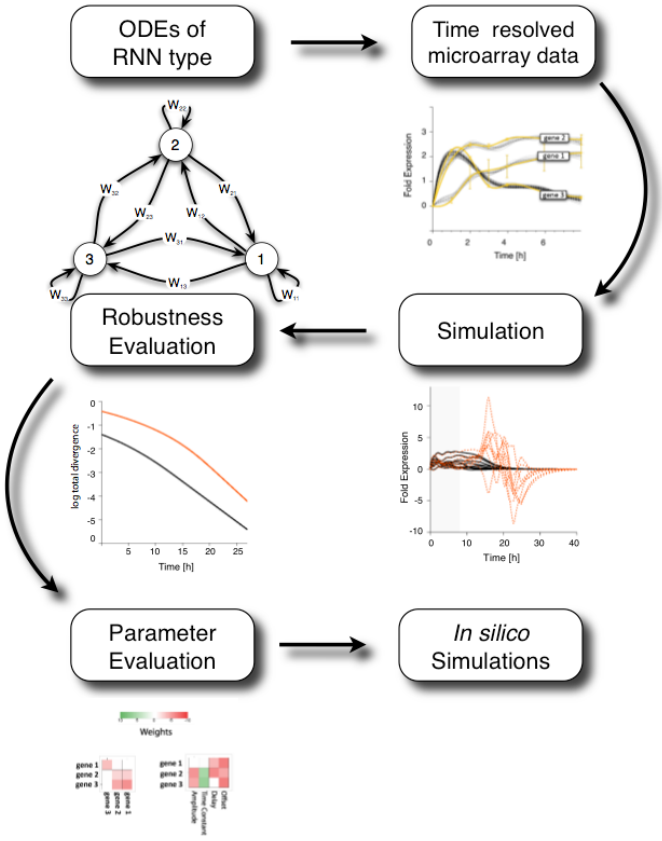


Figure 20 Modeling and simulations (from a technical introduction by Dr. Hauke Busch, DFKZ)

5.7.4 Integration of models

A neural network of the CTRNN type, as used in this project, is a directed cycle (Figure 21:A), an 'enclosed' entity, capable of receiving external signals, and evolving all parameters to identify interactions between neurons from expression data, providing as a result optimal model parameters fitted to experimental data, but incapable of producing a measurable output (it is not a layered network with output neurons as in Figure 21:B).

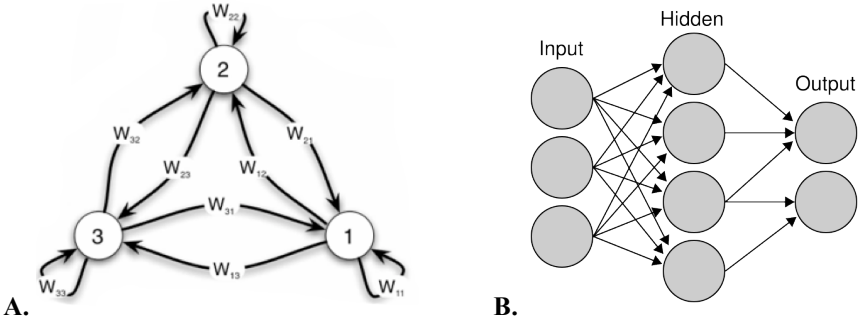


Figure 21 A. CTRNN (Busch et al. 2008); B. feed-forward neural network (Wikimedia Commons, GNU License, created by Colin M.L. Burnett)

Lack of output neurons means that creation of an integrated model consisting of two sub networks, each derived from a different cell type, connected by their extracellular signals is not possible. Such a construct where the output of one network is used dynamically as an input for the other system, and all of them are allowed to freely evolve over time in search of a stable solution is the ultimate goal of reverse engineering; unfortunately no such model exists at the moment. However for the needs of this project, model integration was performed by replacing the theoretical external input functions (e.g. exponentially decaying initializing input, and secondary inputs of various types e.g. positive/negative periodic) with a measured microarray gene expression of known cytokines from the other cell type. This cytokine signature was constructed from the combined and averaged kinetics of all upregulated transcripts of secreted factors identified in the top 500 genes of the cell type of interest forming the intercellular communication.

5.8 Knowledge-driven identification of intercellular signaling

The analysis of signaling between cells, which accounts for intercellular communication is non-trivial, and usually involves experimental screening procedures to distinguish between overlapping and specific interactions. We propose here a knowledge-driven approach, which facilitates the identification of highly specific extracellular signals affecting each of the underlying modeled gene regulatory networks in the cells. Although the identification of genes encoding soluble proteins from microarray data is easy with tools such as GO, the identified signaling profiles are limited to factors inducible in response to stimulation discarding the highly constitutively expressed genes. In addition they are usually unspecific, as proteins secreted by one cell type are often overlapping with the other cell type, forming a distinctive microenvironment that can be considered as background for the actual stimulation. Disturbing communication between cells using those connections is not optimal. More valuable are unique factors, which specifically affect the formation of the proposed GRNs.

The proposed approach involves the coupling of two aforementioned methods (Pathway analysis, and TFBSA) with reverse engineered GRN and microarray data in a four-stage procedure (plus the initial experiment and modeling stages) (Figure 22).

First, TFBSA is used to identify TFs underlying the modeled regulation. **Second**, derived TFs are used in Pathway Studio, where pathways are expanded and extracellular proteins regulating those intracellular gene networks are identified (direction of the regulation is of paramount importance). In the **third stage**, an overlap between the theoretically derived (Pathway Studio) secreted proteins and actual microarray data for each cell type of interest is performed to identify only those proteins that act on the GRN of interest (Figure 22). Finally in the **fourth step**, all derived soluble factors are combined into a single list, and filtering of data is performed to group genes into sets uniquely produced by either cell type, or by both cell types with an additional investigation of factors inducible in response to stimulation.
The last two steps ensure that the literature-derived data from Pathway Studio is verified against the actual microarray experiment providing a tissue-specific context.

As a result of this approach, the modeled GRNs are connected with the soluble factors driving them. TFBSA restores information, which may otherwise be lost due to filtering and ranking. Pathway Studio provides knowledge-based pathways of direct and indirect interactions between

genes and proteins in various compartments of the cell. Depending on cell compartment of interest the procedure may begin at e.g. TF and expand to encompass extracellular factors, which affect them, or it may begin with known secreted proteins in a search for their common targets in the nucleus.

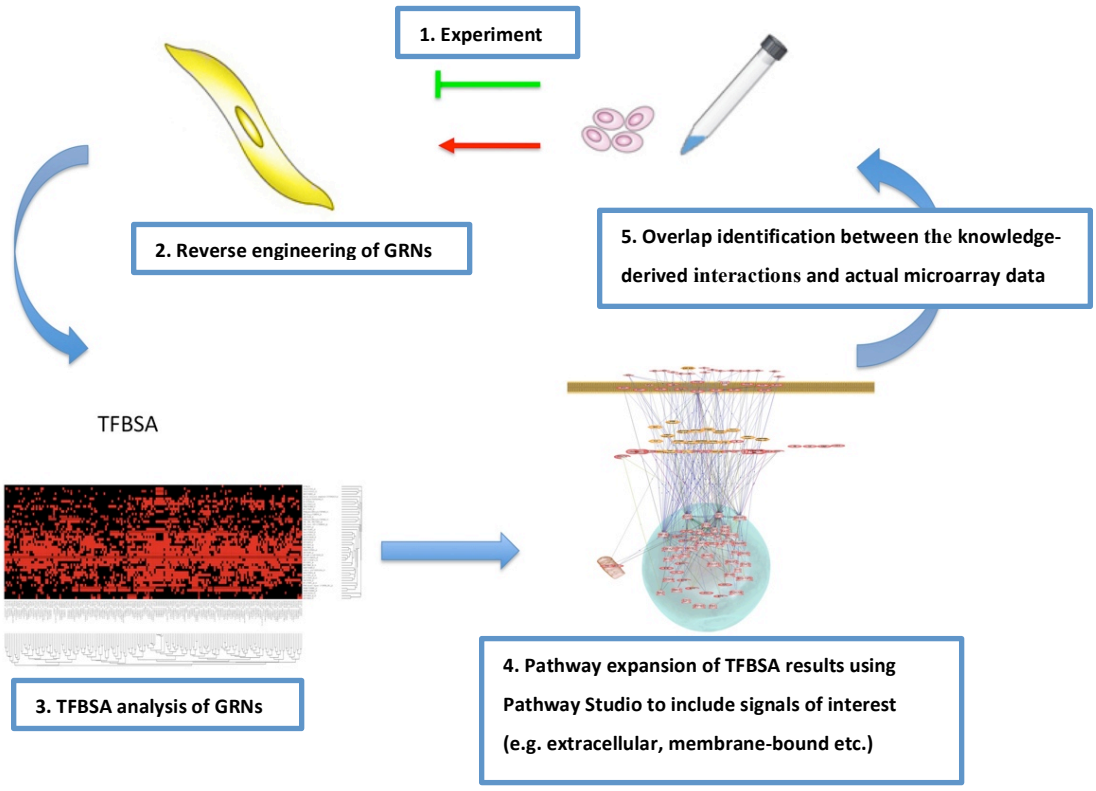


Figure 22 Complete procedure for the identification of extracellular signals in the context of reverse engineered gene regulatory networks

5.9 Brief overview of experimental validation methods

Since neither the design nor the execution of the experiments is part of my work, only a general overview of the procedures will be offered. Detailed analysis of the experimental results was performed by me and can be located in the corresponding sections of the results.

5.9.1 siRNA knockdowns

In order to identify the effect of the selected genes of interest on tumor cells in the context of pancreatic microenvironment, we applied a functional genomics approach based on RNA interference also known as loss-of-function screening. RNA interference (RNAi) is a robust method of posttranscriptional silencing of genes using double-stranded RNA (dsRNA) in the form of either siRNA (short interfering RNA as used here) or shRNA (short hairpin RNA) with sequence homology driven specificity. RNA interference was previously shown to be an effective way of evaluating sensitizing targets in pancreatic cancer (Azorsa et al. 2009).

5.9.2 Controls

Two sets of controls were applied to each experimental setup:

- non-transfected control: normal, non-transfected tumor cells from the corresponding cell line (MiaPaCa2 or Panc1);
- negative control: tumor cell transfected with scrambled siRNA.

Comparing cells transfected with a scrambled sequence siRNA control to non-transfected cells reveals changes caused by the process of siRNA delivery. Therefore a complete analysis involves comparisons between the non-transfected tumor cells, cells transfected with negative control siRNA and cells transfected independently with gene specific siRNAs. This provides the ultimate control over specificity of siRNA effects. While additional control may be provided with rescue experiments in which the RNAi effect is reversed through expression of a target gene refractory to silencing by a particular siRNA, due to limited material availability this approach was not pursued here.

5.9.3 qRT-PCR profiling

In order to confirm the microarray data results gathered from all time series experiments (replication of the microarray), as well as validate simulations, knockdown predictions of the core network, and identified intercellular interactions quantitative reverse transcription PCR experiments have been performed by PD Dr. Thomas Giese, Institute of Immunology, University of Heidelberg.

5.9.4 Functional assays

An additional set of assays was used to investigate biological implications of the knockdowns on tumor and stellate cells including a clonogenic, MTT, and an invasion assay. Clonogenic and invasion assays have been setup as co-culture systems to investigate the effects of the investigated cell types on each other and the effect of GRN predicted knockdowns. While the phenotypic implications remain outside of the scope of the presented work, samples were gathered from all co-culture systems for qRT-PCR evaluations and were used in the experimental validation section of this thesis (see 6.5).

5.9.4.1 Statistical testing of validation experiments

The readouts of assays were of quantitative order (numerical), groups were compared using non-parametric Mann-Whitney and paired Wilcoxon test, as well as parametric two-way ANOVA wherever the corresponding replicated data sets show an approximately Gaussian distribution and pass the d'Agostino-Pearson normality test. Post-tests were performed using the Bonferroni method to address the problem of multiple comparisons.

6. Results

6.1 Stellate cells

6.1.1 Array normalization and filtering

Normalization of all nine arrays in the PSC experiment was performed using RMA expression measure to unify data distribution (see 5.4.1) and visualized with box plots (Figure 23) to verify the quality of the processing. Distribution histogram showed a strong positive skewing (Figure 23) therefore an additional step of IQR filtering has been applied with a filter cut-off of 0.25 ($\log_2$) to achieve a more Gaussian distribution (Figure 24) (5.4.2).

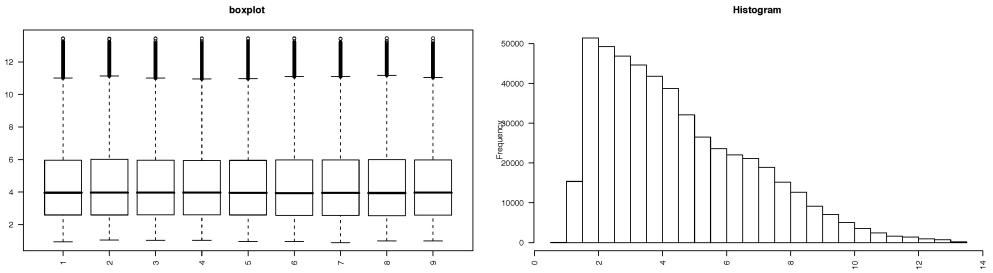


Figure 23 RMA normalization results of the stellate cell microarrays, before filtering.

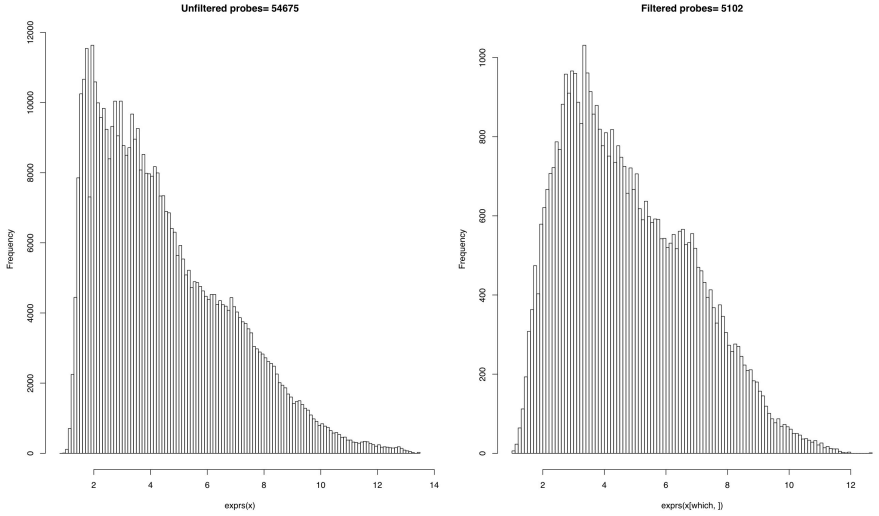


Figure 24 Filtering procedure results in a more normal distribution of the data

6.1.2 Gene ranking

Initial IQR filtering reduced the initial microarray dataset from 54 676 down to 5101 probe sets, a number that still translates into a complex set of gene expression kinetics as shown in Figure 25.

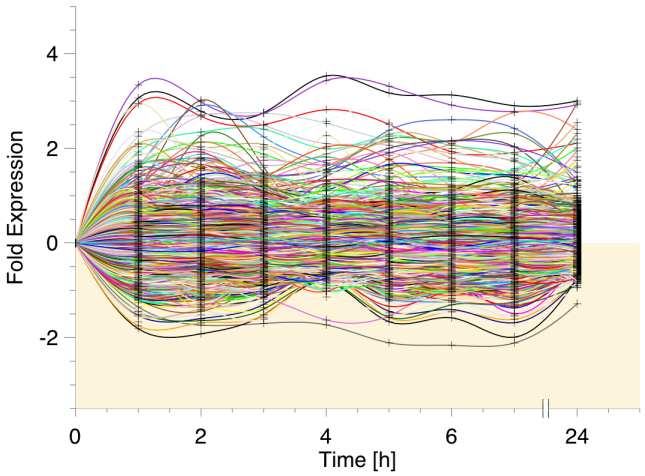
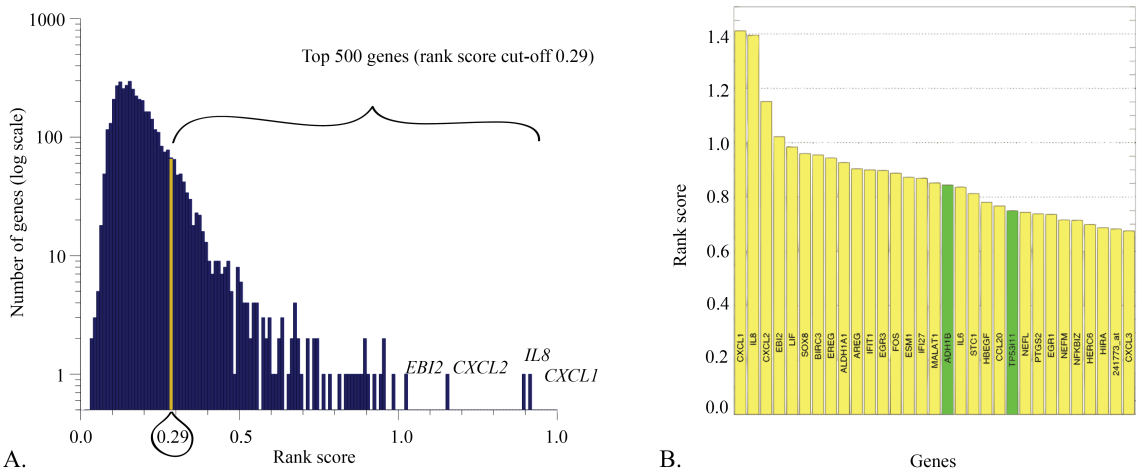


Figure 25 Filtered microarray data retains its time-resolved complexity

Genes were ranked (according to 5.5.1) and sorted from the most to the least differentially changing. The distribution of the rank scores is presented in Figure 26a, and shows that we have a relatively small number of genes, that are very highly upregulated, and a majority of genes in the low score range of 0.0 to 0.5. Top 500 ranked genes (absolute) fall within score range above 0.29 (Euclidian metric), and only 38 genes achieve a mean $\log_2$ FC ratio of > 1.0 $\log_2$. A quick overview of top 30 deregulated genes shows that the majority of them is upregulated in response to the treatment as shown in Figure 26.b.

A two-fold change is represented by a log2 ratio of 1.0 (up-regulation) or -1.0 (down-regulation), a three-fold change in gene expression is represented by a $\log_2$ ratio of 1.58 or -1.58, 0.0 means that there is no change.

PSC Setup #1



Experiment	Rank score cut-off (Euclidian distance)	Mean abs FC cut-off (log_2)	Number of genes with abs mean FC>1.0	Lowest ranked gene in top 500	Highest ranked gene (mean FC log_2)
Experiment 1 PSC	0.29	0.24	38	PKD2 (+0.24)	CXCL1 (+2.7)

Figure 26 A. Distribution of rank scores for both up- and down-regulated genes. B. Top 30 ranked upregulated (yellow) and down-regulated (green) genes in stellate cells.

6.1.3 Statistical analysis of PSC

Statistical analysis using ***Edge*** (5.5.2), was applied to compare FC ranking, with cubic spline regression under the assumption that the top 30 genes, while reshuffled, should still remain within the investigated range of top 500 FC ranked set (250 up- and 250 down-regulated). A comparison of the statistics (sorted) with FC-based ranking is shown in Table 4 (downregulated genes are shown in green). Only 2 out of top 30 genes ranked with *Edge* are found outside of the top 500 ranked genes confirming a reasonable overlap.

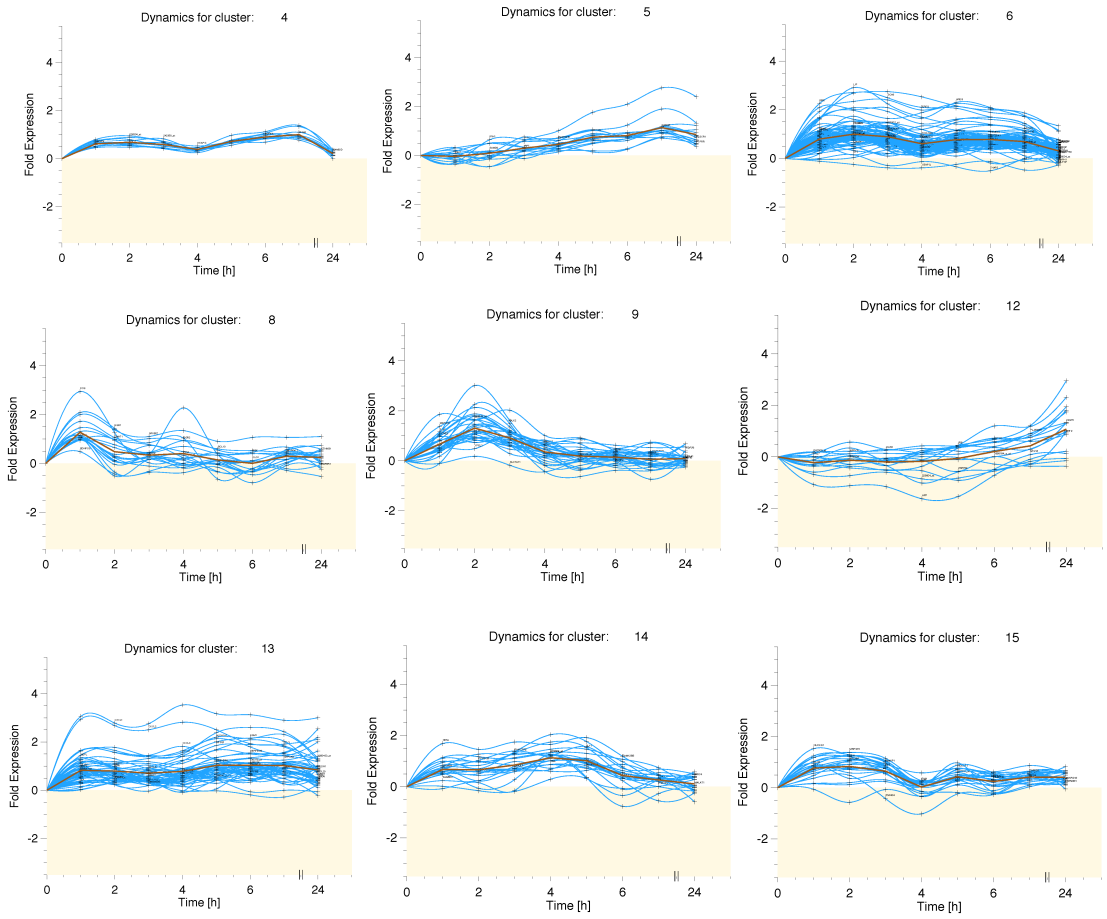
Rank Fold change	Rank EDGE	P-Value EDGE	Q-Value EDGE	ProbeID	Gene Name
66	1	2.74E+01	0.07935286	218400_at	OAS3
110	2	2.16E-04	0.07935286	209278_s_at	TFPI2
17	3	2.86E-04	0.07935286	202411_at	IFI27
11	4	3.08E-04	0.07935286	212224_at	ALDH1A1
49	5	3.14E-04	0.07935286	34478_at	RAB11B
37	6	3.33E-04	0.07935286	205113_at	NEFM
22	7	3.35E-04	0.07935286	205207_at	IL6
14	8	3.41E-04	0.07935286	205239_at	AREG
23	9	3.51E-04	0.07935286	214974_x_at	CXCL5
6	10	3.68E-04	0.07935286	205419_at	EBI2
228	11	4.16E-04	0.07935286	201601_x_at	IFITM1
26	12	5.04E-04	0.07935286	203821_at	HBEGF
250	13	5.49E-04	0.07935286	201890_at	RRM2
29	14	5.96E-04	0.07935286	38037_at	HBEGF
98	15	7.59E-04	0.07935286	1554026_at	MYO10
146	16	9.58E-04	0.07935286	224657_at	ERRFI1
659	17	1.00E-03	0.07935286	235609_at	235609_at
259	18	1.07E-03	0.07935286	202796_at	SYNPO
126	19	1.08E-03	0.07935286	239629_at	CFLAR
410	20	1.09E-03	0.07935286	202572_s_at	DLGAP4
273	21	1.26E-03	0.07935286	215136_s_at	EXOSC8
246	22	1.28E-03	0.07935286	217066_s_at	DMPK
251	23	1.29E-03	0.07935286	213125_at	OLFML2B
16	24	1.30E-03	0.07935286	209612_s_at	ADH1B
89	25	1.34E-03	0.07935286	230380_at	THAP2
285	26	1.42E-03	0.07935286	226218_at	IL7R
332	27	1.43E-03	0.07935286	229578_at	JPH2
57	28	1.46E-03	0.07935286	202376_at	SERPINA3
82	29	1.52E-03	0.07935286	202644_s_at	TNFAIP3
551	30	1.56E-03	0.07935286	208626_s_at	VAT1

Table 4 A comparison of the cubic spline statistics (sorted) with fold change-based ranking (green: downregulated genes)

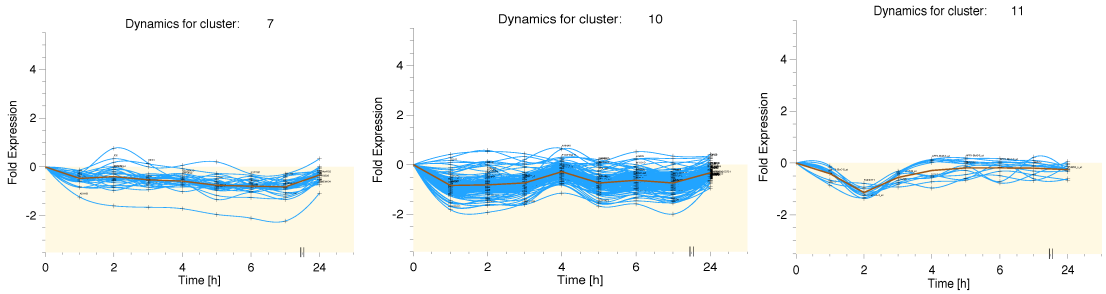
6.1.4 Exploratory analysis

6.1.4.1 Bayesian clustering results

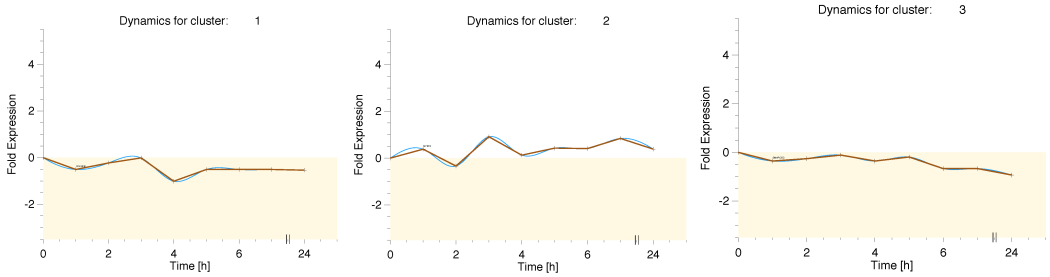
To identify unique gene expression kinetic profiles, which reflect intracellular information flow and responses to extracellular stimuli, we applied a Bayesian clustering algorithm (BC) to the set of top 500 deregulated genes in the PSC experiment. BC divided gene kinetics into **12 major clusters** and **three outliers**: *CDKN2B* (up), *MYEF2* (up), and *SMARCD3* (down) (Figure 27). **Nine** of the twelve clusters are **upregulated**; the remaining **three** (clusters 7, 10 and 11) are **downregulated**.



A. Upregulated



B. Downregulated



C. Outliers

Figure 27 Clustering results for stellate cell experiment (1) divided into three main sets.

Interestingly, while the complexity visible among the upregulated genes in 9 clusters is to be expected in response to the treatment with TC supernatants, the downregulation is somewhat more surprising. A vast majority of the top downregulated genes can be located in a single cluster (10) with the top 10 downregulated genes including: *ARHGDIA, ATN1, HSPB6, TP53I11, RAB11B, SLC9A3R2, GRINA, ADH1B, TLN1, and BCN*.

6.1.5 Gene ontology of stellate cells

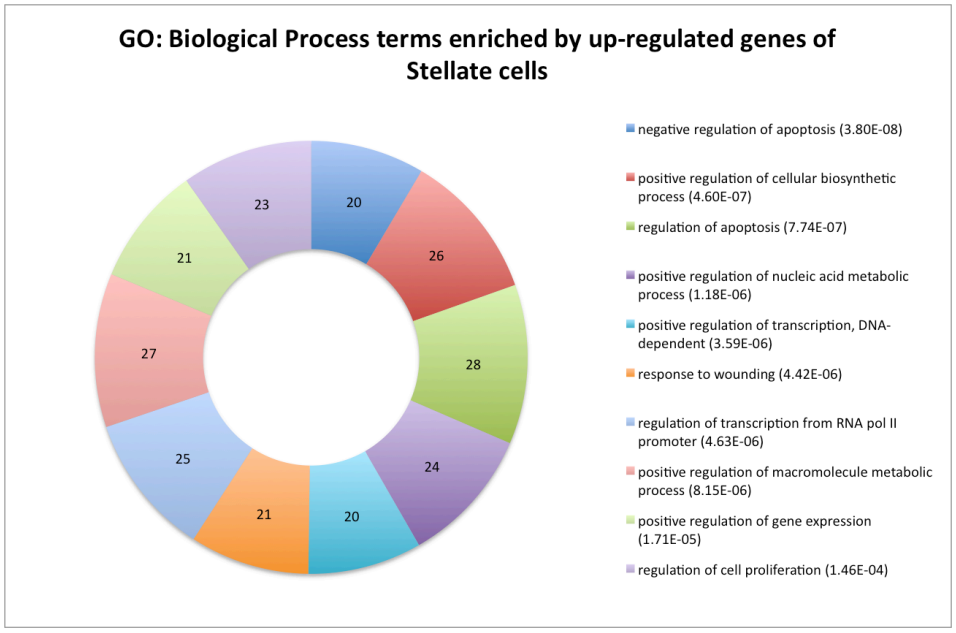
The GO analysis of the PSC experiment was performed using the ***David*** and ***WebGestalt*** (5.6.3). *David* has the benefit over *WebGestalt* in that it provides a set of tools for functional classification and clustering of genes in addition to standard GO enrichment analysis, however *WebGestalt* is far more capable in terms of overviewing the GO data and its visualization. Each analysis step was performed using the standard selection of top 250 up- and 250 down-regulated probe sets (Affymetrix IDs).

6.1.5.1 *David*

Affymetrix IDs were converted into 211 (up-) and 208 (down-regulated) *David* IDs respectively, and their corresponding official gene symbols, discarding duplicates. Functional annotation and functional classifications were performed on both datasets. Using the functional annotation tools in *David* we were able to determine which genes in our dataset share the same functional annotation, and which form patterns of significant regulation in the stellate cells. A combination of the statistic (low p-value) and the number of genes enriching the GO terms provides the most informative results. The most interesting results are located in the Biological Process branch of GO and are presented in the pie chart below (Figure 28), which shows that the PSC actively downregulates pro-apoptotic signaling pathways, increases metabolic processes, and shows wound-healing related signaling.

Functional clustering of the GO terms in stellate cells reveals 40 upregulated clusters (13 with an enrichment score of >1.5), and around 80 downregulated (18 with an enrichment score >1.5).

A.



B.

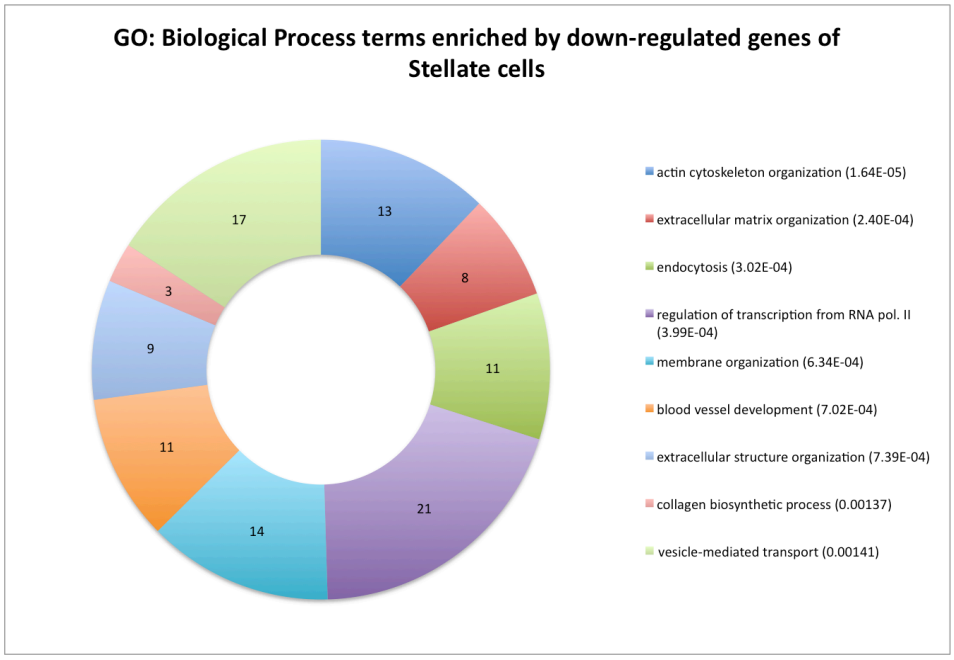


Figure 28 Most enriched, fine-grained, gene ontology terms from top up- (A.) and down-regulated (B.) genes in stellate cells.

6.1.5.2 WebGestalt

WebGestalt was used to perform an enrichment analysis. Default settings used here include the hypergeometric test for enrichment evaluation, and multiple test adjustment with Benjamini & Hochberg (BH).

Upregulated genes: Total number of 250. Unambiguously mapped User IDs to Entrez IDs: 196. Unique User Entrez IDs: 175. The Enrichment Analysis is based upon the unique IDs.

Downregulated genes: Total number of User IDs: 250. Unambiguously mapped User IDs to Entrez IDs: 219. Unique User Entrez IDs: 191. The Enrichment Analysis is based upon the unique IDs.

A visualization of GO enrichment was achieved with a directed acyclic graph as shown for the Molecular Function branch of Gene Ontology in Figure 31. The most fine-grained term of gene ontology within the molecular function branch is chemokine activity. This suggests that the stellate cell likely actively responds to the stimulation by regulating the levels of secreted proteins.

Figure 29 and Figure 30 present an overview of the Biological Process and Cellular Component of GO Slim, for top 250 upregulated, and top 250 downregulated genes respectively. Since the analysis was performed using GO Slim, the overview is not as fine-grained as with the complete *David* results. Interestingly the main GO terms enriched by the upregulated genes in stellate cells are related to response to stimulus, cell communication, and developmental processes (Figure 29 biological process)

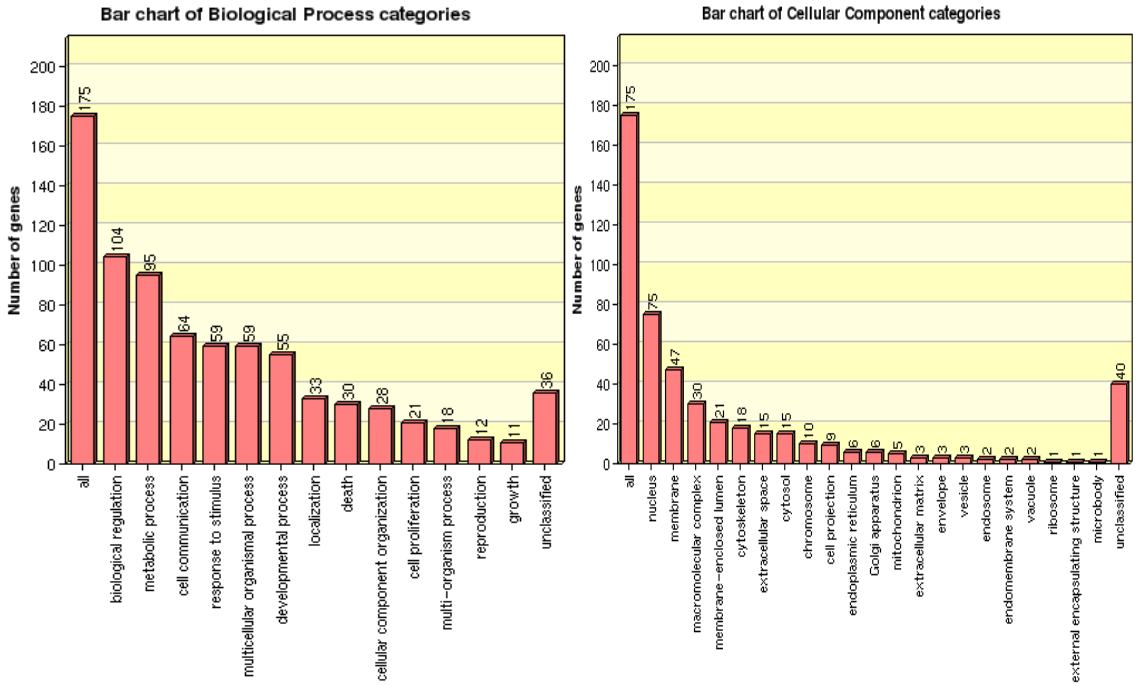


Figure 29 Top 250-upregulated genes enriching the biological process and cellular component of GO slim

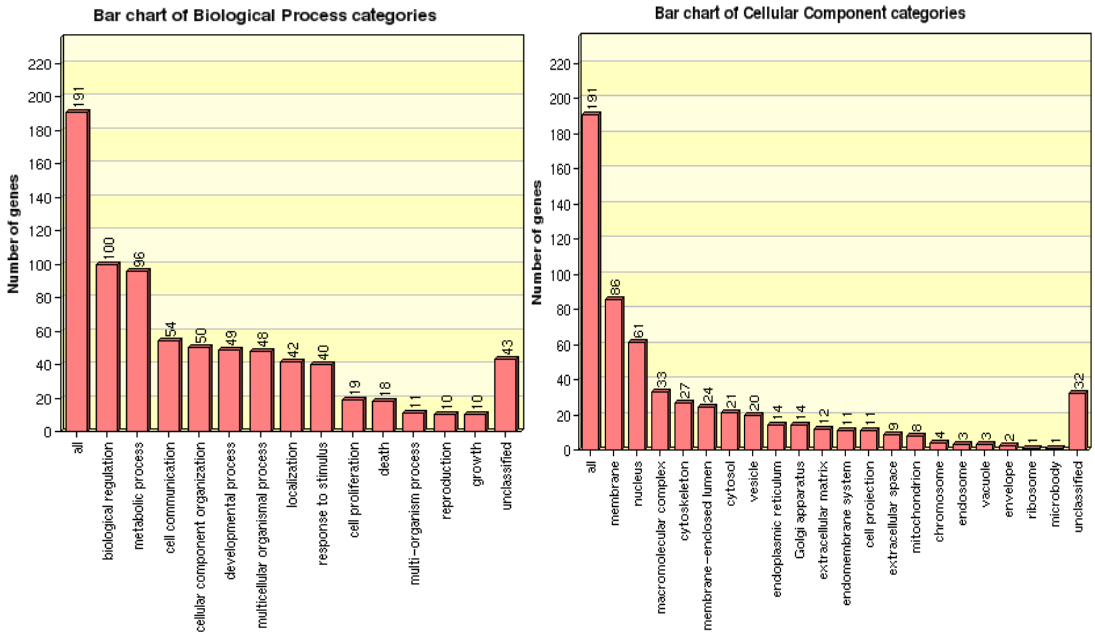


Figure 30 Top 250-downregulated genes enriching the biological process and cellular component of GO slim

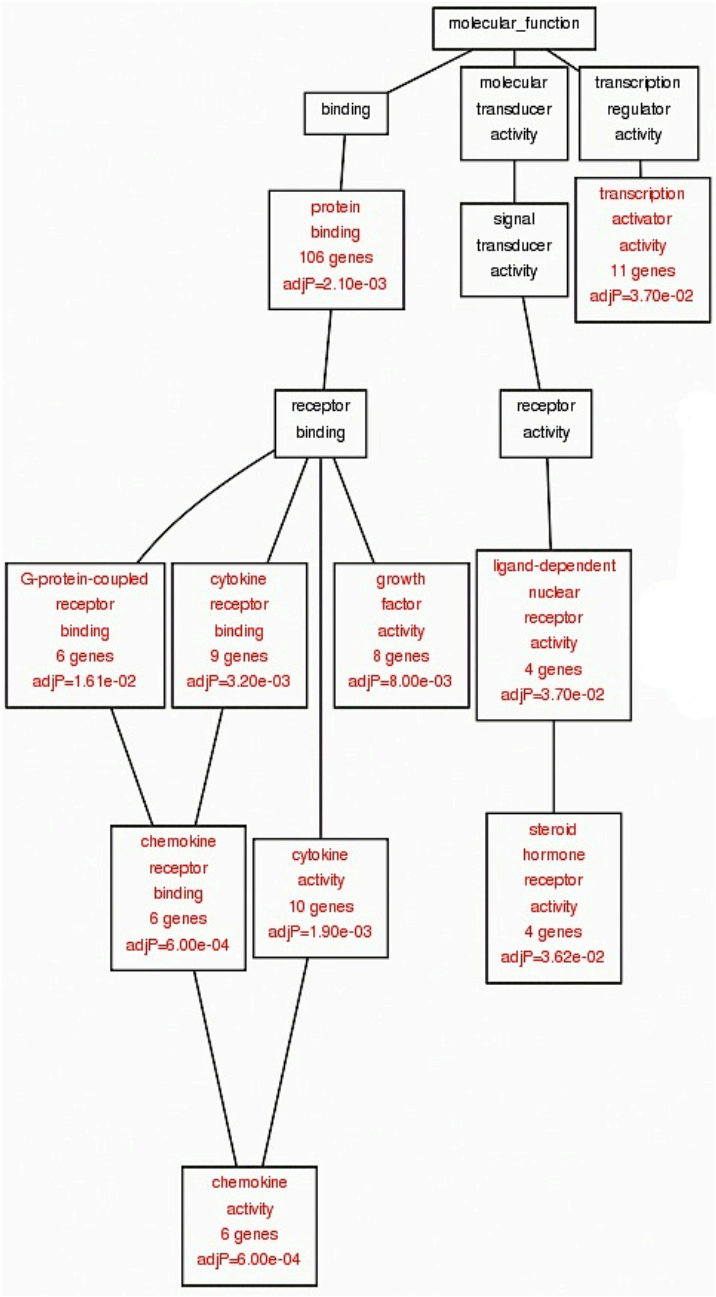


Figure 31 Directed acyclic graph presenting the top 10 terms in the Molecular Function branch of Gene Ontology enriched by top 250 upregulated probe sets (WebGestalt).

6.1.6 Pathway analysis

WebGestalt was used for the initial pathway analysis using KEGG, WikiPathways and Pathway Commons analysis options (5.6.5). Gene selection and analysis parameters were identical to the earlier GO analysis (6.1.5). Top 10 KEGG results are shown (Table 5).

Upregulated		
NOD-like receptor signaling pathway	adjP=1.43e-07	HSP90B1, TNFAIP3, BIRC3, IL6, NFKBIA, IL8, XIAP, CXCL1, CXCL2
Small cell lung cancer	adjP=0.0012	CCNE2, PTGS2, BIRC3, NFKBIA, ITGA2, XIAP
Cytokine-cytokine receptor interaction	adjP=0.0012	CXCL5, LIF, IL11, IL7R, CCL20, IL6, CXCL3, IL8, CXCL1, CXCL2
Pathways in cancer	adjP=0.0079	CCNE2, PTGS2, HSP90B1, BIRC3, IL6, NFKBIA, IL8, ITGA2, XIAP, FOS
Chemokine signaling pathway	adjP=0.0079	CXCL1, CXCL5, CXCL2, CCL20, CXCL3, NFKBIA, IL8
Apoptosis	adjP=0.0079	BIRC3, IRAK2, NFKBIA, CFLAR, XIAP
Hematopoietic cell lineage	adjP=0.0079	IL11, CD44, IL7R, IL6, ITGA2
Toll-like receptor signaling pathway	adjP=0.0102	FOS, IL6, NFKBIA, MAP3K8, IL8
Epithelial cell signaling in Helicobacter pylori infection	adjP=0.0126	CXCL1, HBEGF, NFKBIA, IL8
T cell receptor signaling pathway	adjP=0.0126	NFKBIE, FOS, NFKBIA, MAP3K8, BCL10

Downregulated		
Focal adhesion	adjP=4.36e-05	PARVB, PARVB, ZYX, PXN, COL4A2, TLN1, ITGA7, COL1A1, COL6A1, COL6A1, RAC1, COL3A1, VEGFB, GRLF1
ECM-receptor interaction	adjP=0.0006	COL4A2, AGRN, HSPG2, ITGA7, COL3A1, COL1A1, COL6A1
VEGF signaling pathway	adjP=0.0144	MAPKAPK2, NFATC4, PXN, MAP2K2, RAC1
Notch signaling pathway	adjP=0.0162	NOTCH1, NCOR2, DVL3, JAG1
Neurotrophin signaling pathway	adjP=0.0162	MAPKAPK2, CALM3, YWHAE, MAP2K2, RAC1, ARHGDIA
Pathways in cancer	adjP=0.0162	CDKN2B, COL4A2, RUNX1T1, RAC1, DVL3, TGFB2, MAP2K2, FGFR1, VEGFB, RXRB
Wnt signaling pathway	adjP=0.0216	DVL3, LRP5, NFATC4, PRKACA, DAAM2, RAC1
Metabolism of xenobiotics by cytochrome P450	adjP=0.0216	GSTA4, ADH1B, EPHX1, CYP1B1
Melanogenesis	adjP=0.0216	CALM3, CREB3L1, DVL3, PRKACA, MAP2K2
Prion diseases	adjP=0.0255	NOTCH1, PRKACA, MAP2K2

Table 5 Top 10 KEGG pathways identified among the top expressed genes in PSC

A closer inspection of pathways significantly enriched by genes in both sets reveals the intercellular interactions through the cytokine and chemokine terms, as well as an enrichment of pathways related to immune and antiviral responses (Table 5: Upregulated).

6.1.7 Model of the stellate cell gene regulatory network

6.1.7.1 Gene selection

Gene selection was knowledge-driven, and used the results of Bayesian clustering to form network modules (6.1.4.1).

Module 1	Module 2	Module 3	Module 4	Module 5
FOS	ZC3H12A	BHLHB3	NR4A2	ATF3
EGR1	ZC3H12C	ZEB1	NR4A3	FOSB
	PTGS2	KLF7	FGFR2	JUNB
	KLF6		KLF4	EGR3
	KLF13		KLF5	
			KLF11	
Module 6	Module 7	Module 8	Module 9	
HIRA	IFIT1	RGS2	GPRC5A	
MALAT1	IFIT2	RGS3	MTSS1	
PHACTR2	IFIT3	RGS10		
APOL6	IFITM1	BIRC2		
	IFITM2	PTPRE		
	OAS3			
	MX1			
	ICAM1			

Table 6 Gene regulatory network modules selected for reverse engineering

In general each module for network reconstruction represents an individual expression profile identified using BC. PSC gene selection was augmented by manual pre-selection performed by dr. Axel Szabowski, therefore the division of genes into modules is based on expression profile structure derived from BC and supported by functional classification. Each gene from the final selection was analyzed in detail (Table 7) and alterations to the BC results were performed manually whenever functional reassignment was necessary e.g. *KLF6* (BC cluster 8 along *FOS* and *EGR1*) fits functionally better with Module 2 genes related to apoptosis.

- Module 1 Immediate early gene family of transcription factors (*FOS, EGR1*);
- Module 2 Apoptosis related (ZC3H12a, ZC3H12c, PTGS2, KLF6/13);
- Module 3 Proliferation and differentiation (*BHLHB3, ZEB1, KLF7*);
- Module 4 Proliferation, apoptosis, invasion (*KLF4, KLF5*), nuclear receptors with pleiotropic effects (*NR4A2, NR4A3*)
- Module 5 AP-1 transcriptional complex (*EGR3, ATF3, FOSB, JUNB*);
- Module 6 generally unknown function, significantly upregulated with a common kinetic profile of interest (*APOL6, HIRA, MALAT1, PHACTR2*)
- Module 7 Inflammatory state response, interferon induced proteins (*IFIT1, IFIT2, IFIT3, IFITM1, IFITM2, ICAM1, OAS3, MX1*);
- Module 8 (Family of *RGS* genes, *BIRC2, PTPRE*);
- Module 9 Differentiation and metastasis (*GPRC5A, MTSS1*)

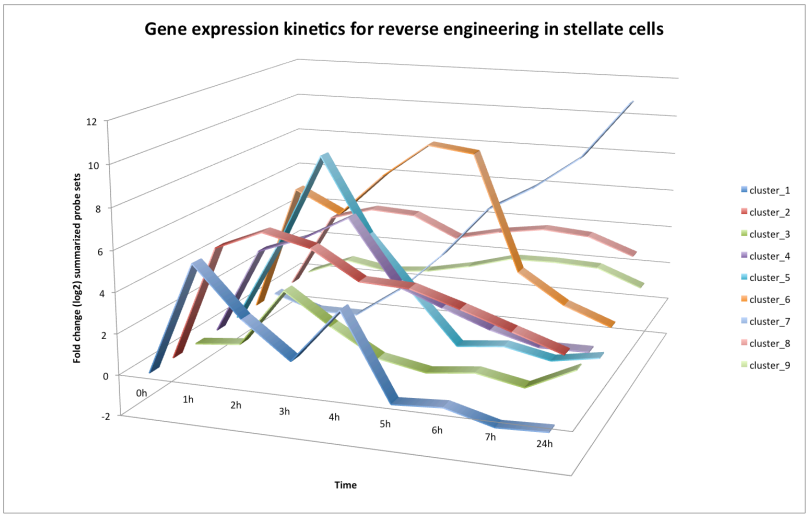


Figure 32 Gene expression kinetics for summarized probe sets in experiment 1 stellate cells

IL8	CXC chemokine family, one of the major mediators of the inflammatory response, functions as a chemoattractant, is a potent angiogenic factor, contributes to the aggressive biology of human pancreatic cancer.
CXCL1	A small cytokine implicated in melanoma pathogenesis. Expressed by macrophages, neutrophils and epithelial cells, has neutrophil chemoattractant activity. Plays a role in developmental processes and is involved in angiogenesis, inflammation, wound healing, and tumorigenesis. Its up-regulation has been attributed to constitutive activation of NFκB, which is an emerging hallmark in various types of tumors including breast, colon, pancreatic, ovarian, as well as melanoma.
Malat1	(Metastasis-associated lung adenocarcinoma transcript) a noncoding RNA, associated with metastasis in early-stage non-small cell lung cancer (NSCLC)
EBI2	Superfamily of rhodopsin-like 7TM receptors (seven-transmembrane segment receptors), also known as G-protein-coupled receptors. Known to control leukocyte movements.
GPRC5A	A retinoic acid-induced protein 3. Significant in light of the fact that PSC activation is associated with loss of cytoplasmic vitamin A (retinol) stores. Metabolites of retinol include all-trans retinoic acid and 9-cis retinoic acid (9-RA). It may be a link between retinoid acid and G protein signaling pathways. May play a role in embryonic development and epithelial cell differentiation, and has been implicated in tumor suppression in lung cancer.
AREG	(amphiregulin) a ligand of the EGF receptor, an autocrine growth factor, as well as a mitogen for astrocytes, Schwann cells, and fibroblasts. Interacts with the EGF/TGF-alpha receptors to promote the growth of normal epithelial cells and inhibits the growth of certain aggressive carcinoma cell lines.
ESM1	Implicated in the regulation of the LFA-1/ICAM-1 pathway, may therefore influence both the recruitment of circulating lymphocytes to inflammatory sites, and LFA-1-dependent leukocyte adhesion and activation. ESM1 binds to integrin and blocks binding to intercellular adhesion – reasonable since one of the most profound features of PDAC is desmoplasia
IL6	Cytokine Interleukin-6 (IL-6) is a pro-inflammatory cytokine secreted by T cells and macrophages to stimulate immune response to trauma, especially burns or other tissue damage leading to inflammation.
LIF	(Leukemia inhibitory factor) a cytokine of the same family as IL6. Has been shown to induce macrophage differentiation.
STC1	(Stanniocalcin) is a glycoprotein hormone involved in calcium and phosphate homeostasis. STC1 is present in breast ductal epithelium, and its expression is induced by BRCA1, a tumor suppressor gene that has an important role in breast and ovarian cancer. STC1 is differentially expressed in a number of cancers compared with the relevant normal tissues.
IFIT1	(Interferon induced protein) is a cytokine most commonly produced by the cells of the immune system in response to challenges such as viruses, bacteria, parasites and tumor cells. It is not surprising to see a great increase in interferon-induced proteins, as PSC are known to produce cytokines during formative stages of chronic pancreatitis.
DUSP6	a dual specificity protein phosphatase that negatively regulates ERK2 member of the mitogen-activated protein (MAP) kinase superfamily, which are associated with cellular proliferation and differentiation.

Table 7 Sample set of genes of interest in the pancreatic stellate cells according to GeneCards (as accessed at the Weizmann Institute of Science)

6.1.7.2 Model parameters

Following parameters (Table 8) were used to build the model of stellate cell GRN (see 5.7).

Model parameters:	
Number of generations	2000
Population size	700
Runs	2000
Mutation rate	0.01
Euler integration parameters:	
Time delay	15
Time step size	0.05
Interpolation points resulting from time step size	480
CTRNN parameters:	
Initialization of parameters	Randomized
Sigmoidal function interpolation points	1000
Range of interaction weights	10
Offset range	2.5
Decay range	2.5
Minimal time constant of decay	0.1
Delay range	20
Range of initial input	15
Fitness points (how many interpolation points are evaluated with the mean square error function)	140
Second input beginning	40
Period of the second input	5.0
Amplitude of the second input relative to the learned first	0.1
Input halflife in minutes	0.5
Tension of the spline interpolation of the data	1.0 (<<1 cubic, >>1 linear)

Table 8 PSC CTRNN model parameters

Out of a total of 480 interpolated points in the time series, 9 were experimentally measured points (0h-7h, 24h), the remaining were spline interpolated. 340 interpolated values between 7h and 24h were of limited use in the fitness evaluation as they only loosely approximated the actual signal in the cells. Two strategies were applied to this part of data. Fitness evaluation was performed either on all interpolated points for the initial 7 hours (140 points), or in a second approach on all 480 points, but with an additional modification of the fitness function (C code implementation). This modification introduced a 10% penalty to the interpolated values ensuring that the experimental points get a higher score. Both approaches produced equally satisfying results (fitness >90%), and both have been evaluated for the final model. The first one is most often used for visualization, because the fitness is higher in the initial data points (0-7h) where the resolution of experimental values is higher.

6.1.7.3 Fitness evaluation

Fitness of the model to the initial 7 experimental data point in PSC experiment is shown in (Figure 33), the total fitness over 24h was > 95% .

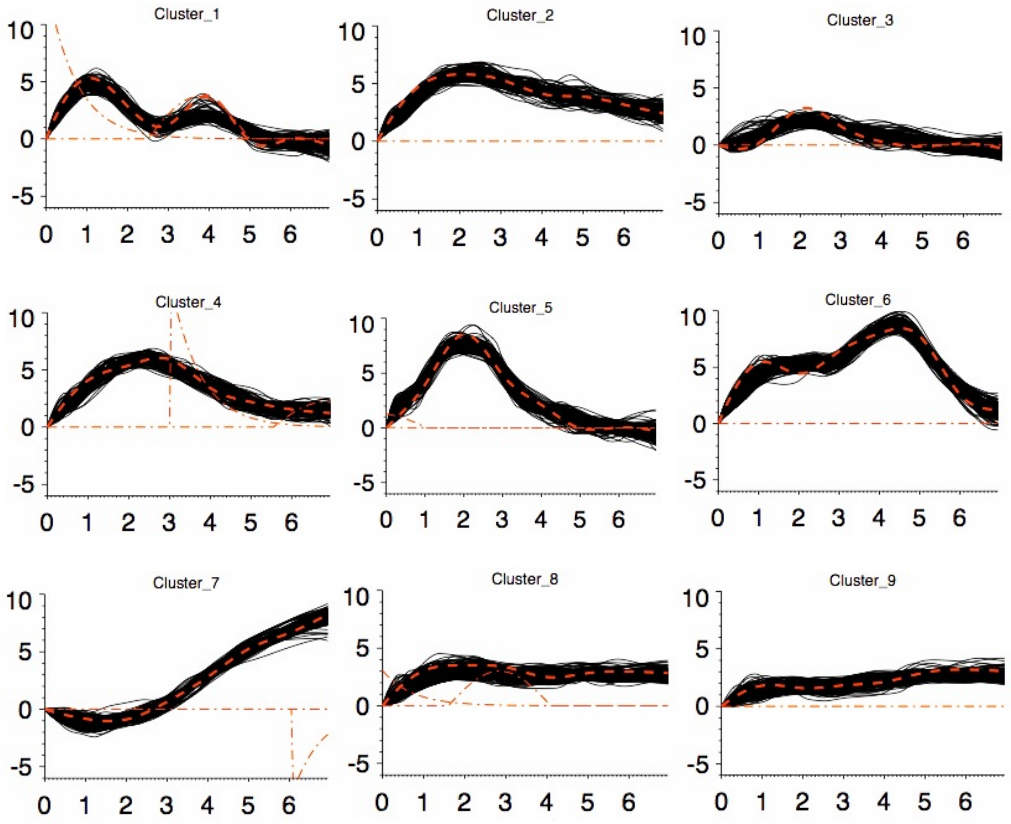


Figure 33 PSC model fitted solutions plotted for each cluster. Dashed red line represents the experimentally measured and interpolated values, black lines are the modeled solutions, and dashed-and-dotted red lines represents the external input functions as well as a 0 input value line for reference.

6.1.7.4 System selection using LLE

Largest Lyapunov Exponent has been used to identify systems, which show the most robust response to perturbation (see section 5.7.3 and Supplementary materials of Busch et al. 2008). The final systems used for *in silico* simulations and knockdowns are selected by ranking them using the Euclidian distance between the robustness and fitness criterions. The LLE method is described in more detail in our paper, Busch et al. (2008).

6.1.7.5 Interaction matrix and the resulting network visualization

Results of the modeling are presented in the form of a weight matrix, where weights over all time points were summarized for visualization purposes, and each rectangle depicts a summarized interaction of one gene (module/cluster) with another.

Figure 34 PSC GRN weight matrix. Positive (activation) interactions are in red, negative (inhibition) in green.

Subsequently the weight matrix was converted into a weighed and directed interaction network (Figure 35). The visualization is organized in sequence of clusters from 1-9 from top to bottom. The first 5 clusters contain the early and intermediate response genes. They are responsible for receiving the external signal (paracrine), and are driven by it. The majority of those genes transduce the signal downstream towards their targets and into what we call a 'central hub of genes' in this network (*HIRA, APOL6, MALAT1, PHACTR2*). 'Central hub' is a module, which integrates and distributes this signaling into the 'effector' clusters. Those are distinctly different, as they show a switch-like behavior, once initialized - they remain permanently upregulated in response, and result in the changed behavior of the cells.

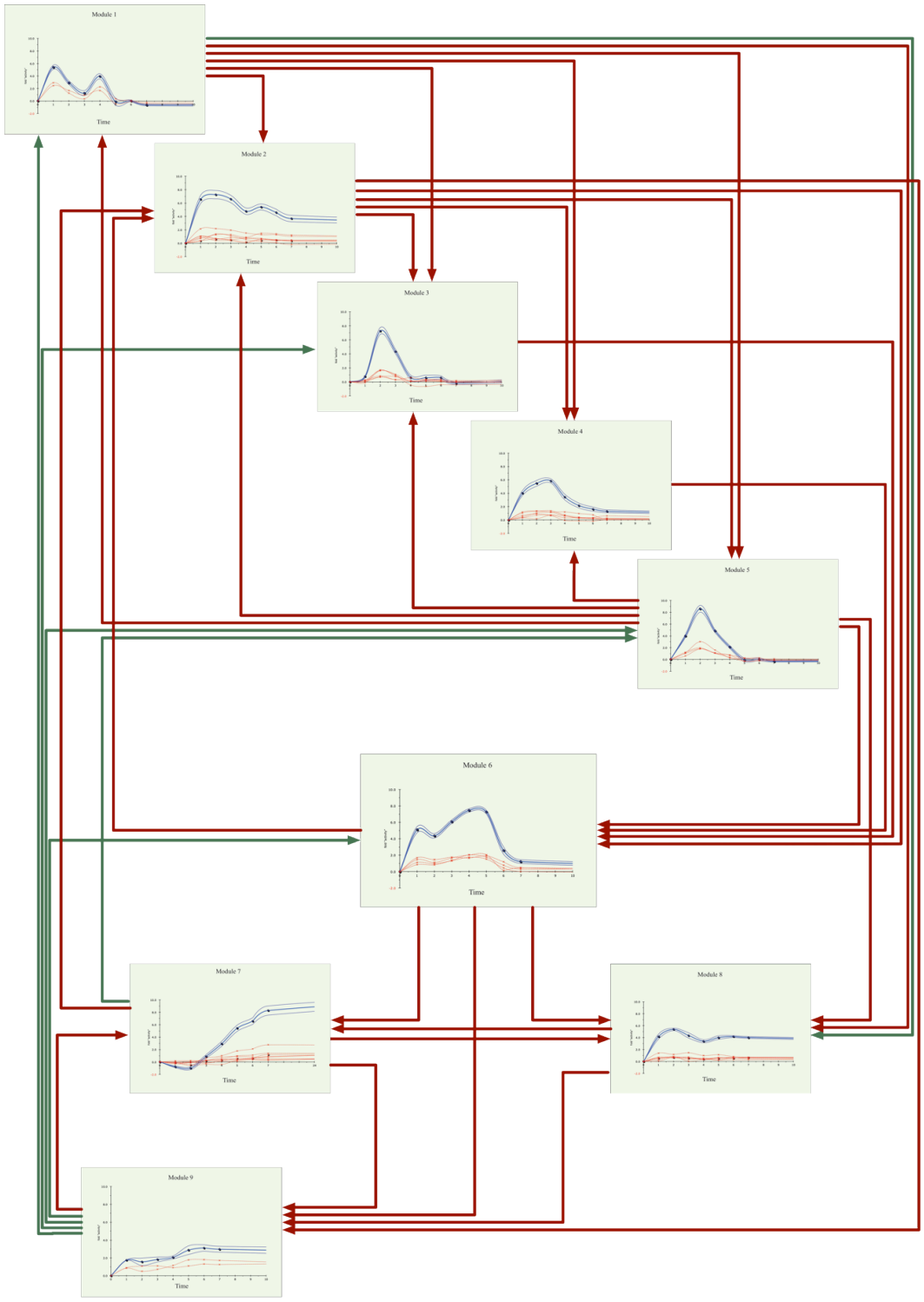


Figure 35 Interaction network (red for activation, and green for inhibition). Only effective weights are plotted accounting for the offset parameter value corresponding to the identified system noise.

6.1.7.6 *In silico* simulations and perturbations

We assume that reconstructed GRNs are formed in response to the PSC-TC crosstalk and remain stable. We next hypothesize that interruption of elucidated interactions should disturb network formation, and we set out to predict which module knockout would lead to a system-wide response breakdown. To elucidate those central breakpoints, we used the derived model for *in silico* simulations and knockdowns. An *in silico* knockdown screen was performed for all modules in each network. Two points were revealed in PSC (Figure 36) including AP-1 complex, and module 6 genes (*HIRA, APOL6, PHACTR2, MALAT1*). Simulations show that the AP1 knockdown downregulated immediate early and delayed gene responses, but system recovers its activity after approx. 7h, proving AP1 redundancy. Module 6 knockdowns resulted in a system-wide breakdown, which suggests its central role in the PSC signaling hierarchy, positioning it as a potential experimental target.

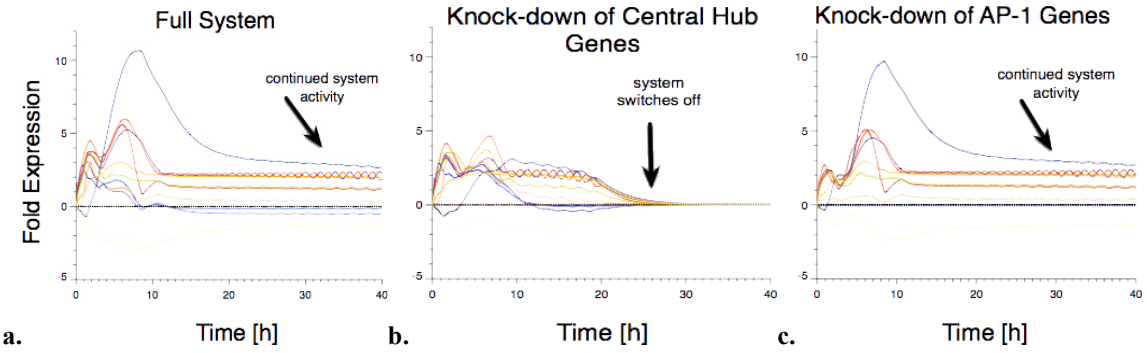


Figure 36 In silico simulation of the native system (a.) and knockdowns of AP1 (b.) and central hub of genes (c.)

6.1.7.7 Stellate cell output

The biological output of the PSC experiment is in the form of a changed profile of soluble factors. Cytokines and chemokines affect the formation of the microenvironment, act on TC in a paracrine feedback loop, and on PSC in an autocrine manner. Identification of the most upregulated genes encoding those proteins was performed using GO among top 250 upregulated ranked genes resulting in a list of 20 factors clustering into 4 distinct profiles (Figure 37).

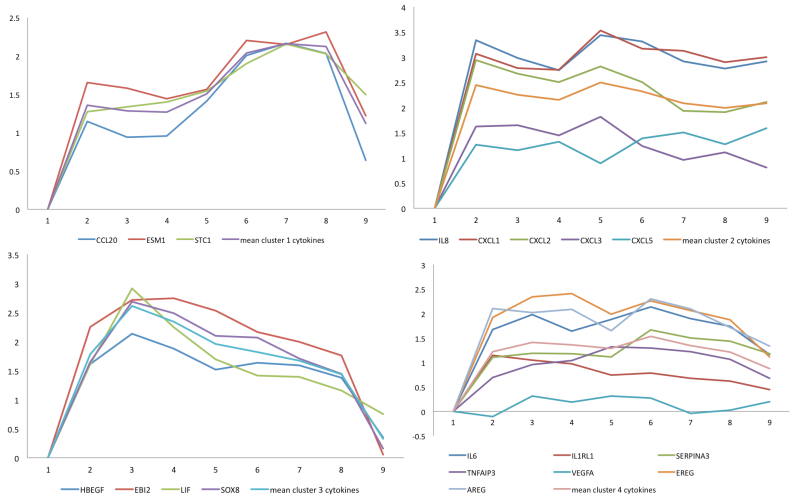


Figure 37 Four main clusters of secreted proteins produced by stellate cells in response to stimulation with tumor cell supernatant, resolved over nine time points.

Expanding the list to top 500 upregulated probe sets increases the number to a total of 39 identified soluble factors (Table 9).

CXCL1	CLCF1	STC1	KITLG
CXCL2	EBI2	TIMP3	VMO1
CXCL3	LIF	ESM1	ST3GAL1
CXCL5	S100A6	TFPI2	SERPINA3
CCL20	EREG	TNFAIP3	ISG15
IL1RL1	AREG	ADAMTS6	LAMC2
IL7R	PLAUR	FAM55C	BDNF
IL6	EGFR	DKK3	GNPTG
IL8	HBEGF	ANGPTL4	COL17A1
IL11	VEGFA	PCOLCE2	

Table 9 Secreted factors and membrane bound proteins identified among top 500 upregulated genes in stimulated stellate cells using *David* Gene Ontology analysis.

6.2 Tumor cells

Having successfully established that the stimulation of PSC by tumor cell supernatant results in a strongly altered expression profile of PSC, we used experiments 2 and 3 to determine what effect the stimulation of PSC has on TC. Both experiments were performed with the same TC line (MiaPaCa2), treated with supernatants from either quiescent (experiment 2), or stimulated PSC (experiment 3).

Processing of the two TC microarray experiment data, which includes normalization, filtering, ranking as well as high-level analysis with gene ontology, and pathway analysis, was performed with the same stringency as in the PSC experiment analysis. Additional steps were introduced to specifically focus on potential differences between TC experiments.

6.2.1 Array normalization and filtering

Normalization and filtering procedures were performed as described in section 5.4 of Materials and Methods resulting in a unified data distribution.

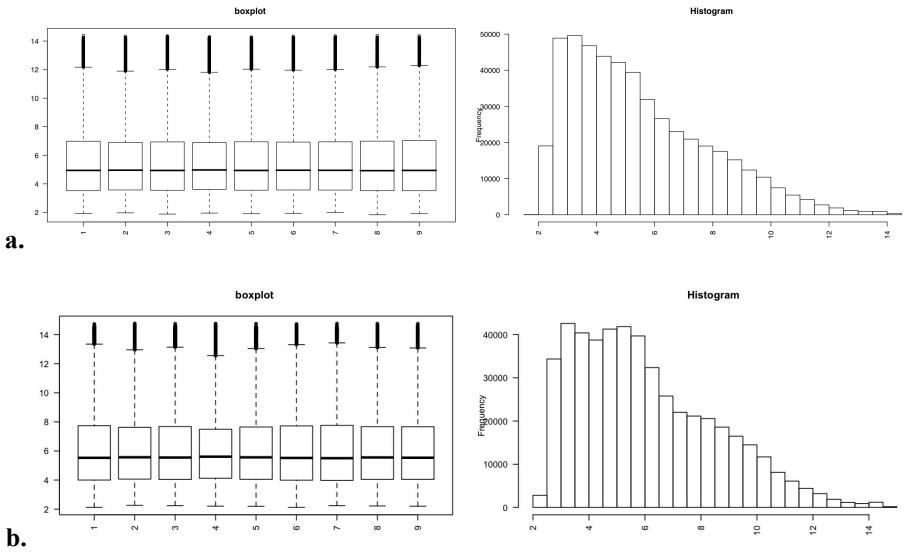


Figure 38 RMA normalization results of the tumor cell experiment 2 (a) and experiment 3 (b) microarrays, before filtering.

An IQR filter was applied to achieve a more Gaussian distribution (Figure 39).

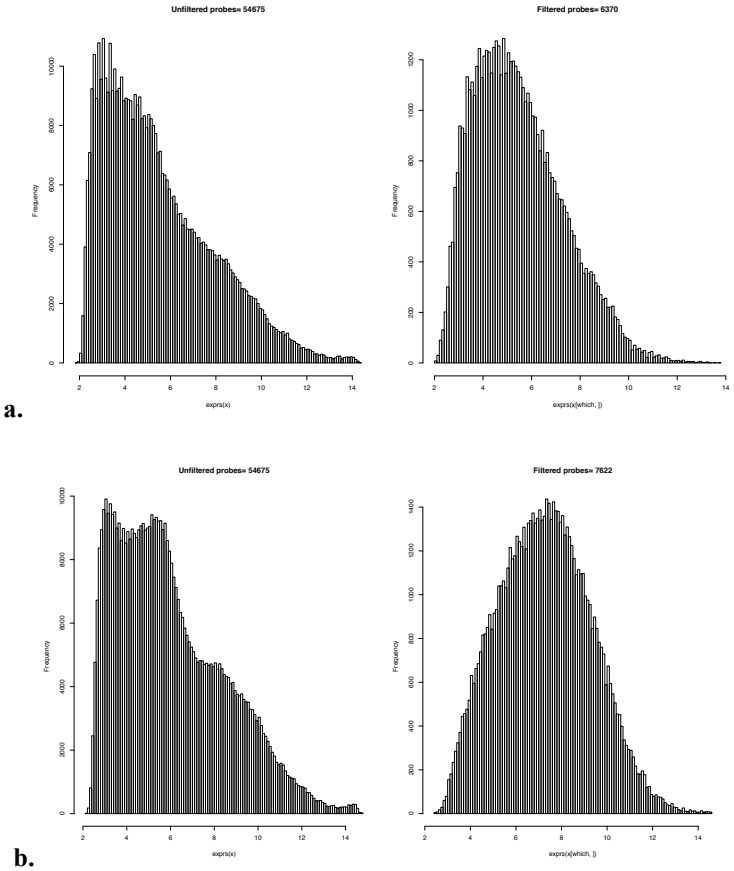
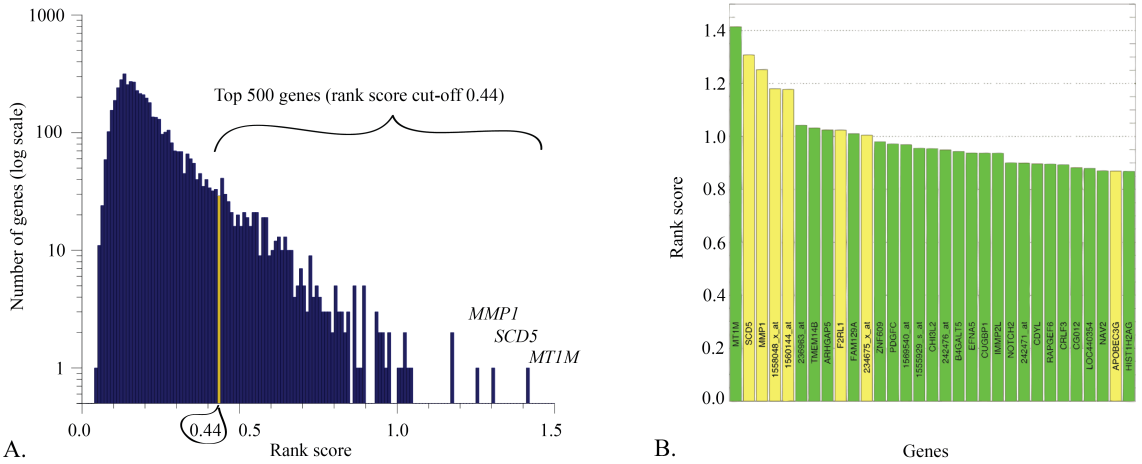


Figure 39 IQR filtering procedure with a cut off value 0.25 for experiment 2 (a) and 0.35 for experiment 3 (b)

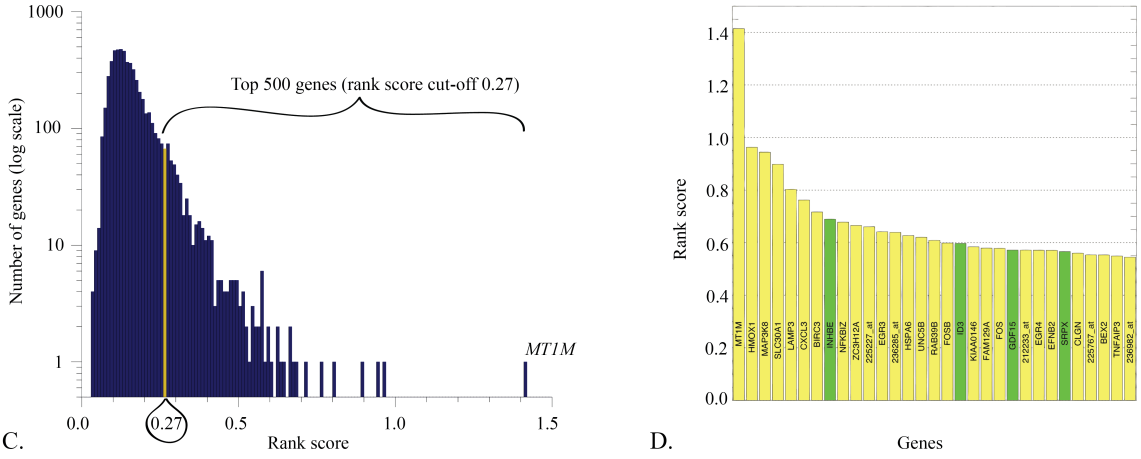
6.2.2 Gene ranking

Gene ranking according to the Euclidian distance between the mean and peak expression was performed using $\log_2$ fold change values as described in section 5.5.1.

TC Setup #2



TC Setup #3



Experiment	Rank score cut-off (Euclidian distance)	Mean FC cut-off (abs. FC $\log_2$)	Number of genes with mean FC>1.0	Lowest ranked gene in top 500 (abs. FC $\log_2$)	Highest ranked gene (mean FC $\log_2$)
TC experiment 2	0.44	0.43	235	225239_at (+0.43)	MT1M (-2.54)
TC experiment 3	0.27	0.44	157	IDI1 (+0.44)	MT1M (+4.75)

TC#2	
Rank	GeneName
0	MT1M
1	SCD5
2	MMP1
3	1558048_x_at
4	1560144_at
5	MBNL1
6	IMMP2L
7	ZCCHC7
8	PIAS1
9	MGAT5
10	ATP1B3
11	MBNL1
12	236963_at
13	TMEM14B
14	JMJD1C
15	ARHGAP5
16	F2RL1
17	FAM129A
18	234675_x_at
19	KIAA1267
20	IMMP2L

E.

TC#3	
Rank	Gene Name
0	MT1M
1	HMOX1
2	MAP3K8
3	SLC30A1
4	SLC30A1
5	LAMP3
6	BIRC3
7	CXCL3
8	HSPA6
9	NFKBIZ
10	RAB39B
11	INHBE
12	WIPF3
13	ZC3H12A
14	225227_at
15	NPL
16	EGR3
17	236285_at
18	BIRC3
19	UNC5B
20	SERPINE1

F.

Figure 40 Gene ranking of TC experiments reveals striking differences. A/C. Rank score distribution; B/D. Top ranked genes (green: downregulated, yellow: upregulated), E/F. top 20 ranked genes (green: downregulated, yellow: upregulated)

Gene ranking has allowed us to select the top 500 differentially regulated genes in the time series – a cut-off, which ensures that we capture all representative gene expression kinetics and yet retain a reasonably concise subset for further investigation. Looking at the distribution of ranked genes in Figure 40 we notice that the tumor cell response to stellate cells is clearly stimulus-dependent. Quiescent PSC elicit a broader response with a greater number of genes upregulated within the rank score range of 0.5-1.0. Stimulated PSC not only elicit a much more focused response, but more importantly geared towards upregulating the majority of top ranked genes (Figure 40.D/F), whereas quiescent PSC effect is nearly reversed (Figure 40.B/E).

A comparison of top 500 ranked genes in all three experiments reveals an overlap of only 8 genes (*240013_at, ZC3H12C, ZBTB24, FOS, ID2, C20orf117, CXCL5, IGFBP4*), however disregarding their upregulated or downregulated state (Figure 41.A.). The highest unique overlap of 48 genes is between stellate cells and tumor cells in experiment 3 suggesting at least a partial stimulation of similar pathways. A closer inspection reveals a clear-cut separation between upregulated and downregulated genes in PSC#1 and TC#2, whereas TC#3 show an additional group of ambiguous genes with a pattern of change from down- to up- and reversed. Venn diagram in Figure 41.B shows an overlap of downregulated genes, and in C upregulated. Downregulated genes overlap only between TC experiments (9 genes). The greatest overlap in upregulated genes is conserved between PSC#1 and TC#3 (41 unique genes in total).

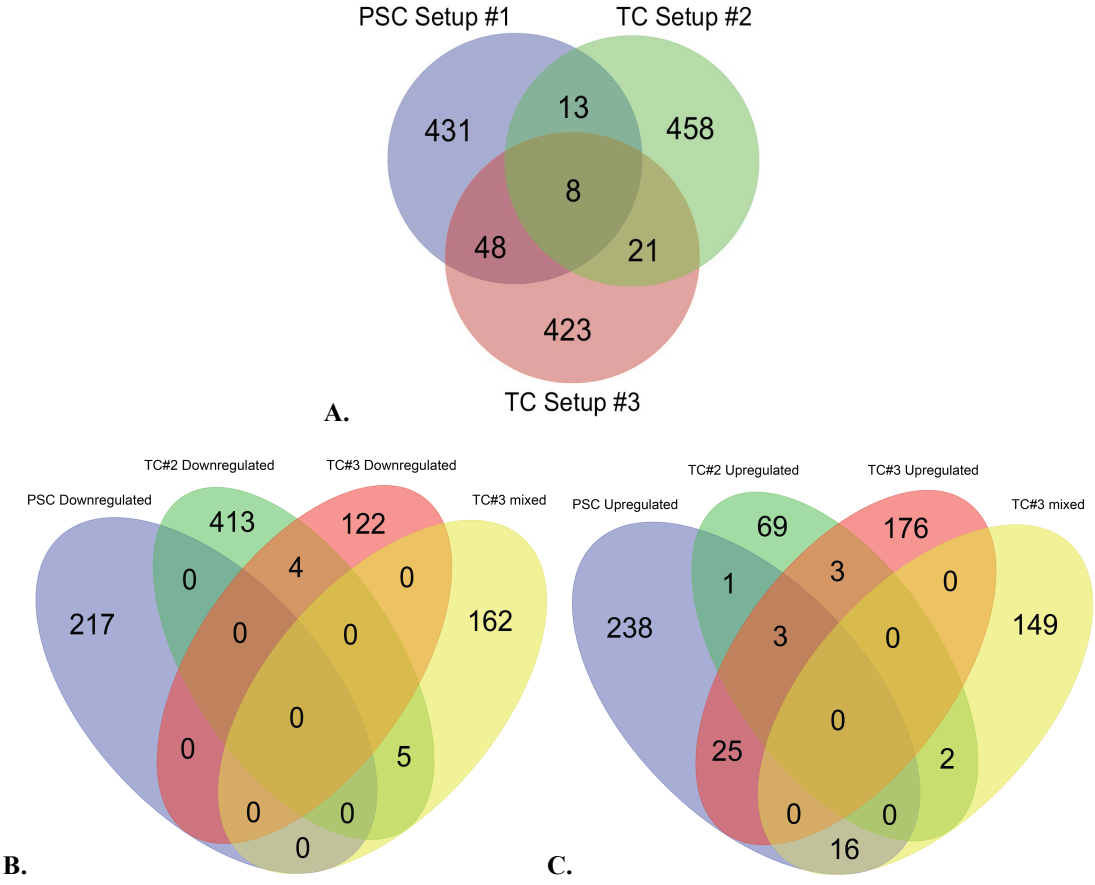


Figure 41 Venn diagram showing the total overlaps of top 500 ranked genes in all three experiments A. regardless of the state of up/down-regulation, B. downregulated, C. upregulated.

6.2.3 Statistical analysis of the tumor cell experiments

Statistical evaluation was performed with *Edge* (see section 5.5.2) and compared to the FC-based ranking. In case of TC#2 (quiescent PSC supernatant)(Table 10) 6 out of the 20 genes fall outside of the top 500 FC-based ranking.

Rank Fold change	Rank EDGE	P-Value EDGE	Q-Value EDGE	Probe ID	Gene Name
592	1	0.00056044	0.05804103	201272_at	AKR1B1
82	2	0.000992151	0.05804103	229011_at	229011_at
16	3	0.001015699	0.05804103	213506_at	F2RL1
45	4	0.001100471	0.05804103	218330_s_at	NAV2
293	5	0.00111146	0.05804103	212023_s_at	MKI67
2	6	0.001130298	0.05804103	204475_at	MMP1
3	7	0.001144427	0.05804103	1558048_x_at	1558048_x_at
322	8	0.001202512	0.05804103	201508_at	IGFBP4
78	9	0.001296703	0.05804103	214079_at	DHRS2
568	10	0.001328101	0.05804103	204950_at	CARD8
77	11	0.001503925	0.05804103	1554036_at	ZBTB24
17	12	0.001516484	0.05804103	217967_s_at	FAM129A
586	13	0.001591837	0.05804103	240126_x_at	BPTF
276	14	0.001802198	0.05804103	205547_s_at	TAGLN
319	15	0.001852433	0.05804103	232965_at	LOC400684
140	16	0.001858713	0.05804103	1554333_at	DNAJA4
47	17	0.002029827	0.05804103	207156_at	HIST1H2AG
738	18	0.002061225	0.05804103	208080_at	AURKA
18	19	0.002142857	0.05804103	234675_x_at	234675_x_at

Table 10 Comparison of Edge and fold change ranking methods for tumor cell experiment 2 (quiescent PSC supernatant)

In TC#3 (stimulated PSC supernatant)(Table 11) only 5 genes (*FDFT1, PHF19, BIRC5, PLK1, and ANLN*) fall below the top 500 FC-based rank.

Rank Fold change	Rank EDGE	P-Value	Q-Value	Probe ID	Gene Name
34	1	2.23E-05	0.02541207	226517_at	BCAT1
99	2	4.20E-05	0.02541207	206632_s_at	APOBEC3B
9	3	4.20E-05	0.02541207	223218_s_at	NFKBIZ
568	4	6.17E-05	0.02541207	210950_s_at	FDFT1
109	5	7.08E-05	0.02541207	209774_x_at	CXCL2
39	6	7.87E-05	0.02541207	202643_s_at	TNFAIP3
7	7	9.05E-05	0.02541207	207850_at	CXCL3
42	8	1.05E-04	0.02541207	224367_at	BEX2
180	9	1.13E-04	0.02541207	205047_s_at	ASNS
64	10	1.55E-04	0.02541207	225285_at	BCAT1
40	11	1.64E-04	0.02541207	205830_at	CLGN
282	12	1.68E-04	0.02541207	204470_at	CXCL1
331	13	1.76E-04	0.02541207	209608_s_at	ACAT2
679	14	1.77E-04	0.02541207	227212_s_at	PHF19
639	15	1.81E-04	0.02541207	210334_x_at	BIRC5
49	16	1.88E-04	0.02541207	202644_s_at	TNFAIP3
159	17	1.95E-04	0.02541207	201041_s_at	DUSP1
174	18	2.09E-04	0.02541207	209146_at	SC4MOL
1454	19	2.28E-04	0.02541207	202240_at	PLK1
608	20	2.44E-04	0.02541207	222608_s_at	ANLN

Table 11 Comparison of Edge and fold change ranking methods for tumor cell experiment 3 (stimulated PSC supernatant)

Those values were expected, but the performance was suboptimal, as we expected at most 10% genes to fall outside of our top 500, for that reason we performed an additional check of all significant genes according to *Edge* (Figure 42).

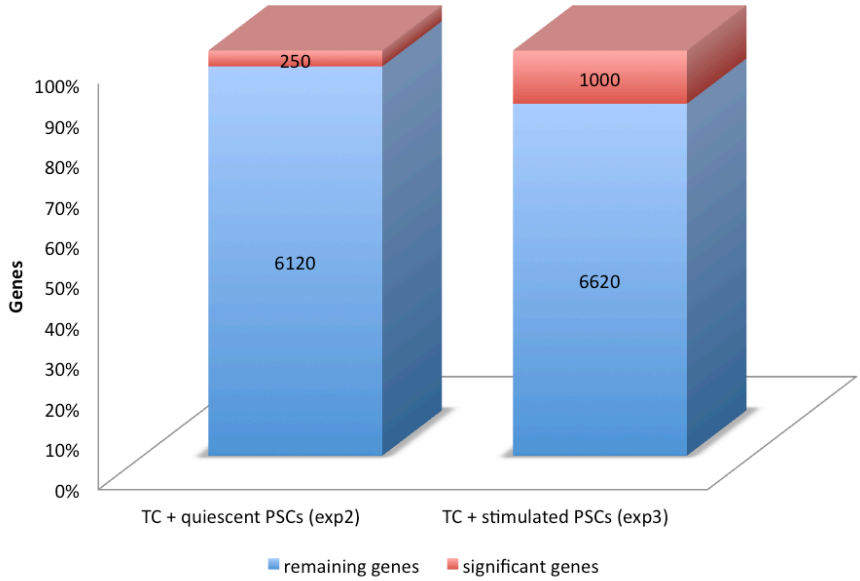


Figure 42 Stacked columns showing the fraction of significant (*Edge*) genes (red) in relation to the total number of genes (entire column) in both filtered TC data sets

Edge has identified approximately 1000 significant genes in TC#3 dataset (with a FDR adjusted p-value <1%). This 1000 significant genes covers 89% of our top 500 FC-ranked genes, which is a much more reasonable result. In contrast to that, there are 'only' approximately 250 genes identified as significant in TC#2 (with a FDR adjusted p-value <1%).

This combined with the earlier overview of rank distribution and FC, offers strong evidence of significant differences between the two TC experiments.

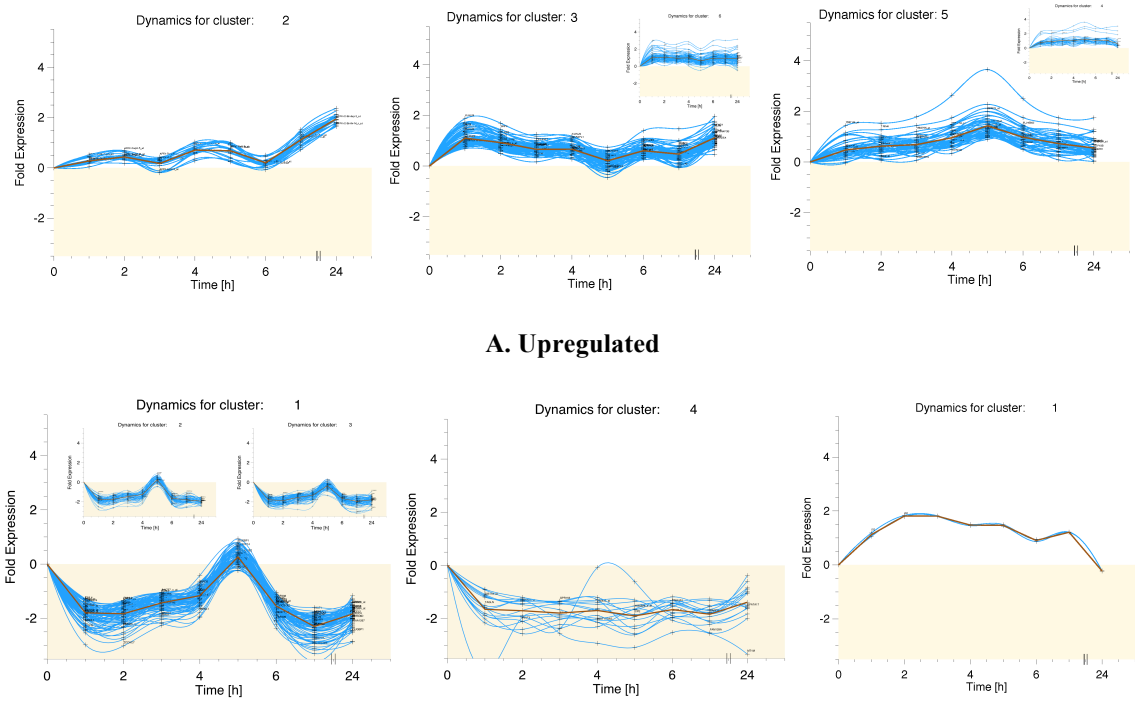
6.2.4 Exploratory analysis

6.2.4.1 Clustering results

Bayesian gene expression clustering (see 6.1.4.1) has revealed striking differences between the TC experiments. TC#2 (quiescent PSC) cluster into a significantly smaller number of profiles (6), then TC#3 (stimulated PSC) (15). The latter are additionally showing a much higher complexity, and are more difficult to organize into a smaller number of unique clusters, with visible shifts and delays in gene kinetics.

6.2.4.1.1 TC#2 (supernatant of quiescent PSC) clustering

Clustering of the top 500 genes in TC#2 resulted in a total of 5 upregulated clusters, 4 downregulated, and one upregulated outlier - *ID2*. Visual inspection and analysis of time resolved behavior allowed us to reduce the numbers into 3 and 2 respectively as shown in Figure 43:



A. Upregulated

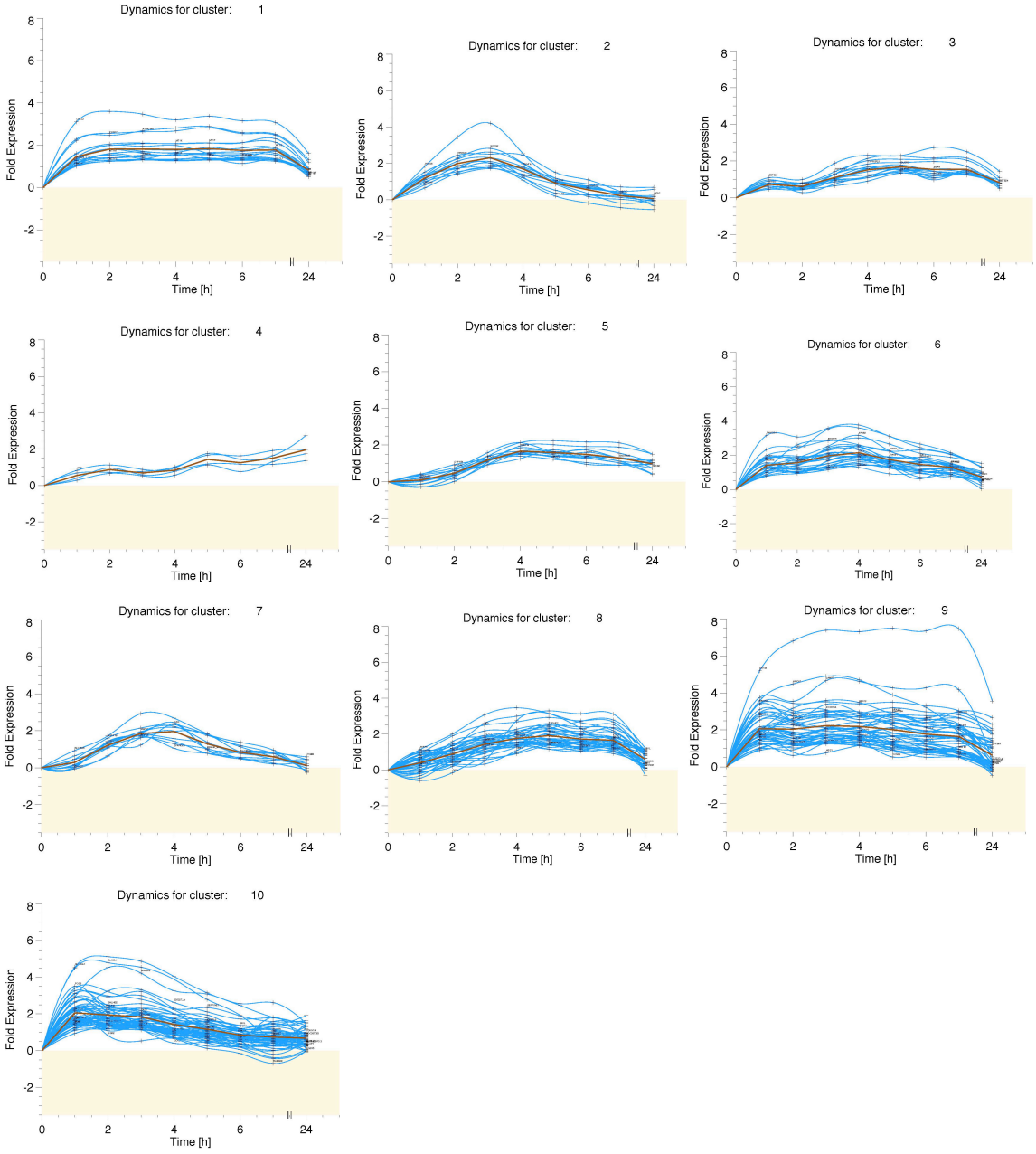
B. Downregulated

C. Outlier

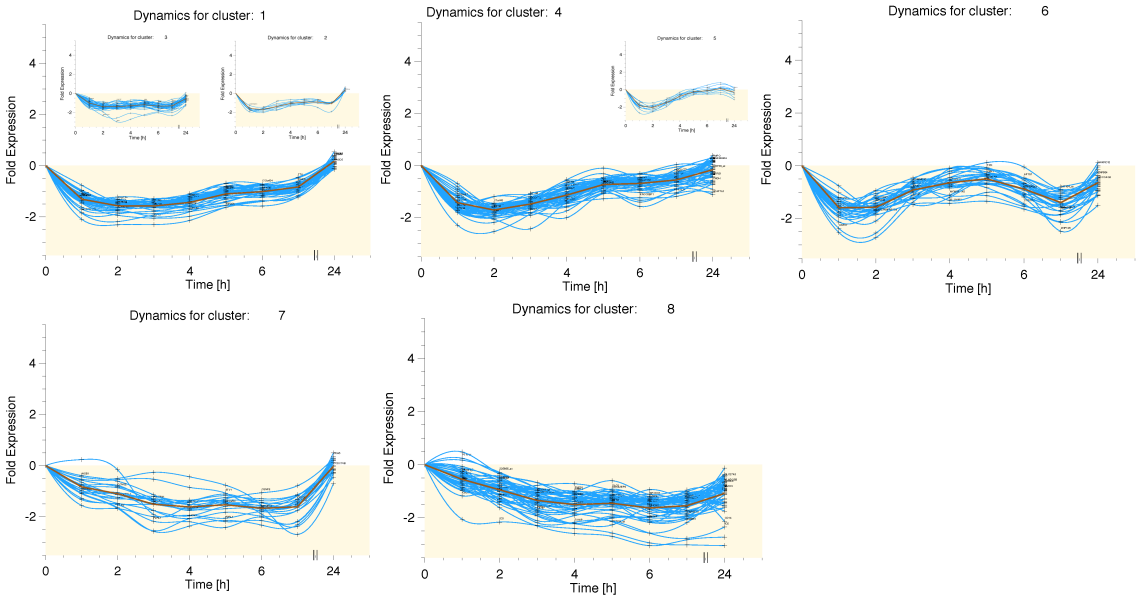
Figure 43 Experiment 2 tumor cell clustering of gene expression profiles

6.2.4.1.2 TC#3 (supernatant of stimulated PSC) clustering

Where there were only 3 upregulated and 2 downregulated clusters in experiment 2, now with stimulated PSC we see an organization of top genes into 10 significantly activated and 5 inhibited clusters (Figure 44).



A. Upregulated



B. Downregulated

Figure 44 Experiment 3 tumor cell clustering of gene expression profiles

6.2.5 Correlation between tumor cell experiments

In order to clearly define the differences between gene kinetics in the TC experiments we performed a correlation analysis using the *Correlate* function of IDL (see 5.6.4). Correlation analysis is the simplest way to quantitatively compare two separate data sets, and is especially useful to perform comparisons between control experiments. The analysis works by applying a specific distance metric to calculate the distance between the points of interest (also in a multidimensional space). The results as presented in Figure 45 clearly show that we have a very low correlation for most of the top ranked genes, with some of them highly anti-correlated (e.g. *MT1M, FAM129A*).

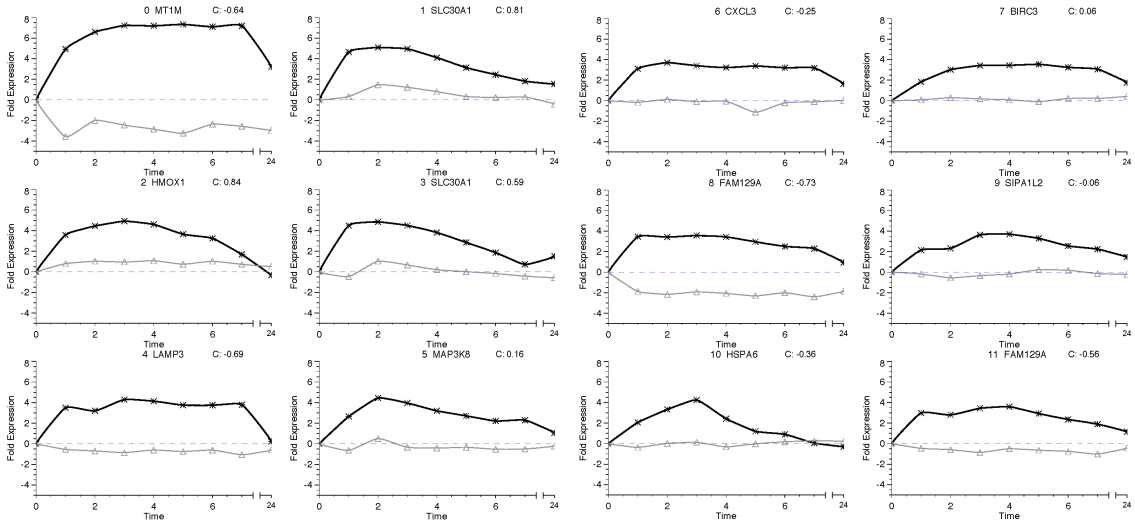


Figure 45 A sample of top ranked genes of experiment 3 (black) plotted against experiment 2 (blue) with the calculated correlation value (C).

Stimulus-dependent differences between TC become clearer as we consider the top ranking genes i.e. on one hand the most upregulated gene in TC#3 – metallothionein *MT1M* – is also the most downregulated gene in TC#2, on the other hand, the most downregulated gene in TC#3 (*ID3*) shows little to no response in TC#2. Many other genes of interest show similar differences, but few are as anti-correlated as *MT1M*. Correlation was performed on a cross-experimentally normalized data set without filtering to preserve genes, which show no variability over time and would otherwise be removed. In addition, there is not a single gene, in the top 250, that has at the same time, exactly the same expression level and exactly the same time-resolved behavior in both experiments.

6.2.6 Gene ontology analysis of TC experiments

GO analysis was performed as described in materials and methods (5.6.3), as well as in the results section for PSC#1 (6.1.5), with all parameters identical across all experiments to ensure comparability of data. TC data were organized as follows:

- Ranked gene lists of each experiment were divided into top 250 up- and 250 down-regulated genes, identified with Affymetrix IDs.
- Each data set was loaded into *David* and *WebGestalt* and converted into official gene symbols and Entrez ID's – annotation results were compared between the two to select the most complete.

Experiment 2 upregulated genes:

Out of 250 Affymetrix IDs, 178 were unambiguously mapped to Entrez IDs, 152 unique IDs were used for analysis.

Experiment 2 downregulated genes:

Out of 250 Affymetrix IDs, *WebGestalt* mapped only 57 Entrez IDs, a poor result of the *WebGestalt* conversion engine. An intermediate stage of conversion via *David* was performed. *David* successfully converted the same set to 227 official gene symbols, which then unambiguously mapped to 154 Entrez IDs, producing 142 unique identifiers usable for both *David* and *WebGestalt*.

Experiment 3 upregulated genes:

Out of 250 Affymetrix IDs, *WebGestalt* unambiguously mapped 190 to Entrez IDs, and 154 unique identifiers were used for the enrichment analysis.

Experiment 3 downregulated genes:

Out of a total of 250 Affymetrix IDs, *WebGestalt* unambiguously mapped 203 to Entrez IDs and 174 unique IDs were used for the enrichment analysis.

A direct comparison of the top 10 enriched GO terms in the Biological Process branch is shown in Table 12, accompanied by the results of the PSC#1 GO analysis.

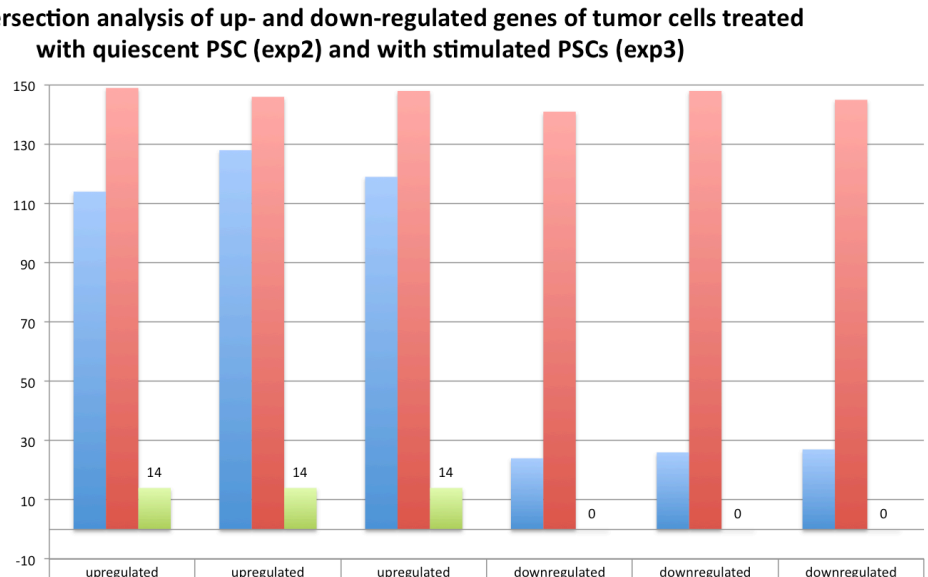
PSC exposed to TC	TC exposed to PSC	TC exposed to PSC*
Upregulated	**Upregulated**	**Upregulated**
Response to wounding	Negative regulation of viral reproduction	Response to chemical stimulus
Response to stress	Regulation of transcription factor activity	Response to external stimulus
Defense response	Negative regulation of viral genome replication	Response to organic substance
Response to external stimulus	Blood vessel maturation	Immune system process
Pos. reg. of transcription from RNA polymerase II promoter	Healing during inflammatory response	Response to stimulus
Positive reg. of macromolec. Biosynthetic process	TGFB receptor signaling pathway	Negative regulation of programmed cell death
Regulation of defense response	Regulation of transcription regulator activity	Negative regulation of apoptosis
Positive regulation of transcription, DNA-dependent	Response to oxidative stress	negative regulation of cell death
Positive reg. of nucleic acid metabolic process	Response to toxin	negative regulation of cellular process
Inflammatory response	Transcription from RNA polymerase II promoter	Negative regulation of biological process
Downregulated	**Downregulated**	**Downregulated**
cortical cytoskeleton organization	organelle organization	cellular component morphogenesis
cortical actin cytoskeleton organization	cellular macromolecule biosynthetic process	cell morphogenesis
membrane organization	macromolecule biosynthetic process	maintenance of sister chromatid cohesion
cellular component organization	chromatin organization	maintenance of mitotic sister chromatid cohesion
actin cytoskeleton organization	vesicle mediated transport	heart development
extracellular matrix organization	microtubule cytoskeleton organization	mitotic sister chromatid cohesion
actin filament-based process	regulation of macromolecule biosynthetic process	cellular component organization
membrane invagination	regulation of Rho protein signal transduction	cytoplasmic microtubule organization
endocytosis	cellular component organization	sister chromatid cohesion
regulation of cellular component organization	Rho protein signal transduction	anatomical structure development

Table 12 Biological process (top 10 enriched terms, with varying adjusted p-value ranges <<0.05)

The presented result shows a strong contrast between the two TC experiments clearly suggesting stimulus-dependent induction of different pathways.

6.2.6.1 Intersection of GO terms in TC experiments

Both TC experiments are significantly different once compared with ranking, clustering, and GO term enrichment. In order to answer a question of how many of the top-ranked genes in both TC experiments enrich the same GO terms, and how many of them overlap between experiments we performed an intersection analysis using *eGOn* (5.6.3). 4 reporter lists were used, 1 from each experiment containing top 250 upregulated, and 1 containing top 250 downregulated genes. Intersection results are presented in overview Figure 46, with more detailed enrichment presented for the biological process branch in Figure 47. GO analysis follows the branches down into finely-grained terms, and only the terms, which are significantly enriched by genes in both experiments, are considered, which explains why the total number of genes in the graphs never exceeds 150 (the remaining genes fall into incongruent categories).



	upregulated biological process	upregulated cellular component	upregulated molecular function	downregulated biological process	downregulated cellular component	downregulated molecular function
TC teated with quiescent PSC	114	128	119	24	26	27
TC treated with stimulated PSC	149	146	148	141	148	145
common genes in both experiments	14	14	14	0	0	0

Figure 46 GO intersection of experiments 2 and 3 tumor cells

The same GO terms are enriched by a very different number of genes (in both, up- and down-regulated state). The differences become even more apparent when we consider how many genes are identical for both experiments among those enriched terms – for the upregulated gene set only 14 genes overlap, for the downregulated – none.

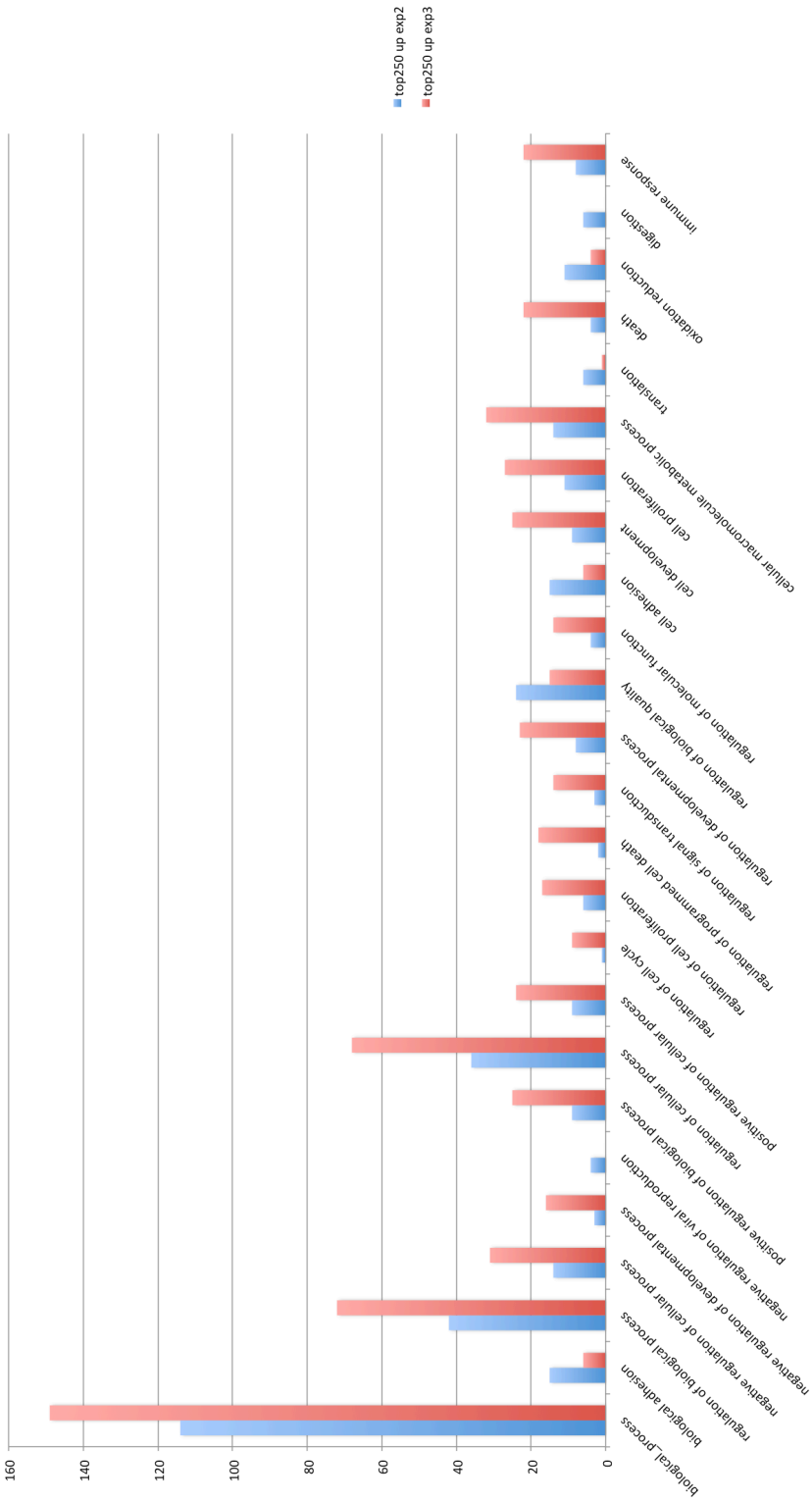


Figure 47 Sample of a detailed intersection analysis for only one branch of GO (Biological Process) between the upregulated genes in experiments 2 and 3.

6.2.7 Pathway analysis

Functional enrichment of pathways was analyzed using the *WebGestalt* with KEGG, Pathway Commons, and WikiPathways (5.6.5). The same gene lists as for GO analyses were used here. A sample of the outcome, in the form of a comparison between the top upregulated pathways in TC#2 and TC#3 is portrayed (Table 13).

TC#2 (supernatant of quiescent PSC)	
KEGG	Adj. p-value
PPAR signaling pathway	0.2871
mTOR signaling pathway	0.2871
Adipocytokine signaling	0.2871
Amyotrophic lateral sclerosis	0.2871
Neurotrophin signaling	0.2871
Pathogenic E.coli infection	0.2871
Aminoacyl-tRNA biosynth.	0.2871
Axon guidance	0.2871
TGF-beta signaling pathway	0.3436
MAPK signaling pathway	0.4016

TC#3 (supernatant of stimulated PSC)	
KEGG	Adj. p-value
Cytokine-cytokine receptor interaction	1.19e-07
MAPK signaling pathway	4.81e-06
NOD-like receptor signaling	6.71e-06
Chemokine signaling pathway	0.0009
Jak-STAT signaling pathway	0.0085
Toll-like receptor signaling	0.0085
Pathways in cancer	0.0085
Prostate cancer	0.0264
T cell receptor signaling pathway	0.0465
Complement and coagulation cascades	0.0552

Table 13 Comparison of top upregulated pathways from KEGG analysis of experiment 2 and 3

Interesting in this context is that out of all top 10 upregulated pathways in experiment 2 none meet the significance cut-off for adjusted p-values (5%), which suggests that a broad range of different pathways is getting activated in response to stimulus, and too few are significantly enriched by a large number of genes. This general deregulation of pathways in TC is a common theme in all tumors, but a comparison with experiment 3 shows that the latter are being pushed in a more directed manner towards survival with proinflammatory signals, and a wide-range of intercellular signaling.

6.2.8 GRN model of tumor cells

All previous analysis stages provided evidence of a clear treatment-dependent separation between TC, at the same time confirming what we saw in the first microarray experiment: Stellate cells are significantly altered when moving from the quiescent into the stimulated state, and this alteration has a strong feedback effect on TC. It is therefore no longer a question whether there is an effect of exposure in TC, but rather how this effect is achieved. We address this question by applying our reverse engineering to discover the underlying GRN responsible for signal integration and its propagation in TC. Tumor experiment 3 was chosen as the logical object for reverse engineering as we have proven it to be a continuation and the last stage of the hypothesized sequence of signaling from TC→PSC→TC.

6.2.8.1 Gene selection

Gene choice was dictated by BC results (see 6.2.4.1.2), and knowledge-driven interest of collaborating partners (Dr. N.Giese, 5.7.1 Materials and Methods). BC provided us with a division of the expression kinetics, and the manual gene selection from each cluster was based on the strength of gene overexpression, existing knowledge of its function, on its potential for novel regulation, and possible accessibility as targets of existing therapeutics, with a main focus on genes with transcriptional activity. The final selection provided a set of genes, which were united into a mean kinetic profile, a module, represented by a neuron in the neural network. Cluster 4 has been removed from the data set as it contained only members of the Complement System, which were of no interest at this point, and a total of 39 genes (Table 14) were selected from the remaining clusters.

Module	Selected genes (TC treated with stimulated PSC → Exp.3)				
1	CEBPD				
2	GRHL1	TRIM16			
3	PCGF5	DUSP2			
4	-				
5	ARNTL2				
6	AGA	FAM129A	PHF14	SAT1	WDR78
7	FGF18				
8	SPHK1	SQSTM1	RELB	BCL3	
9	ANK2	DUSP1	FOS	HES1	
10	EGR1	EGR2	EGR3	EGR4	ATF5
.	BHLHB2	FOSB	JUN	GADD45A	KLF10
.	TRIM36	ZEB1	ZFP36	SKIL	LMO4

Table 14 Gene selection for the GRN model of tumor cells

Each of the modules was investigated to identify potential interactions between them with GO, and pathway analysis. An evaluation of kinetic profiles allowed us to divide clusters into categories based on the sequence of their activation signals over time e.g. genes activated via paracrine or autocrine signals. While the gene content in each of the evaluated modules is significantly altered from the complete clusters, the gene expression dynamic over time is preserved (see 6.2.4.1.2 to review the expression profiles).

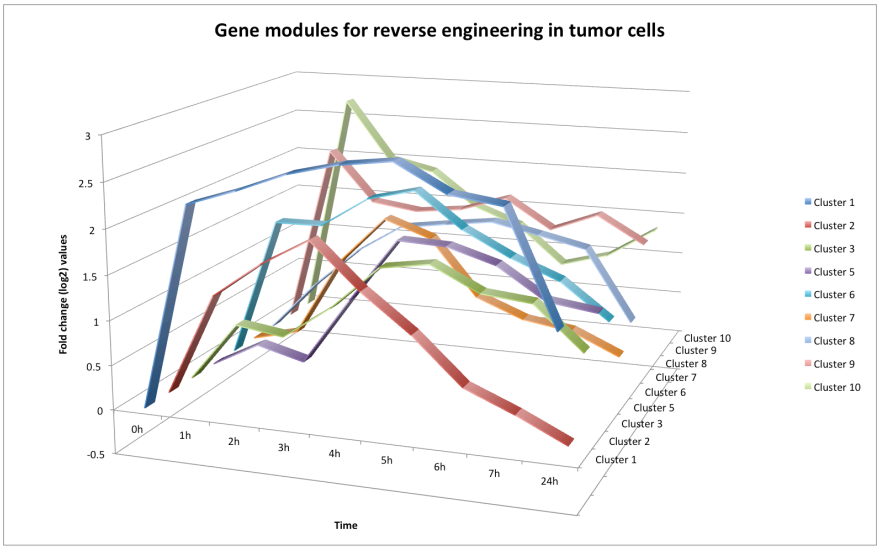


Figure 48 Gene expression kinetics among the top upregulated genes in tumor cells

- Module 1 genes show a sustained activity, beginning with an early upregulation and decreasing only after 24 hours, suggesting that either the initial paracrine stimulus is sufficient to drive the responses, or that a combined paracrine and autocrine signal is necessary to keep them in this state. *CEBPD* has been linked to cell fate determination, immune and inflammatory responses.
- Genes in modules 2 are transiently stimulated and judging from their expression kinetic can be categorized as intermediate response genes to an extracellular paracrine signal. Upregulation starts here at around 2h. Selected genes are related to cell growth and differentiation.
- Module 3 genes show two waves of signaling - initial almost simultaneous with the early responding genes within an hour, and then a secondary amplification at 2.5h or later. This suggests a cooperative upregulation via both paracrine, and autocrine signaling. *PCGF3* gene has been implicated in chromatin remodeling; *DUSP2* is involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development.
- Module 5 genes are upregulated after 2.5h (either induced delayed primary, or secondary response to PSC stimulus), and remains upregulated throughout the experiment, suggesting additional autocrine sustaining. *ARNTL2* was linked to circadian clock and regulation of cell proliferation.
- Module 6 genes show upregulation at all signaling stages, most likely due to the changes in the microenvironment occurring from both paracrine and autocrine signaling.
- Module 7 genes are only transiently upregulated after 2.5h. *FGF18* has broad mitogenic and cell survival activities.
- Module 8 genes are significantly stimulated and judging from their expression kinetic are either primary delayed or secondary response to paracrine stimulus with earliest upregulation starting already within 2h, and an amplification of signal in an autocrine manner past 4h. Selected genes are related to development and growth.
- Module 9 genes are continuously upregulated, and include immune response, response to stimulus, and regulation of cell proliferation, differentiation, and apoptosis.
- Module 10 shows the earliest response to paracrine stimulus. *EGR* family, *FosB, ATF5, KLF10, TGIF*. Cell fate determination, growth factor stimulation.

Cluster	Gene	Function
1	CEBPD	Enhancer, implicated in cell cycle control and/or cell fate determination. C/EBP family members recognize similar DNA sequences in their target genes and form homo- or hetero-dimers with other C/EBPs, as well as with transcription factors of the NF-κB and Fos/Jun families. Important in regulation of genes involved in immune and inflammatory responses, may be involved in the regulation of genes associated with activation and/or differentiation of macrophages. (provided by RefSeq)
2	GRHL1	TF, which may play a role in development. May be involved in epidermal differentiation (By similarity).
2	TRIM16	Regulator of cell growth, a transcriptional regulator in the retinoid signaling. The estrogen-responsive B box protein (EBBP) restores retinoid sensitivity in retinoid-resistant cancer cells via effects on histone acetylation.
3	PCGF5	Part of the polycomb family of proteins, which can remodel chromatin such that transcription factors cannot bind to promoter sequences in DNA.
3	DUSP2	A phosphatase inactivating ERK1 and ERK2 - a negative regulator of MAPK activity. In the massive signaling upon cell stimulation this phosphatase is highly upregulated to recycle the kinases of the MAP signaling pathway. It is involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development.
4	-	Gene cluster 4 contains immune response-related proteins such as complement members CFB, C3, CHI3L2, and has been removed from the selection.
5	ARNTL2	Transcriptional regulator and signal transducer, functionally linked to circadian clock-mediated activities and to the regulation of cell proliferation
6	AGA	Involved in the catabolism of N-linked oligosaccharides of glycoproteins.
6	FAM129A	Regulates phosphorylation of a number of proteins involved in translation regulation including EIF2A, EIF4EBP1 and RPS6KB1. May be involved in stress response.
6	PHF14	Unknown function.
6	SAT1	Acetyltransferase, a rate-limiting enzyme in the catabolic pathway of polyamine metabolism. It catalyzes the acetylation of spermidine and spermine, and is involved in the regulation of the intracellular concentration of polyamines and their transport out of cells.
6	WDR78	Unknown function.
7	FGF18	Fibroblast growth factor (FGF) family member with broad mitogenic and cell survival activities. Involved in a variety of biological processes, including embryonic development, cell growth, morphogenesis, tissue repair, tumor growth, and invasion. It is a pleiotropic growth factor that stimulates proliferation in a number of tissues, most notably the liver and small intestine
8	SPHK1	Sphingosine kinase 1, mediates VEGF induced activation of Ras and MAPK. SPHK1 governs the subtle balance between the sphingolipids ceramide and sphingosine 1-phosphate (S1P) levels. Both of which are key regulators of cell death and proliferation. SPHK1 is an oncogene, overexpressed in many tumors, protects cancer cells from apoptosis.
8	SQSTM1	A multifunctional protein that binds ubiquitin and regulates activation of NF-kB signaling pathway. Functions as a scaffolding/adaptor protein in concert with TNF receptor-associated factor 6 to mediate activation of NF-kB in response to upstream signals. Mutations result in sporadic and familial Paget disease of bone. (provided by RefSeq)
8	RELB	interacts with NF-κ-B 1 and 2 pleiotropic transcription factors, involved in inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis. NF-κ-B is a homo- or heterodimeric complex formed by the Rel-like domain-containing proteins RELA/p65, RELB, NFKB1/p105, NFKB1/p50, REL and NFKB2/p52. It is controlled by various mechanisms of post-translational modification and subcellular compartmentalization as well as by interactions with other cofactors or corepressors. RelB-p50 and RelB-p52 complexes are transcriptional activators. RELB neither associates with DNA nor with RELA/p65 or REL. Stimulates promoter activity in the presence of NFKB2/p49
8	BCL3	A proto-oncogene candidate. Protein encoded by this gene contains seven ankyrin repeats, which are most closely related to those found in I κ B proteins. This protein functions as a transcriptional co-activator that activates through its association with NF-κ B homodimers. The expression of this gene can be induced by NF-κ B, which forms a part of the autoregulatory loop that controls the nuclear residence of p50 NF-κ B. (provided by RefSeq)
9	ANK	Ankyrins family of proteins link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play key roles in activities such as cell motility, activation, proliferation, contact and maintenance of specialized membrane domains. (provided by RefSeq)
9	DUSP1	DUSP1 gene is induced in human skin fibroblasts by oxidative/heat stress and growth factors. The bacterially expressed and purified DUSP1 protein has intrinsic phosphatase activity, and specifically inactivates mitogen-activated protein (MAP) kinase in vitro. Furthermore, it suppresses the activation of MAP kinase by oncogenic ras in extracts of Xenopus oocytes. Thus, DUSP1 may play an important role in the human cellular response to environmental stress as well as in the negative regulation of cellular proliferation. (provided by RefSeq)
9	FOS	The Fos gene family consists of 4 members: FOS, FOSB, FOSL1, and FOSL2. These genes encode leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1. As such, the FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. In some cases, expression of the FOS gene has also been associated with apoptotic cell death. (provided by RefSeq)
9	HES1	Basic helix-loop-helix family of transcription factors. It is a transcriptional repressor of genes that require a bHLH protein for their transcription.
10	EGR1	EGR family of C2H2-type zinc-finger proteins. It is a nuclear protein and functions as a transcriptional regulator. The products of target genes it activates are required for differentitation and mitogenesis.
10	EGR2	Defects in this gene are associated with Charcot-Marie-Tooth disease type 1D (CMT1D), Charcot-Marie-Tooth disease type 4E (CMT4E), and with Dejerine-Sottas syndrome (DSS). Multiple transcript variants encoding two different isoforms have been found for this gene.
10	EGR3	An immediate-early growth response gene, which is induced by mitogenic stimulation, participates in the transcriptional regulation of genes in controlling biological rhythm.
10	EGR4	Transcriptional regulator. Recognizes and binds to the DNA sequence 5'-GCGGGGGCG-3' (GSG). Activates the transcription of target genes whose products are required for mitogenesis and differentiation (By similarity)

Table 15 Entrez summary for model genes in tumor cells

6.2.8.2 Model construction

Clear-cut expression kinetics in PSC experiment 1 produced a likewise streamlined gene regulatory model (Figure 35); such straightforward gene selection was not possible for TC kinetics, due to signal ambiguity with many additional shifts and delays in gene kinetics, additionally complicated by a strong overrepresentation of transcription factors in kinetic profile 10. Therefore, in a multi-stage approach, first, a TC CTRNN model was constructed using a metagene, where a mean-kinetic of each cluster was taken and the resulting system was used to fine-tune parameters and input functions of the neural network (5.7). Subsequently 4 TC models were constructed, which differed in module 10 gene content (Table 16). Network topology was preserved in general across all models with minor deviations (<10 altered weights on a 9x9 summarized weight matrix) resulting from the variations in selection-dependent module kinetics, which modulated interaction weights sufficiently to cross the noise level and become either effective (connection gain) or ineffective (connection loss) (as defined by the offset parameter cf. Materials and methods).

The most connected network of effective weights was achieved with Model 2 containing a total of 24 genes (Table 16, cluster 10 *EGR*-only). All resulting models showed a fit to the experimental data of at least 90%. The identified gene interactions were plotted based on the summarized effective weights in Figure 55.

TC Models	Metagene	Model 1	Model 2	Model 3	Model 4
Genes from (BC) kinetic profiles	9 (1 mean per cluster)	35 (complete preselection)	24 (EGR-only cluster 10)	31 (cluster 10 without EGR)	21 (randomized selection)

Table 16 TC evaluation of gene content in network modules during GRN model construction

6.2.8.3 Model parameters

Parameters were initialized randomly and then evolved over 2000 generations with a genetic algorithm. Following parameters were used to build each model of tumor cell GRN:

Model parameters:	
Number of generations	2000
Population size	700
Runs	1500
Mutation rate	0.01
Euler integration parameters:	
Time delay	15
Time step size	0.05
Interpolation points resulting from time step size	480
CTRNN parameters:	
Initialization of parameters	Randomized
Sigmoidal activation function interpolation points	1000
Range of interaction weights	10
Offset range	2.6
Decay range	2.4
Minimal time constant of decay	0.1
Delay range	20
Range of initial input	15
Fitness points (how many interpolation points are evaluated with the mean square error function)	480
Second input beginning	50 interpol. points (=2.5 hours)
Period of the second input	5 minutes
Amplitude of the second input relative to the learned first	0.1
Input halflife in minutes	0.5
Tension of the spline interpolation of the data	1.0 (<<1 cubic, >>1 linear)

Table 17 TC GRN model parameters

6.2.8.4 Fitness evaluation

Fitness was evaluated with the MSE function for the 140 as well as 480 interpolated points in the time series for the same reason as it was the case for the PSC model (6.1.7.3).

The dashed line represents the experimental values, black lines are the modeled solutions and dashed-and-dotted lines represent the external input functions. Fitting for all 4 models was successful with a rate of >96% fit to experimental data (Figure 49).

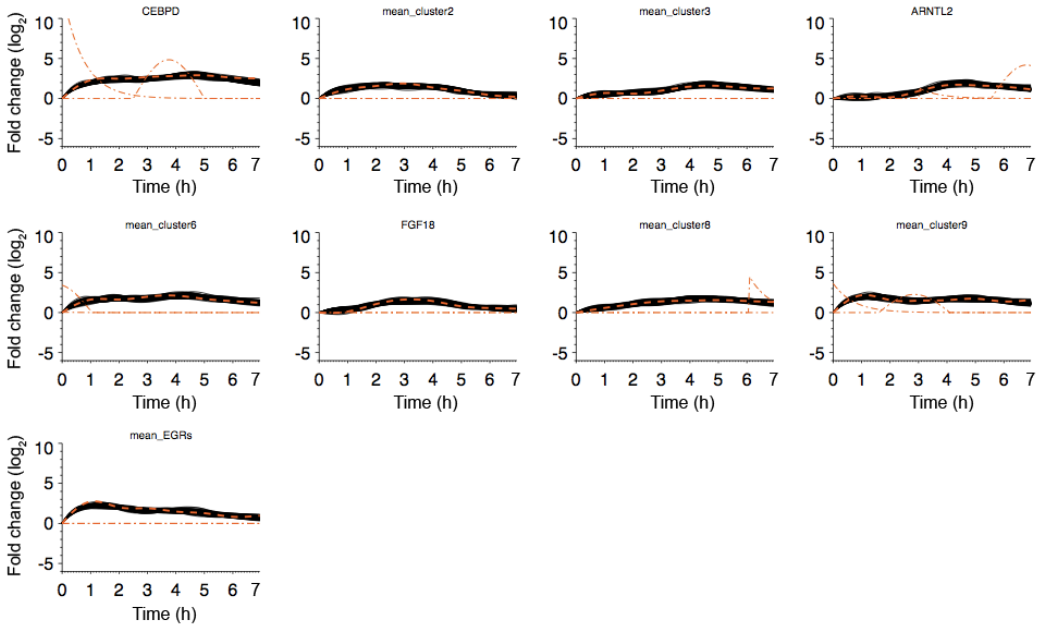


Figure 49 High fitness of model to experimental data has been achieved

6.2.8.5 Selection-driven model evaluation

Interaction networks produced from each of the models' weight matrices provide a visual comparison (Figure 50-54). The most complete set of interactions is achieved with model 2 (Figure 51). Green arrows represent inhibition, red activation. In each of the sub-models, the unchanging interactions are either black or gray (inhibition or activation), while lost interactions are plotted in blue (regardless of what type of an interaction it was).

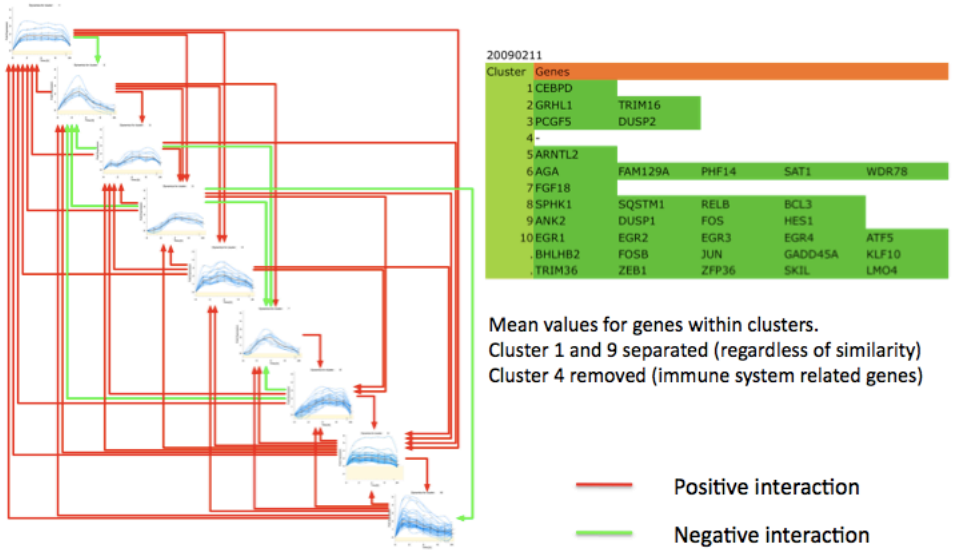


Figure 50 Complete gene selection for reverse engineering of the TC gene regulatory network

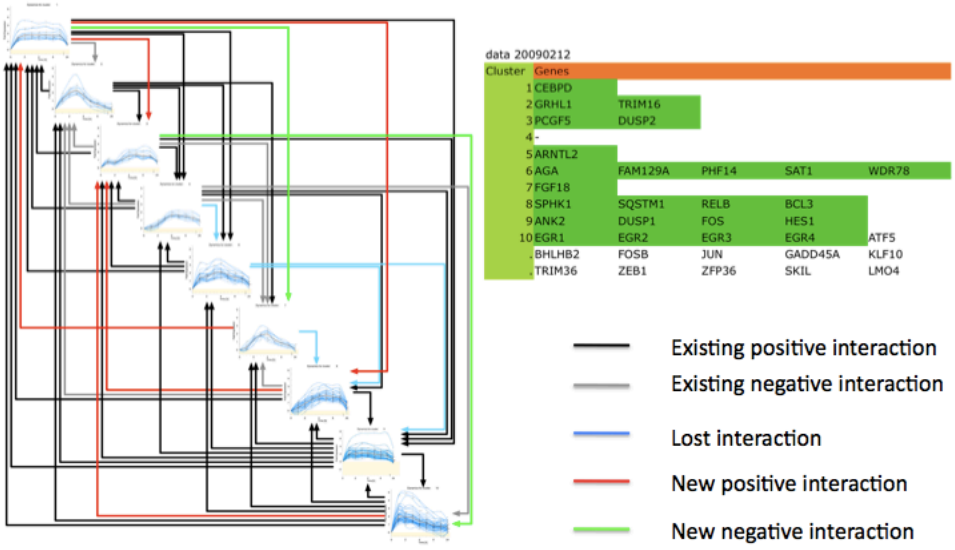


Figure 51 Evaluation of interactions with an EGR-only gene cluster 10

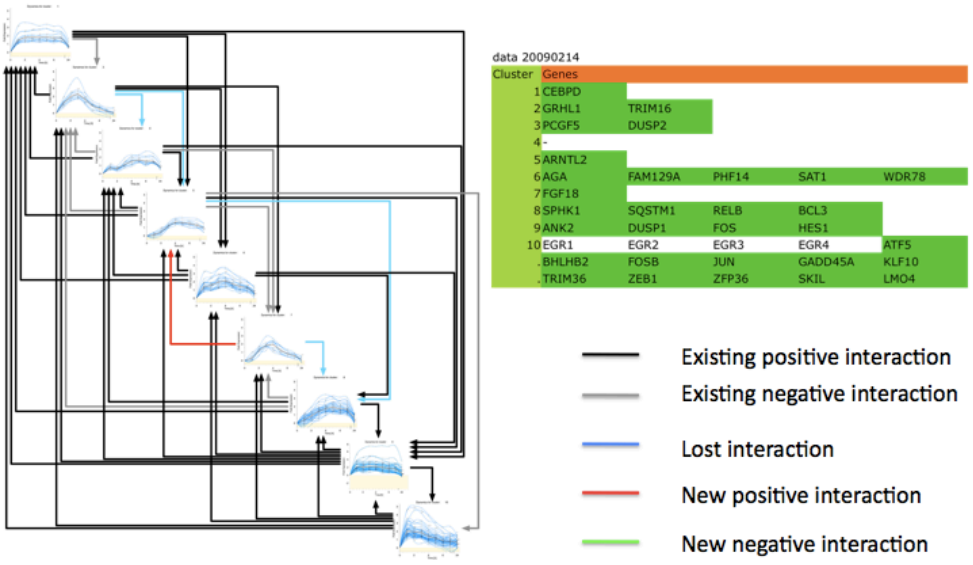


Figure 52 Evaluation of interactions with all genes except EGR's in cluster 10

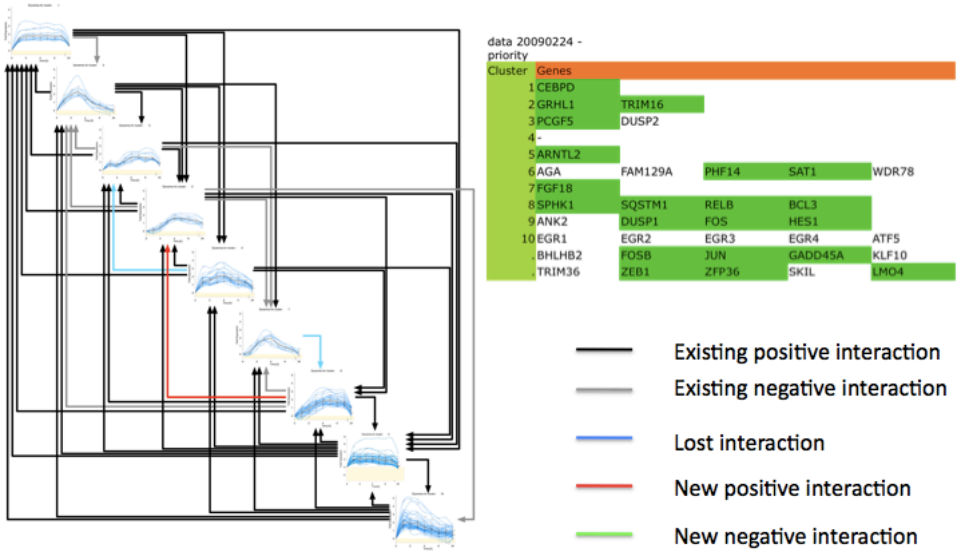


Figure 53 Randomized gene selection

6.2.8.6 Interaction matrix and the resulting network visualization

Model 2 genes (Table 18, Figure 54) formed the modules in the TC#3 GRN, interactions reverse engineered with CTRNN were converted from a matrix of effective weights into an interaction network (Figure 55).

Cluster	Selected genes (TC treated with stimulated PSC → Exp.3)				
1	CEBPD				
2	GRHL1	TRIM16			
3	PCGF5	DUSP2			
4	-				
5	ARNTL2				
6	AGA	FAM129A	PHF14	SAT1	WDR78
7	FGF18				
8	SPHK1	SQSTM1	RELB	BCL3	
9	ANK2	DUSP1	FOS	HES1	
10	EGR1	EGR2	EGR3	EGR4	

Table 18 Tumor cell model gene selection

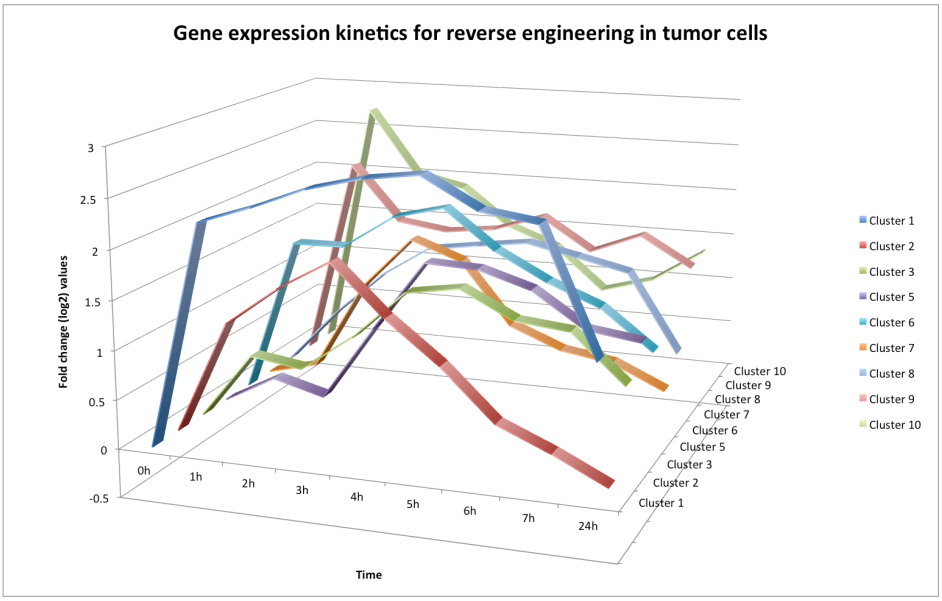


Figure 54 Gene expression kinetics used for the final TC model

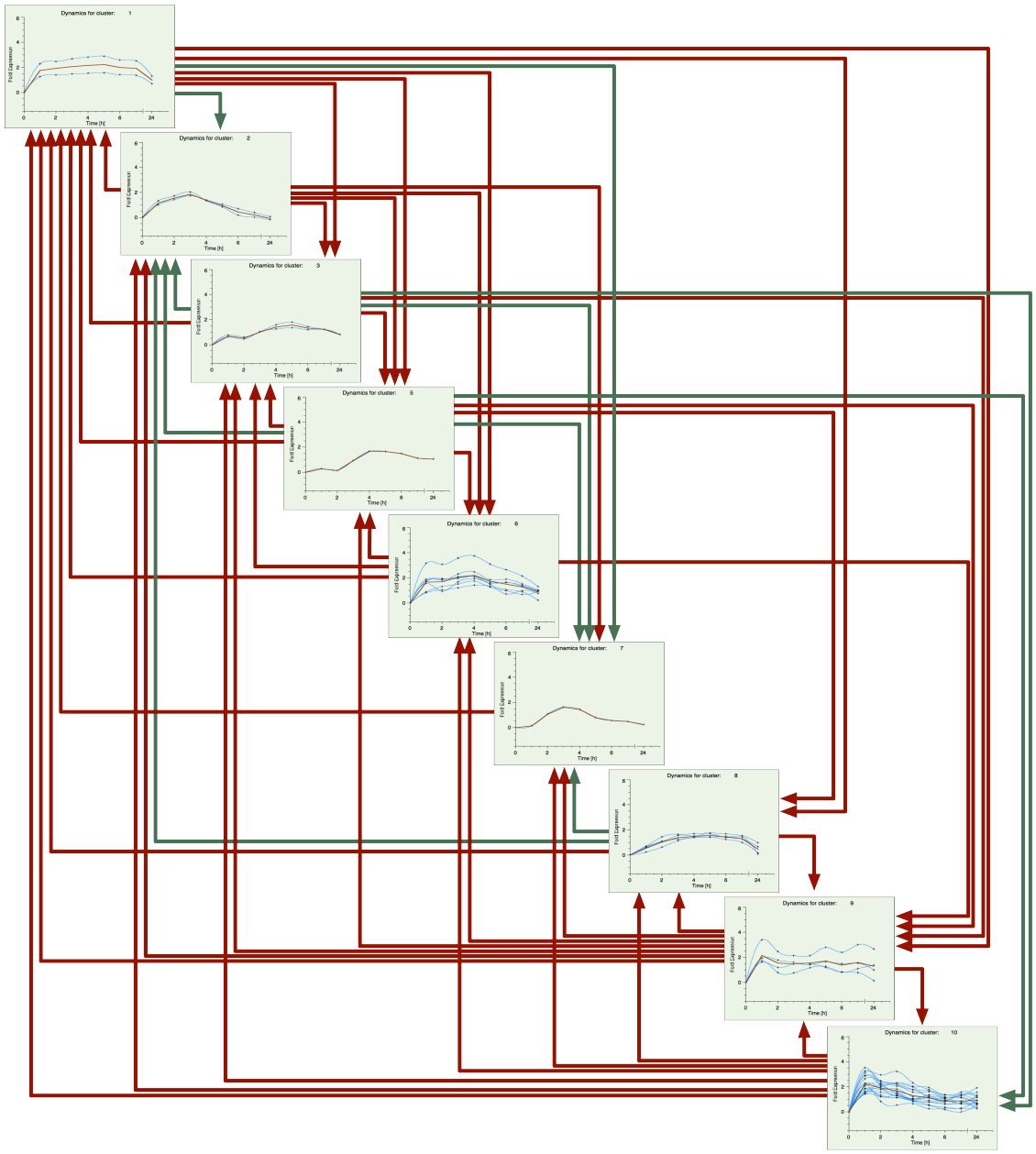


Figure 55 Interaction network (red for activation, and green for inhibition). Only effective weights are plotted accounting for the offset parameter value corresponding to the identified system noise

As gene clustering analysis (6.2.4.1) and model show (Figure 55), unlike in stellate cells, where the main role of the early response genes is to receive, integrate and divert the initial stimulus into a single central hub, which passes it to its downstream targets, in TC, the early response genes (module 10 EGRs) seem to be both, the recipient of the initiating stimulus, and the main response hub, as they transduce the signaling to downstream targets, and receive only a very limited feedback from them. The first targets of that signaling are primary delayed and secondary response genes, which GO describes as responsible for cell-cell communication, regulation of growth and proliferation, as well as metabolic processes, cellular processes and response to stress and stimulus. These modules feed the signaling further towards the intermediate genes (transition module containing genes responsible for signal transduction, regulation of gene expression and ion binding). Finally a module of late genes seems to integrate not only the direct initial stimulus from the early genes, but is most likely amplified by the autocrine feedback in TC (immune response, response to stimulus, regulation of proliferation, differentiation and apoptosis). The result of TC treatment with stimulated PSC is a response, which can be defined as a cytokine profile, that likely acts *in vivo* in an autocrine manner on TC as well as a perpetuating stimulus on PSC. Although we do see genes upregulated past 24 hours e.g. cluster 4 complement members (removed from the final model), the remaining clusters show a decline at the 24-hour time point. This overall lack of a permanent switch in TC may mean that the interactions with PSC are one of many signals necessary to drive them in the direction of a cell-fate decision.

6.2.8.7 Model system selection using LLE

Largest Lyapunov Exponent was used to identify systems, which show the most robust response to perturbation (see section 5.7.3 and Supplementary materials of Busch et al. 2008). The final systems used for *in silico* simulations and knockdowns were selected by ranking them using both the robustness and fitness criterion at the same time, and taking those that rank the highest on both scales. The LLE method is described in more detail in our paper, Busch et al. (2008).

6.2.8.8 *In silico* simulations

6.2.8.8.1 Inputs

Extracellular input evaluation was performed to establish whether tumor cell GRN is capable of retaining its state in the absence of supportive secondary inputs. At the modeling stage we clearly predefine the inputs to account for the *in vitro* situation, where the initial signal is of paracrine nature coming from stellate cells, in our case from supernatant stimulation, and secondary signal is the autocrine feedback produced by TC. Mathematically they are defined as an exponentially decaying function (Figure 56 A) and a sine function (Figure 56 B) respectively. Simulations allow us to alter those functions on the fly to evaluate the system behavior in changing conditions. The vital parameters for those functions are their amplitudes (derived from the neural network model, evolved over many generations with the genetic algorithm), and input decays (either evolved or preset). Results of input modifications show that to achieve an optimal fit of model to the experiment a decaying input mimicking the paracrine signal has to be trailed by a continuous secondary (periodic) function. This suggests that the tumor-stellate cell interactions are forming a perpetuating cycle and are not of switch-like nature.

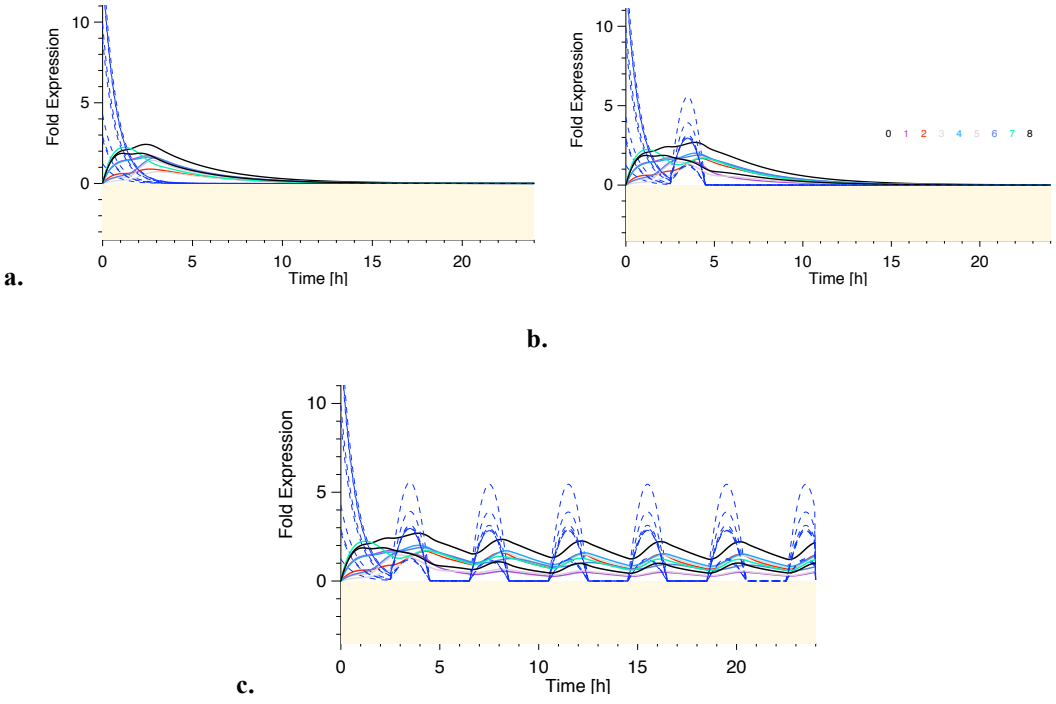


Figure 56 Evaluating input functions allows us to achieve the optimal fit of tumor cell model to the experimental data: A. Single exponentially decaying input; B. Two inputs; C. Continuous secondary input

6.2.8.8.2 TC Knockdowns

Knockdown screen was performed for all modules in the network. In general TC knockdowns have proven less effective than PSC, mostly due to ambiguous gene kinetics, its greater signaling complexity, and as a result greater stability and overall redundancy.

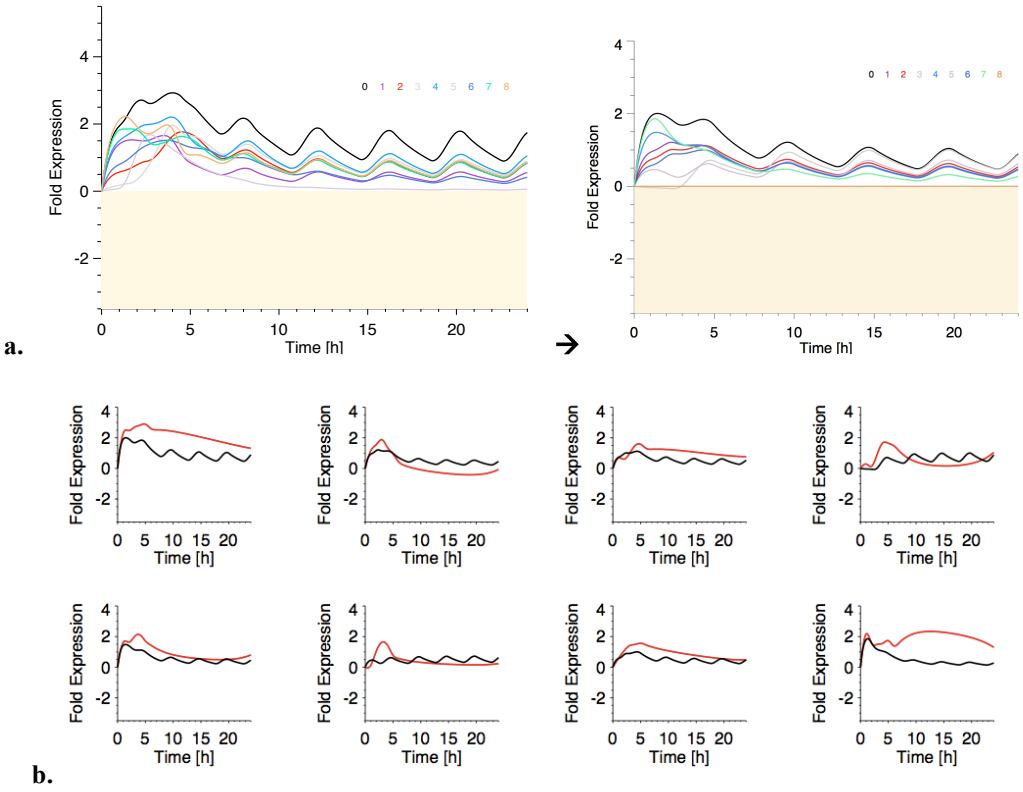


Figure 57 A. Simulation of the native system without knockdowns vs. the knockdown of cluster 10 genes. B. Effect of the cluster 10 knockdown on the remaining genes in the network (red: experimental interpolated data, black: knockdown simulation).

TC GRN knockdowns have revealed module 10 (*EGR1-4*) as the weakest point in the network (Figure 57) of the final model. Knocking it out resulted in the strongest single-module knockdown effect, affecting all the other modules in the network.

All earlier models were also investigated in addition to the final set, and across all of them, knockdowns of modules 1 (*CEBPD*) and 9 (*Fos, Hes1, Ank2, Dusp1*) have had a measurable impact on the system, however depending on the content of the 10th cluster (presence or lack of *EGRs*) the changes were more or less pronounced, in all cases however the system recovered after 24h if the extracellular input (periodic) was preserved. However in case of models without EGRs in module 10, its knockdown had no effect.

The earliest model built with the *metagene* approach, has shown results confirming Model 1 (complete 35 gene selection, see Table 16), assuring the earlier assumption of the *metagene* being an effective, albeit averaged representative of the network modules.

Interestingly, we were able to show that it is possible to achieve a system-wide breakdown of control in TC, by performing a double-node knockdown of *EGR1-4* and *CEBPD* genes (module 10 and 1 respectively) (Figure 58). The genes are 'freed' from constraints of inhibition and activation within the network. This finding while interesting is not usable in practical terms for two reasons: first, we are attempting to break down a network by downregulating its responses and not actually upregulating them, as is visible for some of the genes in that knockdown; second, a multi-module knockdown is not feasible *in vivo* due to its high toxicity.

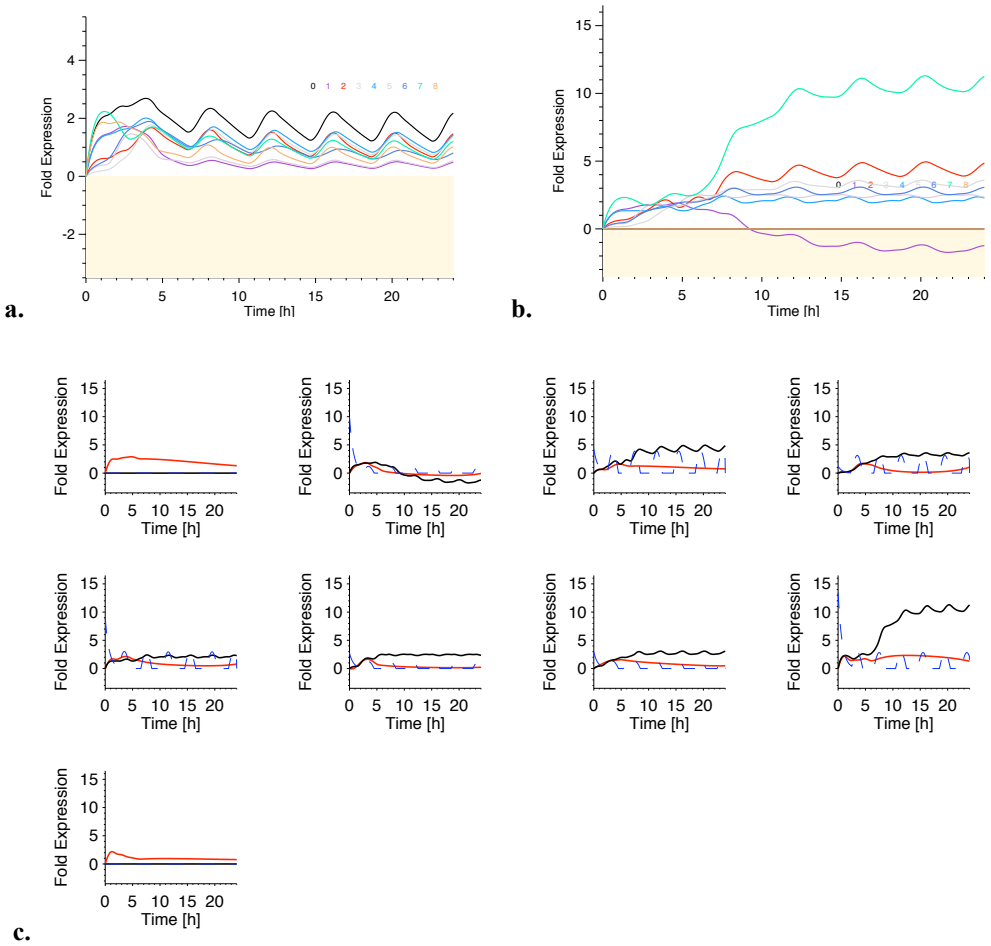


Figure 58 Effect of the *in silico* knockdown of clusters 1 & 10 on the remaining regulatory network. A. native simulation, B. knockdown simulation, C. knockdown simulation for each of the network modules (black: simulated, red: experimental)

Since this correlation of *CEBPD* and *EGR* effect is rather striking, we investigated the known interactions between the *EGR* and the *CEBP* gene families using pathway analysis (Figure 59). The results have confirmed our suspicions of interdependency between those genes. Especially interesting in this context is the work of Carro et al., who have investigated C/EBPβ and linked it to the transcriptional network for mesenchymal transformation (EMT) of brain tumors (Carro et al. 2010) suggesting that there may be potential for cell-fate decisions between these two gene sets.

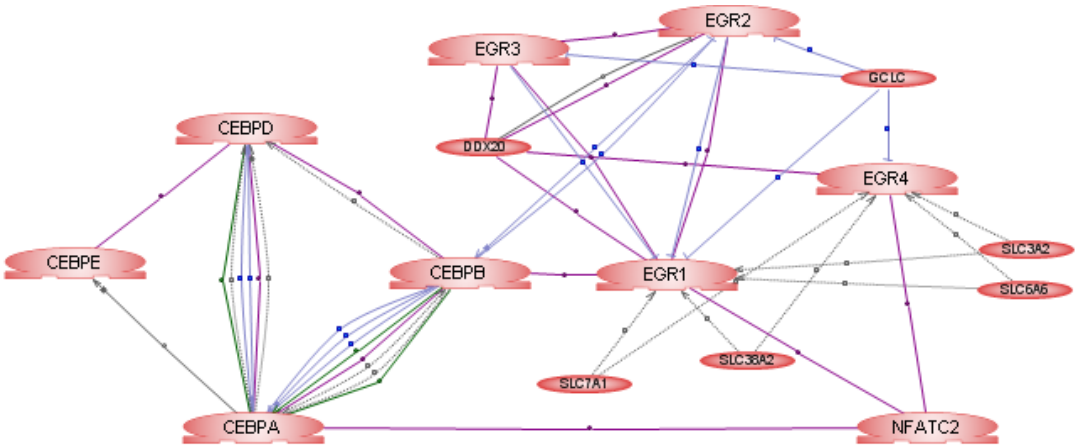


Figure 59 Known (literature-derived) links between *EGR* and *CEBPD* gene families are unveiled via the direct interactions of *EGR1/2* with *CEBPB* and indirect interactions with other genes of both families

Armed with the modeling and simulation results the *EGR* genes became a likely target for *in vitro* validations, however the specificity of those genes had to be evaluated first using our TC#2 control experiment.

6.2.9 Significance and specificity of EGR knockdowns in tumor cells

Selection of *EGRs* as targets was shaped by the outcome of the *in silico* knockdown in TC. Several questions arose as a result: How specific is the *EGR* knockdown? Are the *EGRs* equally as important when quiescent PSC stimulate TC, as they are with stimulated PSC treatment? If they are, then gene selection is invalid, because the changes in expression of *EGRs* are not related to the specific form of stimulus, but are an unspecific reaction, which may be induced by a random stimulus.

6.2.9.1 Control model of tumor cells

GRN model was reconstructed using the control TC dataset (supernatant of quiescent PSC). Preserved parameter ranges and gene selection forced gene re-clustering (BC) to account for significantly different kinetics between the two experiments. Clustering results were altered from the original data set as genes which previously clustered together were separated e.g. cluster 10 was divided into sub-clusters with varying expression profiles (*EGR* genes Figure 60).

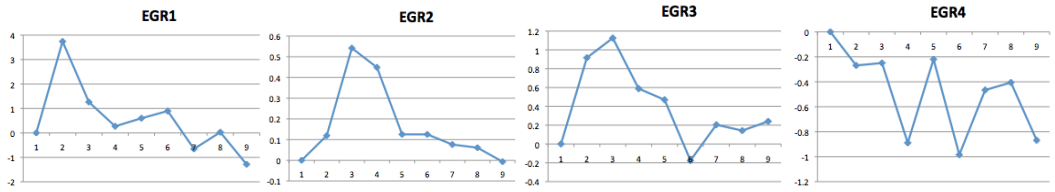
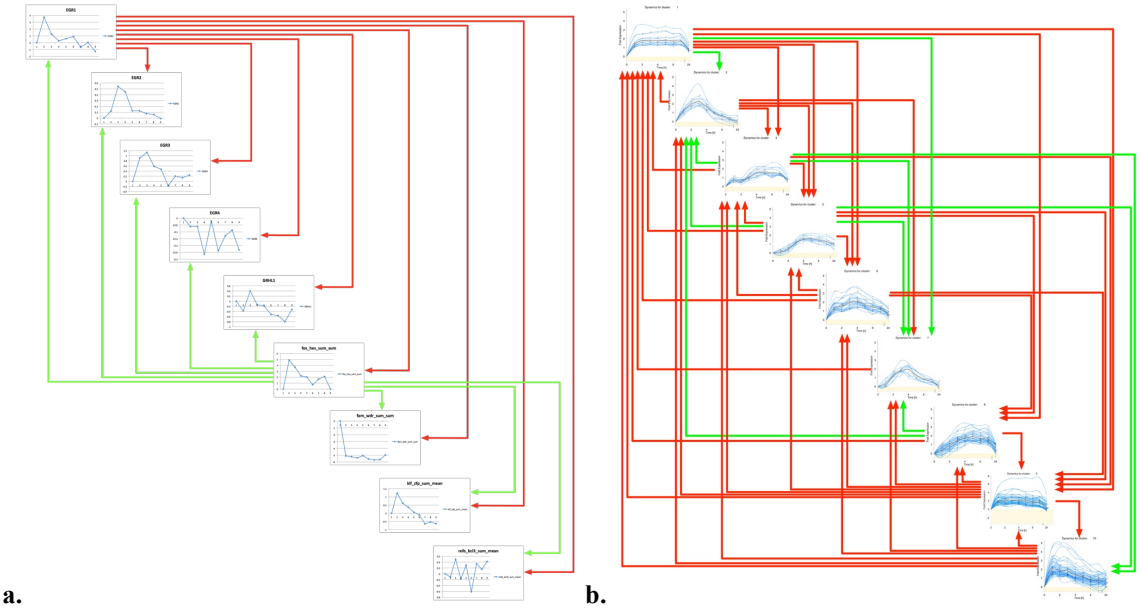


Figure 60 TC#2 *EGR* gene kinetics are significantly altered from TC#3 experiment, notice the different scales

Genes showed often reversed, and always different time-resolved expression profiles. Additionally, expression levels for nearly all genes were also drastically different (usually significantly decreased as shown in 6.2.5). Organizing those genes into the same modules as in experiment 3, without additional evaluation made no biological sense, as the underlying idea of clustering was to identify genes, which may act in a synergistic manner, therefore only non-redundant gene clusters were used. Effectively we retained a set of genes representing unique original clusters **2** (*GRHL1*), **6** (*FAM129A, WDR78*), **9** (*FOS, HES1*), and separated cluster **10** into **4**, one for each *EGR*. Producing the final set of **network modules**: *EGR1, EGR2, EGR3, EGR4, GRLH1, FOS/HES1, FAM129A/WDR78, KLF10/ZFP36, RELB/BCL3.*

6.2.9.2 Interaction networks

Resulting model weight matrix was translated into an interaction network of effective weights (Figure 61.A). Only two effectively connected modules were identified in this sparsely connected network including: *EGR1*, and *FOS/HES1*. As expected this result was in strong disparity with TC#3 effective weights interaction network (Figure 61.B).



a. **b.**

Figure 61 Comparison of TC interaction networks (effective weights) in experiment 2 (a) and experiment 3 (b)

6.2.9.3 *In silico* simulations and knockdowns in TC control experiment

Simulations were performed with the same inputs as defined for experiment 3. Unlike experiment 3, where all genes in the network were up-regulated, here we had a strong presence of downregulated modules, therefore the input functions were allowed to evolve into both: negative as well as positive values (accounting for an inhibitory effect of the extracellular stimulus).

EGR1 silencing compared to the full system simulation (Figure 62) showed no effect (minor changes in *EGR2* and *EGR3* gene kinetics). *FOS/HES1* effect was much stronger as genes, which were downregulated due to the action of this cluster, upon its knockdown became significantly upregulated.

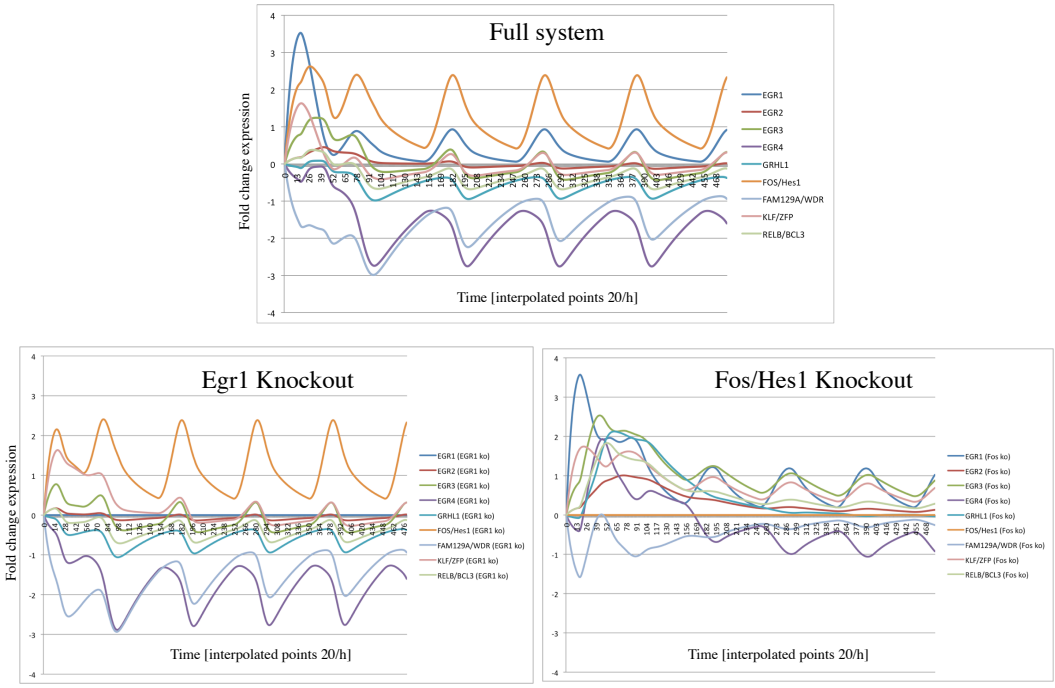


Figure 62 Simulations and knockdown evaluation of tumor cell control experiment

In silico knockdowns showed lack of *EGR* effect on network topology, confirming cell and condition specificity of *EGR* genes identified in TC#3.

6.2.10 Integration of the stellate and tumor cell models

To verify resolved interactions in the TC#3 GRN in the context of actual PSC microarray data, an integration of signaling was performed by replacing the mathematically simulated external stimuli (exponentially decaying function mimicking supernatant stimulation, and a secondary sustaining sine function corresponding to autocrine feedback loops and secondary paracrine stimulation) with an actual signature of secreted factors from PSC (Table 9). That signature was constructed by taking spline interpolated mean of all involved expression kinetics, and a 2.5h delay was added to account for the time necessary for the cytokines to be produced, secreted and accumulated in the extracellular space before they become available to TC (Figure 63). CTRNN parameters were preserved and input amplitude was allowed to evolve in the network.

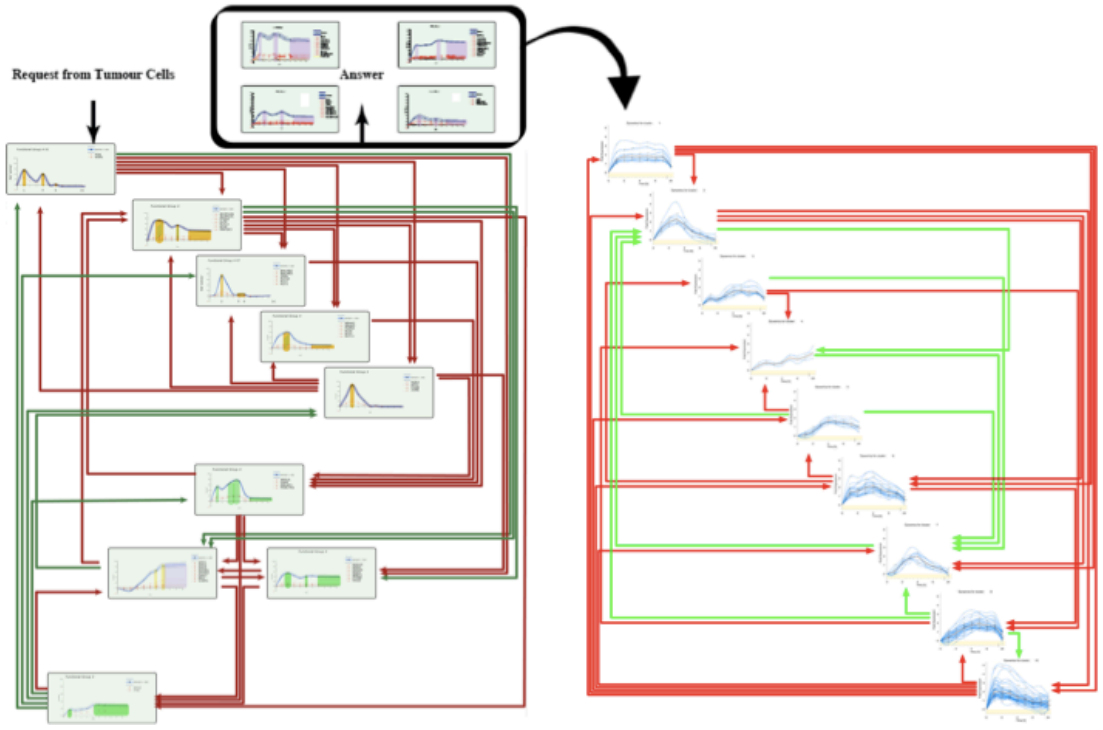


Figure 63 Schematic representation of model integration via the secreted protein gene expression profile of stimulated stellate cells acting on tumor cells.

Data used here were cross-experimentally re-normalized, and re-clustered adding cluster 11 (*SKIL, TRIM36 and ZEB1*), which showed an altered expression profile, and removing *FGF18*, which was lost in the cross-experimental filtering.

Resulting model showed a 95% fit to the experimental data (Figure 64) and preserved topology for a core set of interactions including the *EGR* gene family (Figure 65).

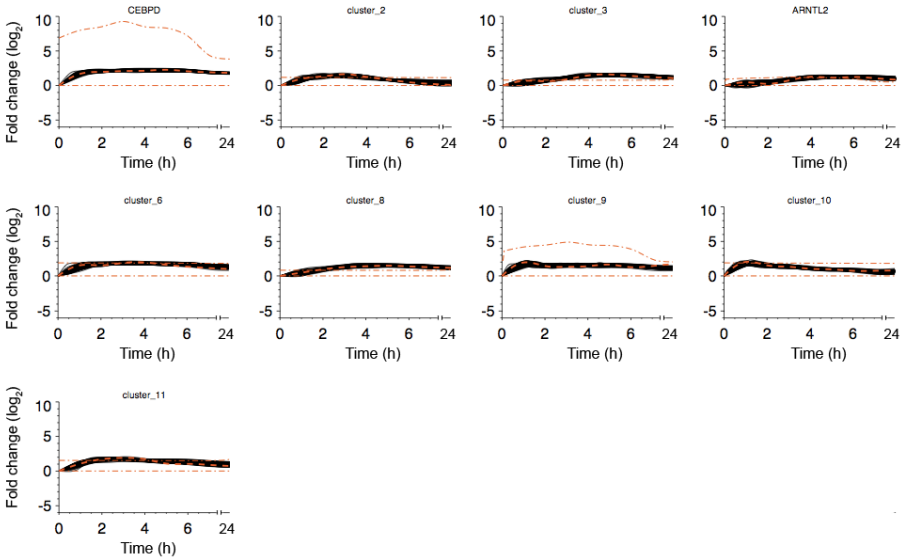
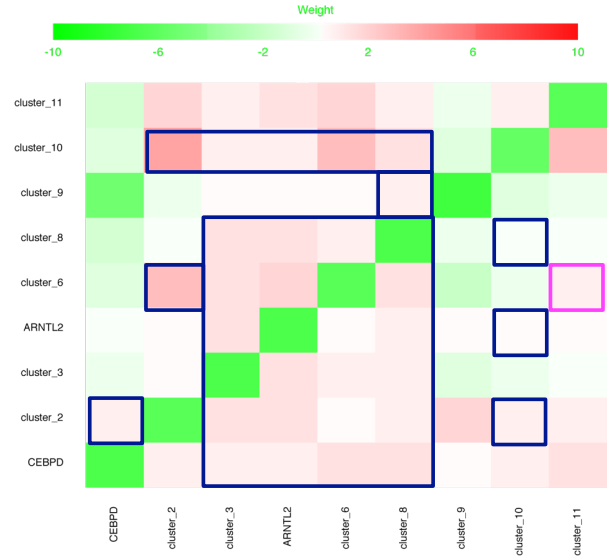


Figure 64 Fitness evaluation over complete time series in the integrated model



Integrated Exp.1 and Exp.3 (20091005)

Figure 65 TC regulatory network reconstructed with a PSC cytokine signature shows a preserved core of interactions (blue frame) including cluster 10 (*EGR* genes)

6.2.10.1 Tumor cell output

Biological output of TC in response to stimulation with PSC supernatant was in the form of an altered GRN, and was also measurable by a changed profile of secreted proteins. Cytokines and chemokines produced by TC affect the microenvironment in a wide range of feedbacks. We have identified 31 soluble factors in TC#2, and 32 in TC#3 using GO (Table 19 and Table 20) among the top 500 upregulated and ranked genes in both sets.

	Response of tumor cells treated with stimulated stellate cell supernatant from top 500 ranked genes
CXCL1	chemokine (c-x-c motif) ligand 1 (melanoma growth stimulating activity, alpha)
CXCL2	chemokine (c-x-c motif) ligand 2
CXCL3	chemokine (c-x-c motif) ligand 3
CXCL5	chemokine (c-x-c motif) ligand 5
CXCL6	chemokine (c-x-c motif) ligand 6 (granulocyte chemotactic protein 2)
CX3CL1	chemokine (c-x3-c motif) ligand 1
CLCF1	cardiotrophin-like cytokine factor 1
CXCL10	chemokine (c-x-c motif) ligand 10
IL6R	interleukin 6 receptor
IL8	interleukin 8
IL15	interleukin 15
IL32	interleukin 32
TNF	tumor necrosis factor (tnf superfamily, member 2)
HBEGF	heparin-binding egf-like growth factor
FGF18	fibroblast growth factor 18
WNT6	wingless-type mmtv integration site family, member 6
CSF1	colony stimulating factor 1 (macrophage)
CSF3	colony stimulating factor 3 (granulocyte)
SERPINE1	serpin peptidase inhibitor, clade e (nexin, plasminogen activator inhibitor type 1), member 1
C3	complement component 3
SHH	sonic hedgehog homolog (drosophila)
PTX3	pentraxin-related gene, rapidly induced by il-1 beta
BTC	betacellulin
SFN	Stratifin
GAL	galanin
ANGPTL4	angiopoietin-like 4
PLAT	plasminogen activator, tissue
FAM20C	family with sequence similarity 20, member c
CFB	complement factor b
PRSS35	protease, serine, 35
HEG1	heg homolog 1 (zebrafish)
PLAUR	plasminogen activator, urokinase receptor

Table 19 32 soluble factors identified with *David* GO among top 500 upregulated genes in TC#3 (supernatant of stimulated PSC)

	Response of tumor cells treated with quiescent stellate cell supernatant from top 500 ranked genes
WNT6	wingless-type mmtv integration site family, member 6
RBP4	retinol binding protein 4, plasma
MMP13	matrix metallopeptidase 13 (collagenase 3)
GDF15	growth differentiation factor 15
IGHG1	immunoglobulin heavy constant gamma 1 (g1m marker)
LOXL2	lysyl oxidase-like 2
CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
PLAUR	plasminogen activator, urokinase receptor
LAMC1	laminin, gamma 1 (formerly lamb2)
SERPINE1	serpin peptidase inhibitor, clade e (nexin, plasminogen activator inhibitor type 1), member 1
CCL5	chemokine (c-c motif) ligand 5
SFRP1	secreted frizzled-related protein 1
ANGPTL4	angiopoietin-like 4
MUC15	mucin 15
IGHG3	immunoglobulin heavy constant gamma 3 (g3m marker)
MMP1	matrix metallopeptidase 1 (interstitial collagenase)
LTBP2	latent transforming growth factor beta binding protein 2
TFPI	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)
KLK7	kallikrein 7 (chymotryptic, stratum corneum)
CTGF	connective tissue growth factor
SERPINA3	serpin peptidase inhibitor, clade a (alpha-1 antiproteinase, antitrypsin), member 3
CYR61	cysteine-rich, angiogenic inducer, 61
GSN	gelsolin (amyloidosis, finnish type)
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)
PSG6	pregnancy specific beta-1-glycoprotein 5
PLAU	plasminogen activator, urokinase
INHBE	inhibin, beta e
SPINLW1	serine peptidase inhibitor-like, with kunitz and wap domains 1 (eppin)
IGFBP4	insulin-like growth factor binding protein 4
PSG9	pregnancy specific beta-1-glycoprotein 9
LIPG	lipase, endothelial

Table 20 31 Soluble factors identified with *David* GO among top 500 upregulated genes in TC#2 (supernatant of quiescent PSC)

6.2.11 Tumor cell summary

Experiments PSC#1 and TC#3 form a logical chain of events in the transformed tissue where the initialization of PSC by TC is closely followed by a stimulation of those TC through the secreted factors coming from PSC in a feedback loop. This loop extends further beyond our experimental setup with TC responding to the stimulation with an altered profile of secreted proteins likely affecting PSC in a paracrine feedback. While evaluating microarray data we investigated only the transcriptional level of organization (assuming that the protein level approximately follows the transcript levels in the cells) and ignoring any posttranslational modifications, which are otherwise known to alter cell regulation, and any mechanistic effects related to the proximity of the cells *in vivo*.

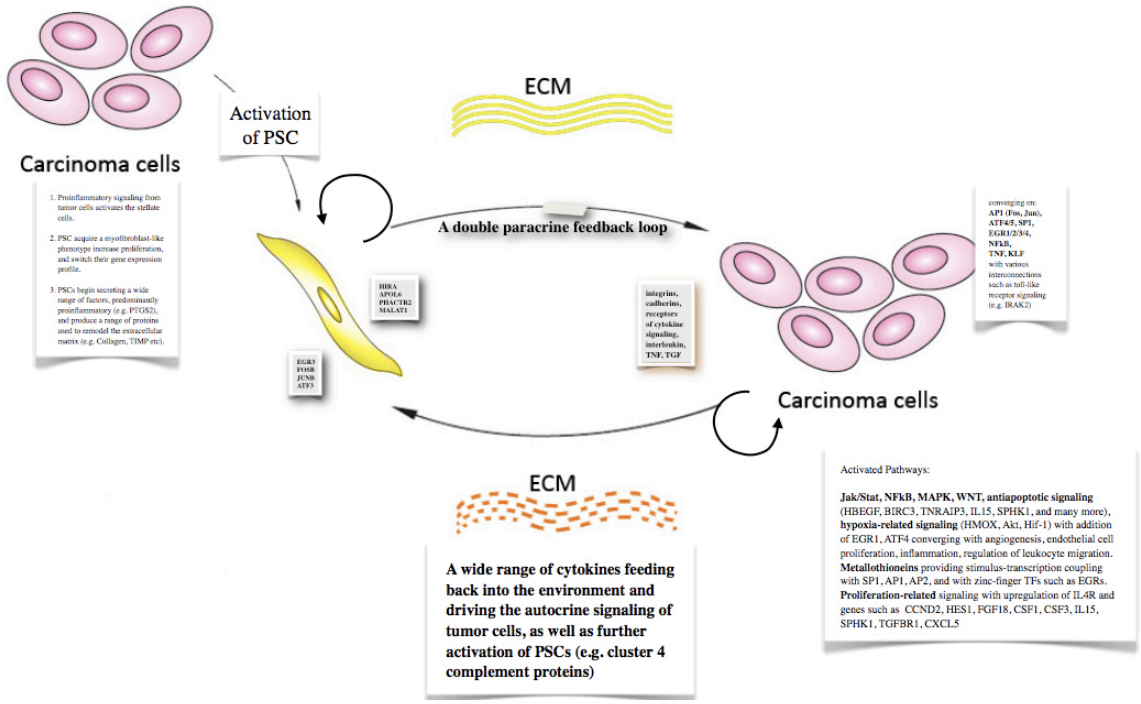


Figure 66 Sequence of events taking place in the tumor microenvironment between TC and PSC

Our complexity reduction approach has allowed us to perform both a quantitative (model) and qualitative (functional genomics) analysis of the inter- and intra-cellular interactions of tumor and stellate cells, with a focus on the GRNs arising in response to the stimulus coming from the other respective cell type. Due to the nature of the tissue organization *in vivo*, it is clear that the impact of TC and PSC on each other is continuous and cannot be completely resolved in an 8-point time series. It is also necessary to understand that in the final stage of intercellular connection identification, we are not investigating the tumor onset, but rather the more stable state of an established system, where PSC have been pre-activated in tissue, and TC have previously seen PSC. Our analysis of time-resolved microarray data, clustering, and GRN models showed that the immediate response of cancer cells to primed stellate cell stimulation comes via the activation of early response transcription factors including:

EGR1 (involved in the MAPK pathway, also mediates up-regulation of epidermal growth factor receptor expression during hypoxia, was implicated in carcinogenesis and cancer progression, especially metastasis, and was also linked to the MEK1/ERK1/2 pathway coupling between the CCK2 receptor and nuclearization and DNA binding of Egr-1 (Leung-Theung-Long et al. 2005));

EGR2 and ***EGR3*** (both have direct links to FasL ligand (CD95L death ligand) and can induce apoptosis in various cancer cell lines by direct transactivation of BNIP3L and BAK (Unoki et al. 2003), although experimental evidence exists suggesting that EGR2 does not directly induce CD95L (Li-Weber et al. 2003); and ***EGR4*** implicated in mitogenesis and differentiation. (RefSeq). These 4 genes feed the signals into the other connected network modules, which include transcriptional factors, repressors and enhancers in the cancer cells, and it seems that functionally the most important among them are:

CEBPD (recognizes similar DNA sequences in their target genes and form homo- or hetero-dimers with other C/EBPs, as well as with transcription factors of the NF-κB and Fos/Jun families);

AP1 member: FOSB, JUNB. Each of the FOS family of proteins can dimerize with proteins of the JUN family, as well as ATF factors forming the transcription factor complex AP-1. Genes of the FOS and JUN families have been implicated as regulators in cell cycle progression, differentiation, and cell death, and were shown to affect transformation (in e.g. squamous cell sarcoma, Hodgkin lymphoma). AP1 has been shown to possess a remarkable capability for decision-making and antagonistic features in nearly all processes it affects depending on the cellular context (Hess et al. 2004).

SOX8 (involved in the regulation of embryonic development and in the determination of the cell fate);

ATF4 Activating transcription factor 4 has been shown to have among others a role in regulating VEGF expression, as well as responses to hypoxic stress (Afonyushkin et al. 2010);

ATF5 shown to increase cisplatin-induced apoptosis through up-regulation of Cyclin D3 transcription (Wei et al. 2006);

NFKB2 implicated in survival, apoptosis, immune and inflammatory responses, and cell growth. Correlates with autoimmune arthritis, asthma, septic shock, lung fibrosis, glomerulonephritis, atherosclerosis (Chen et al. 1999 and Baldwin 1996)

TNFAIP2 induced by TNF signaling interconnects with life and death pathways via JNK, NFκB (RefSeq);

KLF10 – Krueppel-like Factors have a broad range of effects including proliferation, differentiation, development, and programmed cell death. They have guilt by association connection with signaling hubs such as SP1, which while not present in our dataset in significant amounts, may be a significant player in signal distribution. KLF10 has been associated with t-cell differentiation and activation, role in development of the heart (fibrosis and myocyte disarray as a result of KLF knockouts in mice) (McConnell et al. 2010)

ZEB1 a zinc finger transcription factor plays a role in transcriptional repression of interleukin 2 and its best-described biological function is the induction of EMT in epithelial cells (Wellner et al. 2010). Arumugam et al. 2009 suggest that *ZEB1* and other regulators of EMT may maintain drug resistance in human pancreatic cancer cells, and therapeutic strategies to inhibit *ZEB1* and reverse EMT should be evaluated.

In silico simulations of the TC model provided two potential targets for experimental validation (*EGRs*, and *CEBPD*). EGRs showed the strongest single-module knockdown effect *in silico*, therefore they were selected for validation.

6.3 Intercellular signaling

In order to decipher the intercellular signaling patterns a multi-stage approach (see 5.8) was used in which first the soluble factors among the top 500-upregulated genes in each experiment were identified using GO and pathway tools (aforementioned). Altogether 39 factors in PSC (#1), 31 in TC (#2), and 32 in TC (#3) were derived (Table 9, Table 19, and Table 20) with only 2 shared across all (*ANGPTL4, PLAUR*). TC setups provided divergent sets of signatures with only 2 factors shared (*Serpine1, WNT6*), proving a stimulus-dependent induction of signaling (Figure 67).

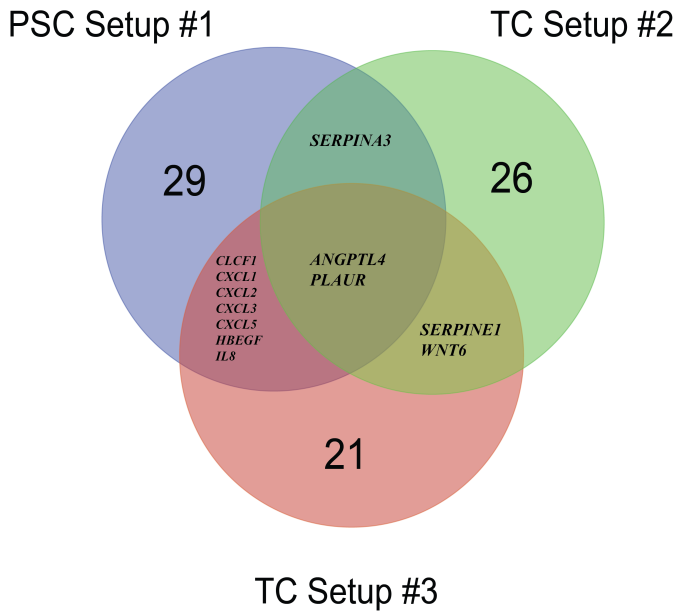


Figure 67 Venn diagram comparing secreted factors identified among top 500 ranked genes in all three experiments reveals only 1 gene overlapping all sets.

In the second stage of this procedure a knowledge-driven approach was applied to link reverse engineered GRN with cell-specific soluble factors as proposed in section 5.8. For this purpose the intercellular signaling was divided into 3 main phases covering the separate outputs of all investigated cells (Figure 68).

The GRNs were explored with transcription factor binding site analysis (TFBSA) to discover the lowest level of regulation, from which data were extrapolated using a non-specific (tissue, cell-type, disease) literature-driven Ariadne Pathway Studio to encompass all levels of signaling. Pathway analysis was applied to each data set with the correct direction of regulation in search of the shortest path connecting either common targets or common regulators of the derived

TF. To ensure condition-specific signal identification, filtering of pathway analysis results was performed against each corresponding microarray data set. Downregulated genes, and in general any genes whose expression never exceeds $\log_2$ fold change value of 0.5 were not taken into consideration with the exception of the search for constitutively expressed factors for which each dataset was taken in its normalized, but unfiltered form.

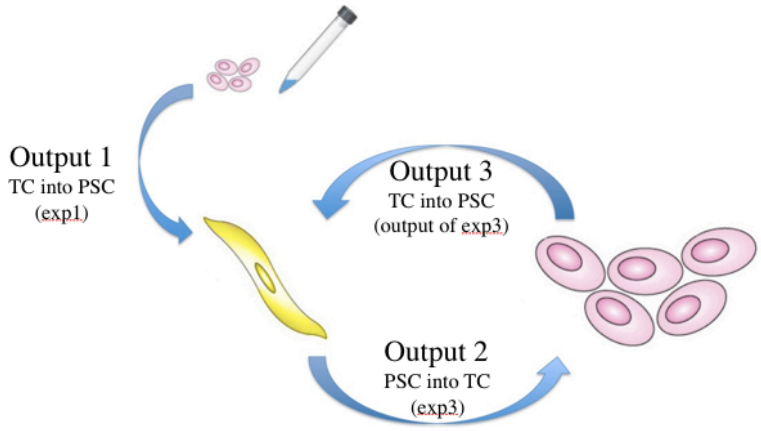


Figure 68 Division of signaling into 3 major phases with specific outputs of each cell type of interest

6.3.1 Output 1: Initialization of communication between TC and PSC

We showed through microarray and model analysis that PSC behavior was significantly altered after initial stimulation with TC (summary effect of 8 cell lines). We therefore recognized that constitutively secreted proteins produced by TC were sufficient to be identified by, and to stimulate PSC. Cytokines constitutively secreted by TC were identified from the controls of experiment 2 and 3. This initiating signal affected PSC and we deciphered (Figure 69) it into:

- the GRN model of PSC (#1)(upregulated genes), and
- a cluster of approximately 150 genes, which is representative of the majority of downregulated genes in this experiment (cluster 10)

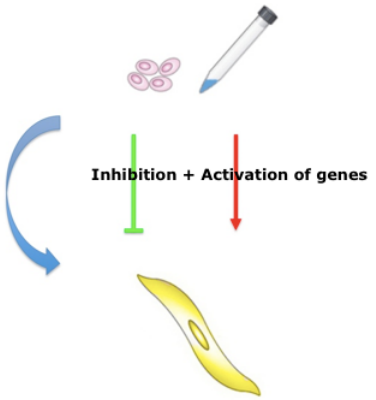


Figure 69 Stimulation elicits inhibition and activation of genes in the target cells (PSC)

The activation of genes resulted in the formation of the GRN as identified with modeling (see section 6.1.7). From this model the most significant module was the central hub containing *HIRA, APOL6, PHACTR2,* and *MALAT1*. Those genes were used for TFBSA analysis to identify the underlying TFs (Figure 70)(3 significant with a p-value < 0.05: *NRSF, CRX, NR5A2* and additional 6 with p-value (0.1 > pVal. > 0.05): *SZF1, RFX, OG2, DBP, NKX25, KID3*).

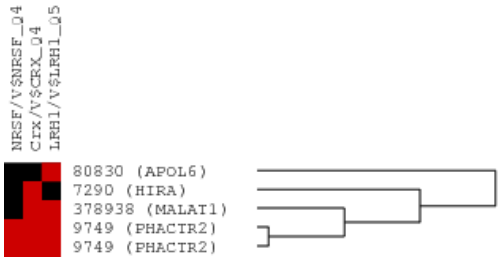


Figure 70 TFBSA heat map result showing the statistically significant p<0.05 (red) transcription factors whose binding sites are overrepresented for the central hub genes

An additional stage of TFBSA analysis was applied to the top 250-upregulated genes in PSC dataset and resulted in a list of 14 TF (*CREB, E2F, EGR1, ETF, ETS1 (c-ETS-1, p54), HIC1, IRF1, PAX-4, SP1, STAT2 (ISGF3), TAX, TEL-2, TFAP2A (AP2), ZF5 (ZFPI161)*).

The derived TFs (Figure 70) were used for pathway reconstruction using Ariadne Pathway Studio (Figure 71), which allowed us to build the shortest pathways by expanding them from the starting set in search of the extracellular and membrane proteins, which regulate them.

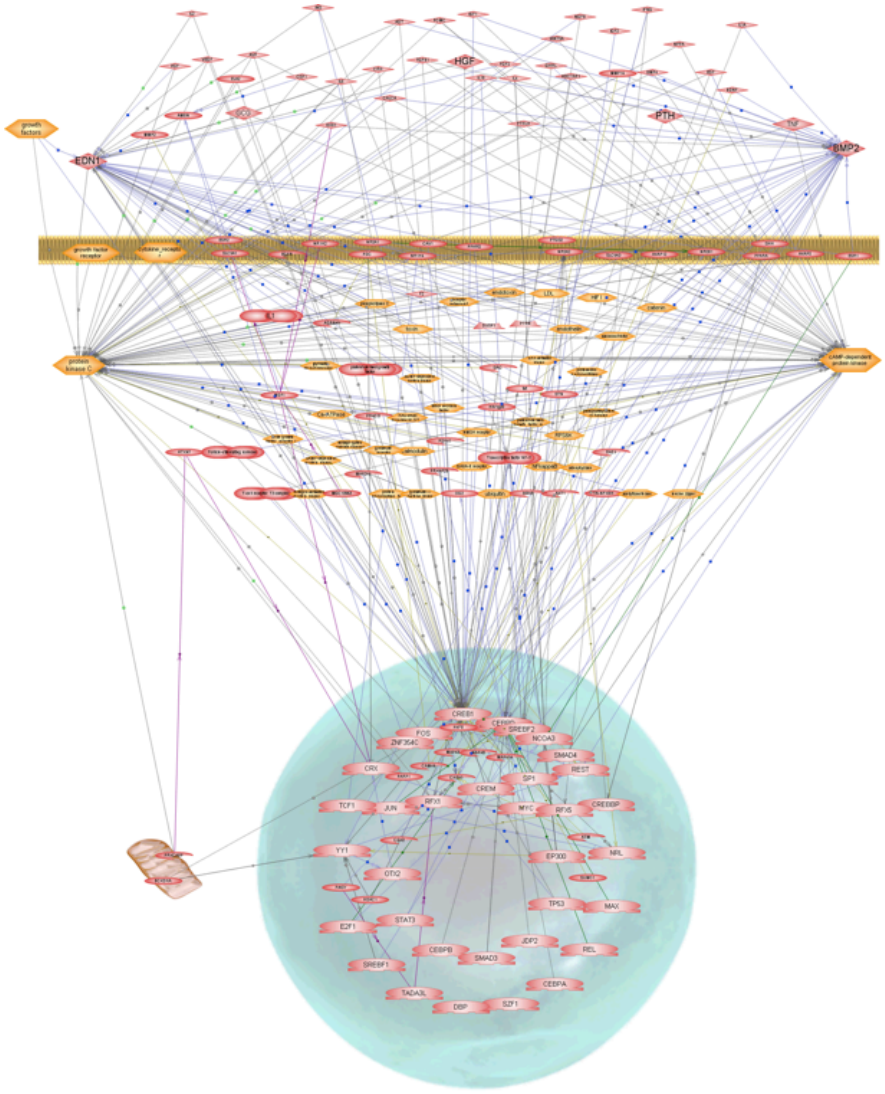


Figure 71 Pathway studio sample result – common regulators of three major transcription factors in stellate cells. Pathway studio provides options to separate signaling into cell compartments and easily identify interactions between any and all of the involved genes.

Analysis of the PSC GRN-derived TFs (TFBSA) in Pathway Studio has yielded a total of 82 soluble factors, all filtered against the experimental microarray datasets to ensure PDAC-specificity (initial Pathway Studio are tissue/cell-type/disease unspecific).

Genes were further filtered to identify unique patterns, where each soluble factor was investigated for its presence in the dataset of the other cell type, both changing among top regulated genes and at a constitutive level in controls. This stage ensured that on one hand no significant genes were lost, as even if they overlapped between cells on a constitutive level they were assigned to the correct group, and on the other hand all factors expressed uniquely by only one cell type were identified providing an overview of stimulus-dependent cues. TC datasets #2 and #3 were generated with MiaPaCa2 TC, whereas PSC setup #1 stimulation was performed with a combined set of 8 TC cell lines, therefore an additional inspection of all genes of interest was carried out against the Oncomine database (Wagner CellLines dataset) for all of the remaining and available lines (MiaPaCa2, Panc1, BxPC3, SU8686, Capan1, Aspc1). Finally genes were divided into three unique groups based on the expression levels in each cell type (Figure 72).

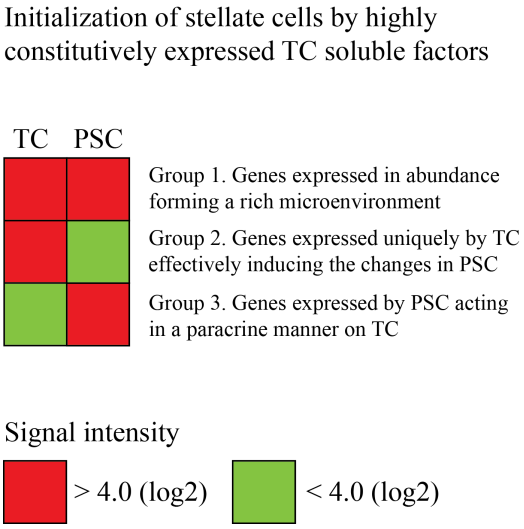


Figure 72 Gene expression grouping in stellate cells

Group 1 genes (TC ↗, PSC ↗) (*VEGF, DKK1, CTGF, TGFβ1*), which were highly constitutively expressed in both cell types (and all TC cell lines of interest) form a homeostatic-like, rich microenvironment, which most likely provides the background for the subsequent signal formation, but is unlikely to affect the regulatory dynamics in the forming GRN reverse

engineered in PSC #1. This group was additionally augmented with all the different genes produced constitutively or inducible in response to treatment from both cell types e.g. genes showing a variance across different TC cell lines (cross-referenced with Oncomine)(*FN1, ADAMTS1, PLAT, BDNF, ADM*).

Group 2 genes (TC ↗, PSC ↘) uniquely target the intracellular GRN dynamics of PSC and were highly expressed by TC in either all (*MMP9, CTSE*), or some of TC cell lines (*LAMA2, CCL5, PGF, NPPA, CSF1, FGF2*), and not by PSC. This group was additionally augmented with genes, which initially showed low expression in MiaPaCa2 TC and PSC, but were upregulated in the other TC lines (*MMP7, LAMC2, PTHLH, CCL8, EGF, BMP2, MMP13*).

Group 3 genes (TC ↘, PSC ↗) in general form the unique signaling path from stellate cell into tumor cells, however when expressed on a constitutive level and not inducible these factors are more likely affecting PSC in an autocrine feedback loop (e.g. *INHBA, CCL2, MMP14*). They are therefore most likely significant for the formation of the stellate cell environment.

Additionally a set of membrane-bound proteins was identified on PSC stimulated by the initialization signal from TC and included:

Membrane/membrane bound proteins in PSC: AKAP12, AKAP5, CAV1, EGFR, ESR1, PDC, PPARG, PTGS2

Nuclear receptors: NR0B1, NR1H2, NR5A1, NR5A2

KEGG Pathways for secreted proteins	
MAPK signaling pathway	adjP=2.90e-06
Cytokine-cytokine receptor interaction	adjP=2.90e-06
Hypertrophic cardiomyopathy	adjP=3.30e-06
Dilated cardiomyopathy	adjP=3.30e-06
Hematopoietic cell lineage	adjP=3.30e-06
Pathways in cancer	adjP=3.30e-06
Graft-versus-host disease	adjP=9.43e-05
Type I diabetes mellitus	adjP=9.43e-05
mTOR signaling pathway	adjP=0.0001
Pancreatic cancer	adjP=0.0001

Table 21 KEGG pathway analysis of secreted proteins

6.3.1.1 Inhibiting signals

An additional evaluation of inhibited genes in PSC was performed for cluster (10) (see 6.1.4.1). Genes were analyzed with TFBSA to identify the underlying TF, which then were used in Pathway Studio to identify extracellular proteins affecting them. Overlap with microarray data provided tissue-specific factors including:

Secreted: AGT, EDN1, EGF, FGF2, FGF8, HGF, IGF1, IL1B, IL4, IL8, INS, LEP, TGFB1, TNF, VEGF.

Membrane/membrane bound: AHR, EGFR, ESR1, KITLG, PPARG, PTGS2, RXRA, VDR,

Nuclear receptors: NR0B1, NR1H4, NR2F1, NR4A1, NR5A2

KEGG pathways analysis was performed using WegGestalt. Top 10 results are shown in Table 22 and Table 23.

KEGG pathways for secreted proteins

Cytokine-cytokine receptor interactions	TGFB1, TNF, HGF, LEP, IL1B, EGF, IL8, IL4	adjP=1.25e-13
Pathways in cancer	TGFB1, HGF, FGF8, FGF2, EGF, IL8, IGF1	adjP=5.51e-11
Melanoma	HGF, FGF8, FGF2, EGF, IGF1	adjP=1.78e-10
MAPK signaling pathway	TGFB1, TNF, IL1B, FGF8, FGF2, EGF	adjP=9.82e-10
Type I diabetes mellitus	TNF, IL1B, INS	adjP=2.03e-06
Regulation of actin cytoskeleton	FGF8, FGF2, INS, EGF	adjP=2.66e-06
NOD-like receptor signaling pathway	TNF, IL1B, IL8	adjP=4.13e-06
Hypertrophic cardiomyopathy	TGFB1, TNF, IGF1	adjP=8.65e-06
Hematopoietic cell lineage	TNF, IL1B, IL4	adjP=8.65e-06
Prostate cancer	INS, EGF, IGF1	adjP=8.65e-06

Table 22 KEGG pathway analysis of secreted proteins

KEGG pathways for membrane proteins

Pathways in cancer	KITLG PTGS2 PPARG EGFR RXRA	adjP=1.17e-05
Thyroid cancer	PPARG RXRA	adjP=0.0007
Non-small cell lung cancer	EGFR RXRA	adjP=0.0014
PPAR signaling pathway	PPARG RXRA	adjP=0.0018
Small cell lung cancer	PTGS2 RXRA	adjP=0.0021
Cytokine-cytokine receptor interaction	KITLG EGFR	adjP=0.0137
MAPK signaling pathway	NR4A1 EGFR	adjP=0.0137

Table 23 KEGG pathway analysis of membrane proteins

The same general procedure was applied to the remaining outputs.

6.3.2 Output 2: Feedback of PSC acting upon TC

PSC response to the initialization signal was shown in the previous step in the form of an identified signature of soluble factors (Table 9). These did not cover the extracellular signaling in its entirety, since they were identified among top 500 ranked genes ignoring highly constitutively expressed factors, which do not change significantly over time and are filtered out. In order to specifically identify the correct subset of soluble factors driving the formation of TC #3 GRN, TFBSA and Pathway Studio analysis were performed as described earlier.

1. Identification of secreted factors regulating the core network genes in TC
Data source: 35 genes used for the TC model analyzed with TFBSA to uncover underlying regulation (see 6.2.8.1 Gene selection). Identified TFs were used for Pathway Studio analysis and the resulting list of genes encoding secreted proteins, was overlapped (intersection search) with stellate cell experiment 1 microarray data set.

2. Identification of TC receptors and membrane-bound proteins affected by secreted proteins from PSC and responsible for transducing signals to the internal network in TC.
Data source: as above, but filtering of the results performed against the TC dataset, in search of TC receptors.

3. Identification of PSC cytokine targets in TC – both receptor and membrane-bound
Data source: genes encoding secreted proteins in stellate cells (experiment 1, Table 9), were used for Pathway Studio analysis to identify membrane bound and intracellular proteins. Results were filtered against the TC data (experiment 3).

Soluble factors from both sets (pathway results and top secreted factors) were combined to produce a list of 61 genes (11 shared between the lists). Subsequent analysis was performed to produce three main groups of soluble factors (Figure 73) with a subset specifically targeting the intracellular dynamics of TC (group 2).

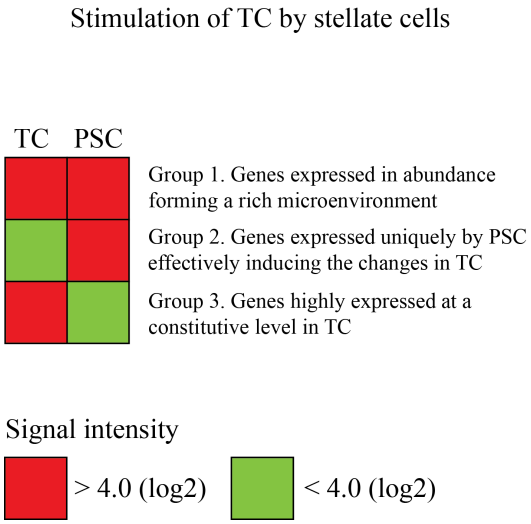


Figure 73 Gene expression grouping of soluble factors in tumor cells experiment #3

Group 1 genes (PSC ↗, TC ↗) are homeostatic factors, which are inducible in PSC, but are also highly constitutively expressed by TC. These factors do not alter the overall homeostasis of the TC, unless they are factors with gradient dependent action (chemoattractants) (*PLAU, S100A6, TFPI, ISG15, PLAUR, GNPTG, VEGFA, LIF, ADM, CLCF1, HBEGF, LCN2, PDGFA, PCOLCE2, ANGPTL4, F3*).

Group 2 genes (PSC ↗, TC ↘) are focusing factors, which are expressed uniquely by PSC and drive the stimulation of TC in setup #3 (*DKK3, TIMP3, TFPI2, IL6, ESM1, EDN1, IL7R, KITLG, CCL20, TNFAIP3, EBI2*). This group was also augmented by selected genes, which were unique to only some of the TC cell lines (downregulated) Constitutively high and unique to PSC (*INHBA, CCL2, DCN, IL6, MMP14* produced by all PSC, and some TC *FN1, ADAMTS1, PLAT, BDNF*)

Group 3 genes (PSC ↘, TC ↗) Autocrine enhancement, and paracrine feedback into PSC (output 3) (*SERPINA3, CXCL1, EREG, CXCL5, CXCL3, COL17A1, AREG, CCL5, CXCL2, LCN2, IL8, IL11*)

In addition we also investigated the presence of receptors and membrane bound proteins in the unfiltered, and unranked microarray dataset in search of highly constitutively expressed genes, which were not changing significantly across the time points (therefore filtered out by the standard processing approaches). Factors were divided into two subgroups – the primary group contained genes identified with both aforementioned methods (targets of cytokines, and common regulators). Secondary group contains factors identified with only one of the two methods.

Membrane/membrane bound (TC):

Primary: *AR, BCL2L11, CEACAM1, HMOX1, ICAM1, IL6R, IL6ST, PLAUR,*
Secondary: *TNFRSF9, IL20RB, CD83, SDC4, ABCG1 ITGB4, SLC7A5, IFNAR2, TGFBR1, SLC3A2, SHH*
Constitutively expressed in TC: *IL8RA, IL8RB (CXCR2), CXCR4, CCR6, EGFR, LIFR, TFR2, IL11RA, PPARA, ITGB1, PDGFRA.*

Intracellular (knowledge-driven, present in tumor cell dataset, but not necessarily the same as the model genes in TC): *BCL2, CASP3, CASP8, CCND1, CCNE1, CDKN1A, CEBPB, CFLAR, CTNNB1, DUSP1, EGR1, EGR3, FAM55C, FOS, FYN, IRF1, JAK1, JUN, JUNB, PCOLCE2, PTK2, RPS6KB1, SHC1, SPHK1, STAT1, TNFAIP3*

KEGG analysis has been performed for all sets combined, top 20 results are shown in Table 24.

KEGG pathways analysis

Pathways in cancer	CASP3, FN1, TGFA, TGFBR1, PTK2, CDKN1A, FOS, CCNE1, JUN, CCND1, PDGFA, CTNNB1, JAK1, IL6, IL8, ITGB1, SHH, STAT1, AR, BCL2, CASP8	adjP=7.73e-20
Cytokine-cytokine receptor interaction	LIF, PDGFA, IL11, IFNAR2, TNFRSF9, CCL20, IL6, IL6ST, IL8, TGFBR1, IL20RB, IL6R	adjP=1.28e-09
Focal adhesion	PDGFA, CCND1, JUN, FYN, CTNNB1, SHC1, FN1, ITGB1, BCL2, PTK2, ITGB4	adjP=1.28e-09
Jak-STAT signaling pathway	LIF, CCND1, IL11, IFNAR2, JAK1, IL6, IL6ST, STAT1, IL20RB, IL6R	adjP=1.47e-09
ErbB signaling pathway	JUN, RPS6KB1, PTK2, CDKN1A, SHC1, AREG, TGFA, EREG	adjP=5.31e-09
Prostate cancer	PDGFA, CCND1, CTNNB1, TGFA, AR, BCL2, CDKN1A, CCNE1	adjP=5.86e-09
Colorectal cancer	CCND1, JUN, BCL2, TGFBR1, CASP3, CTNNB1, FOS	adjP=1.17e-07
Toll-like receptor signaling pathway	JUN, IFNAR2, CASP8, FOS, IL6, IL8, STAT1	adjP=3.23e-07
Small cell lung cancer	CCND1, BCL2, PTK2, FN1, CCNE1, ITGB1	adjP=2.61e-06
p53 signaling pathway	CCND1, CASP3, CDKN1A, CASP8, CCNE1	adjP=1.54e-05
Glioma	PDGFA, CCND1, CDKN1A, SHC1, TGFA	adjP=1.54e-05
Viral myocarditis	CCND1, CASP3, FYN, CASP8, ICAM1	adjP=1.54e-05
Pancreatic cancer	CCND1, TGFBR1, JAK1, TGFA, STAT1	adjP=2.03e-05
NOD-like receptor signaling pathway	TNFAIP3, CASP8, IL6, IL8	adjP=0.0002
Natural killer cell mediated cytotoxicity	IFNAR2, CASP3, FYN, SHC1, ICAM1	adjP=0.0003
Chronic myeloid leukemia	CCND1, TGFBR1, CDKN1A, SHC1	adjP=0.0003
Apoptosis	BCL2, CASP3, CASP8, CFLAR	adjP=0.0006
Prion diseases	FYN, EGR1, IL6	adjP=0.0006
ECM-receptor interaction	ITGB4, FN1, SDC4, ITGB1	adjP=0.0006
MAPK signaling pathway	PDGFA, JUN, TGFBR1, CASP3, FOS, DUSP1	adjP=0.0008

Table 24 KEGG pathway analysis

KEGG analysis provides a clear division of the identified factors into known pathways, corroborating previous GO and pathway results for both cell types. Combined, the unique signals converging on PSC and TC provided a list of potential targets to disrupt not only the cell intracellular network, but in fact the extracellular signaling responsible for the formation of this network, offering the option of breaking the further perpetuation of the TC-PSC signals.

6.3.3 Output 3: Tumor cells response to the stimulation

Analysis of the TC#3 response to PSC stimulation was divided into two phases.

1. Identification of soluble factors produced by TC#3.

Output 3 genes (Table 19) were identified among the most significantly inducible in the time-series experiment using GO among on top 500 upregulated genes. Common factors were joined into the microenvironment (Figure 67), and an additional screen was performed against the PSC genes in search of constitutively highly expressed factors. Each gene identified as highly constitutively expressed in PSC and low in TC was also verified among inducible genes, if found it was considered as inducible, but aiding the formation of the microenvironment rather than the formation of the cell-cell communication. Genes highly constitutively expressed in both cell types additionally induced in TC tip the balance in favor of TC. Finally, genes at a low constitutive level in both cell types, if induced in TC, would present a significant and unique soluble factor most likely inducing further effects on adjacent cells. The last group of primary interest consisted of: *CXCL6, CX3CL1, CXCL10, IL6R, TNF, FGF18, WNT6, CSF3, C3, SHH, BTC, GAL, CFB*.

2. Search for targets of soluble factors.

This analysis step was speculative, as output 3 could not be validated against an experimental target data set (no corresponding measurements of stimulated PSC responding to stimulated tumor cells exist).

Data source: Secreted factors identified in experiment 3, no other dataset can be used, and it is a speculative search for PSC membrane and intracellular targets of TC cytokines. Pathway studio was used to identify targets of tumor cell cytokines.

Intracellular: *BAX, BCL, BIRC3, CASP8, CFLAR, CTNNB1, DUSP1, FOS, MAPK1, MAPK8, PTK2*
Receptors and Membrane-bound proteins: *EGFR, ICAM1, IL6ST, KITLG, LPL, MCL1, PLAUR, PTGS2, TNFRSF11A*

As a result of the analysis a complete schema of interactions has been produced Figure 74.

6.4 Signaling schema

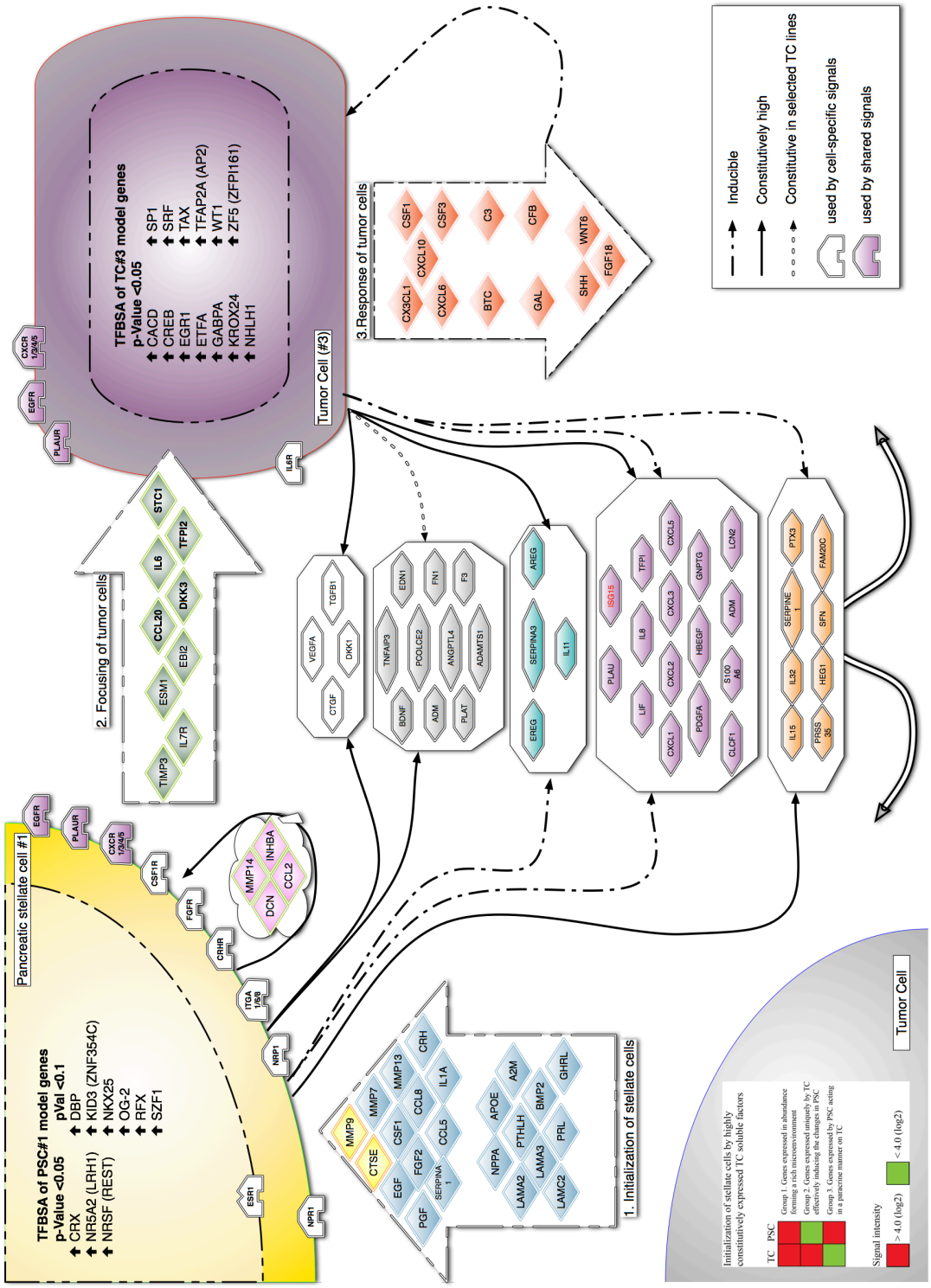


Figure 74 Signaling schema

6.5 Experimental validation

In order to confirm the transcript levels of microarray measured genes in PSC and TC we performed expression profiling using qRT-PCR for selected genes of interest. Moreover, to substantiate *in silico* simulations and predictions of the behavior of the TC core network genes *in vitro* gene silencing and qRT-PCR of selected genes was executed. This project was additionally supplemented by a series of assays including clonogenic, invasion, and MTT, samples were used for qRT-PCR measurements to validate identified intercellular connection. Biological readout of those assays is beyond the scope of the presented work.

6.5.1 qRT-PCR gene expression profiling of stellate cells

qRT-PCR measurements included model genes as well as known markers of stellate cell activation: *SMACTA2, FOS, EGR-1, IFITM1, OAS, ICAM1(CD54), PTGS2 (Cox2), KLF4, KLF5, KLF6, RAGE, HIAP2, IL6, VEGF, IL8, COL1A1, COL4A2, LIF, CXCL2 (MIP2a), CXCL5 (ENA78), CXCL12 (SDF-1), Fibronectin (FN1)*. Time series experiments formed a replicate of experiment 1 microarray and confirmed the used expression profiles.

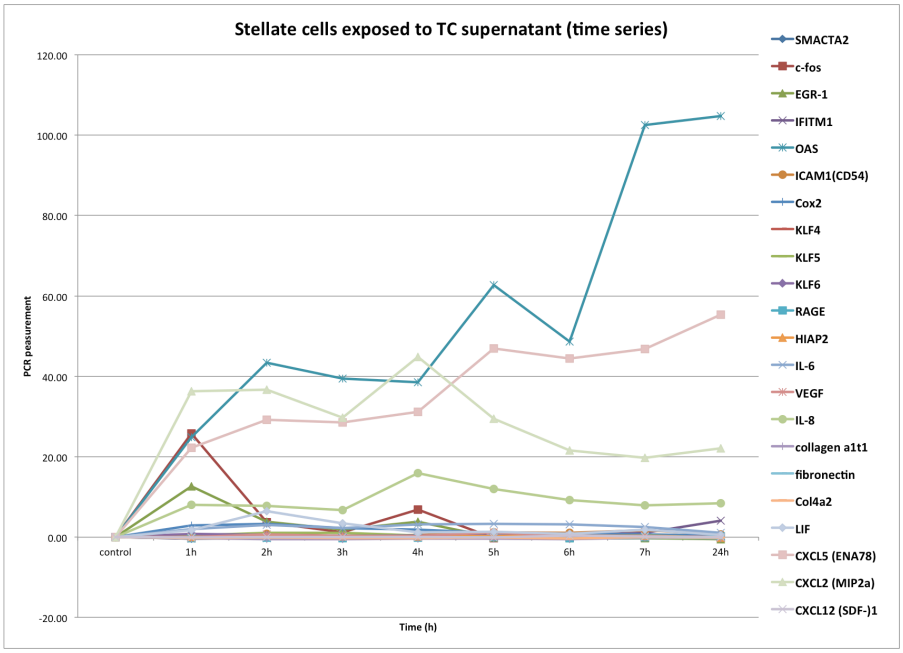


Figure 75 Expression evaluation of stellate cell selected genes of interest in a time series.

Central hub genes *HIRA, APOL6, PHACTR2, MALAT1* measured by means of microarrays (Figure 76) were evaluated using qRT-PCR. The dynamic profiles of *MALAT1 and PHACTR2* were confirmed, unfortunately *HIRA* and *APOL6* measurements were performed using primers targeting alternative splice variants resulting in different gene kinetics (confirmed after blasting the Affymetrix probe sequences, and PCR primers against the target sequences). The high expression levels of the central hub genes in stellate cells were confirmed with snapshot measurements from patient tissue samples. The highest levels were identified for *MALAT1*, and *PHACTR2*, and moderate for *HIRA* and *APOL6*.

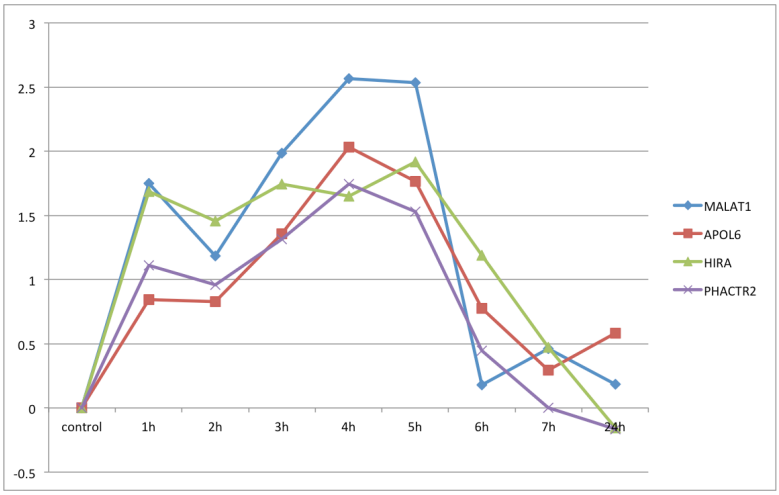


Figure 76 Microarray data for central hub genes *MALAT1, HIRA, APOL6*, and *PHACTR2*

To identify possible relationships in the gene expression we performed the nonparametric Spearman correlation on 3 separate datasets measuring the same set of central hub genes *HIRA, APOL6, PHACTR2, MALAT1* in:

1. 20 histologically verified PDAC patient tissues samples **(Tissue)**
2. Tumor ASPC1, BxPc3, Capan1, Colo357, MiaPaca2, Panc1, Su8384, T3M5 **(TC)**
3. 13 stellate cell samples gathered at various time points **(PSC)**

	TC	PSC	Tissue
correlation	MALAT1	MALAT1	MALAT1
HIRA	0.018	-0.099	-0.030
PHACTR2	0.419	-0.088	0.650
APOL6	0.038	0.863	0.722
p-Value	MALAT1	MALAT1	MALAT1
HIRA	0.948	0.748	0.900
PHACTR2	0.106	0.775	0.002
APOL6	0.888	0.0001	0.0003

	TC	PSC	Tissue
correlation	HIRA	HIRA	HIRA
MALAT1	0.018	-0.099	-0.030
PHACTR2	-0.027	0.310	0.290
APOL6	0.159	0.319	0.220
p-Value	HIRA	HIRA	HIRA
MALAT1	0.948	0.748	0.900
PHACTR2	0.920	0.303	0.214
APOL6	0.556	0.288	0.350

	TC	PSC	Tissue
correlation	PHACTR2	PHACTR2	PHACTR2
APOL6	0.469	0.124	0.642
p-Values	PHACTR2	PHACTR2	PHACTR2
APOL6	0.067	0.687	0.002

Table 25 Spearman correlation coefficients and corresponding p-Values for stellate cell central hub genes across 3 sets – tumor cell lines, stellate cells, and patient tissue.

A high correlation was noticed between *MALAT1* and *APOL6*, in both PSC and tissue samples, but not TC alone. This suggests a lack of functional coupling between the two genes in TC, at the same time confirming their significance in the GRN of PSC.

The known PSC activation markers including α-smooth muscle actin, collagen and fibronectin, showed reduced levels in both microarray and PCR data, which is not surprising as the primary human stellate cells were extracted from malignant tissue, where they were previously fully activated.

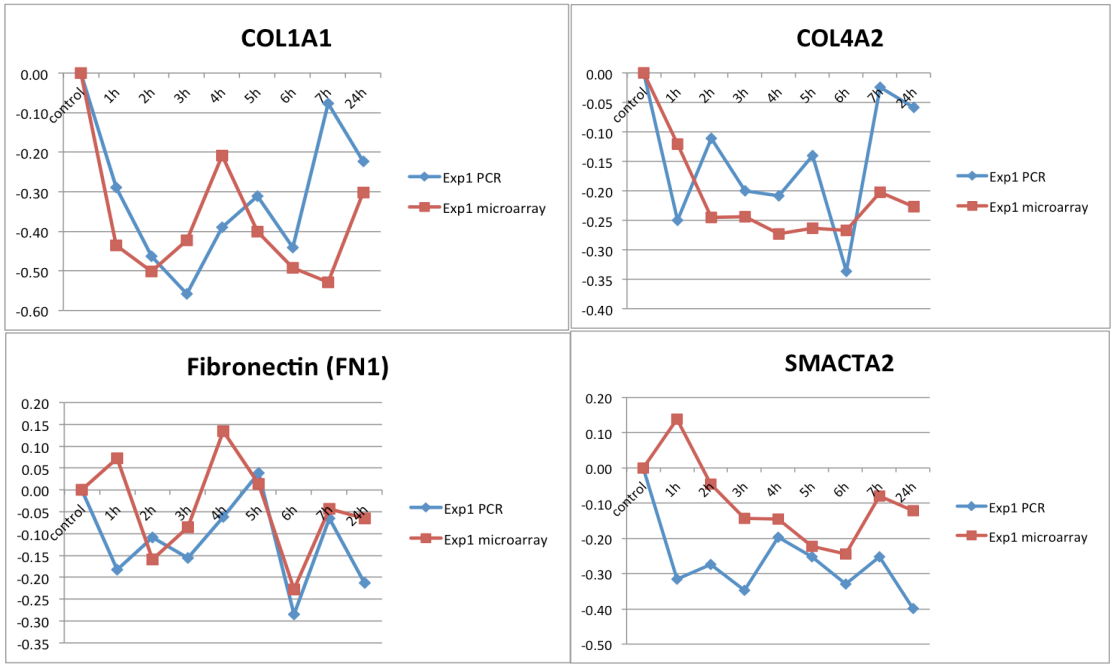


Figure 77 Comparison of marker gene expression levels between microarray and PCR experiments

We confirmed that the original microarray data used for model construction and data analysis was in agreement with the PCR measurements (Figure 78).

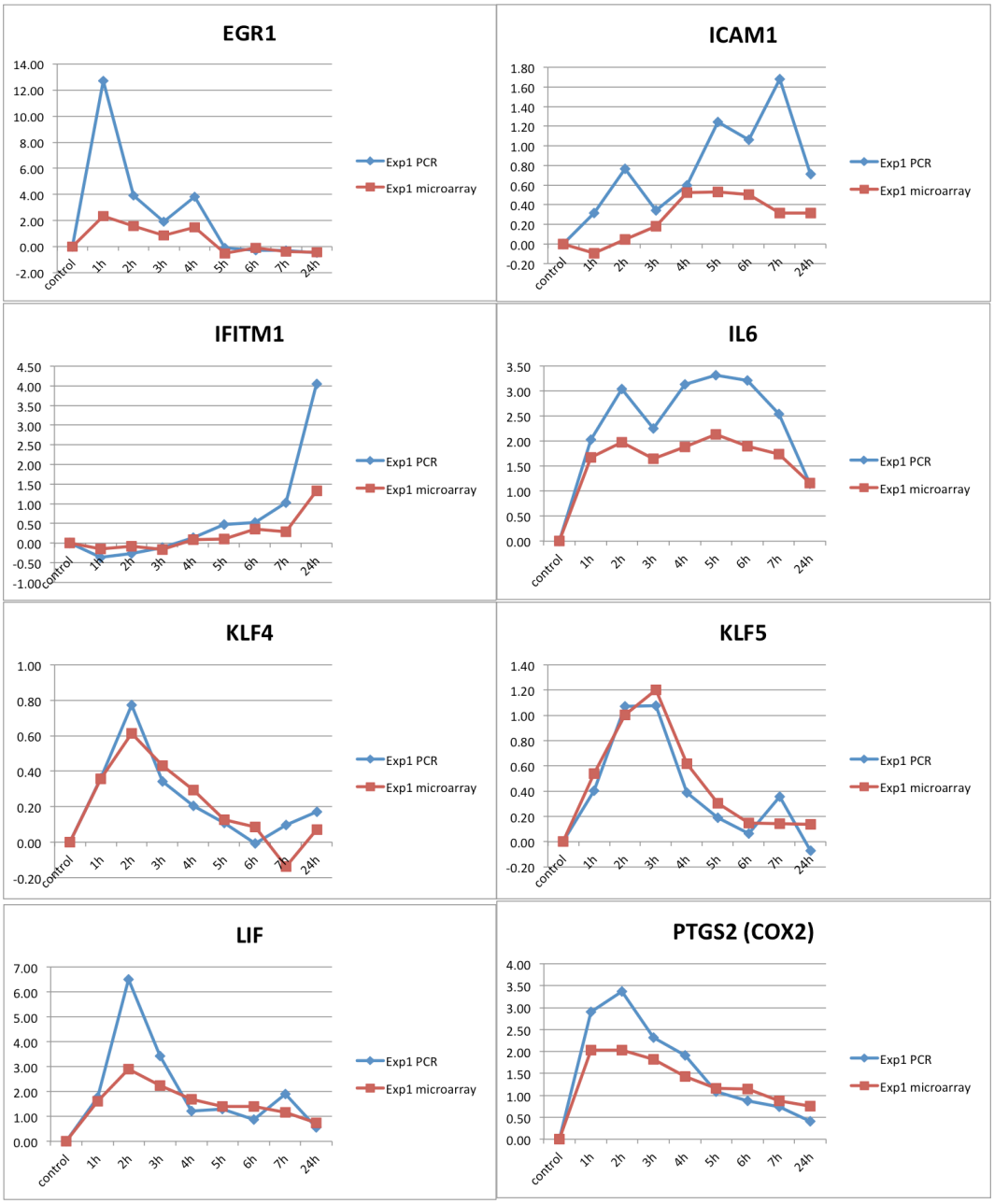


Figure 78 Microarray vs. PCR measurements of selected model genes in stellate cells

6.5.2 qRT-PCR gene expression profiling of tumor cells

qRT-PCR measurements of 16 representative genes used for the TC GRN were evaluated in a time series (Figure 79) including *EGR1, EGR2, EGR3, EGR4, SPHK1, PLAT, BCL3, GADD45A, HES1, JUN, NFKB, c-FOS, GRHL1, DUSP2, ARNTL2, SQSTM1*. This evaluation formed a replicate of the initial microarray experiment, and was supplemented by a set of genes of interest including: *IL6, IL8, VEGF, ICAM, CSF1/M-CSF, CSF3/G-CSF, CXCL2 (MIP2a), CXCL3(MIP2b), CXCL5(ENA78), MALAT1, CEBPB.*

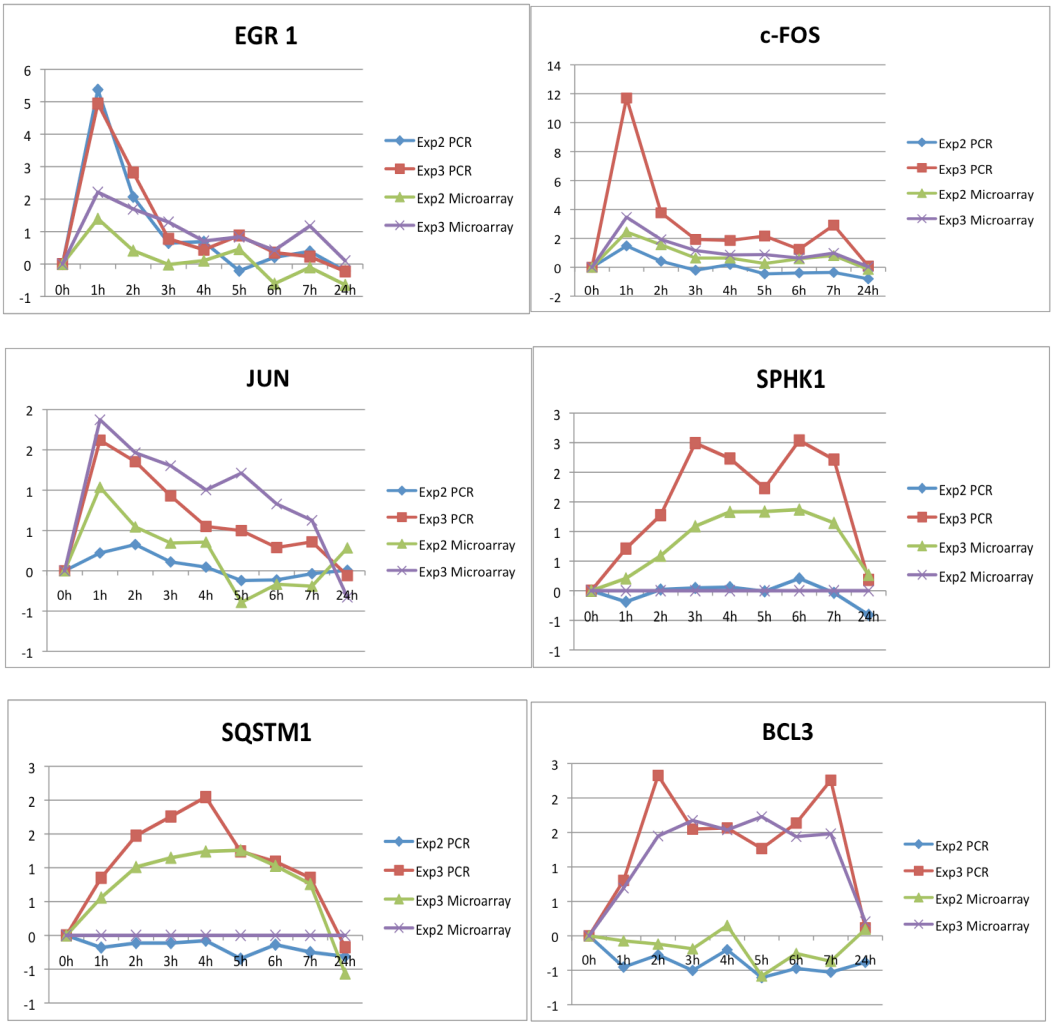


Figure 79 Sample comparison of genes between tumor cell experiments 2 and 3 in microarray and qRT-PCR confirming observations for tumor cell model

6.5.3 Gene silencing in tumor cells (GRN model validation)

Gene silencing in TC (MiaPaCa2) was performed using siRNAs against *EGR1, 2, 3 and 4*. Initial siRNA against *EGR2* was removed from the profiling experiment with stimulated stellate cell supernatant (PSC*) due to low initial transcript levels. siRNA efficiency was tested and confirmed at a concentration of 3nM and 10nM. Knockouts of each and all *EGRs* were evaluated to ensure gene specificity. TC (MiaPaCa2) gene knockouts were performed in medium and after exposure to both quiescent (PSC) and stimulated stellate cell supernatants (PSC*) with aforementioned controls at each of the two concentrations. Three separate sets of experiments were performed at different time points (4h, 24h, 48h) after exposure to stellate cells.

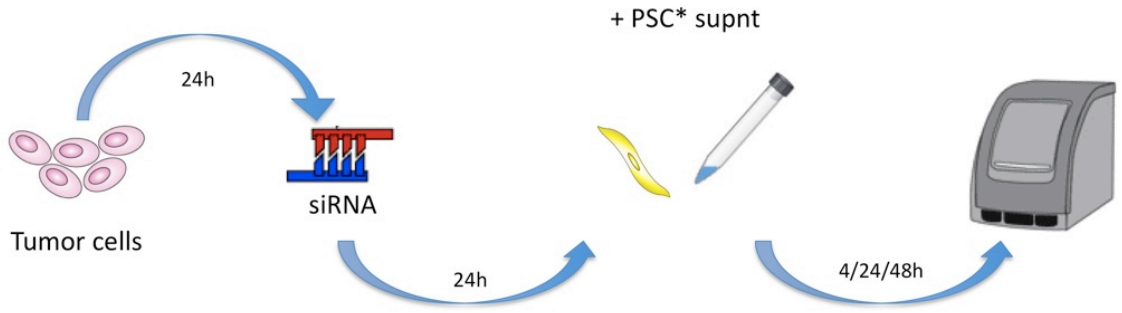


Figure 80 Time constraints in the knockdown experiments

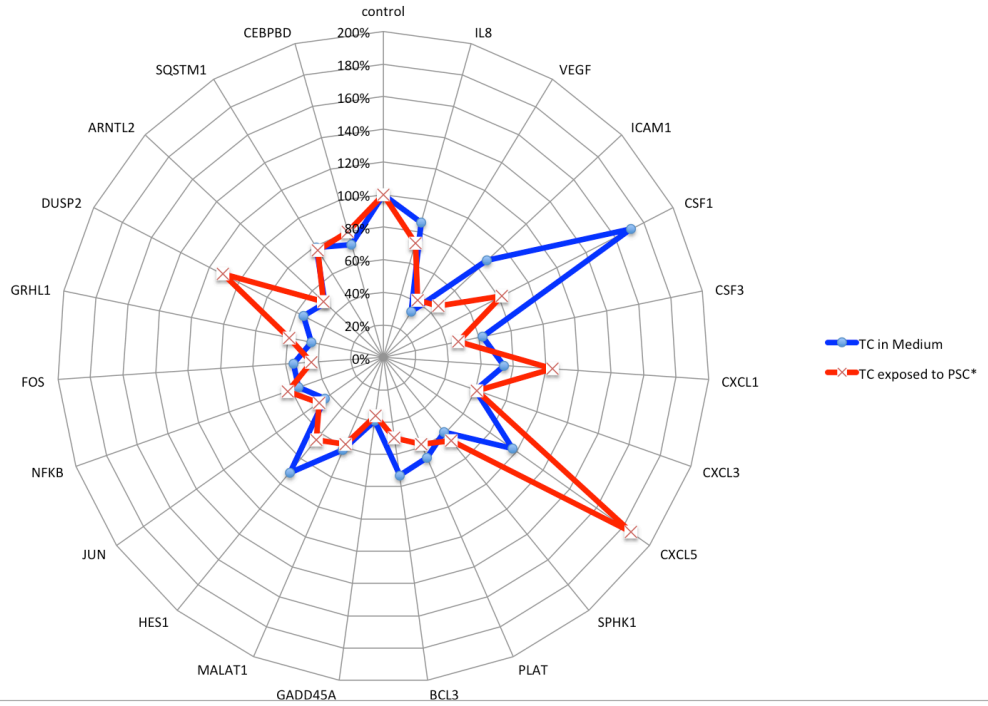
PCR evaluation was performed to confirm:

- siRNA *EGR 1-4* knockdown efficiency
- Knockdown effect on model genes including: *CEBPD, SQSTM1, ARNTL2, DUSP2, GRHL1, c-FOS, NFκB, JUN, HES1, GADD45A, BCL3, PLAT, SPHK1*, as well as a set of genes of interest including *MALAT1, IL8, VEGF, ICAM1, CSF1, CSF3, CXCL1/3/5*.

Figure 81 shows qRT-PCR measurements of selected genes of interest in MiaPaCa2 TC (transfected with siRNA *EGR1/3/4/1-4)* exposed to either: medium (control), or stimulated PSC* after 24h. While the *in silico* knockdown simulations were performed on a combined cluster of *EGR* genes, the *in vitro* experiments were performed against each *EGR* gene separately as well as in combination.

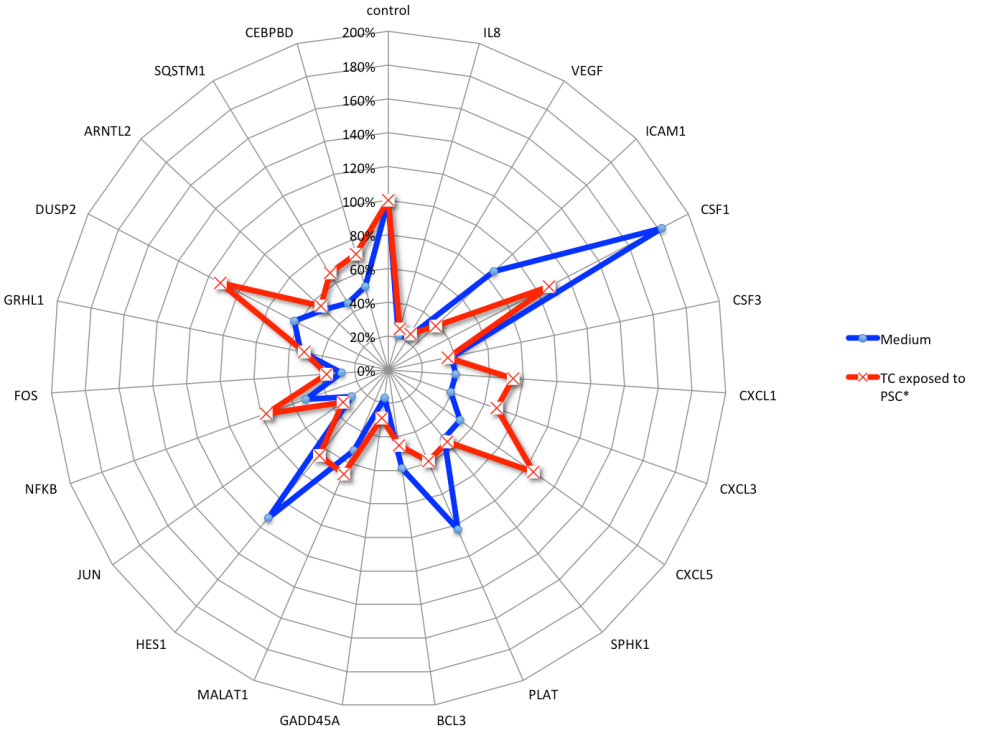
Gene transcript levels upon siRNA EGR1 knockdown of MiaPaCa2 tumor cells (TC) unexposed (medium) and exposed (PSC*) after 24h

(100% - mean control (non-transfected, negative) of TC exposed to PSC)*



Gene transcript levels upon EGR3 knockdown in MiaPaCa2 tumor cells (TC) unexposed (medium) and exposed (PSC*) after 24h

(100% - mean control (non-transfected, negative) of TC exposed to PSC)*



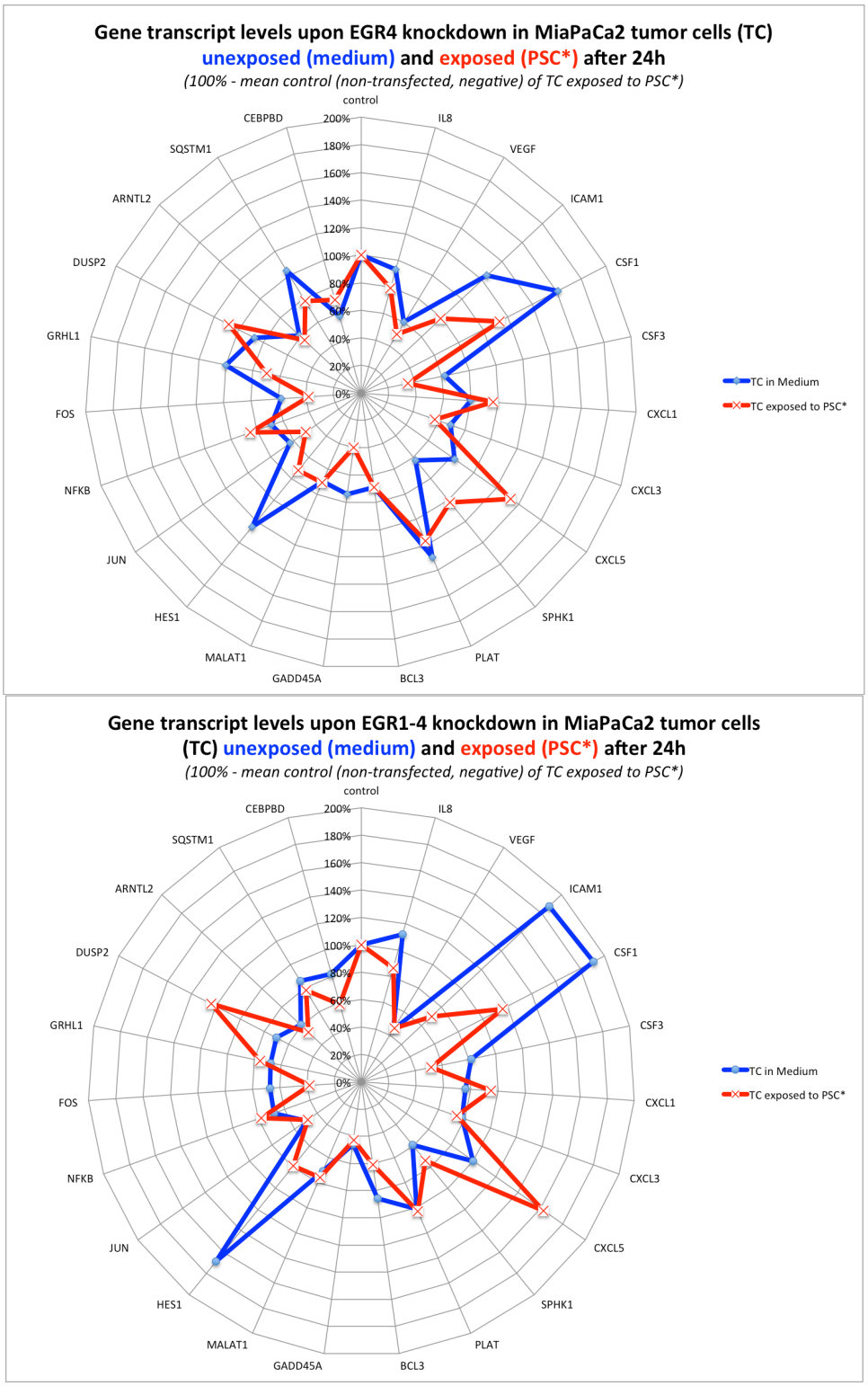


Figure 81 siRNA EGR knockdown effect in tumor cells (MiaPaCa2) in medium and exposed to stimulated stellate cells (measurements taken after 24h from exposure). Blue graph depicts transcript level of TC in medium, red - TC exposed to PSC*.

qRT-PCR analysis showed that the effect of *EGR* knockdowns was different for the various family members (*1-4*), that it depended highly on TC stimulation (medium vs. stimulated PSC), and was strongest for EGR3 gene knockdown in TC treated with PSC supernatants.

The effect of exposure was investigated separately by comparing expression levels in TC (MiaPaCa2) and PSC (both quiescent PSC, and stimulated PSC*) at three time points: 4h, 24h, and 48h from exposure (Figure 82). Due to material constraints only a combined knockout of EGR1-4 was performed here, which was suboptimal since the single-gene knockdowns showed greater efficacy. Profiled genes included: *EGR1, EGR2, EGR3, EGR4, GADD45A, HES1, NFKB, GRHL1, SQSTM1, BCL3, DUSP2, JUN, c-FOS, Sphk1, ARNTL2, CEPBD*. The results (Figure 82) showed that there is a significant difference in gene expression between the three different time points and forms of stimulation. The effect is therefore both stimulus- and time-dependent. The strongest spread between stimulus effects was achieved after 48 hours from siRNA transfection and was clearly seen for genes such as *BCL3 and DUSP2*. Unfortunately the results of this set of PCR measurements were not directly comparable with the results presented earlier (Figure 81) because samples were gathered from a preparation for a clonogenic assay, which as part of the protocol included the trypsinization of cells in order to detach them from the plate surface, this procedure significantly alters the signaling and resulting gene expression profiles in TC. Therefore not the knockdown effect on particular genes, but rather the overall differences arising from the combination of the knockdown, form of cell stimulus, and time of exposure are of interest in that visualization.

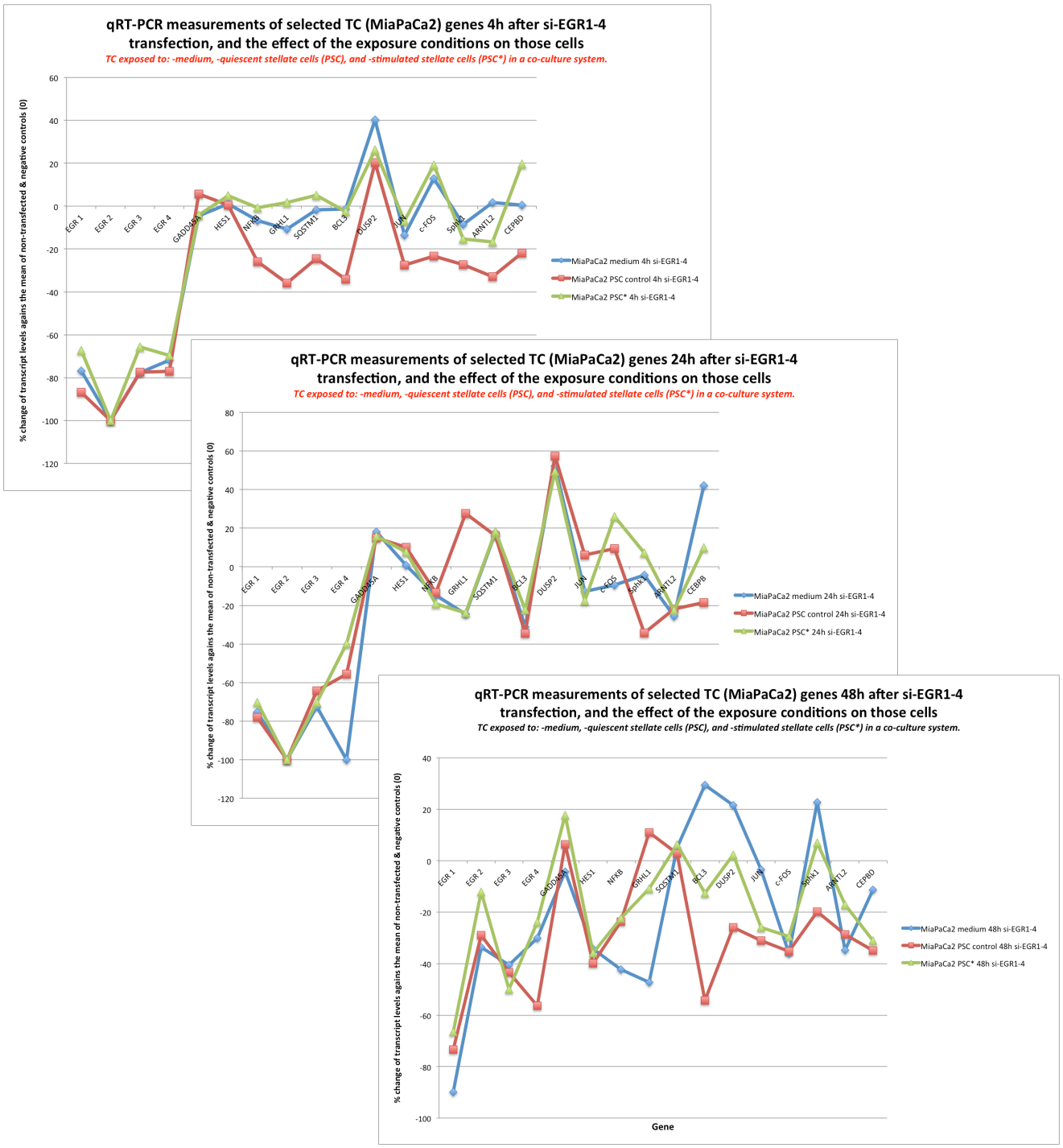


Figure 82 Comparison of gene expression levels in three conditions (medium, quiescent PSC control, and stimulated PSC*) for three separate time points 4h, 24h, and 48h

6.5.4 Predictive power of the tumor cell model

TC model was constructed using cells exposed to conditioned medium of stimulated PSC. *In silico* knockdown simulations were performed on a cluster of *EGR* genes (*1-4*), and the *in vitro* experimental evaluation with siRNA was performed against all *EGRs* separately as well as combined (siRNA *EGR1-4*).

We have shown in the qRT-PCR experiments that the effect of *EGR* genes knockdowns is different for the various family members (*1-4*), that it depends highly on TC treatment (medium, quiescent vs. stimulated PSC), and is strongest for *EGR3* gene knockdown in TC treated with PSC* supernatants. The effect is not only stimulus-, but also time-dependent. The strongest spread between stimulus effects is achieved after 48 hours from exposure to the corresponding supernatant.

Among all *in silico* predictions, of which validation was undertaken, we have failed to correctly identify the behavior of cluster 3 represented by *DUSP2* in MiaPaCa2 cell line. The simulations predicted an effect at both 4h and 24h, whereas the *in vitro* validation showed it to be relatively unchanging, and even slightly up-regulated at the measured time point of 24h (Figure 81), nonetheless the effect of the knockdown becomes visible after 48h, however due to the tripsinization of the cells in that culture it was impossible to state that this effect was due to the EGR knockdown. The remaining gene modules in the network of TC were validated confirming the predictions.

Interesting effect has been noticed for the set of additional genes of interest, containing *IL8, VEGF, ICAM1, CSF1, CSF3, CXCL1, CXCL3, and CXCL5*. Except for *CXCL5* all the other secreted factors showed a clear downregulation in gene expression, for some i.e. *CSF1* and *ICAM1* we saw a major difference between knockdown effect in TC exposed to medium (control) vs. PSC*, which was also dependent on which *EGR* family member was silenced.

6.5.5 Validation of knockdowns in the context of cell-cell communication

The organization of cell-cell communication in the PDAC microenvironment suggests high dependency of signals produced by each cell type on preceding stimuli from the other respective cell type. This in turn may mean that disturbing the intracellular GRN in TC may have a measurable effect on the expression of pre-stimulated PSC genes encoding soluble factors involved in the subsequent double-paracrine regulation.

In order to experimentally validate the established soluble factors forming the connections between both cell types, as well as identify the effect of the TC gene knockdown on those factors, a set of genes (Table 26) was investigated. The experimental data were gathered using qRT-PCR from samples of a co-culture of either uninitialized (PSC), or prestimulated stellate cells (PSC*) with tumor cells (MiaPaCa2, and Panc1) in medium or transfected with siRNA against *EGR* genes. Due to the strongest knockdown effect of *EGR3* gene siRNA shown earlier, this gene was the primary silencing target, additional measurements were performed with *EGR1-4* knockdown.

A total of 24 genes were investigated, 17 were measured in both cell types, 6 were identified as inducible in TC, and 3 in PSC, 1 gene was not expressed (*CXCR2*) in the given conditions. In fact additional FACS analysis showed that this receptor is present (data not shown).

AREG^{2}	CCR6$^{2\ TC}$	CSF1^{TC}	CSF3^{TC}	CXCL1$^{2\ TC}$
CXCL3$^{2\ TC}$	CXCL5	CXCR4^{2}	ERBb1(EGFR)^{2}	HO-1^{2}
ICAM1	IFNaR1^{2}	IFNaR2^{2}	IL8^{2}	LIF^{2}
PLAT	PPARa^{2}	PPARg^{2}	SHH$^{2\ TC}$	VEGF^{2}
IL6$^{2\ PSC}$	COX2 (PTGS2)$^{2\ PSC}$	CXCR2 (IL8R)$^{2\ none}$	ESR1$^{2\ PSC}$	

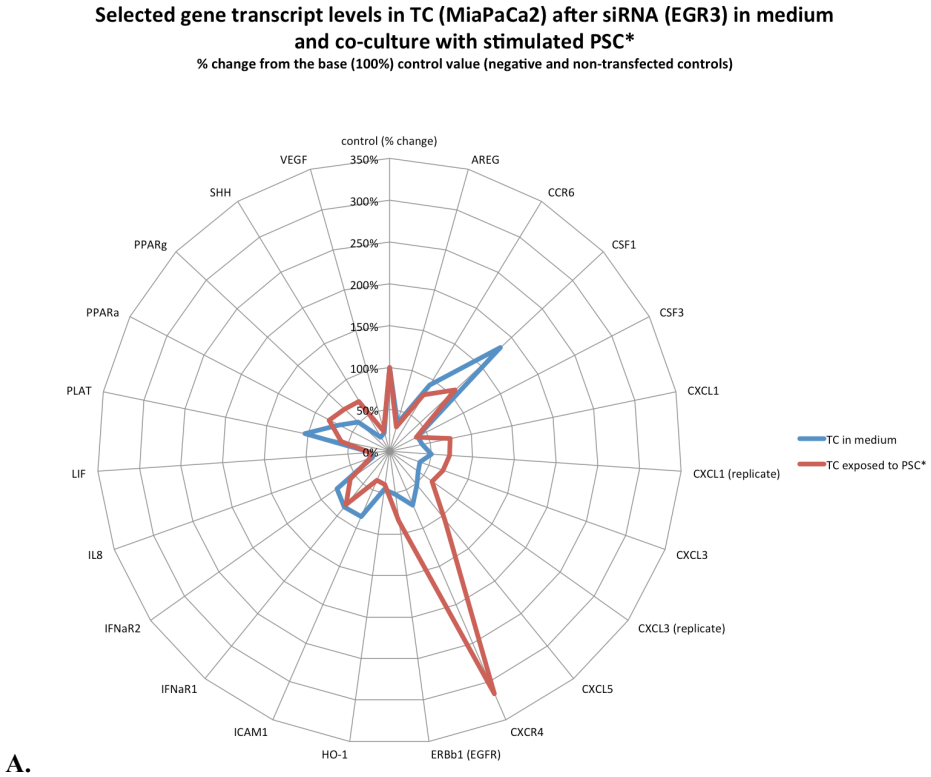
Table 26 Genes selected for co-culture evaluation. ^{2} - measured in both cell types, ^{PSC} - induced only in PSC, ^{TC} - induced only in TC.

MiaPaCa2 *EGR3* knockdown in TC cultured in **medium alone** caused a downregulation of all investigated genes except *CSF1 and PLAT* (Figure 83). CSF1 was not affected by the knockdown, and PLAT transcript level dropped by approximately 50% in TC **exposed to PSC***. *AREG, CCR6, HO-1, ICAM1, IFNaR2, and PLAT*, showed stronger downregulation **upon exposure to PSC*** than in medium alone. *CXCR4* showed a very strong upregulation upon exposure to PSC*, the same knockdown in medium-only caused it's downregulation.

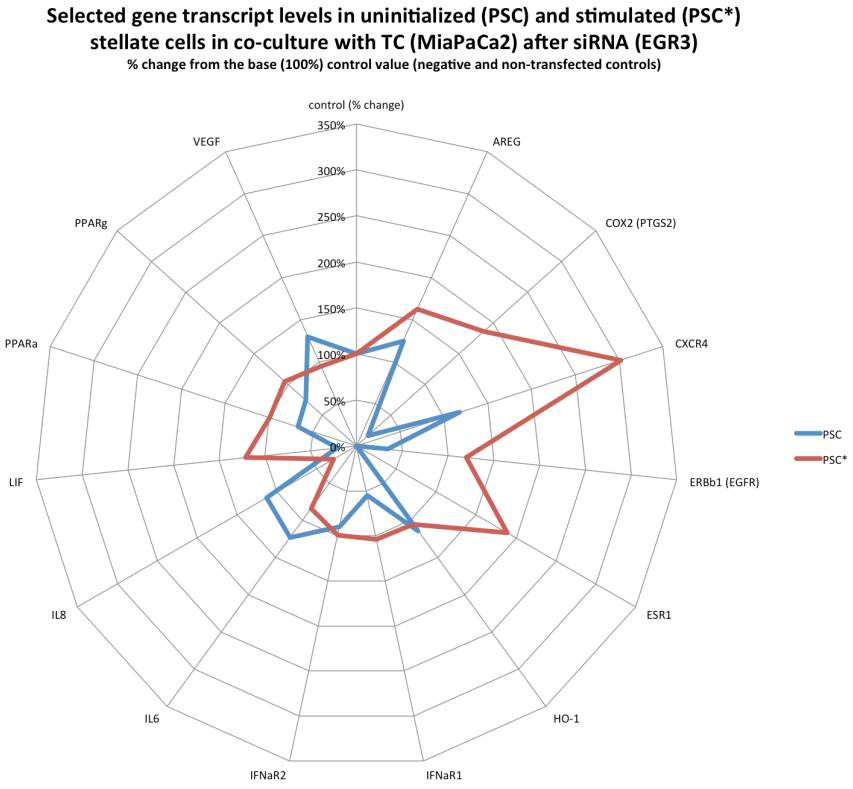
MiaPaCa2 *EGR3* knockdown in TC exposed to **uninitialized PSC** caused a downregulation (minimum 20% up to -80%) of *LIF, CXCL1, PPARg, PPARa, IFNaR1, IFNaR2, ERBb1 (EGFR), CXCR4, CCR6, SHH*.

Uninitiated stellate cells (PSC) were sensitive to tumor cell knockdown, which was revealed in significant downregulation of *PPARa, PPARg, COX2(PTGS2), EGFR, ESR1, IFNAR1, and LIF*. The same knockdown in stimulated stellate cells (PSC*) has had the opposite effect on the aforementioned genes.

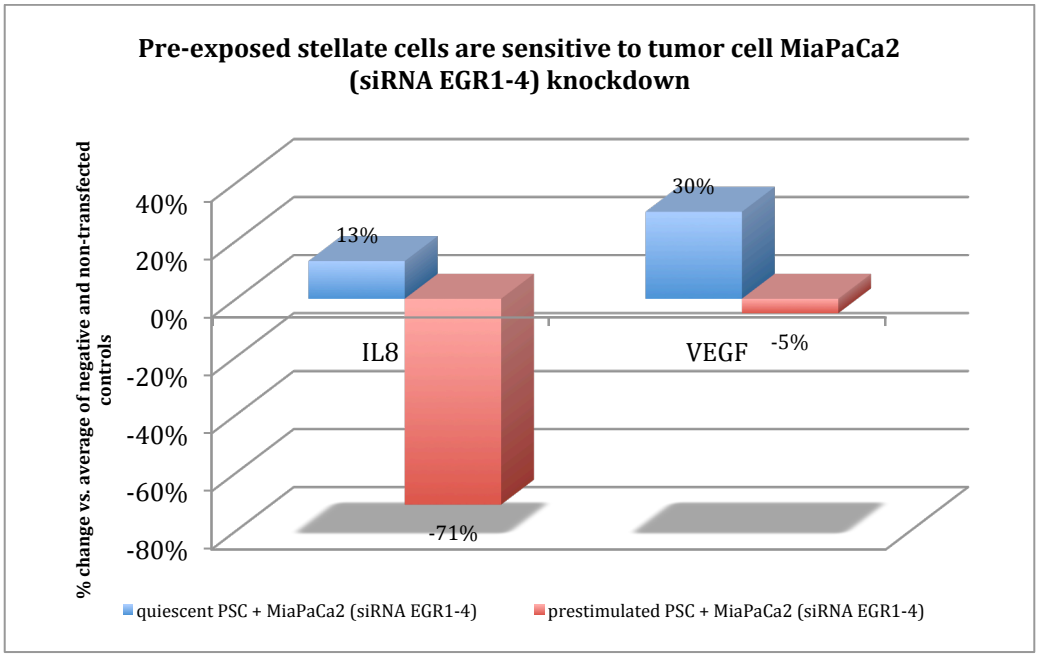
Stimulated PSC* were also sensitive to TC *EGR3* knockdown clearly showed as a downregulation of *VEGF* falling below the corresponding uninitiated PSC levels, strong drop in production of *IL6* and *IL8*. The measurement for *IL8* showed a 71% decrease when compared to the baseline control at 0, *VEGF* drop was much less significant and at 5% was well within the frame of experimental error and biological variation. Nonetheless, a look at the TC knockdown effect on *VEGF* and *IL8* in quiescent PSC revealed significant differences between the two forms of stellate cells.



A.



B.



C.

Figure 83 A. Tumor cells (MiaPaCa2 siRNA *EGR3*) in co-culture with uninitialized (PSC), and stimulated (PSC*) stellate cells. B. Uninitialized and stimulated PSC are sensitive to TC *EGR3* knockdown. C. TC knockdown *EGR1-4* has a strong effect on the behavior of stellate cells

7. Discussion

7.1 Summary

In the presented work we applied Systems Biology approaches to the analysis and modeling of transcriptomic data in pancreatic stellate and tumor cells. We identified the transcript-level dynamic changes in gene regulation of both cell types upon exposure to each other, as they may be occurring *in vivo*. By means of microarray analysis we were able to recognize that tumor cells are capable of stimulating stellate cells, and with the model of PSC we have shown a switch-like gene expression kinetic, which is responsible for PSC response to this stimulus. The fact that TC derived from a cell culture (previously unexposed to PSC) were capable of stimulating quiescent PSC suggested the presence of a sufficiently distinct and recognizable set of secreted proteins at a constitutive level in TC.

The subsequent analysis of TC experiments revealed that their responses are stimulus dependent i.e. quiescent PSC elicited a broad and unspecific response, while pre-stimulated PSC provoked a more focused riposte, and while in both cases TC were driven in many, often conflicting directions, only the latter stimulation produced a picture of progression towards survival with anti-apoptotic, pro-inflammatory, and pro-angiogenic signaling. With the models of both cell types we established the functional links between the genes of interest and general patterns of regulation underlying the responses to stimuli.

The signaling, which began with the initiation of the intercellular communication by TC carried on with PSC responses driving double paracrine feedbacks of TC, forming what seems to be a continuous system with abundant autocrine and paracrine loops between the cells, creating at the same time a rich and, at least transiently, stable environment. Models showed that removal of the sustained stimulation resets the gene expression to the initial state observed in each cell. Since the initiation of the signaling occurs through the recognition of TC by PSC, and further feedbacks seem to enhance the responses of both cell types, a reasonable conclusion is that the TC maintains a form of unstable control over PSC. Breaking down this self-propagating intercellular communication and initialization is of great interest as it may serve as a way to break the TC resistance to therapy which PSC are believed to convey, as well as their ability to progress towards metastasis. Using the derived models and the identified related extracellular signals we were able to identify a set of potential points of interference for the disruption of this communication in both cell types, and have successfully confirmed our *in silico* targets in TC with experimental validation.

7.2 Regulation of cell-cell interactions

As we have stated in the introduction, traditional methods of genome-wide expression analysis in PDAC have not been successful in translating their findings into therapeutical targets, and have failed to dissect the gene regulation in PDAC in a cell-specific manner. Throughout this thesis we have demonstrated a unified, streamlined approach to analyzing gene regulation in each cell type and its resulting intercellular communication. Our experiments were designed to mimic the *in vivo* condition, and the derived GRN's present a reasonable approximation based on experimental evidence. In this section we show how our findings on the intra- and intercellular level correspond to the existing overall knowledge of pancreatic cancer, how they fall within the existing biological framework of cancer biology, and which of them seem to be of importance in the investigated context.

7.2.1 Dynamic regulation of transcriptional responses in PSC

Transcriptional regulation of stellate cells in the experimental conditions is driven by the initial constitutive stimulus from TC, which is intercepted by PSC cells inducing early response genes such as the immediate early gene family of transcription factors *JUN, FOS, EGR1,* all within an hour from stimulation. *FOS*, which is stimulated by growth factors, was previously shown to receive this type of signaling via the Ras-Raf-ERK signaling pathway (Fitzner et al. 2004). This initial stimulus also drives delayed response genes including the *AP-1* transcriptional complex (*EGR3, ATF3, FOSB, JUNB*), as well as e.g. *PTGS2, MAP2k3*, all influenced the most within 2 hours after stimulation, and showing a second wave of activation at approximately 4 hours suggesting an autocrine feedback loop, at a point where the PSC is already secreting various proteins into the ECM. It is believed that the responses to proinflammatory cytokines depend mostly on the c-Jun N-terminal kinase pathway (MAPK), mediating the phosphorylation and activation of c-Jun, and p38 which contributes to both *FOS* and *c-Jun* gene induction (Ip et al.1998). Our model shows that the focal point of the reconstructed network is a hub, represented here by four genes: *APOL6, HIRA, MALAT1, and PHACTR2*, which integrates the initiating signals and diverts them, via its downstream targets, toward the remaining modules in the network, which are represented in our model by three "effector" clusters (containing genes which fall into GO categories of proliferation, differentiation, inflammatory state response, and interferon induced proteins e.g. *ZEB1, KLF7, IFIT1, IFITM1* etc.). All are continuously upregulated past 24 hours, suggesting the presence of a switch-like mechanism driven by the initializing stimulus and kept up by the subsequent autocrine and secondary paracrine responses.

The central hub genes are of special interest in this context as they form the weakest point in this network. *PHACTR2* as a phosphatase has a likely role in modulating the signal transduction through posttranslational protein modifications (reversible protein phosphorylation (kinases) and dephosphorylation (phosphatases) processes). That explains its wide signaling spread in the time series and also a global effect on other genes, as it may easily affect a wide range of signaling pathways in the cell. It has been recently associated with Parkinson's disease as a genetic risk factor (Wider et al. 2009). *MALAT1* is a metastasis associated lung adenocarcinoma transcript, a non-coding regulatory RNA, which has been found to have predictive power in early stage non-small cell lung cancer (Ji et al. 2003). Its function is otherwise unknown. *HIRA* - HIR histone cell cycle regulation defective homolog A is one of only few genes responsible for controlling the localization of histone H3 in the genome. Modifications of this histone have been shown to be important in the epigenetic regulation of gene expression (Goldberg et al. 2010). It is widely recognized that epigenetic modifications play a significant role in human cancers including PDAC (Sato et al. 2006). *APOL6* is a member of the apolipoprotein L gene family, which is pivotal for cholesterol transport. Changes in expression of this gene likely affect the cholesterol content in cellular membranes in turn modulating processes such as gene transcription and signal transduction.

While we could not confirm a transition to a steady state in the model, it does however show a transiently stable state that exists for as long as the information exchange with TC is taking place. The most likely culprit in this transition to this state is a module containing IFN-inducible gene. IFN has been previously shown to induce a state of senescence in PSC (data not shown, paper in preparation). The combined model and experimental observations support the hypothesis that signaling initialization in the ECM, unveiled in a desmoplastic reaction, occurs in response to TC presence. Application of our proposed knowledge-driven method of signal identification allowed us to establish a fairly small subset of soluble factors most likely driving the signaling towards the central hub genes in PSC including: unique soluble factors secreted by TC, PSC surface receptors, as well as potential membrane bound signal transducers. All of these factors seem to be affecting the formation of the GRN in PSC in both direct and indirect manner offering a way to interrupt further responses and stabilization of the system, even without directly affecting the intracellular GRN, which in case of PSC is difficult (inability to transfect PSC remains a major challenge).

7.2.2 A question of activation

Stellate cells, which have been pre-activated *in vivo* do not revert to an inactive state, but as our data showed they are also not being induced to overexpress the standard activation markers by our tumor cells. Fibronectin, laminin, α-smooth muscle actin were all unchanged in the time series. Even so, we still observed increased levels of ECM proteins such as collagen (e.g. *COL1A1, COL6A1* showing minor changes under experimental conditions over time, but remaining at high constitutive levels), along with an abundant expression of *ICAMs* (intercellular adhesion molecules), all accompanied by the secretion of a wide range of cytokines and chemokines - predominantly proinflammatory and proangiogenic. The signaling cues resulting from the initial stimulation seem to take PSC on the path of the aforementioned transition via the interferon-related signaling, as well as show potential for an induction similar to the one during chronic pancreatitis, when the cells become active via the release of pro-inflammatory cytokines (Aoki et al. 2005). It is now well established, from both clinical and experimental studies, that pancreatic necro-inflammation shown in alcohol related pancreatitis is associated with upregulation of cytokines such as *PDGF, TGF-β1, TNF-α, IL-1 and IL-6* (Apte et al. 2006) all of which we find in PSC upon stimulation with TC supernatant. It has also been previously reported that PSC are activated by each of the listed cytokines, therefore autocrine signaling likely plays a significant role in the stimulation of PSC *in vivo*, as well as in paracrine stimulation of surrounding stroma, apparently initializing a series of events altering not only the PSC, but also, and maybe more importantly, the microenvironment (McCarroll et al. 2003, Masamune et al. 2005).

The usual activation of PSC *in vivo* occurs in response to the signaling coming from the immune cells in the pancreatic microenvironment, and offers the standard response to inflammation. As results presented in the section on TC have shown, when treated with stimulated PSC supernatant they produce a range of factors, which are known attractors of immune cells. This suggests a connection between the current state of the PSC as seen in experiment 1 and their potential for induction upon exposure to immune cells invading the neoplasm at later stages of tumor and microenvironment development. This is in line with the findings of Feng et al. (2010), and we believe that the sequence of events in the pancreatic microenvironment may indeed follow a path, in which a form of PSC priming by other cells is necessary before a complete reactivation and progression towards migration by PSC might occur.

In contrast to what we expected, and what was previously shown in literature, our experiments did not yield any significant time-resolved, therefore inducible, changes in

expression levels of genes encoding metalloproteinases in PSC. While PSC have been previously shown to be an important source of MMPs in pancreatic cancer (Schneiderhan et al. 2006) driving tumor desmoplasia and promoting invasiveness (Ellenrieder et al., 2000), here we confirmed the presence of MMPs on a constitutive level only. As later analysis of TC showed, MMPs were upregulated (induced) in TC exposed to PSC. They were also constitutively highly produced by TC (including MMP7, 9, and 13, which we believe are vital for PSC initialization in our system). MMPs are significant in the context of tumor development because they may be used by the TC for ligand/receptor shedding in the process of evading the host immunological response, as well as for the disassembly of the ECM during the desmoplastic reaction. Lack of MMP induction in PSC in our experiment may be attributed to the specific experimental conditions and the fact that they were previously activated *in vivo*. On the other hand one has to keep in mind that the aforementioned publication operates with data gathered from a CAM assay, which is a non-mammalian system that may have a significant impact on the achieved results.

7.2.3 Antiviral responses of stellate cell

Modeling of PSC combined with cluster analysis has revealed 3 effector clusters including one containing interferon-induced genes. Interferons (IFNs) are proteins, which are usually produced by immune cells in response to the presence of pathogens e.g. viruses, bacteria, or tumor cells. Their main function is the initialization of the host immune responses to combat an infection. The initialization of IFN-dependent signaling in PSC suggests they may be recognizing TC and attempt to initialize the host immune response (Figure 84). From TC#3 we know that this response is suboptimal because it does not kill the TC, but it does drive the further increase of inflammation-related signaling, which may enhance TC ability to progress.

In terms of time-resolved gene expression, the IFN-induced genes are either driven by an autocrine feedback, as they become upregulated at the earliest 3h from stimulation, therefore falling into a time frame where the PSC already secretes factors into the microenvironment, or are secondary targets of genes induced by paracrine signals (Figure 84).

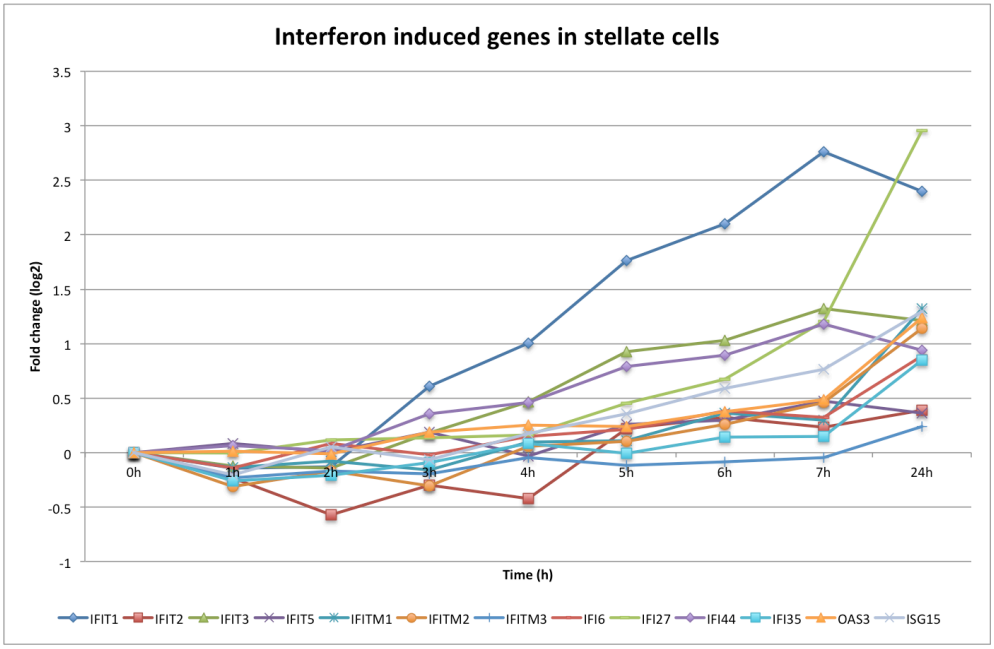


Figure 84 Interferon related genes found in the stellate cells experiment.

The expression of genes encoding IFN receptors in PSC (Figure 85) suggests that the initial stimulus is of paracrine nature and stimulates the interferon α, β, and γ receptor *IFNAR1* inducing the IFN dependent signaling cascades. The time resolved gene expression suggests that a secondary autocrine wave of signaling within PSC enhances those responses after 4h.

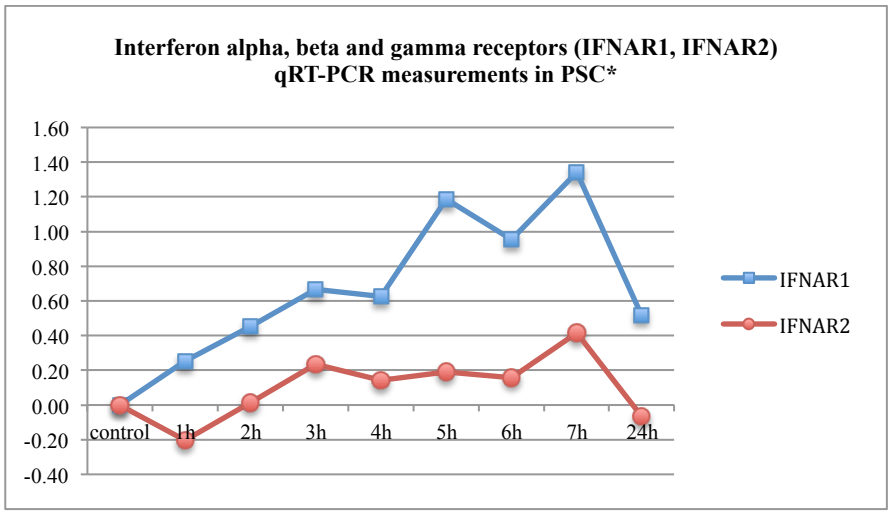


Figure 85 Time resolved gene expression kinetics of the two main interferon receptors in stimulated stellate cells (qRT-PCR)

It is widely recognized that the most common host response to viral infection is disclosed via a range of factors produced to combat it including their encoding genes e.g. *IL1B*, *IL6*, *IL8*, all of which are present in PSC in abundance (Figure 86).

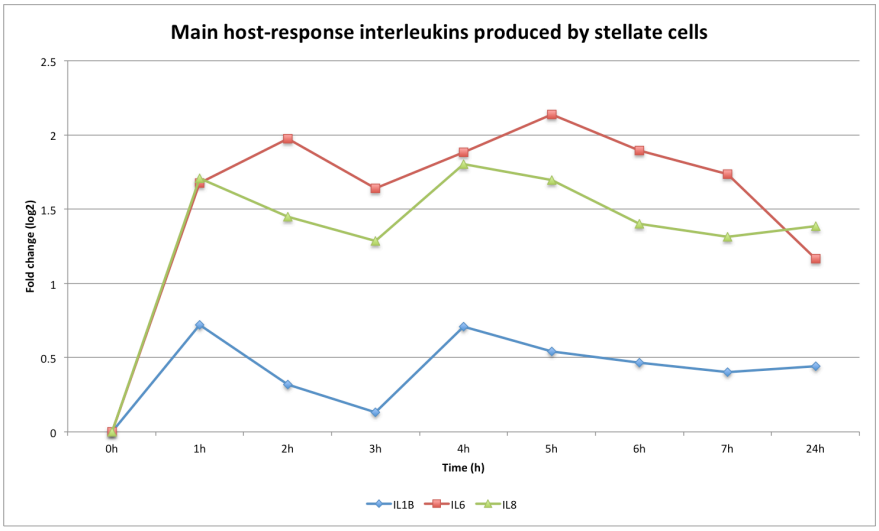


Figure 86 IL1B, IL6 and IL8 gene expression in stellate cells

Monsurro et al. have recently (2010) investigated a set of interferon related genes in PDAC tissue samples, and selected cell lines, and have been able to differentiate between two carcinoma subtypes – one permissive to viral vectors used for gene therapy, and another – resistant to viral infection. They conclude that the detection of these two phenotypes might help the selection of patients enrolled in virally mediated gene therapy trials (Monsurro et al. 2010). Since the set

proposed by Monsurro et al. is the most complete to date, we performed an overlap between their findings and our more specific time-resolved datasets in order to differentiate the signaling between the two cell types.

We used our filtered and ranked microarray data and compared them with Monsurro supplementary data, identifying an overlap shown in Table 27. Genes, which were missing from the Monsurro dataset were found in the unfiltered microarray data set with little to no changes over time, and therefore were removed from the comparison.

↗in PSC	↗in TC	Present and ↗in both	↗ in PSC ↘in TC	↘in PSC ↗in TC
IFI27	GBP1	IFIH1	ISG20	IRF1
IFI35	IFI16	IFIT3		
IFI44	IFI30	OAS3		
IFI44L	IFNAR1	PRKRIR		
IFIT1	IRF1	IFNAR2		
IFIT2	IRF6			
IFIT5	ITPKC			
IFITM1	JAK1			
IFITM2	PCK2			
IFNA10	STAT1			
IFNA14				
IL6				
IRF7				
ISG20				
ISGF3G				
MX1				
PRKRIR				

Table 27 Comparison of interferon-regulated inducible genes in PSC and TC experiments (upregulated ↗, downregulated ↘)

Out of 44 genes overlapping between the signature of Monsurro and our data – 23 were found in PSC only (TC time-resolved expression for those genes does not change significantly, and they are not present in our filtered dataset), 14 were only found in TC, and 7 showed expression changes in the time series of both. Further filtering for only upregulated genes reduced those numbers further to 17, 10, and 5 respectively.

IFNAR2 receptor provided additional insights showing a different expression profile for tumor and stellate cells in response to the treatment. TC response for the receptor was autocrine (with an inhibitory effect of the extracellular stimulus in the early hours), whereas PSC were stimulated by both, the paracrine and autocrine signals, suggesting that the earliest stimulus came from the initializing signals provided by the TC supernatant in the first experiment.

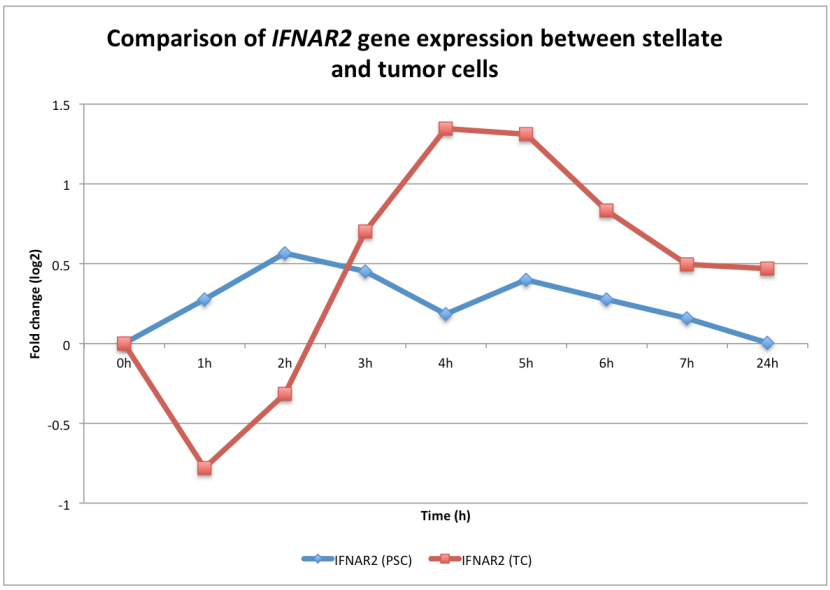


Figure 87 IFNAR2 expression in tumor and stellate cells (microarray data)

Interestingly the TC experiment did not show any upregulation of interferon production in those cells, IFN-induced genes such as *OAS1* were downregulated, there was no significant production of *IL6*, and the only interleukin related to the host immune response was *IL8*, which was not surprising since it may be stimulated in various ways. *STAT1*, which is a known signal transducer in the *IFNα, β, and γ* pathways, was insignificantly upregulated in TC (< 0.5 $\log_2$ fold change). All this seems to point to the fact that PSC initiate the antiviral response, and subsequently their secreted factors affect IFN-related signaling pathways in an autocrine manner in PSC themselves driving the proinflammatory responses. The switch-like behavior of the effector clusters suggests a state transition occurring in the PSC, which is contingent on TC presence.

7.2.4 Autocrine and paracrine signaling resulting from PSC stimulation

GRN reconfiguration in response to perturbation of PSC by TC is a dynamic process, which has an outcome in the form of a cellular response with a predominantly proinflammatory profile arranged into 4 clusters containing significantly upregulated genes encoding soluble factors (Table 9). We also investigated soluble signals apart of the standard ranked list of genes, but in association with the underlying TC GRN, and identified a unique subset of factors likely affecting the TC network (Figure 74). The autocrine component of this second phase of signaling is just as interesting as the PSC initialization because here we identified factors previously reported for pancreatic cancer in general e.g. *PPARG*, and we can confirm their presence in the environment. Our analysis suggests that the *PPARG* signaling is enhanced in both cell types, beginning with the initialization of PSC by TC through recruitment of *PPARG* receptor, PSC seem to respond via *ANGPTL4*. The presence of *PPAR* receptors on TC surface suggests auto- and paracrine signal crosstalk. Interesting is the time-resolved expression of the tumor cell *PPARG* gene, which seems switch-like, and we see a significant peak already 1h after stimulation of TC by PSC supernatant (Exp.3). Then the expression level remains virtually unchanged throughout the duration of the experiment.

A review by Eibl (2008) shows clearly that there is a lot of controversy surrounding the PPAR-γ proteins, but *in vitro* evidence seems to suggest that ligands of this receptor may have a tumor promoting effect by enhancing tumor angiogenesis through the stimulation of VEGF production (Eibl, 2004 and 2008; Margeli et al. 2003, Biscetti et al. 2008). *PPARG* signaling in stellate cells has also been implicated in the maintenance of the quiescent phenotype of hepatic stellate cells HSCs (Galli et al. 2000), however whether this applies to PSC is debatable.

Because of this redundancy of *PPARG* in both cell types, we don't believe that it is such a great experimental or therapeutical target. Our signaling schema (Figure 74) with the unique factors offers a much more precise and cell-specific set of factors.

Other potential autocrine/paracrine components shared in the microenvironment include *EGFR, PLAUR, PTGER4, PTPRE, PTPRR*. Some of those will be discussed in more detail later.

7.2.5 Tumors as wounds that never heal

PSC response to initialization suggests that they perceive TC presence as dangerous, a finding supported by the predominantly proinflammatory signaling (inflammation is a significant player in supporting PDAC progression), initiation of IFN-induced genes in the aforementioned antiviral response, as well as induction of genes, which enrich the "response to wound healing" GO category (in fact the main biological process identified in PSC (Table 12)).

We related these findings to our previous project investigating the transcriptional regulation of migrating keratinocytes in the process of wound healing (Busch et al. 2008), and interestingly keratinocytes seem to share many regulatory features with the tumor-associated fibroblasts (PSC) and only some with pancreatic TC. A rather surprising finding, due to the context and cell type, we expected to see more comparable results between keratinocytes and TC rather than PSC (fibroblasts).

Genes identified as core elements in the transcriptional regulation of keratinocytes included: *FOS, EGR1, EGR3, and PTGS2*. The first three (*FOS, EGR1, EGR3*) were all part of the GRN of PSC and TC, however only the PSC show a significant change in *PTGS2* transcript levels. *PTGS2* is a gene encoding for prostaglandin-endoperoxide synthase 2 (previously reported as overexpressed in PDAC by Yoshida et al. 2005), and in the model of HGF-induced keratinocyte migration it was used as an indicator of migration in conjunction with *AKAP12*.

AKAP12 (encoding A-kinase anchor protein-12) was shown to negatively regulate *PTGS2*, and upregulate *FOS, EGR1, EGR3,* which promote migration *in silico*. Here, we see that *AKAP12* levels are not changing in PSC, but with the knowledge-driven approach we were able to establish that it is a potential signal transducer for TC constitutive factors into PSC GRN (data not shown). *AKAP12* however seems to be of particular interest in light of recent findings by Mardin et al. 2010, where it has been shown that increased mRNA levels correlate with diminished invasive and metastatic capabilities of tumor cells. Tumor cells in our experiments show a downregulation of *AKAP12,* which would correlate with their capacity for invasiveness.

It is however the *EGR* gene family that seems to be of pronounced importance in the investigated GRN in conjunction with wound healing. We have previously shown that *EGR1* and *EGR3* are important for the core network in HGF-induced keratinocytes. In the presented project the high expression levels of both EGRs convinced us to investigate their regulatory effects, and we showed with our TC#3 model that they are in fact the weakest point and a potential target of interference, which is not the case for PSC. This implies the question of why the same genes have a different impact in each of the investigated cell types. We previously showed that the dynamical

time-resolved gene expression changes translate well to the level of proteins, a comparison of *EGR1* and *EGR3* genes between the three pertinent experiments: HGF-induced keratinocytes, PSC, and pancreatic TC reveals differences in transcript levels and distinct time-resolved profiles in each cell type, which suggests that while present in the cells, the stimulation and responses of those genes are significantly different between them.

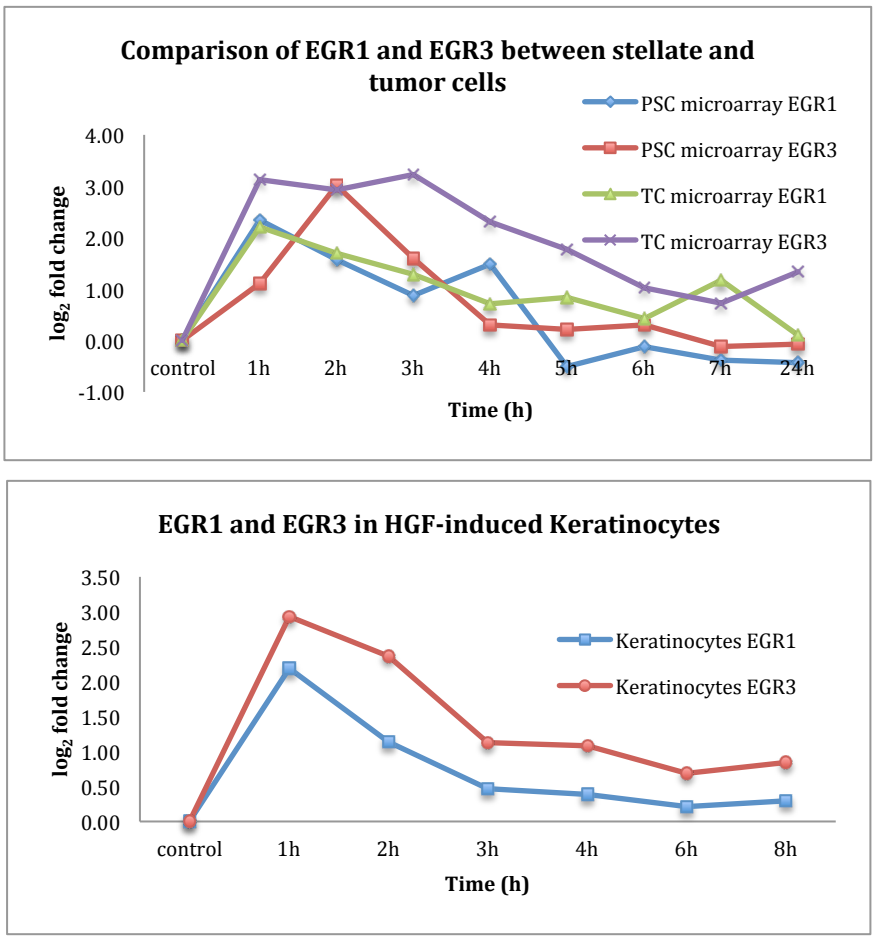


Figure 88 *EGR1* and *EGR3* gene expression is different between the different cells and conditions (microarray data)

A closer inspection of *EGR1* and *EGR3* genes (Figure 88) showed that their time-resolved behavior was condition and cell dependent, and especially *EGR3* showed a significantly altered kinetic in TC. An attempt to overlap *EGR1* and *EGR3* profiles between HGF-induced keratinocytes, PSC and pancreatic TC revealed that *EGR1* expression correlates between keratinocytes and PSC, while *EGR3* profile was different for each of them. *EGR3* was especially interesting in this context as in keratinocytes the peak of stimulation for this gene was around 1h (early response genes) after which we saw a shutdown as kinetics returned to 0. In pancreatic TC

that initial peak was extended to nearly 4 hours, whereas PSC had a focused response with a peak at 2h. The extended signaling range of *EGR3* may be attributed to an autocrine feedback loop, which drives the further activation of this gene in TC, but it may also be a delayed primary response. *EGR* responses differentiating the stimulus effect is not surprising as those TFs can induce different genes through the different cis-regulatory elements. This dependence of gene expression on extracellular stimulus is significant as it suggests that we should be able to attack intracellular signal transduction through alteration of extracellular stimuli.

An additional small subset of genes induced in PSC, TC, and HGF-induced keratinocytes included: *CXCL3, PLAU, PLAUR, and CCL20*. Those genes form part of the feedback-rich signaling between PSC and TC in the pancreatic microenvironment and correlate with wound healing. *PLAUR (CD87)* is a receptor overexpressed in multiple malignancies, where apparently the TC hijacks the plasminogen activation system for its own purposes. It is widely expressed and existing literature evidence suggests its involvement in regulation of cellular adhesion, cell motility, angiogenesis, tumor invasion and metastasis (Kjøller, 2003; Ploug et al. 2002 and 2003; Yimin et al. 2003). The presence of *PLAU/PLAUR* signaling between PSC and TC correlates well with factors we identified in TC including *ICAM1* (intercellular adhesion molecule), and *CEACAM1* (cell adhesion molecule), all transducing signals affecting the TC GRN. *PLAUR* has been shown in a mouse model to be positively correlated with tumor initiation and growth (Ploplis et al. 2007). Its ubiquitous presence on PSC and TC, as well as apparent significance in the regulation of the underlying GRN in TC suggest that it may offer a potential for common stromal targeting, but it's not cell-specific.

All this suggests that the initialization of PSC is not dissimilar from wound healing, at least on the transcript level, and keeping in mind the work of Xu, and Feng, it would seem that the signaling may preserve the outcome – migration – although it is likely that additional interactions are necessary to push the stellate cells into that direction, possibly with immune cells, which are ubiquitous in the PDAC microenvironment, and perhaps an EMT transdifferentiation event is necessary for PSC to achieve that state.

7.2.6 Autocrine and paracrine signaling driving TC stimulation

Stimulation of TC can be divided into two phases, first, the direct stimulation by factors released by PSC in experiment 1, second, an autocrine feedback loop in TC. We reduced the complex interactions into a subset of 9 unique PSC factors affecting the formation of the TC GRN (Figure 74), and interestingly show that the leading paracrine signaling is executed not as one might expect by a factor from the list of top inducible e.g. proinflammatory cytokines *C-X-C*, or interleukin *IL8*, but rather via a set of less expected genes including *TIMP3, IL7R, ESM1, EBI2, CCL20, DKK3, IL6, TFPI2, and STC1*. All of the aforementioned chemokines act here by enriching the complex microenvironment of both cell types, and form a background most likely necessary for the establishment of the GRN, but they do not seem to directly drive the GRN formation. Investigation of the dynamics of extracellular signals shows that while there are genes, which are upregulated past 24 hours, the majority of the ranked secreted factors show a slow decline in their expression levels at the last (24h) time point. This, previously mentioned, overall lack of a permanent switch in TC may mean that these interactions are insufficient to drive TC in the direction of a specific cell-fate decision.

Nonetheless, following the many abundant environmental cues we still arrive at a conclusion that TC in experiment 3 are driven in the direction where the role of inflammation is to promote cancer cell proliferation and stromal/matrix degradation. First and foremost via the pro-angiogenic and proinflammatory cytokines *CXC* along with *LIF*, and interleukins, suggesting that neovascularization is a vital step in progression prior to rapid tumor growth. Pro-angiogenic and inflammation-related chemokines identified in our TC dataset include *CXCL1, CXCL2, CXCL3, CXCL5, CXCL6, CXCL7,* and *CXCL8,* and mediate angiogenic activity through engagement of the common receptor *CXCR2 (IL8RB)* (Romagnani et al. 2004). *CXCL1/3/4* and *5* also, as shown before, overlap with the signals produced by PSC. *CXCL12* in turn mediates angiogenic activity through the *CXCR4* receptor (Liang et al. 2007) in an autocrine feedback in TC. *CXCL1* and *3* have also been shown to exert autocrine control over neoplastic cell proliferation (Richmond 1986, Coussens et al. 2002). *CXCL2*, 5, 6, and 10 all have pleiotropic effects ranging from regulation of proliferation, neutrophil activation, to modulation of adhesion molecule expression. Apparently the supernatant of stimulated PSC causes the destabilization of the fragile immune homeostasis, which normally consists of a succession of pro- and anti-inflammatory signals. Loss of the anti-inflammatory signals is known to lead to chronic inflammation and proliferative signaling (National Cancer Institute Website, accessed Dec.2010).

In 2008 Zhong et al. developed an *in vitro* model, which successfully recapitulates features of lung tumorigenesis *in vivo* to investigate mechanisms by which stromal cells regulate the biological properties of lung adenocarcinoma cells. They made significant discoveries including:

- TC are capable of chemoattracting stromal cells by mimicking inflammation and angiogenesis;
- Co-culture of tumor and stromal cells enhances TC proliferation and clonogenicity;
- Chemokines upregulated *in vitro* (*CCL2, VEGF, CCL5, CXCL1, CXCL2, and CCL11*) correlate with lung tumors in mice and humans;
- CXCL1 secretion was enhanced in the cocultures, and CXCR2 inhibition attenuated cell-cell interactions;
- TC secrete a wide range of factors supporting their development and progression including clusterin and LIF, which are prosurvival molecules; CXCL1, CCL3, LIF, IL-18, TNFα, and IL-15 which are proinflammatory molecules; and CXCL1, CX3CL1, VEGF, VCAM1 which promote tumor angiogenesis.

Comparing those findings to our data we see that much of the interplay between the various cells is similarly modulated in the PDAC tumor-stroma relationship. We identified similar signaling in the PDAC microenvironment using gene ranking and GO analysis, but the factors were often secreted by a different cell type e.g. Zhong et al. have shown lung cancer cells to produce LIF, a pro-survival and proinflammatory molecule, which we show to be a significant regulatory element produced by the PSC in response to stimulation by TC supernatant.

In addition to the aforementioned chemokines, we also identified various interleukins known to mediate the inflammatory gene expression effects. The supernatant of stimulated PSC contains among others *IL1, IL6, and IL8*. Interleukin 8 is actually produced by both, TC and PSC, and is known to amplify inflammation, activate p38 and ERK1/2 MAPK pathways, induce *NFκB* through TRAF6-dependent pathway, and induce c-Jun N-terminal kinase (*JNK*) of the mitogen activated protein kinase pathway (*MAPK*) in a dose- and time-dependent manner (Manna et al. 2005). IL6 produced by PSC and received via IL6R in TC leads to the activation of *NFκB*, and a crucial step of the induction of ICAM-1, which is an intercellular adhesion molecule, also inducible by IL1 and TNF. In fact IL6R receptor on TC initiates multiple effects. The IL6-IL6R complex activates signal transduction pathways involving a protein tyrosine kinase and serine/threonine kinase(s), which are required for the subsequent transcriptional activation of a set of genes including immediate early genes such as *JunB* in a variety of cells, and tissue specific

genes such as acute phase reactants in hepatocytes. TNF related signaling is unveiled via *TNFAIP3* which is induced by multiple stimuli, including the proinflammatory cytokines TNF and IL1, and trigger the toll-like pathogen recognition receptors (Verstrepen et al. 2010).

Not much is known about the remaining unique PSC signals driving TC, including: *TIMP3*, which is a tissue inhibitor of metalloproteinases that inhibit matrix metalloproteinases. It blocks the binding of VEGF to one of its receptors, VEGFR2, and inhibits downstream signaling and angiogenesis. Alterations in the *TIMP-3* have been found to contribute to the tumorigenesis of pancreatic endocrine tumors, causing tumor-specific loss or strong reduction of TIMP-3 protein expression (Wild et al, 2003)(after copewithcytokines.de). *IL7R* is an interleukin receptor, which has not been linked to pancreatic cancer previously. *ESM1* is also known as endocan and was identified by Lassalle et al. (1996) who described its constitutive expression in human umbilical vein endothelial cells and constitutive expression in human lung. *EBI2* is an Epstein-Barr virus-induced G-protein coupled receptor 2, which is interesting in light of the anti-viral response identified by us in PSC. *CCL20* is a strongly chemotactic protein for lymphocytes, which also weakly attracts neutrophils. *DKK3* was shown involved in embryonic development through its interactions with the Wnt signaling pathway. *TFPI2* may play a role in the regulation of plasmin-mediated matrix remodeling. Finally, *STC1* stanniocalcin was linked to increased metabolic rates, and with altered expression to ovarian and breast cancers.

7.2.7 Dynamic regulation of transcriptional responses in TC

Unlike in PSC, we could not identify any potential state transitioning set of genes in TC. Only few genes are showing upregulation past the experimental window e.g. CFB, C3 complement system members, but they do not determine cell fate. This suggests that the TC→PSC→TC interaction axis drives TC-GRN into an unstable condition, which is maintained for only as long as the TC has an immediate contact to PSC. Consequently, once the communication is interrupted cells return to their respective steady states (cf. 6.2.8.8.1).

In silico simulations and knockdowns (cf. 6.2.8.8.2), have provided two points of interference in the underlying GRN: *EGR* genes and *CEBPD*. We showed that the strongest single-module knockdown effect is achieved with the *EGR* genes downregulating the system responses. However unlike in stellate cells the system-wide effect is less pronounced, which may be attributed to the strong transcriptional deregulation of TC, where multiple TF share common targets, as well as to the overall ambiguity of signaling with many additional shifts and delays in the expression profiles. A system-wide effect was achieved through a double-module knockdown (*EGRs* and *CEBPD*) effectively destabilizing the entire TC network. This finding reaffirms the importance of the gene selection. Unfortunately does not seem to be applicable in practical terms, as a double knockdown of this art is highly cytotoxic and cannot be validated with gene silencing. Targeted *in vitro* knockdowns have allowed us to differentiate between the *EGR* gene family members attributing the strongest influence on the other network modules to siRNA against *EGR3* (as measured by qRT-PCR), confirming our *in silico* predictions. An interesting finding in the context of the TC-PSC interactions was offered by an invasion assay, which was established as a co-culture system and used in the presented work for qRT-PCR validations. The biological readout of this assay (data not shown) showed that TC *EGR* knockdown has anti-invasive effect on TC. Since *EGRs* are at the center of the TC GRN, it appears that breaking them down - breaks down the semi-homeostatic, sustained PSC-TC state, and affects the TC capacity to migrate (invasion assays confirmed TC migration reduction of up to 90% upon knockdown, data not shown).

7.2.8 Intracellular dynamics of gene expression in tumor cells

Inflammation-related signaling is a dominant feature of the PSC-TC microenvironment. The local production of cytokines not only stimulates stromal cells in the surrounding microenvironment, but also likely attracts circulating immune cells. In this context, it is not surprising that we find *NFκB* as one of the most enriched signaling pathways in experiment 3 TC. A clear link between *NFκB* and cancer lies in the inflammation as a vast majority of inflammatory genes rely on the activation of *NFκB*. First, it is required for the activation of the innate and adaptive immune responses; second, it is directly connected with cancer through its survival-switch decisions. Although we see a significant upregulation of inhibitors of the *NFκB* pathway such as *NFKBIA*, *NFKBIZ* (previously reported as a significant marker among a six-gene signature by Stratford et al. 2010), others such as *NFKBIL2* are downregulated. *NFKB2* itself sees a peak at a 5-hour time point, and a significant upregulation of *RelA*, and *RelB* has been noticed. Additionally, *TNF* signaling, which we see in tumor cells can be either activating or inhibiting the *NFκB* pathway via:

- *TNFRSF9* (interacts with TRAF adaptor proteins shown to transduce the signals leading to the activation of *NFκB*),
- *TNFRSF21* (shown to activate NFκB and MAPK8/JNK, capable of inducing cell apoptosis through its death domain that interacts with TRADD protein), *TNFAIP2* an angiogenic factor,
- *TNFAIP3* (shown to inhibit TNF-induced NFκB-dependent gene expression by interfering with RIP- or TRAF2- mediated transactivation signal – an inhibitor of programmed cell death (Heyninck et al. 1999).

Additional identified signals acting on NFκB include *IL1R, TLR, CD95L (fasL), IL8, NOD2, SQSTM1, ELL2*, all present in abundance in tumor cells.

Toll-like receptor signaling is especially interesting in this context as it interconnects pathways induced by pro-inflammatory proteins such as *NFκB, TNF, and TRAF*, but also due to it's involvement in pathogen-recognition. *IRAK2* (*IL1R*-associated kinase) is upregulated in both TC and PSC, and is essential for sustaining *TLR*-induced expression of genes encoding pro-inflammatory cytokines and activation of the NFκB, and MAPK. IRAK2 binds to the IL1 type I receptor following IL1 engagement and while it is not necessary for activation of the initial signaling cascades, it has been shown to be indispensable for sustaining *TLR*-induced expression of those genes and pathways (Keating et al. 2007). Another potential point of interest in the inflammatory signaling is disclosed in the form of the HSP70 proteins (molecular chaperones) of

which *HSPA6* is highly upregulated in our tumor cell dataset (others include *HSPA1B, HSPA1A, HSPA4*). HSP70-induced proinflammatory cytokine production is mediated via the MyD88/IRAK/NFκB signal transduction pathway (Rohde et al. 2005, Asea et al. 2002). HSPs are known to be upregulated in highly drug resistant pancreatic cancer as well as in other types of carcinomas. Members of this family promote cancer cell growth by distinct mechanisms. There is an ongoing effort by many laboratories in pursuing HSPs (Ito et al. 2003, Galluzzi et al. 2009, Rérole et al. 2010) and TLRs (O'Neill et al. 2008) as possible cancer therapy targets, and as it seems with good reason.

The proinflammatory response in TC is also exposed in the context of one of the most upregulated genes in our dataset, which is also part of our transcriptional network: *CEBPD*. The CCAAT/enhancer binding protein (CEBP) family of transcription factors includes five genes, and interestingly one of its members – *C/EBPβ* – has been recently linked as a significant player in the transcriptional network for mesenchymal transformation (EMT) of brain tumors (Carro et al. 2010). *CEBPB* shares promoter occupancy with *CEBPD* as was shown with CHiP analysis by Carro. It also binds to the *IL1* response element in the *IL6* gene, as well as to the regulatory regions of several acute phase and cytokine genes, making it an important player in the regulation of genes involved in immune and inflammatory responses. Its distinct expression profile, and significant upregulation have ignited our interest and it turned out to be a very important regulator in the transcriptional network of tumor cells responding to PSC stimulation. A combination *in silico* knockdown of clusters 1 (*CEBPD*) and 10 (*EGR* genes) is the only one that causes a system-wide breakdown of the network in tumor cells.

The inevitable outcome of inflammation is the recruitment of immune cells into the neoplasm, a fact, which caught the attention of Feng et al. 2010. Although tumor associated macrophages (TAMs) are able to kill TC when activated by IL2, IL12 or IFN, they also produce a host of compounds – angiogenic factors, growth factors, proteases and cytokines – that either contribute to cancer progression or blunt the anti-tumor response. As has been shown, the infiltrating immune cells are in fact incapable of destroying the tumors, instead amplifying the inflammation, which supports the TC development (Feng et al. 2010). Infiltrating macrophages have also been shown to amplify the expression of IL8 (abundant in our PSC/TC experiments) in the microenvironment of non-small lung cancer (Chen et al. 2003). The infiltration of tumors with TAMs has been earlier shown to correlate with poor prognosis in several cancers and recent studies suggest an emerging role for TAMs in the remodeling of the tumor microenvironment to

support growth and metastasis and support the concept of modifying TAMs responses as therapeutic approaches (Mantovani et al. 2006, Robinson-Smith et al. 2007).

While the tumor-associated macrophages (TAMs) are outside of the scope of the presented experimental work, they cannot be disregarded in the context of pancreatic microenvironment, and while we do not cover those interactions directly, we do identify early signs of this process in our analysis. IL6 secreted by PSC, and CSF1 (M-CSF), which we identified as secreted by TC, are known to skew monocyte differentiation towards the macrophage lineage (National Cancer Institute Website, accessed Dec.2010). *CSF* specifically, is necessary to recruit mature macrophages into the neoplastic tissue. CSF1 and CSF3 have been reported to positively affect proliferation in TC. Other genes falling into this category found in TC experiment include *CCND2, HES1, FGF18, IL15, SPHK1, TGFBR1, and CXCL5*. The most interesting in the context of immune evasion may be *CSF3* (G-CSF granulocyte colony-stimulating factor), a gene upregulated in TC #3 but downregulated in TC #2, suggesting a stimulus-dependent response. It has been previously shown that this gene is rarely upregulated in pancreatic cancer tumors and has been associated with poor prognosis (Takami et al. 2008). G-CSF (CSF3) as well as M-CSF (CSF1) have been previously reported as a potential prognostic factor for patient survival, although not independent of tumor stage (Groblewska et al. 2007). G-CSF has also been linked with IL6 as the main factors responsible for suppression of dendritic cell differentiation, maturation, and antigen presentation adding to the significant capability of tumor cells to evade the host immune system (Bharadwaj et al. 2007). That capability to actively evade the immune system of the host is usually attributed to the use of metalloproteinases (*MMP*), which may be used for receptor shedding (Schneiderhan et al. 2006). Interestingly one of the main PSC signals uniquely affecting TC is TIMP3, a tissue inhibitor of metalloproteinases. TAMs are at the center of recruitment and response to angiogenic and lymphangiogenic stimuli, and what potentially happens upon infiltration of TAMs into the pancreatic neoplasm is not difficult to predict. Tumor-associated macrophages are known to secrete proteins such as TGFβ, TNFα, IL1α, therefore we can speculate that the TC response to this signaling will involve a further increase in IL8 and VEGFA for vascular angiogenesis. TAMs are known to induce VCAM-1 expression in mesothelial cells, a step that is believed to be key for TC dissemination into peritoneum (lining of the abdominal cavity) during metastasis. Wu et al. 2007 have shown that VCAM-1 expression by tumor cells leads to decreased apoptosis of the TC and a significant decrease in the number of tumor-infiltrating $CD8^+$ T cells in the neoplasm. In summary the inflammatory, and angiogenic signals combined with the speculated infiltration of neoplasm by TAMs as well as the presence of

some strongly upregulated genes such as *CEBPD* and *ZEB1* (related to the Epithelial-Mesenchymal Transition) offer strong signs of the TC metastatic potential enhanced further by the newly discovered (Xu et al. 2010) capability of PSC to co-migrate with TC.

As we have previously shown, there is a gene regulatory connection in TC between the *CEBP* and the *EGR* family of genes, both significant in the reverse engineered regulatory network. From literature we can resolve the connections between *EGR2, 3,* and *4* via *GCLC* interconnecting them (see section 5.2.8.8.2). GCLC is central to the antioxidant capacity of the cell, and may be upregulated by IL1β via the *p38* form of MAPK and NFκB. Their interdependency connects us with *EGR1* and *ATF4*, linking directly to angiogenesis induced in response to hypoxic stress, as well as endothelial cell proliferation, inflammatory response and regulation of leukocyte migration, all significant steps engaged very early in the development of tumors. Hypoxia is known to increase the resistance of human pancreatic cancer cells to apoptosis and genes central to this stress reaction include *Akt, Hif-1,* and *HMOX1*, along with *GCLC.* The last two genes show a significant upregulation in experiment 3 tumor cells. Hypoxia has also been shown to link with EMT inducer Twist (Peinado et al. 2008), which operates under the control of hypoxia signaling in the tumor microenvironment. An even stronger link of tumor pro-survival signaling induced by hypoxia is discovered in our analysis via the strong induction of metallothioneins, which have been previously shown to be capable of attenuating hypoxic cell death in vitro (Tanji et al. 2003).

Especially interesting in the framework of our analysis is *MT1M*, a highly regulated and stimulus-dependent metallothionein produced by TC, which shows strong anti-correlation between TC experiments (high upregulation when treated by stimulated PSC, high downregulation when treated with quiescent PSC). MTs have been shown to correlate with metastasis in certain tumors. Poor prognosis and poor histological grading (poorer glandular differentiation and nuclear anaplasia) has been shown for colorectal cancer, plastic astrocytomas, squamous cell carcinoma of the esophagus, renal cell carcinoma, and small cell carcinoma of the lung. Unfortunately this is not the case for all tumors, which indicates variability of the biological significance according to the tumor type (Palmiter 1998). However, wherever it is being used as a prognostic of poor survival, it is associated with a high proliferative activity of the tumor cells. It has been previously suggested that the expression of metallothioneins could be a prognostic indicator in pancreatic carcinomas (Ohshio et al. 1996). Since the expression of MTs is not always indicative of prognosis (Palmiter 1998), it seems that MTs alone are insufficient as a

marker of progression, but a closer inspection of related genes may provide a set, which correlates together not only functionally, but also expression-wise to offer a stronger marker.

The biological mechanisms underlying metallothionein overexpression in tumors are not fully understood, but from our analysis we recognize a link between MTs, zinc, and its cellular transporters. Zinc is an important trace element for healthy growth and development, which has been previously related to cancer growth and progression (Cherian 1994; Fan et al. 2002). The regulation and maintenance of a "normal" concentration and distribution of cellular zinc is essential to the function, metabolism, growth, proliferation and survival of cells. It is required for many genes like *TNFAIP3* or *ZBTB1* (zinc finger proteins), essential for many transcription factors and metalloenzymes, and critical for metalloproteinases MMPs, which are involved in hypoxia, angiogenesis, cell proliferation and metastasis. Zinc homeostasis is under the control of antagonistic zinc transporters ZIP (intake) and ZnT (removal). In our dataset we recognize a set of upregulated genes related to this regulation including *SLC39A8, SLC39A10 and SLC39A14*, which encode Zrt-, irt-like proteins (ZIP), as well as *SLC30A1*, which encodes a ZnT that removes zinc from the cell. In this context it is interesting that *SLC39A8* has been linked with the regulatory effects of the transcriptional network for mesenchymal transition in brain tumors (Carro et al. 2010). This gene is also part of the transcriptional network governing the angiogenic switch in human pancreatic cancer (Abdollahi et al. 2007).

It is known that too much of ZIP, a molecule that enables the transport of zinc into cells, promotes the growth and spread of pancreatic tumors cells. It has been recently reported that ZIP4 overexpression causes increased IL6 transcription through CREB, which in turn activates STAT3 and leads to increased cyclin D1 expression, resulting in increased cell proliferation and tumor progression in pancreatic cancer (Zhang et al. 2010). On the other hand too much ZnT, which removes zinc, may cause a breakdown in many regulatory processes. Zinc deficiency induces oxidative DNA damage and increases p53 expression. DNA damage combined with an impaired repair system may lead to further cancer progression.

Existing literature suggest regulation mechanism whereby MTs can interact with p53 in the control of cell division (Śliwińska-Mossoń et al. 2009). Hainaut and Milner (1993) found that exposure to a metal chelating agent induces a reversible conformation change in wild type p53 to the mutant form, and they suggest that binding zinc ions to tertiary structure of p53 stabilizes it. Metallothioneins with their high affinity to zinc ions and ability to remove it from the cell suggest a direct connection where reduction of intra-nuclear zinc ion levels induces functional inactivation of p53 providing the tumor cells with the ability to proliferate and accumulate

mutational events. MT promoters have binding sites for the *Sp1, AP-1, and AP-2* transcription factors that mediate the effects of growth factors and protein kinases on transcription processes. All of this combined provides us with a stimulus-transcription coupling between *MTs, SP1, AP1, AP2*, as well as most of the transcription factors of the zinc-finger type including *EGRs*, and through that a connection to signaling cascades such as for example *IGF1R → Src-Mek-Erk-Egr1*.

These potential interactions between the p53 tumor suppressor and MTs are also interesting in light of findings by Yamasawa et al. 2002 who have shown that loss of function p53 correlates with an overexpression of *GADD45A* – a p53 target, and one of the more significantly changing genes in our tumor cell experiment (3). *p53-GADD45A* co-expression has apparently (only an abstract of a Chinese publication is available) been shown to be a poor predictor for patient outcome (Dong et al. 2005). It would be interesting to see how that changes if the profile is expanded to include e.g. MT1M given the biological relationships.

7.2.9 Potential points of interference

The identification of the intracellular gene regulatory networks and unique soluble factors allowed us to propose a set of potential points of interference to disturb the formation of the TC-PSC-TC signaling axis. While we were able to decrypt interactions between cells, and confirm their dependence on the underlying GRN, their physiological impact is yet to be studied. Intracellular target genes are often members of larger clusters, which share time-resolved expression profiles, and a further investigation of those may provide additional interesting venues for potential *in vitro* and *in vivo* validation.

Some of the identified signaling cues are reinforced by literature sources proving the validity of the approaches applied in our work, and emphasizing the potential of the novel factors arising in this context.

Intracellular targets:

- PSC GRN points of interference including *PHACTR2, MALAT1, HIRA, and APOL6*.
- TC GRN points of interference including the *EGR* gene family with emphasis on *EGR3*.

The latter were confirmed with *in vitro* qRT-PCR analysis to have the predicted effect, unfortunately the transfection of PSC remains outside of experimental possibility.

Additionally we offer a relatively short list of intracellular signal transducers, which may be of potential interest including the toll-like receptor signaling in the PSC antiviral responses, and signal transduction in the inflammation-related pathways of TC. Furthermore many signals of autocrine and paracrine nature are transduced through common receptors such as EGFR, PLAUR, CXCR1/3/4/5. Blocking those receptors may offer another way to block the cell-cell interactions.

Intercellular targets:

Signaling established between the PSC and TC is a dynamic process, which is clearly depicted in the temporal gene expression profiles. We were able to divide them into subgroups dependent on autocrine, paracrine and mixed intercellular signals. The points of interference at the extracellular level aim to interrupt not only the direct signal exchange, but also the actual formation of the underlying GRN induced in response to the stimuli. They are in detail presented in Figure 74 and are grouped into three separate phases with TC factors initializing PSC, PSC factors stimulating TC, and TC response into the microenvironment, most likely uniquely affecting PSC in another paracrine feedback:

1. Phase1: *MMP9, MMP7, MMP13, CTSE, EGF, CSF1, PGF, FGF2, CCL8, CCL5, CRH, SERPINA1, IL1A, NPPA, APOE, LAMA2, LAMA3, LAMC2, PTHLH, A2M, BMP2, PRL, GHRL*
2. Phase 2: *TIMP3, IL7R, ESM1, EBI2, CCL20, DKK3, IL6, TFPI2, STC1*
3. Phase 3: *CX3CL1, CXCL6, CXCL10, CSF1, CSF3, BTC, GAL, C3, CFB, SHH, FGF18, WNT6*

Our findings overlap and confirm data reported by other groups including:

- sonic hedgehog SHH produced by TC in the final response phase, recently reported by Bailey et al. 2008, Yauch et al. 2008, Olive et al. 2009, and Xu et al. 2010.
- AKAP12 identified as a signal transducer during PSC initialization, reported by us in the context of the GRN of migrating keratinocytes, where the model has shown that it is important to cell migration (Busch et al. 2008).
- CSF1 produced by TC in the final response, aforementioned in this discussion (Groblewska et al. 2007), also shown by Zhong et al. 2008 for lung cancer.
- PTHLH as a TC signal initializing PSC, reported by us (Busch et al. 2008) as one of the abundantly overexpressed proteins in migrating keratinocytes.

Rich environmental setup:

- TGF-β reported by Ellenrieder et al. 2001, Jungert et al. 2007, Ijichi et al. 2006, and Bardeesy et al. 2006.
- VEGF – Yang et al. 2006, Tang et al. 2006
- EGF via EGFR and HGF both involved in migratory response of keratinocytes in wound healing (Busch et al. 2008)
- ESR1 (Konduri et al. 2007)
- PPAR (Eibl et al. 2004, Eibl et al. 2008)
- PLAU/PLAUR (Kjøller, 2003; Ploug et al. 2002 and 2003; Yimin et al. 2003; Ploplis et al. 2007).

8. Conclusions

Our global gene expression analysis, ranking and modeling allowed us to capture and predict temporal dynamics of gene regulatory networks and to define the interactions between genes of interest in both cell types. *In silico* simulations provided a list of experimental targets, which were successfully validated in tumor cells. From the experimental context and data analysis we were able to recognize that the initial stimulus axis of TC→PSC→TC interactions induces a PSC state transitions from the initial stable steady state (corresponding to PSC, which are activated, but unexposed to TC) into an unstable state contingent on exposure to TC. A finding in line with the *in vivo* situation where cells in the desmoplastic stroma are motile, and transiently exposed to tumor cells. We investigated how the transition is induced (constitutive TC stimulus), what regulatory mechanisms are stimulated during that induction (PSC GRN, soluble output), what effect they have on the associated TC (TC GRN), and how those TC maintain the PSC state through soluble factors. Through the proposed methodology we have gained a deeper understanding of how TC-PSC microenvironment is established, how the extracellular signaling cues are formed, and how they affect the intracellular networks in a time-ordered manner. The identified potential points of interference offer valid and in case of extracellular factors easily accessible targets for the development of therapeutic agents, which would aim to disturb the formation of this complex cellular interplay that drives pancreatic tumor resistance.

9. Future Work

Quiescent PSC were recently isolated by Vonlaufen et al. (2010), with that it is now possible to observe the complete set of state transitions in the stroma including the move from a quiescent into an activated PSC steady state, which in our case was experimentally approximated. To investigate a complete set of interactions the experiments will have to be extended to encompass also immune cells. This is especially important since the occurrence of TC may be either preceded by chronic inflammation and immune cells presence, or trailed by immune cell infiltration of tissue in response to tumor attractors.

Stellate cells were shown to be unique in many ways when investigated in the context of PDAC. The signaling cues put them into the range of: inflammation, wound healing, and antiviral responses. Therefore a change of perspective may be necessary, from looking at PSC as fibroblasts, to looking at them as multi-functional cells sharing features with other cell types e.g. immune cells such as macrophages (antiviral responses). The identification of shared signaling in the area of wound healing also suggests that an investigation of PSC in the context of findings by Xu et al. 2010, and Busch et al. 2008 might be beneficial in order to limit the stellate cell capability to accompany TC in the formation of distant metastases.

An expansion of the existing CTRNN modeling approach to include an additional layer of output neurons (similar to one described by Arie et al. 2007) may allow us to create models of multiple cell types interacting with each other, which would greatly improve model integration.

A one step Midpoint method will be implemented in the future to improve the ODE integration, which is currently achieved with forward Euler method, and avoid the implementation issues of multiple iterative loops of the Runge-Kutta method.

10. Literature

Abdollahi et al. (2007) Transcriptional network governing the angiogenic switch in human pancreatic cancer. PNAS July 31, 2007; vol. 104; no.31:12890-12895

Afonyushkin T, Oskolkova OV, Philippova M, Resink TJ, Erne P, Binder BR, Bochkov VN. (2010) Oxidized phospholipids regulate expression of ATF4 and VEGF in endothelial cells via NRF2-dependent mechanism: novel point of convergence between electrophilic and unfolded protein stress pathways. Arterioscler Thromb Vasc Biol.;30(5):1007-13

Albini A. (1998) Tumor and endothelial cell invasion of basement membranes. Pathology & Oncology Research 1998 Volume 4, Number 3, 230-241

Algül H., Treiber M., Lesina M. & Schmid R.M. (2007) Mechanisms of Disease: chronic inflammation and cancer in the pancreas—a potential role for pancreatic stellate cells? Nature Reviews Gastroenterology and Hepatology 4, 454-462

American Cancer Society Pancreatic Cancer Tumor Staging (2010) http://www.cancer.org/docroot/cri/content/cri_2_4_3x_how_is_pancreatic_cancer_staged_34.asp, as accessed on 10.09.2010

Anand M.K.N., Boylan C., Gupta N. (2010) Pancreatic Adenocarcinoma Imaging, Medscape as accessed on 25.10.2010 http://emedicine.medscape.com/article/370909-overview

Angelini C., De Canditiis D., Mutarelli M., Pensky M. (2007) A Bayesian Approach to Estimation and Testing in Time-course Microarray Experiments. Stat Appl Genet Mol Biol 2007, 6:Article 24

Aoki N., Hiraide K. (1994) Topological Theory of Dynamical Systems. Amsterdam, Netherlands: North-Holland Math Library, vol. 52, Amsterdam, 1994, viii + 416 pp., ISBN 0-444-89917-0

Aoki H. et al. (2005) Autocrine Loop between Transforming Growth Factor-β and Interleukin-1β through Smad3- and ERK-dependent Pathways in Rat Pancreatic Stellate Cells. Am J Physiol Cell Physiol. 2005 Dec 21; 16371439

Apte M.V., Haber P.S., Applegate T.L., Norton I.D., McCaughan G.W., Korsten M.A., Pirola R.C., Wilson J.S. (1998) Periacinar stellate shaped cells in rat pancreas: identification, isolation, and culture. Gut 1998, 43:128-133

Apte M.V., Haber P.S., Darby S.J., Rodgers S.C., Pirola R.C., Korsten M.A., McCaughan G.W., Wilson J.S. (1999) Pancreatic stellate cells are activated by proinflammatory cytokines: implications for pancreatic fibrogenesis. Gut, 44:534-541

Apte M.V., Pirola R.C., Wilson J.S. (2006) Battle-scarred pancreas: Role of alcohol and pancreatic stellate cells in pancreatic fibrosis. J Gastroenterol Hepatol.;21 Suppl 3:97-101

Arie H., Ogata T., Tani J., Sugano S. (2007) Reinforcement learning of a continuous motor sequence with hidden states Advanced Robotics, Vol. 21, No. 10, pp. 1215–1229

Argani P., Iacobuzio-Donahue C., Ryu B. (2001) Mesothelin is overexpressed in the vast majority of ductal adenocarcinomas of the pancreas: identification of a new pancreatic cancer marker by serial analysis of gene expression (SAGE). Clin. Cancer Res. 7(12), 3862–3868

Arumugam T., Ramachandran V., Fournier K.F., Wang H., Marquis L., Abbruzzese J.L., Gallick G.E., Logsdon CD, McConkey D.J., Choi W. (2009) Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. Cancer Res. 2009 Jul 15;69(14):5820-8

Asea et al. (2002) Novel Signal Transduction Pathway Utilized by Extracellular HSP70 Role of Toll-like Receptor (TLR) 2 and TLR4. J. Biol. Chem. 277 (17): 15028-15034

Azorsa D.O. et al. (2009) Synthetic lethal RNAi screening identifies sensitizing targets for gemcitabine therapy in pancreatic cancer J Transl Med. 2009 Jun 11;7:43

Bachem M.G., Schneider E., Gross H., Weidenbach H., Schmid R.M., Menke A., Siech M., Beger H., Grunert A., Adler G. (1998) Identification, culture, and characterization of pancreatic stellate cells in rats and humans. Gastroenterology, 115:421-432

Bachem M.G., Schünemann M., Ramadani M., Siech M., Beger H., Buck A., Zhou S., Schmid-Kotsas A., Adler G. (2005) Pancreatic carcinoma cells induce fibrosis by stimulating proliferation and matrix synthesis of stellate cells. Gastroenterology,Apr;128(4):907-21

Bachem M.G., Zhou S,. Buck K., Schneiderhan W., Siech M. (2008) Pancreatic stellate cells – role in pancreas cancer. Langenbecks Arch Surg; 393: 891–900

Bailey J.M., Swanson B.J., Hamada T., Eggers J.P., Singh P.K., Caffery T., Ouellette M.M., Hollingsworth M.A. (2008) Sonic Hedgehog Promotes Desmoplasia in Pancreatic Cancer, Clin Cancer Res. 2008 Oct 1;14(19):5995-6004

Bailey J.M., Mohr A.M. and Hollingsworth M.A., (2009) Sonic hedgehog paracrine signaling regulates metastasis and lymphangiogenesis in pancreatic cancer. Oncogene 28, 3513-3525 (8 October 2009), doi:10.1038/onc.2009.220

Baldwin AS Jr. 1996, The NF-kappa B and I kappa B proteins: new discoveries and insights. Annu Rev Immunol.;14:649-83.

Barcellos-Hoff, M.H. (1998) The potential influence of radiation- induced microenvironments in neoplastic progression. J. Mammary Gland Biol. Neoplasia 3, 165–175

Bardeesy N., Cheng K.H., Berger J.H., Chu G.C., Pahler J., Olson P., Hezel A.F., Horner J., Lauwers G.Y., Hanahan D., and DePinho R.A. (2006) Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. Genes Dev;20:3130- 3146

Basso K., Margolin A. A., Stolovitzky G., Klein U., Dalla-Favera R., and Califano A. (2005) Reverse engineering of regulatory networks in human B cells. Nature Genetics, vol. 37, no. 4, pp. 382–390

Beal M.J., Falciani F., Ghahramani Z., Rangel C., and Wild D.L. (2005) A Bayesian approach to reconstructing genetic regulatory networks with hidden factors," Bioinformatics, vol. 21, no. 3, pp. 349–356

Beer R.D. (1995). On the dynamics of small continuous-time recurrent neural networks. Adaptive Behavior 3(4):471-511

Beer M.A. and Tavazoie S. (2004) Predicting Gene Expression from Sequence. Cell, Volume 117, Issue 2, 185-198, 16 April 2004, doi:10.1016/S0092-8674(04)00304-6

Beisvag et al. (2006) GeneTools - application for functional annotation and statistical hypothesis testing. BMC Bioinformatics 2006, **7:**470

Bharadwaj et al. (2007) Elevated Interleukin-6 and G-CSF in Human Pancreatic Cancer Cell Conditioned Medium Suppress Dendritic Cell Differentiation and Activation. Cancer Res June 1, 67; 5479; doi: 10.1158/0008-5472.CAN-06-3963

Bhowmick, N.A. et al. (2004) Stromal fibroblasts in cancer initiation and progression. Nature 432, 332–337

Biscetti F., Gaetani E., Flex A., et al. (2008) Selective activation of peroxisome proliferator-activated receptor (PPAR)α and PPARγ induces neoangiogenesis through a vascular endothe- lial growth factor-dependent mechanism. Diabetes, vol. 57, no. 5, pp. 1394–1404

Bissell M.J., Radisky D. (2001) Putting tumours in context. Nat Rev Cancer 1:46-54

Boden, M. (2001) A guide to recurrent neural networks and backpropagation. School of Information Science, Computer and Electrical Engineering Halmstad University. Source:http://www.hh.se/staff/mibo/publications/rn2.pdf

Bolstad B.M., Irizarry R. A., Astrand M., and Speed T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. Bioinformatics 19(2):185-193

Boros, S. (2009). Introduction: Dynamic organizations and organizational dynamics. In S. Boros (Ed.), Exploring organizational dynamics (pp. 1-7). London: Sage

Bosotti R, Locatelli G, Healy S, et al. (2007) Cross platform microarray analysis for robust identification of differentially expressed genes. BMC Bioinformatics 2007; 8(Suppl 1):S5

Boutanaev A.M., Kalmykova A.I., Shevelyov Y.Y. and Nurminsky D.I. (2002) Large clusters of co-expressed genes in the Drosophila genome, Nature 420, 2002, pp. 666–669

Bramhall S.R., Stamp G.W., Dunn J., Lemoine N.R., Neoptolemos J.P. (1996) Expression of collagenase (MMP2), stromelysin (MMP3) and tissue inhibitor of the metalloproteinases (TIMP1) in pancreatic and ampullary disease. Br J Cancer. 1996 Apr;73(8):972-8

Buchholz M., Braun M., Heidenblut A. et al. (2005) Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene 24(44), 6626–6636

Burris HA 3rd, Moore MJ, Andersen J, et al. (1997) Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol 1997;15:2403-13

Busch H., Camacho D., Rogon Z., Breuhahn K., Angel P., Eils R., Szabowski A. (2008) Gene Network Dynamics regulating Keratinocyte Migration, Mol Syst Biol. 2008; 4: 199

Calogero R.A., Cordero F., Sanges R. (2010) "oneChannelGUI Package Vignette" as accessed in the BioConductor resources on October 2010, section 9: Filtering, located at http://bioconductor.org/packages/2.6/bioc/vignettes/oneChannelGUI/inst/doc/

Campagna D., Cope L., Lakkur S.S., Henderson C., Laheru D., Iacobuzio-Donahue C.A. (2008) Gene expression profiles associated with advanced pancreatic cancer. Int. J. Clin. Exp. Pathol. 1(1), 32–43

Campbell P.J., Yachida S., Mudie L.J. et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature;467(7319):1109-13

Caron H. et al., (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains, Science 291 (2001), pp. 1289–1292

Carro M.S. et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature 463, 318-325

Cartmel B., Reid M.: Cancer control and epidemiology. (1997) In: Groenwald S.L., Frogge M.H., Goodman M., Yarbro C.H. Cancer Nursing Principles and Practice. 4th ed. Boston, MA: Jones and Bartlett; 1997:50–74

Center for Disease Control and Prevention - National Center for Health Statistics: "Cancer Mortality Surveillance" as accessed on 05.09.2010

Chaffer C.L., Weinberg R.A. (2011) A Perspective on Cancer Cell Metastasis, Science 25 March 2011: Vol. 331 no. 6024 pp. 1559-1564

Chambers A.F., Matrisian L.M. (1997) Changing views of the role of matrix metalloproteinases in metastasis. J Natl Cancer Inst; 89:1260-70

Chambers A.F., Groom A.C., MacDonald I.C. (2002) Metastasis: Dissemination and growth of cancer cells in metastatic sites. Nature Reviews Cancer 2, 563-572

Chelala C., Hahn S.A., Whiteman H.J. et al. (2007) Pancreatic expression database: a generic model for the organization, integration and mining of complex cancer datasets. BMC Genomics 8, 439

Chelala C., Lemoine N.R., Hahn S.A., Crnogorac-Jurcevic T. (2009) A web-based platform for mining pancreatic expression datasets. Pancreatology 9(4), 340–343

Chen F, Castranova V, Shi X, Demers LM. 1999, New insights into the role of nuclear factor-kappaB, a ubiquitous transcription factor in the initiation of diseases. Clin Chem. 1999 Jan;45(1):7-17

Chen T., He H. L., and Church G. M. (1999) Modeling gene expression with differential equations. In Proceedings of the 4th Pacific Symposium on Biocomputing (PSB '99), pp. 29–40

Chen J.J., Yao P.L., Yuan A. et al. (2003) Up-regulation of tumor interleukin-8 expression by infiltrating macrophages: its correlation with tumor angiogenesis and patient survival in non-small cell lung cancer. Clin Cancer Res;9:729–37

Cherian MG (1994) The significance of the nuclear and cytoplasmic tumor of metallothionein in human liver and tumor cells. Environ Health Perspect 102(suppl 3):131–135

Coussens L.M., Werb Z. (2002) Inflammation and cancer. Nature 420, 860-867

Crnogorac-Jurcevic T., Efthimiou E., Capelli P. et al. (2001) Gene expression profiles of pancreatic cancer and stromal desmoplasia. Oncogene; 20: 7437–46

Crnogorac-Jurcevic T., Efthimiou E., Nielsen T. et al. (2002) Expression profiling of microdissected pancreatic adenocarcinomas. Oncogene 21(29), 4587–4594

Culhane AC, Thioulouse J, Perriere G, Higgins DG. (2005) MADE4: An R package for Multivariate Analysis of Gene Expression Data. Bioinformatics; 21(11): 2789-90

Cunha G.R. et al. (2002) Role of stroma in carcinogenesis of the prostate. Differentiation 70, 473–485

Cutts R.J., Gadaleta E., Hahn S.A., Crnogorac-Jurcevic T., Lemoine N.R. and Chelala C. (2010) The Pancreatic Expression database: 2011 update. Nucleic Acids Research (Database Issue). 2010, 1-6.

D'haeseleer P., Wen X., Fuhrman S., and Somogyi R. (1999) Linear modeling of mRNA expression levels during CNS development and injury," in Proceedings of the 4th Pacific Symposium on Biocomputing (PSB '99), pp. 41–52

de Jong H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. Journal of Computational Biology, vol. 9, no. 1, pp. 67–103

Delbrück M. (1948) Discussion. In Unités biologiques douées de continuité génétique Colloques Internationaux du Centre National de la Recherche Scientifique. CNRS, Paris

Dickman, P. W. and Adami, H. O. (2006) Interpreting trends in cancer patient survival. J Intern Med, 260: 103-117

DiMagno EP, Reber HA, Tempero MA. (1999) AGA technical review on the epidemiology, diagnosis, and treatment of pancreatic ductal adeno- carcinoma. American Gastroenterological Association. Gastroenterology;117:1464 – 84

Dong M, Zhou JP, Kong FM, Guo KJ, Tian YL, Dong YT. Zhongguo (2005) Expressions of p53 and Gadd45a proteins in human pancreatic cancer and their clinicopathological significance. [Article in Chinese] Yi Xue Ke Xue Yuan Xue Bao. 2005 Oct;27(5):628-32

Duncan D.T., Prodduturi N., Zhang B. (2010) WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. BMC Bioinformatics, 11(Suppl 4):P10
WebGestalt: http://bioinfo.vanderbilt.edu/webgestalt

Dunn G.P., Old L.J., Schreiber R.D. (2004) The three Es of cancer immunoediting. Annu Rev Immunol;22:329-60

Dvorak H.F. (1986) Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N. Engl. J. Med. 315, 1650–1659

Egeblad M., Werb Z. (2002) New functions for the matrix metalloproteinases in cancer progression. Nat Rev Cancer;2(3):161-74

Eibl G., Reber H.A., Hines O.J., Go V.L. (2004) COX and PPAR: possible interactions in pancreatic cancer. Pancreas; Nov;29(4):247-53

Eibl G. (2008) The Role of PPAR-gamma and Its Interaction with COX-2 in Pancreatic Cancer. PPAR Res.;2008:326915

Eisen M., Spellman P., Brown P., Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. 95:14863–14868

Ellenrieder V., Alber B., Lacher U., Hendler S.F., Menke A., Boeck W., et al. (2000) Role of MT-MMPs and MMP-2 in pancreatic cancer progression. Int J Cancer; 85:14-20

Ellenrieder V., Hendler S.F., Boeck W., Seufferlein T., Menke A., Ruhland C., Adler G. and Gress T.M. (2001) Transforming growth factor beta1 treatment leads to an epithelial-mesenchymal trans- differentiation of pancreatic cancer cells requiring extracellular signal-regulated kinase 2 activation. Cancer Res; 61:4222- 4228

Fan L.Z., Cherian M.G. (2002) Potential role of p53 on metal- lothionein induction in human epithelial breast cancer cells. Br J Cancer 87:1019–1026

Farmer P., Bonnefoi H., Anderle P., Cameron D., Wirapati P., et al. (2009) A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. Nat Med.;15:68–74

Farrow B., Albo D., and Berger D.H. (2008) The Role of the Tumor Microenvironment in the Progression of Pancreatic Cancer, Journal of Surgical Research 149, 319–328

Feng Y., Santoriello C., Mione M., Hurlstone A., & Martin P. (2010) Live Imaging of Innate Immune Cell Sensing of Transformed Cells in Zebrafish Larvae: Parallels between Tumor Initiation and Wound Inflammation. PLoS Biol. 2010 Dec 14;8(12):e1000562.

Ferrazzi F., Magni P. and Bellazzi R. (2005) Random walk models for bayesian clustering of gene expression profiles. Appl Bioinformatics 4(4):263-76

Fitzner B. (2004) Involvement of AP-1 proteins in pancreatic stellate cell activation in vitro. Int J Colorectal Dis. 2004 Sep;19(5):414-20

Forcada M.L. (2011) Neural Networks: Automata and Formal Models of Computation, Universitat d'Alacant, Dept. Llenguatges i Sistemes Informàtics. http://www.dlsi.ua.es/~mlf/nnafmc/pbook/ as accessed on 01.01.2011

Friedman N., Linial M., Nachman I., and Pe'er D. (2000) Using Bayesian networks to analyze expression data," Journal of Computational Biology, vol. 7, no. 3-4, pp. 601–620

Friedman N., Murphy K., and Russell S. (1998) Learning the structure of dynamic probabilistic networks. In Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence (UAI '98), pp. 139–147

Fu P., Latterich M., Panke S. (2009) Systems Biology and Synthetic Biology, John Wiley & Sons, 2009

Galli A., Crabb D., Price D., Ceni E., Salzano R., Surrenti C., Casini A. (2000) Peroxisome proliferator-activated receptor γ transcriptional regulation is involved in platelet-derived growth factor-induced proliferation of human hepatic stellate cells. Hepatology, 31:101-108

Galluzzi L., Giordanetto F. and Kroemer G. (2009) Targeting HSP70 for Cancer Therapy, Molecular Cell, Volume 36, Issue 2, 176-177

Gardner T. S. and Faith J. J. (2005) Reverse-engineering transcription control networks. Physics of Life Reviews, vol. 2, no. 1, pp. 65–88

GeneCards The Human Gene Compendium (2008-2011) Weizmann Institute of Science, as accessed at: http://www.genecards.org/

Gentleman R. (2010) Annotate, R package, Gene Ontology Consortium, Bioconductor Project, accessed 2010 http://www.bioconductor.org/help/bioc-views/release/bioc/html/annotate.html

Goldberg et al. (2010) Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions, Cell, Volume 140, Issue 5, 678-691

Golubitsky M. (1997) Introduction to Applied Nonlinear Dynamical Systems and Chaos. New York: Springer-Verlag

Gress T.M., Menke A., Bachem M. et al. (1998) Role of extracellular matrix in pancreatic diseases. Digestion;59:625-37

Grivennikov S.I., Greten F.R., Karin M. (2010) Immunity, inflammation, and cancer. Cell. 19;140(6):883-99

Groblewska M., Mroczko B., Wereszczyńska-Siemiatkowska U., Myśliwiec P., Kedra B., Szmitkowski M. (2007) Serum levels of granulocyte colony-stimulating factor (G-CSF) and macrophage colony-stimulating factor (M-CSF) in pancreatic cancer patients. Clin Chem Lab Med.;45(1):30-4

Grutzmann R., Pilarsky C., Ammerpohl O. et al. (2004) Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays. Neoplasia 6(5), 611–622

Gudjonsson B. (1987) Cancer of the pancreas: 50 years of surgery. Cancer;60:2284 –303

Haber P.S., Keogh G.W., Apte M.V., Moran C.S., Stewart N.L., Crawford D.H., Pirola R.C., McCaughan G.W., Ramm G.A., Wilson J.S. (1999) Activation of pancreatic stellate cells in human and experimental pan- creatic fibrosis. Am J Pathol, 155:1087-1095

Hache et al. (2009) Reverse Engineering of Gene Regulatory Networks: A Comparative Study, EURASIP Journal on Bioinformatics and Systems Biology Volume 2009, ArticleID 617281

Hache H., Wierling C., Lehrach H., and Herwig R. (2007) Reconstruction and validation of gene regulatory networks with neural networks. In Proceedings of the 2nd Foundations of Systems Biology in Engineering Conference (FOSBE '07), pp. 319–324

Hainaut P, Milner J (1993) A structural role for metal ions in the “wide-type” conformation of the tumour suppressor pro- tein p53. Cancer Res 53:1739–1742

Han H, Bearss DJ, Browne LW, Calaluce R, Nagle RB, Von Hoff DD. (2002) Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. Cancer Res. 62(10), 2890–2896

Harsha HC, Kandasamy K, Ranganathan P et al. (2009) A compendium of potential biomarkers of pancreatic cancer. PLoS Med. 6(4), e1000046

Hecker M. (2009) Gene regulatory network inference: Data integration in dynamic models – a review; Biosystems;96(1):86-103

Hess, J., Angel, P., Schorpp-Kistner, M. (2004) AP-1 subunits: quarrel and harmony among siblings December 1, 2004 J Cell Sci 117, 5965-5973

Heyninck K. et al. (1999) The zinc finger protein A20 inhibits TNF-induced NF-kappaB-dependent gene expression by interfering with an RIP- or TRAF2-mediated transactivation signal and directly binds to a novel NF-kappaB-inhibiting protein ABIN. J. Cell Biol.;145 (7): 1471–82

Hruban RH, Takaori K, Klimstra DS, Adsay NV, Albores-Saavedra J, et al. (2004) An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms. Am J Surg Pathol 28: 977–987

Hruban R.H., Iacobuzio-Donahue C., Wilentz R.E., Goggins M., Kern S.E. (2001) Molecular pathology of pancreatic cancer. Cancer J., 7: 251-258

Hu, M. and Polyak, K., Curr. (2008) Microenvironmental regulation of cancer development (Review). Opin. Genet. Dev., 18, 27-34

Hu, X., Maglia, A., Wunsch, D. (2005). A general recurrent neural network approach to model genetic regulatory networks. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. pp. 4735–4738.

Huang S. and Ingber D.E. (2006) Self-organizing attractors in cell regulatory networks. Breast Disease 26:27-54

Huang D.W., Sherman B.T., Lempicki R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat Protoc. 4(1):44-57

Hwang RF, Moore T, Arumugam T. et al. (2008) Cancer-associated stromal fibroblasts promote pancreatic tumor progression. Cancer Res.; 68; 918

Ide T, Kitajima Y, Miyoshi A, et al. (2006) Tumor-stromal cell interaction under hypoxia increases the invasiveness of pancreatic cancer cells through the hepatocyte growth factor/c-Met pathway. Int J Cancer;119:2750e9

Ijichi H, Chytil A, Gorska AE, Aakre ME, Fujitani Y, Fujitani S, Wright CV and Moses HL. (2006) Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-beta signaling in cooperation with active kras expression. Genes Dev; 20:3147-3160

Ikejiri N. The vitamin-A storing cells in the human and rat pancreas. (1990) Kurume Med J, 37:67-81

Ip YT, Davis RJ. (1998) Signal transduction by the c-Jun N-terminal kinase (JNK)--from inflammation to development. Curr Opin Cell Biol; 10: 205-19

Irizarry, RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Scherf, U, Speed, TP (2003) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistics .Vol. 4, Number 2: 249-264

Irizarry R.A., Bolstad B.M., Collin F., Cope L.M., Hobbs B. and Speed T.P. (2003), Summaries of Affymetrix GeneChip probe level data Nucleic Acids Research 31(4):e15

Ito A., Matsuoka F., Honda H. and Kobayashi T., Cancer Gene Therapy (2003) 10, 918–925. doi:10.1038/sj.cgt.7700648 Heat shock protein 70 gene therapy combined with hyperthermia using magnetic nanoparticles

Jaenisch R. & Bird A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics 33, 245

Jain, R.K., (2005) Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy (Review). Science, **307,** 58-62

Janes K.A., Gaudet S,. Albeck J.G., Nielsen U.B., Lauffenburger D.A., Sorger P.K. (2006) The response of human epithelial cells to TNF involves an inducible autocrine cascade. Cell 124: 1225–1239

Jaskiewicz K., Nalecz A., Rzepko R., Sledzinski Z. (2003) Immunocytes and activated stellate cells in pancreatic fibrogenesis. Pancreas; 26: 239–42

Jemal A., Siegel R., Xu J., Ward E. (2010) Cancer statistics, 2010. CA Cancer J Clin. doi: 10.3322/caac.20073

Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Müller-Tidow C. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene;22(39):8031-41

Jones L, Ghaneh P, Humphreys M, et al. (1999) The matrix metalloproteinases and their inhibitors in the treatment of pancreatic cancer. Ann NY Acad Sci;880:288e307

Jones S, Zhang X, Parsons DW et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321(5897), 1801–1806

Jordan, D. W. and Smith, P. (1999) Nonlinear Ordinary Differential Equations: An Introduction to Dynamical Systems, 3rd ed. Oxford, England: Oxford University Press

Joshi T, Chen Y, Becker JM, Alexandrov N, Xu D. (2004) Genome-scale gene function prediction using multiple sources of high-throughput data in yeast Saccharomyces cerevisiae. OMICS.;8:322–333

Jungert K, Buck A, von Wichert G, Adler G, Konig A, Buchholz M, Gress TM and Ellenrieder V. (2007) Sp1 is required for transforming growth factor-beta-induced mesenchymal transition and migration in pancreatic cancer cells. Cancer Res; 67:1563-1570

Kalmykova A.I., Nurminsky D.I., Ryzhov D.V. and Shevelyov Y.Y. (2005) Regulated chromatin domain comprising cluster of co-expressed genes in Drosophila melanogaster, Nucleic Acids Res. 33, pp. 1435–1444.(28)

Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol, 22(3):437-467

Keating S.E., Maloney G.M., Moran E.M., Bowie A.G. (2007) IRAK-2 Participates in Multiple Toll-like Receptor Signaling Pathways to NFκB via Activation of TRAF6 Ubiquitination. Journal of Biological Chemistry, 282, 33435-33443

Kim, S., Kim, J., Cho, K.-H. (2007). Inferring gene regulatory networks from temporal expression profiles under time-delay and noise. Comput Biol Chem 31, 239–245.

Kjøller L. (2002) The urokinase plasminogen activator receptor in the regulation of the actin cytoskeleton and cell motility. Biol Chem.;383(1):5-19

Klein, A. P., Hruban, R. H., Brune, K. A., Petersen, G. M., and Goggins, M. (2001) Familial pancreatic cancer. Cancer J, 7: 266-273

Koenig A, Mueller C, Hasel C, Adler G, Menke A. (2006) Collagen type I induces disruption of E-cadherin-mediated cell-cell contacts and promotes proliferation of pancreatic carcinoma cells. Cancer Res; 66: 4662–71

Konduri S., Schwarz R.E. (2007) Estrogen receptor beta/alpha ratio predicts response of pancreatic cancer cells to estrogens and phytoestrogens. J Surg Res.;140(1):55-66

Kopfstein L, Christofori G. (2006) Metastasis: cell-autonomous mechanisms versus contributions by the tumor microenvironment. Cell Mol Life Sci.;63(4):449-68

Korc, M. (2007) Pancreatic cancer–associated stroma production, The American Journal of Surgery 194; S84–S86

Kornmann M, Beger HG, Link KH. (2003) Chemosensitivity testing and test-directed chemotherapy in human pancreatic cancer. Recent Results Cancer Res.;161:180-95

Leek JT, Monsen E, Dabney AR, Storey JD. (2006) EDGE: extraction and analysis of differential gene expression. Bioinformatics.;22(4):507–8

Leisering, W., Alonzo, T. and Pepe, M. S. (2000) Comparisons of Predictive Values of Binary Medical, Diagnostic Tests for Paired Designs, Biometrics 56: 345–351

Lercher M.J., Urrutia A.O. and Hurst L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome, Nat. Genet. 31, pp. 180–183

Leung-Theung-Long S, Roulet E, Clerc P, Escrieut C, Marchal-Victorion S, Ritz-Laser B, Philippe J, Pradayrol L, Seva C, Fourmy D, Dufresne M., 2005, Essential interaction of Egr-1 at an islet-specific response element for basal and gastrin-dependent glucagon gene transactivation in pancreatic alpha-cells. J Biol Chem.;280(9):7976-84

Li C-P, Chao Y, Chi K-H, et al. (2003) Concurrent Chemoradiotherapy Treatment of Locally Advanced Pancreatic Cancer: Gemcitabine versus 5-Fluorouracil, A Randomized Controlled Study. International Journal of Radiation Oncology Biology Physics;57:98-104

Li D., Xie K., Wolff R., and Abbruzzese J. L. (2004) Pancreatic cancer. Lancet, 363: 1049-1057

Lieber M et al. (1975) Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. Int J Cancer 15: 741-7

Li-Weber M, Krammer PH. 2003, Function and regulation of the CD95 (APO-1/Fas) ligand in the immune system. Semin Immunol 15(3):145-57.

Lo Iacono M., Di Costanzo A., Calogero R.A., et al. (2006) The Hay Wells syndrome-derived TAp63alphaQ540L mutant has impaired transcriptional and cell growth regulatory activity. Cell Cycle; 5:78–87

Logsdon CD, Simeone DM, Binkley C et al. (2003) Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. Cancer Res. 63(10), 2649–2657

Lohr M, Schmidt C, Ringel J et al. (2001) Transforming growth factor-beta1 induces desmoplasia in an experimental model of human pancreatic carcinoma. Cancer Res; 61: 550–5

López-Casas P.P., López-Fernández L.A. (2010) Gene expression profiling in pancreatic cancer, a 5 year review, Expert Rev Mol Diagn.;10(5):591-601

Lowenfels A. B., Maisonneuve P., Cavallini G., Ammann R. W., Lankisch P. G., Andersen J. R., Dimagno E.P., Andren-Sandberg A., and Domellof L. (1993) Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. N Engl J Med, 328: 1433-1437

Luo J., Epidemiological Studies of the Etiology of Pancreatic Cancer (2010) Karolinska Institutet, as accessed on 30.10.2010
http://publications.ki.se/jspui/handle/10616/38508

Luttenberger T, Schmid-Kotsas A, Menke A, Siech M, Beger H, Adler G, et al. (2000) Platelet-derived growth factors stimulate proliferation and extracellular matrix synthesis of pancreatic stellate cells: implications in pathogenesis of pancreas fibrosis. Lab Inves; 80:47-55. (PMID 10653002)

Magni P., Ferrazzi F. (2008) TimeClust: a clustering tool for gene expression time series. Bioinformatics; 24 (3): 430-432. doi: 10.1093/bioinformatics/btm605

Mahadevan D, Von Hoff DD. (2007) Tumor–stroma interactions in pancreatic ductal adenocarcinoma. Mol. Cancer Ther. 6(4), 1186–1197

Manna S.K., Ramesh G.T. (2005) Interleukin-8 Induces Nuclear Transcription Factor-κB through a TRAF6-dependent Pathway. The Journal of Biological Chemistry, 280, p.7010-7021

Mantovani A., Porta C., Rubino L., Allavena P. and Sica A. (2006) Tumor-associated macrophages (TAMs) as new target in anticancer therapy. Drug Discovery Today: Therapeutic Strategies Volume 3, Issue 3 , Pages 361-366

Margeli A., Kouraklis G., and Theocharis S. (2003) Peroxisome proliferator activated receptor-γ (PPAR-γ) ligands and angio- genesis. Angiogenesis, vol. 6, no. 3, pp. 165–169

Masamune A, Kikuta K, Watanabe T, Satoh K, Hirota M, Hamada S, Shimosegawa T. (2009) Fibrinogen induces cytokine and collagen production in pancreatic stellate cells. Gut.;58(4):550-9

Masamune A. (2005) Activation of the Jak-Stat pathway is required for the PDGF induced proliferation of pancreatic stellate cells. World J. gastroenterology;11(22):3385-91

McCarroll J.A. (2003) Pancreatic stellate cell activation by ethanol and acetaldehyde is it mediated by the mitogen-activated protein kinase signaling pathway. Pancreas;27(2):150-60

McConnell BB, Yang VW, 2010, Mammalian Krüppel-like factors in health and diseases. Physiol Rev;90(4):1337-81

Mews P., Phillips P., Fahmy R., Korsten M., Pirola R., Wilson J., Apte M. (2002) Pancreatic disease Pancreatic stellate cells respond to inflammatory cytokines: potential role in chronic pancreatitis. Gut;50:535-541

Mollenhauer J, Roether I, Kern HF. (1987) Distribution of extracellular matrix proteins in pancreatic ductal adenocarcinoma and its influence on tumor cell proliferation in vitro. Pancreas.;2(1):14-24

Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, et al. (2007) Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol; 25:1960-6

Morgan R. T. et al. (1980) Colo357: Human cell line (COLO 357) of metastatic pancreatic adenocarcinoma. Int. J. Cancer 25: 591-598

Mulcahy, N. (2009) American Society of Clinical Oncology (ASCO) 45th Annual Meeting: Abstract LBA4505. Presented May 31, 2009.

National Cancer Institute Think Tanks in Cancer Biology, accessed December 2010 http://www.cancer.gov/think-tanks-cancer-biology

Neesse A., Michl P., Frese K.K., Feig C., Cook N., Jacobetz M.A., Lolkema M.P., Buchholz M., Olive K.P., Gress T.M., Tuveson D.A., Stromal biology and therapy in pancreatic cancer, Gut gut.2010.226092 Published Online First: 21 October 2010

Nelander S. et al. (2005) Predictive screening for regulators of conserved functional gene modules (gene batteries) in mammals, BMC Genomics 6, p. 68

Nelson CM, Bissell MJ. (2006) Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. Annu Rev Cell Dev Biol.;22:287-309

Nguyen T.H. (2004) Mechanisms of metastasis, Clin Dermatol.;22(3):209-16

O'Neill L.A.J., (2008) Toll-like receptors in cancer. Oncogene 27, 158-160

Ohshio et al. (1996) Immunohistochemical study of metallothionein in pancreatic carcinomas. Journal of cancer research and clinical oncology, vol. 122, no6, pp. 351-355

Ohuchida, K. et al. (2004) Radiation to stromal fibroblasts increases invasiveness of pancreatic cancer cells through tumor-stromal interactions. Cancer Res. 64, 3215–3222

Olive K.P., Jacobetz M.A., Davidson C.J., et al. (2009), Inhibition of Hedgehog Signaling Enhances Delivery of Chemotherapy in a Mouse Model of Pancreatic Cancer, Science: Vol. 324 no. 5933 pp. 1457-1461

Olivero M., Ruggiero T., Saviozzi S., et al. (2006) Genes regulated by hepatocyte growth factor as targets to sensitize ovarian cancer cells to cisplatin. Mol Cancer Ther; 5:1126–35

Orimo, A. and Weinberg, R.A. (2006) Stromal fibroblasts in cancer: a novel tumor-promoting cell type. Cell Cycle 5, 1597–1601

Ott, E. (1993) Chaos in Dynamical Systems. New York: Cambridge University Press

Palmiter R.D. (1998) The elusive function of metallothioneins. PNAS; vol. 95 no. 15 8428-8430

Peinado H., Cano A. (2008) A hypoxic twist in metastasis., Nature Cell Biology 10, 253 - 254 doi:10.1038/ncb0308-253

Phillips P.A., Wu M.J., Kumar R.K., Doherty E., McCarroll J.A., Park S., Pirola R.C., Wilson J.S., Apte M.V. (2003) Pancreas Cell migration: a novel aspect of pancreatic stellate cell biology. Gut;52:677-682 doi:10.1136/gut.52.5.677

Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29: 153–159

Pineda, F. J. (1987) Generalization of back-propagation to recurrent neural networks. Physical Review Letters, 59(19):2229-2232

Ploplis VA, Tipton H, Menchen H, Castellino FJ (2007) A urokinase- type plasminogen activator deficiency diminishes the frequency of intestinal adenomas in apcmin/+ mice. J Pathol 213: 266–274

Ploug M. (2003) Structure-function relationships in the interaction between the urokinase-type plasminogen activator and its receptor. Curr Pharm Des.;9(19):1499-528

Ploug M, Gårdsvoll H, Jørgensen TJ, Lønborg Hansen L, Danø K. (2002) Structural analysis of the interaction between urokinase-type plasminogen activator and its receptor: a potential target for anti-invasive cancer therapy. Biochem Soc Trans.;30(2):177-83

Polyak K. et al. (2009) Co-evolution of tumor cells and their microenvironment (Review). Trends Genet. 25, 30-38

Purmann A. et al. (2007) Genomic organization of transcriptomes in mammals: coregulation and cofunctionality, Genomics 89, pp. 580–587

Quackenbush J. (2003) Genomics. Microarrays-guilt by association. Science.;302:240–241

Rangel C. et al. (2004) Modeling T-cell activation using gene expression profiling and state-space models. Bioinformatics, vol. 20, no. 9, pp. 1361–1372

Rasband, S. N. (1990) Chaotic Dynamics of Nonlinear Systems. New York: Wiley

The Reference Sequence (RefSeq) (2008-2011) National Center for Biotechnology Information: as accessed at: http://www.ncbi.nlm.nih.gov/RefSeq/

Rérole A.L., Joly A.L., Thuringer D. and Garrido C. (2010) Hsp70 and Hsp27: Emerging Targets in Cancer Therapy, Apoptosome, 169-202, DOI: 10.1007/978-90-481-3415-1_9

Richmond, A. & Thomas, H. (1986) Purification of melanoma growth stimulatory activity. J. Cell. Physiol. 129, 375-384

Robinson-Smith TM, Isaacsohn I, Mercer CA, Zhou M, Van Rooijen N, Husseinzadeh N, McFarland-Mancini MM, Drew AF. (2007) Macrophages mediate inflammation-enhanced metastasis of ovarian tumors in mice. Cancer Res.;67(12):5708-16

Rohde et al. (2005) Members of the heat-shock protein 70 family promote cancer cell growth by distinct mechanisms. *Genes & Dev*. 19: 570-582 doi: 10.1101/gad.305405

Romagnani P., Lasagni L., Annunziato F., Serio M., Romagnani S. (2004) CXC chemokines: the regulatory link between inflammation and angiogenesis. Trends in Immunology Volume 25, Issue 4, Pages 201-209

Ropers D., de Jong H., Geiselmann J. (2008), Mathematical modeling of genetic regulatory networks : Stress responses in Escherichia coli, In : P. Fu, M. Latterich, S. Panke, Systems and Synthetic Biology, Wiley-Intersciences, 235-271

Rothenberg ML, Moore MJ, Cripps MC, et al. (1996) A phase II trial of gemcitabine in patients with 5-FU-refractory pancreas cancer. Ann Oncol;7:347e53

Sadlonova A, Novak Z, Johnson MR, et al. (2005) Breast fibroblasts modulate epithelial cell proliferation in three-dimensional in vitro co-culture. Breast Cancer Res;7:R46

Sato N, Maehara N, Goggins M. (2004) Gene expression profiling of tumor-stromal interactions between pancreatic cancer cells and stromal fibroblasts. Cancer Res;64:6950e6

Sato N., Goggins M. (2006) The role of epigenetic alterations in pancreatic cancer. Journal of Hepato-Biliary-Pancreatic Surgery, Volume 13, Number 4, 286-295, DOI: 10.1007/s00534-005-1057-1

Schafer, M. and Werner, S. (2008) Cancer as an overhealing wound: an old hypothesis revisited. Nat. Rev. Mol. Cell Biol. 9, 628–638

Schneider E, Schmid-Kotsas A, Zhao J, Weidenbach H, Schmid RM, Menke A, Adler G, Waltenberger J, Grünert A, Bachem MG. (2001) Identification of mediators stimulating proliferation and matrix synthesis of rat pancreatic stellate cells. Am J Physiol Cell Physiol.;281(2):C532-43

Schneiderhan W., F. Diaz F, et al. (2006). Pancreatic stellate cells are are important source of MMP-2 in human pancreatic cancer and accelerate tumor progression in a murine xenograft model and CAM assay., J Cell Sci 120 (3): 512-519

Schäfer J. and Strimmer K. (2005) An empirical Bayes approach to inferring large-scale gene association networks," Bioinformatics, vol. 21, no. 6, pp. 754–764

SEER Training Modules, Cancer Classification. U.S. National Institutes of Health, National Cancer Institute. Dec.2010
http://training.seer.cancer.gov/disease/categories/classification.html.

Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet.;34(2):166-76

Sener SF, Fremgen A, Menck HR, et al. (1999) Pancreatic cancer: a report of treatment and survival trends for 100,313 patients diagnosed from 1985-1995, using the National Cancer Database. J Am Coll Surg;189:1–7

Seymour AB, Hruban RH, Redston M, Caldas C, Powell SM, Kinzler KW, Yeo CJ, Kern SE. (1994) Allelotype of pancreatic adenocarcinoma. Cancer Res.;54(10):2761-4

Sémon M. and Duret L. (2006) Evolutionary origin and maintenance of coexpressed gene clusters in mammals, Mol. Biol. Evol. 23, pp. 1715–1723.(35)

Shibue T., Weinberg R.A. (2010) Metastatic colonization: Settlement, adaptation and propagation of tumor cells in a foreign tissue environment. Seminars in Cancer Biology Volume 21, Issue 2, April 2011, Pages 99-106 The Biology of Cancer Metastasis

Singer G.A., Lloyd A.T., Huminiecki L.B. and Wolfe K.H. (2005) Clusters of co-expressed genes in mammalian genomes are conserved by natural selection, Mol. Biol. Evol. 22, pp. 767–775.(51)

Spellman P.T. and Rubin G.M. (2002) Evidence for large domains of similarly expressed genes in the Drosophila genome, J. Biol. 1, p.5

Spugnini EP, Cardillo I, Verdina A, et al. (2006) Piroxicam and cisplatin in a mouse model of peritoneal mesothelioma. Clin Cancer Res.12:6133–43

Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. (2005) Significance analysis of time course microarray experiments. PNAS, 102(36):12837-42

Storey JD, Dai JY, Leek JT. (2007) The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Biostatistics 8: 414-432

Storey JD. (2007) The optimal discovery procedure: A new approach to simultaneous significance testing. Journal of the Royal Statistical Society, Series B, 69: 347-368

Strogatz, S. H. (1994) Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering. Reading, MA: Addison-Wesley

Surowiak P, Suchocki S, Gyorffy B, et al. (2006) Stromal myofibroblasts in breast cancer: Relations between their occurrence, tumor grade and expression of some tumour markers. Folia Histochem Cytobiol;44:111

Śliwińska-Mossoń M., Milnerowicz H. Immunohistochemical localization of metallothionein and p53 protein in pancreatic serous cystadenomas. (2009) Arch. Immunol. Ther. Exp., 57, 295–301

Takami et al. (2008) Granulocyte-colony stimulating factor-producing pancreatic cancer: Report of a case. Surgery Today; Volume 38, Number 5, 453-457, (DOI: 10.1007/s00595-007-3636-z)

Tanji K, Irie Y, Uchida Y, Mori F, Satoh K, Mizushima Y, Wakabayashi K. (2003) Expression of metallothionein-III induced by hypoxia attenuates hypoxia-induced cell death in vitro. Brain Res.;976(1):125-9

Tang RF, Wang SX, Peng L, Wang SX, Zhang M, Li ZF, Zhang ZM, Xiao Y and Zhang FR. (2006) Expression of vascular endothelial growth factors a and c in human pancreatic cancer. World J Gastroenterol;12:280-286

Theory Clusters, Communication Processes, Network Theory and System Theory as provided by the University of Twente, December 2010.
http://www.utwente.nl/cw/theorieenoverzicht/Theory%20clusters/

Tlsty, T.D. and Coussens, L.M. (2006) Tumor stroma and regulation of cancer development. Annu. Rev. Pathol. 1, 119–150

Tullai JW, Schaffer ME, Mullenbrock S, Sholder G, Kasif S, Cooper GM. (2007) Immediate-early and delayed primary response genes are distinct in function and genomic architecture. J Biol Chem. 2007 Aug 17;282(33):23981-95

Tusher V, Tibshirani R, Chu C. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences, 98:5116-5121

Unoki M, Nakamura Y. 2003, EGR2 induces apoptosis in various cancer cell lines by direct transactivation of BNIP3L and BAK. Oncogene 22(14):2172-85.

Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE. (2003) PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. OMICS.;7(3):235-52

van Someren E. P., Wessels L. F. A., Backer E., and Reinders M. J. T. (2002) Genetic network modeling," Pharmacogenomics, vol. 3, no. 4, pp. 507–525

Verstrepen L, Verhelst K, van Loo G, Carpentier I, Ley SC, Beyaert R. (2010) Expression, biological activities and mechanisms of action of A20 (TNFAIP3). Biochem Pharmacol.;80(12):2009-20

Vogel J.H., von Heydebreck A., Purmann A. and Sperling S. (2005) Chromosomal clustering of a human transcriptome reveals regulatory background, BMC Bioinformatics 6, p. 230

von Heydebreck A, Huber W, Gentleman R. (2004) Differential expression with the Bioconductor Project. Bioconductor Project Working Papers

Vonlaufen A, Joshi S, Qu C et al. (2008) Pancreatic stellate cells: partners in crime with pancreatic cancer cells. Cancer Res; 68: 2085–93

Vonlaufen A, Phillips PA, Xu Z, Goldstein D, Pirola RC, Wilson JS, Apte M.V. (2008) Pancreatic stellate cells and pancreatic cancer cells: an unholy alliance. Cancer Res, 68:7707–7710

Vonlaufen A. et al. (2010) Isolation of Quiescent Human Pancreatic Stellate Cells: A Promising in vitro Tool for Studies of Human Pancreatic Stellate Cell Biology. Pancreatology;10:434-443 (DOI: 10.1159/000260900)

Walker M.G., Volkmuth W., Sprinzak E., Hodgson D., Klingler T. (1999) Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res.;9:1198–1203

Warshaw AL, Ferandez-Del Castillo C. (1992) Pancreatic carcinoma. N Engl J Med;326:455– 65

Watari N, Hotta Y, Mabuchi Y (1982) Morphological studies on a vitamin A-storing cell and its complex with macrophage observed in mouse pancreatic tissues following excess vitamin A administration. Okajimas Folia Anat Jpn, 58:837-858

Websters Online Dictionary, access 01.01.2011 http://www.websters-online-dictionary.org

Wells RG, Crawford JM (1998) Pancreatic stellate cells: the new stars of chronic pancreatitis? Gastroenterology, 115:491-493

Wellner U., Brabletz T. and Keck T., 2010, ZEB1 in Pancreatic Cancer, Cancers, 2, 1617-1628

Wei Y, Jiang J, Sun M, Chen X, Wang H, Gu J. (2006) ATF5 increases cisplatin-induced apoptosis through up-regulation of cyclin D3 transcription in HeLa cells. Biochem Biophys Res Commun.;339(2):591-6.

Weisburger JH, Williams GM (1995) Causes of cancer. In: Murphy GP, Lawrence W, Lenhard RE, eds: American Cancer Society Textbook of Clinical Oncology. 2nd ed. Atlanta, GA: American Cancer Society;10–39

Werhli V., Grzegorczyk M., and Husmeier D. (2006) Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks. Bioinformatics, vol. 22, no. 20, pp. 2523–2531

Whipple A.O, et al. (1935) Treatment of carcinoma of the ampulla of Vater. Ann Surg;102:763–779

Wider C. et al., (2009) Phactr2 and Parkinson's disease. Neurosci Lett.; 453(1): 9–11. doi: 10.1016/j.neulet.2009.02.009.

Wikipedia, The online community-driven encyclopedia. (2011) http://www.wikipedia.org

Williams E.J. and Hurst L.D. (2002) Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes, J. Mol. Evol. 54, pp. 511–518.(12)

World Health Organization “Data and Statistics” 2010: http://www.who.org accessed Oct.2010

Wu TC. (2007) The role of vascular cell adhesion molecule-1 in tumor immune evasion. Cancer Res;67:6003–6

Xu J, Kochanek KD, Murphy SL, Tejada-Vera B (2010) Deaths: final data for 2007. Natl Vital Stat Rep;58(19)

Xu Z., Vonlaufen A., Phillips P.A., Fiala-Beer E., Zhang X., Yang L., Biankin A.V., Goldstein D., Pirola R.C., Wilson J.S., and Apte M.V. (2010) Role of Pancreatic Stellate Cells in Pancreatic Cancer Metastasis. The American Journal of Pathology, Vol. 177, No. 5, DOI: 10.2353/ajpath.2010.090899

Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature.;467(7319):1114-7.PMID: 20981102

Yamasawa K., Nio Y., Dong M., Yamaguchi K., Itakura M. (2002) Clinicopathological significance of abnormalities in Gadd45 expression and its relationship to p53 in human pancreatic cancer. Clin. Cancer Res., 8: 2563-2569

Yanai I, Korbel JO, Boue S, McWeeney SK, Bork P, Lercher MJ. (2006) Similar gene expression profiles do not imply similar tissue functions. Trends Genet.;22:132–138

Yang A.D., Camp E.R., Fan F., Shen L., Gray M.J., Liu W., Somcio R., Bauer T.W., Wu Y., Hicklin D.J. and Ellis L.M. (2006) Vascular endothelial growth factor receptor-1 activation mediates epithelial to mesenchymal transition in human pancreatic carcinoma cells. Cancer Res;66:46-51

Yauch, R.L. et al. (2008) A paracrine requirement for hedgehog signalling in cancer. Nature 455, 406–410

Yazhou C, Wenlv S, Weidong Z, et al. (2004) Clinicopathological significance of stromal myofibroblasts in invasive ductal carcinoma of the breast. Tumour Biol;25:290

Yen T.W., Aardal N.P., Bronner N.P., Thorning D.R., Savard C.E., Lee S.P., Bell R.H.Jr. (2002) Myofibroblasts are responsible for the desmoplastic reaction surrounding human pancreatic carcinomas. Surgery, 131:129-134

Yimin G.E., Tarek Elghetany M. (2003) Urokinase Plasminogen Activator Receptor (CD87): Something Old, Something New; Laboratory Hematology 9:67-71

Yoshida S. (2005) Pancreatic Stellate Cells (PSCs) express Cyclooxygenase-2 (COX-2) and pancreatic cancer stimulates COX-2 in PSCs. Molecular Cancer, 4:27 doi:10.1186/1476-4598-4-27

Yu J., Smith V. A., Wang P. P., Hartemink A. J., and Jarvis E. D. (2004) Advances to Bayesian network inference for generating causal networks from observational biological data," Bioinformatics, vol. 20, no. 18, pp. 3594–3603

Yunis A.A., Arimura G.K., Russin D.J. (1977) Human pancreatic carcinoma (mia paca-2) in continuous culture: Sensitivity to asparaginase. International Journal of Cancer Volume 19, Issue 1, pages 128–135

Zhang, B., Kirov, S.A., Snoddy, J.R. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res, 33, W741-748

Zhang Y., Bharadwaj U., Logsdon C.D., Chen C., Yao Q. and Li M. (2010) ZIP4 Regulates Pancreatic Cancer Cell Growth by Activating IL-6/STAT3 Pathway through Zinc Finger Transcription Factor CREB. Clinical Cancer Research; doi: 10.1158/1078-0432.CCR-09-2405

Zhou Y, Young JA, Santrosyan A, Chen K, Yan SF, Winzeler EA. (2005) In silico gene function prediction using ontology-based pattern identification. Bioinformatics.;21:1237–1245

Zhong L., Roybal J., Chaerkady R., Zhang W., Choi K., Alvarez C.A., Tran H., Creighton C.J., Yan S., Strieter R.M., Pandey A. and Kurie J.M. (2008) Identification of Secreted Proteins that Mediate Cell-Cell Interactions in an In vitro Model of the Lung Cancer Microenvironment. Cancer Res 2008; 68: (17)

11. List of Figures

12. List of tables

Erklärung

Ich, Zbigniew Rogon-Lamparski, habe diese Dissertation selbst verfasst und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient. Experimentelle Daten bzw. Materialien, die nicht von mir selbst erhoben bzw. hergestellt wurden, habe ich besonders kenntlich gemacht.

Ich habe an keiner anderen Stelle ein Prüfungsverfahren beantragt und diese Dissertation auch nicht anderweitig in dieser oder anderer Form bereits als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt.

Teile der vorliegenden Arbeit wurden im Vorfeld in Absprache publiziert:

Busch, H., Camacho-Trullio, D., Rogon, Z., Breuhahn, K., Angel, P., Eils, R. and Szabowski, A. (2008) Gene network dynamics controlling keratinocyte migration, *Mol Syst Biol*, 4, 199.

Heidelberg, den _____

Zbigniew Rogon-Lamparski