



TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Genomorientierte Bioinformatik

**Regulatory networks of hematopoietic stem
cells and their micro-environment**

Baiba Vilne

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr.rer.nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. W. Wurst

Prüfer der Dissertation: 1. Univ.-Prof. Dr. H.-W. Mewes
2. Priv.-Doz. Dr. Robert A. J. Oostendorp

Die Dissertation wurde am *09.05.2014* bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München am *15.09.2014* angenommen.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Hematopoietic stem cells (HSCs), their origin, development and characteristic properties | 1 |
| 1.2 | The bone marrow niche | 2 |
| 1.2.1 | Cellular composition | 4 |
| 1.2.2 | Biophysical properties | 4 |
| 1.2.3 | Biochemical properties of the niche and extrinsic regulation of HSCs . | 5 |
| 1.3 | Intrinsic regulation of HSCs | 8 |
| 1.4 | HSCs regulate their niche | 10 |
| 1.5 | Biological networks and systems biology of hematopoiesis | 11 |
| 1.5.1 | Different types of biological networks | 13 |
| 1.5.2 | Computer readable formats describing biological networks and network visualization | 15 |
| 1.5.3 | Topological structure analysis of biological networks | 16 |
| 1.5.4 | Mathematical modeling | 17 |
| 1.6 | Computational modeling of the hematopoietic system | 19 |
| 1.7 | Motivation | 22 |
| 2 | Materials, methods and data | 23 |
| 2.1 | Materials | 23 |
| 2.2 | Methods and data | 31 |
| 2.2.1 | Experimental methods | 31 |
| 2.2.2 | Computational methods and data | 38 |
| 3 | Results | 43 |
| 3.1 | Time-series (TS) gene expression data generation | 43 |
| 3.2 | Microarray data pre-processing and quality control | 46 |
| 3.3 | Computational analysis of the time-series (TS) gene expression data | 47 |
| 3.4 | Independent confirmation of the time-series gene expression data | 50 |
| 3.5 | Phenotypic and functional comparison of freshly isolated vs. 24 h co-culture-derived LSK cells | 61 |
| 3.6 | Identifying Ctgf using candidate gene prioritization | 62 |
| 3.6.1 | Candidate gene prioritisation in LSK cells after 24 h co-culture (Day1; d1) | 64 |
| 3.6.2 | Candidate gene prioritisation in UG26-1B6 cells after 24 h co-culture | 67 |

| | | |
|----------|--|------------|
| 3.7 | Experimentally investigating the functional role of UG26-1B6-derived Ctgf in hematopoiesis | 71 |
| 3.7.1 | Independent confirmation of microarray results | 71 |
| 3.7.2 | Biological description and role in hematopoiesis | 72 |
| 3.7.3 | Generation of Ctgf shRNA knockdown in UG26-1B6 stromal cells | 77 |
| 3.7.4 | Ctgf regulates hematopoietic stem/progenitor cell (HSC/P) activity and engraftment potential | 79 |
| 3.8 | Network model predicting the role of Ctgf in regulating the LSK cell-cycle status | 80 |
| 3.8.1 | The top-down approach to Ctgf network modeling | 80 |
| 3.8.2 | The bottom-up or “seed-gene” approach to Ctgf network modeling | 88 |
| 4 | Discussion | 111 |
| 4.1 | Time-series (TS) gene expression data of co-cultured LSK and UG26-1B6 stromal cells | 111 |
| 4.1.1 | Gene expression (microarray) data generation | 111 |
| 4.1.2 | Computational analysis of the time-series (TS) gene expression data | 113 |
| 4.1.3 | Independent confirmation of the microarray data | 114 |
| 4.1.4 | Phenotypic and functional comparison of freshly isolated vs. 24 h co-culture-derived LSK cells | 116 |
| 4.2 | Identifying Ctgf as a novel regulator of hematopoiesis | 117 |
| 4.2.1 | Identifying Ctgf using candidate gene prioritization | 117 |
| 4.2.2 | Ctgf regulates hematopoietic stem/progenitor cell (HSC/P) activity and engraftment potential | 118 |
| 4.2.3 | Construction and dynamic analysis of the literature-derived Boolean network of Ctgf-regulated HSC cell cycle progression | 119 |
| 4.2.4 | Ctgf and its down-stream targets in the regulation of cell cycle and hematopoiesis | 125 |
| 5 | Conclusions | 131 |

List of Figures

| | | |
|------|--|-----|
| 1.1 | Overview of the hematopoietic hierarchy model. | 3 |
| 1.2 | Topological parameters of a network. | 18 |
| 3.1 | Workflow representing computational analysis vs. experimental validation of the results. | 44 |
| 3.2 | Time-course gene expression data generation from co-cultured Lin-Sca1+cKit+ (LSK) and UG26-1B6 stromal cells. | 45 |
| 3.3 | The quality assessment of time-course gene expression data. | 48 |
| 3.4 | STEM clustering and GO enrichment analysis of the two most significant clusters. | 51 |
| 3.5 | Microarray validation results using RT-qPCR for selected target genes. | 55 |
| 3.6 | Microarray validation results using RT-qPCR for selected differentially expressed genes in LSK cells. | 56 |
| 3.7 | Microarray validation results using RT-qPCR for selected differentially expressed genes in UG26-1B6 stromal cells. | 57 |
| 3.8 | Receiver operating characteristic (ROC) curve of the microarray performance in terms of the fold change and moderate t-test statistics. | 58 |
| 3.9 | Phenotypic and functional characterization of freshly isolated vs. 24 h co-culture-derived LSK cells. | 63 |
| 3.10 | The Venn diagram showing the overlap of ToppGene results between LSK and UG26-1B6 stromal cells. | 71 |
| 3.11 | A diagram showing the arrangement of CCN domains. | 74 |
| 3.12 | Connective tissue growth factor (Ctgf) is a putative novel regulator of hematopoiesis. | 75 |
| 3.13 | Expression of Ctgf mRNA in hematopoietic cells compared to UG26-1B6 stromal cells | 76 |
| 3.14 | Expression of Ctgf protein in the hematopoietic system compared to the stromal cells | 78 |
| 3.15 | From top-down to bottom-up approach to Ctgf network modeling. | 81 |
| 3.16 | The graphical network map depicting Ctgf regulated G0-to-G1 transition, G1/S block, as well as its auto-induction. | 93 |
| 3.17 | Attractor analysis of the Boolean networks modeling Ctgf regulated G0-to-G1 transition, G1/S block, as well as its auto-induction. | 100 |
| 3.18 | Attractor analysis of the Boolean networks modeling Ctgf regulated G0-to-G1 transition and its auto-induction without fixing Ctgf, WNTs and Tgf- β states. | 101 |
| 3.19 | Experimental validation of the Ctgf loss-of-function simulation results. | 104 |
| 3.20 | Experimental validation of the Ctgf loss-of-function simulation results. | 105 |

List of Tables

| | | |
|------|--|-----|
| 2.1 | Instruments | 23 |
| 2.2 | Consumables | 24 |
| 2.3 | Chemicals | 25 |
| 2.4 | Home-made solutions, buffers and media | 26 |
| 2.5 | Commercial buffers and media | 27 |
| 2.6 | Kits | 27 |
| 2.7 | Flow cytometry antibodies | 28 |
| 2.8 | Immunofluorescence (IF) antibodies | 29 |
| 2.9 | Secondary detection reagents for flow cytometry | 30 |
| 2.10 | Secondary antibodies for immunofluorescence (IF) | 30 |
| 2.11 | Mice | 30 |
| 2.12 | Cell lines | 30 |
| | | |
| 3.1 | Microarray validation results using RT-qPCR for selected target genes. | 53 |
| 3.2 | The contingency table containing the counts of the 4 combinations of classification. | 58 |
| 3.3 | ToppGene candidate gene prioritization results in LSK cells showing the top 15 highly ranked genes. | 65 |
| 3.4 | ToppGene candidate gene gene prioritisation results in UG26-1B6 stromal cells showing the top 15 highly ranked genes. | 69 |
| 3.5 | ToppGene candidate gene gene prioritization results in UG26-1B6 stromal cells showing the top 15 highly ranked genes. | 70 |
| 3.6 | Ctgf interactome topology analysis using NetworkAnalyzer. | 83 |
| 3.7 | Gene module detection within Ctgf interactome using the Maximal cliques algorithm provided by GraphWeb. | 85 |
| 3.8 | Gene module detection within Ctgf interactome using the Hub-based algorithm provided by GraphWeb. | 87 |
| 3.9 | Compiling a list of seed genes from the Ctgf interactome. | 89 |
| 3.10 | Boolean rules underlying the definition of the logical parameters describing Ctgf promoted G0-to-G1 transition, G1/S block and auto-induction. | 98 |
| 3.11 | Agreement of Boolean attractors for the three sub-processes with experimental protein measurements. | 106 |

List of Abbreviations

| | |
|----------|--|
| 2D | Two-dimensional |
| 5-FU | 5-fluorouracil |
| 7-AAD | 7-amino-actinomycin D |
| Akt(PKB) | Protein Kinase B |
| Ang-1 | Angiopoietin-1 |
| ATRA | All-trans retinoic acid |
| AUC | area under the ROC curve |
| BM | bone marrow |
| BMP | Bone morphogenetic protein |
| BrdU | Bromodeoxyuridine |
| CaR | Calcium-sensing receptor |
| CDK | Cyclin-dependent kinase |
| CellML | Cell Markup Language |
| CFSE | 5- and 6-carboxyfluorescein diacetate succinimidyl ester |
| CFU-Ba | Basophil progenitor |
| CFU-Eo | Eosinophil progenitor |
| CKI | Cyclin-dependent kinase inhibitor |
| CLP | Common lymphoid progenitor |
| CML | Chemical Markup Language |
| CMP | Common myeloid progenitor |
| Ctgf | Connective tissue growth factor |
| DNA | Deoxyribonucleic acid |
| Dnmt1 | DNA methyltransferase 1 |

| | |
|---------------|--|
| Dnmt3a | DNA methyltransferase 3a |
| Dnmt3b | DNA methyltransferase 3b |
| EXCERBT | Extraction of Classified Entities and Relations from Biomedical Texts |
| FAO | Fatty-acid oxidation |
| FC | Fold change |
| FCS | Fetal calf serum |
| FDR | False discovery rate |
| FPR | False positive rate |
| G0 | Gap 0 phase of cell cycle |
| G2 | Gap 2 phase of cell cycle |
| GC | Gas chromatography |
| GMP | granulocyte/macrophage |
| GO | Gene Ontology |
| GRN | Gene-regulatory network |
| Gsk3- β | Glycogen synthase kinase-3 beta |
| H3K27me3 | Trimethylation at lysine 27 of histone 3 |
| H3K4me3 | Trimethylation at lysine 4 of histone 3 |
| HPRD | Human Protein Reference Database |
| HSC | Hematopoietic stem cells |
| Hz | Hertz, the unit of frequency in the International System of Units (SI) |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| KO | Knock-out |
| LC | Liquid chromatography |
| LIMMA | Linear Models for Microarray Data |
| LMPP | lymphoid primed multi-potent progenitors |
| LSC | Leukemic stem cell |
| LSK | Lin-Sca1+cKit+ stem cells |

List of Tables

| | |
|----------------|---|
| LT-HSC | Long-term haematopoietic stem cell |
| LTBMC | Long-term bone marrow culture |
| M | Mitotic phase of cell cycle |
| MA | Bland–Altman |
| MEP | megakaryocyte/erythroid |
| MINT | Molecular INTeraction database |
| miRNA | Mature micro Ribonucleic acid |
| MM | Mismatch oligonucleotide probes |
| MPP | Hematopoietic multipotent progenitor |
| mRNA | Messenger Ribonucleic acid |
| MS | Mass spectrometry |
| MSC | Mesenchymal stem cell |
| mTOR | The mammalian target of rapamycin |
| NK | Natural killer cell |
| NLP | Natural language processing |
| NRQ | Normalized relative quantity |
| OPN | Osteopontin |
| OPP | oligopotent progenitors |
| PcG | Polycomb-group proteins |
| PM | Perfect match oligonucleotide probes |
| PML | promyelocytic leukemia |
| PPAR- δ | Peroxisome proliferator-activated receptor |
| PPI | Protein- protein interaction |
| PRC1 | Polycomb Repressive Complex 1 |
| PRC2 | Polycomb Repressive Complex 2 |
| PSI-MI | Proteomics Standards Initiative Interaction |
| PTEN | Phosphatase and tensin homolog |

| | |
|----------|--|
| Ptn | Pleiotrophin |
| Ptn | Pleiotrophin |
| RA | Retinoic acid |
| RAR | Retinoic acid receptor |
| Rb | Retinoblastoma |
| RDF | Resource Description Framework |
| RNAi | RNA interference |
| ROC | Receiver operating characteristic |
| ROS | Reactive oxygen species |
| RT-qPCR | Real time quantitative reverse transcription polymerase chain reaction |
| SBML | Systems Biology Markup Language |
| Sfrp1 | Secreted frizzled-related protein 1 |
| Sfrp1 | Secreted frizzled-related protein 1 |
| ST-HSC | Short-term haematopoietic stem cell |
| STEM | Short Time-series Expression Miner |
| TAP | Tandem affinity purification |
| TF | Transcription factor |
| Thpo | Thrombopoietin |
| TPR | True positive rate |
| TS | Time-series |
| UG26-1B6 | Urogenital ridge-derived stromal cell clone |
| XML | Extensible Markup Language |
| Y2H | Yeast two-hybrid |

Abstract

Hematopoietic stem cells (HSC) reside in a specific supporting micro-environment termed the bone marrow '*niche*', which is thought to regulate their *self-renewal* vs. *differentiation into mature blood cells*, including lymphocytes, red blood cells, and platelets. Currently, precise interaction networks between HSCs and their micro-environment have remained poorly understood. Efforts to examine the reciprocal influence of these two entities have led to the generation of *in vitro* culture systems, including the UG26-1B6 stromal cell line, which supports the maintenance of HSCs, mostly via secreted factors, such as Secreted frizzled-related protein 1 (Sfrp1) and Pleiotrophin (Ptn). In the present work, by combining high-throughput 'omics' assay and phenotype data with bioinformatics and systems biology approaches, we hope to substantially facilitate and drive the discovery of novel molecular players, and start to unravel the complexity of biological networks controlling HSC behaviour within their 'niche'. As a starting point, we performed microarray gene expression analysis of three different time-points of Lin-Sca1+cKit+ (LSK, stem cells) co-cultured with UG26-1B6 to investigate the initial interactions during culture stress. Our analysis indicated that the most changes in gene expression in LSK and stromal cells occurred already during the first **24 h of co-culture**. In LSKs, **gene function enrichment analysis** revealed up-regulation of transcripts associated with cell migration and proliferation, whereas epigenetic modifiers mediating gene silencing were among the down-modulated transcripts. In UG26-1B6, enrichment of molecular signatures localized to mitochondria and were associated with metabolism, such as the mTOR signaling, point at the **metabolic stress induced** changes in gene expression, which may be LSK-independent. Further candidate gene **prioritisation using a training set of hematopoiesis-related genes** ranked high a secreted matrix remodeller, **Connective tissue growth factor (Ctgf)** which was among the most significantly **up-regulated** transcripts in both LSK and stromal cells, although in UG26-1B6 this activation was LSK-independent. Since Ctgf has been reported to auto-induce its own expression, we hypothesized that its up-regulation in LSK (stem cells) may be attributed to extrinsic, UG26-1B6-derived Ctgf. Therefore, we used RNAi perturbation of secreted Ctgf levels in UG26-1B6 in order to investigate the role of extrinsic Ctgf in hematopoiesis. Our phenotypic and functional assays using HSC co-culture with UG26-1B6^{siCtgf} demonstrated increased hematopoietic progenitor activity *in vitro* and decreased engraftment potential *in vivo*. To delineate the possible underlying molecular mechanisms, we constructed a **literature-based network map delineating Ctgf auto-induction and, considering the importance of cell cycle regulation in HSC fate decisions, also linking Ctgf to HSC cell cycle progression**, in particular G0/G1 transition and G/S block. Dynamic simulation using the Boolean logic followed by experimental measurements of **mRNA and/or (phospho-)protein levels** of several network species in 24 h UG26-1B6^{siCtgf} co-culture-derived LSK cells suggested a possible **cross-talk between the Wnt/ β -catenin and PTEN/Akt(PKB)/Gsk3- β pathways**, leading to the inhibition

of glycogen synthase kinase-3-beta ($Gsk3-\beta$) via Akt(PKB)-dependent phosphorylation of Ser9, which, on its turn, facilitated the nuclear accumulation of β -catenin, resulting in the induction of its downstream target Cyclin D1 and cell cycle progression. Whereas, the absence of extrinsic Ctgf led to the activation of the tumor suppressor PTEN and the cell cycle inhibitor p27Kip1, resulting in G1/S cell cycle block, which may be accompanied by HSC differentiation, as indicated by our phenotypic and functional experiments. Hence, this approach allowed us for the identification of functionally relevant interactions and new molecular players, supporting the value of bioinformatics-assisted hypothesis-driven target discovery.

Zusammenfassung

Hämatopoetischen Stammzellen (HSZ) befinden sich in einer speziellen unterstützenden Mikroumgebung des Knochenmarks, - der so genannten "Nische", die vermutlich ihre Selbsterneuerung versus Differenzierung in reife Blutzellen, einschließlich Lymphozyten, rote Blutkörperchen und Blutplättchen, reguliert. Derzeit sind präzise Interaktionsnetzwerke zwischen HSZ und deren Mikroumgebung noch nicht erfasst worden. Die Bestrebung die gegenseitige Beeinflussung der beiden Einheiten zu untersuchen hat zu der Generierung von in-vitro-Kultur-Systemen geführt, einschließlich der UG26-1B6 Stromazelllinie, die die Aufrechterhaltung von HSCs unterstützt, meistens über sekretierte Faktoren wie das Sekretierte frizzled-related protein 1 (Sfrp1) und Pleiotrophin (Ptn). In der vorliegenden Arbeit, durch die Kombination von Hochdurchsatz-("Omics")-Assays und Phänotypdaten mit bioinformatischen und systembiologischen Ansätzen, hoffen wir, die Entdeckung neuer molekularer Spieler voranzutreiben sowie die Komplexität von biologischen Netzwerken die das Verhalten der HSZ in dessen "Nische" steuern anfangen zu enträtseln. Als Ausgangspunkt, führten wir Microarray-Genexpressionsanalyse von drei verschiedenen Zeitpunkten der Lin-Sca1+cKit+ Zellen (LSK, Stammzellen) co-kultiviert mit UG26-1B6 durch, um die ersten Interaktionen während des Zellkulturstress zu untersuchen. Unsere Analyse zeigte, dass die meisten Veränderungen in der Genexpression in den LSK- und Stromazellen bereits während der ersten 24 St. der Co-Kultur aufgetreten worden sind. In den LSK-Zellen, zeigten die Anreicherungsanalyse der Genfunktion eine Hochregulierung der Transkripten, die mit der Zellmigration und Zellproliferation assoziiert sind. Gleichzeitig waren die Gen-Silencing-vermittelnden epigenetischen Regulatoren bei den geringer exprimierten Transkripten zu finden. In den UG26-1B6-Zellen wurden Bereicherung molekulare Signaturen die in den Mitochondrien lokalisiert und mit dem Stoffwechsel verbunden sind, wie z.B. mit den mTOR Signalweg, weisen wieder darauf hin, dass die Veränderungen in der Genexpression möglicherweise von dem metabolischen Stress induziert sein könnten und daher unabhängig von den LSK-Zellen auftreten würden. Weitere Kandidatengen-Priorisierung mit einem Trainings-Set von Hämatopoese-relevanten Genen rangierte hoch einen sezernierten Matrixremodeler, *Connective tissue growth factor* (Ctgf), der zu den meist hochregulierten Transkripten in den LSK- und Stromazellen gehörte, wobei in UG26-1B6 wahr diese Hochregulation LSK-unabhängig. Da es vorher berichtet wurde, dass Ctgf seine eigene Genexpression induzieren kann, stellten wir die Hypothese auf, dass dessen Hochregulierung in den LSK-Stammzellen eventuell ein Ergebnis des extrinsischen, UG26-1B6-sekretierten Ctgf-Moleküls sein könnte. Um die Rolle des extrinsischen Ctgf in der Hämatopoese zu untersuchen, verwendeten wir die experimentelle Perturbation der sekretierten Ctgf-Proteinmenge in den UG26-1B6-Stromazellen mit Hilfe der RNAi-Technik. Unsere phänotypischen und funktionellen Assays unter Verwendung von HSC-Co-Kultur mit UG26-1B6^{siCtgf} demonstrierten eine erhöhte Aktivität der hämatopoetischen Vorläufer *in vitro*, sowie ein verringertes engraftment Potential *in vivo*. Um die möglichen zugrunde liegenden molekularen Mechanismen zu beschreiben, haben wir eine Netzwerk-Karte konstru-

iert, die eine Verknüpfung zwischen Ctgf und der Regulation von Zellzyklus, insbesondere dem G0/G1-Übergang und der G1/S-Unterbrechung, sowie der Autoinduktion, erstellte. Die Dynamische Simulation unter Verwendung der Booleschen Logik deutete darauf hin, dass Ctgf möglicherweise mit dem kanonischen Wnt/ β -Katenin, sowie dem PTEN/Akt(PKB)/Gsk3- β Signalwegen in Verbindung gebracht werden kann. Die experimentelle Messungen der mRNA- und/oder (Phospho-)Protein-Menge von einigen ausgewählten Netzwerk-Molekülen in den aus der 24 St. UG26-1B6^{siCtgf} Co-Kultur-gewonnenen LSK Zellen ermöglichten uns die Simulationsergebnisse zu validieren. Im Einklang mit der Vorhersage konnte eine signifikante Abnahme der Cyclin-D1-Proteinmenge, p21Cip1- und Ctgf-mRNA und Proteinmenge, sowie eine signifikante Erhöhung der p27Kip1-Proteinmenge beobachtet werden. Wir konnten ebenfalls eine deutlich erniedrigte Menge von phospho-GSK3- β (Ser9) experimentell bestätigen, während die Abundanz von phosphorylierten Ser33/Ser37/Thr41 β -Katenin signifikant erhöht war. Daher erlaubte uns dieser Ansatz die Identifizierung von funktionell relevanten Interaktionen und neuen molekularen Spielern. Damit konnten wir den Wert bioinformatischer Methoden für die Generierung von Hypothesen zur HSZ/Stroma-Wechselwirkung zeigen.

1 Introduction

1.1 Hematopoietic stem cells (HSCs), their origin, development and characteristic properties

The lifelong maintenance of adequate numbers of mature blood cells ($\sim 4 \times 10^{15}$ cells over a lifetime)¹ depends on a rare ($\sim 0.05\%$) subset of bone marrow (BM) cells² - *hematopoietic stem cells (HSCs)*. These cells are endowed with a dual capacity to self-renew, generating a genetically identical copy of itself upon cell division, and to undergo multi-lineage differentiation. *Self-renewal* is the capacity of the HSC to generate a genetically and functionally identical copy of itself upon cell division. This can occur either asymmetrically, retaining HSC potential in one daughter cell and generating further differentiated progeny in the other daughter cell or symmetrically, expanding the number of HSCs. Alternatively, divisions that generate two *differentiated progeny* daughter cells diminish the HSC pool. The decision of self-renewal vs. differentiation is thought to be determined stochastically.³ At different times of ontogeny and in different environments, the probability of an HSC dividing either symmetrically, asymmetrically or fully differentiating is believed to vary.⁴ Postnatal HSC self-renewal is closely related to a common property of stem cells - *quiescence* in terms of the cell cycle. Adult HSCs have been shown to cycle with very slow kinetics: approximately 8% of *long-term haematopoietic stem cells (LT-HSCs)* asynchronously enter the cell cycle per day and 99% of LT-HSC divide on average every 57 days. Under steady-state conditions, most ($\sim 75\%$) *HSC reside in G0 (quiescent)* phase of cell cycle, 20% are in G1 phase, while only a small fraction ($\sim 5\%$ of LT-HSCs) is in S or G2/M phase.^{5,6} It is likely that maintenance of HSC quiescence and slow cell-cycle kinetics are critically involved in sustaining a self-renewing HSC compartment for life, allowing HSCs to avoid mutation accumulation.^{7,8} Indeed, when quiescence is disrupted (e.g., in case of p21Cip1 deficiency), HSCs long-term repopulating ability is lost.⁹ Moreover, quiescent HSCs are resistant to 5-fluorouracil (5-FU)-induced myelosuppression, suggesting protection of HSCs from various stresses induced by myelotoxic insults.⁷ In contrast to senescence, where the ability to undergo cell divisions is lost, under certain physiological conditions (e.g., after the balance of blood cells or HSC pool is disturbed)¹⁰ a cell can reawaken from the state

1 Introduction

of quiescence and enter the cell cycle in order to undergo different fate decisions, including differentiation,^{11–17} upon which HSCs rapidly lose their self-renewal potential, producing a hierarchy of increasingly **committed progenitor cells** (see Figure 1.1). In particular, long-term HSCs (LT-HSCs) give rise to **short-term HSCs (ST-HSCs)**, which in turn generate **common myeloid progenitors (CMPs)**, **common lymphoid progenitors (CLPs)** and **lymphoid primed multi-potent progenitors (LMPPs)**. CLPs are the precursors of **B and T cells** and also to **natural killer (NK)** and dendritic cells. CMPs give rise to **megakaryocyte/erythroid (MEPs)**, **granulocyte/macrophage (GMPs)** and **eosinophil (CFU-Eo)** and **basophil (CFU-Ba)** progenitors. GMPs differentiate into the committed precursors of **neutrophils and macrophages**. In fact, the distinction between the myeloid and lymphoid lineages has recently been challenged by the finding that T cell precursors retain the ability to ultimately give rise to macrophages,^{18–20} pointing at the fact that the “fate decision window” may actually be wider than once thought.

Remarkably, the stem cell model of hematopoiesis has also been extended to the concept of **‘leukemic stem cells’ (LSCs)** as critical components in a leukemic cell hierarchy. Stem and cancer cells share certain signaling pathways, regulating essential HSC attributes of self-renewing, replication and differentiation into progenies of leukemic blasts. Despite recent advances, treatment of leukemia is often not curative, which may be attributable to this small population of therapy-resistant malignant cells leukemia-initiating cells, commonly referred to as leukemia stem cells (LSCs). Contemporary leukemia research has focused on ways to specifically eliminate LSCs, since these are regarded as the root of leukemia origin and leukemia recurrence after seemingly successful therapy.^{12,16,21–24}

The *ex vivo* expansion of human hematopoietic stem cells (HSCs) has remained an important goal to develop advanced cell therapies for bone marrow transplantation and many blood disorders. Over the last several decades, there have been numerous attempts to expand HSCs in vitro using purified growth factors that are known to regulate HSCs. However, these attempts have been of limited success for clinical applications and a routine method for ex vivo expansion of human HSCs is still not available. It is known that *in vivo* adult hematopoietic stem cells in the bone marrow (BM) reside within micro-environmental niches created by non-stem cells, where they remain undifferentiated. Therefore, the signaling pathways occurring in this niche are important to understand and examine for the *ex vivo* expansion of HSCs.²⁵

1.2 The bone marrow niche

The concept of a ‘niche’ as a specific supporting tissue locale or micro-environment housing stem cells was first proposed for the human hematopoietic system by Schofield²⁶ more than

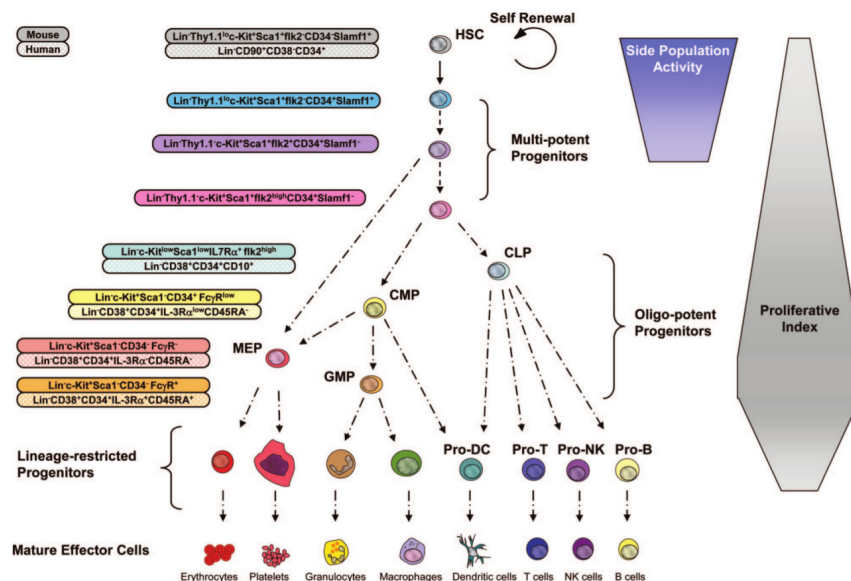


Figure 1.1: Overview of the hematopoietic hierarchy model. In the hematopoietic system, self-renewing HSCs reside at the top of the hematopoietic hierarchy, giving rise to at least 14 types of functional effector cells produced in successive differentiation processes of increasingly committed progenitor cells. Long-term HSCs (LT-HSCs) give rise to short-term HSCs (ST-HSCs), which in turn generate common myeloid progenitors (CMPs), common lymphoid progenitors (CLPs) and lymphoid primed multi-potent progenitors (LMPPs). CLPs are the precursors of B and T cells and also to natural killer (NK) and dendritic cells. CMPs give rise to megakaryocyte/erythroid (MEPs), granulocyte/macrophage (GMPs) and eosinophil (CFU-Eo) and basophil (CFU-Ba) progenitors. GMPs differentiate into the committed precursors of neutrophils and macrophages.¹⁸

30 years ago. Shortly thereafter, a connection between bone and blood dyscrasias was suggested.²⁷ The existence and importance of such micro-environments was later proven for the germ stem cell system of *Drosophila melanogaster*. Here, germline stem cells were demonstrated to be attached to differentiated somatic cells, essential for maintaining their survival and divisions.²⁸ Today, hierarchical stem cell systems and their niches have been identified in different mammalian tissues, including in hematopoiesis, muscle cells, central nervous system, intestinal epithelium, bulge region of the hair follicles, interfollicular epidermis and spermatogonial stem cells.^{29,30} For a long time, the haematopoietic stem cell (HSC) niche was poorly defined and considered a discrete site within the bone marrow. However, current evidence has pointed to the existence of different types of *niches* in the BM: most prominently, (i) an *endosteal osteoblastic* and (ii) a *vascular/perisinusoidal* niche. The osteoblastic niche is located near the endosteum and consists of osteoblasts derived from mesenchymal precursors. Sinusoidal endothelial cells in bone marrow form the vascular niche.^{7,31–33} Although the functional differences between these spatially distinct niches are

1 Introduction

still unclear, they may play a complementary role in the regulation of HSCs in the BM. For example, it has been proposed that the osteoblastic niche, which is a hypoxic niche, maintains hematopoietic stem cells in a quiescent (slow cycling or G0) state, while the vascular niche, an oxygenic niche, supports stem/progenitor cell proliferation, differentiation and mobilization.³⁴ Indeed, recently, using real-time imaging, it has become possible to explore the localization of HSCs in relation to their function. It has been observed that HSCs lodge in the endosteal surface, osteoblasts or osteoprogenitor cells and blood vessels, particularly in trabecular regions, in the mouse calvaria, whereas more mature hematopoietic cells reside away from the endosteum.^{28,35}

1.2.1 Cellular composition

Bone marrow is a soft, spongiform tissue composed of *blood and blood vessels, fat, connective tissue* and small segments of bone (termed *trabecular bone*), where the connective tissue forms a meshwork of delicate bony plates and strands, permeated by numerous thin-walled blood vessels. Within the spaces of this tissue, hematopoietic cells at different stages of differentiation and their stromal cells are suspended. The trabecular surfaces are covered by a layer of *endosteal cells, osteoblast (OB) lineage cells* at many stages of maturation, *osteoclasts* and their precursors (developed possibly from *mesenchymal stem cells (MSCs)*).^{4,30,36} All these diverse cell types have been implicated in regulation of HSC maintenance within the niche, however the precise cellular and molecular contribution of each cell population to the HSC-supportive microenvironment is still unclear.³⁷

1.2.2 Biophysical properties

More recently, the stem cell field has begun to appreciate that stem cell micro-environments present specific biophysical cues that may influence stem cell behavior. An important component of the niche are diverse *mechanobiological inputs*, resulting from *stiffness* (elastic modulus), which varies widely both between different tissues and within individual tissues. For example, cells near vasculature experience a 1 Hz cyclic strain or repetitive stretch, due to pulsatile blood flow. It has been shown that this strain is able to interact synergistically with secreted factors in conditioned media such as members of the Tgf- β /activin signaling pathway, thus, transducing the mechanical signal (cyclic strain) into intracellular biochemical signals, resulting in inhibition of cell differentiation. More recently, the contribution of contractile forces to the determination of the stem cell fate was also reported. Shin et. al. showed that myosin-II isoforms sense matrix stiffness in hematopoietic stem and progenitor cells, with polarized myosin-IIB promoting asymmetric self-renewal and constitutive myosin-IIA activation promoting cytokine-triggered differentiation.³⁸ Similarly,

the spatially inhomogeneous presentation of extracellular ligands and surrounding cells contributes to the induction and maintenance of **cell and tissue polarity**, relevant to cell division, homeostasis and tumorigenesis and differential segregation of stem cell fate determinants to daughter cells. Finally, the role of **temporally dynamic signaling** is already well recognized in developmental biology, as distinct morphogen gradients regulate tissue patterning at different stages of development. Recent work has also shown that cells in general and stem cells in particular respond not only to static concentrations and gradients, but can also be strongly influenced by exposure to temporally evolving ligand fields. In addition, improved imaging technologies have allowed observations of intracellular signaling fluctuations at the time scale of minutes and seconds, providing further evidence that cells can track and respond to these temporally encoded signals.³⁹ These observations may be of relevance when modeling the interactions between HSCs and their micro-environment and should be considered in future.

1.2.3 Biochemical properties of the niche and extrinsic regulation of HSCs

Direct physical **cell-cell** and **cell-extracellular matrix** interaction between HSCs and specific supportive micro-environmental cells, expressing diverse adhesion molecules may localize stem cells within specific niche compartments, where they are in close proximity to locally **secreted** or **membrane-bound cytokines, chemokines, hormones and growth factors** forming gradients that can initiate specific signal transduction within the HSC.⁴⁰ Moreover, the role of **miRNAs** (such as *miR-125b*, *miR-126*) in coordinating these processes is also beginning to emerge.^{41,42}

Tie2/Angiopoietin-1 (Ang-1) HSCs expressing the Tie2 receptor tyrosine kinase are quiescent and anti-apoptotic and comprise a side population of HSCs, which adheres to the **osteoblastic niche** in adult BM. Angiopoietin-1 (Ang-1), a ligand of Tie2, is predominantly expressed by osteoblastic cells in endosteum. Functionally, it has been reported that Tie2/Ang-1 signaling activates **cell adhesion** molecules, such as β 1-integrin and N-cadherin, promoting HSC interactions with extracellular matrix and cellular components of the niche. In addition, it has been also found that Ang-1 inhibits HSC division *in vitro* and **promotes quiescence** of HSCs *in vivo*.⁷

Mpl/Thrombopoietin (THPO) Mpl/Thrombopoietin (Thpo) signaling synergistically induced **HSC proliferation** with other cytokines. When focusing on the effects of exogenous Thrombopoietin (Thpo) on HSCs in mouse long-term bone marrow cultures (LTBMC), it was shown that it can mediate the self-renewal of HSCs.⁴³ In addition, Thpo or Mpl deficient

1 Introduction

mice showed fewer HSCs in the BM,^{44–46} suggesting that Mpl/Thpo signaling is crucial for the ***maintenance of LT-HSCs***. It was observed that Mpl expression in LT-HSCs was closely correlated with cell cycle quiescence and Mpl+ HSCs were in close contact with THPO producing ***osteoblastic cells*** at the endosteal surface in trabecular bone area.⁴⁷

Wnt/ β -catenin signaling Wnt ligands have been identified as a key-signaling pathway for normal ***HSC homeostasis in vitro*** and *in vivo*.⁴⁸ Wnt/ β -catenin signaling can induce the ***expansion of HSCs*** and is activated in both myeloid and lymphoid malignancies. Constitutively active nuclear β -catenin signaling reduces quiescence of HSCs and blocks multi-lineage differentiation. Loss of hematopoietic stem cell function was associated, among others, with decreased expression of *Cdkn1a* gene encoding the cell cycle inhibitor ***p21Cip1 and altered*** integrin expression in LSK cells. Hence, activation of β -catenin enforced cell cycle entry of hematopoietic stem cells, leading to exhaustion of the long-term stem cell pool.^{49,50} It has also been reported that Wnt-inhibited microenvironments, created by the osteoblast-specific overexpression of *Dkk1*, result in increased numbers of proliferating HSCs and reduced ability to reconstitute the hematopoietic system of irradiated recipient mice, indicating that microenvironment-related Wnt/ β -catenin activity is crucial for the maintenance of HSC quiescence. Recently, the results from our group demonstrated that microenvironment-derived *Sfrp1* is required to maintain HSC homeostasis through extrinsic regulation of β -catenin, and that the general level of *Ctnnb1* is decreased in *Sfrp1*^{-/-} LSKs, MPPs and CLPs.⁵¹ Hence, the fine-tuning of Wnt/ β -catenin activity seems to be crucial for the long-term maintenance of stem cell ***quiescence***.⁵²

Tgf- β signaling The transforming growth factor beta (Tgf- β) superfamily is a large family of structurally related growth factors, including Tgf- β , activins and bone morphogenetic proteins (BMPs). In the context of hematopoietic system, *Bmp4* and Tgf- β 1, have emerged as important regulators of HSCs.⁵³ *Bmp4* is mainly known to be involved in the development of the ***hemangioblast***,^{54,55} however it also may act on self-renewal of HSCs as part of their microenvironment.^{56,57} In contrast, ***the role of Tgf- β 1 on HSCs is more controversial***. *In vitro*, it has been shown to inhibit the growth of primitive hematopoietic progenitor cells and maintain HSC properties^{58,59} by regulating cell cycle molecules, including the induction of cell cycle inhibitors from the CIP/Kip (e.g., p57Kip2, p21Cip1) or INK4 families^{60,61} and repression of *Cdk4* or *c-Myc*.⁶² Experiments performed at the single-cell level have revealed that Tgf- β 1 maintains HSCs in quiescence by inhibiting cytokine-mediated lipid raft clustering.⁶³ On the other hand, *in vivo* experiments using knockout animals have failed to conclusively demonstrate the involvement of Tgf- β 1 in HSC maintenance, in part due to the embryonic lethality of Tgf- β 1 KO models.⁵³

Cell adhesion molecules Physical interaction between stem cells and their niche components are thought to participate in stem cell regulation through processes such as *contact dependent inhibition of proliferation*. N-cadherin, β 1-integrin, osteopontin and other cell adhesion molecules might not only be required for HSC anchoring to the niche, but also involved in the regulation of cell cycle status of HSCs. For example, in *Drosophila* germline stem cell, cadherin-mediated cell adhesion also regulates asymmetric cell division.⁶⁴ The increase in the number of spindle-shaped N-cadherin+ Cd45- osteoblastic (SNO) cells was shown to correlate with an increase in the number of HSCs and the long-term HSCs were found to be attached to SNO cells.⁶⁵ Tie2/Ang-1 signaling induces β 1-integrin and N-cadherin dependent HSC adhesion⁷. Mpl/THPO signaling also up-regulates β 1-integrin in LT-HSCs.⁴⁷ Moreover, it has been shown that the conditional inactivation of c-Myc induces excessive expression of integrins and N-cadherin in HSCs and Myc-deficient HSCs are not able to proliferate and detach from the niche due to uncontrollable cell adhesion.⁶⁶ Osteopontin (OPN) negatively regulates HSC number in the BM niche, and the lack of OPN results in an increase in the number of HSCs.^{67,68}

Metabolites and other small molecules The bone marrow is a highly complex system in which gradients of nutrients and other chemicals exist between areas of bone and areas of vascularization. Thus far, the contribution of these possible metabolic cues to hematopoietic stem cell function has *not been well understood*.⁶⁹

Reactive oxygen species (ROS) Quiescent stem cells reside in the low-oxygenic niche, hypoxic regions of tissues not rich in vasculature, such as the trabecular zone for hematopoietic stem cells (HSCs) with the lowest end of an oxygen gradient within the bone marrow. Here, within the osteoblastic niche, in contact with osteoblasts, HSCs remain quiescent. Such *low-oxygenic milieu in bone marrow limits reactive oxygen species (ROS)* production, thus providing long-term protection from ROS-related oxidative stress. In addition, stem cells have developed a unique mechanism to cope with the cumulative ROS load, which involves increased antioxidant defenses and unique redox-dependent effects on growth and differentiation. At the same time, in the relatively more oxygenic vascular niche, due to the proximity to blood circulation, stem cells actively proliferate and differentiate.⁷⁰

Retinoic acid (RA) The retinoic acid receptor (RAR) agonist, all-trans retinoic acid (ATRA), has been demonstrated to *enhance the maintenance and self-renewal* of short- and long-term re-populating hematopoietic stem cells from Lin-Sca1+cKit+ (LSK) hematopoietic precursors cultured in liquid suspension for 14 days. ATRA also prevented the differentiation of these primitive stem cells into the more mature pre-CFU-S population during the 7 to 14 days of culture.⁷¹

Calcium and calcium-sensing receptor (CaR) High concentration of calcium ions at the HSC-enriched endosteal surface is among the features of bone that contribute to a micro-environmental niche for stem cells. HSCs have been shown to express the seven-transmembrane-spanning calcium-sensing receptor (CaR), needed to respond to changing extracellular ionic calcium concentrations, which, on its turn, *dictate the preferential localization of adult mammalian haematopoiesis in bone*. For example, CaR-/- HSCs from mouse fetal liver were highly defective in localizing anatomically to the endosteal niche, due to defective adhesion to the extracellular matrix protein, collagen I.⁷²

1.3 Intrinsic regulation of HSCs

Cell cycle regulators

The distinct hematopoietic stem cell fates of self-renewal and differentiation likely depend on the entry into the cell cycle and the cell division, necessitating the presence of appropriate cell cycle machinery to effect passage into and through G1 phase. This depends on *competing actions of Cyclin-dependent kinase (CDKs)*, driving cell cycle progression, *and CDK inhibitors (CKIs)*, which hinder progression through the cell cycle.^{73,74} For example, conditional deletion of retinoblastoma (Rb), family of transcriptional repressors (consisting of the pRb, p107 and p130 proteins) results an increase in both HSC proliferation and absolute cell numbers, as well as by severe defects in HSC self-renewal. Mice deficient in all three D-cyclins have lower numbers of HSCs and progenitor populations in the fetal liver, with decreased frequency of HSCs in S and G2-M stages of the cell cycle. Deletion of p18Ink4c (a cell cycle inhibitor (CKI) of the Ink4 family), the expression of which is highest in quiescent HSCs, results in increased numbers of actively cycling HSCs, without affecting HSC self-renewal activity. Initially, a role for p21Cip in regulating HSC quiescence was suggested,⁹ however more recent reports indicate that the function of p21Cip in regulating HSC cell cycle activity may be restricted to periods of stress rather than during homeostasis.^{75,76} p27Kip1 deficiency appears to affect the cell cycle activity of more committed progenitor populations.⁷⁷ A more recent work also suggests that p57Kip2 is critical for maintaining HSC quiescence.⁷⁸

Transcription factors

Transcription factors (TFs) have attracted much attention, since the key regulators of both HSC self-renewal and differentiation into the various mature hematopoietic lineages are commonly encoded by transcription factor genes.^{79,80} Recently, genome-wide analysis of hematopoietic TFs has provided new evidence that they operate in a complex combinatorial manner.³³ For example, Gfi1, Pten and Foxo have been found to restrain HSCs from excessive

cycling, whereas Zfx and Tel/Etv6 are critical in suppressing HSC apoptosis.^{8,81-83}

Epigenetic modifiers

Findings from studies on transcription factors have also increasingly focused attention on the epigenetic modifications as a way of coordinating the expression and activity of transcription factors and their target genes by changing the chromatin structures and hence influencing the accessibility of transcription factors to DNA.^{33,84,85} In this context, it has been demonstrated that hematopoietic differentiation correlates to a stepwise decrease in the transcriptional accessibility of multi-lineage-affiliated genes, as a result of concerted epigenetic modifications by DNA methylation and histone modification.⁸⁶⁻⁸⁹ Interestingly, myeloid commitment involved less global DNA methylation than lymphoid commitment.⁹⁰ In general, undifferentiated human and murine hematopoietic cells display less-condensed chromatin structures and exhibit a higher rate of histone acetylation, indicating a state of higher turn-over of chromatin structures.⁹¹

Polycomb-group (PcG) proteins are histone modifiers residing in two multi-protein complexes: Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2). PRC2 catalyses and maintains trimethylation at lysine 27 of histone 3 (**H3K27me3**) responsible for the recruitment of a second complex, PRC1.⁹² H3K27me3 mark has been generally associated with **transcriptional repression**.⁹³ In the context of hematopoiesis, (PcG) proteins have been shown to be involved in the regulation of balance between self-renewal and differentiation of hematopoietic stem cells (HSCs) by repressing genes involved in cell-cycle regulation and differentiation.⁹⁴ Likewise, aberrant expression of, and mutations in, PcG genes have been associated with hematopoietic malignancies, such as hematopoietic neoplasms, where these histone modifiers display both tumor-suppressor and oncogenic activities.⁹⁵

H3K4me3 trimethylation at lysine 4 of histone 3 has in general been associated with **gene expression**.⁹³ In fact, several H3K4me3 regulators have been implicated in the ability of stem cells to self-renew and differentiate.⁹⁶ For example, an H3K4me3 methyltransferase - **Mll1** - has been shown as essential for the maintenance and self-renewal of fetal and adult hematopoietic stem cells (HSC).⁹⁷

DNA methyltransferases *Dnmt3a and Dnmt3b* are mainly involved in *de novo* establishment of methylation patterns during cellular differentiation, whereas **DNA methyltransferase 1 (Dnmt1)** is responsible for maintaining genomic methylation. Dnmt3a silences hematopoietic stem cell self-renewal and is essential for hematopoietic stem cell differentiation, as in Dnmt3a-null HSCs up-regulation of HSC multipotency genes and down-regulation of differentiation factors has been observed.⁹⁸ In Dnmt1-deficient HSCs, defects were observed in self-renewal, niche retention, as well as in the ability of HSCs to give rise to myeloid

progenitor cells.⁹⁹

micro-RNAs

More recent studies are also beginning to reveal the role of microRNAs in the regulation of hematopoietic stem/progenitor cells (HPSCs) in the hematopoietic system.^{41,42} An evolutionarily conserved microRNA cluster consisting of miR-99b, let-7e and miR-125a was found to be preferentially expressed in long-term hematopoietic stem cells. MicroRNA miR-125a controls the size of the stem cell population by regulating HPSC apoptosis, possibly through targeting multiple proapoptotic genes.¹⁰⁰ Another miRNA, expressed in HSCs and early progenitors, miR-126 has been reported to play a role in restraining cell-cycle progression of HSCs *in vitro* and *in vivo*, most likely by regulating multiple targets within the PI3K/Akt(PKB)/Gsk3- β pathway.¹⁰¹

Lipid rafts and fatty acids

Lipid rafts The regulation of lipid raft clustering on the surface of HSCs may be a critical determinant of HSC quiescence by dictating the level of Akt activation induced by cytokine receptors. Quiescent HSCs show minimal amounts of lipid raft clustering, while actively proliferating hematopoietic progenitor cells have high levels of clustering and Akt pathway activation. Interestingly, lipid raft clustering in HSCs has been demonstrated to be suppressed by Tgf- β signaling.^{102,103}

Fatty acids Recently, a previously unknown promyelocytic leukemia (PML) -peroxisome proliferator-activated receptor δ (PPAR- δ) -fatty-acid oxidation (FAO) pathway was identified to be involved in the maintenance of hematopoietic stem cells (HSCs) by controlling the *asymmetric division* (retaining HSC potential in one daughter cell and generating further differentiated progeny in the other daughter cell) of HSCs.⁶⁹

1.4 HSCs regulate their niche

Currently, little is known about whether and how HSCs themselves regulate the maintenance or development of the niche. Although reciprocal cooperation between HSCs and the elements of their microenvironments in establishing the niche has been proposed, there is little direct experimental evidence that would support the concept. For example, it has been suggested that HSCs may guide mesenchymal differentiation toward the osteoblastic lineage under basal conditions and that HSC-derived bone morphogenic factors such as Bmp2 and Bmp6 could be responsible for these activities¹² In another study, HSCs in Nf2-deficient mice were increased in number and demonstrated a marked shift in location to the circulation. with an associated *increase in trabecular bone mass and stromal cell numbers*, as well as vascularity and Vegf levels.¹⁰⁴

1.5 Biological networks and systems biology of hematopoiesis

“If there is any area in which a network thinking could trigger a revolution, I believe that biology is it.”

–Albert László Barabási

Like other cells, hematopoietic stem cells (HSCs) constantly receive environmental cues in many forms: soluble cues such as mitogens and cytokines,^{7,8,105–110} small molecules^{71,111,112} and nutrients,^{70,113,114} as well as ‘solid phase’ cues such as cell-cell contacts and the biochemical and mechanical properties of the extracellular matrix.^{115–117} These signals guide the stem cell towards specific fate decisions: quiescence, differentiation, self-renewal, migration, senescence or apoptosis.¹⁷ Hence, HSC behavior is guided by molecular interactions and reactions involving receptors, signaling intermediates and transcription factors. In particular, signal processing networks relaying input signals from the extracellular space and cell surface to the nucleus feature complex, non-linear components such as feed-forward and feedback loops, signal amplification cascades and cross-talk between multiple signaling pathways. Thereafter, information processing further continues within the cell nucleus where transcription factor networks control the expression of themselves and each other, as well as of their corresponding target genes that are required for execution of the particular fate decisions. This results in a complex, multi-level, non-linear system, which can exhibit a number of different behaviors, including switches and oscillations. Such behaviors are difficult to investigate and interpret intuitively without the aid of systems-level analysis.¹¹⁸

Systems biology approaches offer advantages that complement and enhance traditional experimental strategies tending to focus more on individual components than on interactions occurring within a larger system. Systems biology analyses often rely on computational models that: (1) summarize our knowledge of and assumptions about a system into formal, mathematical statements; (2) highlight gaps in our knowledge of a system; (3) help to generate hypotheses about the behavior of the system that guide experimentation and further modeling; (4) aid in the analysis of large datasets, such as those generated by genomic, transcriptomic, proteomic and kinomic experiments, thus summarizing the data and highlighting important, potentially unintuitive behavior for future experimentation and, finally, (5) highlight critical loci within a system that can be manipulated to generate a desired outcome. Hence, a combination of experimental approaches with mathematical, engineering and computational

1 Introduction

tools is being used in order to gain global insights into complex biological systems and phenomena.¹¹⁸⁻¹²⁰

The *experimental techniques* utilized for systems biology studies usually tend to have high-throughput capabilities, being able to determine the abundance and/or activity of large numbers of components simultaneously. For example, the abundance of mRNA transcripts of thousands of genes can be profiled by microarrays or RNA-seq. Quantitative protein concentrations and post-translational modifications can be determined using proteomics and phosphoproteomics approaches, such as mass spectrometry (MS) or two-dimensional (2D) gels. Metabolomic profiles generated by gas chromatography (GC)-MS or liquid chromatography (LC)-MS can measure the composition and concentration of both targeted and untargeted metabolites. Moreover, high-throughput techniques exist that can detect the interactions among components such as protein-protein interactions, transcriptional regulations (TF-DNA interactions) and genetic interactions. In addition to high-throughput assays, decades of genetic and molecular analyses using small-scale experiments studying fewer components and interactions involved in specific biological processes provide high-quality and reliable focused knowledge for biological systems. Altogether, these data provide a rich source for understanding the system-level mechanisms of biological processes and the identification and characterization of the molecular components and their reciprocal interactions involved in cell signaling has become possible in a systematic way.

Moreover, the systematic collection of this molecular information into web-accessible databases, has facilitated the reconstruction and mapping of ever larger and more complex biological networks.¹¹⁹ In addition, a large amount of knowledge about functional regulatory interactions and the components involved in these interactions is embedded in the biomedical research literature. Text mining is used to extract interactions using natural language processing (NLP) and information retrieval technologies.¹²¹

Using *network modeling*, experimentally obtained signaling pathway and protein-protein interaction (PPI) information can be translated into a graph (network) by representing proteins, transcripts and small molecules as network nodes and denoting the interactions between them as edges. The direction of edges follows the direction of the mass or information flow, from the upstream (source) node to the downstream (product or target) node. In addition, the edges are characterized by signs, where a positive sign indicates activation, whereas a negative sign indicates inhibition.^{119,122}

Computational methods that have been used in systems biology to analyze biological networks can be classified into *top-down and bottom-up approaches*. Top-down approaches, including statistical analyses and static network models, are commonly applied to high-throughput omics data and aim to decipher the organization of the underlying systems

and mine information specific to a biological process under study. Such methods do not require kinetic parameters and can be used for the analysis of genome-scale data with tens of thousands of components or interactions to obtain coarse-grained knowledge about biological systems.¹¹⁹ By this, important clues about the topological organization of the networks can be obtained and relationships between the topological characteristics and biological properties of the involved molecules elucidated. For example, studies on protein-protein-interaction (PPI) networks have revealed complex relationships between the number of neighbors of a node or vertex degree, network modularity (organization into connected subnetworks), gene essentiality and pleiotropy, and are being used increasingly to predict functionality of individual molecules in the network, membership in protein complexes, association with signaling pathways and disease-associated subnetworks.^{123–126} On the other hand, bottom-up methods model how interacting components, e.g., genes, proteins and metabolites achieve the dynamic behaviors of cellular systems. In this case, one usually starts with hypotheses of biological mechanisms generated from individual small-scale experiments.^{127–130} This class of methods is further described in the section *Mathematical modeling*.

Taken together, as opposed to the traditional biological studies dealing with relatively few components and using intuitive reasoning to guide hypotheses and experiments, systems biology approaches, although still in their infancy, allow the collection of molecular information in a systematic way and formulation of a hypothesis that can be a powerful source in directing targeted experiments.

1.5.1 Different types of biological networks

Protein-protein interaction (PPI) networks, stored as undirected graphs, mainly hold information of how different proteins operate in coordination with others to enable the biological processes within the cell. Several large-scale and high-throughput techniques have been developed enabling to detect interacting proteins within an organism. Among them, the most well-known are the tandem affinity purification (TAP),¹³¹ yeast two-hybrid (Y2H),⁶⁹ protein microarrays¹³² and flow cytometry enabling single cell proteomics.¹³³ To make this information concerning PPI data more readily available, a number of ***publicly available databases have set out to collect and store protein-protein interaction data***, such as The Human Protein Reference Database (HPRD)¹³¹ the Molecular INTeraction database (MINT)¹³⁴ and IntAct.¹³⁵ Additionally, well-documented services infer and store mammalian protein-protein interactions using orthologs, meaning that PPIs identified in lower organisms are identified to also exist in, mammalian cells, for example, IntNetDB¹³⁶ and STRING.¹³⁷ A number of data warehouses also consolidate different protein-protein databases by merging networks stored in different formats, including Pathway Commons¹³⁸ and UniHI.¹³⁹ Systems

1 Introduction

such as BioWarehouse¹⁴⁰ and the Gaggle¹⁴¹ provide integration and querying capabilities with other types of biological data in addition to protein-protein interactions, metabolic, gene-regulation and cell-signaling networks. Several studies have attempted to ***compare these different mammalian PPI databases*** to assess their overlap and coverage.^{142,143}

Gene-regulatory networks (GRNs), abstracted as directed graphs with activation/inhibition links, contain information concerning the control of gene expression in cells. Here, genes are translated to ***transcription factors (TFs)***, i.e., more distal regulators, ***regulating the expression of other genes***. Besides TFs, this process is modulated by many other variables, such as post-translational modifications of TFs or their association with other molecular factors. TFs often exhibit specific motifs and patterns concerning their topology. Data collection, data integration and rapidly emerging high-throughput technologies that can experimentally map gene-regulatory networks give now the possibility to study them in a larger scale. Gene expression microarrays time-series data derived from perturbation studies, ***ChIP-chip and ChIP-seq***, comparative genomics (identifying conserved non-coding sequences as potential binding sites) or purely computational approaches that use known consensus DNA binding motifs are typically used to reconstruct gene-regulatory networks.^{121,144} Several databases and tools are developed to collect and integrate such data for example, JASPAR¹⁴⁵ and TRANSFAC,¹⁴⁶ while ***post-translational modification*** can be found in databases like Phospho.ELM¹⁴⁷ and PHOSIDA.¹⁴⁸

Signal transduction networks, commonly represented as directed graphs with three types of links: activation, inhibition and neutral, in contrast to protein-protein interaction networks, capture ***functional relationships between different bioentities***. Besides proteins, signaling networks also include small molecules such as calcium and cAMP. Signal transduction networks investigate how signal transmission is performed either from the outside to the inside of the cell, or within the cell in order to regulate the response of cells to changes in the extracellular environment where signals, received at the cell surface by receptors, transduce information to effector proteins through cascades of coupled biochemical reactions (most commonly, phosphorylation). Thus, environmental parameters change the homeostasis of the cell and, depending on the circumstances, different responses can be triggered. Similarly to GRNs, these networks also exhibit common patterns and motifs concerning their topology.^{121,144} Databases that store information about signal transduction pathways include the Cancer Cell Map (<http://cancer.cellmap.org/cellmap/>), KEGG (Kyoto Encyclopedia of Genes and Genomes)¹⁴⁹ and BioCarta (<http://biocarta.com>).

Metabolic and biochemical networks are collections of pathways, holding information about ***a series of biochemical events*** and the way they are correlated and are being

constructed in order to study the metabolic pathways of an organism. Metabolic pathways consist of a series of biochemical reactions occurring within a cell at different time points. In general, metabolic networks are more complete and rich in quantitative information as compared with protein-protein interaction (PPI), cell-signaling and gene-regulatory networks. Within a metabolic network, the main role is played by the enzymes, since they are the main determinants in catalyzing biochemical reactions. Enzymes are often dependent on other co-factors such as vitamins for proper functioning.^{121,144} Currently, modern sequencing techniques have allowed the reconstruction of the network of biochemical reactions in many organisms, from bacteria to human.^{150,151} Several public databases exist holding information about biochemical networks in many organisms, including KEGG¹⁵², BioCyc¹⁵³ and metaTIGER.¹⁵⁴ In addition, analysis methods have also been proposed to elucidate the pathway structure of metabolic networks.^{155,156}

MicroRNA networks In addition, there is also a growing appreciation for non-canonical metabolites, non-protein biomolecules and non-conventional post-translational modifications functioning in intracellular regulation, for example, miRNA networks. miRNAs are short (~22 nucleotide) transcripts that pair with (full-length) mRNAs of transcribed and translated genes, thereby *suppressing their translation into proteins*. As microRNA sequences are usually known, it is computationally possible to construct the network of interactions between *miRNAs and their putative targets* within the expressed genome.¹²¹ Such examples include the work of Shalgi *et al.*,¹⁵⁷ who developed and analyzed a network of transcription factors and miRNAs. Cui *et al.*¹⁵⁸ used a large-scale signaling network extracted manually from the literature in order to assess how endogenous miRNAs target and regulate components in the cell-signaling system.

1.5.2 Computer readable formats describing biological networks and network visualization

Many computer readable formats have been proposed to describe biological networks in an attempt to develop standards for data sharing and exchange between isolated data sets and analysis tools. The Systems Biology Markup Language (SBML)¹⁵⁹ is an XML-like machine-readable language, designed to represent network models to be analyzed by a computer. SBML can represent metabolic networks, cell signaling pathways, as well as regulatory networks. Biological networks can also be described using file formats, such as the Proteomics Standards Initiative Interaction (PSI-MI),¹⁶⁰ Chemical Markup Language (CML)¹⁶¹ for chemicals or BioPAX¹⁶² for pathways. Several secondary network formats, which can also be used in for similar purposes include the Cell Markup Language (CellML),¹⁶³ which is an XML-like machine-readable language mainly developed for the exchange of computer-based

mathematical models, the Resource Description Framework (RDF),¹⁶⁴ which is a language for the representation of information about resources on the World Wide Web. Each of these storage schemata is designed for handling different types of biomolecular networks. For example, PSI-MI is most appropriate for describing details about experiments, SBML is can be used in order to directly explore biological networks into quantitative modeling tools such as the SBMLToolbox,¹⁶⁵ BioPAX has the advantage that it does not require quantitative information, therefore it is useful for network visualization and data exchange. In addition, in recent years several desktop and web-based applications for pathway and network visualization have emerged, for example, Cytoscape¹⁶⁶ and CellDesigner,¹⁶⁷ supporting different network storage formats.

1.5.3 Topological structure analysis of biological networks

The availability of large-scale curated interaction datasets has given rise to the opportunity to investigate *topological organization* of these interactomes using graph theoretic analysis.¹⁶⁸ Such analysis can be particularly useful in large signaling networks, where a simple visual inspection is not possible and at the same time the construction of precise quantitative models is practically infeasible due to the huge amount of required, but generally unknown, kinetic parameters and concentration values.¹⁶⁹ The topological structure of a network is thought to contain significant biological properties and plays an important role in understanding network architecture and performance. Several commonly used topological parameters include (Figure 1.2): (1) **Node degree** or the number of links connected to that node. For directed networks, a separation between the “in-degree” or the number of edges that end at the node and the “out-degree“, which is the number of edges that start from the node is being made. Functionally, a node with high degree is better connected in the network and therefore may play a more important role in maintaining the network structure. (2) **Distance** or the shortest path length between two nodes, where the maximum distance between any two nodes is termed as the graph diameter. The average distance and diameter of a network measure the approximate distance between nodes in a network. A network with a small diameter is often termed as a “small world“ network, in which any two nodes can be connected with relatively short paths. Actually, many real world networks such as metabolic networks have a small world architecture, which is thought to serve to minimize transition times between metabolic state. (3) **Clustering coefficient** of a node is calculated as the number of links between the nodes within its neighborhood divided by the number of links that are possible between them. A high clustering coefficient for a network is another indicator of a small world. (4) **Betweenness** is the fraction of the shortest paths between all pairs of nodes that pass through one node or link. Betweenness estimates the traffic

load through one node or link assuming that the information flows over a network primarily following the shortest available paths. Biological networks significantly differ from random networks, often exhibiting ubiquitous properties in terms of their structure and organization. For example, biological networks have a **”scale-free“** format containing hubs with many connections and a large number of nodes that have one or a small number of connections. In contrast to a bell-shaped degree distribution in random networks, scale-free networks have a typical **”power law“** distribution, $P(k) \sim k^{-\gamma}$, in which k is the node degree and $P(k)$ is the probability that a randomly selected node will have a degree k . Functionally, the advantage of this type of organization is that the system is more robust, meaning that random loss of individual non-hub or peripheral nodes is less disruptive. On the other hand, hub nodes are extremely important and therefore usually play essential roles in biological systems.¹⁷⁰ Many real-world networks, were demonstrated to be modular and hierarchically organized.¹⁶⁸ Such modules can be identified using network clustering algorithms, for example, betweenness centrality clustering.¹⁷¹ Moreover, after modules have been identified with only considering the network topology, the modules can be further validated by exploring whether the components in the module also share similar GO terms¹⁷² or GO terms and network connectivity can be combined for module identification.¹⁷³ Several tools have been developed to identify modules in networks. For example, MoNet,¹⁷⁴ MCODE,¹⁷⁵ MCL.¹⁷⁶

1.5.4 Mathematical modeling

Experimentally and computationally derived biological networks such as protein-protein interaction networks provide static depictions of the dynamically changing cellular environment. Clearly such large-scale interaction maps, usually including numerous intertwined feedback circuits, are not directly interpretable and sufficient by themselves, as they do not provide insights on the **logic of signaling networks and their spatio-temporal behavior**.¹²³ Moreover, we still lack technology that would allow high-throughput detailed measurement of activity of all signaling molecules and their interactions. This necessitates developing methods to prioritize selection of the molecules such that measuring their activity would be most informative for understanding the molecular crosstalk within the network. In this respect, computational modeling and simulation tools become a necessary complement, allowing the formulation of a systems-level hypothesis that can be a powerful source in directing targeted experiments. Generally, the roles of mathematical models for gene regulatory networks include: (1) describing genetic regulations at a system level; (2) enabling artificial simulation of network dynamic behavior; (3) predicting new structures and relationships among the network components; and, finally, (4) making it possible to analyze or intervene in the network through signal processing methods.¹²⁵ Different mathematical approaches

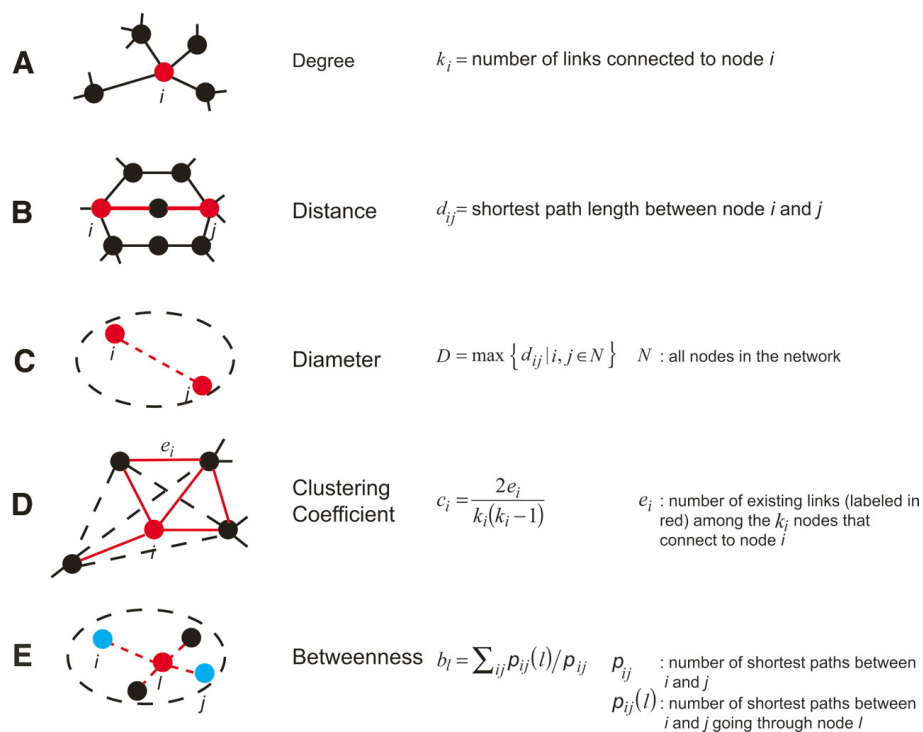


Figure 1.2: Topological parameters of a network. Five commonly used topological parameters are illustrated in both graphs and formulae, based on an undirected network. (A) Node degree measures the number of connections of that node. (B) Distance is the length of the shortest path between two nodes. (C) Diameter is the maximum distance between any two nodes in a network. (D) Clustering coefficient measures the percentage of existing links among the neighborhood of one node. (E) Betweenness is the fraction of those shortest paths between all pairs of nodes that pass through one node or link.¹⁷⁰

have been proposed to model such regulatory networks and to simulate their dynamical behavior,^{124–126,177} including also hybrid approaches.¹⁷⁸

Boolean mathematical formalism

Quantitative and continuous modeling (e.g., using differential equations) typically requires precise mechanistic details on molecular mechanisms and experimentally determined kinetic parameters describing the individual reactions such as synthesis and degradation rates. Consequently, such studies have been limited to a couple of small-scale (with only tens of components or less) well-understood sub-networks. On a genome-wide scale and for newly discovered pathways such detailed data may not be available. At the same time, a wealth of molecular level qualitative data on individual components and interactions can be obtained from the biomedical literature and high-throughput technologies, making methods, which are **straightforward, robust and compatible with qualitative data**, highly attractive to model and analyze essential properties of genetic regulatory and signaling

networks.^{119,122,124–130,179}

Boolean network models^{180,181} are among the simplest discrete models that have been applied to systems biology. In terms of complexity, Boolean networks lie between static network models and continuous dynamic models,¹⁸² making them a tractable and powerful approach to modeling also large-scale biological systems. BNs are able to capture some fundamental characteristics of gene regulations, are conceptually simple and their rule-based structures bear physical and biological meanings. Hence such models can lead to predictive testable hypotheses which is especially valuable in poorly understood large-scale systems.^{183,184} Boolean networks have been successfully applied in modeling many gene regulatory and signaling networks in a variety of organisms.^{122,185–187} The attractors or stable states of BNs have been associated with cellular phenotypes, inspiring the development of control strategy to try to increase the possibility of reaching desirable attractors or “good” phenotypes and decrease the likelihood of undesirable attractors or “bad” phenotypes. Clearly, such efforts are especially appealing in the medical community, since they hold potential to guide the effective intervention and treatment in different cancers.¹⁷⁷

A Boolean network model is a directed graph in which the nodes represent elements (e.g. genes or proteins) and edges represent interactions (e.g. transcription, phosphorylation) between these elements. Every node is assigned one of *two possible binary states in the network: ON (above threshold/expressed/high or 1) or OFF (below threshold/unexpressed/low or 0)*, corresponding to the logic values TRUE and FALSE. At each time point, the state of a node is determined by the states of its upstream neighbors (inputs) via a Boolean logical transfer function, depending on the updating schemes used in the model. For k variables, the Boolean logical transfer function can be written as $B : \{0, 1\}^k \rightarrow \{0, 1\}$. The basic logical operations include AND, OR and NOT. For example, $D = (A \text{ OR } B) \text{ AND NOT } C$ is a Boolean function with three variables. A Boolean function can also be represented by a truth table, wherein each row lists a combination of values of Boolean variables and its associated output value. The truth table of a Boolean function with k variables would have 2^k rows and $k + 1$ columns.^{119,122,179}

1.6 Computational modeling of the hematopoietic system

The validity of mathematical modeling in hematopoiesis was established early by the pioneering work of Till and McCulloch.^{188,189} Since then, deterministic, stochastic, statistical and network-based models have been used to better understand a range of topics in hematopoiesis.¹²⁰

Deterministic models In hematopoietic progenitor cells, the transcription factors GATA-1 and PU.1 are known to act as key regulators and potential antagonists in the erythroid vs. myeloid differentiation processes, where low GATA-1 and PU.1 expression maintain

1 Introduction

the cell in an undifferentiated state, whereas dominant expression of GATA-1 promotes the erythroid/megakaryocyte lineage, and PU.1 promotes the myeloid lineage. In addition, GATA-1 and PU.1 both stimulate their own transcription and inhibit that of the other, resulting in a network that generates a bistable, genetic toggle switch¹¹⁸. Roeder and Glauche¹⁹⁰ have proposed a quantitative model which, was able to account for the observed behavior: depending on the parameters used, the model predicted two different possibilities to explain the experimentally observed priming state of low level co-expression, whereas increasing transcription rates was sufficient to induce differentiation in one scenario. The same system was also studied by Huang et al.^{191,192} using a simple mathematical model supplemented by experimental measurements to analyze the dynamics of this binary fate decision governed by a gene-circuit containing both auto-stimulation and cross-inhibition. Their model yielded stable attractors corresponding to erythroid and myelomonocytic fates, as well as an uncommitted metastable state characterized by co-expression of GATA-1 and PU.1, which would explain the phenomenon of “multi-lineage priming”. Recently, a literature-derived 11-factor Boolean network modeling myeloid differentiation from common myeloid progenitors to megakaryocytes, erythrocytes, granulocytes and monocytes has been constructed¹⁹³ and validated by comparing the attractors with gene expression profiles of differentiating and mature myeloid blood cells, leading to a conclusion that the steady states of the model can be directly attributed to known biological cell phenotypes.

Stochastic models Experimental evidence exists that stochastics may be important in networks that control stem cell behavior.¹¹⁸ Intriguingly, it has been established that, in hematopoietic stem cells, after being separated by flow cytometry, high or low Sca-1 expressing populations reestablish the original distribution within several population doublings. Moreover, low Sca-1 expressing cells preferentially differentiate into the erythroid lineage, whereas high Sca-1 expressers favor the myeloid lineage. This process could be described by a Gaussian mixture model that incorporated noise-driven transitions between discrete subpopulations, suggesting hidden multi-stability within one cell type.¹⁹⁴

A single cell-based stochastic model studying the concept of within-tissue plasticity has also been described.¹⁹⁵ It demonstrates how individual cells may reversibly change their actual set of properties depending on the influence of the local growth environment. Stochastic switching between the growth environments introduced fluctuations generating heterogeneity.

Stochastic models have also been used to study the dynamics of clonal repopulation following hematopoietic stem cell transplant, where trajectories of HSC/P counts and differentiated hematopoietic cell counts were simulated and subsequently compared with experimentally observed cell counts. Rates of self-renewal, differentiation and elimination of cells were estimated. As a result, these stochastic trajectories were found to match experimental results.

1.6 Computational modeling of the hematopoietic system

Altogether, these models predicted that hematopoiesis is probabilistic in nature and that clonal dominance can occur by chance. Moreover, stochastic simulation can be used to incorporate elements of the stem cell niche, such as surrounding stromal cells and thus model cell-cell and cell-microenvironment interactions. Such models could identify regulators of stem cell fate and explore the dynamics of this regulation.¹⁶

1.7 Motivation

Recent advances in genome research and gene profiling technologies have resulted in accumulation of global gene expression patterns of primitive hematopoietic stem cells and their more differentiated progeny, such as.^{17,196,197} At the same time, efforts to examine the interactions between HSCs and their micro-environment have led to the generation of *in vitro* culture systems. We have previously established that two midgestation-derived **stromal clones—UG26-1B6, urogenital ridge-derived, and EL08-1D2, embryonic liver-derived—support the maintenance of murine adult BM** and human cord blood hematopoietic repopulating stem cells (HSCs).^{198,199}

We have already identified several UG26-1B6 and EL08-1D2-derived secreted factors, including **Secreted frizzled-related protein 1 (Sfrp1)** and **Pleiotrophin (Ptn)** and demonstrated their critical role in the maintenance of HSCs.^{51,200} Hence, it appears that such co-culture systems can, at least partially, mimic the hypothetic *in vivo* stem cell niche and as such provides a useful model system for investigating stem cell-stromal cell interactions.¹⁵ **Changes in gene expression in both HSC and niche cells over time**, after being in direct contact, to the best of our knowledge, have not been elucidated, so far. Moreover, at present, precise signaling mechanisms coordinating HSC-fate decisions have remained largely elusive.

In this study, by combining time-series (TS) gene expression analysis and phenotype data with **bioinformatics and systems biology approaches**, we aim to get deeper insights into the two-way communication between HSC and UG26-1B6 stromal cells, thus facilitating the discovery of novel molecular players and starting to unravel the complexity of biological networks controlling this “bidirectional dialogue”.

2 Materials, methods and data

2.1 Materials

Table 2.1: Instruments

| Product | Manufacturer |
|---|---|
| Agilent 2100 Bioanalyzer | Agilent Technologies, Santa Clara, CA, US |
| Animal Blood Counter Scil Vet AbcTM | Scil vet academy, Viernheim, Germany |
| Cell incubator Hera Cell 240 | Heraeus Instruments, Hanau, Germany |
| Cell sorter MoFlo High Speed | BeckmanCoulter, US |
| Centrifuge Megafuge 3.0 RS, Multifuge 3S | Heraeus Instruments, Hanau, Germany |
| Flow cytometer CyAn ADP Lx P8 | BeckmanCoulter, US |
| Fluorescent microscope Leica DM RBE | Leica, Wetzlar, Germany |
| Hematocytometer Neubauer improved | Marienfeld Superior, Germany |
| Ice maschine S.-No: 061244 | Ziegra Eismaschinen, Isernhagen, Germany |
| Laminar flow hood ANTAES 48/72 | BIOHIT, Germany |
| Linear accelerator Mevatron KD2 | Siemens, Germany |
| Microscope Axiovert 25 | Carl Zeiss, Jena, Germany |
| NanoDrop ND-1000 UV/Vis-spectrophotometer | NanoDrop Technologies, Wilmington, DE, US |
| Precision scales PLJ 2100-2M | Kern, Germany |
| QuadroMACS Separator | Miltenyi Biotec, Germany |
| Real-Time PCR System ABI PRISM 7900 | Applied Biosystems, Foster City, US |
| Real-Time PCR System StepOne | Applied Biosystems, Foster City, US |
| Spectrophotometer SmartSpec Plus | Bio-Rad, US |
| Thermomixer comfort | Eppendorf, Germany |
| Vortex IKA [®] MS1 minishaker | Werke & Co., Staufen im Breisgau, Germany |
| Water bath SUB | Grant, UK |

Table 2.2: Consumables

| Product | Manufacturer |
|---|--|
| Blood lancets Supra | megro GmbH & Co KG, Wesel, Germany |
| Disposable bags | Carl Roth, Germany |
| Disposable UV cuvettes | Brand, Germany |
| Filters 0.45/30/70/100 μm | BD Falcon™, BD Biosciences, Heidelberg, Germany |
| Filter tips TipOne 10/100/200/1000 μL | Starlab, Germany |
| MACS LS cell separation columns | Miltenyi Biotec, Bergisch Gladbach, Germany |
| GeneChip Mouse Genome 430 2.0 Arrays | Affymetrix, Santa Clara, US |
| MicroAmp® Fast 96-Well Reaction Plate with Barcode | Applied Biosystems, Foster City, US |
| Microcentrifuge safe-lock tubes, 1.5/2 mL | Eppendorf, Germany |
| Monoject blunt cannula needles | Kendall Healthcare, US |
| Needles BD Microlance™, 27/30 gauge | BD, Heidelberg, Germany |
| Poly-L-lysine-coated glass slides | Thermo Fisher Scientific Inc., US |
| Polypropylene conical and round-bottom tubes 5/12/15/50 mL | BD Falcon™, BD Biosciences Heidelberg, Germany |
| Serological Pipets, 2/5/10/25/50 mL | BD Falcon™, BD Biosciences, Heidelberg, Germany |
| S-Monovette ^ö | Sarstedt AG & Co., Nümbrecht, Germany |
| Blood Collection System | Nümbrecht, Germany |
| Syringes BD Plastipak™ 1 mL | BD, Heidelberg, Germany |
| Tissue culture bottles 250 mL/550 mL | Cellstar ^ö , US |
| Tissue culture dishes 10/20 cm | Corning Incorporated, Corning, US |
| Tissue culture plates 6/12 well | Cellstar ^ö , US |

Table 2.3: Chemicals

| Product | Manufacturer |
|---|---|
| Albumin Fraction V; $\geq 98\%$, pulv., bovine (BSA) | Carl Roth, Karlsruhe, Germany |
| Ampicillin | Sigma Aldrich, Taufkirchen, Germany |
| Carboxyfluorescein succinimidyl ester (CFSE) | Invitrogen, Darmstadt, Germany |
| Ciprofloxacin | Fresenius Kabi, Bad Homburg, Germany |
| Dimethyl sulfoxide (DMSO) | SERVA Electrophoresis GmbH, Heidelberg, Germany |
| Ethanol, 99.8% | AppliChem, Darmstadt, Germany |
| Ethidium bromide, 1% solution | Carl Roth, Karlsruhe, Germany |
| Fetal calf serum (FCS) | PAA, Cölbe, Germany |
| Formalin solution 10% | Sigma Aldrich, Taufkirchen, Germany |
| Gelatin | Sigma Aldrich, Taufkirchen, Germany |
| Glutamax | Invitrogen, Darmstadt, Germany |
| HBSS | Invitrogen, Darmstadt, Germany |
| HEPES | Gibco, Germany |
| Horse serum (HS) | BioWhittaker, Vallensbaek, Denmark |
| Isofluran Foreneó 100% | Abbott & Co., Taufkirchen, Germany |
| Isopropanol | Sigma Aldrich, Taufkirchen, Germany |
| Lipofectamine 2000 | Invitrogen, Darmstadt, Germany |
| Pen/Strep | Gibco, Germany |
| Peptone | Carl Roth, Karlsruhe, Germany |
| Polybreneó | Sigma Aldrich, Taufkirchen, Germany |
| Propidium-Jodid (PI) | Invitrogen, Darmstadt, Germany |
| Puromycin | Invitrogen, Darmstadt, Germany |
| Triton X-100 | Sigma Aldrich, Taufkirchen, Germany |
| Trypan blue | Invitrogen, Darmstadt, Germany |
| Trypsin, 10x | Gibco, Germany |
| Türk's solution | Merck, Germany |
| UltraPure DNase/RNase-Free Distilled Water | Invitrogen, Darmstadt, Germany |
| β -Mercapto-ethanol | Gibco, Germany |

Table 2.4: Home-made solutions, buffers and media

| Product | Recipe |
|---------------------------------------|---|
| FACS buffer (500 mL) | 500 mL DPBS 0.5% BSA |
| Gelatin solution (1%, 500 mL) | 5 g Gelatin powder 500 mL deionized H ₂ O |
| Gelatin solution (0.1%, 500 mL) | 50 mL 1% Gelatin solution 450 mL deionized H ₂ O |
| HF2+ buffer (1000 mL) | 100 mL HBSS 10x 20 mL FCS 10 mL HEPES 10 mL Pen/Strep 860 mL deionized H ₂ O |
| LB medium | 6 g Peptone 3 g Yeast extract 1,5 g NaCl Ampicillin (50 μ g/mL) 300 mL deionized H ₂ O |
| Phoenix Eco culture medium | DMEM 10% FCS |
| Trypsin solution (1x, 50 mL) | 5 mL 10x Trypsin 45 mL DPBS |
| UG26-1B6 cell culture medium (500 mL) | 400 mL Alpha MEM 75 mL FCS 25 mL Horse Serum (HS) 5 mL Pen/Strep 100 μ L β -Mercaptoethanol |

Table 2.5: Commercial buffers and media

| Product | Manufacturer |
|-----------------------|--------------------------------|
| ACK Lysing Buffer | Invitrogen, Darmstadt, Germany |
| Alpha MEM | Invitrogen, Darmstadt, Germany |
| DMEM | Invitrogen, Darmstadt, Germany |
| Dulbecco's PBS (DPBS) | PAA, Cölbe, Germany |
| M5300 | Stemcell Technologies, Canada |
| MethoCult M3434 | Stemcell Technologies, Canada |
| Opti-MEM | Invitrogen, Darmstadt, Germany |

Table 2.6: Kits

| Product | Manufacturer |
|---|---|
| APC BrdU Flow Kit | BD Pharmingen, San Diego, CA, US |
| GeneChip Hybridization, Wash, and Stain Kit | Affymetrix, Santa Clara, CA, US |
| HiSpeed [®] Plasmid Maxi Kit | Quiagen Inc, Hilden, Germany |
| Lineage cell depletion kit | Miltenyi Biotec, Germany |
| MessageAmp aRNA Amplification Kit | Ambion, Austin, TX, US |
| Mouse Ct _g f ELISA Kit | Usen Life Science Inc., Wuhan, China) |
| Power SYBR Green PCR Master Mix | Applied Biosystems, Foster City, CA, US |
| QuantiTect Reverse Transcription Kit | Quiagen Inc, Hilden, Germany |
| RNA fragmentation reagent | Ambion, Austin, Tx, US |
| RNeasy Micro Kit | Quiagen Inc, Hilden, Germany |

Table 2.7: Flow cytometry antibodies

| Antigen | Clone | Fluorochrome | Manufacturer |
|---|--------------|---|--------------------------------|
| Cd4 | GK1.5 | PE-Cy5 | eBioscience, San Diego, CA, US |
| Cd4 | GK1.5 | PE-Cy5 | eBioscience, San Diego, CA, US |
| Cd8a | 53-6.7 | PE-Cy5 | eBioscience, San Diego, CA, US |
| Cd11b | M1/70 | APC-eFluor ^δ 780 | eBioscience, San Diego, CA, US |
| Cd16/32 | 93 | PE | eBioscience, San Diego, CA, US |
| Cd34 | RAM34 | FITC Alexa Fluor 647 | eBioscience, San Diego, CA, US |
| Cd45.1 | A20 | PE | eBioscience, San Diego, CA, US |
| Cd45.1 | | FITC | |
| Cd45.1 | | eFluor ^δ 450 | |
| Cd45.2 | 104 | FITC | eBioscience, San Diego, CA, US |
| Cd45.2 | | PE | |
| Cd45r (B220) | RA3-6B2 | PE-Cy7 | eBioscience, San Diego, CA, US |
| Cd117 (cKit) | 2B8 | PE APC APC-eFluor ^δ 780 | eBioscience, San Diego, CA, US |
| Cd127 (Il7r α) | A7R34 | APC PE | eBioscience, San Diego, CA, US |
| Cd150 | TC15-12F12.2 | APC | BioLegend, San Diego, CA, US |
| Gr-1 | RB6-8C5 | eFluor ^δ 450 | eBioscience, San Diego, CA, US |
| Mouse Hematopoi- etic Lineage Panel | | | eBioscience, San Diego, CA, US |
| Sca-1 | D7 | PE-Cy7 | eBioscience, San Diego, CA, US |

Table 2.8: Immunofluorescence (IF) antibodies

| Product | Manufacturer | Catalog # | Dilution | Species |
|--|----------------------------|------------------|-----------------|----------------|
| anti-Cdc25A | Cell Sign. Techn., US | 3652 | 1:50 | rabbit |
| anti-Cdk2 | Cell Sign. Techn., US | 2546 | 1:50 | rabbit |
| anti-Cdk4 | Cell Sign. Techn., US | 2906 | 1:50 | mouse |
| anti-Ctgf | Santa Cruz Biotec., US | sc-25440 | 1:50 | rabbit |
| anti-Cyclin D1 | Cell Sign. Techn., US | 2978 | 1:25 | rabbit |
| anti-Cyclin E2 | Cell Sign. Techn., US | 4132 | 1:100 | rabbit |
| anti-p21Cip1 | Santa Cruz Biotec., US | sc-271532 | 1:50 | mouse |
| anti-p27Kip1 | BD Transduct. Laborat., US | 610242 | 1:100 | mouse |
| anti-p300 | Upstate/Millipore, US | 05-2576 | 1:100 | mouse |
| anti-phospho-Akt (Ser473) | Cell Sign. Techn., US | 9271 | 1:25 | rabbit |
| anti-phospho-Akt (Thr308) | Cell Sign. Techn., US | 2965 | 1:100 | rabbit |
| anti-phospho- β -catenin (Ser33/37/Thr41) | Cell Sign. Techn., US | 9561 | 1:100 | rabbit |
| anti-phospho-FAK (Tyr925) | Cell Sign. Techn., US | 3284 | 1:50 | rabbit |
| anti-phospho-FoxO1 (Ser256) | Cell Sign. Techn., US | 9461 | 1:50 | rabbit |
| anti-phospho-GSK3- β (Ser9) | Cell Sign. Techn., US | 5558 | 1:400 | rabbit |
| anti-phospho-Lrp6 (Ser1490) | Cell Sign. Techn., US | 2568 | 1:200 | rabbit |
| anti-phospho-p44/42 MAPK (Erk1/2) | Cell Sign. Techn., US | 4377 | 1:200 | rabbit |
| anti-phospho-p53 (Ser15) | Cell Sign. Techn., US | 9284 | 1:50 | rabbit |
| anti-phospho-Rb (Ser780) | Cell Sign. Techn., US | 8180 | 1:200 | rabbit |
| anti-phospho-Smad2 (Ser465/467)/ Smad3 (Ser423/425) (Thr202/Tyr204) | Cell Sign. Techn., US | 9510 | 1:200 | rabbit |
| PTEN | Cell Sign. Techn., US | 9552 | 1:100 | rabbit |
| Skp2 | Cell Sign. Techn., US | 4358 | 1:50 | rabbit |

Table 2.9: Secondary detection reagents for flow cytometry

| Reagent | Conjugate | Manufacturer |
|--------------|-----------------|--------------------------------|
| Streptavidin | Alexa Fluor 610 | Invitrogen, Darmstadt, Germany |
| Streptavidin | PE-Cy5.5 | Invitrogen, Darmstadt, Germany |
| Streptavidin | eFluor®450 | Invitrogen, Darmstadt, Germany |
| Streptavidin | APC | Invitrogen, Darmstadt, Germany |

Table 2.10: Secondary antibodies for immunofluorescence (IF)

| Product | Manufacturer | Catalog # | Dilution |
|----------------------------|---------------------|-----------|----------|
| Alexa Fluor®488 | Invitrogen, Germany | A11008 | 1:1000 |
| Goat Anti-Rabbit IgG (H+L) | | | |
| Alexa Fluor®488 | Invitrogen, Germany | A11001 | 1:1000 |
| Goat Anti-Mouse IgG (H+L) | | | |

Table 2.11: Mice

| Strain | Provider |
|-----------------------------------|---|
| C57BL/6.J (B6, B6.Cd45.2) | Harlan Laboratories Inc., The Netherlands |
| B6.SJL-Ptprca Pepcb/BoyJ (Cd45.1) | Taconic, Denmark |

Table 2.12: Cell lines

| Cell line | Description |
|-----------------------|--|
| UG26-1B6 | stromal cell line derived from murine embryonic urogenital ridge |
| NX (Phoenix) Eco 293T | cell line with vectors for retroviral packaging and envelope protein infectious for murine cells |

2.2 Methods and data

2.2.1 Experimental methods

Isolation of hematopoietic stem cells for co-cultures

Bone marrow (BM) cells were harvested from 7-10-week-old male and female (1:1) C57BL/6J mice (Harlan, Indianapolis, US) by flushing from both hind legs the tibiae and femora with ice-cold HF/2 (Hank's balanced salt solution without Ca²⁺ and Mg²⁺; Invitrogen, Germany, supplemented with 2% FCS, 10 mM HEPES, and antibiotics). The obtained cell suspension was passed through a 30 μ m nylon filter (BD FalconTM, Germany) to remove bone debris and clumps, then washed once. The number of viable BM cells was estimated using a Neubauer hemacytometer by counting the number of Trypan blue (Invitrogen, Germany) unstained cells under an optical microscope. Lineage marker positive cells were depleted using the Lineage Cell Depletion Kit (Miltenyi Biotec, Bergisch Gladbach, Germany), according to the manufacturer's recommendation. Briefly, the cells were incubated with biotinylated anti-lineage markers Cd5, Cd45r (B220), Cd11b, Gr-1 (Ly-6G/C), 7-4, and Ter-119 at ratio 10 μ L of antibody cocktail per 10⁷ cells for 10 min at 4 °C and washed once with cold HF/2 at ratio 1 mL buffer per 10⁷ cells. Streptavidin-conjugated magnetic beads were added at ratio 40 μ L beads per 10⁷ cells and incubated for 15 min at 4 °C. The lineage-depleted cell population was then collected after passing the cell suspension through a magnetic separation column and washed once with cold HF/2. The number of viable lineage negative cells was estimated by counting the number of Trypan blue (Invitrogen, Germany) unstained cells under an optical microscope. Lineage negative cells were then incubated with equal amounts (1 μ L antibody per 10⁶ cells) of Mouse Hematopoietic Lineage Biotin Panel antibodies, PECy7-anti-Sca-1 and PE-anti-c-Kit (all purchased from eBiosciences; San Diego, CA, USA) and PE-Cy5.5-streptavidin (Invitrogen, Germany) for 15 min at 4 °C in the dark, the cells were washed once and then resuspended in PBS (supplemented with 0.5% BSA) containing 0.5 μ g propidium iodide (PI) to exclude the dead cells. After antibody staining, Lin-Sca-1+c-Kit+ cells (LSKs) were selected based on surface marker expression by flow cytometry on a MoFlo cell sorter (Cytomation-Beckman Coulter) supplied with Summit 4.3 software (Beckman Coulter). Flow cytometry data were analyzed using the FlowJo 8.8.3 software (Tree Star, Inc. Ashland, OR, USA).

Stromal cell culture

The midgestation-embryo-derived stromal clone UG26-1B6 (urogenital ridge-derived) cells were cultured (as described in¹⁹⁸) on 0.1% gelatin-coated tissue culture plates (Cellstar[®], US) in stromal cell medium (80% α -minimal essential medium (α MEM), supplemented with

2 Materials, methods and data

15% fetal calf serum (FCS), 5% horse serum (HS), antibiotics penicillin and streptomycin (PenStrep; Gibco, Germany), and 10 μ M β -mercaptoethanol (Gibco, Germany)) with 30% conditioned medium. Conditioned medium (CM) was prepared as described by.¹⁹⁸ Briefly, after 1, 2 and 4 days of culture, CM was collected from confluent-grown stromal cell culture plates, centrifuged and passed through a 0.2 μ m filters to remove the dead cell debris. All stromal cultures were maintained at 33 °C, 5% CO₂ in a humid atmosphere.

For freezing, cells were trypsinized briefly and resuspended in stromal cell medium. Subsequently a centrifugation step (1400 rpm for 5 min at RT) was carried out to pellet the cells. These were then resuspended in freezing medium (89% FCS, 11% DMSO), divided into cryotube aliquots of approximately 0.5×10^6 cells and frozen. For long term storage, cells were stored in liquid nitrogen. Cells were thawed by incubating cryotubes in a water bath at 37 °C until the ice thawed. Then the cell suspension was immediately transferred into a falcon tube with 10 mL pre-warmed stromal cell medium. To exclude the cytotoxic DMSO, cells were subsequently centrifuged (1400 rpm for 5 min at RT), resuspended in fresh stromal cell medium and seeded into appropriate culture plates.

LSK and UG26-1B6 stromal cell co-culture

Before seeding with LSK cells, UG26-1B6 stromal cells were plated into 0.1% gelatin-coated 12-well cell culture plates, (2×10^5 cells per well), grown to 100% confluence and irradiated at 30 Gy using a Mevatron KD2 (Siemens, Munich, Germany). Lin-Sca-1+c-Kit+ cells (LSKs) cells (>95% pure) cells were resuspended in stromal cell culture medium and seeded with stromal cells, $\sim 10^5$ cells per well. Plates were incubated at 33 °C, 5% CO₂ in air and 95% humidity for one to three days (Day1-Day3), respectively. In addition, in order to obtain Day 0 (uncultured; Day0; d0; 0h) cells, $\sim 2 \times 10^5$ UG26-1B6 stromal cells collected by trypsinization and washed, as well as freshly sorted 10^5 Lin-Sca-1+c-Kit+ cells (LSKs) cells were pelleted by centrifugation and stored by -80 °C for subsequent RNA isolation. As an additional control, UG26-1B6 cells 24 h after changing the cell culture medium were also used (Day1 medium control; C).

After one (Day 1; d1), two (Day 2; d2) or three (Day 3; d3) days, co-cultured cells were harvested by trypsinization, washed once with HF/2+ buffer and incubated with equal amounts (1 μ L antibody per 10^6 cells) PECy7-anti-Sca-1, PE-anti-c-Kit, APC-anti-Cd150 (from BioLegend; San Diego, CA, USA), FITC-anti-Cd34, Pacific Blue-anti-Cd45 and PE-Cy5.5-streptavidin-biotin-anti-lineage-markers for 15 min at 4 °C in the dark. Thereafter, cells were washed once and suspended with PBS supplemented with 0.5% BSA containing 0.5 μ g propidium iodide (PI) to exclude dead cells. After antibody staining, Cd45+hematopoietic stem cells and Cd45-Sca1+ stromal cells were separated by a MoFlo cell sorter (Cytomation-

Beckman Coulter) supplied with Summit 4.3 software (Beckman Coulter). Flow cytometry data were analyzed using the FlowJo 8.8.3 software (Tree Star, Inc. Ashland, OR, USA).

In case of a long-term (LT)-co-culture, $\sim 1.5 \times 10^4$ LSK or $\sim 5 \times 10^3$ Lin- cells/well were cultured on stroma in long-term culture (LTC) medium (LTC) medium (Stem Cell Technologies, M5350, Vancouver, Canada) with $1 \mu\text{M}$ hydrocortisone at 33°C , 5% CO_2 in a humid atmosphere. Each week, half of the medium was removed and replaced with fresh medium and hydrocortisone. After four days, one week, two weeks and three weeks, non-adherent and adherent cells were harvested, pooled, and separated into Cd45+ hematopoietic cells and Cd45-Sca1+ stromal cells by cell sorting, as described above.

RNA extraction

Total RNA was isolated from sorted LSKs, Cd45+ hematopoietic cells (Cd45+HCs) and UG26-1B6 stromal cells using RNeasy Micro Kit (Quiagen Inc, Hilden, Germany), according to the manufacturer's recommendation. RNA concentration, purity and integrity was assessed using NanoDrop ND-1000 UV/Vis- spectrophotometer (NanoDrop Technologies, Wilmington, DE, US) or/and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), respectively.

aRNA amplification and array hybridization

Biotin-labeled aRNA was obtained using MessageAmp aRNA Amplification Kit (Ambion, Austin, TX, US), according to the manufacturer's recommendation and fragmented in RNA fragmentation reagent (Ambion, Austin, TX, US) by heating to 94°C for 35 minutes. Subsequently, biotinylated and fragmented aRNA was hybridized to the GeneChip Mouse Genome 430 2.0 Arrays (Affymetrix, Santa Clara, US) using the GeneChip Hybridization, Wash, and Stain Kit (Affymetrix, Santa Clara, US), according to the manufacturer's recommendation.

Quantitative real-time PCR (qPCR)

Gene-specific primer design Gene-specific primers were designed using the NCBI primer design tool Primer-BLAST <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>, which implements Primer3²⁰¹ and BLAST²⁰² thus ensuring that the primers made are specific to the input PCR template. Primer-BLAST was run using the default parameters, except that the PCR product size was restricted to 100-150 bp and the primers were required to span an exon-exon junction in order to eliminate genomic DNA amplification. Primer sequences, melting temperatures (T_m) and the PCR product sizes are summarized in the Table S11.

cDNA synthesis Total RNA (see *RNA extraction*) was immediately reverse transcribed into cDNA using the QuantiTect Reverse Transcription Kit (Quiagen Inc, Hilden, Germany) following the manufacturer's instructions. Briefly, purified RNA was incubated in gDNA

2 Materials, methods and data

Wipeout Buffer (2 min at 42 °C) in order to remove contaminating genomic DNA. The RNA sample was then subjected to reverse transcription (15 min at 42 °C) in a 20 μ l reaction volume, using a master mix prepared from Quantiscript Reverse Transcriptase, Quantiscript RT Buffer, and RT Primer Mix containing a mix of oligo-dT and random primers that enable cDNA synthesis from all regions of RNA transcripts. The reaction was then inactivated at 95 °C for 3 min.

RT-qPCR reaction and thermal cycling conditions The resulting cDNA was quantified by real time PCR (RT-qPCR). The reaction was performed in a 20 μ l reaction volume containing 1 \times Power SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, US), primers at 100 nM concentrations and 1 μ l of the sample cDNA (dilution factor 1:6 or 1:8 for the hematopoietic or stromal cells, respectively). RT-qPCR run was performed on the ABI PRISM 7900 or StepOne Real-Time PCR System (Applied Biosystems, Foster City, CA, US) with the following conditions: 95 °C for 10 min, 40 cycles of 95 °C for 15 sec and 60 °C for 1 min. This was followed by the default dissociation cycle for melt curve analysis in order to determine the specificity of the amplification reaction. Each experiment was done in triplicate, including also a minus template control for each gene. Cycle threshold (Ct) values were calculated using the SDS software v.2.4 or the StepOneTM Software v2.2.2.

RT-qPCR data analysis Rpl39, Gorasp2 and Rpl13a (hematopoietic cells) or Rplp0 (UG26-1B6 stromal cells) were used as endogenous control genes to convert Ct values into normalized relative quantities (NRQs), using the formula 2.1 as described by:²⁰³

$$NRQ = \frac{E_{goi}^{\Delta Ct, goi}}{\sqrt[f]{\prod_o^f E_{ref_o}^{\Delta Ct, ref_o}}} \quad (2.1)$$

where PCR efficiency (E) was assumed to be 100%.

Immunocytofluorescence (IF) microscopy

Since the amount of protein extracted from sorted Lin-Sca-1+c-Kit+ (LSK) cells would be too low to quantify protein expression and/or activation by Western blot, immunofluorescence staining was utilized to analyze intracellular protein expression and phosphorylation status. For this purpose, 10³ sorted freshly isolated or 24 h co-culture-derived LSK cells were spotted on poly-L-lysine-coated glass slides (Thermo Fisher Scientific Inc., US) and incubated on ice for 30 min. After fixation in PBS-4% PFA for 10 min and permeabilization with blocking buffer (10% FCS, 0.1% Triton-X in PBS) for 1 h at room temperature, cells were incubated with primary antibodies diluted in blocking buffer overnight at 4 °C. The cells were then washed three times with PBS and were incubated with a secondary antibody overnight

at 4°C. The primary and secondary antibodies used are listed in Tables 2.8 and 2.10, respectively. After final washes, the cover-slips were mounted in SlowFade® Gold Antifade Reagent supplemented with DAPI (4,6-diamino-2-phenylindole, dihydro-chloride) nuclear stain (Invitrogen, Germany). Fluorescence digital images were taken using constant settings on an Leica DM RBE fluorescent microscope (Leica, Wetzlar, Germany) using AxioVision software (Carl Zeiss, <http://www.zeiss.com>). For each particular sample, images at 100x magnification of at least 30 randomly captured cells were taken.

Fluorescence digital images, were then analyzed using the digital image processing software ImageJ (NIH, Bethesda, US). The obtained measurements were expressed in arbitrary units called normalized relative quantities (NRQ). For this purpose, the mean fluorescence intensity (MFI, average intensity of pixels per cell) and the cell area (number of pixels) were first determined. Fluorescence intensity values were expressed as fluorescence density (protein per pixel) after dividing MFI by the area, thereby normalizing for cell size.²⁰⁴ Background was calculated using pixels around the perimeter of the area being quantified and was removed from pixels/area measurements. In order to compare measurements from separate experiments, they were additionally normalized to the mean of a set of control samples and expressed as fold changes in relation to the control samples.

Enzyme-linked immunosorbent assay (ELISA) for quantification of conditioned media Ctgf protein levels

For this assay, UG26-1B6 stromal cells were cultured for 24 h as described above (see *Stromal cell culture*). The resulting condition media was decanted, filter-sterilized through a 0.45 µm filter, and stored at -80°C until further use. The thawed conditioned media was used for quantitative measurements of secreted Connective tissue growth factor (Ctgf) protein with a sandwich ELISA (USCN Life Science, Wuhan, China). Samples were diluted at least 100 times with PBS and assayed in duplicate, according to the manufacturer's instructions. As a background control, blank culture medium sample was used. The concentration of Ctgf in the samples was then calculated by comparing the O.D. of the samples to the standard curve.

Cell cycle analysis

S-phase analysis using bromodeoxyuridine (BrdU) uptake The BrdU assay is used to quantitate cell proliferation (i.e., the percentage of cells entering and progressing through the S (DNA synthesis) phase of the cell cycle) based on the measurement of bromodeoxyuridine (BrdU) incorporation into newly synthesized DNA. Here, the APC BrdU Flow Kit (BD Pharmingen, San Diego, CA, US) was used to profile the cell cycle status of co-culture derived hematopoietic cells following the manufacturer's instructions. Briefly, during the final 30 min of co-culture, BrdU was added to the culture medium at a final concentration

2 Materials, methods and data

of 10 μM . Immediately following this pulse of BrdU labeling, cells were harvested and Cd45+hematopoietic stem cells and Cd45-Sca1+ stromal cells were separated by a cell sorting as described above. Thereafter, cells were fixed with paraformaldehyde and permeabilized with saponin. After washing and re-fixation, to expose incorporated BrdU, treatment of cells with 30 μg of DNase for 1 hour at 37 °C was performed. Finally, immunofluorescent staining for cell-associated BrdU was done by incubating cells with fluorescent anti-BrdU overnight at 4 °C protected from light. After washing cells once and adding 20 μL of the 7-amino-actinomycin D (7-AAD) to stain for total DNA, flow cytometric analysis was used to enumerate and characterize cells in terms of their cell cycle position (i.e., G0/1, S, or G2/M phases defined by 7-AAD staining intensities).

Cell division tracking using 5- and 6-carboxyfluorescein diacetate succinimidyl ester (CFSE) labeling Lineage negative (Lin-) BM cells (see *Isolation of hematopoietic stem cells for co-cultures*) to be labeled with CFSE (Invitrogen, Darmstadt, Germany) were resuspended in PBS supplemented with 0.2% FCS. CFSE was then added at a final concentration of 1.5 μM . After incubation for 10 min at 37 °C, further dye uptake was prevented by the addition of a quarter volume of ice-cold FCS. Cells were washed once in stromal cell medium and cultured overnight at 37 °C in conditioned medium (CM). Thereafter, CFSE+LSK cells were selected by cell sorting using a narrow gate (\sim 40-channel width on a 1024-channel log amplifier), as previously described²⁰⁵ and co-cultured with UG26-1B6 stromal cells for 4 days. In each experiment, 0.1 $\mu\text{g}/\text{mL}$ of colcemid (Karyomax, Gibco BRL Life Technologies, Grand Island, NY) was added to a separate control culture containing 10^5 cells, which was used to calibrate the fluorescence intensity of cells that did not undergo any divisions during the time course of the experiment. Alternatively, for 24 h co-culture experiments, CFSE-labeled Lin- cells were co-cultured with UG26-1B6 stromal cells either with or without colcemid and the fluorescence distribution of the CFSE+LSK cells was analyzed by comparing to colcemid-treated cells.

Annexin V/Propidium Iodide (PI) apoptosis assay

Propidium iodide (PI) in conjunction with Annexin V is used to discriminate viable (PI-/Annexin V-), apoptotic (PI-/Annexin V+) and necrotic (PI+/Annexin V+) cells. Briefly, after harvesting cells by centrifugation (1400 rpm for 5 min at RT), they were resuspended in 100 μL 1x Annexin V binding buffer (10x Annexin V Binding buffer consisting of 0.1 M HEPES (pH 7.4), 1.4M NaCl and 25 mM CaCl_2 was diluted to 1x with deionized water prior to use), 5 μL of Annexin V-FITC (BD Pharmingen, US) and 5 μL of a 50 $\mu\text{g}/\text{mL}$ PI solution was added to each sample and incubated for 15 min at RT in the dark. Thereafter, 400 μL of 1x Binding Buffer was added to each tube and analyzed immediately by flow cytometry.

Colony-forming cell (CFC) assay

For LSK cells, ~ 250 fresh vs. 24 h co-culture-derived sorted PI- cells were plated in duplicate in 35 mm culture dishes in MethoCult M3434 (Stemcell Technologies, Vancouver, Canada) according to the manufacturer's instructions. Cultures were grown at 37°C in a 5% CO₂ environment, and colonies containing more than 30 cells were subjected for morphological examination and scoring after 12 days. Burst forming unit-erythrocyte (BFU-E), colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte (CFU-GEMM) and colony-forming unit-granulocyte, monocyte (CFU-GM) were scored using standard scoring criteria.

In vivo transplantation assay

The congenic Cd45.1/Cd45.2 mouse model was used for the *in vivo* transplantation assay. Briefly, 10³ fresh and 24 h co-culture-derived sorted LSK (C57BL/6-Cd45.2; for the phenotypic and functional validation of the microarray data) or the input equivalent of 2500 1 week co-cultured (with pLKO.1 and siCtgf stroma) Lin- (B6.SJL-Ptprca Pepcb/BoyJ-Cd45.1; to investigate the putative functional role of Ctgf on hematopoiesis) donor cells were injected intravenously into lethally irradiated (9.0 or 8.5 Gy) 129SvxB6.SJL-Ptprca Pepcb/BoyJ (129xCd45.1) or C57BL/6-Cd45.2 congenic mice, respectively, along with 10⁵ freshly isolated syngeneic bone marrow (BM) and 5x10⁵ spleen cells. A total of 3 independent experiments were performed, using 5 8-week-old male mice for each treatment group as recipients. The reconstitution of donor myeloid and lymphoid cells was monitored by analyzing peripheral blood (PB) samples (100-200 μL) at 5 and 10 and at 16 weeks post-transplantation, mice were sacrificed, and BM, spleen and blood cells were analyzed by flow cytometry, using the following antibodies: anti-Cd45.2-FITC, anti-Cd45.1-PE, anti-Cd4-PE-Cy5, anti-CD8a-PE-Cy5, anti-CD11b-APC-Cy7, anti-Cd45r (B220)-PE-Cy7 and anti-Ly-6G (Gr1, RB6-8C5) (all purchased from eBiosciences; San Diego, CA, USA). For stem cell analysis, Lin- cells were isolated and analyzed as described above (see *Isolation of hematopoietic stem cells for co-cultures*). Positively reconstituted mice were defined as having a minimum of 1% total donor cells (Cd45.2+ or Cd45.1+, respectively), from which a minimum of 1% had to be of the myeloid (Gr1+/med, Cd11b+) or lymphoid (B220+, Cd4/Cd8a+) lineages.

Stable knock-down cells for Ctgf

Stable knock-down of Ctgf in UG26-1B6 stromal cells was made using the lentiviral shRNAir system (Open Biosystems, Huntsville, AL, USA). The vectors containing Ctgf shRNA sequences (TRC Mouse Ctgf shRNA; Clone ID: TRCN0000109665), as well as the empty pLKO.1 control vector were packaged into lentiviral envelopes using the NX (Phoenix) Eco packaging cell line and two additional vectors a packaging vector psPax2 and an envelope

vector pMD2.G. For this purpose, 2×10^6 Phoenix Eco cells were seeded on 6 cm culture dishes 24 h prior to transfecting equal amounts ($4 \mu\text{g}$ of each vector) of either the pLK0.1-shCtgf or the empty pLKO.1 vector, as well as the additional vectors pMD2.G and psPax2 with the help of Lipofectamine 2000 (Invitrogen, Germany), following the manufacturer's instructions. Briefly, vectors and Lipofectamine 2000 were both separately diluted with OptiMEM (Invitrogen, Germany) and subsequently mixed. The mixture was incubated for 30 min at RT to allow for the complex formation. These complexes were then added to the Phoenix Eco cells. After 4-6 h of transfection, the medium was changed to stromal cell medium. Thereafter, the virus containing supernatant was harvested three times with an interval of about 12 h. Each time, the harvested supernatant was filtered ($0.45 \mu\text{m}$), polybrene (Sigma-Aldrich, Germany) was added to a final concentration of $8 \mu\text{g}/\text{mL}$ and the supernatant was added to the stromal cells to be infected (10^5 cells/6 cm cell culture dish). Finally, 12 h after the last infection, the virus supernatant was replaced with fresh stromal cell medium, followed by puromycin-selection ($5 \mu\text{g}/\text{mL}$) for three days.

2.2.2 Computational methods and data

Microarray data pre-processing and quality control

Microarray data analysis was performed using R²⁰⁶ and Bioconductor packages.²⁰⁷ Probe intensity data from Affymetrix GeneChip .CEL files were accessed using the `affy` package.²⁰⁸ Pre-processing of the microarray chips, including background correction, quantile normalization and summarization of the probe set values into expression measure was carried out using the GeneChip RMA (gcRMA)²⁰⁹ algorithm, as implemented into the `gcrma` package.²¹⁰ The log₂ scale data from gcRMA was used in statistical testing. The quality assessment of the data, both prior to and after the normalization, was performed using the Bioconductor package `arrayQualityMetrics`.²¹¹

Clustering of time-series gene expression data

Short Time-series Expression Miner (STEM) <http://www.cs.cmu.edu/~jernst/stem/>²¹² time-series gene expression data clustering algorithm was chosen to search for the most relevant time point, at which the most changes in gene expression patterns could be observed. STEM was designed for the analysis of short (3-8 time points) time series microarray gene expression data and it also implements the gene ontology (GO) enrichment analyses for sets of genes having the same temporal expression pattern, providing the means for an efficient and statistically rigorous biological interpretation of significant temporal expression patterns. Default parameters were used. As input, the complete dataset without prior non-specific filtering or any other pre-selection was used. In case of the GO enrichment analysis, processes

corrected for multiple hypothesis testing (at false discovery rate (FDR) ≤ 0.05) were selected for representation.

Two-way comparisons of consecutive time points

For the two-way analysis, gcRMA-normalized gene expression data were first divided into pairs of consecutive time points, such as freshly isolated LSK cells or UG26-1B6 stromal cells cultured separately (0 h) vs. 24 h-co-culture derived cells, as well as 24 h vs 48 h and 48 h vs 72 h-co-culture- derived hematopoietic or stromal cells, respectively. For stromal cells, an additional control 24 h (Day1; d1) after changing cell culture medium was used, in order to also take into account gene expression differences possibly resulting from the stromal cell medium components. Co-culture-derived transcripts that did not show significant positive (p-Value ≤ 0.05) associations with medium-control-derived transcripts in terms of Pearson's correlation coefficient, as well as transcripts that were part of our microarray validation set (Table 3.1, for which we explicitly tested for this association using qPCR, as summarized in Figure 3.7 and Table S4) were considered for further analysis. Thereafter, the empirical Bayes test statistics, as implemented in the `limma` package²¹³ was used to select genes whose expression differed ($-1 \geq \log_2FC \geq 1$, p-Value ≤ 0.05) across the two consecutive time points being compared.

Signal detection analysis of the microarray data

Using RT-qPCR measurements as the gold standard, a 2×2 contingency table was constructed against the microarray data, containing the counts of the 4 combinations of classification: true positives (TP, differential expression defined as $\log_2FC \geq 1.0$ or $\log_2FC \leq -1.0$ and p-Value ≤ 0.05 detectable by both RT-qPCR assay and microarray), true negatives (TN, differential expression not detectable by either RT-qPCR assay and microarray i.e., $\log_2FC < 1.0$ or $\log_2FC > -1.0$ and/or p-Value > 0.05), false positives (FP, differential expression detectable by microarrays but not by RT-qPCR), and false negatives (FN, differential expression detectable by RT-qPCR but not by MOE430.2). Based on this matrix, the following statistics were calculated for the microarray: *SENSITIVITY* (i.e., recall rate or true positive rate; TPR) (Equation 2.2) assesses the probability that the test correctly classifies a positive subject as positive. *SPECIFICITY* (Equation 2.3) assesses the probability that the test correctly classifies a negative subject as negative. *ACCURACY* (Equation 2.4) quantifies the difference between a measurement and the true value. *PRECISION* (i.e., reproducibility or repeatability)(Equation 2.5) quantifies the variability of a measurement under unchanged conditions.²¹⁴

$$SENSITIVITY = \frac{TP}{TP + FN} \quad (2.2)$$

$$SPECIFICITY = \frac{TN}{TN + FP} \quad (2.3)$$

$$ACCURACY = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.4)$$

$$PRECISION = \frac{TP}{TP + FP} \quad (2.5)$$

The trade-offs between the measures were then represented graphically as a receiver operating characteristic (ROC) curve (ROCR package)²¹⁵ which is a plot of *SENSITIVITY* on the y axis against $(1 - SPECIFICITY)$ on the x axis for varying values of the threshold t . The 45° diagonal line connecting (0,0) to (1,1) is the ROC curve corresponding to random chance. The area under the ROC curve (AUC) is a summary measure that averages detection accuracy across the spectrum of test values, where AUC of 1 represents a perfect test; an area of 0.5 represents a worthless test.²¹⁶

Functional enrichment analysis and candidate gene prioritisation using ToppGene

In order to perform functional enrichment analysis of the differentially expressed genes (DEGs) after `limma`²¹³ two-way comparison, we utilized ToppFun from the ToppGene suite <http://toppgene.cchmc.org>,²¹⁷ which detects functional enrichment of a given gene list based on 14 different features (e.g., GO, human and mouse phenotype, protein domains, pathways, pubmed abstracts, PPIs, cytoband, TF binding site, co-expression, gene family, microRNA and drug target and disease). Default parameters were used. Candidate gene prioritisation was performed using either functional annotation-based algorithm (ToppGene) or network analysis (ToppNet), both implemented also from the ToppGene Suite. Both algorithms were run using the default settings.

ToppGene: Functional annotation-based candidate gene prioritisation method first generates a representative profile of the training genes (a set of genes known to be associated with a particular phenotype) using ToppFun to first identify over-representative terms from the training genes. Next, for each test gene, a similarity score to the training profile for each of the 14 features is derived, using a fuzzy-based similarity measure to compute the similarity between any two genes based on semantic annotations, as described in.²¹⁷

ToppNet: protein–protein interaction network (PPIN)-based candidate gene prioritisation. Based on the observation that biological networks share many properties with Web and social networks, ToppNet uses modified versions of three algorithms - PageRank with Priors, HITS with Priors and K-step Markov- to prioritize candidate genes by estimating their relative importance in the PPIN to the phenotype-related genes. Here, the K-step Markov algorithm with a step size 6 (default) was used. For detailed description of the algorithm

see.²¹⁷

Literature mining using EXCERBT

The text-mining system EXCERBT (Extraction of Classified Entities and Relations from Biomedical Texts) <http://mips.helmholtz-muenchen.de/excerbt/>²¹⁸ was used for automated extraction and inference of relations between entities of interest such as protein-protein interactions or regulatory interrelations, which is done by performing an extensive search of the biomedical literature. EXCERBT is based on a sophisticated neural network-based natural language processing approach Semantic Role Labeling (SRL), which explores the syntactic constituents of a sentence and determines their semantic roles in relation to a certain predicate. This implicates ‘predicate- argument-structure’ (PAS) sets of sentences, containing the predicate (e.g. a verb) and its corresponding semantic arguments (e.g. noun phrases) with their semantic roles. For PAS generation, a modified variant of the SENNA algorithm, a deep convolutional neural network architecture has been applied.

Network visualization and topological analysis

Cytoscape <http://www.cytoscape.org/>²¹⁹ bioinformatics package was used for biological network visualization.

yEd Graph Editor <http://www.yworks.com/> desktop application (version 3.9.2) was utilized to draw network diagrams using the modified Edinburgh Pathway Notation (mEPN) scheme <http://www.mepn-pathway.org/>,²²⁰ designed to allow the logical depiction of process diagrams for a diverse range of biological pathways.

NetworkAnalyzer is a Cytoscape plug-in, was used for the topological analysis of biological networks.²²¹

GraphWeb <http://biit.cs.ut.ee/graphweb>²²² web server was used for the detection of gene modules from networks.

Dynamical network analyses using Boolean logic

R package BoolNet²²³ was used for the analysis of Boolean networks. A logical model is defined by a regulatory graph, in which the nodes represent the regulatory components and edges define the interactions among them. Logical functions can then be used to define the dynamical behavior of each component, depending on the activity level of its regulators: OR represents the combined effect of independent upstream regulators on a downstream node, whereas AND indicates the conditional dependency of upstream regulators to achieve a downstream effect. NOT represents the effect of inhibitory regulators and can be combined with activating regulations by using either OR or AND. The dynamics of the system is

represented by a state transition graph. In this case, the nodes denote states of the system (i.e., a vector giving the levels of activity of all components), and the edges denote their state transitions (i.e., a change in the value of one or several component(s), depending on the values of the relevant logical functions or parameters). In state transition graphs, terminal nodes correspond to "stable states" or attractors. Here, the original network was splitted into three sub-networks, each containing less than 29 variables, therefore steady state analysis could be implemented as exhaustive search of the state space, meaning that the both possible initial levels (ON or OFF) for each network element were considered. Steady-state attractors are the same in asynchronous and synchronous networks. Due to this, in all cases, the simplest update mode, the synchronous scheme was used. In all cases, the initial levels of Ctgf, WNTs and Tgf- β 1 were set to 1, in the unperturbed/wt system, whereas for the Ctgf loss-of-function simulations (extrinsic) Ctgf was set to 0.

Statistics

Unless otherwise indicated, data are presented as the mean \pm the standard error associated with the mean. The two-tailed Mann-Whitney t-test with a level of significance set at 0.05 was performed to compare the differences between the samples under study. All analyses were performed in the R²⁰⁶ statistical environment (v2.14.1).

3 Results

3.1 Time-series (TS) gene expression data generation

In this study, insights on the reciprocal influence of HSCs and their 'niche' cells on each other during initial stem cell activation events, and to determine possible key extracellular and intracellular molecular players governing these behavioral responses, we *performed time-series (TS) gene expression analysis*, in which a purified population of stem/progenitor cells, defined as *Lin-Sca1+c-Kit+ (LSK) were co-cultured with a confluent layer of the urogenital ridge-derived UG26-1B6 stromal cells*, as described in.¹⁹⁸ For this purpose, LSK cells were selected from the bone marrow (BM) by flow cytometry cell sorting and seeded on a feeder layer of irradiated (30 Gy) UG26-1B6 stromal cells. After one (Day 1; d1), two (Day 2; d2) or three (Day 3; d3) days in co-culture, cell sorting was used again to separate Cd45+LSK cells from Cd45-Sca1+ stromal cells. As a control (Day 0; d0; uncultured cells), freshly isolated LSKs and UG26-1B6 stromal cells prior to co-culture were used. For stromal cells, an additional control 24 h (Day 1 medium control; C) after changing the culture medium was used, in order to account also for the effects of undefined medium components (Figure 3.1; step 1 and Figure 3.2). This yielded in total four cell populations in LSK and five cell populations in stromal cells for analysis in biological triplicate, however, some of the samples had to be discarded due to insufficient RNA quality or quantity. The remaining 22 samples (9 LSK and 14 UG26-1B6) were subjected to gene expression analysis using Affymetrix Mouse430.2 microarrays. Interestingly, LSK cells seem to change their surface marker expression pattern already during the first three days in co-culture. After 24 h (Day 1; d1), the expression of the Sca-1 marker decreases, followed also by partial loss of c-Kit on Day 2 (d2), whereas on Day 3 (d3) three distinct cell populations already can be observed: Lin-Sca1+c-Kit+, Lin-Sca1-c-Kit+, as well as Lin-Sca1-c-Kit-, corresponding to LSKs, multipotent progenitors (MPPs), as well as oligopotent progenitors (OPPs) , respectively.

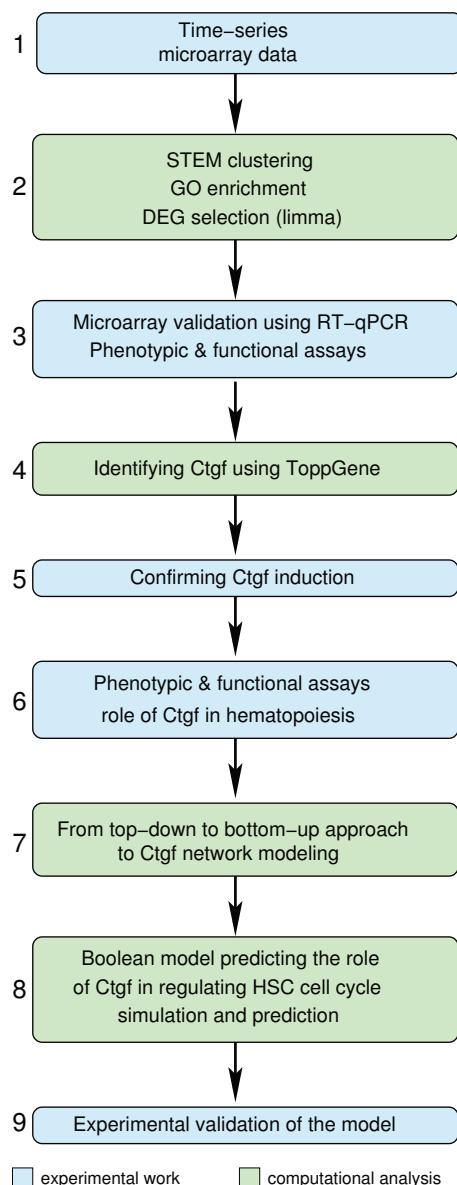


Figure 3.1: Workflow representing computational analysis (in green) vs. experimental validation of the results (in blue). Gene expression time-series data of co-cultured LSK and UG26-1B6 stromal cells were independently confirmed by RT-qPCR and using phenotypic and functional assays. STEM clustering and GO enrichment analysis followed by ToppGene candidate gene prioritization identified Connective tissue growth factor (Ctgf) as a co-culture induced gene in both LSK and stromal cells, which could be confirmed experimentally. Phenotypic and functional assays of HSCs co-cultured with UG26-1B6^{siCtgf} stromal cells demonstrated that Ctgf promotes hematopoietic progenitor activity, leading to decreased engraftment potential *in vivo*. In order to explore the molecular mechanisms, a network map was constructed linking Ctgf to the HSC cell cycle progression, as well as its auto-induction. Boolean logic was employed to simulate the behaviour of the network and repeated co-culture experiments followed by measurements of the network node mRNA and/or protein levels or phosphorylation states were used to validate the prediction results.

3.1 Time-series (TS) gene expression data generation

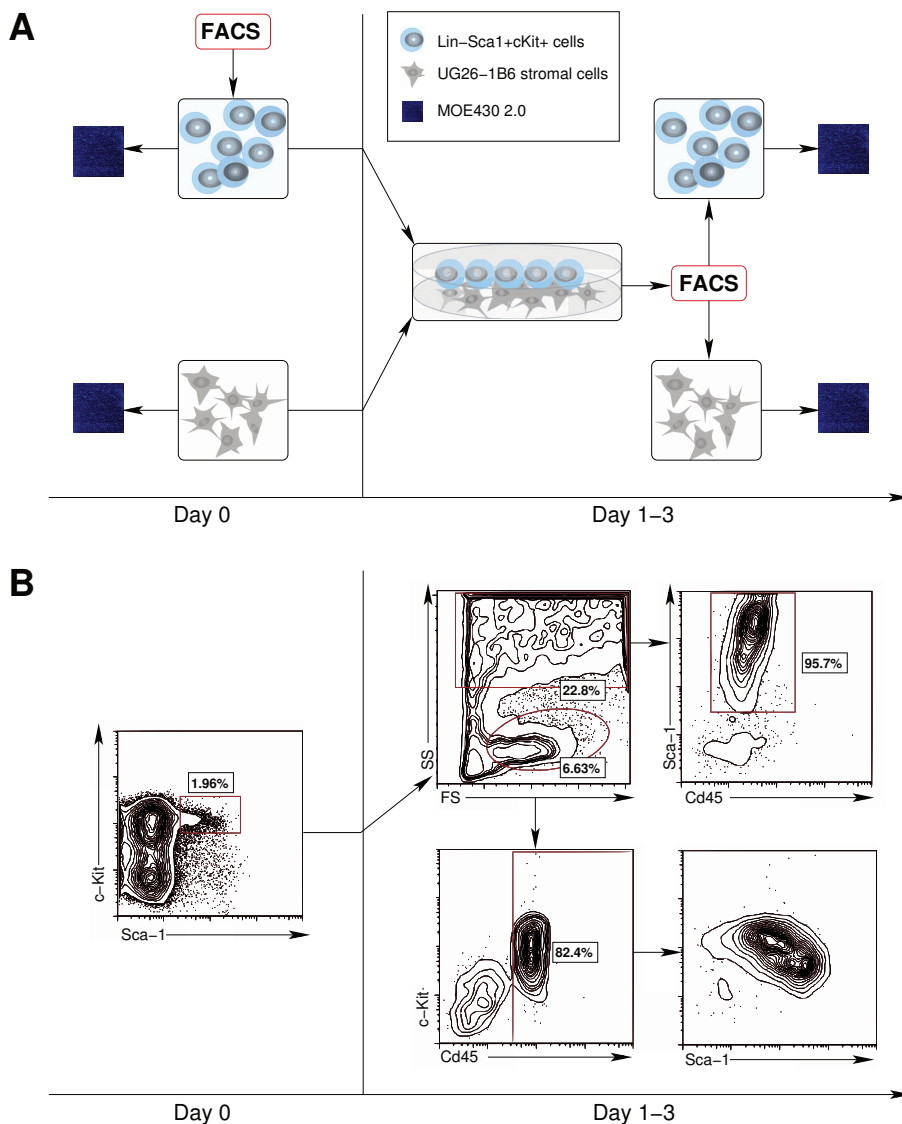


Figure 3.2: Time-series (TS) gene expression data generation from co-cultured Lin-Sca1+cKit+ (LSK, stem cells) and UG26-1B6 stromal cells. LSK cells were purified from bone marrow (BM) by flow cytometry and seeded on a previously irradiated (30 Gy) UG26-1B6 stromal cell feeder layer. Both cell populations were then co-cultured for one (Day 1; d1), two (Day 2; d2) or three (Day 3; d3) days. After this period of time, Cd45+ LSK cells and Cd45-Sca1+ stromal cells were separated by flow cytometry. As a control (Day 0; d0; uncultured cells), we used freshly isolated LSK cells or UG26-1B6 stromal cells prior to co-culture. For stromal cells, an additional control 24 h (Day 1 medium control; C) after changing the culture medium was used, see SI). (A) Schematic representation of the experimental setup showing the cell populations and time points, from which gene expression data were generated. (B) FACS gating strategy demonstrating the selection of LSK cells on Day 0(d0; uncultured cells) and the separation of co-culture into Cd45+LSK cells and Cd45-Sca1+ stromal cells on day one (Day 1; d1), day two (Day 2; d2) or day three (Day 3; d3).

3.2 Microarray data pre-processing and quality control

Microarray data analysis was first pre-processed as described in the *Materials and methods* section, including background correction, quantile normalization and summarization of the probe set values into expression measure, using the GeneChip RMA (gcRMA)²⁰⁹ algorithm. The quality assessment of the microarray data, both prior to and after the gcRMA normalization, was performed using `arrayQualityMetrics`²¹¹ and can be inferred from the Figure 3.3. As already described before, the UG26-1B6 cell samples also included an additional control 24 h (Day1; d1) after changing cell culture medium.

Individual array quality As indicated by the so called Bland–Altman (MA) plots (where M is the log intensity ratio and A is the average log intensity), in LSK cells, prior to normalization (Figure 3.3(A)), there were a couple of scatter plots, where in MA-plot the dependence between the intensity levels and the distribution of the log-ratios tended to be non-linear, however, as expected, *in most cases, the gcRMA normalization seems to have corrected for these intensity-dependent biases* (Figure 3.3(B)). In general, the same was also true for the stromal cells (Figure 3.3(C) and (D), respectively). Nevertheless, there were still a couple of arrays, which may suffer from quality problems also when considering the normalized data: #4 in LSK cells, corresponding to the 24 h (Day1; d1) co-culture derived cell sample and #7 in UG26-1B6, one of the Day2 (d2) co-culture samples. In either case, the variability of the M values (the log-ratio intensity of each array to the reference median array, i.e., the median intensity of the same probe across all arrays) seemed to be greater than that of other arrays in the data set. Other quality metrics, such as feature intensities were forming a uniform distribution, sets of features with particularly high or low intensities were not observable; boxplots and density plots of the log₂-intensities, assessing the homogeneity between arrays, applied to raw array intensities indicated rather variable median and midspread, however, in most cases, this variance decreased after the gcRMA normalization of the signal intensities.

Between array comparison Heatmap representation of the distance between arrays indicated that in LSK cells, both on raw data and on the normalized data (Figure 3.3(A) and (B), respectively), most arrays did not seem to cluster accordingly to their biological replicates. An exception to this were the samples derived from freshly isolated LSKs, which had similar distance matrix entries in the heatmap. However, there was no distinction between the consecutive time points of the co-culture. The variance mean dependency plot demonstrated that on raw data, in both LSK and stroma (Figure 3.3(A) and (C), respectively) a convex curvature on the right hand of the x-axis could be observed, indicating that higher intensities may have a higher variance, symptomatic of a saturation of the intensities. The RNA degradation plots showed that in either LSK or UG26-1B6 stromal cells, a rather strong

shift towards lower signal values for the probes closer to the 5' end was observed, meaning that *the effect of degradation was rather strong, nevertheless, the shift seemed to be systematic* and there were no outlier arrays with a slope being very different from the others. The relative log expression (RLE) plot that compares the expression levels on each array to a reference median of all arrays in the data set again revealed that the boxplot of the arrays #4 in LSK and #7 in UG26-1B6 deviating among the samples, as they had an apparently larger midspread and were not centered on zero. Finally, as expected, the mismatch (MM) oligonucleotide probes had poorer hybridization than the perfect match (PM) probes (Figure 3.3).

3.3 Computational analysis of the time-series (TS) gene expression data

In order to reduce the data set to those expressed genes which might have relevance to the activation of HSC and stromal cells in co-cultures, The Short Time-series Expression Miner (STEM) <http://www.cs.cmu.edu/~jernst/stem/> algorithm with default parameters was used (see²¹² for details), which clusters short time-series (TS; 3-8 time points) gene expression data and performs functional characterization of the clusters using Gene Ontology (GO) Term enrichment analyses for sets of genes having the same temporal expression pattern, providing the means for additional biological interpretation of significant temporal expression patterns. In each cells, two most significant clusters were selected for presentation. In LSK cells, this algorithm identified 17 significant model profiles, which were further grouped together based on similarity to form five clusters of significant profiles (Figure 3.4(A)). According to the analysis results, the largest cluster (1434 genes, 878 annotated genes) (profiles 12,1,9 and 0; C1 in Figure 3.4(A)) contained genes, whose expression was suppressed in co-culture derived LSK cells compared to freshly isolated ones. More specifically, it seems that the *most intense molecular cross-talk between LSK and stromal cells occurs already during the first 24 h of the co-culture*. Gene Ontology (GO) enrichment analysis provided by STEM of the (annotated) cluster genes, showed significant over-representation ($FDR \leq 0.05$) of genes whose products were associated with biological processes (BP) such as *chromatin and histone organization* and modification. The second largest gene cluster (see profiles 42, 48 and 49; C2 in Figure 3.4(A)) (651 genes, 504 annotated) comprised genes whose expression was induced during the co-culture. GO BP enrichment ($FDR \leq 0.05$) indicated over-representation of cell activation (4.45%), communication (18.59%), motility (5.53%) and death (9.06%), response to stress (11.83%), wounding (6.25%) and chemical stimulus (13.36%), (regulation of) cytokine production (4.3%), phagocytosis (2.15%) and signal transduction

3 Results

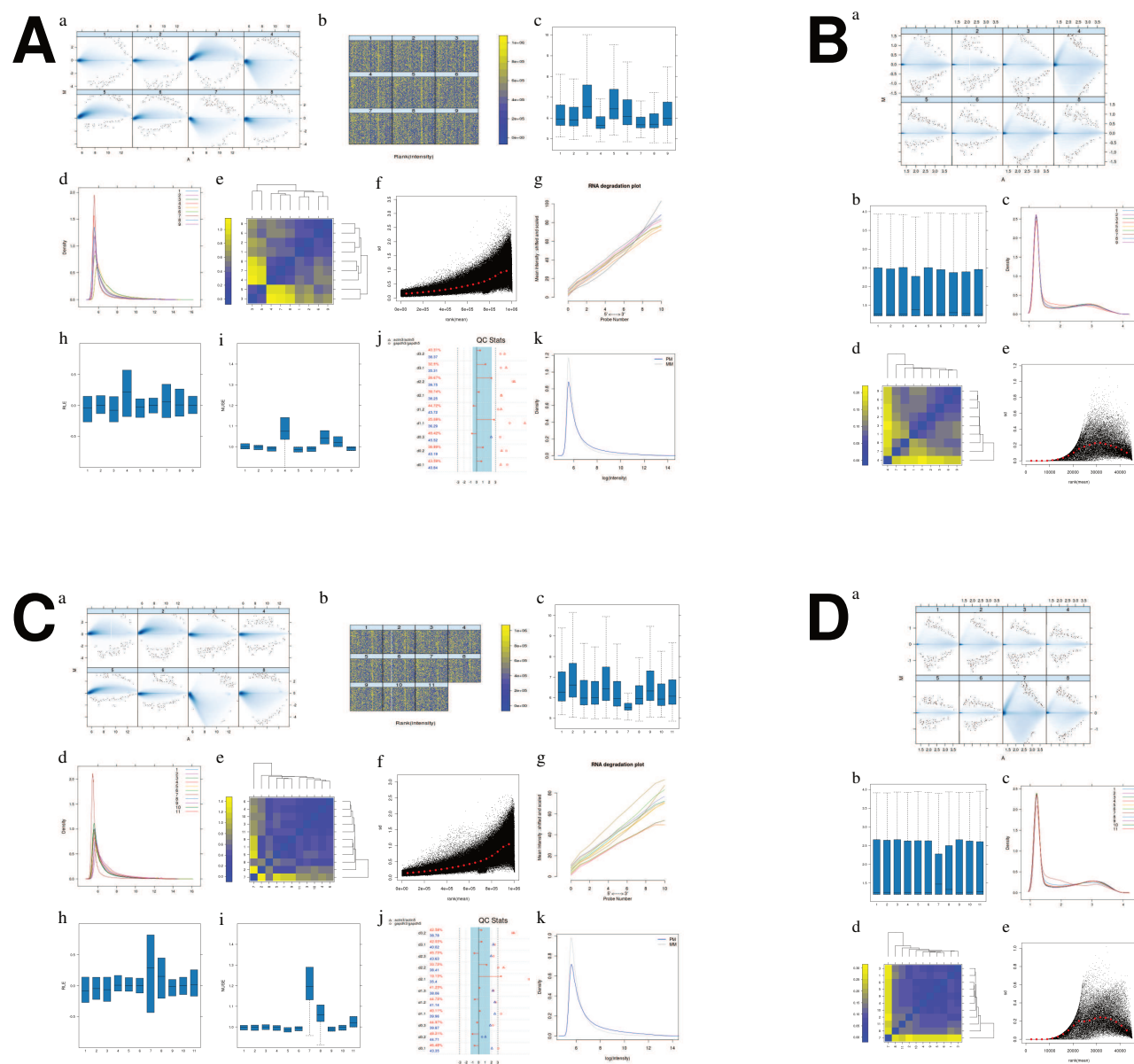


Figure 3.3: The quality assessment of time-course gene expression data of co-cultured Lin-Sca1+cKit+ (LSK) and HSC-supportive UG26-1B6 stromal cells before and after gcRMA normalization. (A) LSK and (C) UG26-1B6 cell samples prior to normalization: (a) the MA plot for each array; (b) spatial distribution of feature intensities; (c) the boxplot of the log₂-intensities; (d) density plot; (e) heatmap representation of the distance between arrays; (f) variance mean dependence; (g) RNA degradation plot; (h) relative log expression (RLE) plot; (i) normalized unscaled standard error (NUSE) plot; (j) diagnostic plot recommended by Affymetrix; (k) perfect matches and mismatches. (B) LSK and (D) UG26-1B6 cell samples after gcRMA normalization: (a) the MA plot for each array; (b) the boxplot of the log₂-intensities; (c) density plot; (d) heatmap representation of the distance between arrays; (e) variance mean dependence.

3.3 Computational analysis of the time-series (TS) gene expression data

17.35%), as well as members of various metabolic processes.

In UG26-1B6 stromal cells, when using culture medium filtered time-series (TS) data (see the Methods and materials section for details), STEM identified 12 significant model profiles, forming three clusters of significant profiles (Figure 3.4(B)). Also in stroma cells, the largest cluster (profiles 12, 0, 1 and 9; C1 in 3.4(B); 1404 genes, 1191 annotated) unified genes whose expression was down-regulated after the first 24 h of co-culture. The same GO-based functional enrichment analysis in the category biological process ($FDR \leq 0.05$) revealed over-representation of metabolism (43.59%) and gene expression (22.65%). The second largest cluster (profiles 48, 49 and 42; C2 in Figure 3.4(B); 1139 genes, 589 annotated) also resembled that of the LSK cells and contained genes whose expression was 24 h co-culture-induced and was further increasing during the consecutive two days of contact with stromal cells. Association with biological processes (FDR not controlled, $p\text{-Value} \leq 0.01$) immune response (3.78%) and **bone remodeling** (0.44%) was found. Hence, clustering of the transcripts revealed that the most prominent changes in gene expression levels occurred already during the **first 24 h of co-culture (Day 1; d1)** (Figure 3.4). **We focused our downstream analyses on this time point** (Figure 3.1; step 2). For two-way comparisons, we utilized the Linear Models for Microarray Data (Limma, as implemented in the R/limma package)²¹³ t-statistic approach, which fits a linear model to the expression data for each gene, using Empirical Bayes to borrow information across genes making the analyses stable even for experiments with small number of arrays. By this, we selected the differentially expressed genes (DEGs) from the gcRMA-normalized (and medium-control pre-filtered in case of UG26-1B6, as described above) gene expression data. Genes were defined as differentially expressed if they had a $-1 \geq \log_2FC \geq 1$ and a $p \leq 0.05$ across the two consecutive time points being compared. This analysis yielded 176 up- and 455 down-regulated transcripts in LSK cells, and in UG26-1B6 stromal cells 930 up- and 1907 down-regulated transcripts were found (see Tables S1 and S2, respectively). In order to interpret a set of differentially expressed genes obtained after the two-way comparison of consecutive time points, again, **functional enrichment analysis** was performed using ToppFun from the ToppGene²²⁴ suite <http://toppgene.cchmc.org>. A complete list of the results is given in the Table S3. In **LSK cells**, among the up-regulated genes, among the significantly over-represented Gene Ontology (GO) Biological Process categories we observed GO:0043067-regulation of **programmed cell death** (Bonferroni adjusted $p\text{-Value} = 1.202E-9$, 85 genes) and a Mouse Phenotype MP:0001819-abnormal immune cell physiology ($p\text{-Value} = 1.313E-2$, 64 up-regulated genes). The down-regulated signature was most significantly enriched for the GO: Biological Processes such as GO:0000278-**mitotic cell cycle** ($p\text{-Value} = 9.548E-7$, 60 Day1 co-culture down-regulated genes) and GO:0016568-**chromatin modification** ($p\text{-Value} = 1.414E-10$, 47 genes). Interestingly, in

UG26-1B6 cells, although filtered for the possible metabolic effects arising from the cell culture medium change (as described above and in *2.2.2.3 Two-way comparisons of consecutive time points* in the *Materials and methods* section, page 40), we observed significant enrichment of the GO: Cellular Components GO:0044429-mitochondrial part (p-Value=1.172E-2, 57 genes) and GO:0005759-**mitochondrial matrix** (p-Value =1.256E-2, 29 genes), as well as the GO: Biological Processes GO:0006396-RNA processing (Bonferroni adjusted p-Value = 1.054E-2, 65 genes) and GO:0042254-ribosome biogenesis (p-Value = 3.617E-2, 29 genes). As mitochondria constitute the most prominent source of ATP and are known to be implicated in various anabolic and catabolic activities, associated with cellular response to metabolic stress,²²⁵ this analysis may suggest that the **filtering of transcripts** based on their expression profile after adding fresh culture medium for one day (Day1 culture medium control; C) **may not have given the desired result**, and the observed changes in gene expression in UG26-1B6 cells may still be LSK-independent. Among the down-regulated genes, the same functional enrichment analysis yielded a significant over-representation of GO: Biological Process GO:0000278-mitotic cell cycle (p-Value=2.058E-31, 229 genes) and Mouse Phenotype MP:0001672-abnormal embryogenesis/development (p-Value =1.449E-11, 192 genes).

3.4 Independent confirmation of the time-series gene expression data

Commonly, independent confirmation of microarray data is needed using an independent gene expression profiling method, usually **RT-qPCR, which is then considered as the "gold standard" for gene expression measurements to estimate the performance of the microarrays**, due to its detection sensitivity, sequence specificity, large dynamic range, as well as high precision and reproducible quantitation.²²⁶⁻²²⁸ Here, to validate the *limma* results, we performed RT-qPCR on mRNAs generated from independent co-culture experiments (Figure 3.1; step 3). Since in both LSK and UG26-1B6 stromal cells most changes in gene expression seemed to occur already during the 24 h (Day1; d1) of co-culture (Figure 3.4), differentially expressed genes (DEGs) from these two time points (freshly isolated LSK or uncultured UG26-1B6 stromal cells (Day0; d0) vs. 24 h co-culture-derived (Day1; d1) cells) were selected for microarray data validation by quantitative RT-PCR (qPCR). In order for the validation conducted to be as comprehensive and unbiased as possible, the target genes for RT-qPCR validation were selected based on the following strategies: (1) ensure a large enough number of validation targets with diverse biological functionality to provide a potentially representative overview of the microarray performance;

3.4 Independent confirmation of the time-series gene expression data

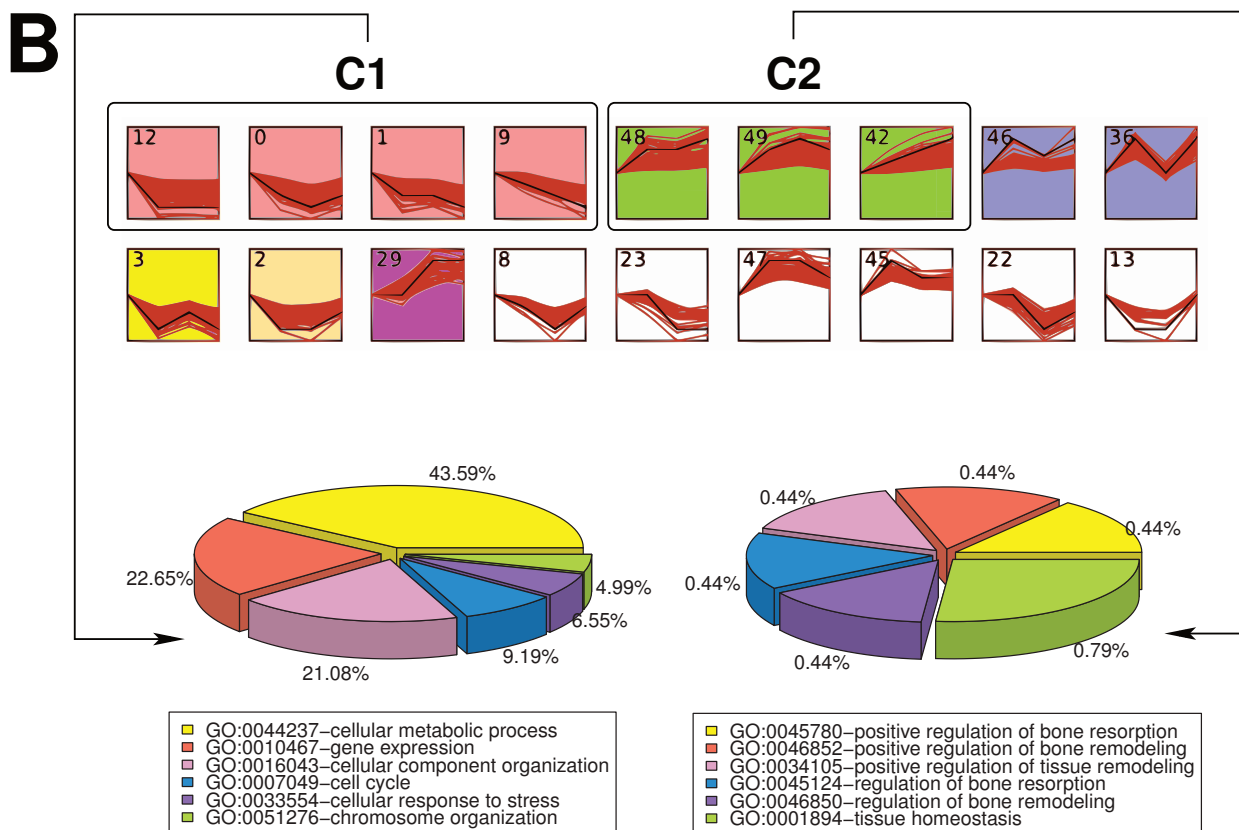
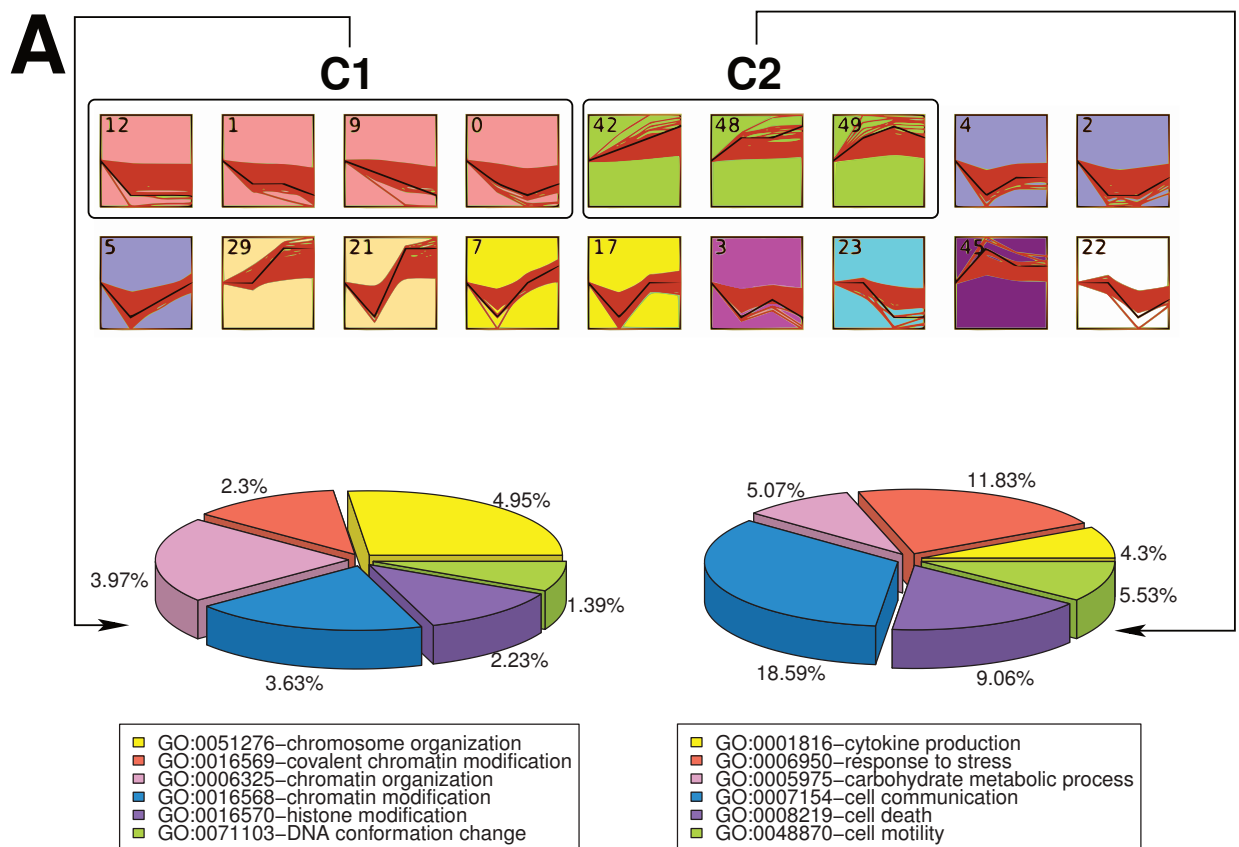


Figure 3.4

Figure 3.4: STEM²¹² clustering and Gene Ontology (GO) Term enrichment (biological process; BP) analysis of the two most significant clusters. Clusters are ordered based on number of genes and profiles are ordered by their p-Value significance, considering the number of genes assigned to clusters versus the number of genes expected to be assigned (default parameters). (A) In LSK cells, STEM identified five clusters which included 17 significant model profiles. The largest cluster (C1; 1434 transcripts, 878 annotated) contained genes, whose expression was suppressed after 24 h (Day1; d1) in co-culture. Functional classification using Gene Ontology (GO) enrichment analysis revealed significant over-representation (FDR ≤ 0.05) of biological processes (BP) related to chromatin and histone organization and modification. The second largest cluster (C2; 651 transcripts, 504 annotated) yielded an immediate-induction response of gene expression changes and significantly enriched GO BP categories included cytokine production and cell activation, response to stress, as well as various metabolic activities. (B) In UG26-1B6 stromal cells, three gene clusters including 12 significant model profiles were identified by STEM. Similarly to LSKs, the largest cluster (C1; 1404 transcripts, 1191 annotated) unified genes down-regulated in response to co-culture. Over-representation (FDR ≤ 0.05) of metabolism, gene expression, cell cycle and chromosome organization could also be observed. The second largest cluster (C2; 1139 genes, 589 annotated) contained 24 h (Day; d1) co-culture-induced transcripts mainly associated with bone homeostasis.

(2) select genes with expression levels and statistical significance values spanning a wide dynamic range. As a result, a total of **75 genes** (46 in LSK cells and 29 in stromal cells, respectively; Table 3.1) with mRNA expression levels ranging from log2 fold change (FC) 7.24 (or FC 151.17) to -3.97 (or FC 0.064) and statistical significance (p-Value < 0.001 to 1.0) were selected for real-time PCR (RT-qPCR) validation. Thus, several up-regulated ($\log_2\text{FC} \geq 1.0$) and down-regulated ($\log_2\text{FC} \leq -1.0$) genes at different levels of significance were tested. Again, for UG26-1B6 stromal cells, an additional control 24 h (Day1; d1) after changing cell culture medium was used, in order to also take into account gene expression differences possibly resulting from the stromal cell medium components. To do so, *for each co-culture-derived transcript (Day1; d1; co-culture) we calculated the Pearson's correlation coefficient and its significance (p-Value ≤ 0.05) with the corresponding medium-control-derived transcript* (medium control; C, Table S4).

The performance of the microarrays was evaluated focusing on the following criteria: (1) fold change ($\log_2\text{FC}$) and p-Value correlation with real-time qPCR data in gene expression profiles determined; (2) sensitivity, specificity, accuracy and precision in (see Equations 2.2, 2.3, 2.4 and 2.5 for definitions) detection of differential expression. The results of this analysis are summarized in Figure 3.5, showing the $\log_2\text{FC}$ comparison of the microarray data (MOE430.2; black bars) to the real-time qPCR (RT-qPCR; light gray bars) measurements for

3.4 Independent confirmation of the time-series gene expression data

Table 3.1: Microarray validation results using RT-qPCR for selected target genes. Positives (i.e., genes that are differentially expressed) and negatives (i.e., genes below differential expression threshold) were determined using the microarray results, according to the differential expression thresholds defined as $\log_2FC \geq 1.0$ or $\log_2FC \leq -1.0$ and $p\text{-Value} \leq 0.05$ and directly compared to their corresponding \log_2FC and $p\text{-Values}$ obtained using RT-qPCR ("ground truth"). Concordance with microarrays was determined by counting the number of true positives (TP, differential expression detectable by both RT-qPCR and microarray), true negatives (TN, differential expression not detectable by either method), false positives (FP, differential expression detectable by microarrays but not by RT-qPCR) and false negatives (FN, differential expression detectable by RT-qPCR but not by MOE430.2).

| | Cells | Gene | MOE430.2 | | RT-qPCR | | TP | FP | TN | FN |
|----|-------|---------|------------|---------|------------|---------|----|----|----|----|
| | | | \log_2FC | p-Value | \log_2FC | p-Value | | | | |
| 1 | LSK | Ctgf | 5.58 | 0.00 | 3.60 | 0.00 | * | | | |
| 2 | LSK | Fos | 5.33 | 0.01 | 2.55 | 0.00 | * | | | |
| 3 | LSK | Lgals3 | 5.03 | 0.00 | 2.46 | 0.11 | | * | | |
| 4 | LSK | Itgb3 | 4.13 | 0.00 | 2.44 | 0.00 | * | | | |
| 5 | LSK | Cebpb | 3.01 | 0.00 | 2.37 | 0.02 | * | | | |
| 6 | LSK | Pak1 | 2.48 | 0.08 | -1.15 | 0.00 | | | | * |
| 7 | LSK | Itgav | 2.29 | 0.04 | -0.11 | 0.71 | | * | | |
| 8 | LSK | Cdkn1a | 2.22 | 0.05 | 2.83 | 0.00 | * | | | |
| 9 | LSK | Foxo1 | 2.20 | 0.03 | 0.71 | 0.06 | * | | | |
| 10 | LSK | Ddit3 | 1.98 | 0.01 | 0.93 | 0.13 | | * | | |
| 11 | LSK | Atf4 | 1.48 | 0.08 | 1.49 | 0.00 | | | | * |
| 12 | LSK | Smad4 | 1.03 | 0.06 | 0.64 | 0.24 | | | * | |
| 13 | LSK | Stat6 | 1.03 | 0.01 | 0.57 | 0.17 | | * | | |
| 14 | LSK | Cxcr4 | 0.65 | 0.46 | -1.06 | 0.02 | | | | * |
| 15 | LSK | Kdm5d | 0.18 | 0.67 | -0.25 | 0.20 | | | * | |
| 16 | LSK | Ccnd1 | 0.17 | 0.26 | 1.31 | 0.04 | | | | * |
| 17 | LSK | Axin2 | 0.11 | 0.38 | -0.81 | 0.23 | | | * | |
| 18 | LSK | Tgfbr1 | 0.02 | 0.62 | -0.20 | 0.08 | | | * | |
| 19 | LSK | Lef1 | 0.00 | 1.00 | 1.59 | 0.05 | | | | * |
| 20 | LSK | Tgfbr2 | -0.01 | 0.72 | 1.07 | 0.04 | | | | * |
| 21 | LSK | Cdkn1b | -0.07 | 0.94 | 0.16 | 0.71 | | | * | |
| 22 | LSK | Hdac2 | -0.14 | 0.74 | -0.49 | 0.11 | | | * | |
| 23 | LSK | Kdm6b | -0.14 | 0.19 | 1.93 | 0.06 | | | | * |
| 24 | LSK | Ep300 | -0.30 | 0.46 | 0.42 | 0.14 | | | * | |
| 25 | LSK | Smarca4 | -0.54 | 0.38 | -1.06 | 0.07 | | | * | |
| 26 | LSK | Meis1 | -0.58 | 0.51 | -0.20 | 0.49 | | | * | |
| 27 | LSK | Hdac1 | -0.81 | 0.13 | 0.06 | 0.79 | | | * | |
| 28 | LSK | Kdm6a | -1.03 | 0.22 | -0.15 | 0.87 | | | * | |
| 29 | LSK | Hoxa9 | -1.05 | 0.10 | -0.32 | 0.10 | | | * | |
| 30 | LSK | Pbx1 | -1.07 | 0.41 | 0.37 | 0.47 | | | * | |
| 31 | LSK | Ccne1 | -1.29 | 0.00 | 1.30 | 0.00 | | * | | |
| 32 | LSK | Ezh2 | -1.29 | 0.17 | -1.45 | 0.00 | | | | * |
| 33 | LSK | Fzd7 | -1.30 | 0.08 | 0.37 | 0.69 | | | * | |

Continued on next page

Table 3.1: – continued from previous page

| | Cells | Gene | MOE430.2 | | RT-qPCR | | TP | FP | TN | FN |
|----|----------|--------|----------|---------|---------|---------|----|----|----|----|
| | | | log2FC | p-Value | log2FC | p-Value | | | | |
| 34 | LSK | Stat1 | -1.67 | 0.06 | -1.66 | 0.00 | | | | * |
| 35 | LSK | Dnmt3a | -1.89 | 0.03 | 0.03 | 0.53 | | * | | |
| 36 | LSK | Suz12 | -1.96 | 0.13 | 0.34 | 0.33 | | | * | |
| 37 | LSK | Mll1 | -2.03 | 0.22 | -0.37 | 0.43 | | | * | |
| 38 | LSK | Eed | -2.03 | 0.23 | -0.23 | 0.54 | | | * | |
| 39 | LSK | Cdc25a | -2.04 | 0.02 | 1.35 | 0.01 | | * | | |
| 40 | LSK | E2f1 | -2.57 | 0.00 | -0.74 | 0.10 | | * | | |
| 41 | LSK | Pcbd1 | -2.62 | 0.03 | 2.14 | 0.13 | | * | | |
| 42 | LSK | Rad54l | -3.15 | 0.04 | -1.59 | 0.00 | * | | | |
| 43 | LSK | Rad51 | -3.25 | 0.07 | -1.76 | 0.03 | | | | * |
| 44 | LSK | Ccne2 | -3.30 | 0.03 | 0.14 | 0.59 | | * | | |
| 45 | LSK | Pbrm1 | -3.52 | 0.06 | -0.08 | 0.46 | | | * | |
| 46 | LSK | Cdk2 | -3.97 | 0.11 | -0.94 | 0.04 | | | | * |
| 47 | UG26-1B6 | Sphk1 | 7.24 | 0.00 | 6.42 | 0.01 | | * | | |
| 48 | UG26-1B6 | Ctsg | 5.13 | 0.00 | 2.27 | 0.01 | * | | | |
| 49 | UG26-1B6 | Ctgf | 3.53 | 0.00 | 5.95 | 0.007 | * | | | |
| 50 | UG26-1B6 | Mbd1 | 3.26 | 0.00 | 1.98 | 0.006 | * | | | |
| 51 | UG26-1B6 | Prtn3 | 3.14 | 0.00 | -0.36 | 0.68 | | * | | |
| 52 | UG26-1B6 | Wnt2 | 2.77 | 0.00 | 1.92 | 0.12 | | * | | |
| 53 | UG26-1B6 | Med1 | 2.44 | 0.00 | 0.18 | 0.41 | | * | | |
| 54 | UG26-1B6 | Plaur | 2.33 | 0.00 | 3.41 | 0.03 | * | | | |
| 55 | UG26-1B6 | Slit3 | 2.28 | 0.01 | 1.82 | 0.01 | * | | | |
| 56 | UG26-1B6 | Esrra | 2.05 | 0.03 | 1.79 | 0.14 | | * | | |
| 57 | UG26-1B6 | Mmp15 | 1.70 | 0.20 | 2.03 | 0.11 | | | * | |
| 58 | UG26-1B6 | Hmga2 | 1.69 | 0.06 | 4.63 | 0.00 | | | | * |
| 59 | UG26-1B6 | Ltbp2 | 1.50 | 0.01 | 1.94 | 0.02 | * | | | |
| 60 | UG26-1B6 | Nfkbia | 1.30 | 0.00 | 2.54 | 0.02 | * | | | |
| 61 | UG26-1B6 | Tgfb2 | 0.05 | 0.31 | 0.83 | 0.08 | | | * | |
| 62 | UG26-1B6 | Tgfb1 | 0.00 | 0.77 | 0.89 | 0.01 | | | | * |
| 63 | UG26-1B6 | Tgfb3 | -1.51 | 0.04 | 0.21 | 0.66 | | * | | |
| 64 | UG26-1B6 | Nfib | -1.84 | 0.02 | 0.32 | 0.74 | | * | | |
| 65 | UG26-1B6 | Cxcl12 | -1.85 | 0.01 | 0.89 | 0.00 | | * | | |
| 66 | UG26-1B6 | Thbs1 | -1.93 | 0.01 | 0.25 | 0.21 | | * | | |
| 67 | UG26-1B6 | Brca1 | -2.01 | 0.03 | -0.83 | 0.02 | * | | | |
| 68 | UG26-1B6 | Jag1 | -2.11 | 0.00 | 1.18 | 0.25 | | * | | |
| 69 | UG26-1B6 | Vcam1 | -2.17 | 0.01 | 0.65 | 0.23 | | * | | |
| 70 | UG26-1B6 | Cited2 | -2.48 | 0.03 | -0.73 | 0.14 | | * | | |
| 71 | UG26-1B6 | Fbn1 | -2.55 | 0.02 | -0.98 | 0.01 | * | | | |
| 72 | UG26-1B6 | Adam10 | -2.74 | 0.00 | 0.53 | 0.23 | | * | | |
| 73 | UG26-1B6 | Nrp1 | -2.91 | 0.03 | -0.84 | 0.01 | * | | | |
| 74 | UG26-1B6 | Npr3 | -2.93 | 0.01 | -1.03 | 0.056 | * | | | |
| 75 | UG26-1B6 | Igf1 | -3.79 | 0.02 | -1.08 | 0.03 | * | | | |

3.4 Independent confirmation of the time-series gene expression data

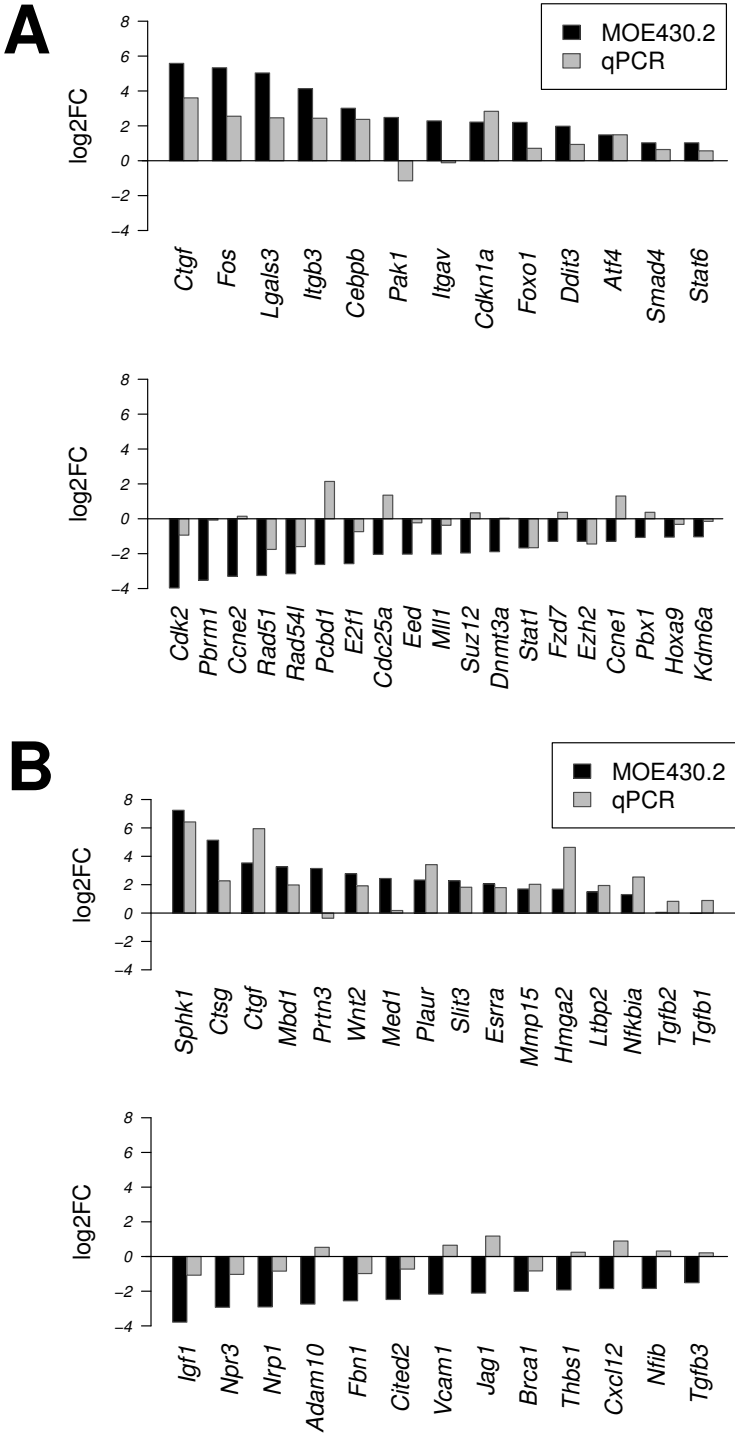


Figure 3.5: Microarray validation results using RT-qPCR for selected target genes, performed on mRNAs from independent co-culture experiments: log₂FC comparison of microarray data MOE430.2 vs. RT-qPCR measurements. Results are sorted in descending order based on the log₂FC reported by the microarray, without taking into account the statistical significance (p-Value) of the measurements. (A) LSK cells and (B) UG26-1B6 stromal cells: (a) up-regulated genes; (b) down-regulated genes.

3 Results

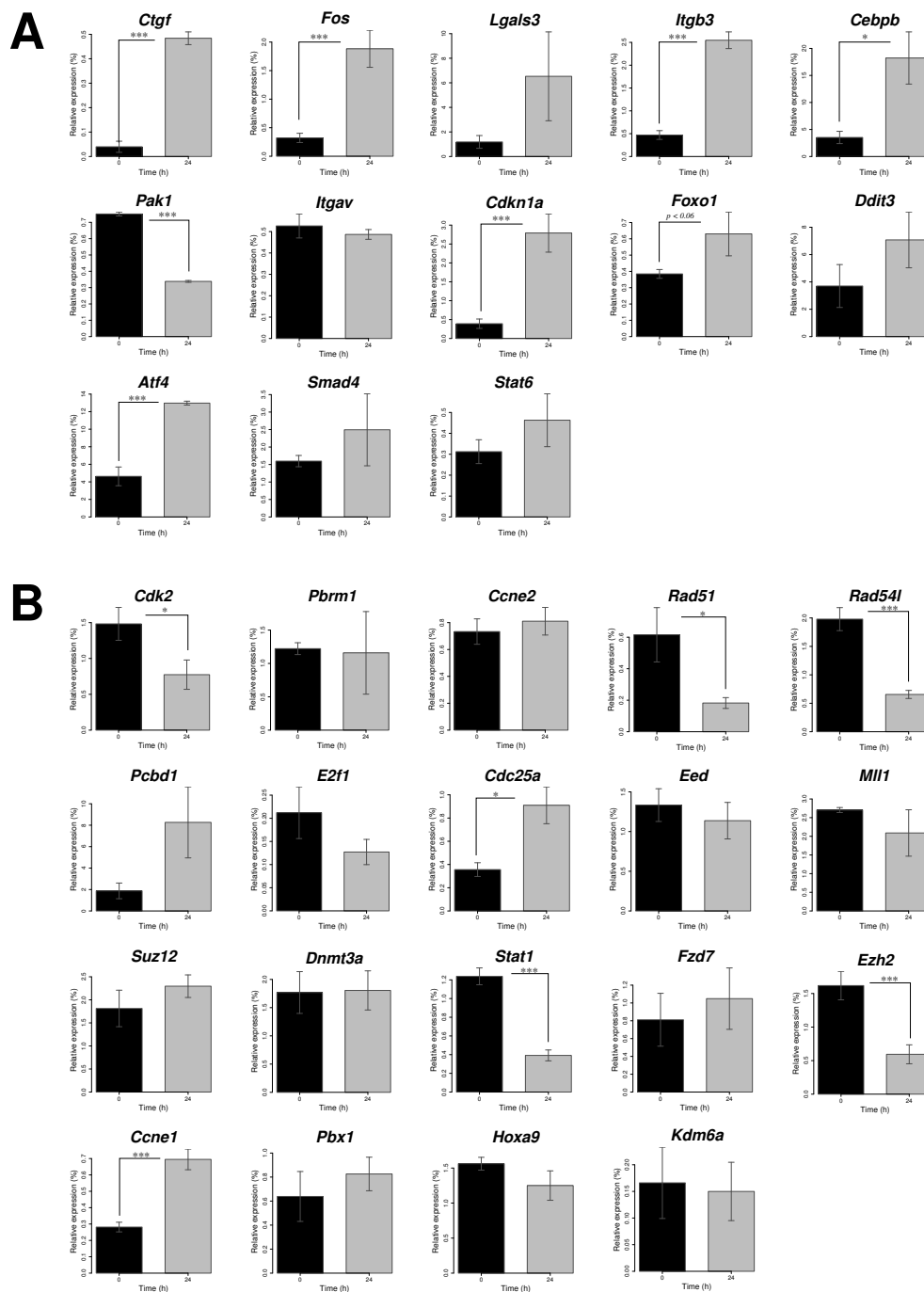


Figure 3.6: Microarray validation results using RT-qPCR for selected (A) up- and (B) down-regulated genes in LSK cells. Results for a single gene, represented as its relative expression (%) of the housekeeping gene (HKG), comparing freshly isolated LSK (0 h; Day0; d0, black bars) to the 24 h co-culture-derived cells (24 h; Day1; d1, light gray bars). *Rpl39*, *Gorasp2* and *Rpl13a* were used as HKGs. Primer sequences used are given in Table S11. Results are shown as mean and standard error of at least three independent samples. The one-tailed Mann-Whitney t-test was used in order to test for the statistical significance in the one direction of interest, reported by microarrays.

3.4 Independent confirmation of the time-series gene expression data

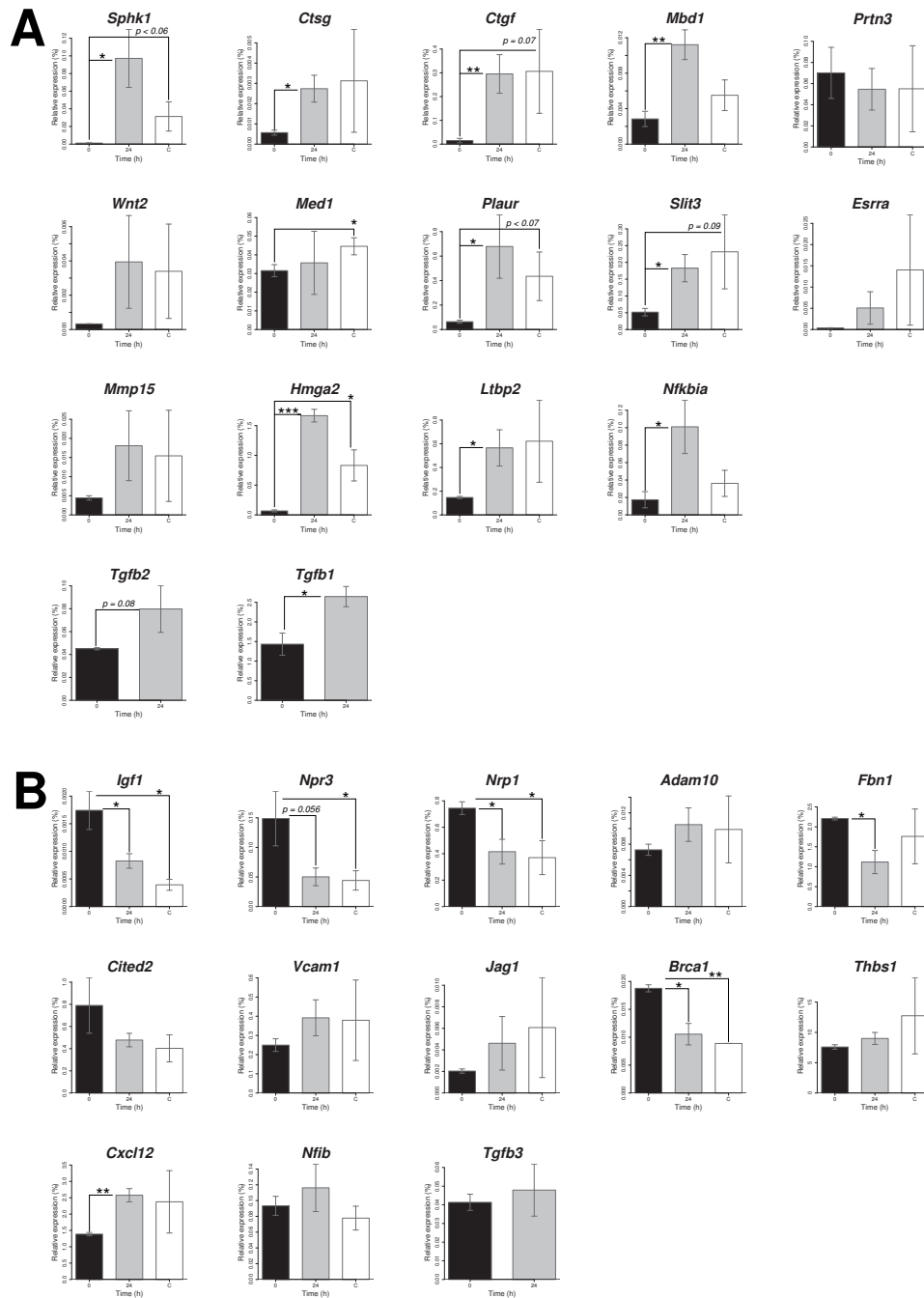


Figure 3.7: Microarray validation results using RT-qPCR for selected (A) up- and (B) down-regulated genes in UG26-1B6 stromal cells. Results for a single gene, represented as its relative expression (%) of the housekeeping gene (HKG), comparing separately cultured UG26-1B6 stromal cells (0 h; Day0; d0, black bars) to the 24 h co-culture-derived cells (24 h; Day1; d1, light gray bars), as well as to the medium control (control; C, white bars) 24 h after changing cell culture medium. *Rpl39*, *Gorasp2* and *Rplp0* were used as HKGs. Primer sequences used are given in Table S11. Results are shown as mean and standard error of at least three independent samples. The one-tailed Mann-Whitney t-test was used in order to test for the statistical significance in the one direction of interest, reported by microarrays. For the medium control (control; C, white bars), we also tested the statistical significance in the same direction of interest, in order to compare those with the Pearson's correlation coefficient analysis, as described in *Methods and Data*.

Table 3.2: The contingency table containing the counts of the 4 combinations of classification. Using RT-qPCR measurements as the "gold standard", a 2×2 contingency table was constructed against the microarray data, containing the counts of the 4 combinations of classification: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN).

| | |
|---|--|
| True positives (TP) | False Positives (FP) (Type I error) |
| 21 | 21 |
| False Negatives (FN) (Type II error) | True negatives (TN) |
| 13 | 20 |

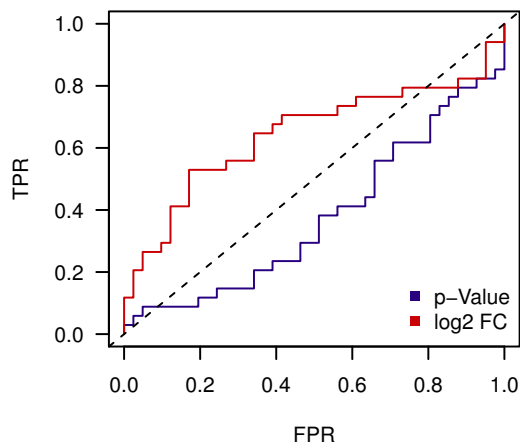


Figure 3.8: Receiver operating characteristic (ROC) curve of the microarray performance in terms of the fold change (\log_2 FC) and moderate t-test statistics (p-Value). The (ROC) curve shows *SENSITIVITY* on the y axis against $(1 - \textit{SPECIFICITY})$ on the x axis for varying values of the \log_2 FC and p-Values indicating the significance of the differential expression measures by the microarray. The 45° diagonal line connecting (0,0) to (1,1) is the ROC curve corresponding to random chance.

3.4 Independent confirmation of the time-series gene expression data

the selected target genes in (A) LSK and (B) UG26-1B6 stromal cells, respectively, as well as Figures 3.6 and 3.7 showing the results for a single gene, represented as normalized relative quantity (NRQ) (see *Methods and Data* for details) $\times 100\%$, comparing freshly isolated LSK or uncultured UG26-1B6 stromal cells (0 h; Day0; d0 black bars) to the same cells after 24 h co-culture (24 h; Day1; d1; light gray bars) or, in UG26-1B6 cells, also to the medium control (control; C, white bars), for which we also tested the statistical significance in the same direction of interest, in order to compare those with the Pearson's correlation coefficient analysis (Table S4).

In LSK cells, a total of 46 genes were selected for validation. From these, according to Affymetrix chip profiling results, 13 had a positive log₂ fold change ($FC \geq 1.0$) and 10 of them also a p-Value ≤ 0.05 (Table 3.1). In 11 out of 13 cases (85%), real-time qPCR analysis (Figure 3.6) confirmed the direction of expression, however, only in 7 cases (54%) the result was statistically significant (p-Value ≤ 0.05) or just below the significance threshold (p-Value ≤ 0.06). In addition to this, for the majority of the genes examined, there were significant quantitative differences between microarray- and qPCR-based data (Figure 3.5(A)). In most cases, the real time PCR found a much lower fold change than the microarrays, and only for one gene (*Cdkn1a*) the opposite was true (log₂FC 2.83 vs. 2.22). Interestingly, this is in contrast to a previous observation that Affymetrix microarrays underestimated the relative changes in mRNA expression between experimental and control samples.²²⁹ In the down-regulated ($FC \leq -1$) group, the measurements were less consistent. Only in less than half of the cases (47%), real-time qPCR analysis (Figure 3.6) confirmed the direction of expression, in 5 (26%) cases the result was statistically significant. Indeed, a similar trend of higher correlations among up-regulated genes than among down-regulated ones has already previously been reported,^{228,230} and it was proposed that this effect may be due to the increased variability observed in low-intensity array spots, i.e. down-regulated genes.²³⁰ In addition, microarray results again over-estimated the fold-change detected by qPCR. Actually, the quantitative differences between microarray- and RT-qPCR-based data were more profound (Figure 3.5(A)). At the same time, in 6 of 14 (43%) negatives (i.e., genes below differential expression threshold) compared, DNA microarrays failed to detect the expression changes that were revealed by RT-qPCR, which can be attributed to the higher detection sensitivity and large dynamic range of qPCR, as compared to the microarrays.²²⁷

In UG26-1B6 stromal cells, analyses of the microarray showed that, of the 29 target genes selected, 14 genes were induced, whereas 13 genes were suppressed at least two-fold during the co-culture with LSK cells. In most induced genes (in 12 out of 14 cases (86%)), data obtained with the two methods were consistent (Figure 3.5). In 9 cases (64%) the result was also statistically significant (p-Value ≤ 0.05), however, in a number of cases, ***these changes***

seem to be LSK-independent, as also observed after changing the cell culture medium (control; C; white bars in Figure 3.7). Moreover, as already observed for LSK cells, in most cases microarray results overestimated the fold-change detected by qPCR. However, in 6 cases (*Ctgf*, *Plaur*, *Mmp15*, *Hmga2*, *Ltbp2* and *Nfkbia*) the opposite was true. Similarly, **for the down-regulated genes, the comparison of microarray-based results with qPCR yielded less agreement**: 6 out of 13 (46%) of the measurements were consistent in terms of the direction of the change in expression. In all cases, significant quantitative differences were observed (Figure 3.5(B)).

Finally, to further evaluate the performance of our microarrays, **we also calculated the signal detection sensitivity, specificity, accuracy and precision of the microarrays**. First, positives (i.e., genes that are differentially expressed) and negatives (i.e., genes below differential expression threshold) from the microarray data set were determined, according to the differential expression thresholds defined as $\log_2FC \geq 1.0$ or $\log_2FC \leq -1.0$ and $p\text{-Value} \leq 0.05$. Thereafter, for each positive and negative, the microarray results were directly compared to their corresponding \log_2FC and $p\text{-Values}$ obtained using RT-qPCR assay, the latter being regarded as the "ground truth". Table 3.5 lists both \log_2 fold changes and $p\text{-Values}$ (MOE430.2 Affymetrix chip vs. RT-qPCR) for the selected target genes. Concordance with microarrays \log_2FC and $p\text{-Values}$ was determined by counting the number of true positives (TP, differential expression detectable by both RT-qPCR assay and microarray), true negatives (TN, differential expression not detectable by either RT-qPCR assay and microarray i.e., $\log_2FC < 1.0$ or $\log_2FC > -1.0$ and/or $p\text{-Value} > 0.05$), false positives (FP, differential expression detectable by microarrays but not by RT-qPCR), and false negatives (FN, differential expression detectable by RT-qPCR but not by MOE430.2). Finally, a 2×2 contingency table was constructed (Table 3.2) for the microarray data, containing the counts of the 4 combinations of classification. Based on this matrix, the following statistics were calculated for the microarray: *SENSITIVITY*, *SPECIFICITY*, *ACCURACY* and *PRECISION* (see *Methods and Data* for definitions).

As a result, we estimated that the microarray detection sensitivity and specificity was 62% and 49%, respectively, meaning that more than a half of all differentially expressed genes are detected by the microarray and a half of the negatives are genes expressed at a constant level in either 0 h (Day0; d0; unclutured cells) or 24 h (Day1; d1; co-culture-derived cells). An accuracy of 55% indicates that in more than a half of the cases the values measured by the microarray are the same as those obtained by qPCR and 50% precision indicates that in a half of the cases repeated measurements under unchanged conditions show the same results. The trade-offs between the measures are represented graphically as a receiver operating characteristic (ROC) curve (see Figure 3.8), showing the true positive rate (TPR)

3.5 Phenotypic and functional comparison of freshly isolated vs. 24 h co-culture-derived LSK cells

against the false positive rate (FPR) for the different possible fold changes (\log_2 FC) and p-Values. According to the ROC curve, \log_2 fold change (FC) values from the microarray analysis should be preferentially used to accurately select the differentially expressed genes (DEGs) false positive rate. At the same time, selecting DEGs using their p-Values would be worse than a random selection. This is reflected also by the area under the ROC curve (AUC), measuring the accuracy (where $AUC = 1.0$ represents a perfect test and an AUC of 0.5 represents a worthless test), which is 0.64 and 0.38 for \log_2 FC and p-Values, respectively.

In conclusion, rigorous filtering of the microarray data is definitely necessary, both in terms of fold change (\log_2 FC ≥ 1.0 or \log_2 FC ≤ -1.0) and p-Value (≤ 0.05) or even the p-Value corrected for multiple testing such as using the Benjamini–Hochberg procedure (BH ≤ 0.05).²³¹ In addition, independent validation of each selected target gene using qPCR is necessary, whereas, for stromal cells, it is also essential to include an additional control 24 h after changing the culture medium (medium control; C), since this may strongly influence the observed gene expression patterns in UG26-1B6 cells.

3.5 Phenotypic and functional comparison of freshly isolated vs. 24 h co-culture-derived LSK cells

In order to *relate gene expression patterns observed after 24 h (Day1; d1) co-culture to phenotypic and functional changes occurring within LSK cell compartment*, the cell cycle status and apoptosis rates, as well as the progenitor producing property and *in vivo* engraftment potential was examined (Figure 3.1; step 3 and Figure 3.9).

Cell proliferation quantitation using bromodeoxyuridine (BrdU) uptake The BrdU assay was used to quantitate cell proliferation, i.e., the percentage of cells entering and progressing through the S (DNA synthesis) phase of the cell cycle. As it can be seen in Figure 3.9(A), the analysis revealed that in both fresh (Day0; d0) LSK and 24 h co-culture-derived (Day1; d1) Cd45+LSK cells, the distribution of the cell cycle positions and active DNA synthetic activities of cells were approximately the same: 24.3 vs 21.8% resided in G0/G1, 5.16 vs 3.2% in S and 0.71 vs 0.62 % of the cells resided in G2/M phase of the cell cycle. Thus, it can be concluded that LSK cells, most probably, have not entered the S phase of the cell cycle (i.e., no recently synthesized DNA) during the first 24 h of co-culture, since there is no shift in the distribution of the different cell cycle phases, as compared to fresh LSK cells.

Cell kinetics tracking using 5- and 6-carboxyfluorescein diacetate succinimidyl ester (CFSE) labeling The cell proliferation kinetics tracking using 5- and 6-carboxyfluorescein diacetate succinimidyl ester (CFSE) labeling was used to compare the proliferation kinetics of 24 h co-culture derived LSK cells vs. colcemid treated cells (i.e., cells that did not

undergo any cell divisions) (Figure 3.9(B)). This analysis demonstrated that 24 h (Day1; d1) co-culture-derived CFSE+ Cd45+ LSK culture-derived cells form a single CFSE-high cell population which overlaps with that of colcemid treated, that is undivided, control cells, indicating that LSK cells have not undergone any cell divisions during this time in co-culture.

Apoptosis assay using Annexin V staining In order to get an estimate of the apoptosis rates during the 24 h (Day1; d1) co-culture with stromal cells, Annexin V apoptosis assay was performed on both, freshly isolated and culture-derived Cd45+ LSK cells. As a positive control, cells treated with 2 mM H_2O_2 for 4 h were used. As it can be inferred from Figure 3.9(C), no early apoptotic cells (PI-, AnnexinV+) cells could be detected within the LSK and Cd45+ LSK cell compartment. The identification of late apoptotic and necrotic cells LSK cells (PI+, AnnexinV+) was hindered by the fact that PI+ cells are excluded prior to the cell surface marker analysis.

Colony forming cell (CFC) assay *In vitro* colony-forming cell (CFC) assay was performed to assess the numbers of CFU-GM, CFU-GEMM, and BFU-E colonies after LSK cells were co-cultured with stromal cells for 24 h. Freshly isolated LSK cells gave rise to greater numbers of colonies than 24 h co-culture-derived LSK cells (~59% vs. 38% out of 250 cells seeded; Figure 3.9(D)), indicating a decreased progenitor forming capacity of LSK cells after the co-culture.

In vivo* transplantation assay** To examine the HSC activity of the 24 h co-culture-derived Cd45+ LSK cells, we performed the *in vivo* transplantation assay, in which 1000 sorted fresh or cultured cells were transplanted together with 10^5 helper BM and 5×10^5 helper SP cells into lethally irradiated recipients. We observed that the repopulating capacity of Cd45+ LSK cells was significantly increased (48.6% vs. ~25 % donor cells) in the bone marrow (BM), 16 weeks after transplantation (Figure 3.9(E)). Hence, it seems that ***although the progenitor forming capacity of LSK cells has decreased, their repopulating capacity has even increased after the 24 h in co-culture, indicating that co-culture-derived cells have not lost their stem cell properties.²³²

3.6 Identifying Ctgf using candidate gene prioritization

As our time-series (TS) microarray data clustering using Short Time-series Expression Miner (STEM)²¹² <http://www.cs.cmu.edu/~jernst/stem/> and the `limma`²¹³ analysis of differential expression still resulted in ***hundreds of potential candidate genes*** (see Tables S1 and S2), we thought to further ***reduce this to a manageable number of target genes for further experimental validation*** and downstream analysis. Traditional candidate gene selection approach is largely limited by its reliance on the priori knowledge about

3.6 Identifying *Ctgf* using candidate gene prioritization

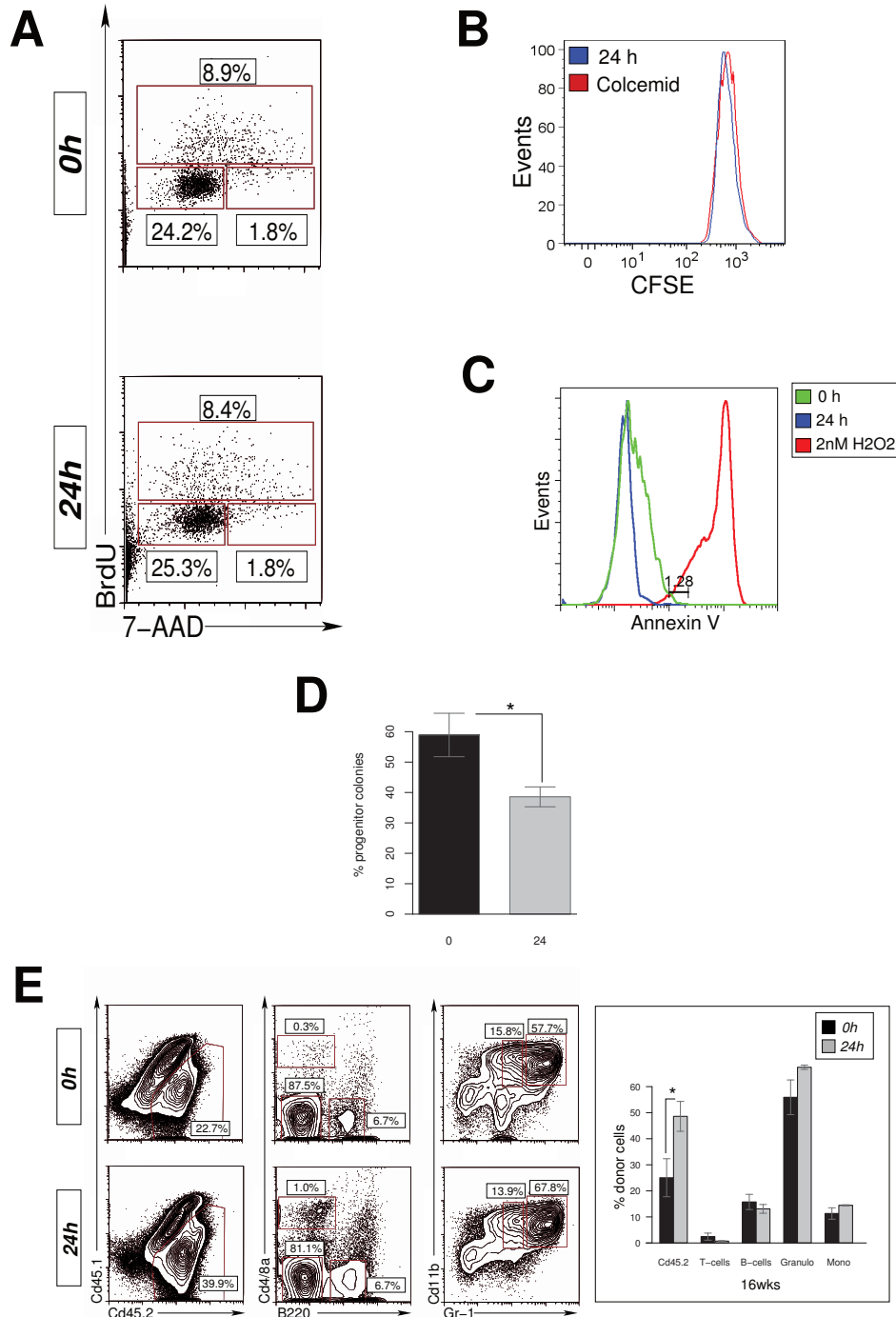


Figure 3.9: Phenotypic and functional characterization of freshly isolated (Day0; d0) vs. 24 h co-culture-derived (Day1; d1) LSK cells: (A) Cell incorporated BrdU and total DNA content (with 7-AAD) to enumerate cells in G0/1, S and G2/M phases of the cell cycle. (B) CFSE fluorescence profile of CFSE+LSK cells recovered 24 h after co-culture with UG26-1B6 stromal cells either with (red peak) or without (blue peak) colcemid. (C) Annexin V fluorescence profile of fresh (green peak) vs. co-culture-derived (blue peak) LSK cells as compared to cells treated with 2 mM H_2O_2 for 4 h (red peak). (D) Percentage of progenitors scored from 250 fresh or 24 h co-culture-derived LSK cells ($n = 3$). (E) Frequency of donor (Cd45.2+), myeloid (Gr1+/med, Cd11b+) and lymphoid (B220+, Cd4/Cd8a+) lineage cells in the bone marrow (BM) of transplanted mice 16 weeks after transplantation.

3 Results

the physiological, biochemical or functional aspects of possible candidates²³³. However, we wanted to select the candidates in a possibly unbiased way. To do so, several gene prioritisation methods have been developed,^{224,234,235} most of them relying on the assumption that similar phenotypes are caused by genes with similar or related functions. Therefore these algorithms make use of functional annotations, gene-expression data or sequence-based features. Several recent studies²³⁶⁻²⁴³ also utilize protein-protein interactions (PPI) for candidate gene prioritisation, motivated by the observation that genes responsible for the same phenotype, e.g. causing the same or similar diseases, tend to lie close to one another in a network of protein-protein interactions.²⁴¹ In addition, also mouse phenotype data have been utilized.²²⁴ In this study, candidate gene prioritisation was performed using the ToppGene Suite <http://toppgene.cchmc.org>,²¹⁷ which enables both functional annotation-based and PPI network analysis-based candidate gene prioritisation.

3.6.1 Candidate gene prioritisation in LSK cells after 24 h co-culture (Day1; d1)

For LSK cells, the training gene set was retrieved by performing extensive biomedical literature search using the text-mining tool EXCERBT (Extraction of Classified Entities and Relations from Biomedical Texts)²¹⁸ <http://mips.helmholtz-muenchen.de/excerbt/>. Co-occurrence search was employed in order to retrieve all the *genes previously associated with hematopoiesis*. Thereafter, false positives (e.g., due to the intrinsic ambiguity in most acronyms) were discarded by manual curation. By this, we compiled a list of 374 “seed” training genes shown to modulate hematopoietic stem cells (HSCs) or hematopoiesis in general. Thereafter, ToppFun from the ToppGene Suite was utilized to generate a representative profile of the training genes (i.e., hematopoiesis-associated genes, in this case) using 14 different features and identifying over-representative terms from the training genes. As a result, this analysis also yielded several mouse phenotypic data listing additional genes associated with phenotypes such as 'Leukemia' (HP:0001909), 'Acute leukemia' (HP:0002488), 'Hematological neoplasia' (HP:0004377), 'abnormal hematopoiesis' (MP:0002123), 'abnormal hematopoietic cell number' (MP:0011180) and 'abnormal hematopoietic stem cell morphology' (MP:0004808). Genes from these categories were added to the seed gene list, yielding a set of training (hematopoiesis-associated) 1737 genes (Table S5). The test gene set consisted of 84 transcripts demonstrating a $\log_2FC \geq 1.0$ or ≤ -1.0 at the FDR threshold of ≤ 0.25 after 24 h co-culture (Day1; d1) with stromal cells.

As it can be inferred from Table 3.3, in both scoring categories, the highest ranked gene was a cytoskeletal element vimentin (Vim) (ToppGene score: 0.72 and p-Value: 4.74e-05; ToppNet PPI count: 118 and p-Value: 6.7e-04). In the context of hematopoiesis, alterations

Table 3.3: ToppGene candidate gene gene prioritization results in LSK cells showing the top 15 highly ranked genes, when trained against a set of hematopoiesis-related genes (Table S5). ToppGene: functional annotation based disease candidate gene prioritization, which uses fuzzy-based similarity measure to compute the similarity between differentially expressed genes (DEGs; test set) and hematopoiesis-related genes (training set) based on semantic annotations. ToppNet: protein-protein interaction network (PPIN) based candidate gene prioritization uses K-Step Markov method (step size = 6, default) to estimates the relative importance of each test set gene in the PPIN to the training genes.

| Rank | ToppGene | | | | ToppNet | | |
|------|----------|---------|--------|----------|-----------|-------|---------|
| | Gene | Score | pValue | Gene | PPI | Score | |
| 1 | 1 | Vim | 0.72 | 4.74e-05 | Vim | 118 | 6.7e-04 |
| 2 | 2 | Ctgf | 0.70 | 1.28e-04 | Igfbp3 | 28 | 3.3e-04 |
| 3 | 3 | Lpl | 0.63 | 1.58e-04 | Ascc2 | 38 | 2.0e-04 |
| 4 | 4 | Igfbp3 | 0.67 | 1.73e-04 | Rrp1b | 28 | 1.8e-04 |
| 5 | 5 | Ccng1 | 0.61 | 3.47e-04 | Ctgf | 10 | 1.5e-04 |
| 6 | 6 | Tyms | 0.51 | 1.16e-03 | Igfbp4 | 7 | 1.1e-04 |
| 7 | 7 | Trem1 | 0.48 | 1.72e-03 | Cd200r1 | 1 | 1.1e-04 |
| 8 | 8 | Cd163 | 0.47 | 3.99e-03 | Htra1 | 13 | 9.7e-05 |
| 9 | 9 | Csnk1g3 | 0.57 | 4.47e-03 | Tbxa2r | 11 | 8.3e-05 |
| 10 | 10 | Pkm2 | 0.48 | 4.56e-03 | Lpl | 15 | 8.2e-05 |
| 11 | 11 | Zfand5 | 0.49 | 4.64e-03 | Golga3 | 10 | 7.3e-05 |
| 12 | 12 | Tbxa2r | 0.53 | 5.18e-03 | Hif3a | 5 | 6.8e-05 |
| 13 | 13 | Hif3a | 0.60 | 5.80e-03 | Cttnbp2nl | 16 | 6.6e-05 |
| 14 | 14 | Igfbp4 | 0.51 | 6.52e-03 | Pkm2 | 12 | 5.7e-05 |
| 15 | 15 | Degs1 | 0.36 | 9.03e-03 | Ccng1 | 9 | 4.9e-05 |

of vimentin intermediate filament (IF) expression have been observed in human hemopoietic committed precursors as they differentiate into mature cells of the erythroid, granulomonocytic, megacaryocytic and lymphoid lineages.²⁴⁴ ***Connective tissue growth factor (Ctgf/Ccn2)(ToppGene score: 0.70 and p-Value: 1.28e-04) is the second and fifth highest ranked gene based on its functional annotation and protein-protein interactions (PPIs) to known hematopoiesis regulators, respectively.*** Ctgf has been found to be the highest over-expressed gene in B-cell ALL (acute lymphoblastic leukemia), suggesting it might have prognostic relevance.²⁴⁵ Moreover, another CCN family member, nephroblastoma over-expressed (Nov/Ccn3), has been connected with the modulation of self-renewal and maturation of a number of cell lineages including hematopoietic, osteogenic and chondrogenic, and its expression has been shown to be disrupted in chronic myeloid leukemia (CML) as a consequence of the BCR-ABL oncogene and allows the leukemic clone to evade growth regulation.²⁴⁶ Interestingly, connective tissue growth factor (Ctgf) and

3 Results

insulin-like growth factor binding proteins (Igfbp3 and Igfbp4) were shown to be higher expressed in primitive hematopoietic progenitors supportive feeder layers of human mesenchymal stromal cells.²⁴⁷ Similarly, insulin-like growth factor-binding proteins 3 and 4 were also found among the genes expressed at a higher level in our HSC-supporting stromal cell lines, EL08-1D2 and UG26-1B6.¹⁹⁹ Interestingly, insulin-like growth factor-binding protein 3 (Igfbp3; ToppGene score: 0.67 and p-Value: 1.73e-04; ToppNet PPI count: 28 and p-Value: 3.3e-04) and insulin-like growth factor-binding protein 4 (Igfbp4; ToppGene score: 0.51 and p-Value: 6.52e-03; ToppNet PPI count: 7 and p-Value: 1.1e-04) are also among the top 15 highest scoring candidate genes from the 24 h-co-culture-derived LSK cell signature. Igfbp3 is a hypoxia-regulated factor, inducing growth inhibition.²⁴⁸ Igfbp4 is an inhibitor of canonical Wnt signaling.²⁴⁹ Lipoprotein lipase (Lpl) (ToppGene score: 0.63 and p-Value: 1.58e-04; ToppNet PPI count: 15 and p-Value: 8.2e-05) promotes binding of lipoproteins to cell surface heparan sulfate proteoglycans and LDL receptors.²⁵⁰ Lpl would be interesting to further investigate in the context of lipid rafts, which have been shown to play a critical role in HSC fate decisions, as freshly isolated HSCs from the BM niche lack lipid raft clustering, accompanied by repression of the AKT-FOXO signaling pathway and abundant p57Kip2 cyclin-dependent kinase inhibitor expression. Conversely, lipid raft clustering induced by cytokines was essential for HSC re-entry into the cell cycle.¹⁰³ In the pool of myeloid progenitors, differentiation toward CMP and MEP displays accumulation of a limited number of mitotic cyclins (mostly cyclin G1 (Ccng1)(ToppGene score: 0.61 and p-Value: 3.47e-04; ToppNet PPI count: 9 and p-Value: 4.9e-05)) that likely contribute to their intrinsic proliferation index.²⁵¹ Scavenger receptor cysteine-rich type 1 protein M130 (Cd163), the hemoglobin-haptoglobin receptor, has been mostly reported to be expressed on monocytes/macrophages, however, it has also been demonstrated that a sub-population of hematopoietic stem/progenitor cells is also expressing it.²⁵² Pyruvate kinase (Pkm2) is a glycolytic enzyme that has been associated with metabolic regulation of hematopoietic stem cells in the hypoxic niche.²⁵³ Thromboxane A2 receptor (Tbxa2r) is known as a potent stimulator of platelet aggregation.²⁵⁴ Hypoxia-inducible factor 3-alpha (Hif3a) together with Hypoxia-inducible factor 1-alpha (Hif1a) and Hypoxia-inducible factor 2-alpha (Hif2a) constitutes the HIF-a family of transcription factors. Hif1a is part of the hypoxia response system in hematopoietic stem cells.²⁵⁵ Hence, based on knowledge from the literature, the top ranked genes seem to be involved in diverse aspects of the hematopoietic system.

3.6.2 Candidate gene prioritisation in UG26-1B6 cells after 24 h co-culture

A similar strategy was adapted to rank the differentially expressed genes in stromal cells. In all cases, the test gene set consisted of genes differentially expressed ($\log_2\text{FC} \geq 1.0$ or ≤ -1.0 , $\text{FDR} \leq 0.25$) after 24 h in co-culture with LSK cells (Day1; d1), and additionally filtering for metabolic effects, as described in previous sections. In addition, we included also genes from our microarray validation set, for which we explicitly tested for the culture medium effects (see Table S4 and Figure 3.7). This resulted in a set of 2837 test genes. As already mentioned above, two midgestation-derived stromal clones UG26-1B6 (urogenital ridge-derived) and EL08-1D2 (embryonic liver-derived) have been earlier demonstrated to preserve the maintenance of repopulating HSCs in an *in vitro* co-culture, without added cytokines, for periods of at least four weeks.^{198,199} In order to search for factors that might be involved in HSC maintenance, the gene expression profile of ***EL08-1D2 and UG26-1B6 stromal cells was also compared with that of four HSC-non-supportive clones (UG15-1B7, AM20-1B4, EL28-1B3 and AM30-3F4)***.²⁵⁶ This list of 450 factors being higher expressed ($\log_2\text{FC} \geq 1.0$, $\text{p-Value} \leq 0.05$) in HSC maintaining cell lines was used as the first training data set. ToppGene again first generated a representative profile of the training genes using 14 different features and identified over-representative terms from the training genes. In HSC-supporting factor list, following GO categories were found to be enriched: (1) *Biological processes* embryonic limb morphogenesis (GO:0030326, p-Value: 1.580e-3, 11 factors) and response to wounding (GO:0009611, p-Value: 2.863e-3, 37 factors) and (2) *Cellular Component* extracellular region part (GO:0044421, p-Value: 4.119E-4, 38 factors). Over-represented Mouse Phenotypes were mostly associated with abnormalities in morphology: abnormal jaw morphology (MP:0000454, p-Value: 1.184e-2, 16 factors), abnormal mandible morphology (MP:0000458, p-Value: 1.713e-2, 13 factors) and abnormal sphenoid bone morphology (MP:0000104, p-Value: 3.545e-2, 9 factors). In addition, the list contained 14 Cebp targets (term 'TF binding site', p-Value: 4.332e-2) and 47 Eed targets (term 'Co-expression', source 'MSigDB: C2.cgp', p-Value: 5.309e-9), 40 Suz12 targets (term 'Co-expression', source 'MSigDB: C2.cgp', p-Value: 1.636e-5), as well as 44 genes possessing the trimethylated H3K27 (H3K27me3) mark in their promoters (term 'Co-expression', source 'MSigDB: C2.cgp', p-Value: 1.256e-6). Again, the top 15 highest ranked genes after both prioritisation analysis are shown (Table 3.4). In this case, scoring results differ between the two (ToppGene vs. ToppNet) methods. Among the ToppGene highest ranked co-culture induced genes, that could also be validated using RT-qPCR analysis (Figure 3.7(A), however, note that in several cases the induction seems to be LSK-independent) were a component of the latent

3 Results

TGF- β complex Latent-transforming growth factor beta-binding protein 2 (Ltbp2) (ToppGene score: 0.32 and p-Value: 3.43e-08) and a matrix remodeling protein Connective tissue growth factor (Ctgf) (ToppGene score: 0.32 and p-Value: 3.52e-05, also among the highest scored genes in LSK cells). Stromal cell-derived factor 1 (Cxcl12)(ToppGene score: 0.23 and p-Value: 2.66e-04), a factor associated with HSC maintenance²⁵⁷ and mobilization,²⁵⁸ was also among the significantly induced genes after the co-culture, as opposed to the microarray results (Figure 3.5(B)). Among the ToppGene highest ranked co-culture suppressed genes that could also be validated using RT-qPCR (Figure 3.7(B)) were Atrial natriuretic peptide receptor 3 (Npr3) (ToppGene score: 0.24 and p-Value: 2.04e-05), a factor which, among others, may regulate skeletal development,²⁵⁹ Insulin-like growth factor I (Igf1) (ToppGene score: 0.25 and p-Value: 2.09e-04), known for its growth-promoting effects on hematopoietic cells,²⁶⁰ Fibrillin-1 (Fbn1) (ToppGene score: 0.29 and p-Value: 3.88e-04), a modulator of endogenous TGF- β and BMP bioavailability during bone formation²⁶¹ and Neuropilin-1 (Nrp1) (ToppGene score: 0.20 and p-Value: 6.59e-04), proposed to act as a receptor on stromal cells mediating interactions between stroma and primitive hematopoietic cells.²⁶² The differential expression of adhesion-associated molecule²⁶³ Thrombospondin-1 (Thbs1) (ToppGene score: 0.23 and p-Value: 3.11e-05) could not be confirmed by RT-qPCR (Figure 3.7(B)).

Alternatively, the previously compiled literature-derived list of hematopoietic regulators, supplemented with ToppGene-provided mouse and human prophenotypic data (as described above) was searched for factors with GO CC annotations 'GO:0005615:extracellular space' and 'GO:0044421:extracellular region part', in an attempt to select possible external e.g., niche-associated regulators of hematopoiesis. This yielded a sub-set of 245 training genes (Table S5). As it can be inferred from Table 3.5, also when trained against a literature-derived set of extrinsic (secreted) hematopoiesis regulators, high rankings achieve previously found factors such as Neuropilin-1 (Nrp1) (ToppGene rank 5, score: 0.58 and p-Value: 2.10e-08; ToppNet rank: 26, score: 8.93e-04, PPIs: 20), Connective tissue growth factor (Ctgf) (ToppGene rank 6, score: 0.66 and p-Value: 2.93e-08; ToppNet rank: 40, score: 5.83e-04, PPIs: 10) and Latent-transforming growth factor beta-binding protein 1 (Ltbp1) (ToppGene rank: 34, score: 0.66 and p-Value: 1.18e-05). Among the ToppGene highest ranked co-culture induced genes, that could also be validated using RT-qPCR analysis (Figure 3.7(A)) were Urokinase plasminogen activator surface receptor (uPAR/Plaur) (ToppGene rank 23, score: 0.57 and p-Value: 2.28e-06; ToppNet rank: 21, score: 5.75e-04, PPIs: 21), demonstrated to regulate the proliferation, marrow pool size, engraftment, and mobilization of mouse hematopoietic stem/progenitor cells²⁶⁴ and a coordinator of the commitment of mesoderm to hematopoietic, endothelial, and cardiac lineages in embryoid bodies²⁶⁵ and Wnt2 protein (ToppGene rank

Table 3.4: ToppGene candidate gene gene prioritisation results in UG26-1B6 stromal cells showing the top 15 highly ranked genes, when trained against the gene set of 450 factors being higher expressed ($\log_2FC \geq 1.0$, $p\text{-Value} \leq 0.05$) in HSC maintaining EL08-1D2 and UG26-1B6 stromal cell lines as compared to the HSC-non-supportive clones (UG15-1B7, AM20-1B4, EL28-1B3, and AM30-3F4).²⁵⁶ ToppGene: functional annotation based disease candidate gene prioritization, which uses fuzzy-based similarity measure to compute the similarity between differentially expressed genes (DEGs; test set) and hematopoiesis-related genes (training set) based on semantic annotations. ToppNet: protein-protein interaction network (PPIN) based candidate gene prioritization uses K-Step Markov method (step size = 6, default) to estimate the relative importance of each test set gene in the PPIN to the training genes.

| Rank | ToppGene | | | ToppNet | | |
|------|----------|-------|----------|-----------|-----|----------|
| | Gene | Score | pValue | Gene | PPI | Score |
| 1 | Ltbp1 | 0.32 | 3.43e-08 | Egfr | 256 | 1.82e-03 |
| 2 | Prrx1 | 0.30 | 3.09e-07 | Tgfb1 | 163 | 1.79e-03 |
| 3 | Nr3c1 | 0.25 | 3.61e-07 | Esr1 | 308 | 1.74e-03 |
| 4 | Pde4b | 0.26 | 2.23e-06 | Hdac2 | 195 | 1.65e-03 |
| 5 | Tpm1 | 0.26 | 2.88e-06 | C1galt1c1 | 3 | 1.57e-03 |
| 6 | Gja1 | 0.28 | 2.96e-06 | Prkca | 187 | 1.47e-03 |
| 7 | Prep1 | 0.27 | 4.54e-06 | Cttnb1 | 199 | 1.45e-03 |
| 8 | Cdh13 | 0.25 | 1.68e-05 | Calm1 | 124 | 1.41e-03 |
| 9 | Npr3 | 0.24 | 2.04e-05 | Csnk2a1 | 195 | 1.36e-03 |
| 10 | Tnc | 0.30 | 2.91e-05 | Ywhaz | 185 | 1.27e-03 |
| 11 | Thbs1 | 0.23 | 3.11e-05 | Mapk1 | 190 | 1.18e-03 |
| 12 | Ctgf | 0.32 | 3.52e-05 | Pik3r1 | 146 | 1.17e-03 |
| 13 | Cdh2 | 0.25 | 3.64e-05 | Akt1 | 155 | 1.13e-03 |
| 14 | Zeb1 | 0.26 | 3.74e-05 | Fyn | 168 | 1.06e-03 |
| 15 | Fut8 | 0.20 | 6.17e-05 | Mark2 | 34 | 1.01e-03 |

Table 3.5: ToppGene candidate gene gene prioritization results in UG26-1B6 stromal cells showing the top 15 highly ranked genes, when trained against a set of hematopoiesis-related genes with GO CC annotations 'GO:0005615:extracellular space' and 'GO:0044421:extracellular region part' (Table S5). ToppGene: functional annotation based disease candidate gene prioritization, which uses fuzzy-based similarity measure to compute the similarity between differentially expressed genes (DEGs; test set) and hematopoiesis-related genes (training set) based on semantic annotations. ToppNet: protein-protein interaction network (PPIN) based candidate gene prioritization uses K-Step Markov method (step size = 6, default) to estimates the relative importance of each test set gene in the PPIN to the training genes.

| Rank | ToppGene | | | | ToppNet | | |
|------|----------|--------|--------|----------|----------|-------|----------|
| | Gene | Score | pValue | Gene | PPI | Score | |
| 1 | 1 | Col1a2 | 0.71 | 2.06e-10 | Sumo2 | 1165 | 2.55e-03 |
| 2 | 2 | Fgf7 | 0.67 | 1.88e-08 | Kiaa0101 | 853 | 2.21e-03 |
| 3 | 3 | Jag1 | 0.58 | 3.91e-08 | Cul3 | 1115 | 2.13e-03 |
| 4 | 4 | Nrp1 | 0.63 | 1.37e-07 | Jak1 | 78 | 1.71e-03 |
| 5 | 5 | Timp3 | 0.69 | 1.57e-07 | Pik3r1 | 168 | 1.65e-03 |
| 6 | 6 | Pik3r1 | 0.62 | 3.57e-07 | Sirt7 | 657 | 1.45e-03 |
| 7 | 7 | Ctgf | 0.68 | 5.31e-07 | Rad21 | 225 | 1.44e-03 |
| 8 | 8 | Tgfbr1 | 0.61 | 5.65e-07 | Fyn | 188 | 1.34e-03 |
| 9 | 9 | Wnt2 | 0.54 | 7.28e-07 | Traf6 | 201 | 1.30e-03 |
| 10 | 10 | Cd44 | 0.62 | 8.24e-07 | Fbxo6 | 149 | 1.28e-03 |
| 11 | 11 | Actn1 | 0.62 | 1.14e-06 | Cand1 | 645 | 1.20e-03 |
| 12 | 12 | Cdk6 | 0.57 | 1.27e-06 | Prkca | 209 | 1.18e-03 |
| 13 | 13 | Ctss | 0.58 | 1.38e-06 | Ywhaz | 341 | 1.13e-03 |
| 14 | 14 | Mmp14 | 0.63 | 1.52e-06 | Ctnnb1 | 243 | 1.11e-03 |
| 15 | 15 | Plaur | 0.59 | 1.82e-06 | Ikbkg | 228 | 1.09e-03 |

3.7 Experimentally investigating the functional role of UG26-1B6-derived *Ctgf* in hematopoiesis

25, score: 0.53 and 2.69e-06: 2.28e-06; ToppNet rank: 31, score: 7.46e-04, PPIs: 4). The up-regulation of HSC mobilization- associated molecule²⁶⁶ Proteinase 3 (Prtn3) (ToppNet rank: 29, score: 8.29e-04, PPIs: 14) could not be confirmed by RT-qPCR (Figure 3.7(A)).

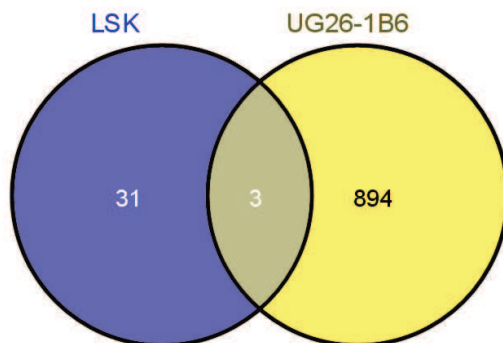


Figure 3.10: The Venn diagram showing the overlap of the TopGene <http://toppgene.cchmc.org>²²⁴ candidate gene prioritization results in LSK and UG26-1B6 stromal cells. Significant hits ($p\text{-Value} \leq 0.05$) from the functional annotation-based analysis were selected (27 and 897 in LSK and UG26-1B6 cells, respectively). There were 3 overlapping genes: *Ctgf*, *Ccng1*, *Degs1*. The diagram was generated using VENNY <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.²⁶⁷

Interestingly, *the overlap between test genes prioritized in this manner in LSK and UG26-1B6 stromal cells was minimal: *Ctgf*, *Ccng1*, *Degs1** (Figure 3.10). Connective tissue growth factor (*Ctgf*) was one of the genes induced in both LSK ($\log_2\text{FC}$ 5.6, $\text{FDR} \leq 0.05$) and stromal ($\log_2\text{FC}$ 3.5, $\text{FDR} \leq 0.03$) cells after 24 h in co-culture. According to the TopGene functional annotation-based prioritization results, *Ctgf* was the second and seventh highest ranked gene in LSK and stromal cells, respectively (Figure 3.1; step 4). Therefore, we first wanted to see, whether we can independently confirm our microarray results, reporting the up-regulation of *Ctgf* in both LSK ($\log_2\text{FC}$ 5.6, $\text{FDR} \leq 0.05$) and stromal ($\log_2\text{FC}$ 3.5, $\text{FDR} \leq 0.03$) cells after 24 h in co-culture (Figure 3.1; step 5).

3.7 Experimentally investigating the functional role of UG26-1B6-derived *Ctgf* in hematopoiesis

3.7.1 Independent confirmation of microarray results

In LSK cells, the induction of *Ctgf* could be confirmed both on mRNA and relative protein level, using RT-qPCR and Immunofluorescence (IF) staining, respectively (Figure 3.12(A)(a,c,e)). In particular, it seems that *Ctgf* is barely expressed by freshly isolated Lin-Sca1+c-Kit+ cells, however, its expression increases ~ 10 -fold after the co-culture. In stromal cells, *Ctgf* has

been found previously among the genes highly expressed by UG26-1B6 in the extracellular space.²⁶⁸ Indeed, our experiments also demonstrate that UG26-1B6 culture medium contains high levels (~ 14 ng/mL) of secreted Ctgf already prior to the co-culture (Figure 3.12 (A)(d); 0h). The presence of LSK cells does not seem to change its secreted protein levels (Figure 3.12(A)(d); 24h). On the mRNA level, we could confirm an up-regulation of Ctgf after 24 h co-culture (Figure 3.12(A)(b)), however, the up-regulation of Ctgf mRNA seems to be LSK (co-culture) independent (Figure 3.7(A)), as its gene expression appears also to be activated 24 h after adding fresh culture medium to the cells (metabolic effects). In fact, Ctgf transcription levels are known to be modulated by factors such as high glucose.²⁶⁹ In addition, Ctgf seems to be more highly expressed in stromal cells, compared to hematopoietic cells (Figures 3.13 and 3.14, for mRNA and protein measurements, respectively), and it has been shown previously that human MSCs, a component of the hematopoietic stem cell 'niche', express high levels of Ctgf compared with leukemia cells.²⁷⁰

3.7.2 Biological description and role in hematopoiesis

Clearly, HSC behavior within their 'niche' involves complex multi-level bidirectional signal processing networks, where multiple environmental cues (cytokines, small molecules cell-cell contacts) guide the stem cell towards specific fate decisions. An ideal experimental technology would be able to measure accurately the concentrations of hundreds and even thousands of different mRNAs and proteins, each possibly subject to a variety of post-translational modifications and should be able to measure all this in a time-dependent, cell and compartment specific manner, under various conditions.²⁷¹ Unfortunately, we currently lack such sophisticated experimental technologies which necessitates selection of candidate molecules, in an ideal case, such that measuring their activity would be most informative for understanding the molecular crosstalk within the network.

In this study, after providing independent confirmation that Ctgf is indeed differentially expressed under the circumstances we used to generate the gene expression profiles, we will focus our downstream analysis on Ctgf and its regulatory networks. Connective tissue growth factor (Ctgf) is a secreted 36-38 kDa cysteine-rich matrix remodeling protein that was first identified in conditioned medium of human umbilical vein endothelial cells. It belongs to a *CCN family of secreted proteins consisting of six members as follows: Cyr61 (cysteine-rich protein 61, Ccn1), Ctgf (connective tissue growth factor, Ccn2), Nov (nephroblastoma over-expressed gene, Ccn3), Wisp1 (Wnt-1-induced secreted protein 1, Ccn4), Wisp2 (Ccn5), and Wisp3 (Ccn6)*. These structurally conserved proteins share an NH₂-terminal signal peptide and *four modular domains* with sequence similarities to insulin-like growth factor-binding proteins, von Willebrand factor

3.7 Experimentally investigating the functional role of UG26-1B6-derived Ctgf in hematopoiesis

type C repeat, thrombospondin type 1 repeat and growth factor cysteine knots characteristic of other growth factors, including platelet-derived growth factor, nerve growth factor and transforming growth factor β (see Figure 3.11).

Each of these domains is encoded by a separate exon, suggesting that CCN genes arose through exon shuffling of pre-existing genes to form proteins with multiple functional domains. The recent emergence of these matricellular regulators has called attention to their functional versatility and mechanisms of actions, which have been ascribed to activities encoded within the four component modular domains, ***each of which can bind several ligands***. Amongst these are other growth factors (e.g., Tgf- β , Bmp4 and -7, Igf1, Vegf), whose function is then modified; cell surface proteins (e.g. integrins, Lrp1, heparan sulphate proteoglycans, tyrosine receptor kinase, through which intracellular signaling may be initiated; and extracellular matrix proteins (e.g. fibronectin) which may act as a sink for CCN proteins and modify their turnover. With so ***many potential interactions*** it is predictable that CCN proteins will ***influence many different biological events***.²⁷³⁻²⁷⁶ Ctgf is typical in this respect. Many responses triggered by it, or by fragments derived proteolytically from it, have been described since it was discovered in 1991. Inasmuch as Ctgf was first identified as a growth factor, it has been tempting to postulate that it might function as a classical growth factor, although a cell surface receptor for Ctgf that resembles a classical growth factor receptor has not been identified to date. Alternatively, the term matricellular coordinator has also been used to designate Ctgf²⁷⁵ and probably more properly characterizes its functionality.^{273,274}

Although, Ctgf is best known as ***a downstream mediator of Tgf β , which also regulates its synthesis and secretion***, it has been associated also with ***Wnt, BMP and p42/p44 MAP kinase signaling***.^{245,277} Other signaling pathways and transcription factors directly stimulated by Ctgf and mediating pertinent biological effects include ***Akt/PKB, JNK and NF- κ B pathways***. In addition, the expression of Ctgf is also regulated by various other stimuli including ***hypoxia, shear stress, bio-mechanical deformation***^{273,274} and, as already mentioned above, ***high glucose***.^{269,278}

In bone development, Ctgf is ***produced from a small population of chondrocytes and acts on all of the mesenchymal cells inside the bone callus to promote the integrated growth of the bone***,²⁷⁹ at the same time haematopoietic stem cells derive regulatory information from bone,³¹ and Ctgf was also found to be more highly expressed in human mesenchymal stromal cells maintaining primitive HPC.²⁴⁷ Moreover, *in vivo* studies suggest a role of Ctgf in ***cell cycle control and proliferation***,^{280,281} and several studies have demonstrated that Ctgf may have an important role in a variety of human cancers. Over-expression of Ctgf is found in prostate cancers, gliomas, and breast cancers.²⁷³

In leukemia, when comparing the gene expression profile of adult acute lymphoblastic

3.7 Experimentally investigating the functional role of UG26-1B6-derived *Ctgf* in hematopoiesis

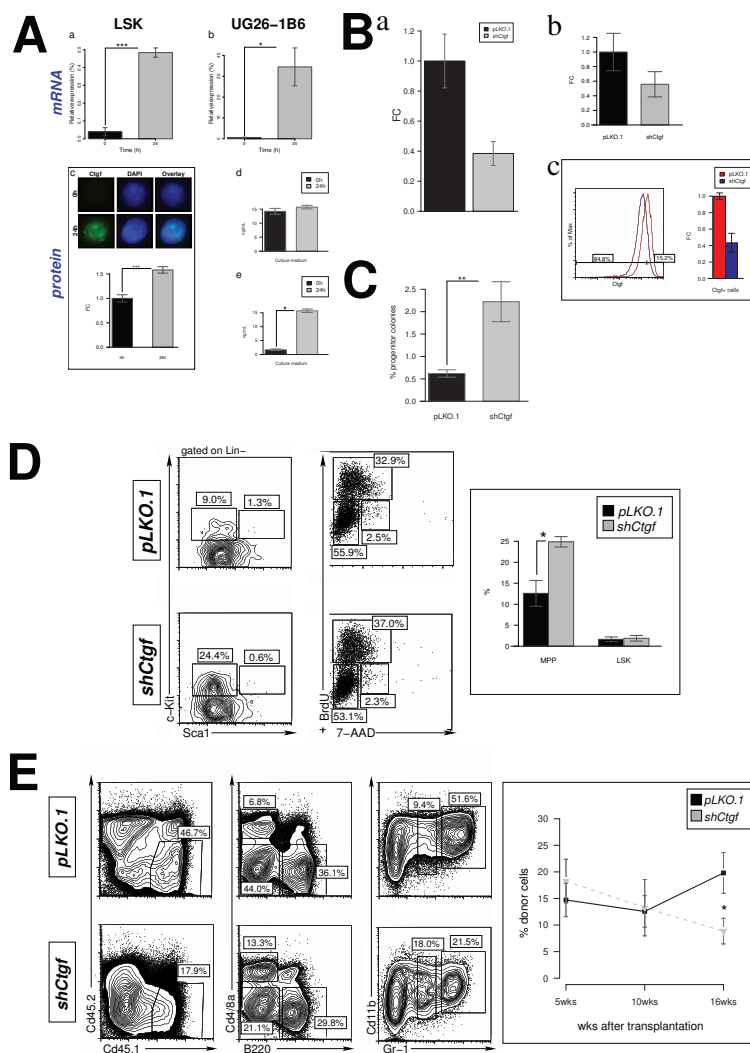


Figure 3.12

leukemia (ALL) to normal hematopoietic and non-ALL samples using microarray analysis, *Ctgf* was the relatively highest *over-expressed gene in precursor B-acute lymphoblastic leukemia (ALL)* compared with the other groups and that increased expression of *Ctgf* is associated with inferior outcome in B-ALL.²⁴⁵ More recently, Lu and Battula²⁸² have also characterized the functional role and down-stream signaling pathways of *Ctgf* in ALL cells by utilizing lentiviral shRNA knock-down of *Ctgf* in RS4;11 and REH ALL cells expressing high levels of *Ctgf* mRNA. Their experiments demonstrated that *silencing of Ctgf resulted in significant suppression of leukemia cell growth* compared to control vector, which was *associated with AKT/mTOR inactivation and increased levels of cyclin-dependent kinase inhibitor p27*. Moreover, *Ctgf* knock-down sensitized ALL cells to vincristine and methotrexate. On the other hand, treatment with an anti-*Ctgf* monoclonal antibody, FG-3019, significantly prolonged survival of mice injected with pri-

Figure 3.12: *Ctgf* regulates hematopoietic stem/progenitor cell (HSC/P) activity and engraftment potential. (A) Independent confirmation of microarray results for *Ctgf*. RT-qPCR results confirming the up-regulation of *Ctgf* mRNA levels in LSK cells after co-cultured with UG26-1B6 stromal cells for 24 h (Day1; d1) and vice versa (a, b). Immunofluorescence (IF) staining results confirming the up-regulation of *Ctgf* relative protein levels in co-culture-derived LSK cells and ELISA results comparing secreted *Ctgf* protein levels in 24 h conditioned culture media when culturing LSK and stromal cells separately or after the co-culture (c-e). (B) Generation of *Ctgf* siRNA knockdown and determination of knockdown efficiency in UG26-1B6 stromal cells at mRNA and secreted protein levels (a-c), using RT-qPCR, ELISA and intracellular protein staining combined with flow cytometry, respectively. (C) Colony-forming cells (CFCs) number (%) in methylcellulose (n = 4). (D) HSC/P analysis using flow cytometry, where LSKs and MMPs were gated as Lin-Sca-1+c-Kit+ cells and Lin-Sca-1-c-Kit+ cells, respectively. (E) Frequency of donor (Cd45.1+), myeloid (Gr1+/med, Cd11b+) and lymphoid (B220+, Cd4/Cd8a+) lineage cells in the peripheral blood (PB) of transplanted mice 5, 10 and 16 weeks after transplantation.

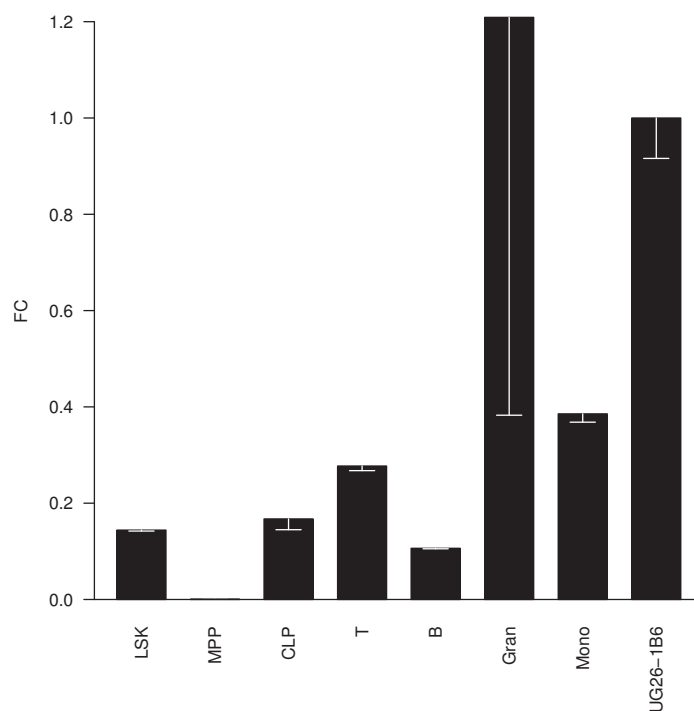


Figure 3.13: Real-time (RT)-qPCR analysis of *Ctgf* mRNA expression in the indicated bone marrow (BM) purified populations, normalized to the expression level in UG26-1B6 stromal cells. LSK: Lin-Sca1+cKit+ hematopoietic stem cells; MPP: Lin-Sca1-cKit+ cells multipotent progenitor cells; CLP: common lymphoid progenitor cells; T: T-cells; B: B-cells; Gran:granulocytes and Mono: monocytes. Primer sequences used are given in Table S11. Results are shown as mean and standard error of three independent experiments.

3.7 Experimentally investigating the functional role of UG26-1B6-derived Ctgf in hematopoiesis

mary xenograft B-ALL cells when co-treated with conventional chemotherapy (vincristine, L-asparaginase and dexamethasone). Therefore, the authors concluded that Ctgf represents a targetable molecular aberration in B-ALL, and blocking Ctgf signaling in conjunction with administration of chemotherapy may represent a novel therapeutic approach for ALL patients. In addition, Battula and colleagues²⁷⁰ have also investigated the **role of Ctgf in mesenchymal stromal cells (MSCs)**, a major component of the normal as well as leukemia bone marrow (BM) microenvironment, since Connective tissue growth factor (Ctgf) is highly expressed in MSCs. They found that Ctgf knocked down (KD) human BM-derived MSCs exhibited **fivefold lower proliferation compared with control MSCs and had markedly fewer S-phase cells**. Moreover, Ctgf KD MSCs differentiated into adipocytes at a sixfold higher rate than controls *in vitro* and *in vivo*. In order to study the effect of Ctgf on engraftment of leukemia cells into BM, the authors developed an *in vivo* model of humanized extramedullary BM (EXM-BM) in NOD/SCID/IL-2rg(null) mice, and demonstrated that transplanted Nalm-6 or Molm-13 human leukemia cells engrafted at a threefold higher rate in adipocyte-rich Ctgf KD MSC-derived EXM-BM than in control EXM-BM.

Currently, to the best of our knowledge, **Ctgf has not been characterized in the context of normal hematopoiesis**. Although, another CCN family member, Nov (Ccn3) has already been identified as a regulator of human hematopoietic stem and progenitor cells.²⁸³ In addition, from systems biology point of view, Connective tissue growth factor (Ctgf) represents an interesting challenge as its functions and mechanisms of regulation are rather complex, producing non-trivial, at times opposite, phenotypic outcomes in response to various related stimuli, since its actions are highly concentration- and molecular context- and cell type- or tissue-dependent, modulated by its numerous interaction partners.^{284–286} Thus, we decided to select Ctgf for further study and focus on Ctgf, its interactome and biological function as a model for other molecules identified above. Since Ctgf has been reported to auto-induce its own expression,²⁸⁷ **we hypothesized that its up-regulation in LSK (stem cells) may be attributed to extrinsic, UG26-1B6-derived Ctgf**.

3.7.3 Generation of Ctgf shRNA knockdown in UG26-1B6 stromal cells

In order to study the putative functional impact of extrinsic stromal cell-derived Ctgf on hematopoiesis (Figure 3.1; step 6), we utilized RNAi to experimentally induce a **stable knock-down of the corresponding protein in UG26-1B6 stromal cells** (see the *Materials and methods* section, *2.2.1.13 Stable knock-down cells for Ctgf* on page 38 for details). We validated the shRNAs according to their knock-down efficiency at both mRNA and protein levels by RT-qPCR, ELISA and intracellular protein staining combined with flow cytometry, respectively (see Figure 3.12(B)(a-c)). We obtained **~ 60% knock-down**

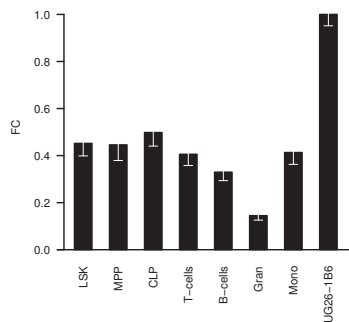
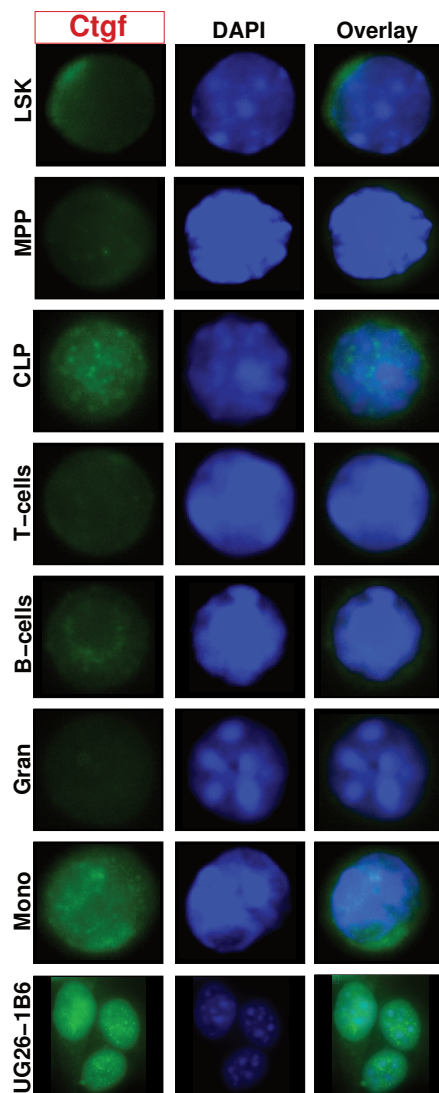


Figure 3.14: Immunofluorescence analysis (IF) of Ctgf relative protein levels in the indicated bone marrow (BM) purified populations, normalized to the expression level in UG26-1B6 stromal cells. LSK: Lin-Sca1+cKit+ hematopoietic stem cells; MPP: Lin-Sca1-cKit+ cells multipotent progenitor cells; CLP: common lymphoid progenitor cells; T: T-cells; B: B-cells; Gran: granulocytes and Mono: monocytes. The primary and secondary antibodies used are listed in Tables 2.8 and 2.10, respectively. Results are shown as mean and standard error of three independent experiments.

efficiency on *Ctgf* at the mRNA and protein level, as compared with control shRNA (pLK0.1) transfection.

3.7.4 *Ctgf* regulates hematopoietic stem/progenitor cell (HSC/P) activity and engraftment potential

As already described above, we hypothesized that the up-regulation of *Ctgf* in 24 h co-culture-derived Lin-Sca1+cKit+ (LSK) hematopoietic stem cells (as shown in Figure 3.12(A)(a,c,e)) could be attributed to extrinsic, UG26-1B6-derived *Ctgf*. To test this hypothesis and to explore the behavior of HSCs within a *Ctgf*-deficient 'niche', we performed a series of *in vitro* co-culture experiments, by considering various different time points (24 h to three weeks). In all cases, the system was perturbed by knocking down (KD) *Ctgf* in UG26-1B6 stromal cells by lentivirally transduced shRNAmir targeting *Ctgf* (see Figure 3.12(B)(a-c)). The cultures using perturbed stromal cells were then compared with those using the unperturbed counterpart (pLK0.1 empty vector carrying stroma). In this regard, the hematopoietic stem/progenitor cells (HSC/Ps) were characterized in terms of their cell surface marker expression, cell cycle status and apoptosis rates, progenitor generating capacity, as well as the *in vivo* engraftment potential. First, we performed co-culture studies of isolated wild-type LSK cells on perturbed and unperturbed UG26-1B6 cells.

As a result, *in vitro*, when comparing 1 week co-cultures of Lineage-depleted (Lin-) cells (see SI *Materials and methods* for details) on si*Ctgf* stromal cells with the control (pLKO.1), our experiments revealed a significant increase ($p \leq 0.05$) in colony-forming cells (CFCs; 2.2% vs. 0.6% , Figure 3.12(C)), suggesting that a decrease of microenvironmental *Ctgf* may **promote hematopoietic progenitor activity** *in vitro*. Further supporting this hypothesis, FACS analysis (Figure 3.12(D)) showed a significant increase in Lin-Sca1-cKit+ early multipotent progenitors (MPPs; 24.9% vs. 12.6%).

In order to also investigate the effect of extrinsic *Ctgf* levels on the *in vivo* HSC activity, the lineage negative (Lin-) fraction of total bone marrow (BM) cells (5000 Lin- cells/well) of Ly5.1 mice co-cultured for 1 week with irradiated UG26-1B6^{si*Ctgf*} or control stromal cells were transplanted into WT B6.Ly5.2 recipients (the input equivalent of 2500 co-cultured Lin- cells) together with 10^5 normal syngeneic bone marrow (BM) and 5×10^5 spleen (SP) cells. Peripheral blood was analyzed after 5, 10 and 16 weeks of transplantation. In addition, bone marrow and spleen were also analyzed after 16 weeks. After 16 weeks, we observed a significantly decreased donor cell compartment (8.9% vs. 19.8% Cd45.1+ donor cells) in the peripheral blood (PB) of recipients (Figure 3.12(E)), indicating that **the absence of extrinsic *Ctgf* significantly reduces the repopulating capacity of HSCs.**

3.8 Network model predicting the role of Ctgf in regulating the LSK cell-cycle status

3.8.1 The top-down approach to Ctgf network modeling

Culling Ctgf interactome from literature and public databases

Ctgf was established as a model molecule possibly playing a role in HSC activation in co-cultures with UG26-1B6 cells. In particular, our results show that the presence or absence of Ctgf significantly affects the maintenance of repopulating HSC in co-cultures of LSK cells and UG26-1B6 stromal cells. As a next step, *to elucidate possible underlying regulatory networks and molecular mechanisms, we first cataloged the complete interactions of Ctgf* (Figure 3.1; step 7). To do so, we first performed an *extensive literature search using the automatic text-mining engine EXCERBT* (Extraction of Classified Entities and Relations from Biomedical Texts)²¹⁸ <http://mips.helmholtz-muenchen.de/excerbt/>. Co-occurrence search was employed in order to retrieve all the molecular species associated with Ctgf. Thereafter, false positives (e.g., due to the intrinsic ambiguity in most acronyms) were discarded by manual curation. By this, we compiled a list of 274 unique interactions (since in some cases controversial results were reported and/or more than one source yielded the association, the total number of interactions was 548), involving 260 interactors, including genes/proteins, microRNAs, pathways, as well as some drugs and chemicals (Table S6). When searching the *Pathway Commons database* <http://www.pathwaycommons.org>¹³⁸ *for PPIs within the interactome of Ctgf, the number of interactions in the network increased to 1742 interactions*, involving 260 interactors, including genes/proteins, miRNAs, pathways, as well as some drugs and chemicals (Table S7 and Figure 3.15(A)). The resulting network will be further referred to as the Ctgf interactome.

Gene set enrichment analysis of the Ctgf interactome

Protein-protein interaction (PPI) information is frequently used as a starting point for the *functional annotations of unknown proteins according to the principle of 'guilty by association'*.²⁸⁸ Similarly, here, in order to get the first insight about the functionality of Ctgf, we performed gene set enrichment analysis of its interactome. Again, we used ToppFun from the ToppGene Suite <http://toppgene.cchmc.org>.²¹⁷ A complete list of ToppFun results is given in Table S8. Briefly, this analysis revealed that Ctgf interactome was mainly associated with biological functions such as cell proliferation, activation, migration and adhesion, as well as programmed cell death (see 'GO: Biological Process' in Table S8).

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

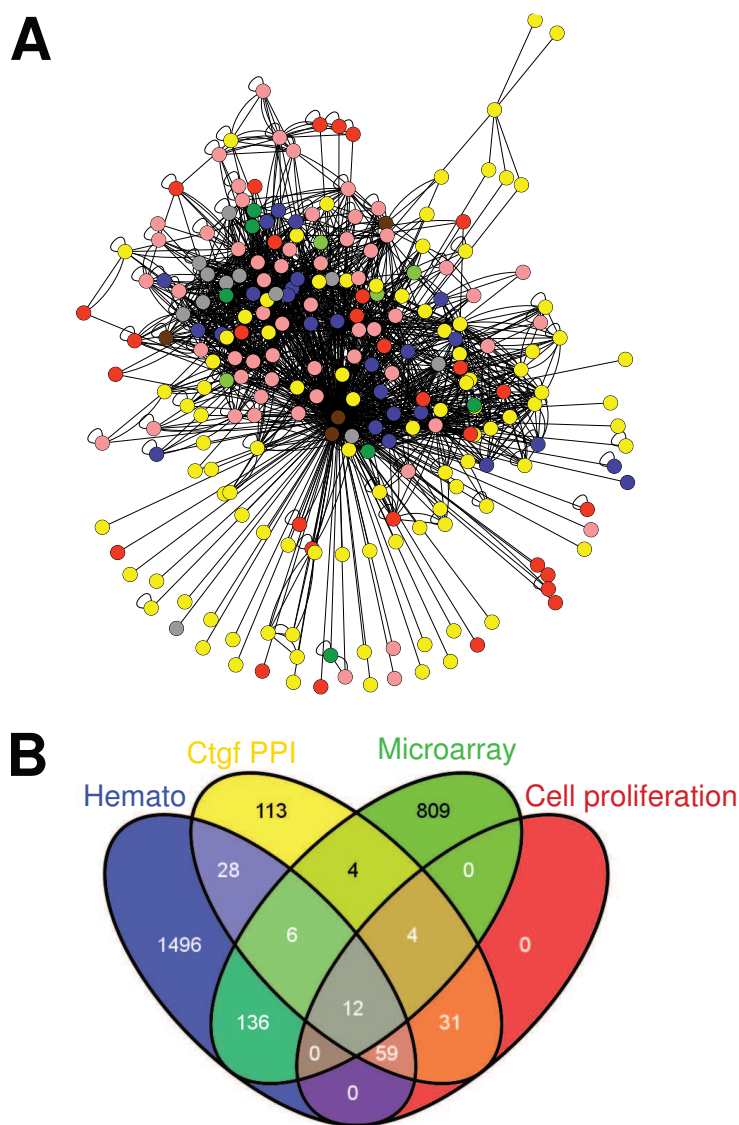


Figure 3.15: From top-down to bottom-up approach to *Ctgf* network modeling. (A) *Ctgf* interactome obtained after an extensive literature search using the text-mining tool EXCERBT followed by a collection of PPIs within the *Ctgf* interactome from the Pathway Commons database. The network diagram was created using Cytoscape²¹⁹ <http://www.cytoscape.org/>. (B) Compiling a list of seed genes from the *Ctgf* interactome (in yellow) by searching for genes: (1) associated with HSC maintenance or hematopoiesis (in blue); (2) genes involved in cell proliferation (Table S8, GO:BP, GO:0008283-cell proliferation p-Value: 2.9e-61, 106 PPIs) (in red); (3) genes present in our differentially expressed gene set (freshly isolated LSK cells vs. 24 h co-culture-derived ones) (in green). The Venn diagram of the corresponding gene sets and their overlaps was generated with the program VENNY <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.²⁶⁷

3 Results

At the molecular pathway level (see 'Pathway' in Table S8), Ctgf and its interaction partners were associated with integrin signaling, ***canonical Wnt and Tgf- β pathways, as well as with the G1/S check point of the cell cycle***. Intriguingly, the analysis yielded also a number of both steady-state and malignant hematopoiesis-related activities, including commitment to the B- and T-lineages, and acute/chronic myeloid leukemia. Moreover, the Hedgehog (HH) pathway, angiotensin/Tie2 signaling, Ifn- γ pathway, as well as osteopontin- and Il6-mediated signaling events have been linked with the self-renewal and maintenance of HSCs.^{7,48,67,289,290} Remarkably, beside the well-described functions of Ctgf such as involvement in wound healing,²⁹¹ angiogenesis²⁹² and bone formation,²⁹³ mouse phenotype enrichment analysis (see 'Mouse phenotype' in Table S8) also revealed a number of hematopoietic phenotypes, including abnormal hematopoietic cell number, morphology and physiology. In addition, abnormalities related to the cell cycle were also over-represented within the Ctgf network.

Topological property analysis and module discovery in the Ctgf interactome

Topological network analysis is motivated by the evidence that the ***topology (architecture) of biological interaction networks is closely related to their function, thereby contributing to better understanding of organization, network-wide interdependencies, causal relationships, and key aspects of network functionality***.^{169,294}

First, several topological parameters of the Ctgf interactome were computed using NetworkAnalyzer.²²¹ The results of this analysis are summarized in Table 3.6. The network was analyzed as an undirected network, and compared to the Erdős-Rényi random graph model²⁹⁵ with the same number of nodes and edges, generated using R package `igraph`.²⁹⁶

Topological property analysis NetworkAnalyzer recognized 236 nodes (interactors) and 1580 edges (interactions). The Clustering coefficient, which measures the degree to which the neighbors of a particular node are connected to each other was calculated to be 0.387 meaning that ***~39% of the possible connections between neighbors existed*** as compared to 0.057 for the random network. Generally, it is thought that the closer the local clustering coefficient is to one, the more likely it is for the network to form clusters. In average, biological networks tend to have a significantly higher average clustering coefficient compared to random networks, indicating their modular nature where cellular processes are governed by subsets of biomolecules that form an interaction module.¹⁴⁴

As it can be seen from the Table 3.6, the Ctgf interactome graph was itself connected and thus had exactly ***one connected component, consisting of the whole graph***. The network diameter or the largest distance between two nodes was five and seven for the undirected and directed network, respectively. Apparently, the topology of the Ctgf

Table 3.6: *Ctgf* interactome topology analysis using NetworkAnalyzer.²²¹

| | Topological parameter | <i>Ctgf</i> interactome | ER random graph |
|----|----------------------------|-------------------------|-----------------|
| 1 | Clustering coefficient | 0.387 | 0.057 |
| 2 | Connected components | 1 | 1 |
| 3 | Network diameter | 5 | 4 |
| 4 | Network radius | 3 | 3 |
| 5 | Network centralization | 0.939 | 0.055 |
| 6 | Shortest paths | 55460 (100%) | 55460 (100%) |
| 7 | Characteristic path length | 2.038 | 2.394 |
| 8 | Avg. # of neighbours | 8.229 | 13.254 |
| 9 | # of nodes | 236 | 236 |
| 10 | # of edges | 1580 | 1580 |
| 11 | Network density | 0.035 | 0.056 |
| 12 | Network heterogeneity | 1.970 | 0.284 |
| 13 | Isolated nodes | 1 | 0 |
| 14 | # of self-loops | 141 | 16 |
| 15 | Multi-edge node pairs | 609 | 0 |

3 Results

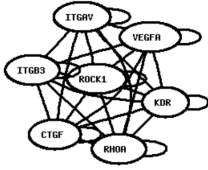
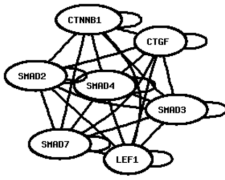
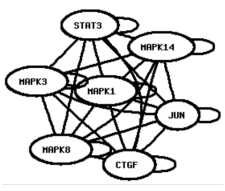
interactome *resembled a star* with centralization close to one (0.939), which could be explained by the fact that the network was built by first doing an extensive literature search to find all Ctgf PPIs, which were then themselves connected by searching for PPIs within the public databases. For the random graph, this number was 0.055. The network heterogeneity or the tendency of the network to contain hub nodes was 1.97 and 0.284 for the Ctgf interactome and the random graph, respectively. The characteristic path length, which offers a measure of a network's overall navigability²⁹⁷ was 2.038 and 2.394, respectively, meaning that given two nodes, two to three links would be needed to pass through to travel between the two nodes. In the Ctgf interactome network, each node had ~ 8.229 neighbors, however, was rather sparsely populated with edges, since its density was only 0.035 (for comparison, the density of a clique would be one). Hence, these parameters were lower as for the random graph (13.254 and 0.056, respectively). Finally, in 609 cases neighboring nodes were linked by more than one edge within the Ctgf network. This was not true for the random graph.

Module discovery GraphWeb <http://biit.cs.ut.ee/graphweb/>²²² was used for module discovery within the Ctgf interactome.

The maximal cliques algorithm First, network module detection was performed using the maximal cliques algorithm, that *finds groups of four or more nodes where each node is connected to every other node*. Cliques in PPI networks have often been *related to protein complexes and common functions*.²²² As a result, 326 modules were found, from which top three are shown in Table 3.7. *Each module consisted of seven nodes and 21 edges and was 100% connected. Functional annotation of the modules revealed key aspects of Ctgf molecular roles within a cell.* Module # one mostly contained genes, whose products are associated with *cell-matrix adhesions* (KEGG, map04510 Focal adhesion, p-Value = 1.84e-08), including integrins Integrin alpha-V (Itgav, Cd51) and Integrin beta-3 (Itgb3, Cd61), Transforming protein RhoA (Rhoa) and Rho-associated protein kinase 1 (Rock1). Module # two contained TGF- β (REACTOME, *Signaling by TGF- β* , p-Value = 1.79e-06) pathway genes, such as Mothers against decapentaplegic homologs (SMADs) 2, 3, 4 and 7, as well as *Wnt pathway* (KEGG, Wnt signaling, p-Value = 9.57e-07) transcription factors Catenin beta-1 (Ctnnb1) and Lymphoid enhancer-binding factor 1 (Lef1). Moreover, this module is also associated with regulation of *cell cycle* (GO:BP, p-Value = 8.79e-03). Module # three mostly contains several *MAPK pathway kinases* (GO:MF, MAP kinase activity, p-Value = 2.51e-009): Extracellular signal-regulated kinases ERK1/2 (Mapk1, Mapk3), Mitogen-activated protein kinase p38 alpha (Mapk14) and Stress-activated protein kinase JNK1 (Mapk8).

3.8 Network model predicting the role of Ctgf in regulating the LSK cell-cycle status

Table 3.7: Gene module detection within Ctgf interactome using the Maximal cliques algorithm provided by GraphWeb <http://biit.cs.ut.ee/graphweb/>.²²² The maximal cliques algorithm finds groups of 4 or more nodes where each node is connected to every other node. For Ctgf interactome, 326 such modules were found, first three of which are shown below, including their functional scores based on g:Profiler annotations, as well as the annotations contributing to the score and their statistical significance. Each module consisted of 7 nodes and 21 edges and was 100% connected.

| Module | Score | g:Profiler annotation | p-Value |
|---|-------|--|----------------------------------|
| # 1  | 110 | GO:BP-cell migration GO:MF-VEGF receptor binding KEGG-Focal adhesion | 2.59e-08 6.41e-05 1.84e-08 |
| # 2  | 183 | GO:BP-regulation of cell cycle KEGG-Wnt signaling REACTOME-Signaling by TGF- β | 8.79e-03 9.57e-07 1.79e-06 |
| # 3  | 110 | GO:MF-MAP kinase activity | 2.51e-009 |

The hub-based algorithm Alternatively, the hub-based algorithm provided by GraphWeb was also utilized in order to *detect groups of genes consisting of a central hub (a node with many connections) and related genes/proteins* within distance d , which was set to one in this case. GraphWeb found 248 such modules, the top nine of them (re-sorted by the functional Score based on g:Profiler annotation) are provided in Table 3.8. The highest scoring module (Score: 190) consisted of *a hub gene Signal transducer and activator of transcription 1 (Stat1)* and 19 of its connections and is associated with regulation of signal transduction (GO:BP, p-Value: $8.94e-17$) and apoptosis (REACTOME, p-Value: < 0.04). The second module is centered around *Nuclear factor NF-kappa-B p105 subunit* (Nfkb1, Score: 170, 20 Nodes) and associated with signaling by NGF (REACTOME, p-Value: $8.59e-06$). Other modules involved in the same pathway include: Signal transducer and activator of transcription 3 (Stat3, Score: 139, Nodes: 28, p-Value: $3.51e-06$) and RAC-alpha serine/threonine-protein kinase (Akt1/PKB, Score: 137, Nodes: 30, p-Value: $5.27e-12$; Table 3.8). This could be explained by the fact that Ctgf interacts with TrkA and $p75^{NTR}$, two receptors that are known to be activated by the neurotrophin nerve growth factor (NGF).²⁹⁸ The top three module involves Catenin beta-1 (Ctnnb1, Score: 161) and 38 of its PPIs, and its functionality is associated with Wnt (KEGG, p-Value: $6.27e-13$) and TGF- β (KEGG, p-Value: $2.37e-40$) signaling pathways, as well as the Cell cycle (KEGG, p-Value: $2.13e-30$) and cell proliferation (GO:BP, p-Value: $1.1e-30$). TGF- β signaling pathway is also associated with several other hubs from the Ctgf interactome, including the SMAD family of signal transducer proteins: Smad2 (Score: 161, Nodes: 23, p-Value: $4.97e-07$)(Table 3.8), Smad3 (Score: 135, Nodes: 28, p-Value: $8.31e-09$) and Smad4 (Score: 131, Nodes: 26, p-Value: $6.39e-07$), as well as Stress-activated protein kinase JNK1 (Mapk8, Score: 121, Nodes: 28, p-Value: < 0.0006), Transcription factor Sp1 (Score: 104, Nodes: 22, Patway: REACTOME-Phospho- R-SMAD forms a complex with CO-SMAD, p-Value: $3.71e-05$), as well as Transforming growth factor beta-1 itself (Tgfb1, Score: 95, Nodes: 18, p-Value: $1.79e-10$). Interestingly, several modules (Akt1/PKB, MAP kinase p38 α /Mapk14, Smad3, Glycogen synthase kinase-3 beta/Gsk3b, ERK1/2, as well as Ctgf itself), seem to involve *Zinc finger protein 161 (Zfp161/ZF5) transcription factor binding motif*. ZF5 is thought to be a putative murine repressor for Myc, with a growth-inhibitory function, also expressed at a very high level in human CD34+ cells.²⁹⁹ Direct association between Ctgf and Zfp161/ZF5 could not be found in the literature, however, Ctgf was reported to be up-regulated in kidneys of Glis2 mutant mice. Gli-similar 1 (Glis1) through Glis3 form a subfamily of Krüppel-like zinc finger proteins that share a highly conserved tandem repeat of five C2H2-type zinc finger (ZF1 to ZF5) motifs.³⁰⁰

Table 3.8: Gene module detection within *Ctgf* interactome using the Hub-based algorithm provided by GraphWebhttp://biit.cs.ut.ee/graphweb/.²²² Hub-based module detection algorithm returns groups of genes consisting of a central hub (a node with many connections) and related genes and proteins within distance *d*. GraphWeb extracts a list of hub-based modules ranked by the central hub degree (number of connections). For *Ctgf* interactome, 248 such modules were found, largest 25 of which are shown below, including their functional scores based on g:Profiler annotations (sorting criterion), as well as the annotations contributing to the score and their statistical significance.

| # | Module | #Nodes | #Edges | Density | Score | g:Profiler annotation | p-Value |
|---|---------------|--------|--------|---------|-------|--|--|
| 1 | # 24 (Stat1) | 20 | 69 | 20 | 190 | GO:BP-regulation of signal transduction REACTOME-Apoptosis | 8.94e-17 0.0361 |
| 2 | # 25 (Nfkb1) | 20 | 68 | 17 | 170 | KEGG-NOD-like receptor signaling REACTOME-Signaling by NGF | 3.65e-09 8.59e-06 |
| 3 | # 3 (Ctnnb1) | 39 | 152 | 2 | 161 | GO:BP-cell proliferation KEGG-TGF- β signaling pathway KEGG-Cell cycle | 1.1e-30 2.37e-40 2.13e-30 |
| 4 | # 17 (Smad2) | 23 | 98 | 24 | 142 | KEGG-Wnt signaling pathway MIRBASE-MI:mmu-miR-363 REACTOME-Signaling by TGF- β | 6.27e-13 0.0048 4.97e-07 |
| 5 | # 13 (Itgav) | 26 | 97 | 5 | 140 | GO:BP-cell adhesion GO:CC-extracellular matrix GO:MF-integrin binding KEGG-Focal adhesion | 6.49e-20 7.29e-09 8.01e-19 4.44e-20 |
| 6 | # 6 (Stat3) | 28 | 117 | 7 | 139 | REACTOME-Signaling by NGF | 3.51e-06 |
| 7 | # 4 (Akt1) | 30 | 128 | 4 | 137 | REACTOME-Signaling by NGF TRANSFAC-Factor: ZF5; motif: NRNGNGCGGCWN | 5.27e-12 0.000117 |
| 8 | # 16 (Mapk14) | 24 | 100 | 19 | 136 | REACTOME-Toll Receptor Cascades TRANSFAC-Factor: ZF5; motif: NRNGNGCGGCWN | 3.28e-05 0.000167 |
| 9 | # 12 (Jun) | 26 | 117 | 18 | 135 | GO:CC-TF complex REACTOME-Phospho-R-SMAD forms a complex with CO-SMAD | 8.64e-08 3.44e-05 |

3.8.2 The bottom-up or “seed-gene” approach to Ctgf network modeling

The topological analysis above provides first insights regarding the organization and key aspects of possible networks operated by Ctgf, as well as their network functionality.^{169,294} However, such *large-scale interactomes* are clearly not directly interpretable and sufficient by themselves, as they *do not provide information on the logic of signaling networks and their spatio-temporal behavior*. Moreover, we still lack technology that would allow high throughput detailed measurement of activity of individual signaling molecules and their interactions. *Computational modeling can aid in simulating the dynamical input/output behavior of a network*, allowing the formulation of a systems-level hypothesis that can be a powerful source in *directing targeted experiments*.^{123,124,178,301} However, the size of a constructed network should be limited, since computationally and mathematically, it is more feasible to model and simulate a network with a *small number of genes (up to 30 species)*. In addition, it is more likely that a small set of genes maintains a specific core regulatory mechanism. Finally, such modeling requires a certain amount of mechanistic detail, which may not be available for larger networks.^{119,122,124–130,179,302} The construction of such models can be better approached in a bottom-up directionality, where *a small number of “seed-genes” are first extracted from within the experimental data and then used to grow the network in several ways*.³⁰³

Here, we first selected a list of “seed genes” from the reference network, according to the following criteria: (i) *hematopoiesis-associated* genes (Table S5); (ii) genes *involved in cell proliferation* (Table S8, GO:BP, GO:0008283-cell proliferation p: 2.9e-61, 106 PPIs); (iii) genes *differentially expressed in LSK cells after the 24 h co-culture* (Day1; d1; see Table S1). Figure 3.15(B) shows a Venn diagram of the corresponding gene sets and their overlaps, while (A) uses the same color code within the Ctgf interactome to illustrate these overlaps graphically. As a result, we obtained 12 genes which satisfied all three criteria (Table 3.9). Seven genes were also part of our limma validation set (Table 3.1, Figure 3.5(A) and Figure 3.6(A)). *From these, we further focused on a subset including: Ctgf, Cyclin D1 (Ccdn1), p21Cip1 (Cdkn1a), FoxO1 Foxo1, Lef (Lef1) and Integrin β -3 Itgb3*. Next, a strategy similar to that described by Hwang and co-workers³⁰⁴ was adopted. Given the seed, we first identified the starting and ending genes/ proteins. According to the definition of signaling pathways by which cells convert extracellular signals into cellular responses, start proteins are defined as ligands or transmembrane receptors; end proteins are transcription factors or some molecular players whose roles are clearly known in cells. In addition, since signals are transmitted from the extracellular space into the nucleus, the right

Table 3.9: Compiling a list of seed genes from the *Ctgf* interactome (Figure 3.15 in yellow) by searching for genes: (1) associated with HSC maintenance or hematopoiesis in general (Figure 3.15 in blue, see Table S5), (2) genes involved in cell proliferation as reported by ToppFun Gene Ontology enrichment analysis of the *Ctgf* interactome (Figure 3.15 in red, see Table S8: GO:BP, GO:0008283-cell proliferation p: 2.9e-61, 106 PPIs); (3) genes present in our differentially expressed gene set in LSK cells after 24 h co-culture (Figure 3.15 in green, Day1; d1, see Table S1). As a result, a "seed" list of 12 genes satisfying all the above mentioned criteria was obtained. For genes, which were also part of our LIMMA²¹³ validation set (Table 3.1, Figure 3.5(A) and Figure 3.6(A)), an arrow indicating the direction of the fold change is added.

| # | Gene | RT-qPCR |
|----|-----------------|---------|
| 1 | <i>Ccnd1</i> | ↑ |
| 2 | <i>Cdkn1a</i> | ↑ |
| 3 | <i>Foxo1</i> | ↑ |
| 4 | <i>Foxo3</i> | |
| 5 | <i>Lef1</i> | ↑ |
| 6 | <i>Stat1</i> | ↓ |
| 7 | <i>Aqp1</i> | |
| 8 | <i>Thbs1</i> | |
| 9 | <i>Serpine1</i> | |
| 10 | <i>Nfatc2</i> | |
| 11 | <i>Cebpb</i> | ↑ |
| 12 | <i>Itgb3</i> | ↑ |

3 Results

orders of factors can be defined: extracellular space \rightarrow plasma membrane \rightarrow cytoplasm \rightarrow nucleus.³⁰⁴ Following this strategy, our starting proteins were Connective tissue growth factor (Ctgf) and Integrin beta-3 (*Itgb3*), since their GO annotations according to SWISSPROT <http://www.uniprot.org/uniprot/> were “secreted” and “receptor activity”, respectively. Since the main regulatory activity in the maintenance of HSC lies in self-renewal, which is a special functional form of cell division, the cell cycle as our terminal node. According to this, we identify the two identified cell cycle regulators in the seed list: G1/S-specific cyclin-D1 (*Ccnd1*) and p21Cip1 (*Cdkn1a*) as terminal nodes. We also considered two transcription factors (TFs): Forkhead box protein O1 (*Foxo1*) and Lymphoid enhancer-binding factor 1 (*Lef1*), which could be possibly involved in the Ctgf-regulated cell cycle progression. The right order of the “seed genes” would be as follows: *Ctgf* \rightarrow *Itgb3* \rightarrow *Foxo1*, *Lef1* \rightarrow *Ccnd1*, *Cdkn1a*.

Literature-based network construction

Next, we performed manual literature search to identify the pathways and major molecular players relaying a signal from our start genes/proteins to the terminal nodes. Important to note, similarly as already described by,¹²⁷ we also consider here only local interactions (e.g., a kinase phosphorylates its substrate). At the same time, in order to keep the size of the network meaningful, parts of it were simplified: for example, the MAPK cascade, in which a series of nodes and edges impinge only on each other, was reduced to FAK \rightarrow Erk1/2.

The resulting network contained four inputs (Ctgf, Tgf β WNTs, and Dkk1) as external stimulus from the ‘niche’, 25 internal regulatory nodes, including membrane receptors (integrin α V β 3, Tgf- β RI/II and Frz/Lrp6), key signaling intermediates (e.g., kinases Gsk3- β , Akt(PKB) and Erk1/2), key transcription factors (SMAD complex, FoxO1 and TCF/LEF), two G1/S transition cyclins and their respective CDKs (Cyclin D1:Cdk4/6 and Cyclin E:Cdk2), as well as CDK inhibitors (p21Cip1 and p27Kip1) and G1/S transition molecules (Rb1 and E2f1). More importantly, the network not only linked molecular pathways regulated by Ctgf, but also inferred three functional outcomes with direct relevance to the description of HSC behavior:

- **G0/G1:** Cyclin D:Cdk4/6 activation is the readout for the G0-to-G1 transition or exit from the quiescence. When active Cyclin D:Cdk4/6 complex accumulates, Rb is activated through hypo-phosphorylation. Active Rb is able to bind E2f and keep a cell in the G1/G0 phase.
- **G1/S block:** cyclin-dependent kinase inhibitors (CKIs) p21Cip1 and p21Kip1 insure that the cell cycle remains arrested at G1/S until an external signal relieves the break. In this case, the activation of p21Cip1 and/or p21Kip1 serves as the readout.

3.8 Network model predicting the role of Ctgf in regulating the LSK cell-cycle status

- **i-Ctgf:** Ctgf can be induced via β -Catenin/Tcf/Lef signaling or TGF- β induces Ctgf by the "classical" Smad pathway. Here, the readout is the induction of intrinsic Ctgf expression.

The modified Edinburgh Pathway Notation (mEPN) scheme²²⁰ was utilized for the graphical depiction of signaling pathways, since it allows the detailed representation of a diverse range of biological entities, interactions and pathway concepts and enables to represent pathway knowledge in a semantically and visually unambiguous manner. A complete network diagram is shown in Figure 3.16. This network will be further referred as the Ctgf signaling network. Table S9 summarizes the information used to construct the network in the form of 32 unique nodes and 95 edges, giving the source (upstream regulator) node, the target (down-stream regulated) node, a qualifier of the nature of the relationship such as 'activates' or 'inhibits' and references where this relationship was reported. The biological description of the biological events occurring within the Ctgf signaling network is given below.

G0/G1; Ctgf promotes G0/G1 transition The current paradigm of G0/G1 cell-cycle progression argues that sufficient amounts of growth factor (GF) stimulation of G0 quiescent cells results in gradual accumulation of Cyclin D:Cdk4/6 complexes that trigger the hypophosphorylation and activation of pRb, promoting its assembly with E2Fs and chromatin remodeling proteins, such as histone deacetylase and SWI/SNF to repress E2F target gene expression.^{73,305-311} Thus, Cyclin D1 serves as a sensor of the mitogenic potential of the micro-environment and its activation seems to be the key event facilitating emergence from quiescence or the G0/G1 transition.³¹²⁻³¹⁴ Ctgf has been reported to induce cell proliferation by enhancing the expression of Cyclin D1 mRNA and protein level,^{273,315,316} and Ctgf shRNA significantly prevented Ctgf and Cyclin D1 expression, arrested cell cycle at G0/G1 phase and suppressed cell proliferation.³¹⁷ Due to this, Ctgf induced G0/G1 transition was attributed to its activating effect on Cyclin D1 expression and the biomedical literature was surveyed for molecular interactions that would lead to this activation. By this, we found that secreted Ctgf binds the integrin α V β 3,²⁷⁴ which leads to the phosphorylation of focal adhesion kinase (FAK)^{274,318} and activation of MAPK pathway resulting in activation of Erk1/2 via phosphorylation on Thr202/Tyr204.^{315,318} As described above, for the sake of simplicity, the activation of the MAPK cascade was compressed to FAK \rightarrow Erk1/2 in this network. Erk1/2 then increases transcription of c-Fos (AP1)³¹⁹ and c-Myc,³²⁰ the transcriptional target of which is Cyclin D1.^{321,322} Turnover of Cyclin D1 also depends on the phosphatidylinositol 3-kinase (PI3K)/Akt (protein kinase B) pathway, which negatively regulates the phosphorylation of Cyclin D1 on Thr286 by glycogen synthase kinase3- β (Gsk3- β).³²³ Ctgf binding to the integrin α V β 3 is known to activate PI3K/Akt(PKB) pathway,³²⁴

3 Results

which may lead to the inhibition of Gsk3- β via Akt(PKB)-dependent phosphorylation of Ser9³²⁵ and thus facilitate the nuclear accumulation of β -catenin. Alternatively, Ctgf requires the Lrp6 receptor to activate the canonical Wnt signaling,³²⁶ resulting in accumulation and nuclear translocation of β -catenin, which is coincident with decreased phosphorylation of β -catenin on Ser33/37 and increased phosphorylation on Tyr142. Thereafter, it engages transcription factors Tcf/Lef to activate expression of downstream genes, such as *Myc* and *Ccnd1*.³²⁷ In early G1, cyclin kinase complexes Cyclin D:Cdk4/6 accumulate and Rb is activated through hypo-phosphorylation (pRb). Active Rb is able to bind E2f and keep a cell in the G0/G1 phase.³²⁸

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

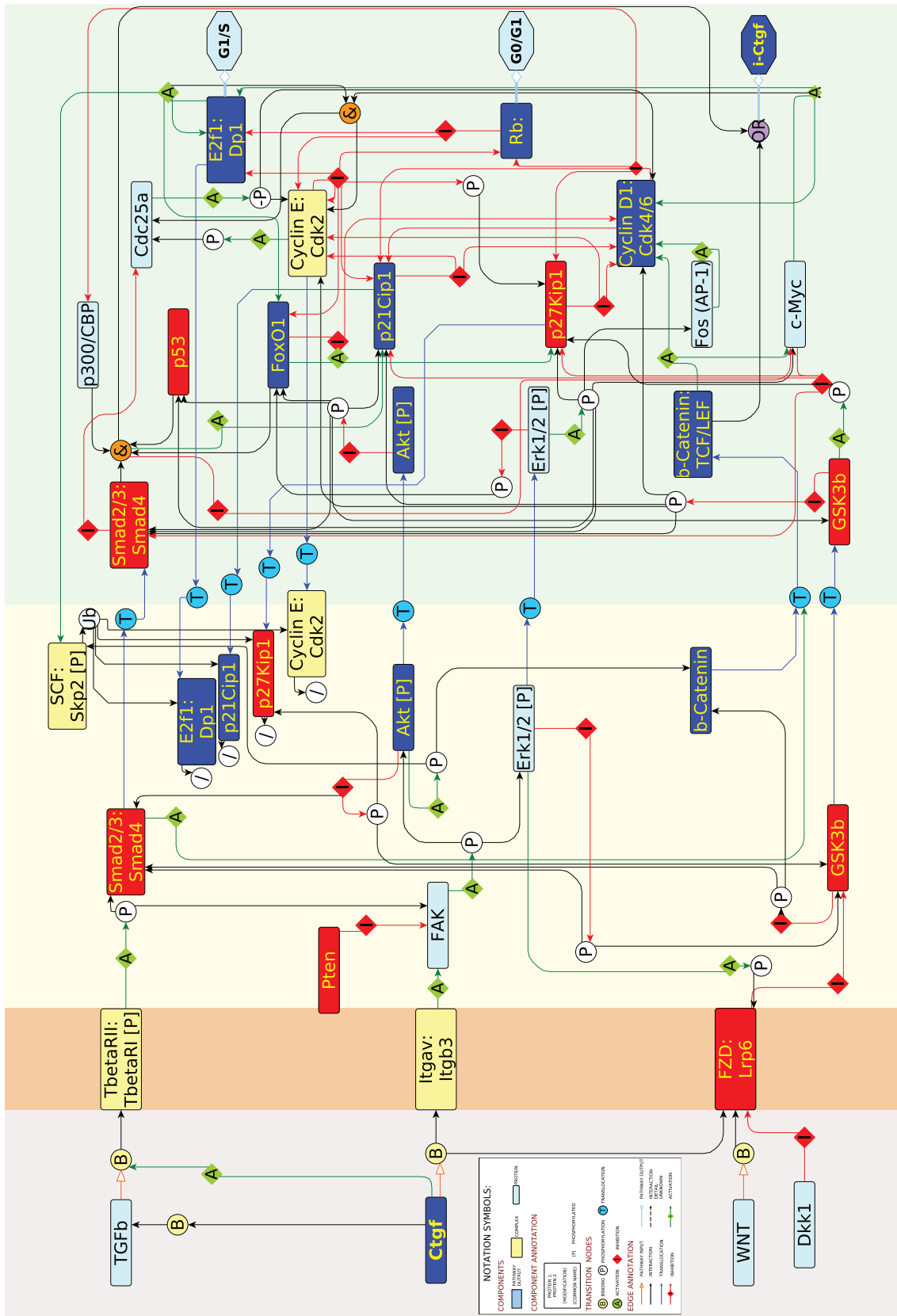


Figure 3.16

Figure 3.16: The graphical network map depicting Ctgf regulated G0-to-G1 transition, G1/S block, as well as its auto-induction derived from the biomedical literature. The "modified Edinburgh Pathway Notation" (mEPN) scheme²²⁰ <http://www.mepn-pathway.org/> was used to graphically visualize the constructed network. Ctgf, Tgf- β , WNTs and Dkk1 constitute the input and G0/G1, G1/S and i-Ctgf represent the outputs (functional readouts) of the network. The interaction information used to construct the network in the form of 32 unique nodes and 95 edges can be inferred from Table S9. The experimental validation results on mRNA (Figure 3.20) and/or protein levels or phosphorylation status (Figure 3.19) were superimposed on the network: nodes, for which a significant increase in mRNA/protein/phosphorylation level was observed after 24 h in Ctgf deficient environment are ON (in red), whereas nodes for which the opposite is true are OFF (in dark blue). Since most biological functions are mediated by proteins and changes in their post-translational modifications (e.g., phosphorylation), we focused more on the protein level data, where available.

G1/S block; Ctgf contributes to the G1/S block via inducing p21Cip1 and p27Kip1

Ctgf induced G1 phase arrest appears to be due to the induction of the cyclin-dependent kinase inhibitors (CKI) p15INK4, p21Cip1, and p27Kip1, which are known to bind and inactivate Cyclin D:Cdk4/6 and the Cyclin E:Cdk2 kinase complexes. To determine the possible role of Ctgf in G1/S block, it is good to briefly have a closer look at the mechanism of G1/S transition regulation. It is well known that active Cyclin D:Cdk4/6 complexes have a second, non-catalytic function, e.g. the sequestration of CKIs, including p21Cip1 and p27Kip1, which lowers their inhibitory threshold and facilitates activation of the Cyclin E:Cdk2 complex. Cyclin E:Cdk2 further antagonizes p21Cip1 by phosphorylating it and triggering its proteolysis. Similarly, Cyclin D- and E-dependent kinases also contribute sequentially to Rb hyper-phosphorylated on Ser-795, canceling its ability to repress E2f family members, which are being activated and can control the transcription of cell division essential genes (e.g., Cyclin E and Cdc25A and E2f). E2f1 can be inactivated by binding to dephosphorylated Rb and by degradation through ubiquitination after phosphorylation by Cyclin E:Cdk2. The activity of the Cyclin E:Cdk2 complex is modulated by transcriptional regulation of its catalytic partner Cyclin E and post-translational modification of the Cyclin E:Cdk2 complex itself, where the transcription of Cyclin E, in addition to E2f1, is also promoted by c-Myc, followed mitogenic stimulation. Post-translationally, another transcriptional target of c-Myc and E2f1- Cdc25A- dephosphorylates and activates the Cdk2 kinase. Interestingly, for Cdc25A to become active it must itself be phosphorylated, which is catalyzed by the active Cyclin E:Cdk2 complex. Finally, the inactivation of the Cyclin E:Cdk2 complex, as well as the Cdc25A phosphatase, is achieved by ubiquitin-mediated degradation.^{323,329-331}

In order for the G1/S arrest to be enforced, high levels of active CKIs p21Cip1 and

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

p27Kip1 have to be maintained, which can be achieved either by increased transcription, reduced degradation or through the release of p21Cip1 and p27Kip1 from complexes with Cyclin D:Cdk4/6.³²⁸ Since we observe elevated *Cdkn1a* mRNA levels in LSK cells after 24 h (Day1; d1) in our co-culture experiments (Table 3.1, Figure 3.5(A) and Figure 3.6(A)), it is tempting to hypothesize that *Ctgf* is responsible for the first one of these, namely, the increased transcription of *Cdkn1a* mRNA. Thus, we further focus our attention to the molecular mechanisms by which *Cdkn1a* transcription could be possibly induced. In this regard, probably the best known is the up-regulation of p21Cip1 by p53 following DNA damage,³³² and it has been shown that p21Cip1 transcription can be induced by *Ctgf* via a phospho-Ser15 p53-dependent mechanism.³¹⁵ Since Erk1/2 is able to phosphorylate p53 at Ser-15,³³³ this mechanism could be involved in the *Ctgf* mediated p53 phosphorylation. Furthermore, evidence exists that p300/CBP-mediated acetylation may be a universal and critical modification for p53 function.⁶⁹ In addition, Akt(PKB) can indirectly hinder p53-dependent p21Cip1 induction by phosphorylating Mdm2 (compressed to Akt(PKB) + p53) or it can directly phosphorylate p21Cip1 and restrict it to the cytoplasm for degradation.³⁴⁹

Alternatively, *Ctgf* can enhance Tgf β -Smad2/3 signaling by binding directly to the growth factor, promoting its interaction with the Tgf β receptor. Tgf- β has been reported to up-regulate *Cdkn1a* mRNA levels.³³⁴ Tgf- β binding to its transmembrane type II receptor (T β RI) initiates formation and activation of a heteromeric complex with the corresponding type I receptor (T β RI). Upon activation, T β RI initiates intracellular signaling by phosphorylation of the receptor-regulated SMAD (R-SMAD) proteins Smad2 and Smad3. Phosphorylated R-SMADs complex with the common mediator Smad4, translocate to the nucleus, and modulate target gene transcription together with numerous other factors.^{335,336} Also Smad2/3/4 complex requires the presence of the co-activator p300/CBP,^{337,338} which itself may be a direct transcriptional target of Tgf- β .^{339,340} SMAD proteins have also been reported to form a complex with FoxO proteins.³⁴¹ In particular, SMADs, together with FoxO and p53 form large transcriptional complexes on the p21Cip1 promoter enhancer.³⁴² These complexes, together with p300/CBP could also be responsible for the inhibition of c-Myc.³⁴³⁻³⁴⁷ Besides the "classical" TGF- β pathway described above, TGF- β may also induce p21Cip1 through mechanisms that involve Erk1/2³⁴². Also Erk1/2 has been reported to act inhibitory on FoxO transcription factors.³⁵⁰ *Ctgf* has been reported to activate Akt(PKB) and Erk1/2.³⁵¹ Furthermore, Tgf- β has also been described to rapidly activate PI3K, as indicated by the phosphorylation of its downstream effector Akt(PKB). This activation appeared to be independent of Smad2/3 activation.³⁵²

Finally, the cell cycle inhibitor p27Kip1 has been shown to be regulated by FoxO.³⁵³ Alternatively, Erk1/2 phosphorylates p27Kip1 at Ser178 increasing its expression.³⁵⁴ On

3 Results

the other hand, Akt(PKB) decreases p27Kip1 by increasing proteolysis and reducing its transcription.³⁵⁵ Cyclin E:Cdk2-dependent phosphorylation of p27Kip1 results in elimination of p27Kip1 from the cell, allowing cells to transit from G1 to S phase.³⁵⁶ In all cases, the nuclear ubiquitin ligase (E3) SCF(Skp2) is implicated in p27Kip1 degradation.³⁵⁷

iCtgf; Ctgf is able to auto-induce its own expression Ctgf has also been reported to auto-induce its own expression.²⁸⁷ A β -catenin:Tcf/Lef-binding site (TBE) was identified in the promoter region of Ctgf and it was found that Ctgf is a transcriptional target of β -catenin:Tcf/Lef signaling, thus the cross-talk between Ctgf and Wnt signaling seems to form a positive feedback loop.²⁷³ In addition, Tgf- β induces Ctgf by the “classical” SMAD pathway via a SMAD binding element located within its proximal promoter.²⁷⁵ Moreover, Ctgf is one of genes that, in addition to p15Ink4b and p21Cip1, are jointly induced by SMAD and FoxO proteins.³⁵⁸

Dynamic network analysis using Boolean mathematical formalism

As a next step, in order to understand the dynamics of the Ctgf signaling network, examine its qualitative stimulus-response behavior, as well as to verify the coherence of the literature-derived interaction graph, we translated it into a computational model by employing the Boolean logic. This mathematical framework was chosen due to its ***straightforwardness, robustness and compatibility with qualitative data***. The regulatory activity of genes/proteins is being simplified by considering them as all or none devices. More precisely, each gene/protein is defined as being either active (ON/value 1) or inactive (OFF/value 0) depending on its abundance or activity level.^{124,125,127,128,305} The Boolean logical framework has already been successfully applied in modeling many gene regulatory and signaling networks in a variety of organisms,^{122,185–187} and it has been shown to lead to predictive testable hypotheses also in poorly understood large-scale systems.^{183,184} For Boolean networks, the major task is the identification of attractors. ***Attractors are stable cycles of states in a Boolean network in which the network resides most of the time.***²²³ Attractors in models of gene-regulatory networks are expected to be ***linked to phenotypes cellular phenotypes.***^{177,223} Transitions from all states in a Boolean network eventually lead to an attractor, as the number of states in a network is finite. ***All states that lead to a certain attractor form its basin of attraction.***²²³ For more details see the 1. Introduction section, 1.5.4.1 Boolean mathematical formalism on page 19.

In this study, the basis of the Boolean model were the known molecular interactions and regulatory relationships from the published literature, as summarized in Table S9 and graphically visualized in Figure 3.16. These interactions between mRNAs/proteins, leading to the above described functional outcomes were then translated into logical (Boolean)

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

functions (Figure 3.1; step 8). To do so, each network node (mRNA/protein or one of the three functional outcomes “G0/G1”, “G1/S” or “i-Ctgf”) was described by one of the two possible states: active (ON/value 1) or inactive (OFF/value 0), representing the transcription and translation of a gene or the activation of a protein or process or the absence of a gene transcript or the inhibition of a protein or process, respectively. The state of each node depends on the states of its upstream regulators and the type of regulation (activation vs. inhibition). The effects of the combinations of interactions on the activity of each network species (node) was defined in terms of logical rules using the Boolean operators AND, OR and NOT, where OR (“|”) represents the combined effect of independent upstream regulators on a downstream node, whereas AND (“&”) indicates the conditional dependency of upstream regulators to achieve a downstream effect. NOT (“!”) represents the effect of inhibitory regulators and can be combined with activating regulations by using either OR or AND. Hence, we transferred the natural-language statements on gene/protein dependencies from literature (Table S9) and expressed them as Boolean rules shown in Table 3.10. According to these rules, for example, for the Smad2/3/4 protein complex to be active (ON/value 1), first of all the Tgf- β receptors I and II (TgfbRI/II) have to be active (ON/value 1), and, at the same time none of its negative regulators (Gsk3- β , Akt(PKB), Erk1/2 or Myc) is allowed to be active (ON/value 1), in other words, they all have to be inactive (OFF/value 0). In case, Gsk3- β or Akt(PKB), or Erk1/2, or Myc would be active (ON/value 1), the state of Smad2/3/4 would be inactive (OFF/value 0).

Moreover, as it also can be seen in Table 3.10, we decided to simulate the three processes separately: (1) *Ctgf* contribution to G0/G1 transition via inducing Cyclin D1, (2) *Ctgf* involvement in the G1/S cell cycle arrest due to the activation of p21Cip1 and/or p27Kip1 cell cycle inhibitors, as well as (3) *Ctgf* induced auto-expression in LSK cells via a positive feedback loop. It has already been recognized earlier³³⁰ that a logical approach is to model the individual cell cycle components (e.g., the G1/S transition) separately before linking them together. In addition, it has been reported³¹⁵ that *Ctgf* stimulates the cells to actively enter the G1 phase from G0, however, they do not then progress further through the cell cycle, which appears to be due to the induction of the cyclin-dependent kinase inhibitors p15INK4, p21Cip1 and p27Kip1. Apart from this, *Ctgf* has also been reported to auto-induce its own expression²⁸⁷, and since we see an increase in *Ctgf* mRNA and protein levels (Figure 3.12(A)(a,c,e)) after 24 h co-culture with UG26-1B6 stromal cells, it was modeled here as a separate sub-network.

The Boolean models of *Ctgf* signaling were defined and analyzed using the R package **BoolNet**.²²³ Since the steady-state attractors are the same in asynchronous and synchronous networks,¹¹⁹ we used the synchronous updating scheme and performed an exhaustive attractor

Table 3.10: Boolean rules underlying the definition of the logical parameters describing Ctgf promoted G0-to-G1 transition, G1/S block and auto-induction. The names of the components of the regulatory graph of Figure 3.16 are listed in the first column. The description is based on the logical formalism, where ”&“ stands for ”AND“, ”|“ denotes ”OR“, and the negation ”NOT“ is written as ”!“. The rules were derived from the references listed in Table S9.

| Process | Product | Boolean rules |
|------------|--------------|--|
| | eCtgf | eCtgf |
| | WNTs | WNTs |
| | TGFb | TGFb & eCtgf |
| | aVb3 | eCtgf |
| | FZD_Lrp6 | ! Dkk1 & ((WNTs & eCtgf) Erk1_2) |
| | Dkk1 | Dkk1 |
| | TgfbRII | TGFb & eCtgf |
| | Smad2_3_4 | TgfbRII & ! (GSK3beta Akt Erk1_2 Myc) |
| | Pten | Pten |
| | FAK | (aVb3 TgfbRII) & ! Pten |
| | beta.Catenin | ! GSK3beta (! GSK3beta & (Akt & Smad2_3_4)) |
| | GSK3beta | ! (FZD_Lrp6 Akt Erk1_2) |
| | Akt | FAK |
| | Erk1_2 | FAK |
| | Fos | Erk1_2 |
| G0/G1 | Foxo1 | ! (Akt & Erk1_2) |
| | CycD_Cdk4_6 | (Myc TCF_LEF Fos) & ! (GSK3beta Foxo1) |
| | Rb | CycD_Cdk4_6 |
| | G0_G1 | Rb |
| G1/S block | Foxo1 | ! (Akt Erk1_2 CycE_Cdk2) |
| | p53 | (Erk1_2 p300) & ! Akt |
| | p300 | Smad2_3_4 & ! CycD_Cdk4_6 |
| | TCF_LEF | beta.Catenin |
| | Myc | (Erk1_2 TCF_LEF) & ! (Smad2_3_4 & p300 & Foxo1 & p53) |
| | p21Cip1 | (Smad2_3_4 & p300 & Foxo1 & p53) & ! (CycD_Cdk4_6 CycE_Cdk2 SCF_Skp2 Akt Myc GSK3beta) |
| | p27Kip1 | (Foxo1 Erk1_2 GSK3beta) & ! (CycD_Cdk4_6 Myc (CycE_Cdk2 & SCF_Skp2) (Akt & SCF_Skp2)) |
| | Cdc25a | (CycE_Cdk2 (E2f1_Dp1 & Myc)) & ! Smad2_3_4 |
| | SCF_Skp2 | E2f1_Dp1 Akt |
| | CycE_Cdk2 | ((E2f1_Dp1 & Myc) Cdc25a) & ! (p21Cip1 p27Kip1 (SCF_Skp2 & CycE_Cdk2 & GSK3beta) RB) |
| | CycD_Cdk4_6 | (Myc TCF_LEF Cdc25a Fos) & ! (Foxo1 p21Cip1 p27Kip1 GSK3beta) |
| | RB | (p21Cip1 p27Kip1) & ! (CycD_Cdk4_6 & CycE_Cdk2) |
| | E2f1_Dp1 | (Myc E2f1_Dp1) & ! (RB (RB & CycE_Cdk2 & SCF_Skp2)) |
| | G1_S_block | RB & ! E2f1_Dp1 |
| i-Ctgf | Foxo1 | ! (Akt & Erk1_2) |
| | p300_CBP | Smad2_3_4 |
| | iCtgf | (Smad2_3_4 & p300_CBP & Foxo1) (beta.Catenin & TCF_LEF) |

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

search in all cases. Importantly to note, since we do not know the states of the network entities *a priori*, we used an “exhaustive” attractor search here, meaning that the starting state can be either ON or OFF for each network node. Of note, “exhaustive” attractors are identified by exhaustive search of all 2^n states, where n is the number of nodes that are not set to a fixed value. Although, we restricted the number of nodes in our models to 29, which was the maximum number of nodes allowed for exhaustive search,²²³ for the G1/S sub-model, R could not allocate memory for the state tables. Since *Ctgf*, WNTs and *Tgf- β* are external factors constantly being secreted by our stromal cells (e.g., *Wnt5a*³⁵⁹ and *Tgfb1* in Figure 3.7(A)), we assumed here that these are always present, and fixed the states of these factors at ON (or in the state “1”) in the wild type. In addition, for the G0/G1 and i-*Ctgf* sub-models, we also did an “exhaustive” attractor search without fixing the state of any of the nodes (see Figure 3.18). In all cases, for loss-of-function simulations, we performed an *in silico* knock-out of extrinsic *Ctgf* by fixing its level to “0”, meaning that the corresponding protein was always inactive or OFF. For detailed description see 2.2.2.8 *Dynamical network analyses using Boolean logic* on page 43 in the 2. Materials, methods and data section.

G0/G1; *Ctgf* promotes G0-G1 transition In the wild type situation (WT HSC cultured on WT stromal cells), the G0-G1 transition model generated four simple attractors, each consisting of one state and having a basin of 65536 states (Figure 3.17, top left panel). Fifty percent of the attractors represented the activation of Cyclin D1 and, hence, the G0/G1 phenotype. According to the Boolean model (Table 3.10), the activation of Cyclin D1 results from the activity of the transcription factors β -catenin-TCF/LEF, c-Myc and Fos, and repression of Gsk3- β and FoxO1. Strikingly, in the second stable state, where the Wnt inhibitor Dkk1 stabilized at ON, leading to an inactive state of the FZD/Lrp6 receptor complex, the simulation still yielded the same G0/G1 output. This can be explained by the OFF state of Pten, resulting in active FAK, Akt(PKB) and/or Erk1/2. In such a case, on one hand, Erk1/2 can induce Fos (AP1), a positive regulator of Cyclin D1, and, on the other hand, Akt (PKB) can inhibit the activity of Gsk3- β , leading to the activation of β -catenin-TCF/LEF and c-Myc. Such an observation would suggest a possible cross-talk between the Wnt and Integrin/FAK(Focal adhesion kinase)/MAPK/PI3K/Akt (PKB) pathways. In addition, there were also two stable states, where no activation of Cyclin D1 could be observed. In both cases, this seemed to result from the activity of Pten, inhibiting the Akt (PKB) and Erk1/2 kinases. As a consequence, Foxo1 was stabilized at ON and would inhibit Cyclin D1. In the fourth attractor, the WNT inhibitor Dkk1 was active and Gsk3- β stabilized at ON, providing an additional inhibition to Cyclin D1. Accordingly, the absence of *Ctgf* did not result in activation of Cyclin D1/Cdk4/6, due to the activity of FoxO1 and Gsk3- β , two downstream targets of Akt (PKB) in this model. In fact, fixing its state at ON was sufficient to restore

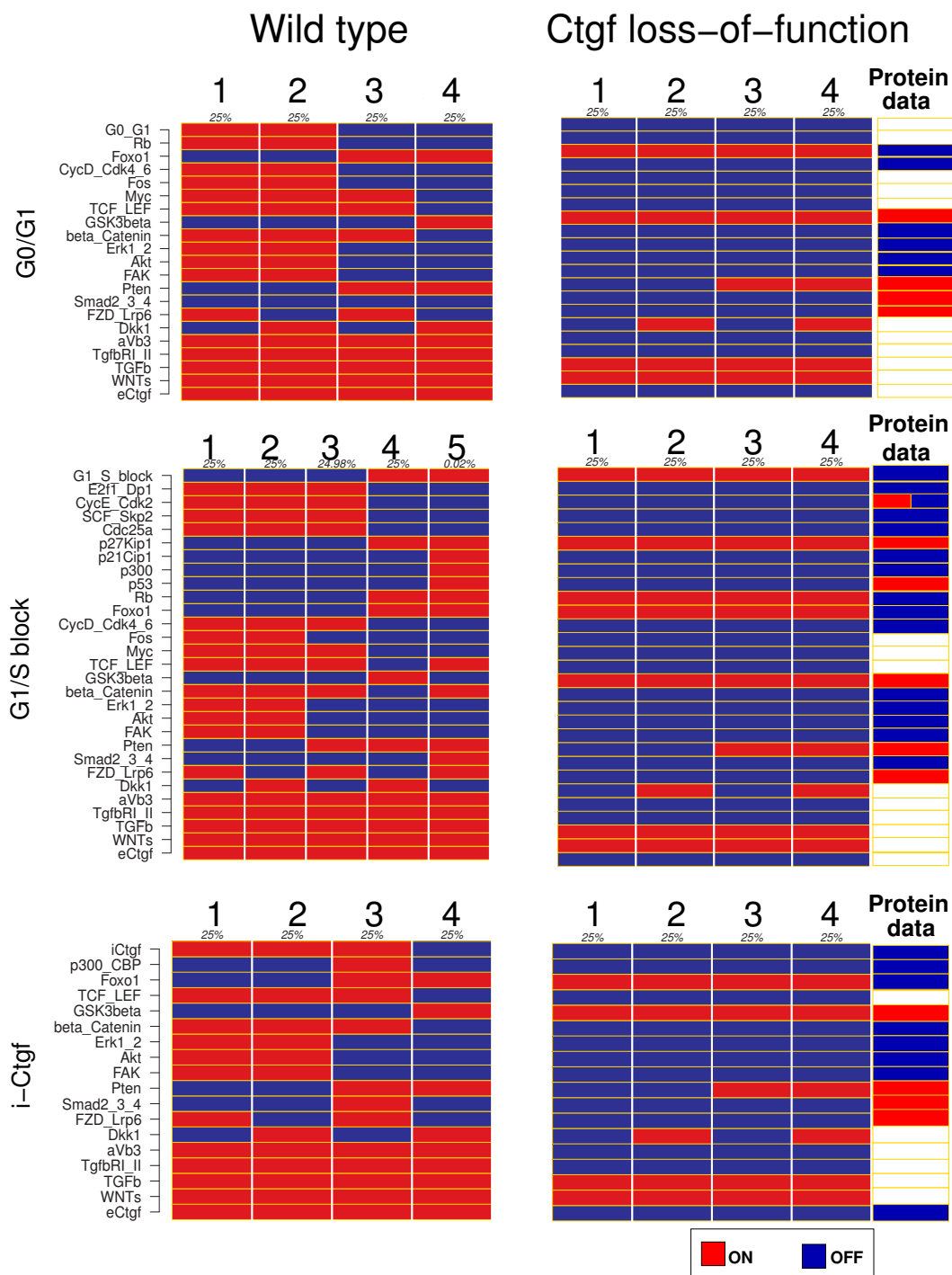


Figure 3.17: Attractor (“stable state“) analysis of the Boolean networks modeling Ctgf regulated G0-to-G1 transition, G1/S block, as well as its auto-induction. Using the synchronous updating strategy and Boolean logical rules listed in Table 3.10 an exhaustive attractor search was performed for either the wild type (left panel) or Ctgf loss-of-function (right panel) case. For the Ctgf loss-of-function case, a comparison (see “Protein data“) of the prediction results (attractors 1-4) to discretized (binarized) values (see Table S10 for details) of the experimental validation results of protein levels or phosphorylation status, were available (Figure 3.19). Otherwise, a white rectangle is left in the column “Protein data“. An exception to this was E2f1_Dp1, where we used the mRNA levels (Figure 3.20; *E2f1*) instead.

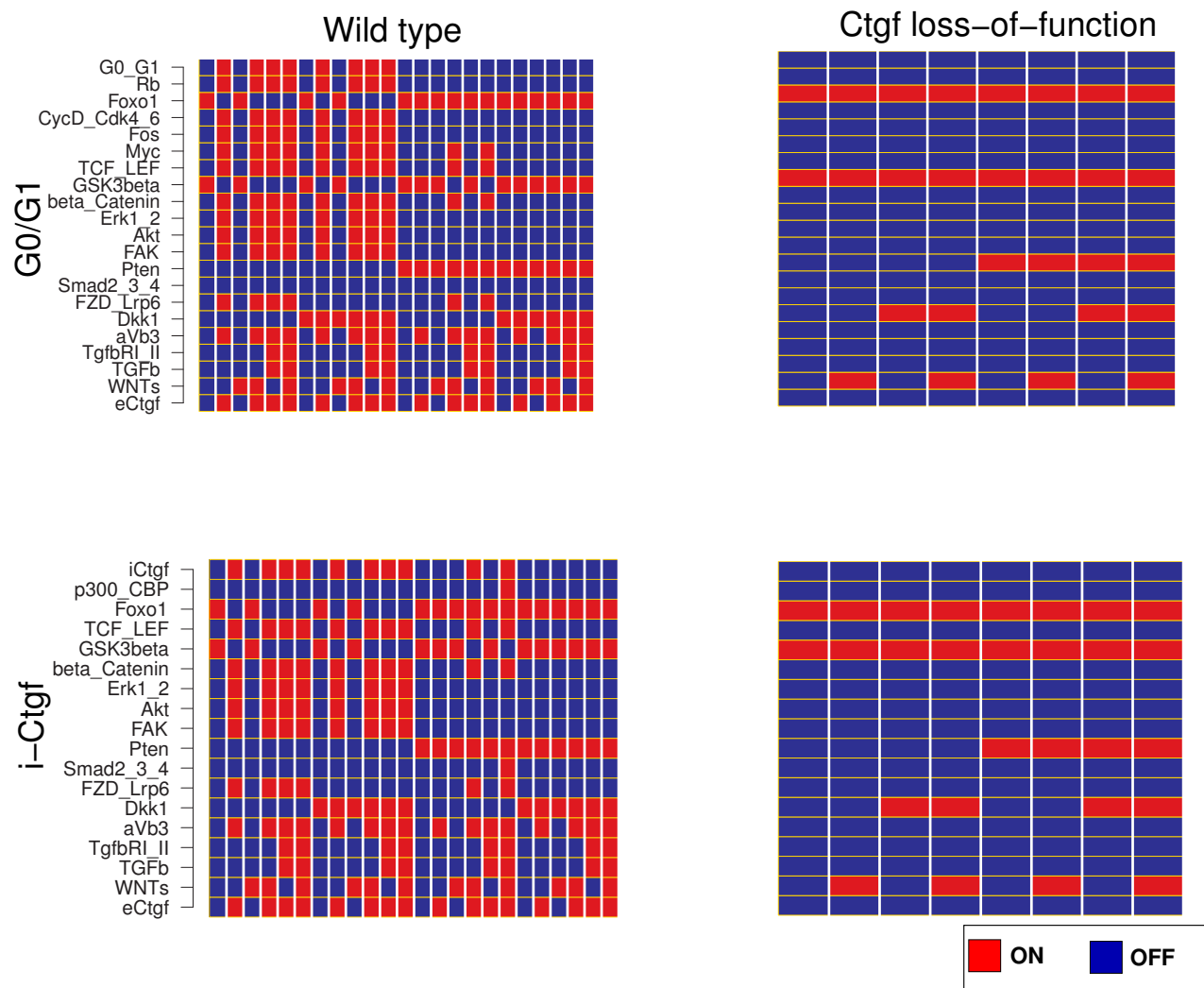


Figure 3.18: Attractor (“stable state”) analysis of the Boolean networks modeling Ctgf regulated G0-to-G1 transition and its auto-induction without fixing Ctgf, WNTs and Tgf- β states. Using the synchronous updating strategy and Boolean logical rules listed in Table 3.10 an exhaustive attractor search was performed for either the wild type (left panel) or Ctgf loss-of-function (right panel) case. Since exhaustive attractors are identified by exhaustive search of all 2^n states, where n is the number of nodes that are not set to a fixed value, for the largest of the sub-models, G1/S-block, also when restricting the number of nodes to 29, which was the maximum number of nodes allowed for exhaustive search,²²³ R could not allocate memory for the state tables.

3 Results

the activation of Cyclin D1 and the subsequent exit from quiescence. Interestingly, Ctgf has been reported to induce Gsk3- β phosphorylation and decreased the active pool of Gsk3- β .³⁶⁰ In the case, where Ctgf, WNTs and Tgf- β states were not fixed, a very large number of attractors was found (Figure 3.18, top panel).

G1/S block; Ctgf contributes to the G1/S block via inducing p21Cip1 and p27Kip1

In the G1/S block model, three out of five attractors reflected the activation of E2f1 leading to the entry into the S phase of the cell cycle, whereas in two stable states the G1/S cell cycle block occurred, due to the activity of solely p27Kip1 or the joint activity of p27Kip1 and p21Cip1. In our model (Table 3.10), for the CKI inhibitor p21Cip1 to stabilize at ON, all its positive regulators (Smad2/3/4, FoxO1, p53 and p300) had to be ON. However, for p27Kip1 to be induced, it was enough that one of its positive regulators, namely, FoxO1 or Gsk3- β was active. Interestingly, previous model revealed that FoxO1 and Gsk3- β could also be involved in the suppression of Cyclin D1 protein levels. Thus, according to the model, both the inhibition of Cyclin D1 and induction of p27Kip1 can be attributed to the activity of FoxO1 and Gsk3- β . Remarkably, in some of the attractors, where G1/S transition would take place, Cyclin D1-Cdk4/6 complex stabilized at ON. Indeed, it has also been shown that Cyclin D1 may stimulate E2f1.^{361,362} The Ctgf loss-of-function simulation yielded four simple attractors each consisting of one state and having a basin of 16777216 states. The activation of p21Cip1, as well as E2f1 and the G1/S transition was completely abolished here. In all stable states, the Gsk3- β kinase was active. According to the literature, Gsk3- β may trigger p21Cip1 degradation.³⁶³ At the same time, Gsk3- β activation is known to prevent the degradation of p27Kip1.³⁶⁴ Indeed, p27Kip1 was still active in all cases. Again, Gsk3- β activity, also FoxO1 stabilized at ON.

iCtgf; Ctgf is able to auto-induce its own expression Finally, we simulated how extrinsic Ctgf might induce its own expression in LSK cells via a positive feedback loop from Wnt, Tgf- β or both signaling pathways (see Figure 3.16). In the wild type, the simulation again yielded four simple attractors, each consisting of one state and having a basin of 8192 states (Figure 3.17, bottom left panel). In three out of four cases, intrinsic Ctgf ("i-Ctgf") stabilized at ON. In two attractors, this induction was due to the activity of the Wnt pathway, in particular the induction of β -Catenin and TCF/LEF. In one case (attractor # 3), in addition to the Wnt pathway, also a positive feedback loop from the Tgf- β pathway could be established, i.e., Smad2/3/4, FoxO1 and p300/CBP stabilized at ON. Attractor # four was a stable state, where no Ctgf auto-induction occurred. Here, the Wnt inhibitor Dkk1 was active, leading to the induction of Gsk3- β . In the Ctgf loss-of-function attractor analysis, no auto-induction of Ctgf could be observed. Similarly to the attractor # four in the wild type, this was due

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

to the activity of Gsk3- β . Again, when the states of *Ctgf*, WNTs and Tgf- β were not fixed, BoolNet returned a rather large number of attractors (Figure 3.18, bottom panel).

Experimental validation of the *Ctgf* loss-of-function simulation results

Above, we established a *Ctgf* signaling network and determined possible ways how *Ctgf* could affect G0/G1-transition, G1/S block, or its auto-induction. To find out which of these possible phenotypes ("stable states") and molecular mechanisms driving them actually occur in stem cells, we performed 24 h co-culture experiments of isolated wild-type Lin-Sca-1+c-Kit+ (LSK) cells with either UG26-1B6^{pLKO} or UG26-1B6^{shCtgf} (Figure 3.1; step 9) and examined mRNA (see Figure 3.20) and/or protein levels or phosphorylation status (Figure 3.19) of several network molecules in the co-culture-derived stem (LSK) cells, including the cell cycle regulators Cyclin D1 (*Ccnd1*), p21Cip1 (*Cdkn1a*), p27Kip1 (*Cdkn1b*), as well as Tgf- β , several Wnt, Integrin/FAK(Focal adhesion)/PI3K/Akt(PKB) and MAPK pathway members, using RT-qPCR and Immunofluorescence staining, respectively. Note that we only experimentally tested the *Ctgf* loss-of-function simulation results (Figure 3.17, right panel). Moreover, we noticed that most changes in the states of the network molecules in LSK cells co-cultured with UG26-1B6^{shCtgf} for 24 h were observed on the level of [phospho]-protein rather than mRNA (Figure 3.20 and 3.19), which is in line with previous findings that the differential expression of mRNA can capture at most 40% of the variation of protein expression.³⁶⁵ Since most biological functions are mediated by proteins and changes in their post-translational modifications (e.g., phosphorylation), conveying information through a network,³⁶⁶ here, we also focused on the protein level data, where available. Since, using the Boolean logical model, we predicted the binary network response (ON/OFF) to the absence of stromal *Ctgf*, a comparison of the experimental data with the predictions requires a discretization of the data, in our case, a binarization. The discretized values were obtained manually, using the following line of reasoning: (i) we first looked at the ratio between the measurement value at time after 24 h of co-culture on UG26-1B6^{shCtgf} stromal cells and the measurement value after 24 h of co-culture on control (pLKO.1) UG26-1B6 stromal cells and its statistical significance; (ii) the biochemical knowledge (e.g., whether the measured phosphorylation is activating or inhibiting) was taken into account. Hence, in order to discretize the measured signal to "ON", a statistically significant (p-Value ≤ 0.05) increase in the relative protein level or an activating phosphorylation, or a decrease in an inhibiting phosphorylation in LSK cells co-cultured with UG26-1B6^{shCtgf} for 24 h would have to be observed, whereas a statistically significant (p-Value ≤ 0.05) decrease in the relative protein level, an increase in an inhibiting phosphorylation, or a decrease in an activating phosphorylation would discretize the measured signal to "OFF". The discretization of all

3 Results

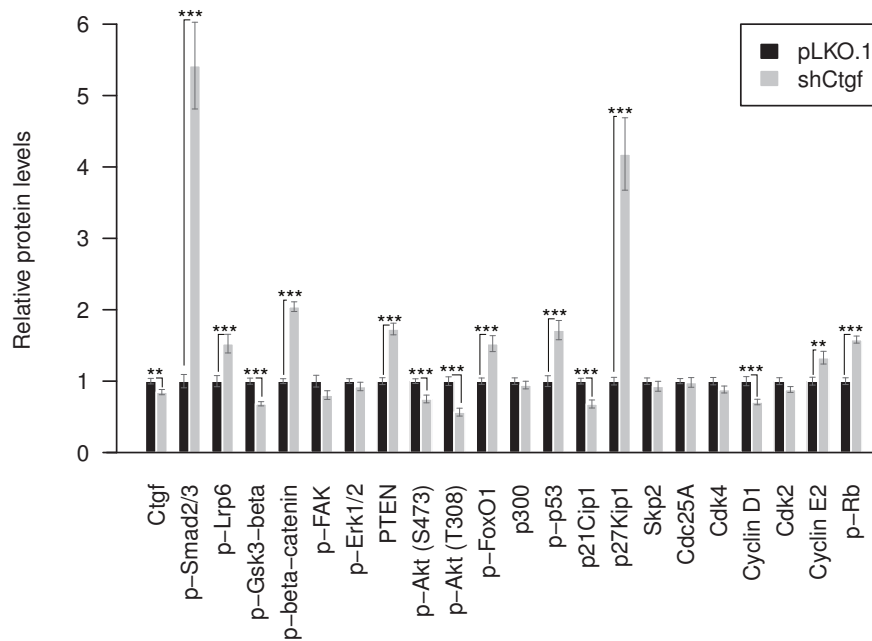
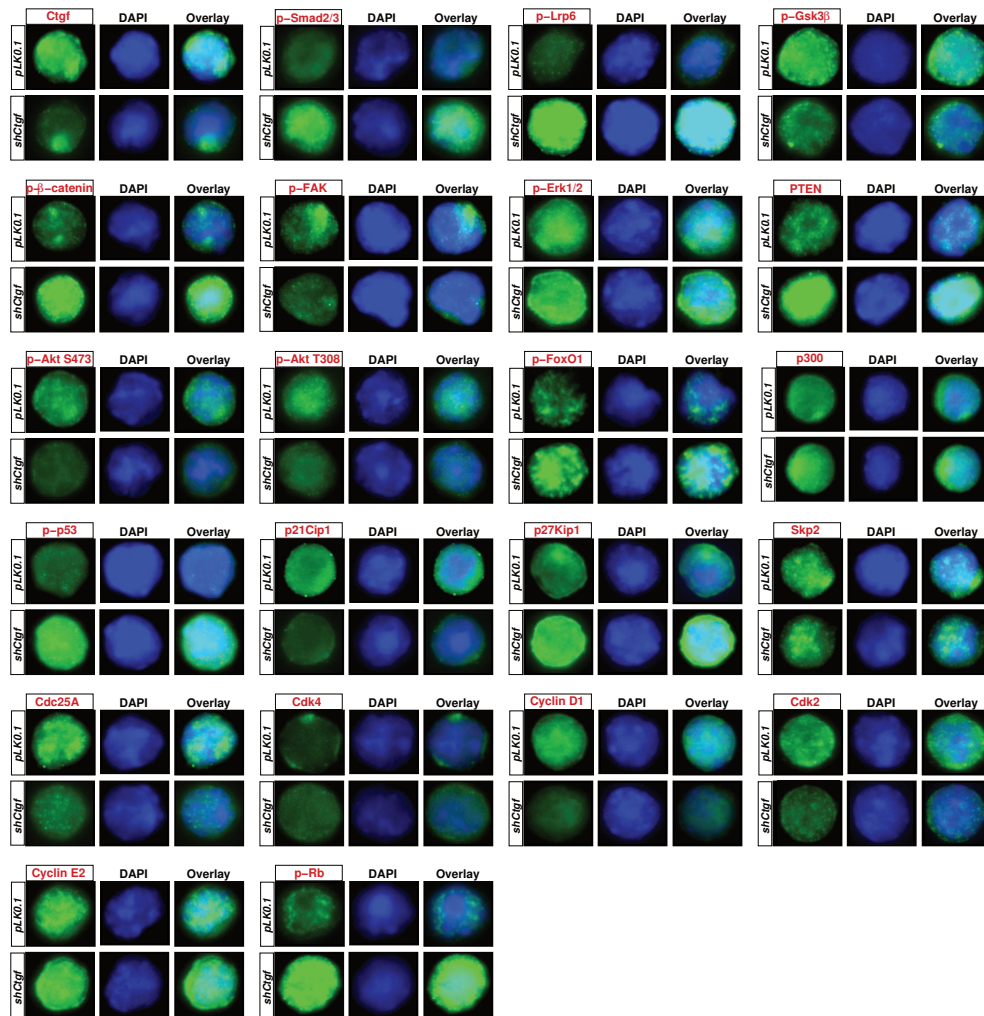


Figure 3.19

Figure 3.19: Experimental validation of the *Ctgf* loss-of-function simulation results. Protein levels or phosphorylation status of selected proteins. LSK (stem cells) were co-cultured for 24 h with UG26-1B6^{si*Ctgf*} or UG26-1B6^{p*LKO.1*} (control) stromal cells. Co-culture-derived Cd45+LSK cells were then separated by flow cytometry and subjected to protein level or phosphorylation state profiling using immunofluorescence staining (IC-IF) and antibodies listed in Tables 2.8 and 2.10 (as described in detail in *Materials and methods, 2.2.1.7 Immunocytofluorescence (IF) microscopy* on page 36).

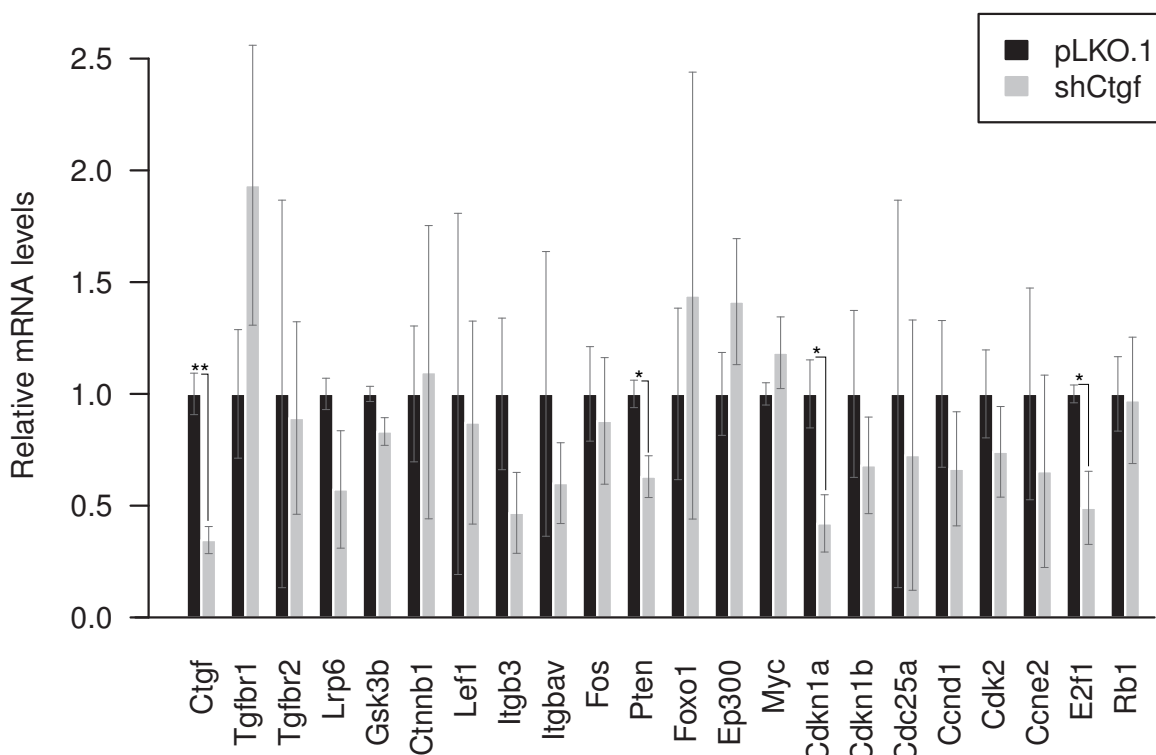


Figure 3.20: Experimental validation of the *Ctgf* loss-of-function simulation results. mRNA expression levels of selected genes. LSK (stem cells) were co-cultured for 24 h with UG26-1B6^{si*Ctgf*} or UG26-1B6^{p*LKO.1*} (control) stromal cells. Co-culture-derived Cd45+LSK cells were then separated by flow cytometry and subjected to mRNA profiling using RT-qPCR and primers listed in Table S11 (as described in detail in *Materials and methods, 2.2.1.6 Quantitative real-time PCR (qPCR)* on page 35).

Table 3.11: Agreement of Boolean attractors for the three sub-processes with experimental protein measurements. For each of the three functional outcomes ("G0/G1", "G1/S block" and "i-Ctgf") the total number of nodes in the network is given ("Total_nodes"). Note that only a sub-set of network nodes were profiled for relative protein levels and/or phosphorylation status ("Protein_tested"), hence, we only calculate the agreement between the simulation results and experimental data for these nodes. Columns "Overlap" "Attractor_1" to "Attractor_4" give, for each of the four Boolean "stable states" (attractors), the number and percentage of nodes, for which the simulation results were in agreement with experimentally measured protein (as shown in Figure 3.19). An exception to this was E2f1_Dp1, where we used the mRNA levels (Figure 3.20; *E2f1*) instead.

| Process | Total_nodes | Protein_tested | Overlap | | | |
|----------|-------------|----------------|-------------|-------------|-------------|-------------|
| | | | Attractor_1 | Attractor_2 | Attractor_3 | Attractor_4 |
| 1 G0/G1 | 20 | 11 | 7 (63.6 %) | 7 (63.6 %) | 8 (72.7 %) | 8 (72.7 %) |
| 2 G1/S | 28 | 19 | 14 (73.7 %) | 14 (73.7 %) | 15 (78.9 %) | 15 (78.9 %) |
| 3 i-Ctgf | 18 | 11 | 7 (63.6 %) | 7 (63.6 %) | 8 (72.7 %) | 8 (72.7 %) |

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

network species, for which experimental measurements of protein levels or phosphorylation status were available (Figure 3.19) is given in Table S10. For the *Ctgf* loss-of-function case, the discretized data were then compared to the predictions (Figure 3.17; right panel; "Ctgf loss-of-function"; "Protein data"). An exception to this was *E2f1_Dp1*, where we used the mRNA levels (Figure 3.20; *E2f1*) instead. A graphical overview of the validation results is also provided in Figure 3.16.

G0/G1; *Ctgf* promotes G0-G1 transition As predicted by the model, our experiments demonstrated that loss of stromal *Ctgf* leads to a significant decrease in Cyclin D1 protein levels (Figure 3.19; Cyclin D1 in row 5 and column 3), as could be confirmed experimentally using Immunofluorescence staining. However, the protein levels of its associated Cdk4 kinase remained unchanged (Figure 3.19; Cdk4 in row 5 and column 2). Moreover, none of the transcription factors, possibly involved in the activation of Cyclin D1 (as predicted by our Boolean model), such as Fos, c-Myc and TCF/LEF demonstrated a significant decrease in their mRNA levels after extrinsic *Ctgf* was being knocked-down (Figure 3.20; see *Fos*, *Myc* and *Lef1*), as it would be expected according to simulation results. Similarly, FoxO1 was defined as a negative regulator of Cyclin D1 (Table 3.10) and it stabilized at ON in all the stable states in the *Ctgf* loss-of-function simulation (Figure 3.17, top right panel). However, no increase in its mRNA levels could be detected (Figure 3.20; see *Foxo1*). On contrary, our Immunofluorescence (IC-IF) experiments revealed a significant increase in phosphorylation of FoxO1-Ser256 (Figure 3.19; p-FoxO1 in row 3 and column 3), which results in nuclear export and inhibition of transcription factor activity.³⁶⁷

In all "stable states" (attractors) of our model (Figure 3.17, top right panel), β -catenin stabilized at OFF and Gsk3- β was active, which also could be confirmed experimentally. As it can be inferred from Figure 3.19 (p-Gsk3 β in row 1 and column 4), the levels of phospho-Gsk3- β (Ser9), an indicative of its inhibition,³⁶⁸ significantly decreased in LSK cells co-cultured with *Ctgf* KD stromal cells. At the same time, the abundance of Ser33/Ser37/Thr41 phosphorylated β -catenin (targeting it towards degradation)³⁶⁹ significantly increased (Figure 3.19; p- β -catenin in row 2 and column 1). RT-qPCR analysis, however, demonstrated no significant changes in mRNA levels for *Ctnnb1* and *Gsk3b* (see Figure 3.20).

The Boolean model also predicted that the activity of Cyclin D1 and Gsk3- β , as well as the CDK inhibitors p21Cip1 and p27Kip1 depends on the activity state of the Erk1/2 and Akt (PKB) kinases (Table 3.10). To test this prediction, we experimentally profiled the levels of phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204), an indicative of its activation³⁷⁰ and the phosphorylation of Akt (PKB) at regulatory residues Thr-308 and Ser-473, known to lead to its full activation.³⁷¹ We observed unchanged levels of phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (Figure 3.19; p-Erk1/2 in row 2 and column 3). Nevertheless, our

3 Results

experiments revealed a significant down-regulation of Akt, as measured by the decrease in p-Akt-Thr-308 and -Ser-473 (Figure 3.19; p-Akt (T308) in row 3 and column 2 and p-Akt (S473) in row 3 and column 1).

It is known that, inactivation of PTEN, leads to the constitutive activation of protein kinase B (PKB)/Akt via enhanced phosphorylation of Thr-308 and Ser-473.³⁷² In line with this, we observed that the protein levels of PTEN, a lipid phosphatase upstream of protein kinase B (Akt),³⁷³ increased when LSK cells were co-cultured with UG26-1B6^{shCtgf} (Figure 3.19; PTEN in row 2 and column 4), whereas its mRNA levels were significantly down-regulated (Figure 3.20; *Pten*). In the *Ctgf* loss-of-function simulation, PTEN took all its possible values (1 and 0). In fact, this was one of the two network species (the other one being *Dkk1*), which differed between the "stable states": PTEN was inactive (OFF/value 0) in the first two cases and active (ON/value 1) in the "stable states" three and four (Figure 3.17, top right panel). *Dkk1*, on the other hand, was inactive (OFF/value 0) in the first and third attractor and active (ON/value 1) in the "stable states" two and four. The mRNA and/or protein levels of *Dkk1* remain to be validated experimentally, yet.

Of note, for the G0/G1 case, the "Protein data" of Rb is marked as not profiled (Figure 3.17, top right panel), since its phosphorylation on Ser-780 by Cyclin D1-Cdk4, which was investigated in our experiments (see Figure 3.19; p-Rb in row 6 and column 2), is specific to late G1 phase of the cell cycle,^{309,374,375} whereas here we are interested in G0 and early G1 phase cells. pRb is present as an unphosphorylated protein in G0 (quiescent) cells and becomes hypo-phosphorylated (activated) in early G1, and hyper-phosphorylated (inactivated) in late G1 phase.^{309,376} In early G1, the Cyclin D1:Cdk4/6 complexes convert unphosphorylated Rb to hypo-phosphorylated pRb, thus activating it. This active (partially phosphorylated or hypophosphorylated) Rb is able to bind and sequester transcription factors, such as E2f, and repress E2f-responsive genes, whereas the release of free E2f is thought to be required for the G1/S transition, and Rb is supposed to keep a cell in the G1/G0 phase as long as it binds E2f.^{308,309,376} At the same time, in late G1, the Cyclin D1:Cdk4/6 complexes are responsible for initiating the inactivation of Rb by hyper-phosphorylation.³⁰⁹ Due to this, different Boolean rules were defined for the two sub-models: in the G0/(early)G1 we define CyclinD1/Cdk4/6 as a positive regulator of Rb, whereas in the (late)G1/S sub-model, CyclinD1/Cdk4/6 is one of its negative regulators (see G0/G1 vs G1/S block in Table 3.10).

Hence, experiments comparing the ratio of unphosphorylated vs. hypo-phosphorylated Rb protein would need to be performed, in order to validate the Boolean model predictions in G0/(early)G1.

G1/S block; *Ctgf* contributes to the G1/S block via inducing p21Cip1 and p27Kip1 In late G1, Rb inactivation and subsequent cell cycle progression requires an initial phosphoryla-

3.8 Network model predicting the role of *Ctgf* in regulating the LSK cell-cycle status

tion by Cyclin D1:Cdk4/6, followed by Cyclin E-CDK2 phosphorylation.³⁷⁷ Cyclin D1-Cdk4, but not Cyclin E-Cdk2 phosphorylate Ser-780 in pRB, marking the conversion of Rb from a transcriptionally repressive, hypo-phosphorylated state to an inactive, hyper-phosphorylated state. Inactivation of pRb by hyper-phosphorylation in late G1 phase causes the release of E2F, allowing transcription of genes important for DNA synthesis (S phase entry).^{309,374,375} Our Boolean model predicted that *Ctgf* loss-of-function would lead to G1/S cell cycle block in all "stable states", due to the ON state of Rb, resulting in E2f1 stabilizing at OFF (Figure 3.17, middle right panel, rows 1, 2 and 10 for G1_S_block, E2f1_Dp1 and Rb, respectively). In agreement to this, our RT-qPCR analysis demonstrated significantly decreased (*E2f1* mRNA levels (Figure 3.20; *E2f1*), however, the abundance of pRb-Ser780 significantly increased (Figure 3.19; p-Rb in row 6 and column 2), independently of the observed decrease in Cyclin D1 (Figure 3.19; Cyclin D1 in row 5 and column 3). In addition, we also used the BrdU assay to quantitate the percentage of cells entering and progressing through the S (DNA synthesis) phase of the cell cycle. However, as it can be seen in Figure 3.12(D) (middle panel), the analysis revealed no significant differences in the distribution of the cell cycle positions in LSK cells after 1 week co-culture on UG26-1B6^{sh*Ctgf*} vs. control (pLKO.1) cells.

Also in the context of G1/S cell cycle block, our Boolean attractor analysis revealed that the activation of p21Cip1 was abolished (Figure 3.17, middle right panel) in the *Ctgf* loss-of-function case. Indeed, a significant decrease in *Cdkn1a* mRNA levels (Figure 3.20; *Cdkn1a*) and p21Cip1 total protein (Figure 3.19; p21Cip1 in row 4 and column 2) could be confirmed experimentally. On the other hand, stable states where p27Kip1 stabilized at ON (active/value 1) were also obtained (Figure 3.17, middle right panel). In agreement to the simulation results, we observed a significant up-regulation of p27Kip1 total protein (Figure 3.19; p27Kip1 in row 4 and column 3), although its mRNA levels remained unchanged (Figure 3.20; *Cdkn1b*). The Boolean model predicted that the activity of the CDK inhibitors p21Cip1 and p27Kip1 depends on the activity state of the Erk1/2 and Akt (PKB) kinases (Table 3.10), the experimental validation results for which were already described above (see "G0/G1; *Ctgf* promotes G0-G1 transition").

Finally, according to our Boolean rules (Table 3.10) and simulation results (Figure 3.17, middle right panel; Smad2_3_4 in row 22), SMAD was inactive (OFF/value 0). Nevertheless, when profiling the levels of phospho-Smad2(Ser465/467)/Smad3(Ser423/425), a significant up-regulation of p-Smad2/3 was observed in LSK cells co-cultured with UG26-1B6^{sh*Ctgf*} (Figure 3.19; p-Smad2/3 in row 1 and column 2).

iCtgf; *Ctgf* is able to auto-induce its own expression Finally, for the Boolean sub-model predicting how *Ctgf* would be able to auto-induce its own expression, a significant decrease in both *Ctgf* mRNA (Figure 3.20; *Ctgf*) and protein (Figure 3.19; *Ctgf* in row1 and column

3 Results

1) levels could be confirmed experimentally.

In general, we noticed that most changes in the states of the network molecules in LSK cells co-cultured with UG26-1B6^{shCtgf} for 24 h were observed on the level of [phospho]-protein rather than mRNA, as it can be seen by comparing Figure 3.20 vs. Figure 3.19, which is actually in line with previous findings that the differential expression of mRNA can capture at most 40% of the variation of protein expression.³⁶⁵ Since most biological functions are mediated by proteins and changes in their post-translational modifications (e.g., phosphorylation), we focused on the protein level data, where available (e.g., in Figure 3.16).

A summary of the above mentioned experimental validation results is given in Table 3.11. Here, for each of the three functional outcomes ("G0/G1", "G1/S block" and "i-Ctgf") we calculated the agreement of the obtained Boolean "stable states" (attractors) to the experimental protein measurements (Figure 3.19). Again, an exception to this was E2f1_Dp1, where we used the mRNA levels (Figure 3.20; *E2f1*) instead. To do so, first we determined the total number of nodes in the network (see "Total_nodes" in Table 3.11). Next, since only a sub-set of network species were profiled for their relative protein levels and/or phosphorylation status, we determine the number of those nodes ("Protein_tested" in Table 3.11), and we only calculate the agreement between the simulation results and experimental data for these nodes. Finally, to determine this agreement, for each "Protein_tested" node within the network, we determine the experimentally (Immunofluorescence; IC-IF) measured activity status (active/ON/value 1 vs. inactive/OFF/value 0) with the fixed state in the Boolean "stable states" (attractors), as shown in Table S10. For each "stable state" (Columns "Overlap" "Attractor_1" to "Attractor_4" in Table 3.11) we then calculate the number and percentage of nodes, for which the simulation results were in agreement with experimentally measured protein. As it can be clearly seen from Table 3.11, in all cases ("G0/G1", "G1/S block" and "i-Ctgf"), the highest overlap with the measurement achieved to "stable states" "Attractor_3" and "Attractor_4": 8 positive nodes or 72.7 % overlap for the G0-to-G1 transition case and for the Ctgf auto-induction case, whereas ***the highest overlap between the Boolean predictions and experimental data was for the G/S block case was 78.9 % (15 positive nodes)***. As already described above, ***the protein levels (activity status) of the dickkopf Wnt signaling pathway inhibitor 1 (Dkk1) remain to be validated experimentally, in order to determine which one of these two Boolean attractors represents the "true phenotype" of the LSK cells***, derived from Ctgf deficient micro-environment (UG26-1B6^{shCtgf}), as, according to the simulation results (Figure 3.17, right panel), Dkk1 was inactive (OFF/value 0) in the "Attractor_3" and active (ON/value 1) in the "Attractor_4".

4 Discussion

4.1 Time-series (TS) gene expression data of co-cultured LSK and UG26-1B6 stromal cells

4.1.1 Gene expression (microarray) data generation

Recent advances in genome research and gene profiling technologies have resulted in accumulation of global gene expression patterns of primitive hematopoietic stem cells and their more differentiated progeny,^{17,32,86,196,197,378–385} as well as terminally differentiated hematopoietic cells.^{386–392} There have also been attempts to compare HSCs to other stem cell populations.^{393–395} At the same time, efforts to examine the interactions between HSCs and their micro-environmental cells have led to the establishment of *in vitro* culture systems. Several stromal cell lines have been generated not only from adherent bone marrow (BM) cells, but also from fetal liver (FL) and the aorta-gonad-mesonephros (AGM) region and have been shown to maintain HSCs *in vitro*.^{396–404} Among others, two midgestation-derived stromal clones UG26-1B6 (urogenital ridge-derived) and EL08-1D2 (embryonic liver-derived)¹⁹⁸ have been earlier demonstrated to preserve the maintenance of repopulating HSCs in an *in vitro* co-culture, without added cytokines, for periods of at least four weeks.¹⁹⁸ Furthermore, we have already identified several UG26-1B6 and EL08-1D2-derived secreted factors, including *Secreted frizzled-related protein 1 (Sfrp1)* and *Pleiotrophin (Ptn)* and demonstrated their critical role in the maintenance of HSCs.^{51,200} Due to this, it appears that such co-culture systems can, at least partially, mimic the hypothetical *in vivo* stem cell 'niche' and as such provides a useful model system for investigating HSC-stromal cell interactions.¹⁵ Gene expression studies profiling various micro-environment elements, such as comparing HSC supporting with non-supporting niche cells, have also been reported in the literature.^{24,405,406}

In this study, we aimed to monitor how hematopoietic stem cells (HSCs) and their micro-environmental stromal cells would influence each others expression pattern over time, after being in direct contact, and to determine possible key extracellular and intracellular molecular players governing these molecular responses. For this purpose, we performed time-course gene expression analysis, in which a purified population of stem/progenitor cells, defined

4 Discussion

as Lin-Sca1+c-kit+ (LSKs) were co-cultured with the urogenital ridge-derived UG26-1B6 stromal cells for 1, 2 and 3 days and compared to day 0 cells, i.e., freshly isolated LSKs and cultured stromal cells (see Figure 3.2). Collection of this high-throughput data was complicated by several obstacles. First, it was essential to identify and purify both cell types unambiguously after the co-culture, which was hampered by the lack of known UG26-1B6 stromal cell-specific surface markers and the autofluorescence signals originating from these cells. Second, it was challenging to isolate sufficient, measurable amounts of intact, clean, and highly concentrated total RNA, needed as input for microarray hybridization, due to several factors, including the extremely low frequency of HSC in the bone marrow ($\sim 0.05\%$),² the fact that most ($\sim 70\%$) HSC reside in G0 (quiescent) phase of cell cycle,⁹ during which little or no mRNA synthesis takes place,⁴⁰⁷ and the substantial cell losses ($\sim 80-90\%$) after the first 24 h of co-culture. Moreover, in contrast to embryonic stem cells (ESCs), which when derived from the inner cell mass can be maintained *in vitro* as cell lines, current attempts to expand or even maintain HSCs *ex vivo* as homogeneous populations have been modest, so far. As a result, successful development of HSC cell lines have not been reported, hampering harvest of large numbers of HSCs, necessary for large-scale experiments.⁴⁰⁸ Interestingly, as it can be inferred from Figure 3.2, LSK cells seem to change their surface marker expression pattern already during the first 3 days in co-culture. After the first 24 h, the expression of Sca-1 decreases: $\sim 50\%$ of cells have lost their Sca-1 surface marker, however, the separation into Sca-1+ and Sca-1- cells is not clear cut, namely, an intermediate Sca-1_{lo} population also seems to be present. Moreover, this is followed also by partial loss of c-Kit on day 2, whereas on day 3 three distinct cell populations already can be observed: Lin-Sca1+c-Kit+, Lin-Sca1-c-Kit+, as well as Lin-Sca1-c-Kit-, corresponding to LSKs, multi-potent progenitors (MPPs), as well as oligopotent progenitors (OPPs), respectively. It has been demonstrated that HSC activity was not detected in the Lin-Sca1- fraction of the bone marrow.² At the same time, however, it is also known that the expression of certain surface markers used in HSC purification procedures may alter their expression profile when, for example, adult mouse HSCs are stimulated to proliferate.⁴⁰⁹ This raises the question, if 50% of the co-cultured cells have really lost the multipotency characteristic to HSCs or it is the *in vitro* co-culture which modifies their surface marker expression, since it is also known that after facing the selective pressures of *in vitro* culture conditions, changes in cell phenotype can be observed, which impedes HSC tracking.⁴⁰⁹

4.1.2 Computational analysis of the time-series (TS) gene expression data

The emergence of microarray and other high-throughput technologies allows the simultaneous measurement of the expression levels of many thousands of genes. Clearly, time-series gene expression data has the potential to generate a great deal of biological knowledge and the information gained from such studies offers an unprecedented opportunity to fully characterize biological processes. However, as already has been recognized earlier, data analysis still constitutes the most challenging step. In particular, there is little to no consensus in the literature about the best method for analyzing (mainly, clustering) time series microarray data despite the fact that hundreds of algorithms have been developed for the task. Currently, the process of generating biological hypotheses from microarray experiments is still rather complex.^{410–412} The analysis of time-series gene expression data of LSK and UG26-1B6 stromal cells both using the STEM (Short Time-Series Expression Miner) clustering algorithm²¹²(Figure 3.4) as well as the analysis of microarray data by two-way comparisons of consecutive time points (e.g., 24h;Day1;d1 vs. 0h;Day0;d0, 48h;Day2;d2 vs. 24h;Day1;d1, etc.) using the `limma`²¹³ t-statistic approach indicated that the most changes in gene expression levels in LSK and stromal cells occurs already during the first 24 h of co-culture (Day 1; d1, Tables S1 and S2, for Lin-Sca1+c-Kit+ (LSK) and UG26-1B6 cells, respectively). Hence, *we focused our downstream analyses on this time point.*

In UG26-1B6 stromal cells, however, these changes in gene expression may not be LSK (stem cell)-dependent. Instead, arising from the cell culture medium change (Figure 3.7, white bars), despite the *filtering of UG26-1B6 transcripts* based on their expression profile 24 h (Day1 culture medium control; C) after adding fresh cell culture medium (as described in 2.2.2.3 *Two-way comparisons of consecutive time points* in the *Materials and methodn* section, page 40).

Also in LSK cells, as already discussed in the context of HSC surface marker expression, it is known that shortly after being exposed to *in vitro* culture conditions, changes in cell phenotype can be observed.⁴⁰⁹ Moreover, in other systems, there have been reports that cell culture and passaging alters gene expression pattern and proliferation rate.⁴¹³ *This raises the question, whether the molecular cross-talk between LSK and UG26-1B6 stromal cells is indeed the most intense right after the first contact of the cells, i.e., a rapid molecular response occurs, or the observed changes in gene expression are in vitro culture induced, i.e. "cell culture effects"?* In general, may not be a good idea to compare freshly isolated cells to the *in vitro* culture-derived ones. Instead, two culture-derived samples, e.g. wild-type vs. perturbed, such as leukemic stem cells

(LSCs) co-cultured with WT stroma or WT HSCs co-cultured with HSCs-non-supporting stroma, or stromal cells where a knock-down of a putative or previously-demonstrated extrinsic regulator of HSCs (e.g., Sfrp1⁵¹ or Ptn²⁰⁰) has been introduced, could be compared.

4.1.3 Independent confirmation of the microarray data

Unfortunately, the reliability of the microarray results is still being challenged due to both systematic and random errors occurring at different stages along the experimental process, and quantitative real-time PCR (RT-qPCR) is a commonly used validation tool for confirming gene expression results obtained from microarray analysis.^{226–228,414–417} Here, we wanted to address, how well the microarray results correlate with gene expression measurements obtained using RT-qPCR and whether there are differences in sensitivity among the two methods. For follow-up analysis, the selection of the validation gene set is often rather biased, as it usually depends on the aims of the study and may be influenced by factors such as the relative difference in expression among the samples, biological function of the candidate genes, their abundance levels and availability of appropriate reagents (e.g., probes and antibodies) for the genes of interest (GOI). Often genes with the highest differential expression ratios are selected for further study, as such differences are most likely to be valid. To overcome these limitations, here, for the validation, 75 genes (46 in LSK cells and 29 in stromal cells (Table 3.1) with possibly diverse biological functionality and expression levels, and statistical significance values spanning a wide dynamic range were selected and profiled by RT-qPCR, which was then used as the "gold standard" in order to evaluate the performance of the microarrays focusing on the fold change (log₂FC) and p-Value agreement, as well as the sensitivity, specificity, accuracy and precision in detection of differential expression. In general, in both LSK and stromal cells, the expression data obtained with the two methods were more consistent for the up-regulated genes (85 and 86% agreement; Figure 3.5(A) and (B), upper panel, for LSK and UG26-1B6 stromal cells, respectively), whereas the comparison of microarray-based results with qPCR yielded less agreement (47 and 46%; Figure 3.5(A) and (B), bottom panel, for LSK and UG26-1B6 stromal cells, respectively). Of note, however, in UG26-1B6 cells, in a number of cases, ***these changes seem to be LSK-independent, as also observed after changing the cell culture medium*** (control; C; white bars in Figure 3.7).

More consistent validation results for the up-regulated genes have been already reported in the literature.^{228,230} To explain this, it was proposed that the effect may be due to the increased variability observed in low-intensity array spots, i.e. down-regulated genes.²³⁰ Alternatively, it was observed that the data set of down-regulated genes included a greater number of genes exhibiting low levels of change (<1.4 fold),²²⁸ for which the lack of concurrence

4.1 Time-series (TS) gene expression data of co-cultured LSK and UG26-1B6 stromal cells

between methods has been commonly reported.^{418–420} In this comparison, however, only genes exhibiting a fold change of 2.0 were selected as differentially expressed, and the agreement seems to be rather poor also for genes demonstrating a log₂FC of ~ -3.0 (Table 3.1). Here, another scenario seems more likely, namely, the lower correlation between the microarray and qPCR for down-regulated genes may be due to the effects of greater variability associated with decreased reaction efficiencies found in qPCR measurements at later cycles, where genes with low expression levels respond.²²⁸ As already described above, we encountered substantial cell losses ($\sim 80\text{-}90\%$) after the first 24 h of co-culture. Hence, the observed effects are most likely influenced by the low amounts of mRNA that could be extracted from these small numbers of cells. Furthermore, in most cases, microarray results over-estimated the fold-change detected by qPCR, an observation also made by others,⁴²¹ however, the opposite has also been reported.⁴²² Possibly, these differences could be attributed to data normalization, which fundamentally differs between microarray analysis and qPCR, the former requiring global normalization, while the latter generally relies on the expression of one or more reference genes or housekeeping genes (HKGs) against which all other gene expression is calibrated. Therefore, as already recognized earlier,²²⁸ selection and appropriate application of normalization criteria may also play a major role in the correlations between the two methods. The statistical significance of the result could not be always confirmed. Among others, this could be attributed to differences in significance testing. In case of RT-qPCR, a simple t-test was used, however, it is well known that this test needs normally distributed variables and is based in the statistical parameters mean and standard deviation. Since the sample size is usually small (three biological replicates in this case), the standard deviation is not well represented.⁴²³ On the other hand, the `limma` package,⁴²⁴ used for microarray data analysis, implements a better alternative, the so called, moderated t-statistics, which is a variant of the t-test that uses linear models with an empirically moderated estimate of the standard error, effectively borrowing information from across the genes to aid inference about individual genes. This gives improved statistical power for even small sample sizes.⁴²³ On the other hand, a large number of t-tests are performed in such gene-by-gene analyses, and many true-null hypotheses may produce small p-values by chance. As a consequence, numerous false positives, or type I errors, may be obtained if p-Values are compared to standard single-test thresholds.⁴²⁵ Due to this, p-Values adjusted for multiple testing should have been used for this comparison. Studies comparing both methods report several other causes of variability, for example, the differences in probe sequence and thus target location.⁴²⁶ In general, increased distance between the location of the qPCR primers and microarray probes on a given gene also decreased the correlation between the two methods.⁴²⁷ Here, in order to avoid such a bias, when a gene-specific primer pair was designed for qPCR, no efforts were

maid in matching its location to that of corresponding probe on the microarray. In addition, the same mRNA should be used for both microarray and real-time PCR analyses, however, this was not possible in our study since the mRNA amount that could be isolated was limited. Moreover, RNA amplification was done before hybridization to the microarray. Furthermore, it is well documented that both qPCR and microarray analysis have inherent pitfalls that may significantly influence the data obtained from each method.²²⁸ For example, one of the issues about the microarray data is the non-specific and cross-hybridization. A significant number of the DNA sequences being arrayed produce 'non-specific' background signals as a result of, e.g., repetitive elements, poly(A) tails or common motifs. When co-hybridized with two cDNA samples labeled with different fluorophores, such DNA sequences may produce (often strong) signals that are interpreted as 'equally expressed' among the biological samples under study.²²⁶ On the other hand, RT-qPCR also has its sources of error including amplification biases, the exponential amplification of errors, mispriming or the formation of primer dimers, and the changing efficiency of qPCR at later cycles.²²⁸ In summary, similar to other laboratory methods, data derived from microarray-based experiments must be interpreted cautiously and skeptically.²²⁶ Although, microarrays demonstrate acceptable reliability for genome-wide gene expression screening, validation of putative changes in gene expression using alternative methods remains necessary. However, for this purpose uniform validation methods and a more complete understanding of how to compare and contrast results derived by different gene expression profiling approaches need to be developed.²²⁶ At the same time, the cost and effort involved in carrying out follow-up studies on a large scale is rather high. More recently, the development of high-throughput DNA sequencing methods has provided an alternative approach for both mapping and quantifying transcriptomes. Hence, RNA-Seq (RNA sequencing) has clear advantages over hybridization-based approaches.⁴²⁸

4.1.4 Phenotypic and functional comparison of freshly isolated vs. 24 h co-culture-derived LSK cells

In contrast to the pluripotency tags of Oct4, Nanog, AP, or SSEA1 for ESCs, we still lack a stringent surrogate marker to follow the HSC multipotent state. Although, current cell sorting strategies allow for the isolation of HSC population populations of interest based on their surface marker expression, shortly after facing the selective pressures of *in vitro* culture conditions, changes in cell phenotype are observed, which hampers HSC tracking.⁴⁰⁸ In the present study, as already discussed above, LSK cells seem to change their surface marker expression pattern already during the first 24 h in co-culture, where $\sim 50\%$ of cells have lost the Sca-1 surface marker (Figure 3.2). Hence, several functional assays were performed in order to confirm HSC activity. Cell proliferation quantitation using BrdU uptake (Figure

3.9(A)) revealed that LSK cells, most probably, did not enter the S phase of the cell cycle (i.e., had no recently synthesized DNA) during the first 24 h in co-culture. In line with this, labeling experiments with CFSE suggested that LSK cells did not undergo any cell divisions during this time period (Figure 3.9(B)). In fact, no explicit data can be found in the literature regarding the time point of first HSC divisions in culture, however, a report by Jing and collaborates,⁴²⁹ comparing three distinct localizations of HSCs relative to the mesenchymal stromal cell layer: (i) those in supernatant (non-adherent cells); (ii) those adhering to the surface of mesenchymal stromal cells (phase-bright cells) and (iii) those beneath the mesenchymal stromal cells (phase-dim cells), states that at day two, approximately 50% of phase-dim cells had not divided (generation 0), while less than 20% of the other two cell fractions were generation 0 cells. As it can be inferred from Figure 3.9(C), no early apoptotic cells (PI-, Annexin V+) cells could be detected within the LSK or Cd45+LSK cell compartments. At the same time, we encounter substantial cell losses (~ 80-90%) after the first 24 h in co-culture. Most probably, the numbers of necrotic cells (PI+, Annexin V+) should be compared, however, PI+ cells together with cell debris are usually excluded from data prior to selecting LSK cells. Alternatively, instead of comparing the rates of spontaneous apoptosis, apoptosis could be first induced (e.g., by incubating the cells with various doses of cycloheximide) in fresh and co-culture-derived LSKs and then a comparison of both cell type susceptibility to induced apoptosis conducted. Interestingly, colony forming cell (CFC) assay (Figure 3.9(D)) indicated a decreased progenitor forming capacity of LSK cells after the co-culture. At the same time, *in vivo* repopulating capacity in the bone marrow (BM) (Figure 3.9(E)) was significantly increased for the 24 h co-culture-derived LSK cells. These observations could be interpreted in terms of cell resistance to stress and susceptibility to apoptosis, where long-term repopulating cells or LT-HSC are known to be uniquely resistant to stress.¹⁷

4.2 Identifying *Ctgf* as a novel regulator of hematopoiesis

Hematopoietic stem cell (HSC) cell cycle status is thought to be precisely coordinated by a specific combination of 'niche' signals. In the present study, we identified a secreted 36-38 kDa cysteine-rich matrix remodeller, Connective tissue growth factor (*Ctgf*), as a putative novel regulator of early interaction events between HSC and stromal cells from our co-culture experiments of Lin-Sca1+cKit+ (LSK) and HSC-supportive UG26-1B6 stromal cells.

4.2.1 Identifying *Ctgf* using candidate gene prioritization

Connective tissue growth factor (*Ctgf*) was among the highest ranked genes using the ToppGene suite <http://toppgene.cchmc.org>,²²⁴ based on the similarity of its functional

annotations to the genes previously associated with hematopoiesis (Table S5). In LSK cells, its induction could be confirmed both on mRNA and protein levels (Figure 3.12(A)(a,c,e)), whereas the UG26-1B6 culture medium contained high levels (~ 14 ng/mL) of secreted Ctgf already prior to the co-culture (Figure 3.12 (A)(d); 0h). Moreover, on the mRNA level, the up-regulation of Ctgf was LSK (co-culture) independent, since an increase in Ctgf mRNA levels was also observed 24 h after adding fresh culture medium to the cells (Figure 3.7(A), white bars, Ctgf).

In the context of leukemia, Ctgf was shown to be the relatively highest *over-expressed gene in precursor B-acute lymphoblastic leukemia (ALL)*,²⁴⁵ and, in a more recent study, *silencing of Ctgf resulted in significant suppression of leukemia cell growth*, which was *associated with AKT/mTOR inactivation and increased levels of cyclin-dependent kinase inhibitor p27*.²⁸² Moreover, Battula and colleagues²⁷⁰ have also investigated the *role of Ctgf in mesenchymal stromal cells (MSCs)*, finding that Ctgf knocked down (KD) human BM-derived MSCs exhibited *fivefold lower proliferation compared with control MSCs, had markedly fewer S-phase cells and differentiated into adipocytes at a sixfold higher rate than controls*.

4.2.2 Ctgf regulates hematopoietic stem/progenitor cell (HSC/P) activity and engraftment potential

Currently, to the best of our knowledge, *Ctgf has not been characterized in the context of normal hematopoiesis*. We decided to select Ctgf for further study. Since Ctgf has been reported to auto-induce its own expression,²⁸⁷ *we hypothesized that its up-regulation in LSK (stem cells) may be attributed to extrinsic, UG26-1B6-derived Ctgf*. To test this hypothesis and to explore the behavior of HSCs within a Ctgf-deficient 'niche', we utilized RNAi to experimentally induce a *stable knock-down of the corresponding protein in UG26-1B6 stromal cells* (Figure 3.12(B)(a-c)) and performed a series of *in vitro* co-culture experiments, in which we phenotypically and functionally characterized hematopoietic stem/progenitor cells (HSC/Ps) after 1 week co-culture with UG26-1B6^{siCtgf} or control (pLKO.1) stromal cells. Our experiments revealed a significant increase ($p \leq 0.05$) in colony-forming cells (CFCs; 2.2% vs. 0.6% , Figure 3.12(C)) and Lin-Sca1-cKit⁺ early multipotent progenitors (MPPs; 24.9% vs. 12.6% , Figure 3.12(D)). Moreover, *in vivo*, 16 weeks after transplantation, we observed a significantly decreased donor cell compartment (8.9% vs. 19.8% Cd45.1⁺ donor cells) in the peripheral blood (PB) of recipients (Figure 3.12(E)). These data demonstrated that *stromal cell-derived Ctgf is involved in the regulation of hematopoietic progenitor activity and numbers in vitro and affects the overall engraftment potential of hematopoietic stem cells in vivo*.

Interestingly, the *in vivo* results of Lin⁻ cell co-cultures of with UG26-1B6^{siCtgf} stromal cells are similar to those we previously obtained when studying two other UG261-B6 stromal cell secreted factors, Secreted frizzled-related protein 1 (Sfrp1)⁵¹ and Pleiotrophin (Ptn).²⁰⁰ Here, the colony number of WT Lin⁻ cells after 2 weeks of co-culture on UG26-1B6^{siSfrp1} or UG26-1B6^{siPtn}, respectively, was also significantly increased. At the same time, in contrast to the significantly decreased regenerative capacity of 1 week UG26-1B6^{siCtgf} co-culture-derived WT Lin⁻ cells, found in this study (Figure 3.12(E)), when analyzed engraftment of wild-type Ly5.1 HSC in Sfrp1(-/-) and control 129Ly5.1 mice, 16 weeks after transplantation, a significant increase of engrafted wild-type cells in peripheral blood (PB) and also bone marrow (BM) in Sfrp1(-/-) mice was observed.⁵¹ Similarly, also the transplantations of wild-type (Ptn(+/+)) HSCs into Ptn(-/-) mice revealed increased donor cell production in serial transplantations.²⁰⁰ On the molecular level, in terms of the canonical Wnt signaling, it was demonstrated that the level of β -catenin and Cyclin D1 was decreased in Sfrp1(-/-) LSK and Cd34-LSK cells, respectively,⁵¹ whereas, in the Pleiotrophin study it was concluded that microenvironmental Ptn regulates hematopoietic regeneration through β -catenin-independent down-regulation of Cyclin D1 (Ccnd1).

4.2.3 Construction and dynamic analysis of the literature-derived Boolean network of *Ctgf*-regulated HSC cell cycle progression

The top-down to bottom-up approach to *Ctgf* network modeling

To pinpoint the possible molecular mechanisms tied to *Ctgf*, we first used a top-down approach to catalogue and analyze the complete *Ctgf* interactome (Tables S6 and S7). Functional enrichment analysis of the network (Table S8) revealed association with Tgf β , MAPK, BMP and Wnt signaling, cell cycle control, proliferation and adhesion, as well as extracellular matrix binding. Remarkably, besides well-described *Ctgf* activities, such as involvement in wound healing,²⁹¹ angiogenesis²⁹² and bone formation,²⁹³ the analysis also reported a number of abnormal hematopoiesis-related phenotypes, including changes in hematopoietic cell number, morphology, physiology and chemotaxis. These findings thus ***further underscored the putative role of *Ctgf* in hematopoiesis and shed light on the possible pathways involving Connective growth factor (*Ctgf*), according to the principle of "guilty by association"***.²⁸⁸

However, ***large-scale interactomes***, such as the complete *Ctgf* interaction network, are, clearly, not directly interpretable and sufficient by themselves, as they ***do not provide information on the logic of signaling networks and their spatio-temporal behavior***.^{123,124,178,301} Since, in this study, we were also interested in understanding the dynamical input/output behavior of the *Ctgf* interactome, allowing us to formulation a systems-level

hypothesis and direct targeted experiments, we decided to apply computational modeling to our network. However, such approaches are currently limited to networks with a relatively small number of genes/proteins (up to 30 nodes), therefore, we further used a bottom-up or "seed-gene" approach in order to extract a smaller sub-network from the Ctgf interactome. As the main regulatory activity in the maintenance of HSCs lies in self-renewal, which is a special functional form of cell division, we selected the cell cycle as our terminal node, and set up a literature-derived Boolean model (Figure 3.16 and Table 3.10) to predict the HSC cell cycle status (G0-to-G1 transition and G1/S block) in response to changing levels of extrinsic, UG26-1B6 stromal cell secreted, Connective growth factor (Ctgf). In addition, since Ctgf has been reported to auto-induce its own expression,²⁸⁷ and in 24 h co-culture-derived LSK cells, the induction of Ctgf could be confirmed both on mRNA and relative protein levels (Figure 3.12(A)(a,c,e)), we also modeled its auto-induction.

Dynamic network analysis using Boolean logic and experimental validation of the simulation results

In the present study, we decided to use Boolean logic, since the simple design of such qualitative discrete models has a number of advantages over more complex dynamic models (e.g., quantitative and continuous modeling using differential equations). First of all, such models do not require precise mechanistic details on molecular mechanisms and experimentally determined kinetic parameters describing the individual reactions such as synthesis and degradation rates. Instead, a wealth of molecular level qualitative data on individual components and their molecular interactions obtained from the biomedical literature and high-throughput technologies can be translated into a Boolean logical model. Second, while differential equation-based dynamic models have been limited to a couple of small-scale (only a few genes/proteins)³⁶⁶ well-understood sub-networks the logical approaches enable us to model relatively large-scale (up to 30 nodes) signaling networks allowing, for example, also to study the effects of pathway cross-talk. Several other advantages of qualitative models already recognized earlier¹²⁸ include the fact that such models can be easily expanded, whereas adding a new reaction to a model of differential equations would usually require the elaborate re-estimation of parameters. This flexibility also enables one to generate and validate hypotheses rather quickly. Finally, the qualitative predictions derived from a logical model do not depend on certain parameter values and are more generally valid.

Dynamic simulation using Boolean logic revealed that Ctgf mediated regulation of LSK cell cycle and its auto-induction may depend on the activity of Gsk3- β . In all cases, Gsk3- β stabilized at ON in the Ctgf loss-of-function simulations (Figure 3.17, right panel). Indeed, Ctgf has been reported to induce Gsk3- β phosphorylation on Ser9, decreasing the active

4.2 Identifying *Ctgf* as a novel regulator of hematopoiesis

pool of Gsk3- β .³⁶⁰ As it can be inferred from Figure 3.19 (p-Gsk3- β), the levels of phospho-Gsk3- β (Ser9), indicating the inhibition of the kinase,³⁶⁸ significantly decreased in LSK cells co-cultured with siCtgf stromal cells. At the same time, the abundance of phosphorylated Ser33/Ser37/Thr41 β -catenin significantly increased (Figure 3.19; β -catenin), which is known to target β -catenin towards degradation.³⁶⁹ In the G0/G1 sub-network, the induction of Cyclin D1 was abolished, when Gsk3- β stabilized at ON. According to the current literature, the phosphorylation of Cyclin D1 at Thr286 by Gsk3- β promotes its nuclear export and degradation.⁴³⁰ We observed a significant decrease in Cyclin D1 protein levels (Figure 3.19; Cyclin D1). In the G1/S block case, when Gsk3- β was active, p21Cip1 was OFF, however, p27Kip1 was ON. Indeed, it has been reported that Gsk3- β triggers p21Cip1 degradation.³⁶³ At the same time, Gsk3- β activation prevents p27Kip1 degradation.³⁶⁴ In line with this, a significant decrease in *Cdkn1a* mRNA levels (Figure 3.20; *Cdkn1a*) and p21Cip1 protein levels (Figure 3.19; p21Cip1) could be confirmed experimentally. Remarkably, a significant increase in p27Kip1 protein levels could be observed (Figure 3.19), although the mRNA levels remained unchanged (Figure 3.20; *Cdkn1b*). Intriguingly, in the i-Ctgf sub-network also the Ctgf auto-induction was lost, when Gsk3- β kinase stabilized at ON. A significant decrease in both Ctgf mRNA (Figure 3.20; *Ctgf*) and protein (Figure 3.19; Ctgf) levels could be also confirmed experimentally. The Boolean model also predicted that the activity of Gsk3- β depends on the activity state of Akt(PKB) (Table 3.10). Indeed, our experiments revealed significant down-regulation of phospho-Akt (Thr308/Ser473), whereas we observed that the protein levels of PTEN, a lipid phosphatase upstream of protein kinase B (Akt),³⁷³ increased when LSK cells were co-cultured with UG26-1B6^{shCtgf}. Although we lack detailed experiments to show the precise regulatory relationships between the above described network species, it is tempting to hypothesize that there may be a cross-talk between the Wnt and Akt(PKB) pathways, leading to the inhibition of glycogen synthase kinase-3-beta (Gsk3- β) via Akt(PKB)-dependent phosphorylation of Ser-9,³²⁵ which, on its turn, facilitates the nuclear accumulation of β -catenin, resulting in the induction of its downstream target Cyclin D1 and cell cycle progression. Whereas, the absence of extrinsic Ctgf leads to the activation of the tumor suppressor PTEN and the cell cycle inhibitor p27Kip1, leading to G1/S cell cycle block, which may be accompanied by HSC differentiation, as indicated by our phenotypic and functional experiments discussed above.

Nevertheless, there were also several inconsistencies between the Boolean logical model predictions and the experimental data. By contrast to what was predicted and expected for the G1/S block sub-model, in the Ctgf loss-of-function case, we observed a significant increase in retinoblastoma protein (Rb) phosphorylated on Ser780 (Figure 3.19; p-Rb in row 6 and column 2), marking the conversion of Rb from a transcriptionally repressive,

hypo-phosphorylated state to an inactive, hyper-phosphorylated state. Inactivation of pRb by hyper-phosphorylation in late G1 phase causes the release of E2F, allowing transcription of genes important for DNA synthesis (S phase entry).^{309,374,375} Our Boolean model predicted that Ctgf loss-of-function would lead to G1/S cell cycle block in all "stable states", due to the ON state of Rb, resulting in E2f1 stabilizing at OFF (Figure 3.17, middle right panel, rows 1, 2 and 10 for G1_S_block, E2f1_Dp1 and Rb, respectively). Of note, our RT-qPCR analysis actually demonstrated significantly decreased (*E2f1* mRNA levels (Figure 3.20; *E2f1*), however, phosphorylation and subsequent inactivation of pRb represents a key event governing cell proliferation.³⁰⁵ Functionally, we used the BrdU assay to quantitate the percentage of cells entering and progressing through the S (DNA synthesis) phase of the cell cycle. However, as it can be seen in Figure 3.12(D) (middle panel), the analysis revealed no significant differences in the distribution of the cell cycle positions in LSK cells after 1 week co-culture on UG26-1B6^{shCtgf} vs. control (pLKO.1) cells. Another discrepancy between the experimental data and the model predictions in the G1/S sub-model, was the increase in phospho-Smad2(Ser465/467)/

Smad3(Ser423/425) in LSK cells co-cultured with UG26-1B6^{shCtgf}. This seems to support another previously reported mechanism, where Ctgf stimulating the Tgf- β type III receptor (T β RIII) antagonizes TGF- β 1-induced Smad2/3 phosphorylation.⁴³¹ Furthermore, in our Boolean model, FoxO1 was predicted to be in the ON state. Moreover, it is known that Akt (PKB) phosphorylates FoxO1 at Ser-256, negatively regulating its function,⁴³² however, the levels of active phospho-Akt (Thr308/Ser473) decreased in UG26-1B6^{shCtgf}-co-culture-derived LSK cells. Indeed, a simultaneous increase in phosphorylation of Akt-Ser473 and FoxO1-Ser256 has been previously detected, when stimulating rat primary skeletal muscle satellite cells with insulin-like growth factor I (IGF-I).⁴³³ Finally, the increase in p-Lrp6 (Ser1490) after 24 h co-culture with UG26-1B6^{shCtgf} represents a discrepancy between the simulation results and data. However, it is known that, upon stimulation with Wnt, Lrp6 activated by Glycogen synthase kinase 3 (Gsk-3) by phosphorylating it at Ser-1490,⁴³⁴ indicating missing edges in our initial model.

Taken together, the highest agreement of the Ctgf loss-of-function Boolean model predictions and experimental protein data in our logical analysis, in all cases (sub-models; "G0/G1", "G1/S block" and "i-Ctgf") was achieved by two "stable states" "Attractor_3" and "Attractor_4": 8 positive nodes or 72.7 % overlap for the G0-to-G1 transition case and for the Ctgf auto-induction case, and 15 positive nodes (78.9 % overlap) for the G/S block case (Table 3.11). As already discussed, *the protein levels (activity status) of the dickkopf Wnt signaling pathway inhibitor 1 (Dkk1) remain to be validated experimentally, in order to determine which one of these two Boolean attractors represents*

the "true phenotype" of the LSK cells, derived from *Ctgf* deficient micro-environment (UG26-1B6^{sh*Ctgf*}), as, according to the simulation results (Figure 3.17, right panel), *Dkk1* was inactive (OFF/value 0) in the "Attractor_3" and active (ON/value 1) in the "Attractor_4", in all three sub-models "G0/G1", "G1/S block" and "i-*Ctgf*". Hence, ***the highest overlap (78.9 %) between the Boolean predictions and experimental data was for the G/S block case*** (Table 3.11), despite the observed inactivation of pRb, indicating the opposite.^{309,374,375} ***Detailed cell-cycle studies profiling and quantifying the percentage of cells in different cell cycle phases (including the distinction between G0 and G1; S and G2/M, i.e. the staining with anti-Ki67) at different time points may still be necessary to experimentally determine which of the two cell cycle functional outcomes (G0-to-G1 transition, G/S transition or cell cycle block, or G0-to-G1 transition followed by G/S cell cycle block) would be observed in LSK cells-residing in Ctgf deficient micro-environment, and whether the auto-induction of Ctgf in these cells plays a role in the process.*** Importantly to note, Lin-Sca1+cKit+ (LSK) cells, used in our experiments, are actually a mixture of hematopoietic stem cells (HSCs) and progenitors. Within the LSK population, more than 70% of highly purified HSCs (Cd34-Cd48-Cd150hiLSK) are in the G0 phase, whereas less than 10% of Cd34+LSK cells (differentiated progenitors) are in the quiescent phase.⁴³⁵ Moreover, even after cell sorting based on their surface marker expression, the obtained cell sub-population is dynamic, where some cells divide rapidly and others more slowly; some differentiate, others self-renew; some can give rise to more lineages than others. Due to this variation, population studies of hematopoietic stem cells are unable to accurately address essential questions.⁴³⁶ HSC cell cycle progression and the molecular changes coordinating and accompanying it could be better addressed using a single-cell tracking approach enabling also automatic detection of cell cycle phases.⁴³⁷

The Boolean model constructed and employed in this study is, to the best of our knowledge, the first existing computational model on the *Ctgf*-regulated cell cycle progression. The model was capable to compile and structure the existing knowledge on cell proliferation and functional interactions of the proteins and their regulatory effects. Given the high level of complexity and missing information, the model is clearly far from being complete and inevitably does not allow for a perfect interpretation of the experimental data collected in our study. Inconsistencies observed allow for the expansion and improvement of the model searching for yet unknown molecular players (e.g., additional cell surface receptors and signaling pathways associated with *Ctgf*) or revising relations that might not be confirmed in experiments or conditionally modified. It has already been acknowledged earlier that it is much more difficult to extract information about combinatorial regulatory effects of different proteins affecting a given

component than to extract information about individual interactions from the literature.³⁰⁵ Moreover, much of current available and rapidly accumulating experimental data attempting to capture intracellular regulation is qualitative (limited mechanistic knowledge), noisy (conflicting hypotheses), inaccurate and incomplete. For example, in spite of the intensive study of NF- κ B and TGF- β signaling pathways, new components of these pathways continue to be discovered, indicating that our analysis of even the most well-studied pathways is likely to be incomplete.^{121,170,438} Therefore, the observed discrepancies between the network models and the experimental data could be due to the incorporation of incorrect logical rules, missing interactions or even missing components in the literature-based sub-networks. Another important challenge with network inference and dynamical modeling methods is the fact that many alternative model realizations can fit the same data. Hence, the question remains, whether the model represents the real system under investigation.^{121,170,438} Furthermore, the cellular reality is far from being boolean. Although regulatory models are handicapped by unknowns, missing information on known relations, and the complexity of the biological processes, they are very well suited to move away from the intuitive interpretation of complex data such as expression profiles towards a rational, structured interpretation of experimental data. Models generate clear hypotheses to solve inconsistencies as shown in our work, eventually by precise perturbations using siRNA or similar approaches. As this and other studies have shown,^{128,305,439–443} despite its simplicity, properties derived from logical models can still provide valuable first insights into the transfer of signals in the cell. As more detailed (qualitative and quantitative) molecular information on Ctgf and the regulatory networks governing the HSC behavior within the micro-environment accumulates, this Boolean logical model may serve as a useful basis for the development of more complex dynamic models. In the meantime, further refinements could be possibly achieved by implementing fuzzy logic description or by considering more precise time delays (early vs. late events, modeled on different time scales) for the interactions and functional outcomes. Finally, we currently lack sophisticated experimental technologies for systematic profiling of the molecule levels and phosphorylation states to better understand the complexity of signaling networks. An ideal technology would have to be able to measure accurately the concentrations of hundred and even thousands of different proteins and their splice variants, each possibly subject to a variety of post-translational modifications and should be able to measure all this in a time-dependent, cell and compartment specific manner, under various conditions.²⁷¹

4.2.4 *Ctgf* and its down-stream targets in the regulation of cell cycle and hematopoiesis

In order to preserve the hematopoietic stem cell (HSC) pool throughout the adult life, it is crucial to maintain the vast majority of HSCs ($\sim 75\%$) at any given time in the quiescent (G0; dormant) phase of the cell cycle.^{6,9} HSC cell cycle status is thought to be precisely coordinated by a specific combination of 'niche' signals. In the present study, we identified a secreted matrix remodeller, connective tissue growth factor (*Ctgf*), as a putative novel regulator of early interaction events between HSC and stromal cells from our co-culture experiments of Lin-Sca1+cKit+ (LSK) and HSC-supportive UG26-1B6 stromal cells. By using RNAi perturbation of secreted *Ctgf* levels in UG26-1B6, we demonstrate that the absence of stromal *Ctgf* deficiency in stromal cells results in increased percentage of myeloid progenitor cells (MPs) and increased colony numbers *in vitro*, whereas the overall engraftment potential of sh*Ctgf* co-cultured cells was significantly reduced *in vivo*.

Nevertheless, on molecular level, it remains unclear, whether *Ctgf* is responsible for the G1/S transition or the block of this transition, due to controversial experimental results. On one hand, we observed up-regulation of the cell cycle inhibitor p27Kip1, but on the other hand, hyper-phosphorylation and inactivation of Rb was also detected. Of note, the retinoblastoma protein (Rb) exists in three general forms: unphosphorylated pRb, present in G0 cells and when pRb is newly synthesized; hypo-phosphorylated pRb, present in contact-inhibited cells and in early G1; and hyper-phosphorylated pRb, that is inactive and present in late G1, S, G2, and M phases of cycling cells. In early G1, the Cyclin D1:Cdk4/6 complexes convert unphosphorylated pRb to hypo-phosphorylated pRb. However, Cyclin D1:Cdk4/6 complexes are responsible for inactivating pRb by hyper-phosphorylation in late G1.³⁰⁹ The retinoblastoma protein Rb is thought to be critical for the regulation of mammalian cell cycle entry, where hypo-phosphorylated Rb is considered to be the active form and to direct G1 block, while hyper-phosphorylated Rb permits the transition from G1 to S phase for cell proliferation.⁴⁴⁴

The role *Ctgf* plays in the hematopoietic system is not well characterized. However, experimental evidence obtained in other cells and tissues, as well as by considering its down-stream targets and their role in the regulation of cell cycle and hematopoiesis, can yield clues about the biological effects of *Ctgf* and its molecular mechanisms. *Ctgf* is produced and secreted by osteoblasts and can regulate osteoblast development and function. In particular, osteoblast proliferation and differentiation.⁴⁴⁵ Osteoblasts play a key role within the hematopoietic stem cell niche, where osteoblastic cells (OBs) are a critical components for sustaining the slow-cycling HSC cell cycle kinetics through inter-cellular signaling pathways.⁴⁴⁶ This leads

4 Discussion

to overall increased survival of HSCs.^{447,448} Generally, loss of long-term hematopoietic stem cell function *in vitro* has been associated with cell cycle progression. Hematopoiesis can be optimally reconstituted by cells in G0 phase, in either G0 or G1 phase, or by cells which have undergone cell division and subsequently entered G1 (but not G0) phase, whereas cells in S, G2 or M phase have minimal engraftment potential.⁷³

In terms of the cell cycle regulation, contradicting reports on Ctgf can be found. On one hand, Ctgf has been described as a potent mitogen for fibroblasts and smooth muscle cells, controlling cell cycle progression through late G1 and S-phase entry.^{280,449–451} At the same time, Ctgf has been demonstrated to control the cell cycle restriction point in late G1, and being responsible for the cell cycle block in very late G1.⁴⁵² In the first scenario, Ctgf induced the S-phase entry by upregulating Cyclin A levels via reduction of p27Kip1, which resulted in hyper-phosphorylation of pRb and release of E2f.²⁸⁰ In the second case, where knocking down Ctgf expression stimulated cell proliferation and enhanced G1/S cell cycle transition, the activation of FAK/PI3K/Akt and its downstream signals regulating the cell cycle was found, including elevated activation of pRb (Ser-780), Cyclin D1 and E2f1, and suppressed expression of p15INK4 and p21Cip1, whereas the expression of Cdk4/6 were not affected.²⁸¹ Yet Ctgf can actually first stimulate the cells to exit from G0 and enter G1, however, they do not then progress further through the cell cycle, due to the induction of the cyclin-dependent kinase inhibitors p15INK4, p21Cip1 and p27Kip1.³¹⁵

The decrease in engraftment potential we observed in HSC/Ps recovered from Ctgf deficient microenvironment (Figure 3.12(E)), points to a possibly accelerated S phase entry. Similarly, increased performance in colony-forming cell (CFC) assay and elevated percentage of myeloid progenitor cells (MPs) is an indicative of increased cell proliferation and differentiation. However, we do not observe any significant differences in the BrdU assay. At the same time, it has been also demonstrated that Ctgf can also act as a growth inhibitor by inducing apoptosis,^{453–456} e.g. via up-regulating Caspase 3.⁴⁵⁴ Interestingly, in LSK cells co-cultured for 24 h with shCtgf stromal cells, we observed a significantly lower *Cdkn1a* mRNA and p21Cip1 protein levels, whereas p27Kip1 was elevated. Inverse correlated of p21Cip1 and p27Kip1 expression levels has been observed earlier.⁴⁵⁷ p27Kip1 and p21Cip1 were initially identified as members of the kinase inhibitor protein (KIP) family of cdk inhibitors, however, further studies showed that they also functions to mediate Cyclin D1:Cdk4/6 assembly and hence play dual roles to both promote and inhibit cell cycle progression.⁴⁵⁸ In particular, in G0, p27Kip1 translation and protein stability are maximal as it binds and inactivates nuclear Cyclin E:Cdk2, whereas in early G1, p27Kip1 promotes assembly and nuclear import of D-type Cyclin-CDKs. The progressive decrease of p27Kip1 in G1 permits Cyclin E:Cdk2 and Cyclin A:Cdk2 to activate the G1-S transition.⁴⁵⁹ For p21Cip1, it has been shown that this regulation

is concentration dependent: at low concentrations, p21Cip1 promotes the assembly of active kinase complexes, whereas at higher concentrations, it inhibits activity⁴⁵⁸. p27Kip1 levels can be affected by transcriptional and translational pathways, but the major mechanisms for p27Kip1 regulation are thought to be post-translational proteolytic degradation.⁴⁶⁰ Similarly, although most of the studies on p21Cip1 regulation have concentrated on its transcriptional regulation by p53-dependent and p53-independent mechanisms, p21Cip1 can also be regulated by post-translational mechanisms. p21Cip1 is an unstable short-lived protein with a half-life of <30 min and with a high proteasomal degradation rate, whereby it is degraded in a ubiquitin-independent manner.⁴⁶¹ At the same time, in endothelial cells, the focal adhesion kinase (FAK)-regulated degradation of p27Kip1 was Skp2 (an F-box protein that targets CDKs) dependent, while levels of p21Cip1 were regulated independent of Skp2 by inhibiting mitogen-induced mRNA.⁴⁶²

Moreover, it has been shown that p21Cip1 promotes assembly of Cyclin D1:Cdk4/6 and increases Cyclin D1 accumulation by direct inhibition of Gsk3- β -triggered nuclear export.⁴³⁰ Gsk3- β phosphorylated p21Cip1 specifically at Thr57 within the Cdk binding domain and overexpression of Gsk3- β decreased p21Cip1 protein levels, indicating that Gsk3- β triggers p21Cip1 degradation. In contrast, stimulation of Akt(PKB) increased p21Cip1 via inhibitory phosphorylation of Gsk3- β .³⁶³ Simultaneously, Gsk3- β activation prevents p27Kip1 degradation, whereas it decreases Cyclin D1 expression.³⁶⁴ These observations are in line with our experiments, indicating a decrease in Akt(PKB), p21Cip1 and Cyclin D1, whereas Gsk3- β and p27Kip1 are both up-regulated (Figure 3.19). In addition, another biological function of p27Kip1 is the protection of cells from apoptosis by constraining Cdk2 activity.⁴⁶³

In hematopoiesis, initially loss of p21Cip1 has been demonstrated to promote HSC entrance into the cell cycle and long-term loss of stem cells.⁹ However, more recent reports suggest that the function of p21Cip1 in regulating HSC cell cycle activity may be restricted to periods of stress rather than during homeostasis.^{75,76} Another member of CIP/KIP family of CKIs, p57Kip2, has also emerged as a critical regulator of HSC quiescence.⁷⁸ p27Kip1 was earlier reported to regulate the proliferation and pool size of more committed hematopoietic progenitor cells (HPCs).⁷⁷ More recently, however, a cooperation between p57Kip2 and p27Kip1 in the maintenance of hematopoietic stem cell quiescence has been suggested, with p27Kip1 compensating for p57Kip2 function, whereas only HSCs null for both p57Kip2 and p27Kip1, were more proliferative and had reduced capacity to engraft in transplantation.⁴⁶⁴ Considering, the TGF- β -induced cell cycle arrest in HPCs⁶⁰ and HSCs⁶³ is correlated with an increase in p57Kip2 expression, p57Kip2 could be included in our model and the expression of p57Kip2 remains to be profiled.

Glycogen synthase kinase-3 (Gsk3- β) regulates both Wnt and mTOR signaling in mouse

4 Discussion

HSCs, with these pathways promoting HSC self renewal and lineage commitment, respectively, such that inhibition of Gsk3- β in the presence of rapamycin expanded the HSC pool *in vivo*.⁴⁶⁵ Similarly, administration of a glycogen synthase kinase-3 inhibitor to recipient mice transplanted with mouse or human HSCs resulted in enhanced sustained long-term repopulation and improved neutrophil and megakaryocyte recovery.⁴⁶⁶ In our experiments, we observed activation of Gsk3- β in HSC/Ps after 24 h in Ctgf deficient environment (Figure 3.19), followed by decreased engraftment potential in cells recovered from 1 week co-cultures (Figure 3.12(E)), linking stromal Ctgf regulated-Gsk3- β activity to increased lineage commitment and hence reduced repopulating capacity of co-culture-derived HSC/Ps. Yet, molecular profiling of mTOR signaling may be necessary to further investigate this connection.

The tumour suppressor PTEN is known as a cell growth inhibitor: it induces apoptosis and cell cycle arrest through phosphoinositol-3-kinase/Akt-dependent and [U+2010]independent pathways.⁴⁶⁷⁻⁴⁷⁰ In particular, PTEN, induces G1 cell cycle arrest by reducing 3' phosphoinositides levels and inhibiting Akt(PKB) activity. Moreover, p27Kip1 is required for PTEN-induced G1 growth arrest, and expression of PTEN is associated with increased expression of p27Kip1.^{469,470} In addition, PTEN also decreases the level and nuclear localization of Cyclin D1.⁴⁶⁸ In agreement to this, we observe down-regulation of Akt/(PKB) and Cyclin D1, whereas PTEN and p27Kip1 are both activated, endorsing the G1/S block Boolean model predicted cell cycle block in the absence of stromal Ctgf (Figure 3.19 and Figure 3.17, right panel). However, these molecular events are commonly accompanied by a decrease in retinoblastoma (Rb) protein phosphorylation on Cyclin D/CDK4-specific sites, due to the negative effect of PTEN on Rb inactivation,⁴⁶⁸ whereas we observe the opposite: a significant increase in pRb-Ser780, which is a Cyclin D/CDK4-specific site (Figure 3.19). Strikingly, however, due to a considerable functional redundancy within Rb family, removal of pRb had no effect on HSC self-renewal as assessed by serial transplantation.⁷⁴ Hence its inactivation we observe in our experiments (Figure 3.19; p-Rb in row 6 and column 2) may have not functional consequences to shCtgf co-culture- derived HSC/Ps. Similar to the Rb proteins, mice deficient for a single D-cyclin, or for only one of the two associated Cdks, have minimal hematopoietic defects.⁷⁴ At the same time, PTEN has essential roles in restricting the activation of HSCs, in lineage fate determination, and in the prevention of leukaemogenesis,⁴⁷¹ as HSCs lacking expression of phosphatase and tensin homolog (Pten) exhibit activation of cell cycling.^{83,471,472}

Ctgf is known as a positive mediator of wound healing, in particular of fibrosis and scar formation, induced by and acting as a downstream effector of Tgf- β .⁴⁷³⁻⁴⁷⁵ In general, stromal wound healing involves activation of quiescent fibroblasts, migration of activated fibroblasts to the site of injury, and differentiation of fibroblasts into myofibroblasts.⁴⁷⁶⁻⁴⁷⁸

4.2 Identifying *Ctgf* as a novel regulator of hematopoiesis

Mice overexpressing *Ctgf* in fibroblasts are susceptible to accelerated tissue fibrosis affecting the skin, lung, kidney, and vasculature, most notably the small arteries. Interestingly, *Ctgf* also induced phosphorylation of Akt in these transgenic mouse fibroblasts,⁴⁷⁹ whereas *Gsk3- β* may oppose *Ctgf* in this process, since mice harboring a fibroblast-specific deletion of *Gsk3- β* exhibited accelerated wound closure, increased fibrogenesis, and excessive scarring compared with control mice.⁴⁸⁰ Loss of PTEN expression by dermal fibroblasts has also been reported to cause skin fibrosis. Moreover, PTEN-deleted fibroblasts showed elevated Akt phosphorylation and increased expression of connective tissue growth factor.⁴⁸¹

Furthermore, tumors have been designated as "wounds that do not heal", due to the similarities between tumor stroma generation and wound healing.⁴⁸² In leukemia, when comparing the gene expression profile of adult acute lymphoblastic leukemia (ALL) to normal hematopoietic and non-ALL samples using microarray analysis, *Ctgf* was the highest ***over-expressed gene in precursor B-acute lymphoblastic leukemia (ALL)*** compared with the other groups and that increased expression of *Ctgf* was associated with inferior outcome in B-ALL.²⁴⁵ More recently, Lu and Battula²⁸² have also characterized the functional role and downstream signaling pathways of *Ctgf* in ALL cells. Their experiments demonstrated that silencing of *Ctgf* resulted in significant suppression of leukemia cell growth, which was associated with Akt/mTOR inactivation and increased levels of cyclin-dependent kinase inhibitor p27Kip1. In addition, Battula and colleagues²⁷⁰ have also investigated the role of *Ctgf* in mesenchymal stromal cells (MSCs), a major component of the normal as well as leukemia bone marrow (BM) microenvironment, since connective tissue growth factor is highly expressed in MSCs. They found that *Ctgf*-KD human BM-derived MSCs exhibited fivefold lower proliferation compared with control MSCs and had markedly fewer S-phase cells. Moreover, *Ctgf* KD MSCs differentiated into adipocytes at a sixfold higher rate than controls *in vitro* and *in vivo*.

Hence, these reports suggest *Ctgf* as a positive regulator of cell growth and proliferation, which is also in line with our G1/S block Boolean sub-model prediction and experimental data, where, similarly to Lu and Battula,²⁸² we also observed that silencing of stromal *Ctgf* resulted in Akt(PKB) inactivation and increased levels of p27Kip1.

Finally, *Ctgf* has been reported to auto-induce its own expression.²⁸⁷ In particular, a β -catenin/TCF/LEF-binding site (TBE) was identified in the promoter region of *Ctgf* and it was found that *Ctgf* is a transcriptional target of β -catenin/TCF/LEF signaling, thus the cross-talk between *Ctgf* and the canonical Wnt signaling appears to form a positive feed-back loop.²⁷³ In addition, TGF- β induces *Ctgf* by the 'classical' SMAD pathway via a SMAD binding element located within the proximal promoter.²⁷⁵ Moreover, *Ctgf* is one of the genes that (in addition to p15Ink4b and p21Cip1) are jointly induced by SMAD and

4 Discussion

FoxO proteins.³⁵⁸ A significant decrease in both Ctgf mRNA (Figure 3.20; *Ctgf*) and protein (Figure 3.19; Ctgf) levels in LSK cells co-cultured for 24 h with Ctgf KD stromal cells suggests a positive feedback loop by which extrinsic Ctgf induces its own expression in LSK cells. This could be regarded as part of a probably much more complex process, by which hematopoietic stem cells may be themselves involved in the modulation of the hematopoietic niche in order to inhibit their own proliferation and exhaustion.

In summary, using our strategy we have successfully identified and validated a novel role for Ctgf in hematopoiesis. Despite a couple of inconsistencies between experimental results and the current knowledge from the literature, our work demonstrates the utility of Boolean logic-based models for explaining experimental observations and predicting phenotypic outcomes of signaling networks. We believe that the results of this study brings us closer to the comprehensive understanding of the molecular mechanisms regulating the function of HSC in steady-state and regenerative hematopoiesis, and will help to pinpoint possible targets of oncogenic transformation.

5 Conclusions

1. Gene expression time-series (TS) of LSK and UG26-1B6 stromal cells at three different time points of co-culture were generated and microarray results were independently confirmed by RT-qPCR, demonstrating $\sim 62\%$ sensitivity and $\sim 49\%$ specificity for the selected candidate genes.
2. Functional assays of freshly isolated vs. 24 h co-culture-derived LSK cells demonstrated no changes in the cell cycle status, however the progenitor generation capacity was decreased, whereas the engraftment potential was increased 16 weeks after transplantation.
3. Clustering analysis of TS data suggested that the most intense molecular cross-talk between LSK and UG26-1B6 cells occurred during the first 24 h of co-culture, whereas in stromal cells these changes may also be culture medium dependent. Gene function enrichment analysis revealed changes in cell adhesion and migration, TGF β signaling, metabolism, MAPK-regulated cell proliferation, as well as epigenetic regulation of gene silencing.
4. Candidate gene prioritisation using a training set of hematopoiesis-related genes ranked high, a secreted 36-38 kDa cysteine-rich matrix remodeller, Connective tissue growth factor (Ctgf). Its interactome analysis revealed association with mouse phenotypes related to abnormal hematopoiesis and topological property analysis and module discovery in the Ctgf interactome indicated modular structure and small-world properties.
5. Phenotypic and functional comparison of shCtgf vs pLK0.1 1 week co-culture-derived HSC/P cells revealed increased numbers of MPP cells, as well as colony-forming cells (CFCs) in methylcellulose, as well as reduced donor cell compartment in the peripheral blood (PB) 16 weeks after transplantation.
6. A bottom-up approach was used to construct a literature-based network model predicting the possible role of Ctgf in (1) G0/G1 by inducing Cyclin D1, (2) G1/S block by inducing p21Cip1 and/or p27Kip1, as well as its (3) auto-induction in LSK cells. Dynamic network analysis using the Boolean logic pointed to a possible association between Ctgf and Akt(PKB)/PTEN, GSK3- β and β -catenin activity, which also could be confirmed experimentally.

Bibliography

1. Rufer, N. *et al.* Telomere fluorescence measurements in granulocytes and t lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory t cells in early childhood. *J Exp Med* **190**, 157–167 (1999).
2. Uchida, N. & Weissman, I. L. Searching for hematopoietic stem cells: evidence that thy-1.1lo lin- sca-1+ cells are the only stem cells in c57bl/ka-thy-1.1 bone marrow. *J Exp Med* **175**, 175–184 (1992).
3. Inaba, M. & Yamashita, Y. M. Asymmetric stem cell division: Precision for robustness. *Cell Stem Cell* (2012).
4. Han, W., Yu, Y. & Liu, X. Y. Local signals in stem cell-based bone marrow regeneration. *Cell Res* **16**, 189–195 (2006). URL <http://dx.doi.org/10.1038/sj.cr.7310026>.
5. Bradford, G. B., Williams, B., Rossi, R. & Bertoncello, I. Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. *Exp Hematol* **25**, 445–453 (1997).
6. Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc Natl Acad Sci U S A* **96**, 3120–3125 (1999).
7. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004). URL <http://dx.doi.org/10.1016/j.cell.2004.07.004>.
8. Qian, H. *et al.* Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* **1**, 671–684 (2007). URL <http://dx.doi.org/10.1016/j.stem.2007.10.008>.
9. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).

10. Chotinantakul, K. & Leraanansaksiri, W. Hematopoietic stem cell development, niches, and signaling pathways. *Bone Marrow Res* **2012**, 270425 (2012). URL <http://dx.doi.org/10.1155/2012/270425>.
11. Eckfeldt, C. E., Mendenhall, E. M. & Verfaillie, C. M. The molecular repertoire of the 'almighty' stem cell. *Nat Rev Mol Cell Biol* **6**, 726–737 (2005). URL <http://dx.doi.org/10.1038/nrm1713>.
12. Jung, Y. *et al.* Hematopoietic stem cells regulate mesenchymal stromal cell induction into osteoblasts thereby participating in the formation of the stem cell niche. *Stem Cells* **26**, 2042–2051 (2008). URL <http://dx.doi.org/10.1634/stemcells.2008-0149>.
13. Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat Rev Genet* **9**, 115–128 (2008). URL <http://dx.doi.org/10.1038/nrg2269>.
14. Kent, D. G. *et al.* Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* **113**, 6342–6350 (2009). URL <http://dx.doi.org/10.1182/blood-2008-12-192054>.
15. Song, Y. *et al.* Stem cell traits in long-term co-culture revealed by time-lapse imaging. *Leukemia* **24**, 153–161 (2010). URL <http://dx.doi.org/10.1038/leu.2009.191>.
16. Kevin S. Tieu, J. A. M.-A., Ryan S. Tieu & Sehl, M. E. Stem cell niche dynamics: From homeostasis to carcinogenesis. *Stem Cells International* **Volume 2012 (2012)**, 9 p. (2012).
17. Forsberg, E. C. *et al.* Differential expression of novel potential regulators in hematopoietic stem cells. *PLoS Genet* **1**, e28 (2005). URL <http://dx.doi.org/10.1371/journal.pgen.0010028>.
18. Bryder, D., Rossi, D. J. & Weissman, I. L. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am J Pathol* **169**, 338–346 (2006). URL <http://dx.doi.org/10.2353/ajpath.2006.060312>.
19. Galan-Caridad, J. M. *et al.* Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* **129**, 345–357 (2007). URL <http://dx.doi.org/10.1016/j.cell.2007.03.014>.
20. Miranda-Saavedra, D., De, S., Trotter, M. W., Teichmann, S. A. & Gtgens, B. Blood-express: a database of gene expression in mouse haematopoiesis. *Nucleic Acids Res* **37**, D873–D879 (2009). URL <http://dx.doi.org/10.1093/nar/gkn854>.

Bibliography

21. ten Cate, B., de Bruyn, M., Wei, Y., Bremer, E. & Helfrich, W. Targeted elimination of leukemia stem cells; a new therapeutic approach in hemato-oncology. *Curr Drug Targets* **11**, 95–110 (2010).
22. Rizo, A., Vellenga, E., de Haan, G. & Schuringa, J. J. Signaling pathways in self-renewing hematopoietic and leukemic stem cells: do all stem cells need a niche? *Hum Mol Genet* **15 Spec No 2**, R210–R219 (2006). URL <http://dx.doi.org/10.1093/hmg/ddl175>.
23. Cheung, A. M. S., Kwong, Y.-L., Liang, R. & Leung, A. Y. H. Stem cell model of hematopoiesis. *Curr Stem Cell Res Ther* **1**, 305–315 (2006).
24. Charbord, P. & Moore, K. Gene expression in stem cell-supporting stromal cell lines. *Ann N Y Acad Sci* **1044**, 159–167 (2005). URL <http://dx.doi.org/10.1196/annals.1349.020>.
25. Schuster, J. A. *et al.* Expansion of hematopoietic stem cells for transplantation: current perspectives. *Exp Hematol Oncol* **1**, 12 (2012). URL <http://dx.doi.org/10.1186/2162-3619-1-12>.
26. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
27. Vaughan, J. Osteogenesis and haematopoiesis. *Lancet* **2**, 133–136 (1981).
28. Xie, T. & Spradling, A. C. decapentaplegic is essential for the maintenance and division of germline stem cells in the drosophila ovary. *Cell* **94**, 251–260 (1998).
29. Fuchs, E., Tumber, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* **116**, 769–778 (2004).
30. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598–611 (2008). URL <http://dx.doi.org/10.1016/j.cell.2008.01.038>.
31. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003). URL <http://dx.doi.org/10.1038/nature02040>.
32. Kiel, M. J. *et al.* Slam family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005). URL <http://dx.doi.org/10.1016/j.cell.2005.05.026>.

33. Oh, I.-H. & Humphries, R. K. Concise review: Multidimensional regulation of the hematopoietic stem cell state. *Stem Cells* **30**, 82–88 (2012). URL <http://dx.doi.org/10.1002/stem.776>.
34. Yin, T. & Li, L. The stem cell niches in bone. *J Clin Invest* **116**, 1195–1201 (2006). URL <http://dx.doi.org/10.1172/JCI28568>.
35. Celso, C. L. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457**, 92–96 (2009). URL <http://dx.doi.org/10.1038/nature07434>.
36. Kiel, M. J. & Morrison, S. J. Uncertainty in the niches that maintain haematopoietic stem cells. *Nat Rev Immunol* **8**, 290–301 (2008). URL <http://dx.doi.org/10.1038/nri2279>.
37. Nakamura, Y. *et al.* Isolation and characterization of endosteal niche cell populations that regulate hematopoietic stem cells. *Blood* **116**, 1422–1432 (2010). URL <http://dx.doi.org/10.1182/blood-2009-08-239194>.
38. Shin, J.-W. *et al.* Contractile forces sustain and polarize hematopoiesis from stem and progenitor cells. *Cell Stem Cell* (2013). URL <http://dx.doi.org/10.1016/j.stem.2013.10.009>.
39. Keung, A. J., Healy, K. E., Kumar, S. & Schaffer, D. V. Biophysics and dynamics of natural and engineered stem cell microenvironments. *Wiley Interdiscip Rev Syst Biol Med* **2**, 49–64 (2010). URL <http://dx.doi.org/10.1002/wsbm.46>.
40. Jeong, M. *et al.* Thioredoxin-interacting protein regulates hematopoietic stem cell quiescence and mobilization under stress conditions. *J Immunol* **183**, 2495–2505 (2009). URL <http://dx.doi.org/10.4049/jimmunol.0804221>.
41. O’Connell, R. M. *et al.* Micornas enriched in hematopoietic stem cells differentially regulate long-term hematopoietic output. *Proc Natl Acad Sci U S A* **107**, 14235–14240 (2010). URL <http://dx.doi.org/10.1073/pnas.1009798107>.
42. Grabher, C. *et al.* Zebrafish microrna-126 determines hematopoietic cell fate through c-myb. *Leukemia* (2010). URL <http://dx.doi.org/10.1038/leu.2010.280>.
43. Yagi, M. *et al.* Sustained ex vivo expansion of hematopoietic stem cells mediated by thrombopoietin. *Proc Natl Acad Sci U S A* **96**, 8126–8131 (1999).

Bibliography

44. Alexander, W. S., Roberts, A. W., Nicola, N. A., Li, R. & Metcalf, D. Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-mpl. *Blood* **87**, 2162–2170 (1996).
45. Carver-Moore, K. *et al.* Low levels of erythroid and myeloid progenitors in thrombopoietin-and c-mpl-deficient mice. *Blood* **88**, 803–808 (1996).
46. Kimura, S., Roberts, A. W., Metcalf, D. & Alexander, W. S. Hematopoietic stem cell deficiencies in mice lacking c-mpl, the receptor for thrombopoietin. *Proc Natl Acad Sci U S A* **95**, 1195–1200 (1998).
47. Yoshihara, H. *et al.* Thrombopoietin/mpl signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. *Cell Stem Cell* **1**, 685–697 (2007). URL <http://dx.doi.org/10.1016/j.stem.2007.10.020>.
48. Reya, T. Regulation of hematopoietic stem cell self-renewal. *Recent Prog Horm Res* **58**, 283–295 (2003).
49. Kirstetter, P., Anderson, K., Porse, B. T., Jacobsen, S. E. W. & Nerlov, C. Activation of the canonical wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. *Nat Immunol* **7**, 1048–1056 (2006). URL <http://dx.doi.org/10.1038/ni1381>.
50. Scheller, M. *et al.* Hematopoietic stem cell and multilineage defects generated by constitutive beta-catenin activation. *Nat Immunol* **7**, 1037–1047 (2006). URL <http://dx.doi.org/10.1038/ni1387>.
51. Renström, J. *et al.* Secreted frizzled-related protein 1 extrinsically regulates cycling activity and maintenance of hematopoietic stem cells. *Cell Stem Cell* **5**, 157–167 (2009). URL <http://dx.doi.org/10.1016/j.stem.2009.05.020>.
52. Suda, T. & Arai, F. Wnt signaling in the niche. *Cell* **132**, 729–730 (2008). URL <http://dx.doi.org/10.1016/j.cell.2008.02.017>.
53. Capron, C. *et al.* A major role of tgf-beta1 in the homing capacities of murine hematopoietic stem cell/progenitors. *Blood* **116**, 1244–1253 (2010). URL <http://dx.doi.org/10.1182/blood-2009-05-221093>.
54. Marshall, C. J., Kinnon, C. & Thrasher, A. J. Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. *Blood* **96**, 1591–1593 (2000).

55. Johansson, B. M. & Wiles, M. V. Evidence for involvement of activin a and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. *Mol Cell Biol* **15**, 141–151 (1995).
56. Bhatia, M. *et al.* Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. *J Exp Med* **189**, 1139–1148 (1999).
57. Goldman, D. C. *et al.* Bmp4 regulates the hematopoietic stem cell niche. *Blood* **114**, 4393–4401 (2009). URL <http://dx.doi.org/10.1182/blood-2009-02-206433>.
58. Keller, J. R. *et al.* Transforming growth factor beta directly regulates primitive murine hematopoietic cell proliferation. *Blood* **75**, 596–602 (1990).
59. Ohta, M., Greenberger, J. S., Anklesaria, P., Bassols, A. & Massagué, J. Two forms of transforming growth factor-beta distinguished by multipotential haematopoietic progenitor cells. *Nature* **329**, 539–541 (1987). URL <http://dx.doi.org/10.1038/329539a0>.
60. Scandura, J. M., Bocconi, P., Massagué, J. & Nimer, S. D. Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57kip2 up-regulation. *Proc Natl Acad Sci U S A* **101**, 15231–15236 (2004). URL <http://dx.doi.org/10.1073/pnas.0406771101>.
61. Ducos, K. *et al.* p21(cip1) mrna is controlled by endogenous transforming growth factor-beta1 in quiescent human hematopoietic stem/progenitor cells. *J Cell Physiol* **184**, 80–85 (2000). URL <http://dx.doi.org/3.0.CO;2-Q>.
62. Ewen, M. E., Sluss, H. K., Whitehouse, L. L. & Livingston, D. M. Tgf beta inhibition of cdk4 synthesis is linked to cell cycle arrest. *Cell* **74**, 1009–1020 (1993).
63. Yamazaki, S. *et al.* Tgf-beta as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation. *Blood* **113**, 1250–1256 (2009). URL <http://dx.doi.org/10.1182/blood-2008-04-146480>.
64. Yamashita, Y. M., Jones, D. L. & Fuller, M. T. Orientation of asymmetric stem cell division by the apc tumor suppressor and centrosome. *Science* **301**, 1547–1550 (2003). URL <http://dx.doi.org/10.1126/science.1087795>.
65. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003). URL <http://dx.doi.org/10.1038/nature02041>.

Bibliography

66. Wilson, A. *et al.* c-myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18**, 2747–2763 (2004). URL <http://dx.doi.org/10.1101/gad.313104>.
67. Nilsson, S. K. *et al.* Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood* **106**, 1232–1239 (2005). URL <http://dx.doi.org/10.1182/blood-2004-11-4422>.
68. Stier, S. *et al.* Osteopontin is a hematopoietic stem cell niche component that negatively regulates stem cell pool size. *J Exp Med* **201**, 1781–1791 (2005). URL <http://dx.doi.org/10.1084/jem.20041992>.
69. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98**, 4569–4574 (2001). URL <http://dx.doi.org/10.1073/pnas.061034498>.
70. Jang, Y.-Y. & Sharkis, S. J. A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. *Blood* **110**, 3056–3063 (2007). URL <http://dx.doi.org/10.1182/blood-2007-05-087759>.
71. Purton, L. E., Bernstein, I. D. & Collins, S. J. All-trans retinoic acid enhances the long-term repopulating activity of cultured hematopoietic stem cells. *Blood* **95**, 470–477 (2000).
72. Adams, G. B. *et al.* Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* **439**, 599–603 (2006). URL <http://dx.doi.org/10.1038/nature04247>.
73. Steinman, R. A. Cell cycle regulators and hematopoiesis. *Oncogene* **21**, 3403–3413 (2002). URL <http://dx.doi.org/10.1038/sj.onc.1205325>.
74. Pietras, E. M., Warr, M. R. & Passegué, E. Cell cycle regulation in hematopoietic stem cells. *J Cell Biol* **195**, 709–720 (2011). URL <http://dx.doi.org/10.1083/jcb.201102131>.
75. Foudi, A. *et al.* Analysis of histone 2b-gfp retention reveals slowly cycling hematopoietic stem cells. *Nat Biotechnol* **27**, 84–90 (2009). URL <http://dx.doi.org/10.1038/nbt.1517>.
76. van Os, R. *et al.* A limited role for p21cip1/waf1 in maintaining normal hematopoietic stem cell functioning. *Stem Cells* **25**, 836–843 (2007). URL <http://dx.doi.org/10.1634/stemcells.2006-0631>.

77. Cheng, T., Rodrigues, N., Dombkowski, D., Stier, S. & Scadden, D. T. Stem cell repopulation efficiency but not pool size is governed by p27(kip1). *Nat Med* **6**, 1235–1240 (2000). URL <http://dx.doi.org/10.1038/81335>.
78. Matsumoto, A. *et al.* p57 is required for quiescence and maintenance of adult hematopoietic stem cells. *Cell Stem Cell* **9**, 262–271 (2011). URL <http://dx.doi.org/10.1016/j.stem.2011.06.014>.
79. Hannah, R., Joshi, A., Wilson, N. K., Kinston, S. & Gtgens, B. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Exp Hematol* **39**, 531–541 (2011). URL <http://dx.doi.org/10.1016/j.expphem.2011.02.009>.
80. Stein, M. I., Zhu, J. & Emerson, S. G. Molecular pathways regulating the self-renewal of hematopoietic stem cells. *Exp Hematol* **32**, 1129–1136 (2004). URL <http://dx.doi.org/10.1016/j.expphem.2004.08.012>.
81. Yu, S. *et al.* Gabp controls a critical transcription regulatory module that is essential for maintenance and differentiation of hematopoietic stem/progenitor cells. *Blood* **117**, 2166–2178 (2011). URL <http://dx.doi.org/10.1182/blood-2010-09-306563>.
82. Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* **431**, 1002–1007 (2004). URL <http://dx.doi.org/10.1038/nature02994>.
83. Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006). URL <http://dx.doi.org/10.1038/nature04703>.
84. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001). URL <http://dx.doi.org/10.1126/science.1063443>.
85. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245–254 (2003). URL <http://dx.doi.org/10.1038/ng1089>.
86. Akashi, K. *et al.* Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003). URL <http://dx.doi.org/10.1182/blood-2002-06-1780>.

Bibliography

87. Tagoh, H. *et al.* Transcription factor complex formation and chromatin fine structure alterations at the murine *c-fms* (*csf-1* receptor) locus during maturation of myeloid precursor cells. *Genes Dev* **16**, 1721–1737 (2002). URL <http://dx.doi.org/10.1101/gad.222002>.
88. Bruno, L. *et al.* Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. *Mol Cell Biol* **24**, 741–756 (2004).
89. Attema, J. L. *et al.* Epigenetic characterization of hematopoietic stem cell differentiation using minichip and bisulfite sequencing analysis. *Proc Natl Acad Sci U S A* **104**, 12371–12376 (2007). URL <http://dx.doi.org/10.1073/pnas.0704468104>.
90. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010). URL <http://dx.doi.org/10.1038/nature09367>.
91. Chung, Y. S. *et al.* Undifferentiated hematopoietic cells are characterized by a genome-wide undermethylation dip around the transcription start site and a hierarchical epigenetic plasticity. *Blood* **114**, 4968–4978 (2009). URL <http://dx.doi.org/10.1182/blood-2009-01-197780>.
92. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011). URL <http://dx.doi.org/10.1016/j.cell.2011.01.032>.
93. Shilatifard, A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* **75**, 243–269 (2006). URL <http://dx.doi.org/10.1146/annurev.biochem.75.103004.142422>.
94. Xie, H. *et al.* Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **Volume 14**, 68–80 (2013).
95. Radulovic, V., de Haan, G. & Klauke, K. Polycomb-group proteins in hematopoietic stem cell regulation and hematopoietic neoplasms. *Leukemia* **27**, 523–533 (2013). URL <http://dx.doi.org/10.1038/leu.2012.368>.
96. Bibikova, M., Laurent, L. C., Ren, B., Loring, J. F. & Fan, J.-B. Unraveling epigenetic regulation in embryonic stem cells. *Cell Stem Cell* **2**, 123–134 (2008). URL <http://dx.doi.org/10.1016/j.stem.2008.01.005>.

97. Jude, C. D. *et al.* Unique and independent roles for *mll* in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* **1**, 324–337 (2007). URL <http://dx.doi.org/10.1016/j.stem.2007.05.019>.
98. Challen, G. A. *et al.* *Dnmt3a* is essential for hematopoietic stem cell differentiation. *Nat Genet* **44**, 23–31 (2012). URL <http://dx.doi.org/10.1038/ng.1009>.
99. Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. Dna methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell* **5**, 442–449 (2009). URL <http://dx.doi.org/10.1016/j.stem.2009.08.016>.
100. Guo, S. *et al.* MicroRNA *mir-125a* controls hematopoietic stem cell number. *Proc Natl Acad Sci U S A* **107**, 14229–14234 (2010). URL <http://dx.doi.org/10.1073/pnas.0913574107>.
101. Lechman, E. R. *et al.* Attenuation of *mir-126* activity expands hsc in vivo without exhaustion. *Cell Stem Cell* (2012). URL <http://dx.doi.org/10.1016/j.stem.2012.09.001>.
102. Yamazaki, S. *et al.* Cytokine signals modulated via lipid rafts mimic niche signals and induce hibernation in hematopoietic stem cells. *EMBO J* **25**, 3515–3523 (2006). URL <http://dx.doi.org/10.1038/sj.emboj.7601236>.
103. Yamazaki, S. *et al.* Cytokine signaling, lipid raft clustering, and hsc hibernation. *Ann N Y Acad Sci* **1106**, 54–63 (2007). URL <http://dx.doi.org/10.1196/annals.1392.017>.
104. Larsson, J. *et al.* *Nf2/merlin* regulates hematopoietic stem cell behavior by altering microenvironmental architecture. *Cell Stem Cell* **3**, 221–227 (2008). URL <http://dx.doi.org/10.1016/j.stem.2008.06.005>.
105. Fleming, H. E. *et al.* Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. *Cell Stem Cell* **2**, 274–283 (2008). URL <http://dx.doi.org/10.1016/j.stem.2008.01.003>.
106. Wilson, A., Laurenti, E. & Trumpp, A. Balancing dormant and self-renewing hematopoietic stem cells. *Curr Opin Genet Dev* **19**, 461–468 (2009). URL <http://dx.doi.org/10.1016/j.gde.2009.08.005>.
107. Nagasawa, T. *et al.* Defects of b-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the cxc chemokine *pbsf/sdf-1*. *Nature* **382**, 635–638 (1996). URL <http://dx.doi.org/10.1038/382635a0>.

Bibliography

108. Puri, M. C. & Bernstein, A. Requirement for the tie family of receptor tyrosine kinases in adult but not fetal hematopoiesis. *Proc Natl Acad Sci U S A* **100**, 12753–12758 (2003). URL <http://dx.doi.org/10.1073/pnas.2133552100>.
109. Nie, Y., Han, Y.-C. & Zou, Y.-R. Cxcr4 is required for the quiescence of primitive hematopoietic cells. *J Exp Med* **205**, 777–783 (2008). URL <http://dx.doi.org/10.1084/jem.20072513>.
110. Thorén, L. A. *et al.* Kit regulates maintenance of quiescent hematopoietic stem cells. *J Immunol* **180**, 2045–2053 (2008).
111. Purton, L. E. *et al.* Rargamma is critical for maintaining a balance between hematopoietic stem cell self-renewal and differentiation. *J Exp Med* **203**, 1283–1293 (2006). URL <http://dx.doi.org/10.1084/jem.20052105>.
112. Ito, K. *et al.* A pml-ppar-delta pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nat Med* (2012). URL <http://dx.doi.org/10.1038/nm.2882>.
113. McLeod, C. J., Wang, L., Wong, C. & Jones, D. L. Stem cell dynamics in response to nutrient availability. *Curr Biol* **20**, 2100–2105 (2010). URL <http://dx.doi.org/10.1016/j.cub.2010.10.038>.
114. Tothova, Z. *et al.* Foxos are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007). URL <http://dx.doi.org/10.1016/j.cell.2007.01.003>.
115. Prosper, F. & Verfaillie, C. M. Regulation of hematopoiesis through adhesion receptors. *J Leukoc Biol* **69**, 307–316 (2001).
116. Thomas, X. & Anglaret, B. [cell adhesion molecules: expression and function in acute myeloid leukemia]. *Bull Cancer* **86**, 265–277 (1999).
117. Verfaillie, C. M., Hurley, R., Lundell, B. I., Zhao, C. & Bhatia, R. Integrin-mediated regulation of hematopoiesis: do bcr/abl-induced defects in integrin function underlie the abnormal circulation and proliferation of cml progenitors? *Acta Haematol* **97**, 40–52 (1997).
118. Peltier, J. & Schaffer, D. V. Systems biology approaches to understanding stem cell fate choice. *IET Syst Biol* **4**, 1–11 (2010). URL <http://dx.doi.org/10.1049/iet-syb.2009.0011>.

119. Wang, R.-S., Saadatpour, A. & Albert, R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol* **9**, 055001 (2012). URL <http://dx.doi.org/10.1088/1478-3975/9/5/055001>.
120. Whichard, Z. L., Sarkar, C. A., Kimmel, M. & Corey, S. J. Hematopoiesis and its disorders: a systems biology approach. *Blood* **115**, 2339–2347 (2010). URL <http://dx.doi.org/10.1182/blood-2009-08-215798>.
121. Ma'ayan, A. Network integration and graph analysis in mammalian molecular systems biology. *IET Syst Biol* **2**, 206–221 (2008). URL <http://dx.doi.org/10.1049/iet-syb:20070075>.
122. Zhang, J. & Li, L. Stem cell niche: microenvironment and beyond. *J Biol Chem* **283**, 9499–9503 (2008). URL <http://dx.doi.org/10.1074/jbc.R700043200>.
123. Bauer-Mehren, A., Furlong, L. I. & Sanz, F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* **5**, 290 (2009). URL <http://dx.doi.org/10.1038/msb.2009.47>.
124. Gonzalez, A. G., Naldi, A., Sánchez, L., Thieffry, D. & Chaouiya, C. Ginsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems* **84**, 91–100 (2006). URL <http://dx.doi.org/10.1016/j.biosystems.2005.10.003>.
125. Zielinski, R. *et al.* The crosstalk between egf, igf, and insulin cell signaling pathways—computational and experimental analysis. *BMC Syst Biol* **3**, 88 (2009). URL <http://dx.doi.org/10.1186/1752-0509-3-88>.
126. Przytycka, T. M. & Kim, Y.-A. Network integration meets network dynamics. *BMC Biol* **8**, 48 (2010). URL <http://dx.doi.org/10.1186/1741-7007-8-48>.
127. Saez-Rodriguez, J. *et al.* A logical model provides insights into t cell receptor signaling. *PLoS Comput Biol* **3**, e163 (2007). URL <http://dx.doi.org/10.1371/journal.pcbi.0030163>.
128. Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K. & Klamt, S. The logic of egfr/erbb signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol* **5**, e1000438 (2009). URL <http://dx.doi.org/10.1371/journal.pcbi.1000438>.

Bibliography

129. Schlatter, R. *et al.* Integration of boolean models exemplified on hepatocyte signal transduction. *Brief Bioinform* **13**, 365–376 (2012). URL <http://dx.doi.org/10.1093/bib/bbr065>.
130. Schlatter, R. *et al.* On/off and beyond—a boolean model of apoptosis. *PLoS Comput Biol* **5**, e1000595 (2009). URL <http://dx.doi.org/10.1371/journal.pcbi.1000595>.
131. Puig, O. *et al.* The tandem affinity purification (tap) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001). URL <http://dx.doi.org/10.1006/meth.2001.1183>.
132. Stoll, D., Templin, M. F., Bachmann, J. & Joos, T. O. Protein microarrays: applications and future challenges. *Curr Opin Drug Discov Devel* **8**, 239–252 (2005).
133. Nolan, J. P. & Yang, L. The flow of cytometry into systems biology. *Brief Funct Genomic Proteomic* **6**, 81–90 (2007). URL <http://dx.doi.org/10.1093/bfpg/elm011>.
134. Chatranyamontri, A. *et al.* Mint: the molecular interaction database. *Nucleic Acids Res* **35**, D572–D574 (2007). URL <http://dx.doi.org/10.1093/nar/gkl950>.
135. Hermjakob, H. *et al.* Intact: an open source molecular interaction database. *Nucleic Acids Res* **32**, D452–D455 (2004). URL <http://dx.doi.org/10.1093/nar/gkh052>.
136. Xia, K., Dong, D. & Han, J.-D. J. Intnetdb v1.0: an integrated protein-protein interaction network database generated by a probabilistic model. *BMC Bioinformatics* **7**, 508 (2006). URL <http://dx.doi.org/10.1186/1471-2105-7-508>.
137. von Mering, C. *et al.* String 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358–D362 (2007). URL <http://dx.doi.org/10.1093/nar/gkl825>.
138. Cerami, E. G. *et al.* Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res* **39**, D685–D690 (2011). URL <http://dx.doi.org/10.1093/nar/gkq1039>.
139. Chaurasia, G. *et al.* Unihi: an entry gate to the human protein interactome. *Nucleic Acids Res* **35**, D590–D594 (2007). URL <http://dx.doi.org/10.1093/nar/gkl817>.
140. Lee, T. J. *et al.* Biowarehouse: a bioinformatics database warehouse toolkit. *BMC Bioinformatics* **7**, 170 (2006). URL <http://dx.doi.org/10.1186/1471-2105-7-170>.

141. Shannon, P. T., Reiss, D. J., Bonneau, R. & Baliga, N. S. The gaggles: an open-source software system for integrating bioinformatics software and data sources. *BMC Bioinformatics* **7**, 176 (2006). URL <http://dx.doi.org/10.1186/1471-2105-7-176>.
142. Futschik, M. E., Chaurasia, G. & Herzel, H. Comparison of human protein-protein interaction maps. *Bioinformatics* **23**, 605–611 (2007). URL <http://dx.doi.org/10.1093/bioinformatics/btl683>.
143. Chaurasia, G., Herzel, H., Wanker, E. E. & Futschik, M. E. Systematic functional assessment of human protein-protein interaction maps. *Genome Inform* **17**, 36–45 (2006).
144. Pavlopoulos, G. A. *et al.* Using graph theory to analyze biological networks. *BioData Min* **4**, 10 (2011). URL <http://dx.doi.org/10.1186/1756-0381-4-10>.
145. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. Jaspar: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91–D94 (2004). URL <http://dx.doi.org/10.1093/nar/gkh012>.
146. Matys, V. *et al.* Transfac and its module transcompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**, D108–D110 (2006). URL <http://dx.doi.org/10.1093/nar/gkj143>.
147. Diella, F. *et al.* Phospho.elm: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**, 79 (2004). URL <http://dx.doi.org/10.1186/1471-2105-5-79>.
148. Gnad, F. *et al.* Phosida (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* **8**, R250 (2007). URL <http://dx.doi.org/10.1186/gb-2007-8-11-r250>.
149. Kanehisa, M. & Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
150. Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B.Ø. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7**, 129–143 (2009). URL <http://dx.doi.org/10.1038/nrmicro1949>.
151. Ma, H. *et al.* The edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* **3**, 135 (2007). URL <http://dx.doi.org/10.1038/msb4100177>.

Bibliography

152. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**, D355–D360 (2010). URL <http://dx.doi.org/10.1093/nar/gkp896>.
153. Karp, P. D. *et al.* Expansion of the biocyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* **33**, 6083–6089 (2005). URL <http://dx.doi.org/10.1093/nar/gki892>.
154. Whitaker, J. W., Letunic, I., McConkey, G. A. & Westhead, D. R. metatiger: a metabolic evolution resource. *Nucleic Acids Res* **37**, D531–D538 (2009). URL <http://dx.doi.org/10.1093/nar/gkn826>.
155. Schilling, C. H., Schuster, S., Palsson, B. O. & Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog* **15**, 296–303 (1999). URL <http://dx.doi.org/10.1021/bp990048k>.
156. Schuster, S., Dandekar, T. & Fell, D. A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol* **17**, 53–60 (1999).
157. Shalgi, R., Lieber, D., Oren, M. & Pilpel, Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol* **3**, e131 (2007). URL <http://dx.doi.org/10.1371/journal.pcbi.0030131>.
158. Cui, Q., Yu, Z., Purisima, E. O. & Wang, E. Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* **2**, 46 (2006). URL <http://dx.doi.org/10.1038/msb4100089>.
159. Hucka, M. *et al.* The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
160. Hermjakob, H. *et al.* The hupo psi’s molecular interaction format—a community standard for the representation of protein interaction data. *Nat Biotechnol* **22**, 177–183 (2004). URL <http://dx.doi.org/10.1038/nbt926>.
161. Peter Murray-Rust, H. S. R. & Wright, M. Development of chemical markup language (cml) as a system for handling complex chemical content. *New Journal of Chemistry* **25**, 618–634 (2001).
162. group, B. W. *BioPAX-biological pathways exchange language* (2004).

163. Lloyd, C. M., Halstead, M. D. B. & Nielsen, P. F. Cellml: its future, present and past. *Prog Biophys Mol Biol* **85**, 433–450 (2004). URL <http://dx.doi.org/10.1016/j.pbiomolbio.2004.01.004>.
164. Boxwala, A. A. *et al.* Glif3: a representation format for sharable computer-interpretable clinical practice guidelines. *J Biomed Inform* **37**, 147–161 (2004). URL <http://dx.doi.org/10.1016/j.jbi.2004.04.002>.
165. Keating, S. M., Bornstein, B. J., Finney, A. & Hucka, M. Sbmtoolbox: an sbml toolbox for matlab users. *Bioinformatics* **22**, 1275–1277 (2006). URL <http://dx.doi.org/10.1093/bioinformatics/btl1111>.
166. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003). URL <http://dx.doi.org/10.1101/gr.1239303>.
167. Funahashi A, M. M. K. H., Tanimura N. Celldesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* **1**, 159–162 (2003).
168. Boon-Siew Seah, C. F. D., Sourav S Bhowmick & Yu, H. Fuse: a profit maximization approach for functional summarization of biological networks. *BMC Bioinformatics* **13** (2012).
169. Klamt, S., Saez-Rodriguez, J., Lindquist, J. A., Simeoni, L. & Gilles, E. D. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* **7**, 56 (2006). URL <http://dx.doi.org/10.1186/1471-2105-7-56>.
170. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. *Genes Dev* **21**, 1010–1024 (2007). URL <http://dx.doi.org/10.1101/gad.1528707>.
171. Yoon, J., Blumer, A. & Lee, K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics* **22**, 3106–3108 (2006). URL <http://dx.doi.org/10.1093/bioinformatics/btl533>.
172. Andreopoulos, B., An, A., Wang, X., Faloutsos, M. & Schroeder, M. Clustering by common friends finds locally significant proteins mediating modules. *Bioinformatics* **23**, 1124–1131 (2007). URL <http://dx.doi.org/10.1093/bioinformatics/btm064>.
173. Lubovac, Z., Gamalielsson, J. & Olsson, B. Combining functional and topological properties to identify core modules in protein interaction networks. *Proteins* **64**, 948–959 (2006). URL <http://dx.doi.org/10.1002/prot.21071>.

Bibliography

174. Luo, F. *et al.* Modular organization of protein interaction networks. *Bioinformatics* **23**, 207–214 (2007). URL <http://dx.doi.org/10.1093/bioinformatics/btl1562>.
175. Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
176. Enright, A. J., Dongen, S. V. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575–1584 (2002).
177. Xiao, Y. A tutorial on analysis and simulation of boolean gene regulatory network models. *Curr Genomics* **10**, 511–525 (2009). URL <http://dx.doi.org/10.2174/138920209789208237>.
178. Singhania, R., Sramkoski, R. M., Jacobberger, J. W. & Tyson, J. J. A hybrid model of mammalian cell cycle regulation. *PLoS Comput Biol* **7**, e1001077 (2011). URL <http://dx.doi.org/10.1371/journal.pcbi.1001077>.
179. Zheng, J. *et al.* Simboolnet—a cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics* **26**, 141–142 (2010). URL <http://dx.doi.org/10.1093/bioinformatics/btp617>.
180. Thomas, R. Boolean formalization of genetic control circuits. *J Theor Biol* **42**, 563–585 (1973).
181. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* **22**, 437–467 (1969).
182. Kahlem, P. & Birney, E. Dry work in a wet world: computation in systems biology. *Mol Syst Biol* **2**, 40 (2006). URL <http://dx.doi.org/10.1038/msb4100080>.
183. Assmann, S. M. & Albert, R. Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. *Methods Mol Biol* **553**, 207–225 (2009). URL http://dx.doi.org/10.1007/978-1-60327-563-7_10.
184. Gross, L. When evidence is scant, mathematical modeling offers a roadmap for discovery. *PLoS Biol* **4**, e323 (2006). URL <http://dx.doi.org/10.1371/journal.pbio.0040323>.
185. Li, S., Assmann, S. M. & Albert, R. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol* **4**, e312 (2006). URL <http://dx.doi.org/10.1371/journal.pbio.0040312>.

186. Saadatpour, A. *et al.* Dynamical and structural analysis of a t cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. *PLoS Comput Biol* **7**, e1002267 (2011). URL <http://dx.doi.org/10.1371/journal.pcbi.1002267>.
187. Albert, R. & Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. *J Theor Biol* **223**, 1–18 (2003).
188. BECKER, A. J., McCULLOCH, E. A. & TILL, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).
189. SIMINOVITCH, L., MCCULLOCH, E. A. & TILL, J. E. The distribution of colony-forming cells among spleen colonies. *J Cell Physiol* **62**, 327–336 (1963).
190. Roeder, I. & Glauche, I. Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors gata-1 and pu.1. *J Theor Biol* **241**, 852–865 (2006). URL <http://dx.doi.org/10.1016/j.jtbi.2006.01.021>.
191. Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* **305**, 695–713 (2007). URL <http://dx.doi.org/10.1016/j.ydbio.2007.02.036>.
192. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* **94**, 128701 (2005).
193. Krumsiek, J., Marr, C., Schroeder, T. & Theis, F. J. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS One* **6**, e22649 (2011). URL <http://dx.doi.org/10.1371/journal.pone.0022649>.
194. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008). URL <http://dx.doi.org/10.1038/nature06965>.
195. Roeder, I. & Loeffler, M. A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Exp Hematol* **30**, 853–861 (2002).
196. Chambers, S. M. *et al.* Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS Biol* **5**, e201 (2007). URL <http://dx.doi.org/10.1371/journal.pbio.0050201>.

Bibliography

197. Forsberg, E. C. *et al.* Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. *PLoS One* **5**, e8785 (2010). URL <http://dx.doi.org/10.1371/journal.pone.0008785>.
198. Oostendorp, R. A. J. *et al.* Stromal cell lines from mouse aorta-gonads-mesonephros subregions are potent supporters of hematopoietic stem cell activity. *Blood* **99**, 1183–1189 (2002).
199. Oostendorp, R. A. J. *et al.* Long-term maintenance of hematopoietic stem cells does not require contact with embryo-derived stromal cells in cocultures. *Stem Cells* **23**, 842–851 (2005). URL <http://dx.doi.org/10.1634/stemcells.2004-0120>.
200. Istvanffy, R. *et al.* Stromal pleiotrophin regulates repopulation behavior of hematopoietic stem cells. *Blood* **118**, 2712–2722 (2011). URL <http://dx.doi.org/10.1182/blood-2010-05-287235>.
201. Rozen, S. & Skaletsky, H. Primer3 on the www for general users and for biologist programmers. *Methods Mol Biol* **132**, 365–386 (2000).
202. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990). URL [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
203. Hellemans, J., Mortier, G., Paepe, A. D., Speleman, F. & Vandesompele, J. qbase relative quantification framework and software for management and automated analysis of real-time quantitative pcr data. *Genome Biol* **8**, R19 (2007). URL <http://dx.doi.org/10.1186/gb-2007-8-2-r19>.
204. Zhang, H. L., Singer, R. H. & Bassell, G. J. Neurotrophin regulation of beta-actin mrna and protein localization within growth cones. *J Cell Biol* **147**, 59–70 (1999).
205. Oostendorp, R. A., Audet, J. & Eaves, C. J. High-resolution tracking of cell division suggests similar cell cycle kinetics of hematopoietic stem cells stimulated in vitro and in vivo. *Blood* **95**, 855–862 (2000).
206. Ihaka R, G. R. R: A language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299?314 (1996).
207. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004). URL <http://dx.doi.org/10.1186/gb-2004-5-10-r80>.

208. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of affymetrix genechip data at the probe level. *Bioinformatics* **20**, 307–315 (2004). URL <http://dx.doi.org/10.1093/bioinformatics/btg405>.
209. Wu Z., G. R. M. M. F.-S. F., Irizarry R.A. A model based background adjustment for oligonucleotide expression arrays. *Johns Hopkins University, Dept. of Biostatistics Working Papers* (2004).
210. Irizarry, R. A. *et al.* Summaries of affymetrix genechip probe level data. *Nucleic Acids Res* **31**, e15 (2003).
211. Kauffmann, A., Gentleman, R. & Huber, W. arrayqualitymetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009). URL <http://dx.doi.org/10.1093/bioinformatics/btn647>.
212. Ernst, J. & Bar-Joseph, Z. Stem: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006). URL <http://dx.doi.org/10.1186/1471-2105-7-191>.
213. Smyth, G. K. *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Springer, 2005).
214. Cork, R. C., Vaughan, R. W. & Humphrey, L. S. Precision and accuracy of intraoperative temperature monitoring. *Anesth Analg* **62**, 211–214 (1983).
215. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. Rocr: visualizing classifier performance in r. *Bioinformatics* **21**, 3940–3941 (2005). URL <http://dx.doi.org/10.1093/bioinformatics/bti623>.
216. Zou, K. H., O'Malley, A. J. & Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **115**, 654–657 (2007). URL <http://dx.doi.org/10.1161/CIRCULATIONAHA.105.594929>.
217. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. Toppgene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305–W311 (2009). URL <http://dx.doi.org/10.1093/nar/gkp427>.
218. Barnickel, T., Weston, J., Collobert, R., Mewes, H.-W. & Stmpflen, V. Large scale application of neural network based semantic role labeling for automated relation extraction from biomedical texts. *PLoS One* **4**, e6393 (2009). URL <http://dx.doi.org/10.1371/journal.pone.0006393>.

Bibliography

219. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011). URL <http://dx.doi.org/10.1093/bioinformatics/btq675>.
220. Freeman, T. C., Raza, S., Theocharidis, A. & Ghazal, P. The mepn scheme: an intuitive and flexible graphical system for rendering biological pathways. *BMC Syst Biol* **4**, 65 (2010). URL <http://dx.doi.org/10.1186/1752-0509-4-65>.
221. Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008). URL <http://dx.doi.org/10.1093/bioinformatics/btm554>.
222. Reimand, J., Tooming, L., Peterson, H., Adler, P. & Vilo, J. Graphweb: mining heterogeneous biological networks for gene modules with functional significance. *Nucleic Acids Res* **36**, W452–W459 (2008). URL <http://dx.doi.org/10.1093/nar/gkn230>.
223. Mssell, C., Hopfensitz, M. & Kestler, H. A. Boolnet—an r package for generation, reconstruction and analysis of boolean networks. *Bioinformatics* **26**, 1378–1380 (2010). URL <http://dx.doi.org/10.1093/bioinformatics/btq124>.
224. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. Toppgene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305–W311 (2009). URL <http://dx.doi.org/10.1093/nar/gkp427>.
225. Galluzzi, L., Kepp, O., Trojel-Hansen, C. & Kroemer, G. Mitochondrial control of cellular life, stress, and death. *Circ Res* **111**, 1198–1207 (2012). URL <http://dx.doi.org/10.1161/CIRCRESAHA.112.268946>.
226. Chuaqui, R. F. *et al.* Post-analysis follow-up and validation of microarray experiments. *Nat Genet* **32 Suppl**, 509–514 (2002). URL <http://dx.doi.org/10.1038/ng1034>.
227. Wang, Y. *et al.* Large scale real-time pcr validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics* **7**, 59 (2006). URL <http://dx.doi.org/10.1186/1471-2164-7-59>.
228. Morey, J. S., Ryan, J. C. & Dolah, F. M. V. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time pcr. *Biol Proced Online* **8**, 175–193 (2006). URL <http://dx.doi.org/10.1251/bpo126>.
229. Yuen, T., Wurmbach, E., Pfeffer, R. L., Ebersole, B. J. & Sealfon, S. C. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res* **30**, e48 (2002).

230. Beckman, K. B., Lee, K. Y., Golden, T. & Melov, S. Gene expression profiling in mitochondrial disease: assessment of microarray accuracy by high-throughput q-pcr. *Mitochondrion* **4**, 453–470 (2004). URL <http://dx.doi.org/10.1016/j.mito.2004.07.029>.
231. Benjamini, Y., Yoav; Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57** (1), 289–300 (1995).
232. Shih, C. C., Hu, M. C., Hu, J., Medeiros, J. & Forman, S. J. Long-term ex vivo maintenance and expansion of transplantable human hematopoietic stem cells. *Blood* **94**, 1623–1636 (1999).
233. Zhu, M. & Zhao, S. Candidate gene identification approach: progress and challenges. *Int J Biol Sci* **3**, 420–427 (2007).
234. Aerts, S. *et al.* Gene prioritization through genomic data fusion. *Nat Biotechnol* **24**, 537–544 (2006). URL <http://dx.doi.org/10.1038/nbt1203>.
235. Bie, T. D., Tranchevent, L.-C., van Oeffelen, L. M. M. & Moreau, Y. Kernel-based data fusion for gene prioritization. *Bioinformatics* **23**, i125–i132 (2007). URL <http://dx.doi.org/10.1093/bioinformatics/btm187>.
236. Khler, S., Bauer, S., Horn, D. & Robinson, P. N. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet* **82**, 949–958 (2008). URL <http://dx.doi.org/10.1016/j.ajhg.2008.02.013>.
237. Ozgr, A., Vu, T., Erkan, G. & Radev, D. R. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* **24**, i277–i285 (2008). URL <http://dx.doi.org/10.1093/bioinformatics/btn182>.
238. Ortutay, C. & Vihinen, M. Identification of candidate disease genes by integrating gene ontologies and protein-interaction networks: case study of primary immunodeficiencies. *Nucleic Acids Res* **37**, 622–628 (2009). URL <http://dx.doi.org/10.1093/nar/gkn982>.
239. Chen, J., Aronow, B. J. & Jegga, A. G. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* **10**, 73 (2009). URL <http://dx.doi.org/10.1186/1471-2105-10-73>.

Bibliography

240. Huang, H., Li, J. & Chen, J. Y. Disease gene-fishing in molecular interaction networks: a case study in colorectal cancer. *Conf Proc IEEE Eng Med Biol Soc* **2009**, 6416–6419 (2009). URL <http://dx.doi.org/10.1109/IEMBS.2009.5333750>.
241. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol* **6**, e1000641 (2010). URL <http://dx.doi.org/10.1371/journal.pcbi.1000641>.
242. Erten, S., Bebek, G. & Koyutrk, M. Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. *J Comput Biol* **18**, 1561–1574 (2011). URL <http://dx.doi.org/10.1089/cmb.2011.0154>.
243. Chen, Y. *et al.* In silico gene prioritization by integrating multiple data sources. *PLoS One* **6**, e21137 (2011). URL <http://dx.doi.org/10.1371/journal.pone.0021137>.
244. Dellagi, K., Vainchenker, W., Vinci, G., Paulin, D. & Brouet, J. C. Alteration of vimentin intermediate filament expression during differentiation of human hemopoietic cells. *EMBO J* **2**, 1509–1514 (1983).
245. Sala-Torra, O. *et al.* Connective tissue growth factor (ctgf) expression and outcome in adult patients with acute lymphoblastic leukemia. *Blood* **109**, 3080–3083 (2007). URL <http://dx.doi.org/10.1182/blood-2006-06-031096>.
246. McCallum, L. & Irvine, A. E. Ccn3—a key regulator of the hematopoietic compartment. *Blood Rev* **23**, 79–85 (2009). URL <http://dx.doi.org/10.1016/j.blre.2008.07.002>.
247. Wagner, W. *et al.* Molecular and secretory profiles of human mesenchymal stromal cells and their abilities to maintain primitive hematopoietic progenitors. *Stem Cells* **25**, 2638–2647 (2007). URL <http://dx.doi.org/10.1634/stemcells.2007-0280>.
248. Huang, S. S. *et al.* Cellular growth inhibition by igfbp-3 and tgf-beta1 requires lrp-1. *FASEB J* **17**, 2068–2081 (2003). URL <http://dx.doi.org/10.1096/fj.03-0256com>.
249. Zhu, W. *et al.* Igfbp-4 is an inhibitor of canonical wnt signalling required for cardiogenesis. *Nature* **454**, 345–349 (2008). URL <http://dx.doi.org/10.1038/nature07027>.
250. Williams, K. J. *et al.* Mechanisms by which lipoprotein lipase alters cellular metabolism of lipoprotein(a), low density lipoprotein, and nascent lipoproteins. roles for low density lipoprotein receptors and heparan sulfate proteoglycans. *J Biol Chem* **267**, 13284–13292 (1992).

251. Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J Exp Med* **202**, 1599–1611 (2005). URL <http://dx.doi.org/10.1084/jem.20050967>.
252. Matthews, K. E., Mueller, S. G., Woods, C. & Bell, D. N. Expression of the hemoglobin-haptoglobin receptor cd163 on hematopoietic progenitors. *Stem Cells Dev* **15**, 40–48 (2006). URL <http://dx.doi.org/10.1089/scd.2006.15.40>.
253. Suda, T., Takubo, K. & Semenza, G. L. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell* **9**, 298–310 (2011). URL <http://dx.doi.org/10.1016/j.stem.2011.09.010>.
254. Thomas, D. W. *et al.* Coagulation defects and altered hemodynamic responses in mice lacking receptors for thromboxane a₂. *J Clin Invest* **102**, 1994–2001 (1998). URL <http://dx.doi.org/10.1172/JCI51116>.
255. Takubo, K. & Suda, T. Roles of the hypoxia response system in hematopoietic and leukemic stem cells. *Int J Hematol* **95**, 478–483 (2012). URL <http://dx.doi.org/10.1007/s12185-012-1071-4>.
256. Ledran, M. H. *et al.* Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches. *Cell Stem Cell* **3**, 85–98 (2008). URL <http://dx.doi.org/10.1016/j.stem.2008.06.001>.
257. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by cxcl12-cxcr4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006). URL <http://dx.doi.org/10.1016/j.immuni.2006.10.016>.
258. Broxmeyer, H. E. *et al.* Rapid mobilization of murine and human hematopoietic stem and progenitor cells with amd3100, a cxcr4 antagonist. *J Exp Med* **201**, 1307–1318 (2005). URL <http://dx.doi.org/10.1084/jem.20041385>.
259. Jaubert, J. *et al.* Three new allelic mouse mutations that cause skeletal overgrowth involve the natriuretic peptide receptor c gene (npr3). *Proc Natl Acad Sci U S A* **96**, 10278–10283 (1999).
260. McCubrey, J. A. *et al.* Growth-promoting effects of insulin-like growth factor-1 (igf-1) on hematopoietic cells: overexpression of introduced igf-1 receptor abrogates interleukin-3

Bibliography

- dependency of murine factor-dependent cells by a ligand-dependent mechanism. *Blood* **78**, 921–929 (1991).
261. Nistala, H. *et al.* Fibrillin-1 and -2 differentially modulate endogenous tgf-beta and bmp bioavailability during bone formation. *J Cell Biol* **190**, 1107–1121 (2010). URL <http://dx.doi.org/10.1083/jcb.201003089>.
262. Tordjman, R. *et al.* Neuropilin-1 is expressed on bone marrow stromal cells: a novel interaction with hematopoietic cells? *Blood* **94**, 2301–2309 (1999).
263. Jurk, K. *et al.* Thrombospondin-1 mediates platelet adhesion at high shear via glycoprotein ib (gpib): an alternative/backup mechanism to von willebrand factor. *FASEB J* **17**, 1490–1492 (2003). URL <http://dx.doi.org/10.1096/fj.02-0830fje>.
264. Tjwa, M. *et al.* Membrane-anchored upar regulates the proliferation, marrow pool size, engraftment, and mobilization of mouse hematopoietic stem/progenitor cells. *J Clin Invest* **119**, 1008–1018 (2009). URL <http://dx.doi.org/10.1172/JCI36010>.
265. Wang, H. *et al.* Wnt2 coordinates the commitment of mesoderm to hematopoietic, endothelial, and cardiac lineages in embryoid bodies. *J Biol Chem* **282**, 782–791 (2007). URL <http://dx.doi.org/10.1074/jbc.M606610200>.
266. Lévesque, J.-P., Hendy, J., Winkler, I. G., Takamatsu, Y. & Simmons, P. J. Granulocyte colony-stimulating factor induces the release in the bone marrow of proteases that cleave c-kit receptor (cd117) from the surface of hematopoietic progenitor cells. *Exp Hematol* **31**, 109–117 (2003).
267. Oliveros, J. Venny. an interactive tool for comparing lists with venn diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html> (2007).
268. Buckley, S. M. *IDENTIFICATION OF MICROENVIRONMENT FACTORS THAT REGULATE HEMATOPOIETIC STEM AND PROGENITOR CELLS*. Ph.D. thesis, THE UNIVERSITY OF MINNESOTA (2009).
269. Murphy, M. *et al.* Suppression subtractive hybridization identifies high glucose levels as a stimulus for expression of connective tissue growth factor and other genes in human mesangial cells. *J Biol Chem* **274**, 5830–5834 (1999).
270. Battula, V. L. *et al.* Connective tissue growth factor regulates adipocyte differentiation of mesenchymal stromal cells and facilitates leukemia bone marrow engraftment. *Blood* **122**, 357–366 (2013). URL <http://dx.doi.org/10.1182/blood-2012-06-437988>.

271. Saez-Rodriguez, J., Alexopoulos, L. G. & Stolovitzky, G. Setting the standards for signal transduction research. *Sci Signal* **4**, pe10 (2011). URL <http://dx.doi.org/10.1126/scisignal.2001844>.
272. Holbourn, K. P., Acharya, K. R. & Perbal, B. The ccn family of proteins: structure-function relationships. *Trends Biochem Sci* **33**, 461–473 (2008). URL <http://dx.doi.org/10.1016/j.tibs.2008.07.006>.
273. Deng, Y.-Z. *et al.* Connective tissue growth factor is overexpressed in esophageal squamous cell carcinoma and promotes tumorigenicity through beta-catenin-t-cell factor/lef signaling. *J Biol Chem* **282**, 36571–36581 (2007). URL <http://dx.doi.org/10.1074/jbc.M704141200>.
274. Chen, C. C., Chen, N. & Lau, L. F. The angiogenic factors cyr61 and connective tissue growth factor induce adhesive signaling in primary human skin fibroblasts. *J Biol Chem* **276**, 10443–10452 (2001). URL <http://dx.doi.org/10.1074/jbc.M008087200>.
275. Abraham, D. Connective tissue growth factor: growth factor, matricellular organizer, fibrotic biomarker or molecular target for anti-fibrotic therapy in ssc? *Rheumatology (Oxford)* **47 Suppl 5**, v8–v9 (2008). URL <http://dx.doi.org/10.1093/rheumatology/ken278>.
276. Mason, R. M. Connective tissue growth factor(ccn2), a pathogenic factor in diabetic nephropathy. what does it do? how does it do it? *J Cell Commun Signal* **3**, 95–104 (2009). URL <http://dx.doi.org/10.1007/s12079-009-0038-6>.
277. Bird, T. G., Lorenzini, S. & Forbes, S. J. Activation of stem cells in hepatic diseases. *Cell Tissue Res* **331**, 283–300 (2008). URL <http://dx.doi.org/10.1007/s00441-007-0542-z>.
278. Ha, Y. M., Lee, D. H., Kim, M. & Kang, Y. J. High glucose induces connective tissue growth factor expression and extracellular matrix accumulation in rat aorta vascular smooth muscle cells via extracellular signal-regulated kinase 1/2. *Korean J Physiol Pharmacol* **17**, 307–314 (2013). URL <http://dx.doi.org/10.4196/kjpp.2013.17.4.307>.
279. Kubota, S. & Takigawa, M. Role of ccn2/ctgf/hcs24 in bone growth. *Int Rev Cytol* **257**, 1–41 (2007). URL [http://dx.doi.org/10.1016/S0074-7696\(07\)57001-4](http://dx.doi.org/10.1016/S0074-7696(07)57001-4).

Bibliography

280. Kothapalli, D. & Grotendorst, G. R. Ctgf modulates cell cycle progression in camp-arrested nrk fibroblasts. *J Cell Physiol* **182**, 119–126 (2000). URL <http://dx.doi.org/3.0.CO;2-4>.
281. Zhen, Y. *et al.* Reduced ctgf expression promotes cell growth, migration, and invasion in nasopharyngeal carcinoma. *PLoS One* **8**, e64976 (2013). URL <http://dx.doi.org/10.1371/journal.pone.0064976>.
282. Lu, H. *et al.* Targeting connective tissue growth factor (ctgf) in acute lymphoblastic leukemia preclinical models: anti-ctgf monoclonal antibody attenuates leukemia growth. *Ann Hematol* (2013). URL <http://dx.doi.org/10.1007/s00277-013-1939-2>.
283. Gupta, R., Hong, D., Iborra, F., Sarno, S. & Enver, T. Nov (ccn3) functions as a regulator of human hematopoietic stem or progenitor cells. *Science* **316**, 590–593 (2007). URL <http://dx.doi.org/10.1126/science.1136031>.
284. Cicha, I. & Goppelt-Struebe, M. Connective tissue growth factor: context-dependent functions and mechanisms of regulation. *Biofactors* **35**, 200–208 (2009). URL <http://dx.doi.org/10.1002/biof.30>.
285. O’Seaghdha, C. M. *et al.* Lower urinary connective tissue growth factor levels and incident ckd stage 3 in the general population. *Am J Kidney Dis* **57**, 841–849 (2011). URL <http://dx.doi.org/10.1053/j.ajkd.2010.11.022>.
286. Arnott, J. A. *et al.* The role of connective tissue growth factor (ctgf/ccn2) in skeletogenesis. *Crit Rev Eukaryot Gene Expr* **21**, 43–69 (2011).
287. Riser, B. L. *et al.* Regulation of connective tissue growth factor activity in cultured rat mesangial cells and its expression in experimental diabetic glomerulosclerosis. *J Am Soc Nephrol* **11**, 25–38 (2000).
288. Schauer, K. & Stingl, K. ‘guilty by association’ - protein-protein interactions (ppis) in bacterial pathogens. *Genome Dyn* **6**, 48–61 (2009). URL <http://dx.doi.org/10.1159/000235762>.
289. Yang, L. *et al.* Ifn-gamma negatively modulates self-renewal of repopulating human hemopoietic stem cells. *J Immunol* **174**, 752–757 (2005).
290. Ueda, T. *et al.* Expansion of human nod/scid-repopulating cells by stem cell factor, flk2/flt3 ligand, thrombopoietin, il-6, and soluble il-6 receptor. *J Clin Invest* **105**, 1013–1021 (2000). URL <http://dx.doi.org/10.1172/JCI8583>.

291. Barrientos, S., Stojadinovic, O., Golinko, M. S., Brem, H. & Tomic-Canic, M. Growth factors and cytokines in wound healing. *Wound Repair Regen* **16**, 585–601 (2008). URL <http://dx.doi.org/10.1111/j.1524-475X.2008.00410.x>.
292. Brigstock, D. R. Regulation of angiogenesis and endothelial cell function by connective tissue growth factor (ctgf) and cysteine-rich 61 (cyr61). *Angiogenesis* **5**, 153–165 (2002).
293. Smerdel-Ramoya, A., Zanotti, S., Stadmeyer, L., Durant, D. & Canalis, E. Skeletal overexpression of connective tissue growth factor impairs bone formation and causes osteopenia. *Endocrinology* **149**, 4374–4381 (2008). URL <http://dx.doi.org/10.1210/en.2008-0254>.
294. Wang, R.-S. & Albert, R. Elementary signaling modes predict the essentiality of signal transduction network components. *BMC Syst Biol* **5**, 44 (2011). URL <http://dx.doi.org/10.1186/1752-0509-5-44>.
295. Erdos, P. & Renyi, A. *On the evolution of random graphs*. (MTA Mat. Kut. Int. K ozl., 1960).
296. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).
297. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004). URL <http://dx.doi.org/10.1038/nrg1272>.
298. Wahab, N. A., Weston, B. S. & Mason, R. M. Modulation of the tgfbeta/smad signaling pathway in mesangial cells by ctgf/ccn2. *Exp Cell Res* **307**, 305–314 (2005). URL <http://dx.doi.org/10.1016/j.yexcr.2005.03.022>.
299. Gomes, I. *et al.* Novel transcription factors in human cd34 antigen-positive hematopoietic cells. *Blood* **100**, 107–119 (2002).
300. Beak, J. Y., Kang, H. S., Kim, Y.-S. & Jetten, A. M. Functional analysis of the zinc finger and activation domains of glis3 and mutant glis3(ndh1). *Nucleic Acids Res* **36**, 1690–1702 (2008). URL <http://dx.doi.org/10.1093/nar/gkn009>.
301. Yener, B. *et al.* Multiway modeling and analysis in stem cell systems biology. *BMC Syst Biol* **2**, 63 (2008). URL <http://dx.doi.org/10.1186/1752-0509-2-63>.

Bibliography

302. Hashimoto, R. F. *et al.* Growing genetic regulatory networks from seed genes. *Bioinformatics* **20**, 1241–1247 (2004). URL <http://dx.doi.org/10.1093/bioinformatics/bth074>.
303. Schlitt, T. & Brazma, A. Current approaches to gene regulatory network modelling. *BMC Bioinformatics* **8 Suppl 6**, S9 (2007). URL <http://dx.doi.org/10.1186/1471-2105-8-S6-S9>.
304. Hwang, S., Kim, S., Shin, H. & Lee, D. Context-dependent transcriptional regulations between signal transduction pathways. *BMC Bioinformatics* **12**, 19 (2011). URL <http://dx.doi.org/10.1186/1471-2105-12-19>.
305. Sahin, O. *et al.* Modeling erbb receptor-regulated g1/s transition to find novel targets for de novo trastuzumab resistance. *BMC Syst Biol* **3**, 1 (2009). URL <http://dx.doi.org/10.1186/1752-0509-3-1>.
306. Haberichter, T. *et al.* A systems biology dynamical model of mammalian g1 cell cycle progression. *Mol Syst Biol* **3**, 84 (2007). URL <http://dx.doi.org/10.1038/msb4100126>.
307. Seville, L. L., Shah, N., Westwell, A. D. & Chan, W. C. Modulation of prb/e2f functions in the regulation of cell cycle and in cancer. *Curr Cancer Drug Targets* **5**, 159–170 (2005).
308. Harbour, J. W., Luo, R. X., Santi, A. D., Postigo, A. A. & Dean, D. C. Cdk phosphorylation triggers sequential intramolecular interactions that progressively block rb functions as cells move through g1. *Cell* **98**, 859–869 (1999).
309. Ezhevsky, S. A. *et al.* Hypo-phosphorylation of the retinoblastoma protein (prb) by cyclin d:cdk4/6 complexes results in active prb. *Proc Natl Acad Sci U S A* **94**, 10699–10704 (1997).
310. Gérard, C. & Goldbeter, A. Temporal self-organization of the cyclin/cdk network driving the mammalian cell cycle. *Proc Natl Acad Sci U S A* **106**, 21643–21648 (2009). URL <http://dx.doi.org/10.1073/pnas.0903827106>.
311. Simone, C. Swi/snf: the crossroads where extracellular signaling pathways meet chromatin. *J Cell Physiol* **207**, 309–314 (2006). URL <http://dx.doi.org/10.1002/jcp.20514>.

312. Resnitzky, D., Gossen, M., Bujard, H. & Reed, S. I. Acceleration of the g1/s phase transition by expression of cyclins d1 and e with an inducible system. *Mol Cell Biol* **14**, 1669–1679 (1994).
313. Assoian, R. K. & Klein, E. A. Growth control by intracellular tension and extracellular stiffness. *Trends Cell Biol* **18**, 347–352 (2008). URL <http://dx.doi.org/10.1016/j.tcb.2008.05.002>.
314. Klein, E. A. & Assoian, R. K. Transcriptional regulation of the cyclin d1 gene at a glance. *J Cell Sci* **121**, 3853–3857 (2008). URL <http://dx.doi.org/10.1242/jcs.039131>.
315. Abdel-Wahab, N., Weston, B. S., Roberts, T. & Mason, R. M. Connective tissue growth factor and regulation of the mesangial cell cycle: role in cellular hypertrophy. *J Am Soc Nephrol* **13**, 2437–2445 (2002).
316. Wu, S.-H., Wu, X.-H., Lu, C., Dong, L. & Chen, Z.-Q. Lipoxin a4 inhibits proliferation of human lung fibroblasts induced by connective tissue growth factor. *Am J Respir Cell Mol Biol* **34**, 65–72 (2006). URL <http://dx.doi.org/10.1165/rcmb.2005-01840C>.
317. Wang, R., Xu, Y.-J., Liu, X.-S., Zeng, D.-X. & Xiang, M. Knockdown of connective tissue growth factor by plasmid-based short hairpin rna prevented pulmonary vascular remodeling in cigarette smoke-exposed rats. *Arch Biochem Biophys* **508**, 93–100 (2011). URL <http://dx.doi.org/10.1016/j.abb.2011.01.019>.
318. Tan, T.-W. *et al.* Ctgf enhances migration and mmp-13 up-regulation via alphavbeta3 integrin, fak, erk, and nf-kappab-dependent pathway in human chondrosarcoma cells. *J Cell Biochem* **107**, 345–356 (2009). URL <http://dx.doi.org/10.1002/jcb.22132>.
319. Glauser, D. A. & Schlegel, W. Sequential actions of erk1/2 on the ap-1 transcription factor allow temporal integration of metabolic signals in pancreatic beta cells. *FASEB J* **21**, 3240–3249 (2007). URL <http://dx.doi.org/10.1096/fj.06-7798com>.
320. Nishimoto, S. & Nishida, E. Mapk signalling: Erk5 versus erk1/2. *EMBO Rep* **7**, 782–786 (2006). URL <http://dx.doi.org/10.1038/sj.embor.7400755>.
321. Hennigan, R. F. & Stambrook, P. J. Dominant negative c-jun inhibits activation of the cyclin d1 and cyclin e kinase complexes. *Mol Biol Cell* **12**, 2352–2363 (2001).
322. Mateyak, M. K., Obaya, A. J. & Sedivy, J. M. c-myc regulates cyclin d-cdk4 and -cdk6 activity but affects cell cycle progression at multiple independent points. *Mol Cell Biol* **19**, 4672–4683 (1999).

Bibliography

323. Sherr, C. J. & Roberts, J. M. Cdk inhibitors: positive and negative regulators of g1-phase progression. *Genes Dev* **13**, 1501–1512 (1999).
324. Zhao, Z. *et al.* Connective tissue growth factor (ctgf) expression in the brain is a downstream effector of insulin resistance- associated promotion of alzheimer's disease beta-amyloid neuropathology. *FASEB J* **19**, 2081–2082 (2005). URL <http://dx.doi.org/10.1096/fj.05-4359fje>.
325. Gärtner, A., Huang, X. & Hall, A. Neuronal polarity is regulated by glycogen synthase kinase-3 (gsk-3beta) independently of akt/pkb serine phosphorylation. *J Cell Sci* **119**, 3927–3934 (2006). URL <http://dx.doi.org/10.1242/jcs.03159>.
326. Rooney, B. *et al.* Ctgf/ccn2 activates canonical wnt signalling in mesangial cells through lrp6: implications for the pathogenesis of diabetic nephropathy. *FEBS Lett* **585**, 531–538 (2011). URL <http://dx.doi.org/10.1016/j.febslet.2011.01.004>.
327. Luo, Q. *et al.* Connective tissue growth factor (ctgf) is regulated by wnt and bone morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. *J Biol Chem* **279**, 55958–55968 (2004). URL <http://dx.doi.org/10.1074/jbc.M407810200>.
328. Féliers, D., Frank, M. A. & Riley, D. J. Activation of cyclin d1-cdk4 and cdk4-directed phosphorylation of rb protein in diabetic mesangial hypertrophy. *Diabetes* **51**, 3290–3299 (2002).
329. Trimarchi, J. M. & Lees, J. A. Sibling rivalry in the e2f family. *Nat Rev Mol Cell Biol* **3**, 11–20 (2002). URL <http://dx.doi.org/10.1038/nrm714>.
330. Qu, Z., Weiss, J. N. & MacLellan, W. R. Regulation of the mammalian cell cycle: a model of the g1-to-s transition. *Am J Physiol Cell Physiol* **284**, C349–C364 (2003). URL <http://dx.doi.org/10.1152/ajpcell.00066.2002>.
331. Jungert, K. *et al.* Smad-sp1 complexes mediate tgfbeta-induced early transcription of oncogenic smad7 in pancreatic cancer cells. *Carcinogenesis* **27**, 2392–2401 (2006). URL <http://dx.doi.org/10.1093/carcin/bg1078>.
332. Macleod, K. F. *et al.* p53-dependent and independent expression of p21 during cell growth, differentiation, and dna damage. *Genes Dev* **9**, 935–944 (1995).
333. She, Q. B., Chen, N. & Dong, Z. Erks and p38 kinase phosphorylate p53 protein at serine 15 in response to uv radiation. *J Biol Chem* **275**, 20444–20449 (2000). URL <http://dx.doi.org/10.1074/jbc.M001020200>.

334. Robson, C. N., Gnanapragasam, V., Byrne, R. L., Collins, A. T. & Neal, D. E. Transforming growth factor-beta1 up-regulates p15, p21 and p27 and blocks cell cycling in g1 in human prostate epithelium. *J Endocrinol* **160**, 257–266 (1999).
335. Johnsen, S. A., Subramaniam, M., Monroe, D. G., Janknecht, R. & Spelsberg, T. C. Modulation of transforming growth factor beta (tgfbeta)/smad transcriptional responses through targeted degradation of tgfbeta-inducible early gene-1 by human seven in absentia homologue. *J Biol Chem* **277**, 30754–30759 (2002). URL <http://dx.doi.org/10.1074/jbc.M204812200>.
336. Guo, X. & Wang, X.-F. Signaling cross-talk between tgfbeta/bmp and other pathways. *Cell Res* **19**, 71–88 (2009). URL <http://dx.doi.org/10.1038/cr.2008.302>.
337. Hu, P. P., Datto, M. B. & Wang, X. F. Molecular mechanisms of transforming growth factor-beta signaling. *Endocr Rev* **19**, 349–363 (1998).
338. Xiao, H., Hasegawa, T. & Isobe, K. p300 collaborates with sp1 and sp3 in p21(waf1/cip1) promoter activation induced by histone deacetylase inhibitor. *J Biol Chem* **275**, 1371–1376 (2000).
339. Bhattacharyya, S. *et al.* Fibroblast expression of the coactivator p300 governs the intensity of profibrotic response to transforming growth factor beta. *Arthritis Rheum* **52**, 1248–1258 (2005). URL <http://dx.doi.org/10.1002/art.20996>.
340. Ghosh, A. K. & Varga, J. The transcriptional coactivator and acetyltransferase p300 in fibroblast biology and fibrosis. *J Cell Physiol* **213**, 663–671 (2007). URL <http://dx.doi.org/10.1002/jcp.21162>.
341. Seoane, J., Le, H.-V., Shen, L., Anderson, S. A. & Massagué, J. Integration of smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell* **117**, 211–223 (2004).
342. Moustakas, A. & Heldin, C.-H. Non-smad tgfbeta signals. *J Cell Sci* **118**, 3573–3584 (2005). URL <http://dx.doi.org/10.1242/jcs.02554>.
343. Yagi, K. *et al.* c-myc is a downstream target of the smad pathway. *J Biol Chem* **277**, 854–861 (2002). URL <http://dx.doi.org/10.1074/jbc.M104170200>.
344. Faiola, F. *et al.* Dual regulation of c-myc by p300 via acetylation-dependent control of myc protein turnover and coactivation of myc-induced transcription. *Mol Cell Biol* **25**, 10220–10234 (2005). URL <http://dx.doi.org/10.1128/MCB.25.23.10220-10234.2005>.

Bibliography

345. Janknecht, R., Wells, N. J. & Hunter, T. Tgf-beta-stimulated cooperation of smad proteins with the coactivators cbp/p300. *Genes Dev* **12**, 2114–2119 (1998).
346. Ho, J. S. L., Ma, W., Mao, D. Y. L. & Benchimol, S. p53-dependent transcriptional repression of c-myc is required for g1 cell cycle arrest. *Mol Cell Biol* **25**, 7423–7431 (2005). URL <http://dx.doi.org/10.1128/MCB.25.17.7423-7431.2005>.
347. Delpuech, O. *et al.* Induction of mxil-sr alpha by foxo3a contributes to repression of myc-dependent gene expression. *Mol Cell Biol* **27**, 4917–4930 (2007). URL <http://dx.doi.org/10.1128/MCB.01789-06>.
348. Greer, E. L. & Brunet, A. Foxo transcription factors at the interface between longevity and tumor suppression. *Oncogene* **24**, 7410–7425 (2005). URL <http://dx.doi.org/10.1038/sj.onc.1209086>.
349. Han, C.-T., Schoene, N. W. & Lei, K. Y. Influence of zinc deficiency on akt-mdm2-p53 and akt-p21 signaling axes in normal and malignant human prostate cells. *Am J Physiol Cell Physiol* **297**, C1188–C1199 (2009). URL <http://dx.doi.org/10.1152/ajpcell.00042.2009>.
350. Shankar, S., Chen, Q. & Srivastava, R. K. Inhibition of pi3k/akt and mek/erk pathways act synergistically to enhance antiangiogenic effects of egcg through activation of foxo transcription factor. *J Mol Signal* **3**, 7 (2008). URL <http://dx.doi.org/10.1186/1750-2187-3-7>.
351. Hayata, N. *et al.* Connective tissue growth factor induces cardiac hypertrophy through akt signaling. *Biochem Biophys Res Commun* **370**, 274–278 (2008). URL <http://dx.doi.org/10.1016/j.bbrc.2008.03.100>.
352. Zhang, Y. E. Non-smad pathways in tf-beta signaling. *Cell Res* **19**, 128–139 (2009). URL <http://dx.doi.org/10.1038/cr.2008.328>.
353. Lees, S. J., Childs, T. E. & Booth, F. W. Age-dependent foxo regulation of p27kip1 expression via a conserved binding motif in rat muscle precursor cells. *Am J Physiol Cell Physiol* **295**, C1238–C1246 (2008). URL <http://dx.doi.org/10.1152/ajpcell.00349.2008>.
354. Wolf, G., Reinking, R., Zahner, G., Stahl, R. A. K. & Shankland, S. J. Erk 1,2 phosphorylates p27(kip1): Functional evidence for a role in high glucose-induced hypertrophy of mesangial cells. *Diabetologia* **46**, 1090–1099 (2003). URL <http://dx.doi.org/10.1007/s00125-003-1163-z>.

355. Clarke, R. B. p27kip1 phosphorylation by pkb/akt leads to poor breast cancer prognosis. *Breast Cancer Res* **5**, 162–163 (2003). URL <http://dx.doi.org/10.1186/bcr596>.
356. Sheaff, R. J., Groudine, M., Gordon, M., Roberts, J. M. & Clurman, B. E. Cyclin e-cdk2 is a regulator of p27kip1. *Genes Dev* **11**, 1464–1478 (1997).
357. Kamura, T. *et al.* Cytoplasmic ubiquitin ligase kpc regulates proteolysis of p27(kip1) at g1 phase. *Nat Cell Biol* **6**, 1229–1235 (2004). URL <http://dx.doi.org/10.1038/ncb1194>.
358. Gomis, R. R. *et al.* A foxo-smad synexpression group in human keratinocytes. *Proc Natl Acad Sci U S A* **103**, 12747–12752 (2006). URL <http://dx.doi.org/10.1073/pnas.0605333103>.
359. Buckley, S. M. *et al.* Maintenance of hsc by wnt5a secreting agm-derived stromal cell line. *Exp Hematol* **39**, 114–123.e1–5 (2011). URL <http://dx.doi.org/10.1016/j.exphem.2010.09.010>.
360. Smerdel-Ramoya, A., Zanotti, S. & Canalis, E. Connective tissue growth factor (ctgf) transactivates nuclear factor of activated t-cells (nfat) in cells of the osteoblastic lineage. *J Cell Biochem* **110**, 477–483 (2010). URL <http://dx.doi.org/10.1002/jcb.22561>.
361. Johnson, D. G. Regulation of e2f-1 gene expression by p130 (rb2) and d-type cyclin kinase activity. *Oncogene* **11**, 1685–1692 (1995).
362. Hiyama, H., Iavarone, A., LaBaer, J. & Reeves, S. A. Regulated ectopic expression of cyclin d1 induces transcriptional activation of the cdk inhibitor p21 gene without altering cell cycle progression. *Oncogene* **14**, 2533–2542 (1997). URL <http://dx.doi.org/10.1038/sj.onc.1201080>.
363. Rssig, L., Badorff, C., Holzmann, Y., Zeiher, A. M. & Dimmeler, S. Glycogen synthase kinase-3 couples akt-dependent signaling to the regulation of p21cip1 degradation. *J Biol Chem* **277**, 9684–9689 (2002). URL <http://dx.doi.org/10.1074/jbc.M106157200>.
364. Wang, Z. *et al.* Dependence of egf-induced increases in corneal epithelial proliferation and migration on gsk-3 inactivation. *Invest Ophthalmol Vis Sci* **50**, 4828–4835 (2009). URL <http://dx.doi.org/10.1167/iovs.08-2983>.
365. Tian, Q. *et al.* Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics* **3**, 960–969 (2004). URL <http://dx.doi.org/10.1074/mcp.M400055-MCP200>.

Bibliography

366. MacNamara, A., Terfve, C., Henriques, D., Bernabé, B. P. & Saez-Rodriguez, J. State-time spectrum of signal transduction logic models. *Phys Biol* **9**, 045003 (2012). URL <http://dx.doi.org/10.1088/1478-3975/9/4/045003>.
367. Arden, K. C. Foxo: linking new signaling pathways. *Mol Cell* **14**, 416–418 (2004).
368. Mishra, P., Senthivinayagam, S., Rana, A. & Rana, B. Glycogen synthase kinase-3beta regulates snail and beta-catenin during gastrin-induced migration of gastric cancer cells. *J Mol Signal* **5**, 9 (2010). URL <http://dx.doi.org/10.1186/1750-2187-5-9>.
369. Dar, A. A., Belkhiri, A. & El-Rifai, W. The aurora kinase a regulates gsk-3beta in gastric cancer cells. *Oncogene* **28**, 866–875 (2009). URL <http://dx.doi.org/10.1038/onc.2008.434>.
370. Rubinfeld, H. & Seger, R. The erk cascade: a prototype of mapk signaling. *Mol Biotechnol* **31**, 151–174 (2005). URL <http://dx.doi.org/10.1385/MB:31:2:151>.
371. Kuo, Y.-C. *et al.* Regulation of phosphorylation of thr-308 of akt, cell proliferation, and survival by the b55alpha regulatory subunit targeting of the protein phosphatase 2a holoenzyme to akt. *J Biol Chem* **283**, 1882–1892 (2008). URL <http://dx.doi.org/10.1074/jbc.M709585200>.
372. Persad, S. *et al.* Inhibition of integrin-linked kinase (ilk) suppresses activation of protein kinase b/akt and induces cell cycle arrest and apoptosis of pten-mutant prostate cancer cells. *Proc Natl Acad Sci U S A* **97**, 3207–3212 (2000). URL <http://dx.doi.org/10.1073/pnas.060579697>.
373. Eng, C. Role of pten, a lipid phosphatase upstream effector of protein kinase b, in epithelial thyroid carcinogenesis. *Ann N Y Acad Sci* **968**, 213–221 (2002).
374. Kitagawa, M. *et al.* The consensus motif for phosphorylation by cyclin d1-cdk4 is different from that for phosphorylation by cyclin a/e-cdk2. *EMBO J* **15**, 7060–7069 (1996).
375. Geng, Y. *et al.* Deletion of the p27kip1 gene restores normal development in cyclin d1-deficient mice. *Proc Natl Acad Sci U S A* **98**, 194–199 (2001). URL <http://dx.doi.org/10.1073/pnas.011522998>.
376. Ezhevsky, S. A., Ho, A., Becker-Hapak, M., Davis, P. K. & Dowdy, S. F. Differential regulation of retinoblastoma tumor suppressor protein by g(1) cyclin-dependent kinase

- complexes in vivo. *Mol Cell Biol* **21**, 4773–4784 (2001). URL <http://dx.doi.org/10.1128/MCB.21.14.4773-4784.2001>.
377. Lundberg, A. S. & Weinberg, R. A. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol Cell Biol* **18**, 753–761 (1998).
378. Lugus, J. J. *et al.* Gata2 functions at multiple steps in hemangioblast development and differentiation. *Development* **134**, 393–405 (2007). URL <http://dx.doi.org/10.1242/dev.02731>.
379. Terskikh, A. V., Miyamoto, T., Chang, C., Diatchenko, L. & Weissman, I. L. Gene expression analysis of purified hematopoietic stem cells and committed progenitors. *Blood* **102**, 94–101 (2003). URL <http://dx.doi.org/10.1182/blood-2002-08-2509>.
380. Kluger, Y. *et al.* Lineage specificity of gene expression patterns. *Proc Natl Acad Sci U S A* **101**, 6508–6513 (2004). URL <http://dx.doi.org/10.1073/pnas.0401136101>.
381. Venezia, T. A. *et al.* Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol* **2**, e301 (2004). URL <http://dx.doi.org/10.1371/journal.pbio.0020301>.
382. Rochon, C. *et al.* Comparison of gene expression pattern in sp cell populations from four tissues to define common "stemness functions". *Exp Cell Res* **312**, 2074–2082 (2006). URL <http://dx.doi.org/10.1016/j.yexcr.2006.03.010>.
383. Månsson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407–419 (2007). URL <http://dx.doi.org/10.1016/j.immuni.2007.02.013>.
384. Jankovic, V. *et al.* Id1 restrains myeloid commitment, maintaining the self-renewal capacity of hematopoietic stem cells. *Proc Natl Acad Sci U S A* **104**, 1260–1265 (2007). URL <http://dx.doi.org/10.1073/pnas.0607894104>.
385. Ficara, F., Murphy, M. J., Lin, M. & Cleary, M. L. Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell* **2**, 484–496 (2008). URL <http://dx.doi.org/10.1016/j.stem.2008.03.004>.
386. Hoffmann, R., Seidl, T., Neeb, M., Rolink, A. & Melchers, F. Changes in gene expression profiles in developing b cells of murine bone marrow. *Genome Res* **12**, 98–111 (2002). URL <http://dx.doi.org/10.1101/gr.201501>.

Bibliography

387. Hoffmann, R., Bruno, L., Seidl, T., Rolink, A. & Melchers, F. Rules for gene usage inferred from a comparison of large-scale gene expression profiles of t and b lymphocyte development. *J Immunol* **170**, 1339–1353 (2003).
388. Terszowski, G. *et al.* Prospective isolation and global gene expression analysis of the erythrocyte colony-forming unit (cfu-e). *Blood* **105**, 1937–1945 (2005). URL <http://dx.doi.org/10.1182/blood-2004-09-3459>.
389. Chen, Z., Hu, M. & Shivdasani, R. A. Expression analysis of primary mouse megakaryocyte differentiation and its application in identifying stage-specific molecular markers and a novel transcriptional target of nf-e2. *Blood* **109**, 1451–1459 (2007). URL <http://dx.doi.org/10.1182/blood-2006-08-038901>.
390. Dudziak, D. *et al.* Differential antigen processing by dendritic cell subsets in vivo. *Science* **315**, 107–111 (2007). URL <http://dx.doi.org/10.1126/science.1136080>.
391. Nykter, M. *et al.* Gene expression dynamics in the macrophage exhibit criticality. *Proc Natl Acad Sci U S A* **105**, 1897–1900 (2008). URL <http://dx.doi.org/10.1073/pnas.0711525105>.
392. Painter, M. W. *et al.* Transcriptomes of the b and t lineages compared by multiplatform microarray profiling. *J Immunol* **186**, 3047–3057 (2011).
393. Ivanova, N. B. *et al.* A stem cell molecular signature. *Science* **298**, 601–604 (2002). URL <http://dx.doi.org/10.1126/science.1073823>.
394. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. "stemness": transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597–600 (2002). URL <http://dx.doi.org/10.1126/science.1072530>.
395. Fortunel, N. O. *et al.* Comment on " 'stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* **302**, 393; author reply 393 (2003). URL <http://dx.doi.org/10.1126/science.1086384>.
396. Kodama, H. A., Amagai, Y., Koyama, H. & Kasai, S. A new preadipose cell line derived from newborn mouse calvaria can promote the proliferation of pluripotent hemopoietic stem cells in vitro. *J Cell Physiol* **112**, 89–95 (1982). URL <http://dx.doi.org/10.1002/jcp.1041120114>.
397. Collins, L. S. & Dorshkind, K. A stromal cell line from myeloid long-term bone marrow cultures can support myelopoiesis and b lymphopoiesis. *J Immunol* **138**, 1082–1087 (1987).

398. Sutherland, H. J., Eaves, C. J., Lansdorp, P. M., Thacker, J. D. & Hogge, D. E. Differential regulation of primitive human hematopoietic cells in long-term cultures maintained on genetically engineered murine stromal cells. *Blood* **78**, 666–672 (1991).
399. Issaad, C., Croisille, L., Katz, A., Vainchenker, W. & Coulombel, L. A murine stromal cell line allows the proliferation of very primitive human cd34⁺⁺/cd38⁻ progenitor cells in long-term cultures and semisolid assays. *Blood* **81**, 2916–2924 (1993).
400. Deryugina, E. I., Ratnikov, B. I., Bourdon, M. A. & Mller-Sieburg, C. E. Clonal analysis of primary marrow stroma: functional homogeneity in support of lymphoid and myeloid cell lines and identification of positive and negative regulators. *Exp Hematol* **22**, 910–918 (1994).
401. Wineman, J., Moore, K., Lemischka, I. & Mller-Sieburg, C. Functional heterogeneity of the hematopoietic microenvironment: rare stromal elements maintain long-term repopulating stem cells. *Blood* **87**, 4082–4090 (1996).
402. Moore, K. A., Ema, H. & Lemischka, I. R. In vitro maintenance of highly purified, transplantable hematopoietic stem cells. *Blood* **89**, 4337–4347 (1997).
403. Ohneda, O. *et al.* Hematopoietic stem cell maintenance and differentiation are supported by embryonic aorta-gonad-mesonephros region-derived endothelium. *Blood* **92**, 908–919 (1998).
404. Xu, M. J. *et al.* Stimulation of mouse and human primitive hematopoiesis by murine embryonic aorta-gonad-mesonephros-derived stromal cell lines. *Blood* **92**, 2032–2040 (1998).
405. Hackney, J. A. *et al.* A molecular profile of a hematopoietic stem cell niche. *Proc Natl Acad Sci U S A* **99**, 13061–13066 (2002). URL <http://dx.doi.org/10.1073/pnas.192124499>.
406. Shimizu, N. *et al.* Identification of genes potentially involved in supporting hematopoietic stem cell activity of stromal cell line mc3t3-g2/pa6. *Int J Hematol* **87**, 239–245 (2008). URL <http://dx.doi.org/10.1007/s12185-008-0048-9>.
407. Ruscetti, F. W., Akel, S. & Bartelmez, S. H. Autocrine transforming growth factor-beta regulation of hematopoiesis: many outcomes that depend on the context. *Oncogene* **24**, 5751–5763 (2005). URL <http://dx.doi.org/10.1038/sj.onc.1208921>.

Bibliography

408. Deneault, E. *et al.* A functional screen to identify novel effectors of hematopoietic stem cell activity. *Cell* **137**, 369–379 (2009). URL <http://dx.doi.org/10.1016/j.cell.2009.03.026>.
409. Uchida, N. *et al.* Abc transporter activities of murine hematopoietic stem cells vary according to their developmental and activation status. *Blood* **103**, 4487–4495 (2004). URL <http://dx.doi.org/10.1182/blood-2003-11-3989>.
410. Curtis, R. K., Oresic, M. & Vidal-Puig, A. Pathways to the analysis of microarray data. *Trends Biotechnol* **23**, 429–435 (2005). URL <http://dx.doi.org/10.1016/j.tibtech.2005.05.011>.
411. Kuenzel, L. Gene clustering methods for time series microarray data. *Biochemistry* **218** (2010).
412. Farfán F., S. M. M. G. J. H., Ma J. Think back: Knowledge-based interpretation of high throughput data. *BMC Bioinformatics* **13** (2012).
413. Neumann, E. *et al.* Cell culture and passaging alters gene expression pattern and proliferation rate in rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther* **12**, R83 (2010). URL <http://dx.doi.org/10.1186/ar3010>.
414. Asyali, M. H. & Alci, M. Reliability analysis of microarray data using fuzzy c-means and normal mixture modeling based classification methods. *Bioinformatics* **21**, 644–649 (2005). URL <http://dx.doi.org/10.1093/bioinformatics/bti036>.
415. Nguyen, D. V., Arpat, A. B., Wang, N. & Carroll, R. J. Dna microarray experiments: biological and technological aspects. *Biometrics* **58**, 701–717 (2002).
416. Mitsuteru Nakao, Y. K. O. & Kanehisa, M. Quantitative estimation of cross-hybridization in dna microarrays based on a linear model. *Genome Informatics* **11**, 231–232 (2000).
417. Taniguchi, M., Miura, K., Iwao, H. & Yamanaka, S. Quantitative assessment of dna microarrays—comparison with northern blot analyses. *Genomics* **71**, 34–39 (2001). URL <http://dx.doi.org/10.1006/geno.2000.6427>.
418. Wurmbach, E., Yuen, T. & Sealfon, S. C. Focused microarray analysis. *Methods* **31**, 306–316 (2003).
419. Williams, T. D., Gensberg, K., Minchin, S. D. & Chipman, J. K. A dna expression array to detect toxic stress response in european flounder (*platichthys flesus*). *Aquat Toxicol* **65**, 141–157 (2003).

420. Rajeevan, M. S., Vernon, S. D., Taysavang, N. & Unger, E. R. Validation of array-based gene expression profiles by real-time (kinetic) rt-pcr. *J Mol Diagn* **3**, 26–31 (2001). URL [http://dx.doi.org/10.1016/S1525-1578\(10\)60646-0](http://dx.doi.org/10.1016/S1525-1578(10)60646-0).
421. Yang, Z.-M., Chen, W.-W. & Wang, Y.-F. Gene expression profiling in gastric mucosa from helicobacter pylori-infected and uninfected patients undergoing chronic superficial gastritis. *PLoS One* **7**, e33030 (2012). URL <http://dx.doi.org/10.1371/journal.pone.0033030>.
422. Resnick, M. B. *et al.* Global analysis of the human gastric epithelial transcriptome altered by helicobacter pylori eradication in vivo. *Gut* **55**, 1717–1724 (2006). URL <http://dx.doi.org/10.1136/gut.2006.095646>.
423. Ramon Goni, P. G. & Foissac, S. The qpcr data statistical analysis. *Integromics White Paper* (2009).
424. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004). URL <http://dx.doi.org/10.2202/1544-6115.1027>.
425. Bourgon, R., Gentleman, R. & Huber, W. Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci U S A* **107**, 9546–9551 (2010). URL <http://dx.doi.org/10.1073/pnas.0914005107>.
426. Canales, R. D. *et al.* Evaluation of dna microarray results with quantitative gene expression platforms. *Nat Biotechnol* **24**, 1115–1122 (2006). URL <http://dx.doi.org/10.1038/nbt1236>.
427. Etienne, W., Meyer, M. H., Peppers, J. & Meyer, R. A. Comparison of mrna gene expression by rt-pcr and dna microarray. *Biotechniques* **36**, 618–20, 622, 624–6 (2004).
428. Wang, Z., Gerstein, M. & Snyder, M. Rna-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009). URL <http://dx.doi.org/10.1038/nrg2484>.
429. Jing, D. *et al.* Hematopoietic stem cells in co-culture with mesenchymal stromal cells—modeling the niche compartments in vitro. *Haematologica* **95**, 542–550 (2010). URL <http://dx.doi.org/10.3324/haematol.2009.010736>.
430. Alt, J. R., Gladden, A. B. & Diehl, J. A. p21(cip1) promotes cyclin d1 nuclear accumulation via direct inhibition of nuclear export. *J Biol Chem* **277**, 8517–8523 (2002). URL <http://dx.doi.org/10.1074/jbc.M108867200>.

Bibliography

431. O'Donovan, H. C. *et al.* Connective tissue growth factor antagonizes transforming growth factor- β 1/smad signalling in renal mesangial cells. *Biochem J* **441**, 499–510 (2012). URL <http://dx.doi.org/10.1042/BJ20110910>.
432. Lee, J.-W., Chen, H., Pullikotil, P. & Quon, M. J. Protein kinase a-alpha directly phosphorylates foxo1 in vascular endothelial cells to regulate expression of vascular cellular adhesion molecule-1 mrna. *J Biol Chem* **286**, 6423–6432 (2011). URL <http://dx.doi.org/10.1074/jbc.M110.180661>.
433. Machida, S., Spangenburg, E. E. & Booth, F. W. Forkhead transcription factor foxo1 transduces insulin-like growth factor's signal to p27kip1 in primary skeletal muscle satellite cells. *J Cell Physiol* **196**, 523–531 (2003). URL <http://dx.doi.org/10.1002/jcp.10339>.
434. Zeng, X. *et al.* A dual-kinase mechanism for wnt co-receptor phosphorylation and activation. *Nature* **438**, 873–877 (2005). URL <http://dx.doi.org/10.1038/nature04185>.
435. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008). URL <http://dx.doi.org/10.1016/j.cell.2008.10.048>.
436. Scudellari, M. A closer look at the single cell. *Nature Reports Stem Cells* (2009).
437. Roccio, M. *et al.* Predicting stem cell fate changes by differential cell cycle progression patterns. *Development* **140**, 459–470 (2013). URL <http://dx.doi.org/10.1242/dev.086215>.
438. Kaltenbach, H.-M., Dimopoulos, S. & Stelling, J. Systems analysis of cellular networks under uncertainty. *FEBS Lett* **583**, 3923–3930 (2009). URL <http://dx.doi.org/10.1016/j.febslet.2009.10.074>.
439. Mendoza, L., Thieffry, D. & Alvarez-Buylla, E. R. Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analysis. *Bioinformatics* **15**, 593–606 (1999).
440. Espinosa-Soto, C., Padilla-Longoria, P. & Alvarez-Buylla, E. R. A gene regulatory network model for cell-fate determination during arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* **16**, 2923–2939 (2004). URL <http://dx.doi.org/10.1105/tpc.104.021725>.

441. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* **101**, 4781–4786 (2004). URL <http://dx.doi.org/10.1073/pnas.0305937101>.
442. Davidich, M. I. & Bornholdt, S. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One* **3**, e1672 (2008). URL <http://dx.doi.org/10.1371/journal.pone.0001672>.
443. Fauré, A., Naldi, A., Chaouiya, C. & Thieffry, D. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22**, e124–e131 (2006). URL <http://dx.doi.org/10.1093/bioinformatics/btl1210>.
444. Guo, J., Sheng, G. & Warner, B. W. Epidermal growth factor-induced rapid retinoblastoma phosphorylation at ser780 and ser795 is mediated by erk1/2 in small intestine epithelial cells. *J Biol Chem* **280**, 35992–35998 (2005). URL <http://dx.doi.org/10.1074/jbc.M504583200>.
445. Safadi, F. F. *et al.* Expression of connective tissue growth factor in bone: its role in osteoblast proliferation and differentiation in vitro and bone formation in vivo. *J Cell Physiol* **196**, 51–62 (2003). URL <http://dx.doi.org/10.1002/jcp.10319>.
446. Arai, F. & Suda, T. Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche. *Ann N Y Acad Sci* **1106**, 41–53 (2007). URL <http://dx.doi.org/10.1196/annals.1392.005>.
447. Hirao, A., Arai, F. & Suda, T. Regulation of cell cycle in hematopoietic stem cells by the niche. *Cell Cycle* **3**, 1481–1483 (2004).
448. Arai, F., Hirao, A. & Suda, T. Regulation of hematopoietic stem cells by the niche. *Trends Cardiovasc Med* **15**, 75–79 (2005). URL <http://dx.doi.org/10.1016/j.tcm.2005.03.002>.
449. Fan, W. H., Pech, M. & Karnovsky, M. J. Connective tissue growth factor (ctgf) stimulates vascular smooth muscle cell growth and migration in vitro. *Eur J Cell Biol* **79**, 915–923 (2000). URL <http://dx.doi.org/10.1078/0171-9335-00122>.
450. Bradham, D. M., Igarashi, A., Potter, R. L. & Grotendorst, G. R. Connective tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the src-induced immediate early gene product cef-10. *J Cell Biol* **114**, 1285–1294 (1991).

Bibliography

451. Kothapalli, D., Frazier, K. S., Welply, A., Segarini, P. R. & Grotendorst, G. R. Transforming growth factor beta induces anchorage-independent growth of nrk fibroblasts via a connective tissue growth factor-dependent signaling pathway. *Cell Growth Differ* **8**, 61–68 (1997).
452. Kothapalli, D., Hayashi, N. & Grotendorst, G. R. Inhibition of tgf-beta-stimulated ctgf gene expression and anchorage-independent growth by camp identifies a ctgf-dependent restriction point in the cell cycle. *FASEB J* **12**, 1151–1161 (1998).
453. Hishikawa, K. *et al.* Connective tissue growth factor induces apoptosis in human breast cancer cell line mcf-7. *J Biol Chem* **274**, 37461–37466 (1999).
454. Hishikawa, K. *et al.* Overexpression of connective tissue growth factor gene induces apoptosis in human aortic smooth muscle cells. *Circulation* **100**, 2108–2112 (1999).
455. Kodama, T. *et al.* Increases in p53 expression induce ctgf synthesis by mouse and human hepatocytes and result in liver fibrosis in mice. *J Clin Invest* **121**, 3343–3356 (2011). URL <http://dx.doi.org/10.1172/JCI44957>.
456. Yang, H.-W., Chen, X.-L., Liu, Z.-L., Liu, J. & Bu, L.-M. Ctgsirna ameliorates retinal cells apoptosis in streptozotocin-induced diabetic rats. *Int J Ophthalmol* **3**, 120–124 (2010). URL <http://dx.doi.org/10.3980/j.issn.2222-3959.2010.02.06>.
457. Jang, T. J. *et al.* Increased expression of cyclin d1, cyclin e and p21(cip1) associated with decreased expression of p27(kip1) in chemically induced rat mammary carcinogenesis. *Jpn J Cancer Res* **91**, 1222–1232 (2000).
458. LaBaer, J. *et al.* New functional activities for the p21 family of cdk inhibitors. *Genes Dev* **11**, 847–862 (1997).
459. Wander, S. A., Zhao, D. & Slingerland, J. M. p27: a barometer of signaling deregulation and potential predictor of response to targeted therapies. *Clin Cancer Res* **17**, 12–18 (2011). URL <http://dx.doi.org/10.1158/1078-0432.CCR-10-0752>.
460. Besson, A. *et al.* A pathway in quiescent cells that controls p27kip1 stability, subcellular localization, and tumor suppression. *Genes Dev* **20**, 47–64 (2006). URL <http://dx.doi.org/10.1101/gad.1384406>.
461. Sheaff, R. J. *et al.* Proteasomal turnover of p21cip1 does not require p21cip1 ubiquitination. *Mol Cell* **5**, 403–410 (2000).

462. Bryant, P., Zheng, Q. & Pumiglia, K. Focal adhesion kinase controls cellular levels of p27/kip1 and p21/cip1 through skp2-dependent and -independent mechanisms. *Mol Cell Biol* **26**, 4201–4213 (2006). URL <http://dx.doi.org/10.1128/MCB.01612-05>.
463. Hiromura, K., Pippin, J. W., Fero, M. L., Roberts, J. M. & Shankland, S. J. Modulation of apoptosis by the cyclin-dependent kinase inhibitor p27(kip1). *J Clin Invest* **103**, 597–604 (1999). URL <http://dx.doi.org/10.1172/JCI5461>.
464. Zou, P. *et al.* p57(kip2) and p27(kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with hsc70. *Cell Stem Cell* **9**, 247–261 (2011). URL <http://dx.doi.org/10.1016/j.stem.2011.07.003>.
465. Huang, J. *et al.* Pivotal role for glycogen synthase kinase-3 in hematopoietic stem cell homeostasis in mice. *J Clin Invest* **119**, 3519–3529 (2009). URL <http://dx.doi.org/10.1172/JCI40572>.
466. Trowbridge, J. J., Xenocostas, A., Moon, R. T. & Bhatia, M. Glycogen synthase kinase-3 is an in vivo regulator of hematopoietic stem cell repopulation. *Nat Med* **12**, 89–98 (2006). URL <http://dx.doi.org/10.1038/nm1339>.
467. Weng, L., Brown, J. & Eng, C. Pten induces apoptosis and cell cycle arrest through phosphoinositol-3-kinase/akt-dependent and -independent pathways. *Hum Mol Genet* **10**, 237–242 (2001).
468. Radu, A., Neubauer, V., Akagi, T., Hanafusa, H. & Georgescu, M.-M. Pten induces cell cycle arrest by decreasing the level and nuclear localization of cyclin d1. *Mol Cell Biol* **23**, 6139–6149 (2003).
469. Sun, H. *et al.* Pten modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and akt/protein kinase b signaling pathway. *Proc Natl Acad Sci U S A* **96**, 6199–6204 (1999).
470. Gottschalk, A. R. *et al.* p27kip1 is required for pten-induced g1 growth arrest. *Cancer Res* **61**, 2105–2111 (2001).
471. Zhang, J. *et al.* Pten maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441**, 518–522 (2006). URL <http://dx.doi.org/10.1038/nature04747>.
472. Miyamoto, K. *et al.* Foxo3a is essential for maintenance of the hematopoietic stem cell pool. *Cell Stem Cell* **1**, 101–112 (2007). URL <http://dx.doi.org/10.1016/j.stem.2007.02.001>.

Bibliography

473. Blom, I. E., Goldschmeding, R. & Leask, A. Gene regulation of connective tissue growth factor: new targets for antifibrotic therapy? *Matrix Biol* **21**, 473–482 (2002).
474. Shi-Wen, X., Leask, A. & Abraham, D. Regulation and function of connective tissue growth factor/*ccn2* in tissue repair, scarring and fibrosis. *Cytokine Growth Factor Rev* **19**, 133–144 (2008). URL <http://dx.doi.org/10.1016/j.cytogfr.2008.01.002>.
475. Shi, L. *et al.* Activation of jnk signaling mediates connective tissue growth factor expression and scar formation in corneal wound healing. *PLoS One* **7**, e32128 (2012). URL <http://dx.doi.org/10.1371/journal.pone.0032128>.
476. Gabbiani, G., Ryan, G. B. & Majne, G. Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. *Experientia* **27**, 549–550 (1971).
477. Garrett, Q. *et al.* Involvement of *ctgf* in *tgf-beta1*-stimulation of myofibroblast differentiation and collagen matrix contraction in the presence of mechanical stress. *Invest Ophthalmol Vis Sci* **45**, 1109–1116 (2004).
478. Cirri, P. & Chiarugi, P. Cancer associated fibroblasts: the dark side of the coin. *Am J Cancer Res* **1**, 482–497 (2011).
479. Sonnylal, S. *et al.* Selective expression of connective tissue growth factor in fibroblasts in vivo promotes systemic tissue fibrosis. *Arthritis Rheum* **62**, 1523–1532 (2010). URL <http://dx.doi.org/10.1002/art.27382>.
480. Kapoor, M. *et al.* Gsk-3beta in mouse fibroblasts controls wound healing and fibrosis through an endothelin-1-dependent mechanism. *J Clin Invest* **118**, 3279–3290 (2008). URL <http://dx.doi.org/10.1172/JCI35381>.
481. Parapuram, S. K. *et al.* Loss of *pten* expression by dermal fibroblasts causes skin fibrosis. *J Invest Dermatol* **131**, 1996–2003 (2011). URL <http://dx.doi.org/10.1038/jid.2011.156>.
482. Dvorak, H. F. Tumors: wounds that do not heal. similarities between tumor stroma generation and wound healing. *N Engl J Med* **315**, 1650–1659 (1986). URL <http://dx.doi.org/10.1056/NEJM198612253152606>.

Acknowledgments

I would like to express my sincere gratitude and deep appreciation to my Ph.D. thesis supervisors Prof. Robert A.J. Oostendorp, Prof. Hans-Werner Mewes, Prof. Christian Peschel and Dr. Volker Stümpflen for the very challenging and interesting project, as well as for the patient guidance, encouragement, advice, support and enthusiasm they have provided throughout this thesis. I thank Prof. Hans-Werner Mewes and Prof. Christian Peschel for providing the necessary infrastructure and contributing the tools and materials for this project.

My thanks are also due to my research group members - Dr. Rouzanna Istvanffy, Christina Eckl, Franziska Bock, Sandra Grziwok and Charlotta Pagel, as well as the members of other resaeach groups in Trogerstrasse 32 for their very kind help with the experiments and for creating a great work atmosphere in the lab.

I also wish to thank the Department of Microbiology and Immunology, TU München,- this collaborration was essential for succeeding with the project. Especially I would like to thank Lynette Henkel who was always willing to help me with the cell sorting experiments throughout the project. My acknowledgment goes also to the Strahlentherapie des Klinikums rechts der Isar, TU München, for kindly helping me with the irradiations.

Finally, I am most grateful to the Graduate School of Information Science in Health (GSISH) for providing me with the financial resources necessary.