INTERCELLULAR FEEDBACK IN HEMATOPOIESIS

By

Daniel Christopher Kirouac

A thesis submitted in conformity with the requirements for the degree of Doctorate of Philosophy
Institute for Biomaterials and Biomedical Engineering
University of Toronto

© Copyright by Daniel Christopher Kirouac 2010
INTERCELLULAR FEEDBACK IN HEMATOPOIESIS

Daniel Christopher Kirouac

Doctorate of Philosophy

Institute for Biomaterials and Biomedical Engineering
University of Toronto

2010

Abstract

Despite the importance of inter-cellular (between cell) communication networks in regulating homeostasis in multicellular organisms, very little is known about their topology, dynamics, or functional significance. Inter-cellular communication networks are particularly relevant in stem cell biology, as stem cell fate decisions (self-renewal, proliferation, lineage specification) are tightly regulated based on physiological demand. Using human blood stem cell cultures as an experimental paradigm, we present an integrated experimental and computational approach to interrogate a hierarchically organized tissue network. We have developed a novel mathematical model of blood stem cell development incorporating cell-level kinetic parameters as functions of secreted molecule-mediated inter-cellular networks. By relation to quantitative cellular assays, our model is capable of predictively simulating many disparate features of both normal and malignant hematopoiesis, relating internal parameters and microenvironmental variables to measurable cell fate outcomes. Through integrated in silico and experimental analyses we show blood stem and progenitor cell fate is regulated by cell-cell feedback, and can be controlled non-cell autonomously by dynamically perturbing inter-cellular signalling.

Furthermore, we have compiled genome-scale molecular profiles (transcriptome and secretome), publicly available databases, and literature mining to reconstruct soluble factor-mediated inter-
cellular signalling networks regulating cell fate decisions. We find that dynamic interactions between positive and negative regulators, in the context of tuneable cell culture parameters, tip the balance between stem cell supportive vs. non-supportive conditions. The cell-cytokine interactions can be summarized as an antagonistic positive-negative feedback circuit wherein stem cell self-renewal is regulated by a balance of megakaryocyte-derived stimulatory factors vs. monocyte-derived inhibitory factors. To understand how the experimentally identified positive and negative regulatory signals are integrated at the intra-cellular level, we define a literature-derived blood stem cell self-renewal network wherein these extracellular signals converge for coherent processing into cell fate decisions. In summary, this work demonstrates the utility of integrating experimental and computational methods to explore complex cellular systems, and represents the first attempt to comprehensively elucidate non-autonomous signals balancing stem cell homeostasis and regeneration.
Acknowledgments

I am grateful for the help and support of many individuals, without whom I could not have completed this work. Much of the experimental work was planned and conducted with the assistance of numerous skilled graduate students, research associates, and technicians, most notably Dr. Gerard Madlambayan, Dr. Caryn Ito, Mei Yu, Ting Yin, Edward Sykes, and Aline Roch, as well as other Stem Cell Bioengineering Laboratory members. I thank my committee members Dr. Julie Audet, Dr. Lilliana Attisano, and Dr. John Dick, as well as Dr. Gary Bader for their counsel and help. The foundation for this work and ideas presented have come from fruitful discussions with my supervisor and mentor over the last six years, Dr. Peter Zandstra, with whom I am grateful for his guidance and indispensable advise. I could not have survived graduate school without the support of my loving wife Santa Luangkham-Kirouac, whom I can be certain did not marry me for money. Finally, I would not have made it to where I am today without the enduring love and support of my parents Roland and Sandra Kirouac, who provided me with the opportunities they never had.

Toronto, July 2009
Table of Contents

Acknowledgments ........................................................................................................................................ iv

Table of Contents ................................................................................................................................. v

List of Tables .......................................................................................................................................... ix

List of Figures .......................................................................................................................................... x

1 INTRODUCTION ..................................................................................................................................... 1
  1.1 Cell therapies as an emerging therapeutic modality ................................................................. 2
  1.2 Safety, regulatory and manufacturing issues specific to cell therapies ................................. 6
  1.3 Cell therapy bioprocess optimization and development ....................................................... 7
    1.3.1 Measurement of cellular properties ........................................................................ 7
    1.3.2 Experimental design and process modeling .......................................................... 10
    1.3.3 Multi-dimensional characterization of cell state ................................................. 12
    1.3.4 Process design and control ................................................................................ 15
    1.3.5 Product quality control .................................................................................... 16
  1.4 Future prospects for cell therapies .......................................................................................... 16
  1.5 Biology of hematopoiesis ........................................................................................................ 17
  1.6 Clinical motivation for HSC expansion bioprocesses ............................................................ 19
  1.7 Challenges to growing human HSCs in culture ..................................................................... 22
    1.7.1 Markers ................................................................................................................. 22
    1.7.2 Assays .................................................................................................................. 23
    1.7.3 Heterogeneity ........................................................................................................ 23
  1.8 Regulation of HSC fate .............................................................................................................. 24
    1.8.1 Exogenous manipulations ..................................................................................... 24
    1.8.2 Endogenous inter-cellular signalling networks ..................................................... 27
    1.8.3 Cell population and microenvironmental dynamics .......................................... 28
    1.8.4 Inter-cellular feedback .......................................................................................... 28
  1.9 HSC cultures as dynamic cellular networks .......................................................................... 33
    1.9.1 Mathematical models of hematopoiesis ................................................................ 33
    1.10 Tools necessary for the development of predictive models of cellular network output ... 36
      1.10.1 Stem cell transcriptomics to identify subpopulation dynamics ........................ 36
      1.10.2 Stem cell proteomics to measure extra- and intra-cellular cytokine signalling .... 37
      1.10.3 Systems-level dataset integration ...................................................................... 38
  1.11 Hematopoietic stem cell bioprocess design ......................................................................... 39
  1.12 Hypothesis and approach ...................................................................................................... 39

2 CELL-CELL INTERACTION NETWORKS REGULATE BLOOD STEM AND PROGENITOR CELL FATE ................................................................................................................................. 41
  2.1 ABSTRACT .................................................................................................................................... 42
  2.2 INTRODUCTION .......................................................................................................................... 42
  2.3 COMPUTATIONAL METHODS ..................................................................................................... 44
    2.3.1 Model Assumptions ............................................................................................... 46
    2.3.2 Explicit Time-Dependent Terms .......................................................................... 48
3.3.15 Model Simulations ........................................................................................................ 105
3.4 RESULTS .......................................................................................................................... 105
  3.4.1 Genome-wide expression patterns correlate with cellular functional activity ..... 105
  3.4.2 Gene expression levels predict cellular phenotypes and secretome profiles ..... 111
  3.4.3 Pairwise sample comparisons identify cell population dynamics and signalling
pathways associated with HSC expansion and depletion in vitro .................................. 113
  3.4.4 Inter-cellular network reconstruction ...................................................................... 115
  3.4.5 Inter-cellular network validation ............................................................................ 120
  3.4.6 Model simulations enable functional classification of endogenous regulatory
ligands .................................................................................................................................. 125
  3.4.7 Intra-cellular molecular network integration ...................................................... 127
  3.4.8 Targeted intra-cellular molecular network perturbation ........................................ 133
3.4.9 DISCUSSION ................................................................................................................ 134
4 CONCLUSIONS & FUTURE WORK .................................................................................. 141
  4.1 Summary of results .................................................................................................... 142
  4.2 Future computational work ....................................................................................... 145
    4.2.1 Evaluation of model assumptions ................................................................... 145
    4.2.2 Evaluation of alternative models ................................................................... 150
  4.3 Future model applications ......................................................................................... 153
  4.4 Future inter-cellular network analysis work .............................................................. 159
  4.5 Future intra-cellular network analysis work ............................................................. 160
  4.6 Future validation studies ........................................................................................... 162
  4.7 Broader implications and significance ...................................................................... 164
5 SUPPLEMENTARY TABLES ............................................................................................. 166
6 REFERENCES .................................................................................................................. 186
7 Appendices ...................................................................................................................... 210
  7.1 MATLAB Code ........................................................................................................... 210
    7.1.1 Cellular Balance ODEs ................................................................................ 210
    7.1.2 Cell balance solver; 8days NSNE condition ................................................. 211
    7.1.3 Cell Balance Solver; 8 days, NS/E condition ................................................. 214
    7.1.4 Cell Balance Solver; 8 days, S/NE condition ................................................. 214
    7.1.5 Cell Balance Solver; 8 days, S/E condition ................................................. 214
    7.1.6 Objective function for Parameter estimation – NS/NE, NS/E, S/NE, & S/E
    data ............................................................................................................................ 214
    7.1.7 Objective Function; 8-week culture ................................................................. 215
    7.1.8 Random Parameter Space Search ................................................................. 215
    7.1.9 Randomize parameters .................................................................................. 217
    7.1.10 Phase Portrait Generator ............................................................................. 218
    7.1.11 Local Parameter Sensitivity Analysis ........................................................... 219
    7.1.12 Bootstrap Algorithm ................................................................................. 219
    7.1.13 Fed Batch, constant dilution rate cell balance ............................................ 220
    7.1.14 Fed Batch, proportional dilution rate cell balance ....................................... 220
    7.1.15 Perfusion cell balance .................................................................................. 221
    7.1.16 Feed-back controlled fed batch cell balance based on SF2 concentration ...... 221
7.1.17 Feed-back controlled Fed Batch cell balance based on total cell density........ 222
7.2 Colony Pictures.................................................................................................................. 223

Copyright Acknowledgements................................................................................................. 227
List of Tables

Table 1-1. HSC expansion cultures – design variables and results ............................................. 26
Table 2-1. Estimated compartment – functional assay relationships .......................................... 46
Table 2-2. Model parameters ...................................................................................................... 63
Table 2-3. Cell population outputs used for parameter training ................................................... 67
Table 4-1. 10 most highly connected nodes in the blood stem cell self-renewal network .......... 161

Supplementary Table S1. Published gene sets used for compositional analysis ....................... 166
Supplementary Table S2. Pairwise sample comparisons ............................................................. 167
Supplementary Table S3. B1 - d0 (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>−</sup> Enriched ................................ 167
Supplementary Table S4. B1 - d0 (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>+</sup> Enriched ................................ 167
Supplementary Table S5. B2 - Lin<sup>−</sup> (d4 vs. d0) - d4 Enriched .................................................. 167
Supplementary Table S6. B2 - Lin<sup>−</sup> (d4 vs. d0) - d0 Enriched .................................................. 168
Supplementary Table S7. B3 - Lin<sup>+</sup> (d4 vs. d0) - d4 Enriched ...................................................... 168
Supplementary Table S8. B3 - Lin<sup>+</sup> (d4 vs. d0) - d0 Enriched ...................................................... 168
Supplementary Table S9. B4 - d4 (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>−</sup> Enriched ............................................. 169
Supplementary Table S10. B4 - d4 (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>+</sup> Enriched ............................................. 169
Supplementary Table S11. B5 - Total (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>−</sup> Enriched ............................................. 169
Supplementary Table S12. B5 - Total (Lin<sup>−</sup> vs. Lin<sup>+</sup>) - Lin<sup>+</sup> Enriched ............................................. 170
Supplementary Table S13. B6 - Total (cultured vs. d0) - Culture enriched .................................... 170
Supplementary Table S14. B6 - Total (cultured vs. d0) - d0 Enriched ......................................... 170
Supplementary Table S15. HSC Expansion-Correlated Biological Processes .......................... 171
Supplementary Table S16. Non-Correlated Biological Processes ............................................. 171
Supplementary Table S17. Depletion-Correlated Biological Processes ..................................... 172
Supplementary Table S18. HSC Expansion-Correlated Pathways ........................................... 172
Supplementary Table S19. Non-Correlated Pathways ............................................................... 173
Supplementary Table S20. Depletion-Correlated Pathways ...................................................... 173
Supplementary Table S21. HSC Expansion-Correlated Ligands (16) .......................................... 174
Supplementary Table S22. Non-Correlated Ligands (6) ............................................................. 174
Supplementary Table S23. HSC Depletion-Correlated Ligands (11) .......................................... 174
Supplementary Table S24. HSC Expansion-Correlated Receptors (11) ....................................... 175
Supplementary Table S25. HSC Depletion-Correlated Receptors (12) ....................................... 175
Supplementary Table S26. Ligand-Receptor Interactions ......................................................... 176
Supplementary Table S27. Secreted ligand summary .............................................................. 180
Supplementary Table S28. HSC Self-renewal associated genes (109) ........................................... 181
List of Figures

Figure 1-1. Roadmap for the development of cell therapies .......................................................... 3
Figure 1-2. An overview of the regulatory framework for cell therapies in the United States .... 5
Figure 1-3. The classic hematopoietic hierarchy ................................................................. 18
Figure 1-4. Dynamic interactions between cytokine concentration, cell subpopulation growth, and protein secretion responses determine culture output ........................................... 31
Figure 2-1. Schematic depiction of blood stem cell development model incorporating functional assays and positive and negative feedback .......................................................... 62
Figure 2-2. Statistical analysis of state space reveals critical roles for non-cell autonomous parameters ............................................................................................................... 66
Figure 2-3. Model training methodology and results reveals endogenous inhibitors as key regulators of cell population outputs ............................................................................ 69
Figure 2-4. Parameter sensitivity analysis reveals antagonism between mature and primitive cell compartments ............................................................................................... 70
Figure 2-5. Phase portraits reveal differences in growth kinetics and dynamic sensitivities between mature and primitive cell compartments ...................................................... 74
Figure 2-6. Cell population dynamics in response to consecutive culture perturbations reveals inter-cellular feedback control ................................................................. 77
Figure 2-7. Stochastic variability in secretion rates $sr1-4$ most closely reproduces experimentally observed distribution in culture outputs .................................................................... 78
Figure 2-8. Inter-cellular signalling regulates long-term culture dynamics .............................. 80
Figure 2-9. Phase portraits depict the effect of media exchange efficiency on simulated 16-day culture outputs ................................................................................................ 81
Figure 2-10. Culture initiation parameters affect cell population expansion via inter-cellular signaling ......................................................................................................................... 83
Figure 2-11. Akaike Information Criterion (AIC)-based ranking of control (C) and systematically altered (S1-9) models .................................................................................... 84
Figure 2-12. Loss of responsiveness to self-renewal inhibitor $SF2$ alone is capable of inducing leukemic transformation ............................................................... 86
Figure 3-1. Functional and molecular profiling of hematopoietic progenitor (Lin-) and differentiated (Lin+) cells in serum-free liquid culture ..................................................... 108
Figure 3-2. Average Activity Scores for 56 Published Gene Sets Averaged across 10 Experimental Samples .................................................................................................................. 110
Figure 3-3. mRNA expression indices correlate with cell surface expression and secretion of proteins ................................................................................................................................. 113
Figure 3-4. Reconstruction of inter-cellular network dynamics from gene expression data ...... 117
Figure 3-5. Emergent cell population heterogeneity in vitro .................................................. 118
Figure 3-6. Reconstructed inter-cellular signalling network .................................................. 120
Figure 3-7. Functional validation of cell population and secreted protein effects on progenitor expansion in vitro ........................................................................................................ 122
Figure 3-8. Functional effects of serotonin (5HT1) and the TGF-β inhibitor SB505124 on culture output ............................................................... 123
Figure 3-9. Schematic representation of co-culture bioassay workflow .................................. 124
Figure 3-10. Simulated activities of theoretical proliferation and self-renewal regulatory factors functionally classify experimentally identified ligands ......................................... 126
Figure 3-11. Simulated dynamic activities of theoretical proliferation and self-renewal regulatory factors. ................................................................. 127
Figure 3-12. Integration of endogenous regulatory signals in the HSC intra-cellular self-renewal network. ....................................................................................................................... 131
Figure 3-13. Reconstructed intra-cellular self-renewal signalling network.......................... 133
Figure 3-14. Schematic summary of experimental findings. ................................................ 134
Figure 3-15. Functional Cytokine Interaction Network......................................................... 138
Figure 4-1. Effect on model simulations of varying the assumed probability of self-renewal in fresh umbilical cord blood cells........................................ 148
Figure 4-2. Batch-Fed & Perfusion culture outputs as a function of dilution rate.............. 155
Figure 4-3. Adaptive feedback-controlled bioreactor.......................................................... 156
Figure 4-4. Adaptive feedback controlled fed-batch cultures........................................... 158
Figure 7-1. 8-day cultured cells at 4x (left) and 40x (right) magnification................................. 223
Figure 7-2. BFU-E at 4x magnification. ................................................................................ 223
Figure 7-3. CFU-E at 4x (left) and 40x (right) magnification. ................................................. 224
Figure 7-4. CFU-G at 4x (left) and 40x (right) magnification................................................ 224
Figure 7-5. CFU-M at 4x (left) and 40x (right) magnification................................................ 224
Figure 7-6. CFU-Meg at 4x (left) and 40x (right) magnification.............................................. 225
Figure 7-7. CFU-MEGG at 4x magnification................................................................. 225
Figure 7-8. Mixed colonies at 4x magnification........................................................................ 225
Figure 7-9. LTCIC wells at 4x (left) and 40x (right) magnification.......................................... 226
Figure 7-10. LTCIC bioassay at 4x magnification; negative (left) and positive (right).......... 226
Figure 7-11. LTCIC bioassay at 40x magnification; negative (left) and positive (right)........ 226
1 INTRODUCTION

Portions of this chapter have been published in *Cell Stem Cell* (Kirouac and Zandstra, 2008) and *Current Opinion in Biotechnology* (Kirouac and Zandstra, 2006), co-authored by Peter Zandstra. Authorization to reproduce this work has been obtained from the publisher and co-author.
1.1 Cell therapies as an emerging therapeutic modality

Breakthroughs in molecular cell biology and recombinant DNA technologies in the 1970s lead to a new class of therapeutics - recombinant proteins and antibodies; cell-based therapies may represent the next wave in the development of biologics to treat or cure disease. The emergence of cellular therapies from experimental research is exemplified by the fact that there are currently over 500 cell therapy-based companies worldwide, and the industry prospects have not gone unnoticed by the venture capital community (Parson, 2008). Despite the huge expectations however, only a handful of cell therapy products have made it to market [tissue engineered skin products, which have largely failed, and a recently approved dendritic cell-based immunotherapy], the vast majority being in pre-clinical and early clinical development. Whether or not these potentially curative technologies ever reach (or are even tested on) patients will untimely depend on technologies to manufacture (and deliver) cells in a robust and cost effective manner.

The use of live human cells as a therapeutic modality is by no means a new concept. Blood cell transfusions have been in clinical practice for almost 200 years; the first successful solid-organ transplant (kidney) was performed over 50 years ago, and bone marrow reconstitution following myeloablative conditioning (the only currently available stem cell therapy) was first performed in 1957 (Thomas et al., 1957). Cell based therapeutics are motivated by the observation that many of these and other tissue or organ transplantation procedures yield significant benefits to patients. Unfortunately however, the cells and/or tissues needed to treat the numbers of patients that could benefit from cell transplantation significantly outstrip availability from donors. What differentiates the new generation of cellular therapeutics is that investigators are no longer simply acquiring and processing cells or tissues, but rather engineering and manufacturing cell-based products. The increasing sophistication of such cellular engineering procedures motivates the development of new technologies for the robust and cost-effective production of cellular therapeutics. These technologies will not be successful unless they incorporate cell-level parameters such as heterogeneity, endogenously produced factors and the local physicochemical microenvironment into robust and cost-effective bioprocess engineering strategies. A roadmap is proposed (Figure 1-1) that frames the issues, technologies and design strategies to enable such cellular systems biology-guided cell therapy bioprocess development.
Figure 1-1. **Roadmap for the development of cell therapies.**

Cell therapy development can be divided into 4 phases; discovery, process optimization, production and therapeutic delivery. In the discovery phase, the main issue is product characterization. Functional assays and cellular and molecular profiling are useful tools in this phase. The process optimization phase involves quantifying the relationship between culture parameters and cell output. The rational design of experiments and high-throughput screening using micro-culture platforms can be used to generate empirical cell-based models, which may be integrated with molecular profiling technologies to develop more mechanistic, molecular-based models. Considerations in the production phase include the scale-up strategy and quality control. The final phase is the therapeutic delivery of the cell product. It should be noted that the development phases feedback on one another – during process optimization biological discoveries may be made, and during production the design space will become better defined.
**Figure 1-2. An overview of the regulatory framework for cell therapies in the United States.**

Cell therapies (HCT/Ps) are broken into two regulatory classes in the United States based on risk. HCT/Ps that have been minimally manipulated, and are intended for homologous use (native function) are regulated under the PHSA section 361, and do not require pre-market review by the FDA. HCT/Ps that are intended for non-homologous use and/or have been “more-than-minimally manipulated” are regulated under the PHSA section 351 via the biological products framework, requiring FDA approval of IND applications prior to initiating clinical studies, and BLA approval prior to market. While both HCT/P classes must be handled according to cGTPs, 351 HCT/Ps must also be produced under cGMPs. After sufficient pre-clinical data is generated to suggest safety and efficacy, an IND submission may be filed with the FDA CBER. The main sections of an IND submission are Chemistry Manufacturing and Controls (CMC), Product Characterization, and Pharmacology and Toxicology, all of which pose significant unaddressed challenges for cell therapies. Following completion of all three phases of clinical trials, the clinical data is compiled into a BLA, and if the CBER feels this demonstrates substantial proof of efficacy, approval is granted to market the NBE. While the regulatory bodies and terms vary between jurisdictions, the basic process is quite consistent internationally (ANZTPA, 2007; EMEA, 2001; HealthCanada, 2005).

1.2 Safety, regulatory and manufacturing issues specific to cell therapies

Stem and progenitor cells play a very important role as the “raw material” in manufacturing cell therapeutics, perhaps analogous to the use of *e.coli, Chinese hamster ovary* (CHO) and hybridoma cell lines as the workhorses of the biopharmaceutical industry. Of course the fact that the product is the living cells themselves, not their isolated and enriched protein products, brings forward a different set of manufacturing and safety challenges more akin to viral vaccine production. Specific issues associated with cell transplantation include biological or donor-to-donor variability, microbiological contamination, immunological responses to alloantigens, and tumourigenicity of the transplanted cells. Many of these regulatory and safety issues are well reviewed and outlined in other material (Burger, 2003; Carpenter et al., 2009; Fink, 2009; Halme and Kessler, 2006; Weber, 2006). While guidelines and regulations pertaining to manufactured cell therapies have been established (FDA, 1998), a number of unresolved issues remain, hence it will be challenging to navigate the regulatory path towards clinical approval (Figure 1-2). In particular, standards and methodologies for pre-clinical safety and efficacy evaluation, product characterization, and process validation and control have yet to be adequately developed (FDA, 2008). From a bioprocess engineering perspective the additional safety and regulatory issues that focus on culture-induced changes to input cells and the monitoring and control aspects of the pre-clinical and clinical production systems are particularly relevant. The extensive proliferative potential which makes stem and progenitor cells an attractive cell source, also confers risk associated with tumourogenesis; particularly since there is evidence that some cancers arise from aberrant tissue-resident stem cells (Lapidot et al., 1994). Indeed, there is even evidence that the proliferative burden placed on hematopoietic stem cells (HSC) following bone marrow transplantation plays a role in the increased risk of secondary hematopoietic malignancies and myeloproliferative disorders (Rocci et al., 2007; Sala-Torra et al., 2006). Growing stem cells and their derivatives thus involves balancing cell growth and proliferation (product yield) with culture induced genetic or epigenetic changes (product quality). Strategies to produce cells for cell therapies will need to address a multi-dimensional optimization problem that maximizes target cell output while reducing production costs (necessary for any relevant clinical translation) and minimizing cell culture-associated selective pressure.
1.3 Cell therapy bioprocess optimization and development

The goal of cell therapy bioprocess optimization is to define the conditions that will be brought forward for pre-clinical validation and clinical approval. Not only does the right cell type have to be produced, strategies for its production in a manner that can yield sufficient numbers of cells, under conditions that are appropriate for regulatory approval and are cost effective, need to be defined. Almost all emerging cell therapy candidates start as “open” laboratory-scale cultures (petri dishes, T-flasks, and multi-well plates) wherein controlling, monitoring and evaluating the impact of key parameters on target cell output and productivity is difficult. In fact, arguably, most cell-based biomedical research involves (hypothesis-driven) experiments into how to control cell fate. Rationalization of approaches to interrogate and optimize cell responses could accelerate this endeavour. So how do we move from “cell X can generate cell Y under culture condition P” (biological query) to “Y cells can be generated from X cells per volume * time as a function of culture parameters P₁ through Pₙ” (engineered system). These issues are emerging as increasingly important given the recent observations that cell fate and developmental potential can be broadly influenced using both genetic [iPS cells (Takahashi and Yamanaka, 2006)] and non-genetic (Chen et al., 2006) strategies. To be effective, cell therapy bioprocess design and optimization needs to incorporate a few basic criteria: a) assessment of relevant cell properties; b) measurement and control of key parameters; c) robust predictive strategies for interrogating and evaluating the many parameters that may impact the culture output; and d) approaches to test the many different parameters that may impact cell output in a high throughput and scale-relevant manner.

1.3.1 Measurement of cellular properties

The utility of a cell therapy product is ultimately based on its function in the human patient. Robust functional assays capable of predicting clinical safety and efficacy are hence required. For example, the immuno-deficient mouse assays (Non-Obese Diabetic / Severe-Combined Immuno-Deficient (NOD/SCID), NOD/SCIDβ2m⁺, NOD/SCIDγc⁺, or Rag2⁺γc⁺) used for human blood stem cell repopulation are considered the gold standard for quantifying human hematopoietic stem cells (HSC). However, increasing evidence points to the fact that these assays read out short-term rather than clinically relevant long-term repopulating cells (Horn et
al., 2003), likely related to the low proliferative demand placed on the transplanted cells. For this reason, small animals may also be inappropriate for long-term safety evaluation. For example, HOXB4-transduced HSCs have been assayed hundreds of times in mouse models with no statistically relevant incidence of transformation, however in large animals studies (dog and macaque) high incidences of myeloid leukemias are observed following latencies of 1-2 years (Zhang et al., 2008b). This highlights the importance of utilizing large animal models and extended follow up periods for stringent cell therapy safety and efficacy evaluation. This is particularly relevant for ESC-based therapies, which may be particularly susceptible to generating tumours if the graft contains any undifferentiated cells (Wong et al., 2008). Ensuring the depletion of undifferentiated cells will hence be an important quality control parameter for ESC-based therapies.

Despite the actual or assumed clinical relevance of large-animal studies, it is untenable to consider using these outputs for the development of cell therapy production processes. Cell therapy bioprocesses will hence need to be optimized and controlled based on “real-time” measurements of predictive surrogate markers or assays of cellular function. Real-time is not necessarily used here in the traditional sense, but rather meaning assays with delay times short enough to allow for feedback-based process control, implemented at standardized sampling rates. It is noteworthy that most surrogate assays of cell function have been developed and validated for uncultured stem and progenitor cells. As many have shown (Gan et al., 1997; Zandstra et al., 1997), the correlation between phenotype and function for de novo isolated cells may not be relevant to culture-produced cells. For HSCs, the cell surface expression of the CD34 antigen is rapidly upregulated, and CD38 downregulated upon culture, making the standard CD34⁺CD38⁻ phenotype uninformative for culture-derived cells. Thus better surrogate measures of cellular function and/or culture performance are required, which must be thoroughly validated under the full range of process operating conditions. Identification of such robust surrogate markers is non-trivial. Part of this is due to the use of markers wherein the biology relevant to the cell function of interest is poorly understood or simply correlative. This raises the worrying possibility that the relationship between phenotype and function can be influenced by cell culture parameters. This may be particularly problematic where small numbers of specific measurements are used.
In some cases, it may be necessary to purify the cell population of interest from heterogeneous cell populations produced in culture. This may be achieved either via negative or positive selection strategies, and clinically relevant examples of such procedures include HSC transplantation using CD34⁺ cell-selected (Platzecker et al., 2004) or CD3⁺ T cell-depleted (Gordon et al., 2002) mobilized peripheral blood (MPB) grafts respectively. If positive cell selection is to be used, the possible biological effects of antibody staining must be considered; either due to the activation / blocking of cell surface receptors on the donor cells (Gilner et al., 2007) or possible administration of the antibodies to the recipient (FDA, 2001). Significant limitations exist in technologies available for isolating defined, typically rare cell populations from large numbers of input cells. Current FACS sorting technologies require many hours to isolate purified blood cells from patient samples, and high yields and recoveries of target cells are important. New high speed sorters or highly parallel microfluidic strategies to enable rapid (less that 1hr) high volume multiparameter isolations are important fundamental technologies that are receiving some attention in the bioengineering community (Nagrath et al., 2007).

For HSCs it is established that engraftment (both in murine assays and clinical transplantation) requires differentiated accessory cells in the graft (Bonnet et al., 1999), and the composition of the accessory cell population effects the cell dose – engraftment response relationship (Bonnet et al., 1999). Clinically, CD34⁺ cell dose and myeloid colony forming cell (CFU-GM) are standard indices of clinical function (Jansen et al., 2007), however it has also been shown that additional measures such as megakaryocyte (CD41⁺ and CD61⁺) (Woo et al., 2007) and cytotoxic T-cell (CD8⁺) (Terakura et al., 2007) doses are also predictive of hematological recovery. Similarly, the functional activity of tumour-infiltrating lymphocytes cannot be explained by individual sub-population frequencies, but require complex functions (“sub-population signatures”) due to multi-cellular basis of immune reactivity (Oved et al., 2009). It can hence be argued that the clinical functionality of a graft is an emergent property of the cell population, not amenable to quantification by a single biochemical feature. Consequently, technologies to enable unbiased monitoring of large panels of cellular markers (i.e. the entire set of CD antigens), ideally in real-time, would greatly enhance cell therapy development. Emerging proteomic technologies are
beginning to address these challenges. In particular, cell surface glycan-linked peptide capture technologies, in combination with quantitative mass-spectrometry can be used to quantify hundreds of low abundance cell surface proteins in parallel (Gundry et al., 2008; Wollscheid et al., 2009). The use of panels of markers, or the determination of relevant cell properties that can be measured in real-time would allow for culture system optimization based on overall emergent properties of the target cells, and/or the maintenance of their relationship to the uncultured input cells.

1.3.2 Experimental design and process modeling

Coupling high-throughput bioprocessing with the rational design of experiments (statistical methods for exploring high-dimensional, non-linear systems) can provide high level screens for groups of culture parameters that have an effect (or not) on culture outputs. Rational experimental strategies are necessary, as the design space can be very large (for example investigation of 4 different cytokines/growth factors at 3 different concentrations and 2 time points will yield an experimental matrix of 162 conditions). Further, this analytical strategy can be used to develop empirical models that describe culture output as a function of multiple culture parameters. This approach provides an indication of the shape of the parameter response surface and therefore, may be used to identify sensitive parameters or parameter groups for control strategies thus making the process more robust and ultimately lead to optimized process settings. As a specific example, factorial design experiments and response surface analysis has been used to predict the dose-dependent effects of cytokines on HSC expansion (Audet et al., 2002; Zandstra et al., 1998). This technique has not only been useful for bioprocess development, but may also provide a clearer understanding of the underlying biological mechanisms that govern stem cell fate regulation (Zandstra et al., 2000).

Defining mathematical models of biological systems (as opposed to qualitative conceptual models) serves multiple useful purposes. In addition to the utility in process design and optimization, formulating models forces one to think logically about the relevant system components, interactions, and underlying assumptions, thereby identifying critical system
features and control points (Bailey, 1998). Numerous empirical and mechanistic cell population-based models have been developed for stem and progenitor cell responses to in vitro and in vivo conditions [reviewed in (Viswanathan and Zandstra, 2003)]. In empirical models, exogenous stimuli (within defined conditions) are linked to cellular responses by fitting internal lumped parameters to experimental data. However, deriving robust, predictive mathematical models of biological systems is by no means trivial. Even if the network structure is defined, (i.e. all the relevant variables are identified and the connections between variables defined) two major hurdles generally persist. First, some (usually many) of the variables involved are either hidden, or indirectly available through observation functions, resulting in a partially observed system (Quach et al., 2007). Second, most kinetic parameters are inaccessible and must be estimated by fitting model simulations to experimental data (i.e. reverse engineering). While a number of global optimization algorithms are available to do so (Mendes and Kell, 1998; Moles et al., 2003), none is guaranteed to find an optimal (or any) solution, particularly for highly non-linear systems with multi-modal cost surfaces. Hybrid methodologies, wherein global optimization algorithms are applied in parallel with other metrics (such as system robustness) (Kurata et al., 2007) can be used to evaluate otherwise equivalent parameter sets for biological relevance.

For ordinary differential equation (ODE)-based state space models, techniques and concepts developed in dynamical systems theory can be applied to characterize system properties (de Jong and Ropers, 2006). Most notably Phase Portraits (plotting system dynamics over a range of parameter values or input variables) and Parameter Sensitivity Analysis (PSA) (quantifying the relationship between individual parameters and parameter interactions on system outputs) are common methods used to systematically explore a system’s behaviour in silico. Parameter sensitivities represent emergent properties; features which arise at the system-level, and cannot be deduced by analysing individual components or reactions (Savageau, 1971). Such analysis can be used to analyze overall system robustness, identify critical control points, and determine key perturbations, measurements, and sampling times for optimal experimental design (van Riel, 2006). As a recent concrete example of the usefulness of such analysis, a temporal ligand delivery strategy was designed to maximize ESC self-renewal via modulating JAK-STAT3 pathway kinetics based purely on in silico analysis (Mahdavi et al., 2007). In silico analyses are in fact becoming an integral tool for drug target development in the pharmaceutical industry.
(Henney, 2008), and such approaches may ultimately make the development of more robust cell production systems more efficient.

1.3.3 Multi-dimensional characterization of cell state

A major challenge in the use of modelling approaches to optimize complex cell production processes will be the incorporation of more global-scale measurements of the cell state and microenvironment into useful and predictive tools (Loging et al., 2007). Many bioinformatic software tools and databases are being developed to systematically integrate and interpret these large and disparate datasets, with Systems Biology Markup Language (SBML) surfacing as the industry-standard programming language (Kitano et al., 2005). These tools are ultimately moving in vitro mammalian cell systems along the “systems biology paradigm”; elements $\rightarrow$ networks $\rightarrow$ models $\rightarrow$ phenotypes (Jamshidi and Palsson, 2008). The human genome is of course fully sequenced, and many genomic-scale networks (gene regulation, protein-protein interaction, metabolism, etc…) are being produced at ever-increasing precision and detail, however attaining the kinetic parameters necessary to convert these conceptual networks and correlative relationships into dynamic models is currently the main bottleneck in realizing the potential of systems biology.

Lauffenburger and Sorger have proposed a modelling continuum from detailed molecular mechanisms to global bioinformatic abstractions (Aldridge et al., 2006), and even in the stem cell field it is clear that examining system properties at these different levels of detail will yield insights that will ultimately assist us in optimizing cell production processes. In the absence of systems-based models, biological insights can still be attained by relating network structure and dynamics to cellular phenotypes, and examples abound from a wide variety of applications. A number of microarray studies have compared different types of stem cell populations and differentiated cells in search of a unique stem cell gene expression signature (Akashi et al., 2003; Forsberg et al., 2005; Georgantas et al., 2004; Huttmann et al., 2006; Ivanova et al., 2002; Jaatinen et al., 2006; Komor et al., 2005; Luckey et al., 2006; Park et al., 2002; Phillips et al., 2000; Terskikh et al., 2003; Toren et al., 2005; Zhong et al., 2005). While a stem cell-specific
A unique application of gene expression profiling in bioprocess monitoring and development may be a method to unravel the cell lineages produced and their dynamics. It has been shown that hematopoietic cells display lineage-specific expression profiles, and unbiased hierarchical clustering of gene expression actually recapitulates the developmental hierarchy (Chambers et al., 2007; Kluger et al., 2004). It is in fact possible to classify a cell population into one of the 100+ adult tissue types solely on the gene expression pattern (Greco et al., 2008; Zilliox and Irizarry, 2007). Such an approach has been used to define the developmental capacity of ESC-derived blast cells (Lu et al., 2007), and could be applied on a wider scale to track cell population dynamics and lineage development in culture.

While gene expression profiling can be an extremely useful tool, it is limited in that mRNA transcript abundance accounts for less than 50% of differential protein expression (Tian et al., 2004). For example, semi-quantitative mass spectrometry-based proteomic profiling has identified metabolic differences in HSCs compared to primitive progenitors not evident at the level of gene expression (Unwin et al., 2006). Importantly, genome-scale protein profiling is not necessarily required to obtain an in-depth understanding of the underlying networks governing cell output. By monitoring the phosphorylation status of a limited set of intracellular signalling...
proteins and cellular phenotypes in response to defined exogenous factors, it is possible to build predictive models linking combinatorial ligand stimulation, via signal transduction networks, to cellular responses such as self-renewal, proliferation, apoptosis, and cytokine secretion (Kumar et al., 2007; Miller-Jensen et al., 2007; Prudhomme et al., 2004). Pharmaceutical companies are currently implementing this experimental approach, in combination with the in silico analyses described above, as a platform for drug target discovery (Schoeberl, 2008).

Genome-scale physical protein-protein interaction (PPI) networks (Futschik et al., 2007; Ramirez et al., 2007; Rual et al., 2005) have been used on their own to define a core self-renewal module active in ESCs (Wang et al., 2006b), and but more importantly are emerging as platforms for integrating and interpreting other high-throughput data sets (Bossi and Lehner, 2009; Cline et al., 2007; Hanisch et al., 2002). For example, mapping the entire knowledge base of cancer-associated genes onto genomic-scale PPI networks has defined common topological characteristics (i.e. shared positive and negative feedback loops) and novel signalling connections between oncogenes and tumour suppressors (Cui et al., 2007; Platzer et al., 2007). Similarly, meta-analyses of gene expression in the context of PPI networks has been used to identify core modules, pathways, and interactions underlying process as diverse as drug addiction (Li et al., 2008a) and aging (Xue et al., 2007). In fact, dysregulated gene interactions, rather than expression patterns, have proven more reliable for predicting clinical outcomes of lymphomas (Mani et al., 2008) and breast cancer (Chuang et al., 2007; Pujana et al., 2007; Taylor et al., 2009). Applying such approaches to cell therapy development should allow for the systematic interrogation of cell signalling dynamics and key control points underlying cell fate decisions.

Understanding and controlling metabolism has played an important role in microbe-based bioprocess engineering (Edwards et al., 2001), and modelling steady state metabolic reaction networks via linear programming (so called “flux balance analysis”) is becoming an essential tool for the rational genetic engineering of microbes for industrial biotechnology applications (Famili et al., 2003; Patil et al., 2005). It is likely that the rational design of mammalian cell production bioprocesses will ultimately require an understanding the interface between cell
signalling, gene regulation, and metabolic networks to optimize target cell outputs (Papin et al., 2005). The human metabolic network has recently been reconstructed (Duarte et al., 2007; Ma et al., 2007), and through integration with the On-line Mendelian Inheritance of Man (OMIM) database and Medicare records, has already proven a useful tool for understanding disease prevalence and comorbidity (Lee et al., 2008a).

1.3.4 Process design and control

Ultimately, regardless of the complexity of the tools used, rational strategies to (1) measure system variables (2) analyze functional relationships between these variables and product quality (target cell output) (3) build mechanistic or correlative relationships (mathematical models) between product quality and key process variables, and (4) implement closed loop process control strategies to manage such key process variables should refine our ability to produce stem cell-derived cell therapies. Process modeling has largely not been used in the biopharmaceutical industry as in most other high tech manufacturing, such as the petrochemical and electronics industries. This may in part explain the poor manufacturing productivity achieved in comparison, as dynamic process models are required for the implementation of advanced optimization and control methods (Dutton and Scharer, 2007). The use of process models in pharmaceutical development and production is however being pushed forward by regulatory agencies via the Process Analytical Technology (PAT) initiative launched by the FDA (FDA, 2004), and Quality by Design (QbD) principles outlined in the ICH Pharmaceutical Development (ICH, 2004) and Risk Management (ICH, 2005) guidelines. PAT is defined as systems for the design, analysis, and control of manufacturing to increase product quality and productivity through enhanced process understanding (a principle component of Quality by Design). This essentially translates to the use of mathematical models, statistical design of experiments, and on-line process monitoring and control tools. The principles apply equally to cell manufacturing processes wherein product quality is directly related to the cell populations produced. In addition to their fundamental utility, demonstrating detailed knowledge of the cell manufacturing design space through predictive models should expedite the regulatory process and allow for increased regulatory flexibility with respect to post-IND manufacturing changes.
1.3.5 Product quality control

Robust markers of relevant cellular properties, including functional outputs, are required for cell therapy product characterization and release criteria. This is particularly important in conditions where the cells must be used soon after production (real-time release) or where the mechanism of action (and thus the relevant measurable parameter) is incompletely understood or inaccessible (i.e., in vivo long term insulin responsiveness). As outlined above, cellular systems biology-based strategies may prove useful through the development of models relating critical process parameters to emergent cell properties (product quality). Essentially the parameters and outputs that were used to develop and validate the cell production processes now become the criteria by which the success or failure of any particular production run is measured (process validation).

1.4 Future prospects for cell therapies

It may seem obvious, but given the excitement around stem cell-based therapies it is worth recognizing that only a (small) subset of diseases will ultimately benefit from cell therapies. Careful thinking is required to determine which ones can be treated by pharmaceutical-based endogenous stem cell manipulations, and which ones require the additional complexities of cell and tissue transplantation. Perhaps an initial criterion for candidate cell therapy production systems are the cell-based (as opposed to organ-based) therapies that show promise in a transplantation setting, however are limited by the availability of cells. HSC transplants serve as the best example; insulin regulating cells or MSCs as adjuvants to endogenous healing and transplantation are other near-term therapies. Simple tissue engineered constructs such as skin, cartilage, and bone may also be clinically viable in the near-term, however significant groundwork is still required before complex tissue engineered devises such as heart or liver grafts reach the clinic.

Producing a cell therapy product suitable for clinical use in an economic manner introduces many regulatory and engineering challenges not normally dealt with by research scientists. Novel bioengineering technologies should be integrated into the cell therapy production process as they evolve to further enhance control of cell product identity, reproducibility and efficacy.
Cellular systems biology-based tools will likely play an important role as a strategy to address both unresolved regulatory and engineering challenges. The window of opportunity for the efficient integration of new strategies and processes into cell therapy production processes is still wide open. If cell therapies are to become an integral part of clinical medicine in the coming decades, basic science must navigate through these regulatory and engineering challenges.

1.5 Biology of hematopoiesis

Cell therapy production processes designs must be based on an intimated knowledge of the underlying biology. A review of the cellular and molecular basis of blood cell development will therefore serve to introduce the concepts and issues involved in the design of HSC expansion bioprocesses. The hematopoietic system can be envisioned series of functional compartments, wherein the hematopoietic stem cell compartment gives rise to a hierarchy of increasingly differentiated and developmentally restricted progenitors, eventually producing all of the mature blood cell types. Due to the short life span of most blood cells, a continuous production of enormous numbers of cells is required, estimated to be approximately \(10^{10}\) red blood cells, and \(10^9\) white blood cells per hour (Handin et al., 2002). These numbers, while impressive, belie the complexity of the hematopoietic system as the number of circulating blood cells is maintained within strict limits. Regulatory mechanisms are required to maintain homeostasis in response to external disturbances such as infection and bleeding.

HSCs must constantly “decide” whether to divide, remain quiescent, or undergo apoptosis. Once a cell divides, it may either self-renew or differentiate. Long-term HSCs (LT-HSC), the most primitive cell, self-renew for the lifetime of an individual. The derivative short-term HSCs (ST-HSC) retain limited self-renewal capacity, and give rise to multipotent progenitors (MPP) which then differentiate into oligo-lineage, and eventually uni-lineage restricted progenitors through a series of functional maturation steps, which proliferate and differentiate into mature blood cells (Figure 1-3). The “classic” model of hematopoeisis, elucidated via purification of progenitor subsets and retrospective analysis of \textit{in vitro} and \textit{in vivo} colony outputs, assumed these maturation steps as discrete, irreversible bifurcations. However, these assumptions have been challenged by a functional continuum model, which allows for a certain degree of experimentally observed plasticity and reversibility in developmental potential (Kirkland, 2004; Quesenberry,
Furthermore, the “classic” lineage topology, based around a myeloid-lymphoid dichotomy, has recently been challenged by a number of studies demonstrating the existence of alternative developmental roadmaps (Ceredig et al., 2009; Iwasaki and Akashi, 2007; Kawamoto and Katsura, 2009).

**Figure 1-3. The classic hematopoietic hierarchy.**

LT-HSCs at the apex of the hematopoietic hierarchy differentiate into ST-HSCs and MPPs. The first lineage bifurcation separates the myelo-erythroid from the lymphoid cells, generated from common myeloid and lymphoid progenitors respectively (CMP vs. CLP). CMPs give rise to megakaryocyte-erythroid progenitors (MEP) and granulocyte-monoocyte progenitors (GMP), differentiating into leukocytes, erythrocytes, and platelets, while CLPs differentiate into lymphocytes (T, B, and NK cells). Dendritic cells represent a heterogeneous population, arising though both myeloid and lymphoid lineages. Reproduced from (Larsson and Karlsson, 2005).

Differing models have been advanced proposing that HSC lineage commitment and self-renewal are driven extrinsically through soluble ligands, cell-cell and extracellular matrix interactions, and other deterministic influences (Metcalf, 1998), intrinsically by stochastic processes and
lineage selection (Ogawa, 1999; Till et al., 1964), or both (Robb, 2007; Zandstra et al., 2000). Stochastic models are based on observation that phenotypically homogenous cells, subjected to identical conditions, read out a huge variability in functional outputs (colony number, size, composition) (Leary et al., 1984; Mayani et al., 1993; Suda et al., 1984). However, it is also clearly evident that cell fate decisions (proliferation, self-renewal, and lineage commitment) can be predictably modulated by exogenous signals (Petzer et al., 1996; Rieger et al., 2009; Takano et al., 2004; Zandstra et al., 1997). Deterministic models are thus based on the average response of cell populations in response to exogenous stimuli, smoothing out the stochasticity observed at the single cell level. While genetic studies have defined functional roles for many genes in controlling HSC self-renewal, differentiation, cycling, and apoptosis, the precise molecular circuitry underlying cell fate decisions and the mechanisms by which these processes are coordinated remain poorly understood (Zon, 2008).

1.6 Clinical motivation for HSC expansion bioprocesses

The clinical utility of HSCs arises from their ability to engraft and sustain multilineage hematopoiesis in hematologically compromised hosts. The first clinical HSCT was performed over 50 years ago, a novel approach to treating cancer consisting of high-dose radiation and chemotherapy followed by the intravenous infusion of bone marrow (Thomas et al., 1957). Our understanding of the underlying biology of hematopoiesis has since advanced enormously and been clinically translated such that approximately 50,000 people have now been treated with this therapy worldwide (Appelbaum, 2007). HSCT is now routinely used as a curative treatment of hematopoietic diseases including multiple leukemias, lymphomas, anemias, and genetic diseases of the hematopoietic system (Thomas et al., 1999). Autologous HSCT is also routinely used as supportive therapy for patients with solid tumours undergoing high-dose therapy, and is being tested in numerous clinical trials for the treatment of autoimmune disorders such as type 1 diabetes (Voltarelli et al., 2007), systemic sclerosis (Nash et al., 2007; Oyama et al., 2007), multiple sclerosis (Saccardi et al., 2006), and systemic lupus erythamatosus (Burt et al., 2006).
Umbilical cord blood (UCB) has become an increasingly attractive source of HSCs for a number of reasons. The cells are easy to harvest with no harm to the donor, it is a renewable source, HSCs are enriched in UCB (Wang et al., 1997) and have enhanced proliferative and self-renewal capacities compared to mobilized peripheral blood (MPB) and bone marrow (BM) (Lewis and Verfaillie, 2000; Ramirez et al., 2005; Theunissen and Verfaillie, 2005). Clinically, there is a lower incidence of graft versus host disease (GvHD) and greater tolerance to HLA-disparity using cord blood due to the immunological naivety of the cells (Locatelli et al., 1999). The lower incidence of GvHD may also be related to the enrichment of potent regulatory T cells (Treg) present in UCB (Godfrey et al., 2005) known to suppress this disease (Hanash and Levy, 2005). In addition, the period between request for a graft and transplantation can be much shorter when using UCB [average of 13.5 days (Barker et al., 2002)] compared to BM or MPB. In response to these advantages both public and private cord blood banks have been established worldwide, with over 250,000 cord blood units in storage in public banks alone (Rocha and Gluckman, 2006).

The major limitation of the use of UCB transplantation (UCBT) is the limited number of stem and progenitor cells obtained in each cord blood collection. There exists a strong correlation between the number of cells infused per body mass and the time to hematopoietic recovery (Schoemans et al., 2006), hence the majority of UCB transplants to date have been limited to children (Rubinstein et al., 1998). Depending on the institution, generally less than a third of adults are able to find a suitable UCB unit meeting the current cell dose requirement of $2.5 \times 10^7$ TNC/kg (Brunstein and Wagner, 2006). However, adults now account for about one third of all UCBT recipients (Steinbrook, 2004) despite the fact that the number of cells infused has typically been 10-fold less than when using BM or MPB (Laughlin et al., 2001; Laughlin et al., 2004; Rocha et al., 2004). The delayed hematopoietic recovery and corresponding increased rates of infection-related mortality are largely accounted for by the smaller size of cord blood grafts (Parody et al., 2006). Compounding this, cord blood progenitors give rise to an intrinsically slower hematological recovery, though providing better long-term restoration of the hematopoietic reservoir (Frassoni et al., 2003).
To overcome the cell dose limitation, clinicians have begun combined transplants of 2 partially HLA-matched UCB units. Initial results show the procedure produces clinical outcomes equivalent to single unit UCBT, with no increased incidence or severity of GvHD, one of the major concerns (Barker et al., 2005; Jaing et al., 2007). In combination with reduced intensity (non-myeloablative) conditioning, double unit UCBT was in fact found to result in higher rates of disease-free survival at 3 years in comparison to single unit grafts (Brunstein et al., 2007). Interestingly, one unit consistently comes to predominate after a transient period of dual-chimerism, however the biological or clinical variables determining which unit predominates remain unclear (Haspel et al., 2008). Low dose UCBT has also been supplemented by co-infusion with third party MPB grafts, wherein early engraftment arises from MPB progenitors, which are gradually replaced by UCB-derived cells supporting long-term hematopoiesis. Initial results show faster rates of neutrophil recovery and lower incidence of neutropenia-related infections compared to historic UCBT controls (Bautista et al., 2009; Magro et al., 2006). Methods to increase the cell dose of UCB grafts thus show clear clinical benefits, however combining multiple units is likely not a sustainable long-term solution.

Ex vivo expansion of hematopoietic stem and progenitor cells from cord blood prior to engraftment thus represents an ideal solution to expand the donor pool and reduce post-transplant complications. If the stem / progenitor cells could be significantly expanded (greater than 10-fold), then a single cord blood unit could be used for multiple transplants. Expansion of HSCs would be clinically useful in other situations were cell numbers are limiting such as poor mobilizers for autologous MPB transplantation. A number of phase 1/2a clinical studies have co-transplanted fresh and cultured HSC grafts, demonstrating safety but no clinical benefit (Boiron et al., 2006; Jaroscak et al., 2003; Kogler et al., 1999; Shpall et al., 2002). It should be noted that the culture methodologies used in these studies were not shown to actually expand, or even maintain HSC numbers. In fact, the single Phase1/2a study conducted to date wherein solely expanded cells were transplanted resulted in engraftment failure in 4/4 patients (Holyoake et al., 1997). Future trials utilizing improved culture methodologies may therefore yet demonstrate clinical efficacy (i.e. faster hematological recovery and reduced incidence of graft failure). Culture methodologies have also been described for the generation of large numbers of mature blood cells such as erythrocytes (Giarratana et al., 2005), platelets (Matsunaga et al., 2006),
neutrophils (Timmins et al., 2009), and mast cells (Lappalainen et al., 2007) from HSCs. If integrated with HSC expansion processes, eventually blood products for transfusion may be produced \textit{in vitro}, reducing or even eliminating the need for continual blood donors.

HSCs are also an ideal target for many gene therapies for the treatment of hematological diseases (Aiuti et al., 2007; Aiuti et al., 2002; Klein and Baum, 2004; Ott et al., 2006), as a vector for the systemic delivery of therapeutic proteins (Chang et al., 2006a), or even as a means to generate HIV-resistant lymphocytes (Cohen, 2007). While major safety concerns were raised following the first HSC gene therapy trials due to the high incidence of malignancies (Hacein-Bey-Abina et al., 2003), a result of proto-oncogene activation (Cattoglio et al., 2007; Kustikova et al., 2007; Wagner et al., 2005a), this issue is currently being addressed via engineering safer retroviral vectors (Aiuti et al., 2009).

1.7 Challenges to growing human HSCs in culture

While hematopoietic stem cells are the most extensively studied and arguably best characterized adult stem cell population, a number challenges associated with \textit{in vitro} HSC propagation remain.

1.7.1 Markers

There are no definitive cell surface markers uniquely expressed on human HSCs, although many markers, or combinations of markers can be applied to significantly enrich HSCs (Lin$^{-}$, CD34$^{+}$, CD38$^{+}$, AC133$^{+}$, CD90$^{+}$, CD45RA$^{+}$, Rho$^{lo}$, ALDH1$^{hi}$, c-kit$^{+}$ (Majeti et al., 2007; McKenzie et al., 2007; Wognum et al., 2003)). Isolation of HSCs from tissue is therefore difficult considering the rarity of such cells, at approximately $1/9 \times 10^5$ total nucleated cells (TNC) in umbilical cord blood or $1/3 \times 10^6$ TNCs in adult bone marrow (Wang et al., 1997). This problem is exacerbated during culture as expression of surface markers is modulated by culture conditions, resulting in a dissociation of phenotype from function (Danet et al., 2001; Zandstra et al., 1997).
1.7.2 Assays

In the absence of reliable phenotypic markers, functional assays are used to retrospectively identify and quantify HSCs. In vitro, the colony forming cell (CFC), long term culture-initiating cell (LTC-IC) and cobblestone area forming cell (CAFC) assays quantitatively measure cellular proliferation and differentiation, and output colony formation and phenotype. The ability to (serially) reconstitute and sustain multi-lineage hematopoiesis in a xenotransplant model using immune-deficient (NOD/SCID, NOD/SCID-β2m-/-, or NOD/SCID-γc-/-) mice can be used to measure stem cell activity in vivo (Coulombel, 2004). While exceedingly useful in the fundamental understanding of adult stem cell biology, the retrospective nature of these analyses significantly limits culture development strategies. Acceptable alternatives are needed as culture technologies move towards the clinic and reliable screening and cell-product release criteria are required.

1.7.3 Heterogeneity

In vitro culture analysis must incorporate solutions to at least two heterogeneity-associated problems. At one level, the input cell population contains multiple cell types; a heterogeneity that dynamically changes as a function of culture properties and time. At the second level, HSCs (and other adult stem cells), are heterogeneous in terms of their proliferation kinetics, homing, self-renewal and differentiation capacities (Dykstra et al., 2007; McKenzie et al., 2006; Muller-Sieburg et al., 2002; Sieburg et al., 2006). While it remains controversial whether this heterogeneity is intrinsically or extrinsically regulated, it can confound analysis of the effects of in vitro culture on stem cell properties.
1.8 Regulation of HSC fate

1.8.1 Exogenous manipulations

HSCs are capable of *in vivo* expansion (calculated to be up to 8000-fold (Iscove and Nawa, 1997)) as evidenced by hematopoietic reconstitution from single transplanted HSCs, serial transplanations (up to tertiary) of limiting cell numbers, and HSC recovery following sub-lethal irradiation. This level of expansion may impact cellular genetic stability and development potential due to a loss in teleomere length (Lansdorp, 2005) and oxidative stress (Ito et al., 2006). However, even modest blood stem cell expansion (10-fold) would have a significant therapeutic impact.

*In vitro* cultures have been designed with the goal of reconstructing the so-called HSC niche (Dexter et al., 1977). Co-culture systems have been refined over the years as hematopoietic-supportive feeder cell lines have been created (Chute et al., 2004; Li et al., 2004; Shih et al., 1999; Wineman et al., 1996). Largely these systems have only demonstrated maintenance of HSC numbers (Mayani et al., 1998), likely because they most closely model of steady-state hematopoiesis. The adult *in vivo* stem cell niche is generally homeostatic – motivating the need to replicate either the environment of the developing embryo (Nolta et al., 2002; Wineman et al., 1996; Zhang et al., 2006a), or that of early post-transplantation wherein dynamic changes in stem cell numbers are evident (Iscove and Nawa, 1997). Maintenance of HSCs on some feeders has been shown not to require direct cell-cell contact (Burroughs et al., 1994; Punzel et al., 1999; Verfaillie, 1992), an observation that has driven the development of serum and feeder-free cultures.

Extensive research has been devoted to defining culture conditions supporting HSC expansion, particularly in identifying optimal cytokine cocktails that simultaneously inhibit cell death, induce mitosis, and prevent differentiation (Heike and Nakahata, 2002). In addition to the growth factor cocktail employed many other culture conditions must be optimized, such as the culture vessel employed, seeding density, the cell population for culture initiation, feeding schedule, and length of culture as well as more general considerations such as the composition of the media.
and oxygen tension (Audet et al., 1998). From the large number of studies conducted to date general conclusions can be drawn; HSCs production is improved using lower seeding densities, enriched progenitor populations, media exchange, and high concentrations of multiple early acting growth factors (Balducci et al., 2003; Emerson, 1996; Kohler et al., 1999). However, because the specific effects of each parameter are dependent upon other culture parameters (such as the cytokine cocktail employed (Xu et al., 2000b)) a finding in one culture condition may not be directly applicable to another condition.

Despite significant effort, HSC growth in serum-free cytokine-supplemented liquid suspension cultures has been quite modest; human HSC expansions above 5-fold (relative to input) and in vitro maintenance for greater than 2 weeks has generally been unattainable without cellular transformation (Sauvageau et al., 2004). While recently there have been some modest improvements through the use of novel growth factors such as immobilized delta-like ligand-1 (Suzuki et al., 2006) or angiopoietin-like 5 (Zhang et al., 2008a) or small molecules such as 5azaD (Araki et al., 2006) (Table 1-1), our ability to elicit robust HSCs growth in vitro using exogenous parameters is currently limited by as yet unidentified mechanisms.
<table>
<thead>
<tr>
<th>Source</th>
<th>$P_o$</th>
<th>Media Supplement</th>
<th>time</th>
<th>TNC</th>
<th>CFC</th>
<th>LTCIC</th>
<th>SRC</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCB CD34+</td>
<td>$10^4$</td>
<td>SCF, TPO, FGF1, IGFBP2, ANGPTL5</td>
<td>10</td>
<td>230</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>(Zhang et al., 2008)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$5 \times 10^4$</td>
<td>SCF, FL, TPO, IL3 + 5azaD</td>
<td>9</td>
<td>-</td>
<td>46</td>
<td>11.5</td>
<td>9.6</td>
<td>(Araki et al., 2006)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>?</td>
<td>SCF, FL, TPO, IL3, IL6, + DL1-Fc</td>
<td>21</td>
<td>500</td>
<td>250</td>
<td>-</td>
<td>5.8</td>
<td>(Suzuki et al., 2006)</td>
</tr>
<tr>
<td>BM CD34+CD38-</td>
<td>$10^5$</td>
<td>SCF, FL, IL3, IL6, GCSF (1.5% O2)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1.5</td>
<td>4.2</td>
<td>(Danet et al., 2003)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^6$</td>
<td>SCF, FL, TPO, IL6</td>
<td>6</td>
<td>45</td>
<td>6</td>
<td>-</td>
<td>4.2</td>
<td>(Ueda et al., 2000)</td>
</tr>
<tr>
<td>UCB CD34+CD38-</td>
<td>$10^4$</td>
<td>SCF, FL, IL3, IL6, GCSF</td>
<td>4</td>
<td>-</td>
<td>4</td>
<td>15</td>
<td>4</td>
<td>(Bhatia et al., 1997)</td>
</tr>
<tr>
<td>UCB CD34+CD38-</td>
<td>$10^6$</td>
<td>SCF, FL, IL3, IL6, GCSF</td>
<td>8</td>
<td>-</td>
<td>100</td>
<td>4</td>
<td>2</td>
<td>(Conneally et al., 1997)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$4 \times 10^4$</td>
<td>SCF, FL, TPO, IL7</td>
<td>14</td>
<td>37</td>
<td>2.5</td>
<td>2.5</td>
<td>2</td>
<td>(Lewis and Verfaillie, 2000)</td>
</tr>
<tr>
<td>MPB CD34+</td>
<td>$5 \times 10^4$</td>
<td>SCF, FL, TPO</td>
<td>4</td>
<td>1</td>
<td>1.3</td>
<td>4.8</td>
<td>1</td>
<td>(Danet et al., 2001)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$2 \times 10^3$</td>
<td>SCF, FL, TPO, IL3, FBS</td>
<td>14</td>
<td>250</td>
<td>15.5</td>
<td>128</td>
<td>-</td>
<td>(Piacibello et al., 1997)</td>
</tr>
<tr>
<td>BM CD34+CD38-</td>
<td>$2 \times 10^3$</td>
<td>SCF, FL, IL3, IL6, GCSF</td>
<td>10</td>
<td>-</td>
<td>280</td>
<td>62</td>
<td>-</td>
<td>(Zandstra et al., 1997a)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^3$</td>
<td>SCF, FL, TPO</td>
<td>6</td>
<td>5.7</td>
<td>3.5</td>
<td>61</td>
<td>+</td>
<td>(Herrera et al., 2001)</td>
</tr>
<tr>
<td>BM CD34+CD38-</td>
<td>$2 \times 10^3$</td>
<td>SCF, FL, IL3</td>
<td>10</td>
<td>-</td>
<td>76</td>
<td>49</td>
<td>-</td>
<td>(Petzer et al., 1996)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$4 \times 10^4$</td>
<td>SCF, IL3, IL1b, FBS</td>
<td>7</td>
<td>88</td>
<td>72</td>
<td>17</td>
<td>-</td>
<td>(Moore and Hoskins, 1994)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$5 \times 10^3$</td>
<td>SCF, FL, TPO, IL3</td>
<td>7</td>
<td>21</td>
<td>18</td>
<td>16</td>
<td>+</td>
<td>(Rossmanith et al., 2001)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^4$</td>
<td>SCF, FL, TPO, GCSF</td>
<td>6</td>
<td>1262</td>
<td>120</td>
<td>10</td>
<td>+</td>
<td>(Kobari et al., 2000)</td>
</tr>
<tr>
<td>BM MNC</td>
<td>$10^6$</td>
<td>SCF, FL, IL3, IL6, IL11</td>
<td>14</td>
<td>20</td>
<td>66</td>
<td>9</td>
<td>-</td>
<td>(Zandstra et al., 1997b)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$5 \times 10^4$</td>
<td>SCF, FL, TPO, IL6</td>
<td>12</td>
<td>100</td>
<td>42</td>
<td>4</td>
<td>+</td>
<td>(Kusadasi et al., 2000)</td>
</tr>
<tr>
<td>BM CD34+CD90+</td>
<td>$10^5$</td>
<td>SCF, FL, TPO, FBS</td>
<td>6</td>
<td>4.7</td>
<td>-</td>
<td>3.2</td>
<td>+</td>
<td>(Luens et al., 1998)</td>
</tr>
<tr>
<td>MPB CD34+</td>
<td>$3 \times 10^4$</td>
<td>SCF, FL, IL3</td>
<td>4</td>
<td>30</td>
<td>60</td>
<td>3</td>
<td>+</td>
<td>(Mobest et al., 1999)</td>
</tr>
<tr>
<td>MPB CD34+</td>
<td>$3 \times 10^4$</td>
<td>SCF, IL1b, IL3, IL6, EPO</td>
<td>14</td>
<td>77</td>
<td>4.6</td>
<td>1</td>
<td>-</td>
<td>(Mobest et al., 1998)</td>
</tr>
<tr>
<td>MPB CD34+</td>
<td>$2 \times 10^4$</td>
<td>SCF, IL3, IL6, GCSF, FBS + HS</td>
<td>10</td>
<td>22.1</td>
<td>11.4</td>
<td>0.69</td>
<td>-</td>
<td>(Sandstrom et al., 1995)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$5 \times 10^2$</td>
<td>SCF, TPO, GMCSF, GCSF, FCS</td>
<td>14</td>
<td>5000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(Xu et al., 2000)</td>
</tr>
<tr>
<td>UCB CD34+CD38-</td>
<td>$2 \times 10^4$</td>
<td>SCF, FL, TPO</td>
<td>9</td>
<td>241</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>(Levac et al., 2005)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^4$</td>
<td>SCF, FL, TPO, IL3</td>
<td>8</td>
<td>237</td>
<td>29</td>
<td>-</td>
<td>-</td>
<td>(Kohler et al., 1999)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$2 \times 10^4$</td>
<td>SCF, FL, TPO + 5HT</td>
<td>8</td>
<td>120</td>
<td>75</td>
<td>-</td>
<td>+</td>
<td>(Yang et al., 2007)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$2 \times 10^4$</td>
<td>SCF, FL, IL3, IL6, IL11 (3% O2)</td>
<td>7</td>
<td>60</td>
<td>50</td>
<td>-</td>
<td>+</td>
<td>(Ivanovic et al., 2004)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^5$</td>
<td>SCF, FL, TPO, IL6 + TEPA</td>
<td>21</td>
<td>43</td>
<td>194</td>
<td>-</td>
<td>+</td>
<td>(Peled et al., 2004)</td>
</tr>
<tr>
<td>UCB CD34+</td>
<td>$10^6$</td>
<td>SCF, FL, TPO, IL, IL6, FBS</td>
<td>7</td>
<td>2.3</td>
<td>3.2</td>
<td>-</td>
<td>+</td>
<td>(Yang et al., 2008)</td>
</tr>
</tbody>
</table>

$P_o$ = cell density; FBS = fetal bovine serum; HS = human serum; time = days; TNC / CFC / LTCIC / SRC = fold expansion of the respective cell populations.
1.8.2 Endogenous inter-cellular signalling networks

The question arises as to why the enormous effort put forward to expand HSCs in vitro has produced such limited success. Observations that hematopoietic cell populations secrete large numbers of regulatory proteins (cytokines, chemokines and growth factors) and corresponding receptors in a lineage and differentiation-stage specific manner (Billia et al., 2001a; Cluitmans et al., 1995; Janowska-Wieczorek et al., 2001; Majka et al., 2001; McKinstry et al., 1997) may provide some clues. In addition to the exogenously supplemented growth factors, in vitro culture may expose HSCs to hundreds of endogenously produced regulatory proteins in an uncontrolled manner. Cytokines interact in a non-linear fashion, wherein the dose-response characteristics of one cytokine are continuously modified by the presence of other cytokines. This results in non-additive interactions that are unpredictable based on the dose–response effects of individual factors (Natarajan et al., 2006; Roeder et al., 1998a).

There exist multiple mechanisms by which cytokines can interact, both directly and indirectly. Direct interactions involve the integration of distinct intracellular signalling events such as signalling pathway cross-talk, transcriptional regulation, or modulation of signalling receptor and/or adaptor protein expression. For example, transforming growth factor-beta1 (TGF-β1) inhibits HSC and progenitor proliferation, while interleukin-1-beta (IL-1β), stem cell factor (SCF), Flt3 ligand (FL), and thrombopoietin (TPO) have opposing stimulatory effects (Jacobsen et al., 1994). IL-1β signalling can suppress the inhibitory effects of TGF-β1 both at the level of signal transduction and transcriptional regulation (Benus et al., 2005). TGF-β1 has conversely been shown to inhibit the proliferative effects of the cytokines SCF, FL, and TPO via the reversible downregulation of their respective receptors c-kit, Flt3, and mpl (Fortunel et al., 2003), demonstrating interaction via the regulation of cell surface receptor expression. Indirect cytokine interactions occur through the regulated secretion of secondary factors. For example, in macrophages IL-7, IL-4, and interferon-α (IFN-α) suppress TGF-β1 secretion while macrophage inflammatory protein-1α (MIP-1α) conversely induces TGF-β1 secretion (Maltman et al., 1996). Cellular responses to exogenous stimuli and even transcription factor dysregulation in a variety of cell systems are in fact mediated indirectly through the induction of autocrine/paracrine signalling cascades (Coppe et al., 2008; Deneault et al., 2009; Janes et al., 2006) Despite the fact
that complex intercellular signalling networks are established in culture, the implications of such endogenous signalling networks on culture output, and cellular responses to exogenous stimuli in general, have largely been ignored.

1.8.3 Cell population and micorenvironmental dynamics

As HSCs cannot be purified to homogeneity, cultures are initiated with heterogeneous populations of HSCs, progenitors, and mature cells. As would be expected for different primary cell types, each population has distinct growth factor requirements and a unique and complex pattern of cytokine receptor expression (Billia et al., 2001b). In addition, the lineage development and population dynamics of hematopoietic cell cultures has been demonstrated to correlate with distinct growth factor consumption rates (Koller et al., 1995). Given the spectrum of cells in the hematopoietic hierarchy, it is not surprising that growth factors that specifically induce self-renewal divisions of HSCs have yet to be identified. In fact, increasingly primitive progenitors typically have intrinsically slower rates of proliferation such that progenitor expansion in culture correlates with the stage of maturation. This results in the consistent observation that \textit{in vitro} net growth rate [a sum of individual proliferation, differentiation, and apoptosis rates (Viswanathan et al., 2005)] of total nucleated cells (TNC), progenitors and stem cells decreases in sequential manner (Mobest et al., 1999). As culture progresses, cell population composition changes significantly and mature cell populations accumulate. A dynamic cell population results in a correspondingly dynamic signalling microenvironment as the cytokine expression profile changes.

1.8.4 Inter-cellular feedback

As culture progresses and increasingly mature cell populations accumulate, the balance of signals acting on HSCs is in flux. It has been repeatedly observed that mature cell populations generated in culture, including CD34\(^{+}\) (Gilmore et al., 2000) and CD14\(^{+}\) cells (Xu et al., 2000b) inhibit the proliferation of progenitor populations. Culture-generated Lin\(^{+}\) cells inhibit the expansion of HSCs, mediated at least in part through the secretion of negative regulators such as TGF-\(\beta\)1 and MIP-1\(\alpha\) (Flores-Guzman et al., 2006; Madlambayan et al., 2005), and progenitor
output in both stroma-supported (Koller et al., 1995; Oh et al., 1994; Schwartz et al., 1991) and liquid suspension cultures (Kohler et al., 1999) is enhanced by frequent media exchange and cell-density reduction. Negative feedback circuits are therefore established in culture, with mature cells inhibiting the proliferation and/or self-renewal of primitive progenitors and stem cells.

The negative feedback observed in culture may represent an in vitro manifestation of the homeostatic control mechanisms active in vivo. HSCT studies demonstrate that numbers of stem cells, progenitors, and total cells in the bone marrow and circulation are regulated via feedback control mechanisms. Feedback control is inferred from observations that there exist non-autonomous established “set points” for cell numbers in each compartment, such that cellular expansion in vivo correlates negatively with cell dose transplanted (Iscove and Nawa, 1997). This feedback is mediated at least partially via soluble factors, as protein extract form normal bone marrow interstitial fluid is inhibits, while extract from regenerating marrow stimulates progenitor expansion in vitro (Lord et al., 1976; Wright et al., 1979). While numbers of circulating blood cells and bone marrow progenitors regenerate to normal levels, there exists an extended (indefinite) decrease in HSCs numbers (Pawliuk et al., 1996; Selleri et al., 1999), likely due to the activation of regulatory mechanisms inhibiting HSC expansion as mature cells reach normal levels [limited niche availability may also contribute to this phenomena (Sauvageau et al., 2004)].

In vivo imaging and functional studies have demonstrated directly that normally quiescent LT-HSCs are induced to rapidly enter cycle and self-renew in response to chemical or radiation-induced progenitor depletion, returning to quiescence following regeneration of the bone marrow (Nygren and Bryder, 2008; Randall and Weissman, 1997; Wilson et al., 2008; Xie et al., 2009). Appropriate hematopoietic regeneration requires independent regulation of stem cell proliferation and self-renewal, and evidence suggests that these processes are in fact regulated independently by distinct cell compartments in vivo; bone marrow progenitors (CFU) inhibit self-renewal independent of proliferation effects (Blackett and Botnick, 1981), while mature cells in the circulation normally suppress stem and progenitor cell proliferation in the bone marrow (Cheshier et al., 2007). However, under appropriate circumstances (i.e. infection)
circulating white blood cells can directly induce the proliferation of HSCs through the secretion of inflammatory cytokines (Essers et al., 2009). Such negative feedback regulatory mechanisms in fact appear to be a conserved feature across other tissue stem cell systems, including muscle, epithelial, and neural tissues (Conboy et al., 2005; Giangreco et al., 2009; Lander et al., 2009), as cell-cell communication is essential for homeostasis in multicellular organisms (Torok-Storb, 1988). Local interactions within the niche alone are insufficient to explain the dynamic responsiveness of stem cells to systemic perturbations. Cells in the marrow must be capable of sensing the density composition of cells in circulation, which may explain the close interaction between stem cells and tissue vasculature (Jones and Wagers, 2008).

A simple example of how such feedback mechanisms can be established in culture is provided by the common use of Thrombopoietin (TPO) in HSC expansion cultures. TPO is a potent early acting cytokine that synergizes with Stem Cell Factor (SCF) and Flt3-ligand (FL) to induce HSC self-renewal divisions (Ohmizono et al., 1997). However, TPO also induces the proliferation and differentiation of progenitor cells to megakaryocytes. Megakaryocytes are documented to secrete a number of cytokines including high levels of TGF-β1 (Wickenhauser et al., 1995), a potent inhibitor of HSC proliferation. Therefore, at early stages of culture (when HSCs are typically enriched) TPO induces HSC expansion, while at later stages of culture (when HSCs are diluted by mature cells) this cytokine may inhibit HSC expansion indirectly via megakaryocytic differentiation and corresponding TGF-β1 secretion. This is of course an oversimplification, as megakaryocytes also secrete stimulatory factors, including TPO (Soslau et al., 1997). Ultimately, the dynamic interactions between cytokine concentration, target cell growth, and protein secretion responses will determine culture output (Figure 1-8).
Figure 1-4. Dynamic interactions between cytokine concentration, cell subpopulation growth, and protein secretion responses determine culture output.

Stem cell (blue) expansion \textit{in vitro} is regulated by interactions between exogenous supplied growth factors (XGF) and culture-generated cell populations (red = inhibitory, green = stimulatory, and yellow = neutral). Interactions between cell populations and exogenous growth factors are indicated by arrows (red = inhibitory, green = stimulatory, blue = other types of interactions such as indirect effects), the thickness of the arrow indicating the strength of the relationship. At early stages of culture HSCs are enriched, and exogenous growth factors stimulate the proliferation and self-renewal of HSCs. As culture progresses, mature cell populations accumulate and begin secreting factors capable of modulating stem cell fate. The combined action of culture-generated inhibitory and stimulatory cell populations determines the overall effect on HSC self-renewal; a predominant production of inhibitory cell populations will lead to HSC differentiation and/or quiescence.

The effects of culture-generated mature cells on HSCs are not likely to be solely inhibitory, but rather more balanced and complex. For example, activated CD4\(^+\) T cells and invariant natural killer T cells (iNKT) are required for maintenance of normal hematopoiesis \textit{in vivo} (Monteiro et al., 2005), and iNKT cells have been shown to enhance progenitor proliferation and clonogenic capacity \textit{in vitro} (Kotsianidis et al., 2006). Mature cell populations in fact display lineage-specific functional effects on stem and progenitor cells; platelets (Foss et al., 2008) and NK cells (Fardoun-Joalland et al., 1994) both secrete factors which enhance progenitor expansion, while
macrophages (Xu et al., 2000a) and red blood cells (Cheshier et al., 2007) secrete factors which inhibit progenitor expansion. Stem and progenitor cell fate decisions are regulated by a balance of antagonistic stimulatory vs. inhibitory soluble factors (Cashman et al., 1990; Jacobsen et al., 1994) produced by the various cell populations comprising the local and systemic microenvironment (Wright et al., 1979).

There also exists mechanisms responsible for the inhibition of HSC expansion in vitro other than direct cytokine secretion, such as the production of growth factor-degrading proteases by the CD34⁻ progeny of CD34⁺ progenitors (Goselink et al., 2006). In addition, intracellular changes in HSCs must be considered during culture. For example, the cellular response to cytokines can be modulated via regulation of receptor internalization and degradation kinetics (Vilar et al., 2006) or the regulated expression of intracellular inhibitors of signal transduction such as Suppressors of Cytokine Signalling (SOCS) (Nicola and Greenhalgh, 2000) or the adaptor protein Lnk (Ema et al., 2005), both known to suppress signalling by the early acting growth factors SCF, FL, IL-3, and TPO. Cytokine signalling in hematopoietic cells has also been shown to generate reactive oxygen species (ROS) as mediators of signal transduction, hence sustained high-dose cytokine stimulation in vitro may induce oxidative damage mediated stem cell exhaustion (Iiyama et al., 2006; Ito et al., 2006). In fact, the maintenance of liquid cord blood cultures at low oxygen tension (Koller et al., 1992) and the addition of catalase to long-term murine marrow cultures (Gupta et al., 2006) have both been shown to enhance stem cell expansion, effects which may be mediated through reductions in ROS. The finding that cytokine cocktails that induce HSC self-renewal and expansion initially during culture result in HSC differentiation and depletion at later time points can therefore be explained with a mechanism whereby the impact of specific intercellular and intracellular signalling pathways dynamically change during culture. These results motivate the development of cytokine delivery strategies that target positive feedback loops while limiting negative feedback loop influences.
1.9 HSC cultures as dynamic cellular networks

The analysis presented above suggests two conceptual pitfalls which may be hindering the development of effective HSC expansion cultures, and are likely relevant in other adult stem cell systems. First, individual parameters (cell enrichment, cell number, etc) are generally considered in isolation rather than in the context of multi-parameter interactions. Second, the dynamic nature of HSC cultures is generally not considered with respect to endogenous internal parameters such as cellular interaction networks, with only the input and output states being considered. Essentially, HSC cultures have been thought of as low-dimensional, linear, single-input single-output systems rather than high-dimensional, non-linear, multi-input multi-output systems. The reason for this conceptual impediment is not due to a lack of insight, but rather due to the complexity of such systems. The development of post-genomic high-throughput measurement technologies and associated computational analysis tools is now revolutionizing biological research and can be applied to deconstruct such complex systems.

1.9.1 Mathematical models of hematopoiesis

Stem cell cultures can be envisioned as dynamic coherent systems, with the multiple cell types functioning as system elements (nodes), interacting via complex networks of secreted factors (edges) in the context of exogenously supplied factors. Suspension HSC cultures have distinct advantages as platforms for cellular network analysis as the additional complexities of cell adhesion, migration, and contact-mediated signalling are minimized. Emergent system behaviour arises from such non-linear interaction networks, hence a greater understanding of the structure and dynamics of these cellular networks are required for the rational design of methods for system control.

Mathematical models of hematopoiesis generally represent the hematopoietic hierarchy as discreet compartments, with compartment transitions (survival vs. apoptosis, proliferation vs. quiescence, self-renewal vs. differentiation, and lineage choice) defined by either stochastic or deterministic rules (Viswanathan and Zandstra, 2003). Early stochastic models were able to accurately simulate kinetics of hematopoietic regeneration, colony formation, and differentiation
patterns following HSCT in mice (McCulloch and Till, 1971; Till et al., 1964). Such models have developed as analytical methods of the field have matured, and are capable of explaining processes such as the dynamically unstable patterns of clonal contribution to blood formation following HSCT (Abkowitz et al., 1996; Mangel and Bonsall, 2008; Roeder et al., 2005) and the development of neoplasms (Catlin et al., 2005).

Hematopoiesis is tightly regulated (blood cells in circulation are maintained at relatively constant numbers) and highly responsive to perturbations (bleeding, infection, cytotoxicity, etc…), hence it is highly unlikely that cell fate is regulated by intrinsically set, stochastic mechanisms. The earliest deterministic models of HSC fate incorporated feedback mechanisms to control HSC population size, wherein the rate of entry into cell cycle was directly inhibited by the local HSC “concentration” (Kirk et al., 1968; Kretchmar, 1966; Lajtha et al., 1962). Age structured models, based on systems of ordinary differential equations (ODEs) incorporating population amplification terms and delay times between successive compartments, gradually developed from these initial theoretical studies (Gidali and Lajtha, 1972; Lajtha, 1971; Loeffler and Wichmann, 1980), and have been extended to incorporate multilineage development and feedback control (Schmitz et al., 1990; Wichmann et al., 1988). Such models have recently been applied to interrogate the dynamics of various haematological pathologies (Adimy et al., 2006; Apostu and Mackey, 2008; Colijn and Mackey, 2005a; Colijn and Mackey, 2005b; Crauste et al., 2008; Michor, 2007), neutrophil recovery following HSCT (Scheding et al., 1999) and multicycle chemotherapy (Engel et al., 2004) and targeted molecular therapies (Abbott and Michor, 2006; Michor et al., 2005).

Age structured compartment models have been used to simulate in vitro hematopoiesis; These were originally applied to simulate culture dynamics observed in long-term bone marrow stroma-supported, or “Dexter-type” cultures (Varma et al., 1992; Wichmann et al., 1984) and later, continuous-perfusion bioreactors (Peng et al., 1996). These later models took the addition step of replacing the amplification and delay terms by adding additional hidden compartments – the number of which is estimated based on fitting the cell population balance to experimental observations.
The models discussed so far are based on the assumption of a constant exogenous stimulus, thus the internal parameters (self-renewal and proliferation rates) are culture condition-dependent. Empirical models have been developed to simulate stem cell proliferation in response to exogenous growth factor signalling based on Hill-type functions (Chaudhry et al., 2004). By incorporating terms for cell population heterogeneity into simple dose-response models, the results are much more biologically relevant in that the output is a cell population distribution rather than a scalar term (Deasy et al., 2003; Zhang et al., 2001). However, the additional complexity associated with the growth factor dose-response functions are compensated for by eliminating the compartment hierarchy and considering only the stem cell compartment. These models are thus much too simple to be of use in modeling real hematopoiesis as feedback regulatory mechanisms and system non-linearities cannot be incorporated.

Hybrid deterministic-stochastic models are capable of accounting for cell population heterogeneity by setting internal parameters as probability distributions rather than scalar terms, and/or incorporating stochastic fluctuations in parameter values for individual cells. By allowing exogenous factors to shift the probability distributions, such hybrid models can thus account for the apparent stochasticity at the single cell level and deterministic regulation at the population level (Zandstra et al., 2000). Such hybrid models have been used to simulate in vitro cell proliferation, self-renewal (Viswanathan et al., 2005) and lineage commitment (Glauche et al., 2007).

Numerous studies in recent years have demonstrated the existence of “alternative road maps” in the hematopoietic developmental hierarchy (Ceredig et al., 2009; Kawamoto and Katsura, 2009) and tissue-plasticity in the developmental potential of different stem cell systems (Loeffler and Roeder, 2002). Additionally, it is known that a fluctuating continuum exists in the functional properties of stem cells, modulated by cell cycle status and other physiological properties (Quesenberry et al., 2002). Combined, this data suggests that stem cell development may be better envisioned as a transition through a high-dimensional phase space toward differentiation.
attractors, rather than through a series of discrete, hierarchically organized compartments (Kirkland, 2004). The structure of the inter- and intra-cellular networks would thus define an attractor landscape through which cell populations transverse – cells may explore various paths dependent upon their initial state and exogenous inputs, however the trajectories would eventually converge toward a limited set of attractors (differentiated cell phenotypes) (Huang et al., 2005). This would allow for flexibility in the developmental trajectory and functional capacity of stem cell systems, as observed experimentally. Indeed, cell culture experiments using hematopoietic cell lines (Chang et al., 2008; Chang et al., 2006b; Huang et al., 2005; Huang et al., 2007) and marrow cultures (Sandstorm et al., 1995) support the existence of both inter-cellular and intra-cellular self-organizing networks regulating *in vitro* hematopoiesis.

Inducing robust HSC expansion *in vitro* will require the ability to predictably control endogenous inter- and intra-cellular regulatory networks using external perturbations. However, cell population dynamics, secretion rates of bioactive factors from individual cell populations, and multi-dimensional relationships between cellular composition, secreted factor profiles, and stem cell output must first be defined. Post-genomic tools are required to generate the necessary high-throughput (and ideally high-content) biological measurements. Such datasets can then be fit into multi-dimensional models of *in vitro* hematopoiesis incorporating cellular interaction networks. The development of such models should serve as a starting point for the rational design of blood stem cell expansion bioprocesses utilizing dynamic system perturbations to achieve the preferential expansion of target cell populations.

### 1.10 Tools necessary for the development of predictive models of cellular network output

#### 1.10.1 Stem cell transcriptomics to identify subpopulation dynamics

Microarray analysis of HSC cultures could be utilized to define gene expression programs associated with expansion and differentiation *in vitro*. Comparison of multiple culture conditions at various time points (system perturbations), and relation to functional HSC assays could reveal transcriptional programs and endogenous mechanisms underlying HSC self renewal
vs. differentiation in culture. This is analogous to analyzing transcriptional profiles of multiple diseased and healthy tissues for the identification of underlying biological features, target pathways for therapeutic intervention, or the prediction of disease outcome based on gene expression signatures (Rhodes et al., 2004). The dynamic cellular heterogeneity and small target cell population presents a problem for transcriptional (or any other molecular level) profiling, as differential gene expression will result from changes in cell population composition. However, if the focus is set on global changes and endogenous regulatory mechanisms in culture rather than isolated HSCs, it may be advantageous to maintain the cell population heterogeneity.

Additional information can be obtained by separating the culture-generated cell populations based on phenotype and transcriptionally profiling the purified cell populations. Such an approach has been taken by Rendl et al. (Rendl et al., 2005) seeking to analyze intercellular regulatory relationships between epithelial stem cells and their niche within the hair follicle microenvironment in vivo. Focusing specifically on the expression of secreted proteins and cell surface receptors, putative intracellular regulatory networks can be constructed from such data and substantiated by the expression of downstream signal transduction components and responsive genes (Graeber and Eisenberg, 2001). Comparison of network structure between multiple culture conditions and time points would provide functional relations between intercellular networks and HSC output.

1.10.2 Stem cell proteomics to measure extra- and intra-cellular cytokine signalling

Several innovative studies have attempted to define the HSC proteome using mass spectrometry-based techniques. However, at the current state such proteomic analysis is limited in terms of breadth and sensitivity, as only a small fraction of highly abundant proteins can be detected when working with limited numbers of primary cells (Tao et al., 2004; Zenzmaier et al., 2005). Thus using current technology, broad screens of intracellular proteins are unlikely to yield information directly applicable to culture development.
The direct identification of proteins secreted in HSC cultures would provide key insights into the structure and dynamics of endogenously established intercellular signalling networks. Proteomic analysis of conditioned media ("secretomics") is technically challenging, primarily due to the high concentrations of exogenous media proteins compared to endogenously produced factors. Consider that culture media typically contains 10mg/ml albumin, 10-100µg/ml insulin and transferrin, and 10-100ng/ml recombinant growth factors, while endogenous secreted cytokines may display activity at low pg/ml levels, representing a dynamic range of nine orders of magnitude. A common method of overcoming this difficulty is to culture cells in serum or albumin-containing media until confluence or a desired cell density is attained, then switch to a protein-free media, which is conditioned for a given amount of time and concentrated prior to analysis (Mbeunkui et al., 2006). Such an approach has been used to define the secretome of skeletal muscle (Gajendran et al., 2002), mouse embryonic fibroblasts used as feeder cells for ESCs (Lim and Bodnar, 2002), and human macrophages (Dupont et al., 2004). However, the majority of proteins identified in such studies are intracellular, raising the possibility that these components are the result of cell death in culture, likely due to the switch to protein-free media for conditioning.

Novel methodologies which specifically enrich glycoproteins (i.e. secreted cytokines) and deplete abundant serum proteins such as albumin and transferrin may now allow for the direct mass spectrometry-based proteomic analysis of albumin-containing conditioned media (Zhang et al., 2003). Stable isotope labelling of the enriched glycoproteins prior to analysis additionally allows for semi-quantitative sample comparison. The ongoing development of antibody array (Barry and Soloviev, 2004) and microsphere-based technologies (Carson and Vignali, 1999) provide alternative solutions for the direct analysis of secreted proteins.

1.10.3 Systems-level dataset integration

A primary challenge for 21st century science remains to assimilate the explosion of genomic-level data to extract biological information; the system structures, dynamics and methods of control (Brent and Bruck, 2006). Incremental steps are being made towards this goal through the
development of increasingly sophisticated bioinformatic platforms (Kersey and Apweiler, 2006; Ng et al., 2006; Toyoda and Wada, 2004). A systems-level model of in vitro hematopoiesis will require the integration of multiple levels of biological data, relating transcriptional activity, intracellular signalling, and intercellular interaction networks to cellular composition and functional outputs.

1.11 Hematopoietic stem cell bioprocess design

Analogous to the comparison of intercellular gene regulatory network structure under diverse stimuli (Luscombe et al., 2004), comparisons of cellular interaction network structure and dynamics under HSC supportive and non-supportive conditions should allow for the functional annotation of network structures to guide bioprocess development. An ideal model would incorporate terms for a heterogeneous cell population with parental relationships, specific growth kinetics for each cell population, and functional intercellular signalling networks between populations. It will be interesting to see whether the global architecture of such cellular interaction networks follow scale-free hierarchical organisation similar to intra-cellular molecular networks (Almaas et al., 2004; Barabasi and Albert, 1999; Barabasi and Oltvai, 2004; Jeong et al., 2000; Rual et al., 2005; Song et al., 2005). Biological network process diagrams, such as those described by Kitano et al. (Kitano et al., 2005) for modelling intracellular signalling, could be adapted with cell populations represented as state nodes and functional interactions represented by transition nodes and edges. The value of such a model will lie in guiding the rational design and in silico optimization of next generation HSC expansion bioprocesses.

1.12 Hypothesis and approach

Hematopoietic stem and progenitor cell fate is regulated by inter-cellular signalling networks. Understanding the structure and dynamics of such networks would allow for their systematic perturbation to control stem cell expansion in culture.
In chapter 2 we present a novel mathematical model of blood stem cell development incorporating cell level kinetic parameters as functions of inter-cellular communication networks. The model connects internal model parameters and microenvironmental variables to measurable cell fate changes, suggests new strategies for culturing blood stem cells in vitro, and describes leukemic transformation as resulting from dysregulated responsiveness to inter-cellular signals. As the intercellular signals in the model are purely theoretical, in chapter 3 we integrate transcriptome and proteomic profiling with extensive database mining to experimentally identify and functionally validate both inter- and intra-cellular networks regulating cell fate in vitro. The networks can be represented as a coupled positive/negative feedback circuit, wherein megakaryocyte-derived growth factors stimulate, and monocyte-derived chemokines and cytokines inhibit the expansion of primitive stem and progenitors in culture. In chapter 4 we outline how these findings may be applied and extended upon for next generation HSC bioprocess designs. In summary, this body of work demonstrates that stem and progenitor cell fate is regulated non-autonomously via interactions with multi-lineage differentiating progeny, and cell fate decisions can be modulated indirectly via targeted manipulation of these interactions.
CELL-CELL INTERACTION NETWORKS REGULATE BLOOD STEM AND PROGENITOR CELL FATE

Portions of this chapter have been published in Molecular Systems Biology (Kirouac et al., 2009a) Authorization to reproduce this work has been obtained from the publisher and co-authors.
2.1 ABSTRACT

Communication networks between cells and tissues are necessary for homeostasis in multicellular organisms. Inter-cellular (between cell) communication networks are particularly relevant in stem cell biology, as stem cell fate decisions (self-renewal, proliferation, lineage specification) are tightly regulated based on physiological demand. We have developed a novel mathematical model of blood stem cell development incorporating cell-level kinetic parameters as functions of secreted molecule-mediated inter-cellular networks. By relation to quantitative cellular assays, our model is capable of predictively simulating many disparate features of both normal and malignant hematopoiesis, relating internal parameters and microenvironmental variables to measurable cell fate outcomes. Through integrated in silico and experimental analyses we show that blood stem and progenitor cell fate is regulated by cell-cell feedback, and can be controlled non-cell autonomously by dynamically perturbing inter-cellular signalling. We extend this concept by demonstrating that variability in the secretion rates of the inter-cellular regulators is sufficient to explain heterogeneity in culture outputs, and that loss of responsiveness to cell-cell feedback signalling is both necessary and sufficient to induce leukemic transformation in silico.

2.2 INTRODUCTION

Systems-biology research to date has primarily focused on elucidating the topological features and dynamics of intra-cellular networks, with the implicit assumption of cell populations as homogenous, autonomous units. Inter-cellular communication networks, represented with cell types as vertices, and functional (rather than molecular) interactions as edges, have been largely unexplored.

Communication networks between cells, tissues, and organ systems are necessary for homeostasis in multicellular organisms. For example, soluble factor-mediated cell-cell networks play a dominant role in orchestrating immune reactions via “cytokine cascades” in response to
infection. Inter-cellular communication networks are particularly relevant in stem cell biology, as stem cell fate decisions (self-renewal, proliferation, lineage specification) are tightly regulated based on physiological demand and responsive to external perturbations. Evidence suggests that stem cell dysregulation is fundamental to the progression of multiple cancers, degenerative diseases, and general aging phenomenon (Rossi et al., 2008). Stem cell-based therapies are hence emerging as a foundational tool in regenerative medicine. One of the key challenges lies in controlling the emergent cellular and microenvironmental complexity that arises as stem cell populations develop in vitro and in vivo (Kirouac and Zandstra, 2006).

Hematopoiesis, the process by which blood cells develop, serves as a prototype for other stem cell systems. Hematopoietic stem cells (HSC), at the apex of a developmental hierarchy, give rise to a series of increasingly differentiated and developmentally restricted progenitor cells, eventually producing all of the mature blood cell populations. In vivo, HSC fate decisions are regulated by cross-talk with neighbouring cell populations either directly or via secreted factors (Wilson and Trumpp, 2006). Much experimental and theoretical work has been conducted to understand the structure and dynamics of the homeostatic control mechanisms in vivo linking blood cells in the circulation to stem cells in the bone marrow (Lajtha et al., 1962). Evidence suggests that mature blood cells suppress proliferation and differentiation of progenitors, and progenitors correspondingly suppress the expansion of stem cells via coupled negative feedback loops (Wichmann and Loeffler, 1985).

Hematopoietic cells are known to secrete and respond to a large numbers of regulatory proteins in lineage- and differentiation stage-specific patterns (Billia et al., 2001a; Majka et al., 2001). This results in complex and dynamic inter-cellular signalling networks, providing a mechanism by which cells interrogate and interpret their local microenvironment (their niche), propagate this information through signal transduction and gene regulatory networks, and respond by modulating cell fate decisions. A number of studies have attempted to reconstruct the intra-cellular molecular networks regulating stem cell fate (Muller et al., 2008), including mathematical modelling of genetic regulatory networks and intra-cellular feedback mechanisms (Glauche et al., 2007). However, inter-cellular regulatory mechanisms remain largely undefined.
Motivated by the fact that hematopoietic stem cell transplantation is a curative therapy for a number of hematopoietic and immunological diseases, herein we explore the behaviour of intercellular regulatory networks as tools to regulate cell fate during in vitro human blood stem cell propagation. We have developed and utilized for predictive purposes a novel mathematical model of in vitro hematopoiesis by linking functional cellular assays to specific model outputs, by defining cell-level kinetic parameters such as cell cycle rates and self-renewal probabilities as functions of culture variables, and by simulating feedback regulation using cell-cell interaction networks. Our resultant model captures many facets of hematopoiesis, connecting internal model parameters and microenvironmental variables to measurable cell fate changes. We show that negative feedback signalling between differentiated cells and stem and progenitor populations is a dominant factor regulating culture output, and that stem cell fate can be controlled non-autonomously by the dynamic perturbation of cell-cell signalling networks. We extend this concept by demonstrating that variability in the secretion rates of inter-cellular regulators are sufficient to explain variability in culture output, and that loss of responsiveness to cell-cell feedback signalling is both necessary and sufficient to induce leukemic transformation in silico. The development of quantitative models incorporating cell-cell regulatory networks should serve as an important tool to understand and control emergent cellular complexity in vitro and in vivo (Kirouac and Zandstra, 2008).

2.3 COMPUTATIONAL METHODS

The hematopoietic hierarchy can be divided into a number of discrete compartments, from long-term repopulating hematopoietic stem cells (LT-HSC) to fully differentiated mature cells. Each compartment can be viewed as representing a cell population at a distinct state of maturation, with unidirectional transition between compartments (differentiation) associated with cell cycling. While recent evidence shows that LT-HSCs may undergo functional transitions (differentiation) prior to mitosis (Kent et al., 2008b), differentiation is generally co-incident with population amplifying cell divisions. For simplicity, lineage specification is considered only after differentiation (Lin’). A cell population balance can be constructed around each compartment \( i \) where the number of cells in the compartment \( X_i \) is dependent upon the
number of cells entering from the previous compartment \((X_{i-1})\), the cell proliferation rate \((u_i)\), and the probability of self-renewal \((f_i)\). This is a deterministic model, which nonetheless can be viewed as incorporating stochastic elements whose impacts are negligible at the population-level.

The cellular growth rate for compartment \(i\) is given by the equation:

\[
\frac{dX_i}{dt} = (1 - f_{i-1})u_{i-1}X_{i-1} + (2f_i - 1)u_iX_i \quad \text{for } i = [1, 2, 3, \ldots, n]
\]  

(1)

A system of ordinary differential equations (ODE) is therefore constructed which describes the growth of each cellular compartment for a total of \(n\) compartments, with compartment 1 \((X_1)\) representing LT-HSCs, and terminally differentiated mature cells represented by compartment \(n\) \((X_n)\). Specific compartments can be ascribed to experimentally measurable cellular assays. The functional measures considered are long-term Non-Obese Diabetic (NOD)-Scid mouse repopulating cells (SRC), long-term culture-initiating cells (LTCIC), and colony forming cells (CFC), which readout stem cells, primitive progenitors, and mature progenitors respectively. We additionally characterize the cells phenotypically as undifferentiated Lin\(^-\) or differentiated Lin\(^+\).

Starting with a stem cell \((i = 1)\), the number of cells in compartment \(i\) that can be generated via symmetric differentiation divisions is given by:

\[
X_i = 2^{i-1}
\]  

(2)

And subsequently the total number of cells \((X_T)\) at differentiation stage \(n\) generated from a stem cell is given by:

\[
X_T \geq \sum_{i=1}^{n} X_i = 2^n - 1
\]  

(3)

Hence the total number of compartments \((n)\) read-out in a given functional assay can be estimated via rearrangement of equation (4):

\[
n \geq \frac{\ln(X_T + 1)}{\ln(2)} \approx \frac{\ln(X_T)}{\ln(2)}
\]  

(4)
2.3.1 Model Assumptions

We assume the SRC assay quantifies LT-HSCs with absolute accuracy, although this assumption may be disputed. Estimates derived directly from equation (4) based on experimentally measured frequencies in UCB total nucleated cells (TNC) are presented in Table 2-1.

Table 2-1. Estimated compartment – functional assay relationships.

<table>
<thead>
<tr>
<th>CELL POPULATION</th>
<th>FREQUENCY</th>
<th>COMPARTMENT(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>~ 1/10^6 TNC</td>
<td>1</td>
</tr>
<tr>
<td>LTCIC</td>
<td>~ 1/4×10^4 TNC</td>
<td>1-8</td>
</tr>
<tr>
<td>CFC</td>
<td>~ 1/5×10^3 TNC</td>
<td>1-11</td>
</tr>
<tr>
<td>Lin'</td>
<td>~ 1/100 TNC</td>
<td>1-13</td>
</tr>
<tr>
<td>TNC</td>
<td>-</td>
<td>1-20</td>
</tr>
</tbody>
</table>

Only the last compartment within a given population is considered, and the first compartment is set to 1 (LT-HSC). While LT-HSCs will not readout in the CFC assay, and it is unclear whether LT-HSCs readout in the LTCIC assay (Coulombel, 2004), placement of the first compartment has negligible impact on resulting calculations. Similarly, the self-renewal potential associated with HSCs and immediate descendants will increase the calculated total number of compartments (n), however a few additional divisions in the early stages of the hierarchy will have negligible effect on the population balance calculations. The simplifying assumption that \( X_{i+1} = 2X_i \) will likely not hold true across all compartments for a system in steady state, and this amplification coefficient may in fact be a dynamic parameter. However, in lieu of experimentally defined in vivo kinetic parameters, it is a reasonable estimate supported by our and others experimental data. It is notable that our calculation of \( n = 20 \) is consistent with previous theoretical and experimental-based estimates ranging from 17 to 30 (MacKey, 2001; Shochat et al., 2002).
The self-renewal probabilities \((f_i)\), proliferation rates \((u_i)\), and cell death rates must be specified for each compartment in order to solve the cell population balance ODEs defined in Equation 1. Because these internal variables are inaccessible to experimental measurement, we chose to estimate parameters based on generalized functions and a reverse engineering strategy to estimate specific values within experimentally or biological constraints.

Stimulation with high concentrations of hematopoietic growth factors such as Stem Cell Factor (SCF), Flt3 Ligand (FL), and Thrombopoietin (TPO) are known to inhibit apoptosis, and cultures are maintained such that concentrations of nutrients (i.e. amino acids) and waste products (i.e. lactate) are not limiting. It is therefore reasonable to assume the cellular death rate to be negligible.

It has been documented that the proliferation rate \((u_i)\) of hematopoietic cells varies with the stage of maturation such that the proliferation rate \((u_i)\) of mature progenitors (CFC) > primitive progenitors (LTCIC) > stem cells (SRC) (Mobest et al., 1999). The in vitro conditions under study do not allow for the functional maturation of differentiating cells, and this can be accounted for by setting a lower proliferation / differentiation rate for the Lin\(^+\) cells. We define the proliferation rate of Lin\(^-\) cells as a function of compartment number by a Gaussian-type function:

\[
u_i = u_{MAX} \exp \left[ \frac{-(i - n_{MAX})^2}{2D_{GR}^2} \right] \tag{5}\]

Where \(n_{MAX}\) is the compartment with the maximum proliferation rate (analogous to the mean), \(u_{MAX}\) is the proliferation rate of this compartment, and \(D_{GR}\) is the growth rate decay term (analogous to the variance), which defines the steepness of the change in cycle rates between successive compartments. Physiological limits bound all three parameters to between 1 and 10.

By definition, only the LT-HSC population has the capacity for self-renewal probability \((f_i)\) to exceed 50% in vivo, however ST-HSCs must by definition also have some capacity to self-renew
(lower than 50% in vivo), and downstream progenitor populations have also been documented to undergo limited self-renewal divisions (Marley et al., 2003). Self-renewal probabilities \( f_i \) should therefore diminish with the stage of differentiation, as described theoretically in (Roeder and Loeffler, 2002). We define the probability of self-renewal \( f_i \) as a function of compartment number by a Gaussian-type function with maximum set at \( i = 1 \):

\[
f_i \sim f_{\text{MAX}} \exp \left[ \frac{-(i - 1)^2}{2D_{SR}^2} \right]
\]

Where \( f_{\text{MAX}} \) is the maximal self-renewal probability of the LT-HSC compartment \( (X_1) \), (limited to \([0, 1]\)) and \( D_{SR} \) is the self-renewal decay term (analogous to the variance), which defines the steepness of the change in self-renewal probabilities between successive compartments, bounded between 1 and 10 based on physiological limits.

### 2.3.2 Explicit Time-Dependent Terms

When purified hematopoietic progenitor cells are placed in culture, there is an initial lag time before the cells enter cycle (G0 to S/G2/M/G1 transition), after which cells cycle with approximately constant doubling time. A lag phase is therefore introduced into the system via the explicit time-dependant function:

\[
U_i \sim \left( \frac{t^{kt}}{\tau_D^{kt} + t^{kt}} \right)
\]

Where \( t \) is the culture duration (days), \( \tau_D \) is the time for 50% of the cells to enter cycle, and \( kt \) is the Hill coefficient defining the rate at which cells are induced to cycle (approaching a step function at \( \tau_D \) as \( kt \to \infty \)). Based on experimental observations (Ko et al., 2007), setting \( \tau_D = 2 \) days and \( kt = 4 \) produces a reasonable kinetic response. For simplicity this lag phase term is applied equally to all cell compartments, while in reality there would likely be differences between HSCs, progenitors, and mature cells in their rate of entry to cycle upon stimulation (Punzel et al., 2002).
2.3.3 Non-Linear Terms

It was our goal to incorporate as few structural assumptions as possible into the model so as to minimize any systemic bias. For models containing few compartments (i.e. 3 or 4) it is possible to use random methods to structure the intercellular regulatory relationships (positive feedback, negative feedback, feed-forward), followed by optimization algorithms to select the top performing network topologies (Socolovsky et al., 2007). However, for a 20-compartment model this approach would be computational unfeasible due to the combinatorial explosion of possible topologies. Hence, we utilized the wealth of experimental and theoretical literature on hematopoietic regulation to define a candidate intercellular network topology.

The intercellular regulatory network topology described in (Wichmann and Loeffler, 1985) consisting of 3 inter-related negative feedback control loops is capable of simulating in vivo hematopoietic response to multiple perturbations. The topology consists of (1) auto-regulatory HSC feedback (2) intra-medulary short-range feedback and (3) blood-bone marrow feedback. While the mechanisms of feedback control are suggested to be soluble factors, the model is formulated by setting proliferation and self-renewal rates as directly responsive to cell densities, implying a linear relationship between cells and secreted factor concentrations. We expanded upon this concept by considering proliferation rates ($u_i$) and self-renewal probabilities ($f_i$) as regulated by a balance of secreted inhibitory and stimulatory factors ($SF1-4$). Mature ($Lin^+$) cell sub-populations secrete factors ($SF1$) that inhibit the proliferation of progenitors ($Lin^-$), and progenitors ($Lin^-$) in turn secrete factors ($SF2$) that inhibit the probability of self-renewal of stem and primitive progenitor cells. These negative feedback loops are coupled as $SF2$ secretion is induced by $SF1$. We additionally consider two factors secreted by mature ($Lin^+$) cell sub-populations, $SF3$ and $SF4$, which stimulate progenitor proliferation and self-renewal respectively (positive feedback). This regulatory architecture is described by the following set of ODEs; secretion of $SF1$ through $SF4$ is given by:

$$\frac{d}{dt}[SF1] = srI\left(\sum_{14}^{20} X_i\right)$$  \hspace{1cm} (8)
\[ \frac{d[SF2]}{dt} = sr2 \left( \frac{[SF1]^{Ks}}{Ls^{Ks} + [SF1]^{Ks}} \right) \left( \sum_{i}^{13} X_i \right) \]  \hspace{1cm} (9)

\[ \frac{d[SF3]}{dt} = sr3 \left( \sum_{i}^{20} X_i \right) \]  \hspace{1cm} (10)

\[ \frac{d[SF4]}{dt} = sr4 \left( \sum_{i}^{20} X_i \right) \]  \hspace{1cm} (11)

And the dose response effects on proliferation rates and self-renewal probabilities given by coupled Hill-type functions:

\[ u_i \sim \left( \frac{L_i^{K1}}{L_i^{K1} + [SF1]^{K1}} \right) \left( \frac{L_i^{K3}}{L_i^{K3} + [SF3]^{K3}} \right) \]  \hspace{1cm} (12)

\[ f_i \sim \left( \frac{L_i^{K2}}{L_i^{K2} + [SF2]^{K2}} \right) \left( \frac{L_i^{K4}}{L_i^{K4} + [SF4]^{K4}} \right) \]  \hspace{1cm} (13)

As factors SF1 through SF4 are theoretical, the corresponding terms L1 through L4, representing ligand concentrations producing half-maximal dose response, are arbitrary and hence set to 1 for simplicity. Equations 13 and 14 then reduce to:

\[ u_i \sim \left( \frac{1 + [SF3]^{K3}}{1 + [SF1]^{K1} + [SF3]^{K3}} \right) \]  \hspace{1cm} (14)

\[ f_i \sim \left( \frac{1 + [SF4]^{K4}}{1 + [SF2]^{K2} + [SF4]^{K4}} \right) \]  \hspace{1cm} (15)

We can then set practical bounds on secretion rates sr1 though sr4 based on dynamic dose-response relationships. For cell densities in the range of \( 5 \times 10^5 \) cell/mL, secretion rates of \( 10^{-7} \)
to $10^{-5}$ pg/cell.day result in 50% maximal response reached between 0.5 to 50 days, and we similarly set $Ls = 10^{-2}$ to $10^1$. Hill coefficients ($k_1-4$ and $ks$) for the terms are set between 0.1 and 10, covering the range of relevant sigmoidal dose response curves. Combining equations (5), (7), and (14), and (6) and (15) to define $u_i$ and $f_i$ respectively as functions of compartment number ($i$), time ($t$) and secreted factor concentrations ($SF1-4$) generates the functions:

$$u_i = u_{MAX} \exp \left[ -\frac{(i - n_{MAX})^2}{2D_{GR}^2} \right] \times \left( \frac{t^k}{\tau_D^{u_t} + t^{k_t}} \right) \times \left( \frac{1 + [SF3]^k}{1 + [SF1]^k + [SF3]^k} \right)$$  \hspace{1cm} (16)$$

$$f_i = f_{MAX} \exp \left[ -\frac{(i - 1)^2}{2D_{SR}^2} \right] \times \left( \frac{1 + [SF4]^k}{1 + [SF2]^k + [SF4]^k} \right)$$  \hspace{1cm} (17)$$

Such that ($0 \leq u_i \leq u_{MAX}$) and ($0 \leq f_i \leq f_{MAX}$). It should be noted that $SF1$ through $SF4$ most likely represent groups of ligands rather than single factors. Additionally, secretion of $SF1$, $SF3$, and $SF4$ by Lin$^+$ cells can be equivalently interpreted as a single homogenous population secreting all three factors, three distinct populations each secreting a single factor, or more likely, a combination thereof. Within the heterogeneous Lin$^+$ population, there are likely sub-populations biased toward $SF1$, $SF3$, and $SF4$-type secretion profiles, however this additional complexity is not considered. Proliferation rates and self renewal probabilities of the progenitor populations will hence be dependent on the exogenous growth factor(s) provided (setting $u_{MAX}$, $u_+$, $n_{MAX}$, $f_{MAX}$) and the balance between inhibitory and stimulatory factor accumulation in the culture microenvironment ($SF1$: $SF3$, $SF2$: $SF4$).

### 2.3.4 Parameter Estimation

The resulting model contains 16 undefined parameters ($P_1$ through $P_{16}$) (see Table II). Data from (Madlambayan et al., 2005) (see Table III) was used as a training set to estimate parameters via a reverse engineering strategy. The proportion of differentiated (Lin$^+$) cells produced ($%Lin^+$) and cell population fold-expansion values ($TNCX$ = total nucleated cell, $CFCX$ = colony forming cell, $LTCICX$ = long term culture-initiating cell, $SRCX$ = scid repopulating cell) represent observable functions ($O_i$) of system variables ($X_i$);
Where \( t_0 \) and \( t_f \) represent culture initiation and analysis time points respectively. The observable variables may be represented as terms in the observation vector (\( O \)):

\[
O = [\%\text{Lin}^+, \text{TNCX}, \text{CFCX}, \text{LTCICX}, \text{SRCX}]
\]  \hspace{1cm} (23)

Which defines the system output (\( O \)). A cost function (\( J \)) to be minimized was defined as a weighted (\( w \)) sum of squared residuals between experimental (\( O_e \)) and simulated (\( O_s \)) system outputs:

\[
J = \sum w_i \frac{(O_{e_i} - O_{s_i})^2}{O_{e_i}}
\]  \hspace{1cm} (24)

A total of 20 system outputs (\( \%\text{Lin}^+ \), TNC, CFC, LTCIC, and SRC expansion \( \times 4 \) culture conditions) are therefore used to fit the 16 internal parameters. Each term in the equation \( J \) is
normalized (division by $Oe^2$), and a weighting factor ($w_i$) is included (set to 1 for all terms initially) introducing a user-defined bias based on which terms are deemed more or less important.

Due to the highly non-linear and multi-modal nature of the cost function ($J$), a hybrid stochastic method was used to solve this non-linear programming problem. A genetic algorithm was first used as a global method to search the entire parameter space, followed by the Nelder-Mead Simplex algorithm to search for a local minimum in the vicinity of output from the genetic algorithm. These were implemented using the MATLAB `gatool` and `fminsearch` functions respectively. Multiple settings for the genetic algorithm were tested and the following were selected, producing a solution that converges within 150 generations; population size = 160 [10 × number of parameters ($N_p$)], initialized as a random log-uniform distribution; elite count = 8; crossover fraction = 0.5; mutation rate = 0.5. Following model validation studies, objective functions were created based on the new data, and a generalized pattern search algorithm (`psearchtool`) was used to adjust parameter values within 95% confidence intervals (± 2 standard deviations) of the estimates defined in Table III.

2.3.5 Model Analysis – Local Parameter Sensitivities

To understand the relationship between system responses and variations in individual parameter values, Local Parameter Sensitivity Analysis (LPSA) was performed. The sensitivity coefficient ($S$) is defined as:

$$S_{p}^{O} = \frac{\partial O / O}{\partial P / P} \approx \frac{\Delta O / O}{\Delta P / P} \quad \text{for small } \Delta P$$

Which is defined for each system output ($O_i$) and system parameter ($P_j$). Individual parameters were altered individually by 1% ($\Delta P = 0.01$) from their estimated values, and resulting changes in system outputs ($\Delta O$) determined. The resulting expression essentially denotes the % change in output $O_i$ resulting from a 1% change in parameter $P_j$. This analysis produces a $5 \times 16$ sensitivity coefficient ($S$) matrix.

$$S_{p_j}^{O_i} \quad \text{for } i = [1: 1: 5] \text{ and } j = [1: 1: 16]$$

(26)
To quantify relationships between parameters and system outputs, this matrix was converted to a heat-map image for visualization, and 2-dimensional, unsupervised hierarchical clustering used to organize rows ($P_j$) and columns ($O_i$). Clustering and visualization was performed using dChip v2006 software (www.dChip.org). Sensitivity values were normalized (mean 0 standard deviation ± 1 across rows) and rows/columns clustered via the centroid-linkage method with a distance metric of $1 - r$ (Pearson correlation coefficient) as described in (Eisen et al., 1998).

2.3.6 Model Analysis – Parameter Identifiability

Parameter estimation accuracies are central to measuring identifiability of mechanistic models. The Fisher Information Matrix (FIM) and Cramer-Rao theorem are commonly used to estimate the lower bound of parameter estimation errors. However, this approach assumes the model is linear with respect to parameters, while our model is highly non-linear. We therefore implemented a bootstrapping approach as described in (Kremling et al., 2004), which is more computationally intensive but requires no underlying assumptions. Based on experiments and results described in Figure 2, we generated a synthetic data set of 50 experiments from the distributions reported in Table III. We then ran our parameter estimation algorithm on each individual data set, resulting in 50 estimated parameter vectors ($P$). We filtered the results for parameter vectors ($P$) producing cost functions ($J$) < 5, and characterize the distribution of individual parameter estimates ($P_i$) by their coefficient of variation (CV). Non-identifiable parameters are defined as those for which it is not possible to determine with 95% confidence (estimated as ± 2 standard deviations) that their values are non-zero (Zak et al., 2003).

2.3.7 Model Analysis – Structural Discrimination

The large number of non-identifiable free parameters persuaded us to examine the sensitivity of model simulations to structural alterations of the regulatory network. The regulatory architecture is based upon theoretical and experimental evidence rather than an exhaustive search of all possible topologies due to computational limitations. We chose to systematically “re-wire” intra-cellular network connections by perturbing the sign of interactions (stimulatory ↔
inhibitory) while conserving the number of free parameters. The following perturbations we define as $A$, $B$, and $C$:

- **A**: SF2 effect on $f_i$ - inhibitory $\rightarrow$ stimulatory
- **B**: SF4 effect on $f_i$ - stimulatory $\rightarrow$ inhibitory
- **C**: SF1 effect on $sr2$ - stimulatory $\rightarrow$ inhibitory

These perturbations were implemented alone and combinatorially, and the resulting structurally altered models defined as $S1$ through $S7$:

- **$S1$**: A
- **$S2$**: B
- **$S3$**: C
- **$S4$**: AB
- **$S5$**: AC
- **$S6$**: BC
- **$S7$**: ABC

Perturbing the sign of interaction for the effect of SF1 and SF3 on progenitor proliferation rate ($u_i$) is equivalent to simply deleting one of the feedback connections; hence we considered two additional structural alterations that have the effect of reducing the number of free parameters:

- **$S8$**: delete SF1 functional effects ($k1 = -\infty$)
- **$S9$**: Delete SF3 functional effects ($sr3 = 0, k3 = 0$)

We ran the parameter estimation algorithm on each of the structurally altered models ($S1$-$S9$) in addition to our original model [control ($C$)] using an extended objective function ($J$) incorporating experimental data reported in Figures 2-3, 2-8, and 2-10. The Akaike Information Criterion ($AIC$) was then computed for each:

$$AIC = 2Np + Ne \left( \ln \left( \frac{2\pi \cdot J}{Ne} \right) + 1 \right)$$ (27)

Where $Np$ = number of parameters and $Ne$ = number of independent experimental measurements. Alternate models are then ranked, with the lowest $AIC$ corresponding to the model best able to describe the data with minimum free parameters (Landaw and DiStefano, 1984).
2.3.8 Software

All model simulations and computational analysis was performed using MATLAB R2008a software (The Mathworks, Natick, MA, USA), and differential equations were solved using the non-stiff numerical solver *ode23* with default error tolerances. To ensure results were not affected by the propagation of rounding errors, select simulations were performed using the non-stiff solver *ode45* and the stiff solver *ode15s* with different error tolerances, and the results were unaffected.

2.4 EXPERIMENTAL METHODS & MATERIALS

2.4.1 Cell Sample Collection and Processing

Umbilical cord blood (UCB) samples were obtained from consenting donors according to procedures accepted by the ethics boards of Mt Sinai hospital (Toronto, ON, Canada), Joseph Brandt hospital (Burlington, ON, Canada) and Credit Valley Hospital (Mississauga, ON, Canada). Mononuclear cells were obtained by first mixing the UCB sample with 10% pentastarch (Bristol-Myers Squibb Canada, Montreal, QB, Canada) at a 1:5 volumetric ratio. The sample was then centrifuged for 10 minutes at 50g, and the upper (leukocyte rich) plasma layer was removed and centrifuged for 10 minutes at 400g to obtain a cell pellet. Red blood cells were depleted by suspending the cells for 10 minutes in Red Blood Cell Lysis Buffer (0.15 M NH₄Cl, 0.01M KHCO₃, 0.1 mM EDTA). Lineage depleted (Lin⁻) cells were isolated from the mononuclear cell fraction using the StemSep system (Stem Cell Technologies, Vancouver, BC, Canada). This process depletes cells expressing cell surface antigens CD2, CD3, CD14, CD16, CD19, CD24, CD56, CD66b, and CD235a.

2.4.2 Cell Culture

UCB Lin⁻ cells were seeded at 10⁵ cells/mL (unless otherwise noted) in serum-free Stem Span media (Stem Cell Technologies, Vancouver, BC, Canada) supplemented with 100 ng/mL Stem Cell Factor (SCF; Amgen, Thousand Oaks, CA, USA), 100 ng/mL Flt3 Ligand (FL; Amgen), and 50 ng/mL Thrombopoietin (TPO; R&D Systems, Minneapolis, MN, USA), 1 μg/mL low-
density lipoproteins (Calbiochem, La Jolla, CA, USA), and penicillin-streptomycin at 100 U/mL & 100 µg/mL respectively (Invitrogen, Carlsbad, CA, USA). The cell suspension either injected into a cell culture bag of appropriate volume (2, 7, or 12 mL) through the self-sealing rubber septum using a sterile syringe attached to a threaded cannula, or placed into wells of a 24-well tissue culture plate (Corning Inc., Corning, NY, USA). Cultures were carried out using a culture bag-based bioprocess (described below) for experiments requiring Lin⁺ cell depletion, otherwise 24-well tissue culture plates were used. Cultures were maintained on an orbital shaker at 37°C in a humidified atmosphere of 5% CO2 in air.

2.4.3 Subpopulation Selection and Media Dilution

The bioprocess used in these studies for depletion of in vitro-generated Lin⁺ cells was described in (Madlambayan et al., 2006). It consists of 2 gas-permeable cell culture bags connected via a magnetic selection element, used to remove the Lin⁻ cells (or any other antibody-labelled cell sub-population). The system is completely closed, sterile, autoclavable, and disposable (single use), making it attractive for clinical applications. Cell selection in the bioprocess was performed in a similar manner to the StemSep system using the reagents provided with the kit, as described in (Madlambayan et al., 2006). The cell culture bag was flushed with Lin⁺ antibody cocktail and magnetic colloid (dextran-coated iron particles) as per manufacturers instructions. This effectively attached the dextran-coated iron particles to the Lin⁻ cells. The cell culture was subsequently allowed to flow the selection element that was placed in a magnetic field, allowing the iron-labelled Lin⁺ to be retained in the element while non-labelled Lin⁻ cells flow through to the secondary culture bag. For flow rate control, a peristaltic pump was attached upstream of the cell culture bag using a septum / cannula connection, and used to drive the cell solution through the selection element at a flow rate of 1.3 mL/min.

Media dilution was performed on the enriched Lin⁻ cells by removing the secondary culture bag from the selection element and placing it into a 50 mL conical centrifuge tube. Paper batting was used to stabilize the culture bag during centrifugation. The bag was centrifuged for 7 min at 200xg, after which a cell pellet was visible at the bottom. Conditioned media was then removed
through the self-sealing septum using a sterile syringe and fresh media added through the same septum. The cell culture bag was then placed back into the incubator.

2.4.4 Phenotypic Analysis

Staining for Lin$^+$ marker expression on culture-generated cells was accomplished by suspending $5 \times 10^4$ cells in 100 μL ice cold Hank’s balanced saline solution containing 2% (v/v) human UCB serum (HBSS-HS). The cells were then incubated with Lin$^+$ antibody cocktail followed by magnetic colloid (dextran-coated iron beads) as described, washed twice in HBSS-HS, and finally stained with saturating amounts of FITC-labelled anti-dextran antibody (Stem Cell Technologies) for 30 minutes on ice. For isotype controls, the Lin$^+$ antibody incubation step was not performed. All samples were washed in HBSS-HS and stored on ice prior to analysis either on a FACSCanto (BD Biosciences, San Jose, CA, USA) or Coulter EPICS XL (Beckman Coulter, Fullerton, CA, USA) flow cytometer.

2.4.5 Progenitor Cell Assays

Cells were assayed for CFC frequency by plating 500 cells into 1.5 mL methylcellulose-based medium (MethoCult H4434, Stem Cell Technologies) containing 1% methylcellulose in Iscove’s Modified Dulbecco’s Medium, 30% fetal bovine serum (FBS), 1% bovine serum albumin (BSA) $10^{-4}$ M 2-mercaptoethanol, 2 mM L-glutamine, 50 ng/mL SCF, 10 ng/mL granulocyte-macrophage colony stimulating factor (GM-CSF), 10 ng/mL interleukin-3 (IL-3) and 3 U/mL erythropoietin. After 14 days of incubation at 37°C in a humidified atmosphere of 5% CO2 in air, duplicate cultures were visually scored for CFC content (colony number and lineage composition). Cells were assayed for LTCIC frequency by plating $10^3$ freshly isolated Lin$^-$ cells, or $5 \times 10^3$ to $2 \times 10^4$ culture-generated cells in triplicate onto irradiated (6000 rad) murine stromal cells (M2-10B4) on collagen-coated 6-well plates in MyeloCult H5100 medium containing $10^{-6}$ M hydrocortisone (Stem Cell Technologies). After 5 weeks of culture at 37°C with weekly half-media exchanges, the contents of each well were harvested using 0.25% Trypsin with 0.38 g/L EDTA in HBSS (Invitrogen, Carlsbad, CA, USA), and plated into methylcellulose-based media. LTC-IC content was determined by enumerating CFCs present after 14 days of incubation.
2.4.6 Conditioned Media Proteome Analysis – Luminex Liquid Chips & ELISA

Conditioned media samples were assayed in triplicate using the Biosource Human Cytokine 30-Plex detection kit (Invitrogen, Carlsbad, CA, USA). These kits utilize Luminex (Luminex Co., Austin, TX, USA) microshperes as a fluid platform for multiplex sandwich ELISA. The “microspheres” consist of 5 μm polysytene beads bar-coded via unique ratios of APC: APC-Cy7 dye. Each colour-coded microsphere contains primary capture antibody against an individual ligand, which in combination with secondary PE-conjugated detection antibody, can be used to quantify the concentration proteins in a test sample via flow cytometry. Briefly, 50 μL antibody-coated microspheres per test were suspended and aliquoted into wells of a 96-well filter plate (Millipore, Billerica, MA, USA) and washed in Wash Buffer. 50 μL of Incubation Buffer plus 100 μL of culture media sample (or appropriately diluted standard) was then added to each well, and the plate was covered and incubated on an orbital shaker at room temperature for 2 hours. Following sample incubation, the wells were washed and 100 μL of Detection Antibody mixture was added to each well and incubated on an orbital shaker at room temperature for 1 hour. Following incubation, wells were washed and incubated with 100 μL streptavidin-PE for 30 minutes. Following this final incubation, wells were washed and resuspended in 5 mL polystyrene falcon tubes (BD Biosciences, San Jose, CA, USA) for analysis using a BD FACSCanto flow cytometer (BD Biosciences, San Jose, CA, USA). APC vs. APC-Cy7 gates for each of the 30 cytokine-specific bead regions were defined based on manufacturer’s datasheets, and cytokine signals quantified by comparing PE fluorescence levels to standard dilution curves prepared using the Human Cytokine 30-Plex Standard mixture provided. Cytokine-specific standard curve equations were developed by fitting data to a 4-parameter logistic curve:

$$\log(PE) = \frac{A}{1 + \exp[-B(\log(x_i) + C)]} + D$$

(28)

Where $PE$ is the raw PE-fluorescence intensity, $x_i$ is the cytokine concentration, and $A$, $B$, $C$, and $D$ are the unknown parameters, fit using the Nelder-Mead Simplex non-linear optimization algorithm.
TGF-β1 is not one of the molecules included in the kit, hence conditioned media samples were separately analysed using a human TGF-β1 Quantikine ELISA Kit (R & D Systems, Minneapolis, MN, USA) as per manufacturers instructions.

2.5 RESULTS

2.5.1 A feedback-based cell-cell interaction network model of hematopoiesis

The hematopoietic hierarchy can be divided into discrete cellular compartments, wherein compartment transitions are typically coincident with compartment size amplifying cell divisions. Taking advantage of differentiation-state-associated in vitro and in vivo assays we have defined functional readouts as overlapping series of consecutive compartments. The functional readouts we consider are the immuno-deficient (Non-Obese Diabetic (NOD)/Scid) mouse repopulating cell (SRC) assay for quantifying stem cells, the long-term culture-initiating cell (LTC-IC) assay for quantifying primitive progenitors, and the colony forming cell (CFC) assay for quantifying committed progenitors (Coulombel, 2004). Hematopoietic cell populations are also broadly classified phenotypically based on their expression (Lin⁺), or lack of expression (Lin⁻), of blood lineage-associated cell surface antigens. Frequency estimates for cell compartment – assay relationships, as described in the methods, are shown in Table 2-1, and a schematic diagram of the model is depicted in Figure 2-1a. For clarity, all model parameters, variables, and simulated cell population outputs will henceforth be indicated with italics.
Model Variables

\[ i \] = compartment number [ ]

\[ X_i \] = cells in compartment \( i \)

\[ u_i \] = proliferation rate (day \(^{-1}\))

\[ f_i \] = self-renewal probability

\( SF1 \) = proliferation inhibitor

\( SF2 \) = self-renewal inhibitor

\( SF3 \) = proliferation stimulator

\( SF4 \) = self-renewal stimulator

Developmental potential

Differentiation
Figure 2-1. Schematic depiction of blood stem cell development model incorporating functional assays and positive and negative feedback.

(a) Stem cells ($X_i$) at the apex of the hematopoietic hierarchy may self-renew with probability $f_i$ to regenerate the stem cell compartment, or differentiate ($1-f_i$), giving rise to a series of increasing differentiated, and developmentally restricted progenitor populations ($X_i$). Transit between cell compartments is associated with mitosis ($u_i$). *In vivo* (SRC) and *in vitro* (LTCIC, CFC) functional assays, and cell surface phenotype ($Lin^-$ vs. $Lin^+$) can be used to quantify different cellular compartments within the hierarchy. Differentiated cell populations secrete factors that inhibit or enhance progenitor proliferation ($SF1$ and $SF3$ respectively), and undifferentiated ($Lin^-$) cells secrete factors that inhibit stem cell self-renewal ($SF2$). The inhibitory effects of $SF2$ on self-renewal are balanced by secretion of the self-renewal stimulator $SF4$ by differentiated cells. Differentiated populations are functionally lumped into those that secreted inhibitory factors (red), those that secrete stimulatory factors (green), and those that secrete factors with no effect on stem and progenitor cell growth (yellow). Phase portraits below the diagram display normalized proliferation rates ($u_i/u_{MAX}$) and self-renewal probabilities ($f_i/f_{MAX}$) as functions of model parameters varied between the constraints given in Table II (low to high values represented by blue to red as indicated). (b) Proliferation and self-renewal vs. differentiation status (compartment number) for parameters $n_{MAX}$, $D_{GR}$, and $D_{SR}$ left to right respectively. (c) Proliferation vs. time for parameters $\tau_D$ (left) and $k_I$ (right). Proliferation or self-renewal vs. secreted regulatory factor ($SF1$-$4$) concentrations for Hill coefficients ($k_1$-$4$). Refer to Table II for parameter definitions.

It is well established experimentally that self-renewal probabilities decrease and cell cycling rates increase with differentiation (Young et al., 1996). We implemented Gaussian-type functions to quantify proliferation rates ($u_i$) and the self-renewal probabilities ($f_i$) as functions of compartment number ($i$) (see Methods for details and Figure 2-1 b and c for phase portraits).

We simulate a branching model of hematopoiesis by lumping differentiated ($Lin^+$) cells into 3 functional classes based on their functional feedback interactions with stem and progenitor cells; populations that secrete inhibitors, populations that secrete stimulators, and populations that
secrete molecules with no net effect. The framework for the inter-cellular signalling network topology is based largely around in vivo experimental data on murine hematopoietic homeostasis (Wichmann and Loeffler, 1985) with the implicit assumption that the network will be at least partially reconstituted in vitro. We have designated compartment-specific self-renewal and proliferation rates as regulated by a balance between endogenously secreted inhibitors (negative feedback) and stimulators (positive feedback) (Figure 2-1d). We consider only soluble factor-mediated signalling between blood cell populations, and thus limit our analysis to liquid suspension cultures; the additional complexities associated with stoma-supported co-cultures or in vivo hematopoiesis are not implicitly considered. The resultant model consists of 24 state variables [20 cell compartments ($X_i$) and 4 secreted regulatory molecules ($SF1-4$), defined in Figure 2-1a], and 16 internal parameters, their definitions and theoretically constrained ranges given in Table 2-2.

**Table 2-2. Model parameters.**

<table>
<thead>
<tr>
<th>P</th>
<th>DESCRIPTION</th>
<th>UNITS</th>
<th>LO</th>
<th>HI</th>
<th>EST</th>
<th>CV</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_{\text{MAX}}$</td>
<td>Maximum proliferation rate of lin$^-$ cells</td>
<td>d$^{-1}$</td>
<td>10$^0$</td>
<td>10$^1$</td>
<td>6.26 $\times$ 10$^0$</td>
<td>0.311</td>
<td>+</td>
</tr>
<tr>
<td>$u_+$</td>
<td>Maximum proliferation rate of lin$^+$ cells</td>
<td>d$^{-1}$</td>
<td>10$^{-1}$</td>
<td>10$^0$</td>
<td>2.04 $\times$ 10$^{-1}$</td>
<td>0.209</td>
<td>+</td>
</tr>
<tr>
<td>$n_{\text{MAX}}$</td>
<td>Compartment with maximal proliferation</td>
<td>-</td>
<td>10$^0$</td>
<td>10$^1$</td>
<td>5.32 $\times$ 10$^0$</td>
<td>0.273</td>
<td>+</td>
</tr>
<tr>
<td>$D_G$</td>
<td>Proliferative decay term</td>
<td>-</td>
<td>10$^0$</td>
<td>10$^1$</td>
<td>3.38 $\times$ 10$^0$</td>
<td>0.368</td>
<td>+</td>
</tr>
<tr>
<td>$f_{\text{MAX}}$</td>
<td>Self-renewal probability of LT-HSC</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^0$</td>
<td>6.34 $\times$ 10$^{-1}$</td>
<td>0.136</td>
<td>+</td>
</tr>
<tr>
<td>$D_S$</td>
<td>Self-renewal decay term</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>1.96 $\times$ 10$^0$</td>
<td>0.355</td>
<td>+</td>
</tr>
<tr>
<td>$sr1$</td>
<td>Secretion rate of $SF1$</td>
<td>pg/cell.d</td>
<td>10$^{-7}$</td>
<td>10$^{-5}$</td>
<td>2.37 $\times$ 10$^{-5}$</td>
<td>0.300</td>
<td>+</td>
</tr>
<tr>
<td>$sr2$</td>
<td>Secretion rate of $SF2$</td>
<td>pg/cell.d</td>
<td>10$^{-7}$</td>
<td>10$^{-5}$</td>
<td>2.93 $\times$ 10$^{-5}$</td>
<td>0.539</td>
<td>-</td>
</tr>
<tr>
<td>$sr3$</td>
<td>Secretion rate of $SF3$</td>
<td>pg/cell.d</td>
<td>10$^{-7}$</td>
<td>10$^{-5}$</td>
<td>5.96 $\times$ 10$^{-6}$</td>
<td>0.237</td>
<td>+</td>
</tr>
<tr>
<td>$sr4$</td>
<td>Secretion rate of $SF4$</td>
<td>pg/cell.d</td>
<td>10$^{-7}$</td>
<td>10$^{-5}$</td>
<td>5.30 $\times$ 10$^{-6}$</td>
<td>0.632</td>
<td>-</td>
</tr>
<tr>
<td>$k1$</td>
<td>Hill coefficient for $SF1$ (Equation 14)</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>6.14 $\times$ 10$^{-1}$</td>
<td>1.22</td>
<td>-</td>
</tr>
<tr>
<td>$k2$</td>
<td>Hill coefficient for $SF2$ (Equation 15)</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>5.55 $\times$ 10$^{-1}$</td>
<td>1.25</td>
<td>-</td>
</tr>
<tr>
<td>$k3$</td>
<td>Hill coefficient for $SF3$ (Equation 14)</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>6.25 $\times$ 10$^{-1}$</td>
<td>0.467</td>
<td>+</td>
</tr>
<tr>
<td>$k4$</td>
<td>Hill coefficient for $SF4$ (Equation 15)</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>5.33 $\times$ 10$^{-1}$</td>
<td>1.15</td>
<td>-</td>
</tr>
<tr>
<td>$L_s$</td>
<td>$[SF1]$ inducing ½ maximal $SF2$ secretion</td>
<td>pg/ml</td>
<td>10$^{-4}$</td>
<td>10$^1$</td>
<td>9.15 $\times$ 10$^{-1}$</td>
<td>0.635</td>
<td>-</td>
</tr>
<tr>
<td>$k5$</td>
<td>Hill coefficient for $SF1$ (Equation 9)</td>
<td>-</td>
<td>10$^{-1}$</td>
<td>10$^1$</td>
<td>1.08 $\times$ 10$^0$</td>
<td>1.04</td>
<td>-</td>
</tr>
</tbody>
</table>

$P$ = model parameter, $LO$ = lower constraint on parameter, $HI$ = upper constraint on parameter, $EST$ = estimated parameter value, $CV$ = coefficient of variation for parameter estimate, $ID$ = identifiability (“+” = identifiable, “-” = non-identifiable)
2.5.2 Global parameter space analysis reveals a critical role for non-cell autonomous parameters

To systematically explore the parameter space in an unbiased manner, functional outputs from theoretical 8-day cultures (Madlambayan et al., 2005) were simulated for 1000 random parameter values log-uniformly distributed within the constraints (“LO” and “HI” values) defined in Table 2-2. Figure 2a shows the statistical relationship between the in vivo, in vitro, and phenotypic assays, and Figure 2b shows the statistical relationship between individual parameter values and culture outputs. Pearson correlation coefficients (r) were calculated for each relationship and are shown on each plot.

Notably, there is a poor correlation between total cell (TNC) and stem cell (SRC) expansion (r = -0.029) and between progenitor (CFC) and stem cell (SRC) expansion (r = 0.433). The model thus captures the recognition that these mature cell assays are poor surrogates for stem cell output (Zandstra et al., 1997). Our analysis suggests that under many conditions the LTCIC assay is a better surrogate for stem cell content (r = 0.703). Production of differentiated Lin+ cells correlates negatively with stem and progenitor cell output (r = -0.322 for CFC and LTCIC, r = -0.187 for SRC) likely due to the negative feedback-biased architecture of the system.
Figure 2-2. Statistical analysis of state space reveals critical roles for non-cell autonomous parameters.

Eight-day culture outputs simulated using 1000 random parameter combinations (within constraints given in Table 2-2) are plotted against each other (a) and against each of the 16 model parameters (b) as correlation matrices. Plots are in log scale. Correlation coefficients (r) are displayed for each individual plot, and red boxes are used to highlight plots with $r^2 > 0.05$.

As one might predict, the progenitor proliferation rate ($u_{MAX}$) is the best positive predictor of total cell ($TNC$), progenitor ($CFC$), and primitive progenitor ($LTCIC$) expansion ($r = 0.695, 0.452$, and $0.274$ respectively). However, surprisingly, this parameter ($u_{MAX}$) is inversely correlated with stem cell ($SRC$) expansion ($r = -0.161$). Non-intuitively, the strongest positive correlate to stem cell ($SRC$) expansion is secretion rate of the proliferation inhibitor $SF1$ ($sr1$) ($r = 0.227$), a finding that suggests a negative relationship between total cell and stem cell growth. The secretion rate of the self-renewal inhibitor $SF2$ ($sr2$) is the best negative correlate with $CFC$, $LTCIC$, and $SRC$ expansion ($r = -0.14, -0.372$, and $-0.479$ respectively). This analysis suggests that non-cell autonomous parameters (soluble factor-mediated cell-cell interactions) may be dominant factors controlling stem cell growth. It is notable that while the parameter correlations are low, particularly with regard to $SRC$ output (the highest $r^2 = 0.0515$), most are statistically significant ($|r| > 0.062$ corresponds to $p < 0.05$ for $n = 1000$ simulations). The wide distribution of outputs observed is indicative of highly non-linear parameter interactions, necessitating the use of stochastic global optimization methods to estimate parameter values. Thus, we next devised an appropriate reverse engineering strategy using previously published data.

2.5.3 Parameter training identifies self-renewal inhibitor ($SF2$) exposure time as a key regulator of stem and primitive progenitor cell output

Our laboratory has previously developed a bioprocess for the clinical-grade expansion of umbilical cord blood-derived HSCs (Madlambayan et al., 2006). This system allows us to dynamically perturb cell-cell signalling networks, and thereby interrogate our theoretical model. Specifically, Lin$^{-}$ cell negative selection (S) reduces the secretion rates $sr1$ to $sr4$, and media exchange (E) reduces the concentration of factors $SF1$ to $SF4$ (Figure 2-3a). Eight-day culture
outputs for combinations of Lin\(^{-}\) cell selection (S) vs. no selection (NS), and media exchange (E) vs. no exchange (NE) at culture day-4, as previously reported (Madlambayan et al., 2005), are shown in Table 2-3. Performing Lin\(^{-}\) cell selection and media exchange in combination (S/E) synergistically enhanced primitive progenitor (LTCIC) and stem cell (SRC) output, while the effect of the described culture manipulations on other system outputs (%Lin\(^{+}\), TNC, or CFC expansion) was not statistically significant (\(p > 0.1\)).

**Table 2-3. Cell population outputs used for parameter training.**

<table>
<thead>
<tr>
<th>READOUT</th>
<th>NS/NE</th>
<th>NS/E</th>
<th>S/NE</th>
<th>S/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Lin(^{+})</td>
<td>(~ 25)</td>
<td>(~ 25)</td>
<td>(~ 25)</td>
<td>25 ± 10</td>
</tr>
<tr>
<td>TNC X</td>
<td>23 ± 9</td>
<td>26 ± 10</td>
<td>17 ± 9</td>
<td>33 ± 14</td>
</tr>
<tr>
<td>CFC X</td>
<td>17 ± 10</td>
<td>12 ± 9</td>
<td>10 ± 5</td>
<td>21 ± 13</td>
</tr>
<tr>
<td>LTCIC X</td>
<td>7 ± 3</td>
<td>7 ± 1</td>
<td>4 ± 3</td>
<td>15 ± 3</td>
</tr>
<tr>
<td>SRC X</td>
<td>(~ 1)</td>
<td>(~ 1)</td>
<td>(~ 1)</td>
<td>5 ± 2</td>
</tr>
</tbody>
</table>

(Madlambayan et al, 2005)

To estimate parameter values, we used the data displayed in Table 2-3 to define an objective function based on the sum of squared residuals between experimental and simulated culture outputs (see Equation 25). As the objective landscape was expected to be multi-modal, a hybrid stochastic method was developed to estimate the unknown parameters by minimizing the objective function. A genetic algorithm was first used to globally search the parameter space, followed by the Nedler-Mead Simplex algorithm for local optimization. The estimated parameter vector is displayed in Table 2-3 ("EST" column), and the results are summarized in Figure 2-3. Coefficients of Variation (CV) for the parameter estimates (Table 2-2 "CV" column) were calculated using a computational bootstrapping approach, and indicate that a number of parameters are non-identifiable (i.e. system dynamics can be simulated using multiple parameter combinations, indicated in Table 2-2 "ID" column; "+" = identifiable, "-" = non-identifiable). Such "sloppy" parameter estimates are in fact a consistent feature of systems biology models, and may reflect underlying robustness of biological networks (Gutenkunst et al., 2007).
Figure 2-3. Model training methodology and results reveals endogenous inhibitors as key regulators of cell population outputs.

(a) Cytokine-supplemented cultures initiated with UCB-derived undifferentiated (Lin\(^-\)) cells (blue) rapidly evolve, producing heterogeneous populations of differentiated cells (red, yellow, and blue) and secreting multiple regulatory factors. At culture day-4, cultures are perturbed via Lin\(^-\) cell negative selection (S), or not (NS), and media exchange (E), or not (NE). These culture manipulations are carried out in combination, producing four conditions in total: NS/NE, NS/E, S/NE, and S/E. Following an additional 4 days in culture, the resulting cell populations are analysed using in vitro, in vivo, and phenotypic assays. (b) Experimental outputs (TNC, CFC, LTCIC, and SRC population expansions) used to train the model are plotted against the resulting model simulations. The S/E manipulation significantly enhances stem (SRC) and primitive progenitor (LTCIC) expansion compared to the remaining 3 conditions, predicted to result from reduced exposure (concentration \(\times\) time) to the proliferation and self-renewal inhibitors \(SF1\) and \(SF2\) (c).

Importantly, the model captures the experimental results, and explains the synergistic and targeted effect of the culture manipulations (S/E) on stem and primitive progenitor populations via reduced exposure (concentration \(\times\) time) to the proliferation and self-renewal inhibitors \(SF1\) and \(SF2\) (Figure 2-3c). To further investigate the relationship between individual parameters and system dynamics, parameter sensitivity analysis was performed.

2.5.4 Phase portraits and parameter sensitivity analysis demonstrate antagonistic relationship between mature and primitive cell compartments

Parameter sensitivities represent emergent systems-level properties which cannot be deduced by analysing individual components or reactions (Savageau, 1971). Parameter Sensitivity Coefficients (S) were calculated for each system output \((O)\) ~ parameter \((P)\) combination (Equations 26 and 27), producing a 5\(\times\)16 sensitivity coefficient matrix, depicted as a heat map in Figure 2-4. Parameters and system outputs were organized via unsupervised hierarchical clustering, and distances \((1 – r)\) represented by associated dendograms.
Parameter sensitivity analysis reveals antagonism between mature and primitive cell compartments. Parameter sensitivity coefficients are represented as a heat-map, subjected to unsupervised 2-dimensional hierarchical clustering of system outputs (rows) and parameters (columns). Values were normalized to 0 ± 1 standard deviation across rows.

The system outputs organize into two primary clusters, one containing differentiated (Lin$^+$) cells, total cells (TNC), and mature progenitors (CFC), and the other containing primitive progenitors (LTCIC) and stem cells (SRC). This clustering pattern emphasizes the modular nature of cell compartments, and the relative independence of primitive and mature cell outputs. The parameters correspondingly fall into two general categories; those that positively influence stem cell (SRC) expansion while repressing total cell (TNC) growth, and vice-versa. This demonstrates an antagonistic relationship between mature and primitive cell compartments, a result of the (negative) feedback interactions.
Figure 2-5. Phase portraits reveal differences in growth kinetics and dynamic sensitivities between mature and primitive cell compartments.

Phase portraits depict the effects of varying the 16 model parameters between the constraints given in Table 2-2, on total cell (TNC), progenitors (CFC), primitive progenitor (LTC-IC), and stem cell (SRC) dynamics in culture.
To highlight the effects of individual parameters on system dynamics, phase portraits are shown in Figure 2-5, wherein system outputs \([TNC, CFC, LTCIC, \text{ and } SRC]\) are plotted against time, for individual parameter values varied over the ranges given in Table 2-2. In all cases there is a clear divergence in the population dynamics for total cells \((TNC)\) vs. stem cells \((SRC)\); while the total cell \((TNC)\) numbers display near-exponential growth throughout the 8-day cultures, primitive progenitors \((LTCIC)\) and stem cell \((SRC)\) numbers plateau mid-culture and then begin to decline. This bi-phasic growth curve has been observed in many blood progenitor cultures, regardless of the choice of cytokine cocktail (Mobest et al., 1999; Zandstra et al., 1997).

As expected, the most sensitive parameter controlling stem cell growth is the probability of self-renewal \((f_{\text{MAX}})\). This is consistent with most current strategies for inducing \textit{in vitro} stem cell expansion via the identification and use of specific modulators of self-renewal (Sauvageau et al., 2004). However, our analysis suggests for the first time that the system dynamics are predominantly regulated by the non-stem cell autonomous negative feedback signals \((SF1\text{ and } SF2)\) rather than stem cell-autonomous factors. Following this, sustained stem cell amplification should depend upon removal or blocking of these signals. In fact, factors inducing total cell expansion [i.e. increasing progenitor proliferation rate \((u_{\text{MAX}})\) via stimulatory growth factors such as IL3] may actually repress stem cell growth via enhanced accumulation of cells secreting inhibitory factors. We next sought to use this insight in a new series of experiments to predictively manipulate culture variables to control stem and progenitor cell output.

2.5.5 Dynamic culture perturbations reveal inter-cellular feedback control of stem and progenitor cell output

Our model simulations predict that cell-cell feedback would produce non-linear cell population dynamics in response to culture perturbations. In a first series of experiments, we measured the \textit{in vitro} production of undifferentiated \((\text{Lin}^-)\) cells, differentiated \((\text{Lin}^+)\) cells, and progenitors \((\text{CFC})\) in response to consecutive culture perturbations. Umbilical cord blood-derived \text{Lin}^- cells were cultured for 8 days un-manipulated (NS/NE), or subject to the \text{Lin}^+ cell selection and media exchange procedure at culture day-4 (S/E-d4), or at day-4 and day-6 consecutively (S/E-d4,d6) (Figure 2-6a,b). The relative frequency of differentiated \((\%\text{Lin}^+)\) cells at culture day-8 was unaffected by the number of times the \text{Lin}^+ cells were depleted. While the production of
undifferentiated (Lin−) cells across conditions was unaffected, total output of differentiated (Lin+) cells was increased in response to the S/E procedures ($p = 0.079$, $t$-test). Additionally, progenitor (CFC) output was enhanced in response consecutive perturbations ($p = 0.075$, $t$-test). These non-intuitive results can be attributed to the dynamic regulation of progenitor cell proliferation and differentiation by the inter-cellular feedback control loops.

The wide sample-to-sample variability in the expansion potential of blood stem cells necessitates a large number of replicate experiments to achieve statistical significance (Koller et al., 1996). We examined what sources of exogenous (culture variables) or endogenous (biological) variability could account for the variability typically observed in culture. To test this, 8-day un-manipulated (NS/NE) culture simulations were run with individual parameter values and select combinations normally distributed [coefficient of variation (CV) = 1] around the values given in Table 2-3. Experimental ($n = 9$) and simulated ($n = 100$) output distributions were compared using a ranking metric based on the sum of squared residuals between total cell (TNC vs. \(TNC\)), progenitor (CFC vs. \(CFC\)), and primitive progenitor (LTCIC vs. \(LTCIC\)) population expansion means and CVs (Figure 2-7). Variabilities in the secretion rates \(sr1-sr4\) most closely simulated the distribution of experimental cell population outputs. As shown in Figure 2-6c, the distribution of culture outputs [CV for TNC = 0.64, CFC = 0.67, and SRC = 0.73] is closely recapitulated \textit{in silico} [CV for \(TNC = 0.60\), \(CFC = 0.83\), and \(SRC = 0.76\)], providing evidence that variability in culture outputs may be attributable to variability in inter-cellular signalling dynamics. That is, microenvironmental noise may play a role in guiding cell fate.
Figure 2-6. Cell population dynamics in response to consecutive culture perturbations reveals inter-cellular feedback control.

(a) Flow cytometric analysis of Lin⁺ composition in 8-day culture-generated cells produced under the three culture conditions [un-manipulated (NS/NE), Lin⁻ selection and media exchange
at day-4 (S/E-d4), or day-4 and day-6 (S/E-d4,d6)] compared with in silico simulations of Lin\(^+\) cell composition dynamics (n = 3).  \(\textbf{(b)}\) Fold expansion of Lin\(^-\) and Lin\(^+\) cell populations (normalized to input Lin\(^-\)) and CFCs under the three culture conditions tested (n = 3).  \(\textbf{(c)}\) Variability in total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) expansions in 8-day un-manipulated (NS/NE) cultures observed experimentally (Exp; n = 9), and simulated (Sim; n = 100) by randomizing secretion rates of the endogenous regulators \((sr1-4)\) with normalized distributions around the values given in Table 2-2.  Error bars represent standard deviation.

\[\text{Figure 2-7. Stochastic variability in secretion rates } sr1-4 \text{ most closely reproduces experimentally observed distribution in culture outputs.}\]

8-day un-manipulated cultures were simulated with individual parameters and select parameter combinations normally distributed around estimated values given in Table 2-2. A metric based on the sum of squared residuals between total cell (TNC vs. TNC), progenitor (CFC vs. CFC), and primitive progenitor (LTCIC vs. LTCIC) population expansion means and CVs was used to rank parameter / parameter combinations.
2.5.6 Inter-cellular feedback signalling regulates long-term culture dynamics.

We next asked whether our model simulations are capable of accurately predicting cell population dynamics for extended (16-day) liquid cultures. While in un-manipulated cultures, stem and progenitor cell numbers rapidly plateau and begin to decline within a week, our model simulations predicted that by performing consecutive culture manipulations, stem and progenitor cell growth would be enhanced. Liquid cultures were maintained as described with Lin– cell selection and media exchange (S/E) performed every four days. As shown in Figure 2-8a, total cells (TNC) continue to proliferate throughout the 16-day culture, while progenitor (CFC) expansion plateaus after day-12, and primitive progenitor (LTCIC) numbers begin to decline after a maximal expansion of approximately 12-fold at day-8. Cultures subjected to media exchange alone every 4-days (NS/E) were predicted to decline faster (Figure 2-8a, dashed lines). The model simulations fit the experimental data for media exchange efficiencies (defined as the % of media replaced) of approximately 90% or greater (Figure 2-9), and suggest that the culture decline was due to increased secreted factor (SF1 and SF2) feedback from the accumulating differentiating cells (Figure 2-8b). To interrogate the predicted dynamic correlation between inhibitory protein secretion (SF1 and SF2) and culture decline, we profiled conditioned media obtained at culture days 4, 8, 12, and 16 for a set of 31 ligands using a liquid chip cytokine array. While the sample-to-sample variability in protein concentrations detected was large, the chemokines MIP-1β (CCL4), IL-8 (CXCL8) and RANTES (CCL5), as well as TGF-β1 display dynamics qualitatively similar to that of the theoretical proliferation and self-renewal inhibitors SF1 and SF2 (Figure 2-8c). The remaining 27 ligands were not consistently detected above background. Many additional factors are likely to be involved, however technical limitations of proteomic technology currently preclude systematic measurement for the small sample volumes obtained from these primary human cell cultures. While the specific functional effects of combinations of these factors on stem cell fate have yet to be elucidated, this data shows that a number of ligands are endogenously secreted with dynamics qualitatively similar to our model predictions.
Figure 2-8. Inter-cellular signalling regulates long-term culture dynamics.

(a) Cultures were initiated \textit{in vitro} (symbols) and \textit{in silico} (lines) with UCB Lin\(^-\) cells at 10\(^5\) cells/mL and cultured in serum-free, cytokine-supplemented liquid media and subjected to consecutive Lin\(^-\) selection and media exchange (S/E) manipulations (solid lines) or media exchange alone (E) (dashed lines), every 4 days. Total cells (TNC) progenitors (CFC) and
primitive progenitors (LTCIC) were enumerated throughout the cultures (n = 3). (b) Simulated concentration of endogenous proliferation and self-renewal inhibitors $SF1$ and $SF2$ in culture media. (c) Concentrations of endogenously produced MIP-1β, IL-8, RANTES, and TGF-β1, measured via multiplexed ELISA, in conditioned media throughout 16-day cultures. Individual data points = red, average = black. (d) Cumulative fold LTCIC expansion *in vitro* (symbols) and *in silico* (lines) from day-8 to day-12 for cultures with S/E procedure performed every four days (S/E-d4,8) or with an additional S/E procedure performed at day-10 (S/E-d4,8,10) (n = 3). Error bars represent standard deviation.

**Figure 2-9. Phase portraits depict the effect of media exchange efficiency on simulated 16-day culture outputs.**

16-day cultures were simulated as in Figure 2-8a, but with media exchange efficiencies (%media replacement) varied from 0% (no exchange) to 99.9%. Experimental data (black symbols ± standard deviation) for total cell (*TNC*, left) and primitive progenitor (*LTCIC*, right) is reproduced at efficiencies of approximately 90% or greater, in line with experimental efficiencies typically greater than 95%.

Model simulations predicted that increased frequency of culture manipulations at later time points (day-8+) would improve primitive progenitor (*LTCIC*) expansion, while having no effect
on the output of total cells (TNC) or committed progenitors (CFC). To test this prediction, we conducted a series of 12-day liquid cultures as described (S/E-d4,8), or with an additional S/E procedure at day 10 (S/E-d4,8,10). As predicted, the additional culture manipulation had no effect on total (TNC) or progenitor cell (CFC) output (not shown), however primitive progenitor (LTCIC) expansion was enhanced from $11 \pm 4$ to $19 \pm 3$ fold ($p = 0.10$, t-test) (Figure 2-8d). Primitive progenitor cell output can thus be specifically manipulated by dynamically perturbing inter-cellular feedback. We next asked whether our inter-cellular feedback model could provide mechanistic insight into the effects of culture variables on cell population outputs.

2.5.7 Culture strategies to manipulate intercellular feedback signalling: Plating density and input cell population enrichment

Thus far, all cell cultures described have been initiated at Lin- cell densities of $10^5$ cell/mL. However, it has been reported in a number of studies that low cell densities and progenitor enrichment are associated with greater stem cell expansions (Kohler et al., 1999; Sandstrom et al., 1995; Xu et al., 2000a), although the underlying mechanisms are largely undefined. To study the effect of plating density on culture outputs, both in vitro and in silico cultures were initiated with UCB Lin- cells at densities of $10^4$, $5 \times 10^4$, $10^5$, and $5 \times 10^5$ /mL. As shown in Figure 2-10a (upper panel), lower cell seeding densities induce greater expansions of total cells (TNC), progenitors (CFC) and primitive progenitors (LTCIC).

In a second series of simulations, cultures were initiated in silico using cell populations with different amounts of stem cell enrichment [from $i = 20$ (~TNC, Lin- plus Lin+) to $i = 8$ (~LTCIC)], and 8-day culture outputs were simulated (Figure 2-10b upper panel). Although the corresponding experiment was not performed, experimental data for similar studies initiated with mononuclear cells (TNC), Lin-, Lin-CD34+, and Lin-CD34+CD38- enriched populations have been published (Conneally et al., 1997; Kohler et al., 1999). Consistent with previous reports, expansion of stem cells (SRC), progenitors (LTCIC, CFC), and total cells (TNC) directly correlate with the degree of progenitor cell enrichment.

Examination of the simulated secreted factor concentrations (Figure 2-10a and b, lower panels) under these conditions yields mechanistic insight. Cultures seeded at lower cell densities and/or
with more enriched populations are cumulatively (concentration × time) exposed to lower levels of endogenously produced proliferation and self-renewal inhibitors (SF1 and SF2), resulting in greater stem and progenitor cell expansions.

![Graphs](image)

**Figure 2-10.** Culture initiation parameters affect cell population expansion via intercellular signaling.

Cultures were initiated *in vitro* (symbols, a only) and *in silico* (lines, a and b) with (a) UCB Lin− cells at cell densities of 10^4 to 5×10^5 cell/mL as indicated (n = 3) or (b) UCB cells enriched sequentially from TNCs (i = 20) to LTC-ICs (i = 8) as indicated, plated at 10^5 cell/mL in each case. Simulated outputs were connected with a continuous intrapolated line. Fold expansion of total cells (TNC), progenitors (CFC), primitive progenitors (LTCIC) and stem cells (SRC) over 8-day unmanipulated cultures are depicted above corresponding simulated concentrations of endogenous proliferation and self-renewal inhibitors SF1 and SF2. Error bars represent standard deviation.
Our inter-cellular feedback model is thus capable of simulating many features of *in vitro* hematopoiesis. As the model contains a large number of (non-identifiable) parameters, we performed analysis to determine if our results could be attributable to parameter “over-fitting”. Intra-cellular network connections were systematically re-wired producing a set of 9 structurally altered models (S1-9) with conserved or decreased number of free parameters. We then calculated the Akaike Information Criterion (AIC) for each, a metric used to rank alternative models based on their ability to explain data with a minimum number of free parameters (Landaw and DiStefano, 1984). Our original model performs significantly better than 8/9 altered models, however deletion of the positive proliferation feedback connection (SF3) produces an incrementally higher ranking (model “S9”), as the data is equivalently described, but with 2 fewer parameters (Figure 2-11). As our culture media is supplemented with high concentrations of proliferation-inducing cytokines, endogenous proliferative signals may be obscured. However, in other *in vitro* or *in vivo* environments, positive feedback loops may play a significant role (Eaves and Eaves, 1988). This connection was hence maintained to enable broader model applicability. We next queried whether the model could be extended to hematopoietic pathologies.

![Figure 2-11. Akaike Information Criterion (AIC)-based ranking of control (C) and systematically altered (S1-9) models.](image)

Based on the AIC ranking metric, our original model (C) scores significantly better than 8/9 of the systematically rewired models, however model S9, where the positive proliferation feedback loop was eliminated, ranks incrementally higher than the control.
2.5.8 Loss of responsiveness to endogenous self-renewal inhibitors simulates stem cell leukemic transformation

Progenitor enriched (Lin⁻ or CD34⁺) umbilical cord blood cells have been used to generate in vitro models of leukemic stem cells (LSC) by transformation with leukemia-associated gene fusions such as MLL-ENL (Barabe et al., 2007), MLL-AF9 (Wei et al., 2008), and TLS-ERG (Warner et al., 2005). The resulting cell lines are immortalized, yet growth factor-dependent and growth factor-responsive, display partially blocked myeloid differentiation patterns, and induce acute myeloid leukemias (AML) in vivo upon transplantation into immunodeficient mice.

We used data from (Warner et al., 2005) as a training set, and the reverse engineering strategy previously described, to determine model parameter sets that most closely simulate long-term leukemic stem cell cultures. The simplest notion of tumourgenesis, as applied in chemo- and radiation therapies, is that cancerous cells have a higher intrinsic proliferation rate. To simulate this, the progenitor proliferation rate ($u_{MAX}$) was increased by 50%. We could account for a partial differentiation block via reducing the Lin⁺ maturation rate ($u_+$) by 50%. Enhanced self-renewal associated with LSCs was simulated by increasing the probability of self-renewal ($f_{MAX}$) by 15%. As malignant transformation involves multiple mutations, we also considered the effect of modulating all three parameters ($u_{MAX}$, $u_+$, and $f_{MAX}$) simultaneously. Finally, we considered a situation wherein stem cell responsiveness to endogenous self-renewal inhibitors ($SF2$) are diminished. Simulated culture outputs are shown in Figure 2-12. Simulations were run using a range of parameter values, however the results were qualitatively unaffected.

Remarkably, modulating any of the three parameters ($u_{MAX}$, $u_+$, or $f_{MAX}$) independently had little effect on the long-term in vitro growth kinetics. Modulating the three parameters simultaneously had additive effects, however the cell line was not immortalized as the stem cells (SRC) were gradually depleted as a result of accumulating differentiated cells and associated feedback inhibition. When the stem cells were simulated to be unresponsive to the endogenous self-renewal inhibitor ($SF2$), they display all of the characteristic features of in vitro-derived LSCs; sustained proliferation (immortalization), reduced differentiation, and constitutive self-renewal. The dynamic oscillations result from the weekly re-plating, as the transformed cells are still
responsive to the endogenous proliferation inhibitor \( SF1 \). Modulating this single topological feature of the model (which may represent a combination of several distinct molecular events) is therefore necessary and sufficient to induce leukemic transformation \textit{in silico}. This is consistent with the parameteric and structural analysis of our model; system dynamics are robust against alterations in kinetic parameters, but sensitive to alterations of the regulatory network structure.

**Figure 2-12.** Loss of responsiveness to self-renewal inhibitor \( SF2 \) alone is capable of inducing leukemic transformation.

Simulated culture dynamics for UCB Lin\(^{-}\) cells re-plated at \( 10^5 \) cell/mL every 7 days. Cumulative fold expansion of total cells (\( TNC \)) \( \text{(a)} \) and stem cells (\( SRC \)) \( \text{(b)} \) are shown for different parameter modifications. Experimental data for total cell output from control (black) and a transformed leukemic cell line (red) as published in Wagner et al. (2006) is overlaid. Error bars represent standard deviation.
2.6 DISCUSSION

The cellular network model described predictively recapitulates many disparate features of \textit{in vitro} hematopoiesis, explaining the enhanced stem and progenitor cell expansions observed at low cell densities, progenitor enrichment, frequent media exchange and progenitor re-selection as a consequence of reduced inhibitory feedback signalling. The model suggests scenarios where stem cell fate is non-cell autonomously controlled by inter-cellular signalling dynamics. This basic concept could be extended to malignant hematopoiesis, as dysregulated responsiveness to inter-cellular signals was sufficient to reproduce characteristic features of \textit{in vitro} leukemic transformation. Additionally, biological variability in the secretion rates of endogenous regulators was sufficient to explain the large sample-to-sample variability observed in culture outputs. While the model was developed specifically for blood stem cell growth \textit{in vitro}, the underlying principles may be more broadly applicable to other stem cell systems both \textit{in vitro} and \textit{in vivo}.

While many individual aspects of our model have been considered previously, to our knowledge this is the first mathematical model of stem cell growth that systematically considers; (1) quantitative functional readouts as lumped and overlapping cell compartments; (2) both proliferation rates and self-renewal probabilities as functions of differentiation state; and (3) proliferation rates and self-renewal probabilities as functions of endogenously secreted positive and negative regulatory molecules. While some of the structural assumptions may be disputed \[i.e.\text{the hierarchy may exist as a functional continuum rather than discrete stages (Quesenberry, 2006) and cells may alter their functional status (undergo compartment transitions) without proliferating (Kent et al., 2008b)}\], these alterations have minimal effects on model outputs. Additionally, the use of differential equations limits the model’s applicability to situations where the stochastic effects of low stem cell numbers are negligible. Modelling the growth of limiting numbers of stem cells requires the use of single cell-based (stochastic) models (Roeder et al., 2008). Further, our analysis has been limited to liquid suspension cultures; extension to stroma-supported culture or \textit{in vivo} hematopoiesis would require additional complexities such as direct cell-cell contact, blood-stroma interactions, and/or hormonal regulatory mechanisms to be incorporated.
Previous models of in vivo hematopoiesis incorporating cell-cell feedback have set kinetic parameters (or probability functions) as directly responsive to neighbouring cell densities (Roeder and Loeffler, 2002; Wichmann and Loeffler, 1985). The limited number of models of in vitro hematopoiesis have either ignored inter-cellular feedback (Varma et al., 1992), or considered it indirectly by setting proliferation kinetics as cell density dependant, assumed to be a result of nutrient depletion as in microbial fermentations (Peng et al., 1996). Hence, our model is the first to explicitly consider a mechanism (secreted regulatory molecules) whereby cells are capable of sensing local cell densities and differentiation status, and applying this regulatory feature to in vitro culture.

Numerous molecular genetic studies have identified a negative relationship between stem cell proliferation and self-renewal - genetic mutations which constitutively enhance HSC cycling generally result in a loss of long-term repopulating cells, and vice versa (Orford and Scadden, 2008). Our findings suggest that this antagonistic relationship may be mediated (at least partially) by non-cell autonomous effects – rapid cycling results in the accumulation of progenitors that inhibit self renewal via feedback signalling. This is supported by recent demonstrations that normally quiescent stem cells are induced to rapidly proliferate in response to progenitor depletion in vivo (Wilson et al., 2008). Data from various in vitro culture systems are also consistent with our model, indicating that that soluble endogenous factors limit stem cell amplification, and progenitor re-enrichment and media exchange/dilution enhance stem and progenitor cell growth (Eaves and Eaves, 1988; Flores-Guzman et al., 2006; Gammaitoni et al., 2004). The antagonistic relationship between mature and primitive cell output hence has direct implications for bioprocess design studies.

Our analysis suggests two complementary approaches to control the inter-cellular signalling networks established in culture. If the molecular regulators are known in sufficient detail, they may be targeted via the use of blocking antibodies or receptor antagonists against inhibitory factors, the exogenous addition of competing stimulatory factors, or by direct the modulation of intra-cellular signalling pathways using small molecules. The specific inhibitory / stimulatory factors may change over time as the culture evolves, likely necessitating dynamic media
supplementation. Conversely, a global culture manipulation strategy may be employed, wherein media exchange or dilution and cell subpopulation-specific selection procedures are performed to indirectly control the inter-cellular signalling network dynamics.

While there are likely to be many exogenous and endogenous sources of culture variability, we demonstrate that variabilities in the secretion rates of the endogenous regulators are sufficient to reproduce the variability observed in culture outputs. It has been reported that donor-to-donor variability in culture output is reduced though the use of stromal feeder cells (Koller et al., 1996), and wide distributions in stem and progenitors numbers is not observed between individuals in vivo under normal physiological conditions. The lack of additional regulatory controls in liquid cultures may result in the amplification of initial biological differences, which would be compensated for by stromal cells in vitro and in vivo. Understanding the emergence of such biological noise in culture will be important for the design of robust cell therapy bioprocesses.

A main limitation of the model in its present form is the secreted regulators SF1-4 remain largely theoretical. We have, none-the-less, identified a limited number of candidate regulatory factors; i.e., TGF-β1, MIP-1β (CCL4), IL-8 (CXCL8), and RANTES (CCL5). It is notable that TGF-β1 is a well-established inhibitor of stem and progenitor expansion in vitro and in vivo (Yamazaki et al., 2009), and IL-8 suppresses myeloid colony formation in vitro (Broxmeyer and Kim, 1999). Many molecules with stimulatory effects on proliferation and self-renewal have been reported as being expressed by hematopoietic cells. Importantly, it is unlikely that any one or even a few of these molecules could be added or blocked to simulate model predictions. Instead, we propose that it is the net balance of opposing signals, integrated by the cell through intracellular signalling pathways, which regulates cell fate decisions.

Our simulations show that while perturbing kinetic constants has little effect on long-term system dynamic, modifying the network structure (deleting the self-renewal negative feedback
loop) is both necessary and sufficient to induce leukemic transformation. This is consistent with experimental data, as leukemic cells are known to display abnormal growth-factor response characteristics, and specifically to be resistant to TGF-β1 inhibition (Lin et al., 2005). A number of other groups have modelled leukemia using both differential equation- and stochastic single cell-based models. While all models rely on a basic assumption that LSCs have a competitive advantage against normal HSCs in vivo, the cell-level parameters responsible for this competitive advantage have yet to be conclusively defined. Either LSCs are assumed to be unresponsive to an intrinsic limit in the stem cell reserve size (Michor et al., 2005) or to have an increased amplification rate (Roeder et al., 2006), both possibilities being conceptually consistent with our model. Unresponsiveness to inhibitory signals would allow the LSCs to expand above their normal physiological limit, resulting in an increased amplification rate and selective advantage against normal HSCs.

The large uncertainties associated with some of the parameter estimates indicate that the model is parametrically non-identifiable, an outcome consistent with the structure of the cellular network (i.e. ratios of competitive interactions may be conserved rather than their characteristic values). Non-identifiability appears to be a universal feature of many systems-level models, and does not necessarily limit their predictive value (Gutenkunst et al., 2007). Consistent with recent models of intra-cellular signalling in mammalian (Chen et al., 2009) and bacterial systems (Barkai and Leibler, 1997), our analysis indicates that key system behaviours are determined primarily by network structure rather than the precise tuning of kinetic parameters.

In summary, our inter-cellular feedback model of blood stem cell growth predictively simulates the dynamic characteristics of both normal and malignant hematopoiesis in vitro. This model may therefore serve as a platform for further experimental interrogation of the regulatory mechanisms controlling stem cell fate in vitro and in vivo, and as a tool for the rational design of stem cell-therapy bioprocesses.
3 DYNAMICS OF A HIERARCHICALLY ORGANIZED TISSUE NETWORK

Portions of this chapter have been submitted to *Cell* (Kirouac et al., 2009b) Authorization to reproduce this work has been obtained from the publisher and co-authors.
3.1 ABSTRACT

Despite the importance of inter-cellular communication networks in regulating stem cell fate decisions, very little is known about their topology, dynamics, or functional significance. Using human blood stem cell cultures as an experimental paradigm, we present an integrated experimental and computational approach to interrogate a hierarchically organized tissue network. Specifically, we have integrated genome-scale molecular profiles (transcriptome and secretome), protein interaction and pathway databases, literature mining, and mechanistic cell population modelling to reconstruct soluble factor-mediated inter-cellular signalling networks regulating cell fate decisions. These can be summarized as a coupled positive-negative feedback circuit wherein stem cell supportive environments are regulated by a balance of megakaryocyte-derived stimulatory factors vs. monocyte-derived inhibitory factors. We define a core intra-cellular self-renewal network wherein these extracellular signals converge for coherent processing into cell fate decisions, representing the first attempt to comprehensively elucidate non-autonomous signals balancing stem cell homeostasis and regeneration.

3.2 INTRODUCTION

Inter-cellular communication networks maintain homeostasis in multi-cellular organisms. These constitute local juxtacrine and paracrine signalling for within-tissue cellular regulation, and systemic endocrine signalling for organism-level tissue regulation. Inter-cellular networks are of particular relevance for stem cell biology, as stem and progenitor cell fate (self-renewal, proliferation, lineage-specification) must be dynamically responsive to physiological demand and external perturbations. Stem cell fate decisions in vivo are regulated via interactions in the bone marrow microenvironment – the stem cell “niche”. Recent progress has been made in elucidating the physical location and cellular components of the niche, including molecular cross-talk between hematopoietic stem cells and niche cells (Kiel and Morrison, 2008). However, local interactions within the microenvironment are insufficient to explain the dynamic responsiveness of stem cells to systemic perturbations.
Under steady state conditions, the niche (both local and systemic) appears to be a largely repressive environment, maintaining stem cells in a dormant state (Foudi et al., 2009). However, following chemical or radiation-induced depletion of cycling progenitors, stem cells are induced to rapidly proliferate, and return to their quiescent state following hematopoietic regeneration (Wilson et al., 2008). Numerous experimental and mathematical studies have indirectly demonstrated hematopoietic stem and progenitor cell fate as responsive to systemic perturbations such as bleeding and irradiation via (undefined) feedback signalling from mature cells (Loeffler and Wichmann, 1980). Such feedback regulatory mechanisms appear to be a conserved feature across other tissue stem cell systems (Conboy et al., 2005; Giangreco et al., 2009; Lander et al., 2009).

Despite the recognized importance of inter-cellular networks in regulating stem and progenitor cell fate, the specific cell populations involved, and underlying molecular mechanisms are largely undefined. Hematopoietic cells are known to secrete and respond to a large numbers of regulatory proteins in lineage- and differentiation stage-specific patterns (Billia et al., 2001a; Majka et al., 2001). In fact, secreted ligands and receptors are expressed in tissue-specific and selective patterns (Kluger et al., 2004). This results in the formation of complex and dynamic inter-cellular signalling networks, providing a mechanism by which cells interrogate and interpret both their local and global environment, propagate this information through signal transduction and gene regulatory networks, and respond appropriately by modulating cell fate decisions. A large number of studies have utilized high-throughput profiling (Muller et al., 2008) and mathematical models (Chickarmane et al., 2009) to reconstruct and analyze intra-cellular molecular networks regulating stem cell fate. While a limited number of studies have applied novel bioinformatic approaches to unravel inter-cellular signalling in other cell systems (Frankenstein et al., 2006; Rendl et al., 2005), a comprehensive analysis of inter-cellular communication in a hierarchical tissue network has yet to be reported.

Recently, we have developed a predictive mathematical model of in vitro hematopoiesis. Through integrated computational and experimental analyses, we have demonstrated that inter-cellular feedback networks regulate blood stem and progenitor cell fate decisions. Furthermore, these networks can be systematically perturbed in vitro by modifying culture parameters to
indirectly control stem and progenitor cell output. (Kirouac et al., 2009a). While this theoretical “top-down” approach is capable of reproducing and explaining the system dynamics, the molecular signals mediating inter-cellular communication remain unknown.

Using our mathematical model and in vitro culture of umbilical cord blood progenitors as an experimental paradigm, herein we describe the structure and dynamics of a hierarchically organized tissue network. By integrating high-throughput molecular profiling (transcriptome and proteome), protein interaction and pathway databases, literature mining, and mechanistic modelling with cell culture experiments, we show that soluble-factor mediated inter-cellular communication networks regulate blood stem cell fate decisions. In particular, primitive progenitor cell fate is modulated by a coupled positive-negative feedback circuit composed of megakaryocyte-derived stimulatory growth factors (EGF, PDGF, EGF) vs. monocyte-derived inhibitory factors (chemokines CCL3, CCL4, CXCL10, TGFB2, and TNFSF9). We reconstruct a core intra-cellular signalling network specifically associated with primitive progenitors and stem cells, and show that this complex milieu of endogenous signals is integrated, and coherently processed by this network, thereby linking extra-cellular and intra-cellular signalling. This work, for the first time, systematically describes how cell fate decisions are regulated non-autonomously through lineage-specific interactions with differentiated progeny. Importantly, this serves as a foundational platform to understand other stem cell systems, and suggests novel strategies to manipulate stem cell fate decisions in a non-cell autonomous manner.

### 3.3 METHODS AND MATERIALS

#### 3.3.1 Cell sample collection and processing

Umbilical cord blood (UCB) samples were obtained from consenting donors according to procedures accepted by the ethics boards of Mt Sinai hospital (Toronto, ON, Canada), Joseph Brandt hospital (Burlington, ON, Canada) and Credit Valley Hospital (Mississauga, ON, Canada). Mononuclear cells were obtained and lineage depleted (Lin⁻) cells isolated from the mononuclear cell fraction using the StemSep system (Stem Cell Technologies, Vancouver, BC, Canada) as described in (Kirouac et al., 2009a).
3.3.2 Cell culture

UCB Lin- cells were seeded at $10^5$ cells/mL (unless otherwise noted) in serum-free Stem Span media (Stem Cell Technologies) supplemented with 100 ng/mL Stem Cell Factor (SCF; R&D Systems, Minneapolis, MN, USA), 100 ng/mL Flt3 Ligand (FL; R&D Systems), and 50 ng/mL Thrombopoietin (TPO; R&D Systems), 1 µg/mL low-density lipoproteins (Calbiochem, La Jolla, CA, USA), and penicillin-streptomycin at 100 U/mL & 100 µg/mL respectively (Invitrogen, Carlsbad, CA, USA). The cell suspension either injected into a cell culture bag of appropriate volume (American Fluoroseal Corporation, Gaithersburg, MD, USA) or placed into wells of a 24-well tissue culture plate (Corning Inc., Corning, NY, USA). Cultures were carried out using a culture bag-based bioprocess described in (Madlambayan et al., 2006) for experiments requiring Lin+ cell depletion, otherwise 24-well tissue culture plates were used. For the ligand functional validation and signal transduction inhibitor studies, 1 mL cultures were supplemented with saturating amounts of recombinant human proteins or small molecules (estimated from the manufacturers data sheets, and indicated below), and subject to a media exchange at day-4. ADIPOQ (1000ng/mL), CCL2 (100ng/mL), CCL3 (50ng/mL), CCL4 (50ng/mL), CCL5 (10ng/mL), CXCL8 (100ng/mL), CXCL10 (50ng/mL), EGF (1ng/mL), FST (500ng/mL), HGF (50ng/mL), IL16 (100ng/mL), PDGFBB (5ng/mL), TGFB2 (10ng/mL), TNFSF9 (10ng/mL), and VEGF (10ng/mL) were purchased from Peprotech (Rocky Hill, NJ, USA), and CXCL7 (100ng/mL) and SPARC (100ng/mL), and were purchased from R&D Systems. The small molecule antagonists LY294002 (10µM), Degulin (10µM), U73122 (10µM), BAY439006 (10µM), and PD0325901 (1µM) were purchased from Axon Medchem (Groningen, The Netherlands), and SB-505124 (1µM) from R&D Systems. Serotonin-creatine sulphide complex (10µM) was purchased from Sigma-Aldrich (St. Louis, MO, USA). All cultures were maintained on an orbital shaker at 37°C in a humidified atmosphere of 5% CO2 in air.

3.3.3 Phenotypic analysis

Analysis of cell surface expression was accomplished by suspending $5 \times 10^4$ cells in 100µL ice cold Hank’s balanced saline solution containing 2% (v/v) human UCB serum (HBSS-HS). The
cells were then incubated on ice for 30 minutes with saturating amounts of fluorescently labelled antibodies; CD34-phycoerthrin (PE) or CD34-fluorescein isothiocyanate (FITC) (Beckman Coulter, Fullerton, CA, USA), and/or CD133-PE (Miltenyi Biotec, Bergisch Gladbach, Germany), and/or CD38-FITC (Beckman Coulter), or appropriate isotype controls for 30 minutes on ice. Staining for Rhodamine123 (Rho123; Molecular Probes, Eugene, OR, USA) was performed as described in (Uchida et al., 1996). A 1mg/mL stock solution of Rho123 (Molecular Probes, Eugene, OR, USA) was prepared in ethanol and stored at –20 °C in the dark. Cells were suspended at < 10^6 cell/ml in HBSS-HS and incubated with Rho123 at 0.1 μg/mL for 30 minutes at 37 °C. The cells were then washed, resuspended in 1 mL HBSS-HS, and incubated for 30 minutes at 37°C to allow cells to efflux the dye. The cells were then either stained with additional antibodies as described above, or stored on ice until analysis. Fluorescence intensity of Rho123 was analyzed in the FITC channel. All samples were washed in HBSS-HS and stored on ice prior to analysis either on a FACSCanto (BD Biosciences, San Jose, CA, USA) or Coulter EPICS XL (Beckman Coulter, Fullerton, CA, USA) flow cytometer.

3.3.4 Progenitor cell assays

Cells were assayed for CFC and LTCIC frequency as described in (Madlambayan et al., 2005) using complete media from Stem Cell Technologies Inc.

3.3.5 Co-culture functional validation bio-assay

Differentiated cells were generated by seeding UCB Lin^- cells at 10^4 cells/mL into serum-free Stem Span media (Stem Cell Technologies) supplemented with the SCF, FL, TPO cocktail described above for myeloid cell production, or 20ng/mL SCF, 2ng/mL IL3, 1ng/mL GM-CSF, and 1U/mL EPO (all from R&D Systems) for erythrocyte production. Cultures were maintained in 25 mL T-flasks, with a media exchange and approximate 5-fold culture dilution performed at day-7. At day-14 the cells were harvested, washed, and suspended in HBSS-HS at 10^6 to 10^7 cell/mL for staining with CD33-PE and CD14-FITC, CD33-PE and CD15-FITC, CD41-FITC, or CD71-FITC and CD235a-PE antibodies (Beckman Coulter). 30 minutes prior to sorting, cells were stained with 7 aminoacitinomycin (7-AAD; Invitrogen) for live cell selection. In parallel,
fresh Lin− cells were first stained with Rho123, washed in ice cold HBSS-HS, then stained with CD34 and CD38 antibodies as described above.

Cells were sorted under sterile conditions at the SickKids-UHN Flow Cytometry Facility (Toronto Medical Discovery Tower, Toronto, ON) on a MoFlo sorter (Dako Cytomation, Glostrup, Denmark). 10 fresh Lin−Rho<sup>lo</sup>CD34<sup>+</sup>CD38<sup>−</sup> cells (Rho<sup>lo</sup> defined as the lowest 25% of the population) were sorted into round-bottom 96-well culture plates (Starstedt, Numbrecht, Germany) containing 100μl standard media. 100 live (7-AAD<sup>−</sup>) CD33<sup>+</sup>CD14<sup>+</sup> (monocytes), CD33<sup>+</sup>CD15<sup>+</sup> (granulocytes), CD41<sup>+</sup> (megakaryocytes), or CD71<sup>+</sup>CD235a<sup>+</sup> (erythrocytes) cells were then subsequently sorted into the wells. Plates were visually checked shortly after sorting using an inverted microscope to ensure the sorting procedure was effective. Culture plates were then incubated for 7-days, after which wells were pooled and re-plated at limiting dilutions (1:1, 1:2.5, and 1:10) onto 96-well flat bottom culture plates (Corning Inc.) with pre-established and irradiated M2-10B4 feeder layers maintained in MyeloCult H5100 media (Stem Cell Technologies). Cultures were maintained for 5 weeks with weekly half-media exchanges, after which media was gently aspirated and replaced with methylcellulose-based media (MethoCult H4435, Stem Cell Technologies). After 14 days of further incubation, wells were visually scored for the presence or absence of CFCs, and LTC-IC frequency determined using the maximum likelihood estimator based on Poisson statistics using L-Calc Software (Stem Cell Technologies).

### 3.3.6 RNA isolation & microarray hybridization

The following time points / conditions were chosen for microarray analysis; d0, d4, d8-S/E, d8-NS/NE, and d12, as these cover a range of HSC culture activity. At each time point cells were separated into Lin− and Lin<sup>+</sup> fractions as described, yielding 10 cell populations in total for analysis. Two biological replicates (n = 2) were used, hence 20 array hybridizations were performed in total. Total RNA was purified using RNeasy Mini Kits (Qiagen GmbH, Hilden, Germany) according to the manufacturer’s instructions. Yield and RNA quality was assessed using the Bioanalyzer 2100 system (Agilent, Waldbrin, Germany). 40 ng RNA per sample (representing 2x10⁴ – 5x10⁴ cells) was subject to linear amplification using the Two-cycle Target
Labelling Kit (Affymetrix, Santa Clara, CA, USA) to generate biotinylated cRNA, which was again assessed for yield and quality using the Bioanalyzer 2100 system. Fragmented cRNA was hybridized to Affymetrix HU133 Plus2.0 GeneChips (Affymetrix) according to manufacturer’s instructions, and scanned using an Affymetrix GeneChip Scanner 3000. Sample amplification, hybridization, and array scanning were performed at The Centre for Applied Genomics (TCAG; The Hospital for Sick Children, Toronto, ON, Canada). Raw data is available from GEO (http://www.ncbi.nlm.nih.gov/projects/geo/; series accession number GSE16589) in compliance with MIAME guidelines.

### 3.3.7 Microarray data processing and analysis

CEL files generated at the TCAG were analysed using dChip v2006 software (www.dchip.org). Arrays were normalized via smoothing spline-least squares regression, artefacts removed, and expression indices (PM/MM) and present/absent calls (P/A) calculated based on the Model Based Expression Index (MBEI) method of Li and Wong (Schadt et al., 2000). For differential expression analysis, probe sets were first filtered as follows:

- 0.5 < Coefficient of Variation (CV) between samples < 1000
- P call in arrays used ≥ 10%
- PM/MM > 20
- CV within replicates < 0.5

Thereby removing genes that are not differentially expressed, genes that are absent in all samples, and genes with high replicate variability. 5,939 of the 54,693 probe sets on the array (10.9%) were thus considered for further analysis. Genome-level expression profiles were visualized using both self organizing maps (SOM) generated in Gene Expression Dynamics Inspector (GEDI) v2.1 (http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm) (Eichler et al., 2003) using a 12x11 grid size (k = 132) with 100 training iterations, and 2-dimensional unsupervised hierarchical clustering heatmaps generated in dChip v2006, via the centroid linkage method with as described in (Eisen et al., 1998).
For pairwise sample comparisons, \( P \)-value cut-offs were adjusted so as to simultaneously minimize the false discovery rate (FDR; calculated via random sample permutations) and maximize sensitivity. The PANTHER database and web-based software program [http://www.pantherdb.org/](http://www.pantherdb.org/) was used to analyze gene-lists for statistical enrichment of biological processes (BP) and 165 expert-curated signalling pathways using the binomial test.

### 3.3.8 Comparison to published gene expression data

To see how our microarray data compares to other published gene expression datasets, we scanned the literature for relevant papers from which gene lists were accessible in the supplementary information. Nine publications were selected, the information from which is summarized in **Supplementary Table S1**. These correspond to a meta-analysis of human Embryonic Stem Cell (HESC) expression profiles (Assou et al., 2007), freshly isolated quiescent vs. cycling mobilized peripheral blood (MPB) CD34\(^+\) cells (Graham et al., 2007), freshly isolated vs. 7d-cultured umbilical cord blood (UCB) CD34\(^+\) cells (Li et al., 2006), freshly isolated UCB CD34\(^+\)CD38\(^-\) vs. CD34\(^+\)CD38\(^+\) and 7d culture-derived slow dividing (SDF) and fast dividing (FDF) cells (Wagner et al., 2004), a shared UCB and MPB CD133\(^+\) cell profile (Chambers et al., 2007), UCB CD34\(^+\) culture-derived erythroblasts (CD235a\(^+\)), monoblasts (CD14\(^+\)), myeloblasts (CD14\(^-\)), and megakaryoblasts (CD41\(^+\)) (Ferrari et al., 2007), bone marrow (BM) CD34\(^+\) culture-derived erythroblasts (CD71\(^+\)), gruanuloblasts (CD15\(^+\)), and megakaryoblasts (CD61\(^+\)) (Komor et al., 2005), 10 primary murine hematopoietic cell populations as defined in the Hematopoietic Fingerprints Database (Chambers et al., 2007), and 37 primary human tissues as defined in (Ge et al., 2005). The murine genes defined in the Hematopoietic fingerprints database were converted to their human orthologues by mapping gene symbols using the Gene ID conversion tool provided in the web-based software and database DAVID [http://david.abcc.ncifcrf.gov/home.jsp](http://david.abcc.ncifcrf.gov/home.jsp) (Dennis et al., 2003), resulting in slightly greater than 50% mapping efficiency.

Bioinformatic software tools such as Gene Set Enrichment Analysis (GSEA) ([Subramanian et al., 2005](#)) or Genomica (Segal et al., 2004), are available to quantify discrete statistical enrichment of gene sets in lists of differentially expressed genes based on hypergeometric or binomial distributions. However, we desired to quantify the expression level of these gene sets
in our microarray expression data in comparison to background levels. We thus calculated the
gene set Activity score \( a_{w_j} \) for each gene set \( (g) \) in sample \( (j) \) as defined in (Lee et al., 2008b):

\[
a_{w_j} = \frac{\sum_{i=1}^{K} Z_{ij}}{\sqrt{K}}
\]

Wherein \( Z_{ij} \) is the Z-score for gene \( i \) in sample \( j \), normalized to background expression across the
total genes comprising each gene set. As the Z-scores are based on a Gaussian
distribution, and PM/MM expression indices follow a log-normal distribution, expression indices
were first log-transformed. Activity scores \( a_{g} \) scores were thus calculated for each of the 55
gene sets defined in Supplementary Table S1 across all 10 samples, producing a 55 \times 10
matrix, with \( |a_g| \geq 2 \) corresponding to \( P \leq 0.05 \) (significantly over or under-expressed). The
average Activity score \( a_g \) for each gene set across all 10 samples was also computed such that
we could compare the relevance of each gene set to our expression data.

3.3.9 Inter-cellular network reconstruction

A list of 270 ligand-receptor interactions (Supplemental Table S11) was manually curated,
largely based on the COPE (Cytokines and Cells Online Pathfinder Encyclopedia) database
(http://www.copewithcytokines.de/). This list was used to define autocrine and paracrine inter-
cellular signalling connections between Lin\(^{-}\) and Lin\(^{+}\) cells at each time point / culture
conditioned profiled (d4, d8-S/E, d8-NE/NE, d12) based on expression patterns. For each ligand
(L) receptor (R) pair, the following Boolean logic operation was performed:

TRUE IF L = P AND L (PM/MM) > 50 AND \[R_1 \text{ OR } R_2 \text{ OR} \ldots R_N = P\]

Where \( R_1, R_2, \ldots R_N \) represent the different possible receptors for a particular ligand. An
expression index of PM/MM > 50 was chosen based on our proteomic data for confidence that
the given ligand will be present at detectable levels in the media. Specific autocrine / paracrine
interactions were constructed between the Lin\(^{-}\) and Lin\(^{+}\) populations based on the choice of L, R
as follows:
For Lin\(^{-}\) autocrine signalling; L = Lin\(^{-}\), R = Lin\(^{-}\)

For Lin\(^{+}\) autocrine signalling: L = Lin\(^{+}\), R = Lin\(^{+}\)

For Lin\(^{-}\) to Lin\(^{+}\) paracrine signalling; L = Lin\(^{-}\), R = Lin\(^{+}\)

For Lin\(^{+}\) to Lin\(^{-}\) paracrine signalling; L = Lin\(^{+}\), R = Lin\(^{-}\)

This analysis produced a list of the inter-cellular signalling loops active for each culture condition (58 in total, ranging from 26 to 41 activate per condition).

The Lin\(^{+}\) population is heterogeneous, comprised of all erythro-myeloid lineages (monocyte, granulocyte, erythrocyte, and megakaryocyte). To estimate which mature cell sub-populations within the Lin\(^{+}\) population were responsible for specific interactions, and thereby better define the inter-cellular network architecture, gene expression profiles of in vitro-generated erythroblasts megakaryoblasts, and monoblasts published in (Ferrari et al., 2007) and described above were downloaded. These expression profiles were chosen as the cell populations were generated under similar conditions (short-term liquid culture of UCB CD34\(^{+}\) progenitors), as reflected in the Activity scores.

As the authors used a different normalization procedure (Robust Multi-array Average (RMA)) rather than Model Based Expression Index (MBEI) applied to our data, the expression indices are not directly comparable. However, the expression distributions are very similar. We applied a conservative cut-off and defined genes with expression indices in the top 40% as positive (P) and bottom 60% and absent (A). For each of the 58 signalling loops scored as active in one or more culture condition, differential expressed genes, and proteins detected in conditioned media, P/A calls for the ligands and receptors in the Lin\(^{-}\) d8-S/E population (our data) as well as the erythroblast megakaryoblast, and monoblast populations (Ferrari et al. (2006) data). This information was converted to a directed graph with 2 classes of vertices (cell populations and ligands), and edge directionality indicating cell population-specific ligand and receptor
expression, producing a set of theoretical cell-cytokine-cell interactions. The graph was visualized in Cytoscape v2.5.

### 3.3.10 Protein-protein interaction network mapping

Entrez Gene IDs for gene lists of interest (HSC self-renewal genes and PANTHER-defined pathways) were searched against the i2D database (Brown and Jurisica, 2005) ([http://ophid.utoronto.ca/ophidv2.201/index.jsp](http://ophid.utoronto.ca/ophidv2.201/index.jsp)), consisting of 60675 proteins and 92545 (non-unique) interactions compiled from multiple independent databases (HPRD, MINT, DIP, etc.).

### 3.3.11 Cytokine functional interaction network mapping

For the set of 77 ligands used to define the inter-cellular network, an extensive literature search was performed using PubMed and the COPE databases for functional interactions between the given ligands. Functional interactions are defined as expression induction and repression (at the level of mRNA and/or protein secretion), intracellular signalling synergy or antagonism, and physical extracellular interactions, identified in at least one of the cell types produced in our cultures (HSCs, progenitors, or mature myeloid cells). This information was converted into partially directed graph with ligands as vertices.

All graphs / networks were visualized using Cytoscape v2.6 (Shannon et al., 2003) ([http://www.cytoscape.org/](http://www.cytoscape.org/)), and densely connected modules detected using the MCODE (Molecular Complex Detection) Cytoscape plugin (Bader and Hogue, 2003).

### 3.3.12 Conditioned media proteome analysis – Raybiotech™ antibody arrays

Conditioned media samples corresponding to each culture condition and time point used for microarray profiling (d4, d8-S/E, d8-NS/NE, d12, and additionally d16) were assayed in
duplicate using the Raybio Human Cytokine Array C Series 2000 (Raybiotech Inc., Norcross, GA, USA) following the manufacturer’s instructions. The arrays contain antibody spots against a set of 120 cytokines, chemokines, proteases, and soluble receptors, functioning as a multiplex sandwich ELISA. Briefly, the membranes were blocked by incubation with the blocking buffer at room temperature for 30 min and incubated with 1 mL sample at room temperature for 1 h. Membranes were washed three times with Wash Buffer I and two times with Wash Buffer II at room temperature for 5 min per wash and incubated with biotin-conjugated antibodies at 4 °C overnight. Finally, the membranes were washed, incubated with HRP-conjugated streptavidin at room temperature for 1 h and with detection buffer for 1 min, and exposed to X-ray film for 40 s. Chemiluminescence images were recorded and quantified using the ChemImager 5000 (Alpha Innotech, San Leandro, CA). Individual spot intensities were quantified by calculating the Z-scores against background (negative) spot intensities, and expressed as % control spot intensity for normalization.

3.3.13 Conditioned Media Proteome analysis – Luminex™ Liquid Chips

Conditioned media samples corresponding to each culture condition and time point used for microarray and antibody array profiling (d4, d8-S/E, d8-NS/NE, d12, and additionally d16) were also assayed in duplicate using the Biosource Human Cytokine 30-Plex detection kit (Invitrogen). These kits utilize Luminex microshperes (Luminex Co., Austin, TX, USA) as a fluid platform for multiplex sandwich ELISA. The “microspheres” consist of 5 μm polystyrene beads bar-coded via unique ratios of APC: APC-Cy7 dye. Each colour-coded microsphere contains primary capture antibody against an individual cytokine, which in combination with secondary PE-conjugated detection antibody, was used to quantify the concentration proteins in a test samples (detection limit ≥ 10 pg/mL) via flow cytometry as described previously (Krouac et al., 2009a).
3.3.14 Statistics

The effect of exogenous ligands on cell population expansions were determined using linear regression ANOVA, implemented using JMP Student Edition v.6 software (SAS Institute Inc., Cary, NC, USA).

The hypergeometric Z-score is given by:

\[
Z = \frac{r - n \left( \frac{R}{N} \right)}{\sqrt{n \left( \frac{R}{N} \right) \left( 1 - \frac{R}{N} \right) \left( 1 - \frac{n-1}{N-1} \right)}}
\]

Where \( N \) = total number of elements (i.e. genes), \( R \) = total number of positive elements, \( n \) = sample size, and \( r \) = number of positive elements in sample, as described in (Doniger et al., 2003).

For transcriptome-secretome comparisons we wished to test whether proteins with highly expressed transcripts were more likely to be detected in conditioned media. We compiled secreted factor gene sets corresponding to the antibodies on the Luminex (28) and Raybiotech (94) arrays (overlap = 68). For each time-point / culture condition at which conditioned media was profiled (d4, d8-NS/NE, d8-S/E, d12), corresponding mRNA expression indices (PM/MM) were calculated based on the average of the Lin\(^+\) and Lin\(^-\) populations. Protein signals were converted to binary values (1/0) based on their detection, and mRNA expression indices (PM/MM) converted to an expression ranking from highest (100 percentile) to lowest (0 percentile). For RayBiotech Antibody arrays \( N = 380 \) and \( R = 22 \); for Luminex Liquid Chip arrays \( N = 112 \) and \( R = 29 \). A quartile sample size (\( n = 95 \) and 28 respectively) was used to calculate Z-scores across the expression ranking (i.e. from 0-25% to 75-100%). Different sample sizes (\( n = 0.05 \times N \) through 0.5\( \times N \)) were tested however the 25% window was found to produce an optimal balance of high resolution and minimal noise.
For the ligand validation studies we wished to test whether the distribution of stimulatory and inhibitory effects on LTCIC output was statistically correlated to our predictions. For stimulatory effects \( N = 17, R = 3, n = 5, \) and \( r = 3; \) for inhibitory effects \( N = 17, R = 7, n = 7, \) and \( r = 5; \) and for null effects \( N = 17, R = 7, n = 5, \) and \( r = 4. \)

For testing test whether the common stimulatory and inhibitory signal transduction molecules were enriched in the self-renewal we considered \( N = \) total genes in the PANTHER pathways database = 2995; \( R = \) overlap with the HSC self-renewal network = 599 (20%); \( n = \) common signal transduction molecules = 35; \( r = \) overlap with the active HSC self-renewal network = 22 (63%).

To test whether the 15 genes from Deneault et al. (2009) were statistically enriched in the HSC self-renewal network, we considered as background gene set the annotated genes (assigned an Entrez Gene ID) represented on the Affymetrix HU133 Plus2 GeneChip, as the authors compiled data from gene expression studies, and only well-annotated genes generally have human-mouse orthologues; \( N = 20177, R = 15, n = 1728, \) and \( r = 6. \)

3.3.15 Model Simulations

The mathematical model described in (Kirouac et al., 2009a) was simulated using MATLAB R2008b (The Mathworks, Natick, MA, USA) software.

3.4 RESULTS

3.4.1 Genome-wide expression patterns correlate with cellular functional activity

To systematically interrogate the regulatory mechanisms in a hierarchically organized tissue network, we used the \textit{in vitro} propagation of umbilical cord blood (UCB)-derived stem and
progenitor cells in serum-free, minimal cytokine supplemented, liquid cultures as a model system. We have developed a mathematical model of hematopoiesis which predictively simulates experimentally measured population expansions in vitro (Kirouac et al., 2009a). The model is implemented as a system of ordinary differential equations, experimental and in silico analyses of which has demonstrated that the cell population dynamics are attributable to intercellular signalling rather than cell-autonomous effects (chapter 2). Additionally, our laboratory has developed a bioprocess for the clinical-grade expansion of umbilical cord blood-derived HSCs (Madlambayan et al., 2006) based on manipulation of feedback signalling in culture (Madlambayan et al., 2005). By performing Lin− cell re-selection and media exchange during culture (“S/E”) primitive progenitor (long term culture-initiating cell; LTCIC) and stem cell (Scid-mouse repopulating cell; SRC) expansions are significantly enhanced in comparison to control un-manipulated (“NS/NE”) cultures.

UCB-derived Lin− cells were cultured in serum-free, cytokine-supplemented media for 12 days in total, subject to culture manipulations (S/E) on culture days 4 and 8, or un-manipulated (NS/NE). On culture days 0, 4, 8 and 12, quantitative functional (total nucleated cell [TNC], progenitor [colony forming cell; CFC], and primitive progenitor [LTCIC]) and phenotypic (CD133, C34, CD38, and Rho123 staining) assays were performed to evaluate culture-derived cells. As depicted in Figure 3-1a, this covers a range of functional activity, including HSC-expansion (d4, d8-SE) and HSC-depletion (d8-NSNE, d12) phases.

In parallel with the functional and phenotypic analyses, cells were physically separated into undifferentiated (Lin−) and differentiated (Lin+) sub-populations for transcriptional profiling using Affymetrix GeneChips. To depict sample relationships, genome-wide expression patterns are displayed as self-organized maps in Figure 3-1b.
Figure 3-1. Functional and molecular profiling of hematopoietic progenitor (Lin⁻) and differentiated (Lin⁺) cells in serum-free liquid culture.

Umbilical cord blood derived-Lin⁻ cells were cultured in serum-free cytokine supplemented media for 12 days with Lin⁻ cell selection and media exchanges (“S/E”) performed every 4 days, or for 8 days without the S/E procedure performed at day-4 (“NS/NE”). Functional assays for total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) content were performed (a) in parallel with transcriptome profiling of Lin⁻ and Lin⁺ cells, and sub-proteome profiling of conditioned media. Genome-wide expression profiles for fresh and culture-derived Lin⁻ (blue) and Lin⁺ (red) cells are represented via self-organized maps, linked by arrows indicating parental relationships (b). Cell population dynamics were extracted from the differential gene expression data by computing Activity scores for the 55 published gene sets (indicated by 1ˢᵗ author – cell population), depicted as a hierarchical clustered heatmap and associated dendogram (c). Only gene sets with Activity scores ≥ 2 in at least one sample are shown (25 in total).

There exists a wealth of published microarray and SAGE data for hematopoietic stem cells, progenitors, and differentiated cell populations isolated under various conditions. We wished to utilize these available resources to extract characteristic gene sets for the various cell populations as a means to interrogate the composition of our fresh and culture-generated cells. We compiled 9 publications covering a wide range of cell populations, both fresh and culture-derived, from hematopoietic and non-hematopoietic tissues. From these we define 55 cell population-characteristic gene sets (Supplementary Table I) based on the statistical filtering methods used in the individual publications. We then used an Activity score (see Methods and Materials) to characterize the distribution of transcript expression levels in the defined gene sets in comparison to the background signal on the arrays. As shown in Figure 3-2, gene sets characterizing hematopoietic cells, particularly cycling and culture-derived hematopoietic cells are highly expressed, whereas gene sets characterizing non-hematopoietic tissues are in general under-expressed. Our expression data hence broadly correlates, in a platform independent-manner, with a range of previously published data sets.
To interrogate the cellular composition of each profiled population, Activity scores \( (a_g) \) for each sample are displayed as a heat map, with gene sets (filtered for \( a_g \geq 2 \), corresponding to \( P \leq 0.05 \) in at least one sample population) organized via unsupervised hierarchical clustering (Figure 3-1c). Gene sets cluster distinctly into fresh vs. culture-derived, up/down-regulated precisely in accordance with functional status of the cell populations profiled. Examining the differentiated cell populations, \textit{in vitro}-derived megakaryocyte gene sets [from both (Komor et al., 2005) and (Ferrari et al., 2007)] are highly expressed in HSC expansion-associated Lin\(^+\) cell populations (d4 and d8-SE), while under expressed in HSC depletion-associated Lin\(^+\) populations (d8-NSNE and d12). \textit{In vitro}-derived erythrocyte gene sets are however constitutively expressed across all Lin\(^+\) cell populations. Megakaryocyte development hence correlates with stem cell output in culture, while erythrocyte development is constitutive, non-correlated with stem cell dynamics.

Cell populations cluster according to function; fresh (d0) cells cluster distinctly from culture-derived cell populations; within the culture-derived populations, Lin\(^-\) (undifferentiated) cells cluster separately from Lin\(^+\) (differentiated) cells; and Lin\(^+\) populations cluster according to stem and primitive progenitor cell activity with HSC-expansion (d4, d8-SE) vs. HSC-depletion (d8-NSNE, d12) sub-groups. These clustering patterns indicate that gene expression activity may be used to elucidate functional information, and that stem cell output in culture may be regulated non-autonomously via Lin\(^+\) differentiated cell populations.
Figure 3-2. Average Activity Scores for 56 Published Gene Sets Averaged across 10 Experimental Samples.

Activity scores were computed for the 56 gene sets described in Table II for each of the 10 profiles, using total gene expression across the array for background comparison. Average Activity scores across the 10 populations ± standard deviations are shown for each gene set.
3.4.2 Gene expression levels predict cellular phenotypes and secretome profiles

To gain confidence on our microarray data we wished to quantify the relationship between gene expression and corresponding protein cell activity. We compared mean fluorescence intensities of the blood progenitor cell surface markers CD34, CD133, CD38, and intracellular Rhodamine 123 dye (Rho123), measured via flow cytometry, to $CD34$, $CD133$, $CD38$, and $ABC-B1$ transcript expression levels [the ABC transporter is responsible for Rho123 efflux property associated with quiescent stem cells (Uchida et al., 1996)]. As shown in Figure 3-3a, positive correlations exist between CD34, CD133, and CD38 gene-protein indices, and a negative correlation between ABC-B1 expression and Rho123 staining, as expected. All four relationships fit well to 2nd degree polynomial curves ($r^2 = 0.874, 0.924, 0.842, and 0.857$ for $CD34$, $CD133$, $CD38$, and $ABC-B1$/Rho123 respectively), likely due to a saturation effect at high gene expression levels.

In parallel with functional, phenotypic, and gene expression analyses of cell populations, condition media samples were collected and analysed for “secretome” (secreted protein) profiles using both RayBiotech (n = 30) and Luminex (n = 120) cytokine antibody array systems. While the majority of secreted proteins scored Present at the transcript level were not detected at protein level in conditioned media, scatter plots show a crude correlation between transcript abundance and protein signal (Figure 3-3b, c). This is not unexpected, as numerous undefined biological and technical factors affect the relationship between intracellular transcript abundance and extracellular protein abundance. Hence, we decided to binarize the secretome data (Present vs. Absent) and statistically quantify the relationship between transcript expression ranking and the probability of protein detection. Z-Scores, based hypergeometric distribution with a quartile sampling size, show that secreted proteins are statistically under-represented at the lower range (< 60% for Raybiotech arrays) and over-represented at the higher-range of expression ranking (> 75% for Raybiotech arrays and > 45% for Luminex arrays) (Figure 3-3d). In summary, these results demonstrate that mRNA expression indices correlate with complex measures of proteome activity (secretion and cell surface expression), lending confidence for further analysis of the array data.
Figure 3-3. mRNA expression indices correlate with cell surface expression and secretion of proteins.
mRNA expression indices (PM/MM) of CD34, CD133, CD38, and ABC-B1 transcripts vs. fluorescence of corresponding cell surface proteins (CD34, CD38, and CD133) and functional activity (Rho123 exclusion) measured via flow cytometry throughout culture (a). 2nd degree polynomial curves were fit to the individual data sets with $r^2$ values of 0.874, 0.924, 0.842, and 0.857 respectively. Scatter plots and trend-lines for mRNA expression index (PM/MM) vs. normalized protein signal intensity (Raybiotech antibody array (b)) and protein concentration (Luminex liquid chips (c)). Hypergeometric Z-Scores for the expression ranking of a secreted protein transcript vs. the probability of detection in conditioned media via Luminex (red) and Raybio (blue) antibody array systems, using a quartile sampling size (d). Z-scores of ± 2 are indicated by dashed lines, corresponding to enrichment / depletion $P$-values < 0.05.

3.4.3 Pairwise sample comparisons identify cell population dynamics and signalling pathways associated with HSC expansion and depletion in vitro

We next wished to identify gene sets associated with changes in functional activity in culture, hence we performed a number of pairwise sample comparisons. As shown in Supplementary Table S2, we performed 6 baseline comparisons (B1-6) to identify transcriptome changes associated with fresh vs. culture-derived cells and differentiated (Lin+) vs. undifferentiated (Lin-) cells as a quality control measure, and 6 experimental sample comparisons (C1-6) to identify transcriptome changes associated with HSC-expansion vs. HSC-depletion in culture. Differential expressed gene lists were analyzed for biological process and pathway enrichment using PANTHER web-based software and database. Biological process changes inferred from gene expression ($P \leq 10^{-4}$) are consistent with what would be expected from the baseline comparisons.

Examining the number of differentially expressed genes for the various sample comparisons alone reveals biologically relevant features (Supplementary Table S2). For example, more than twice as many genes are down-regulated as up-regulated upon culture (B2, B3, and B6).
Stimulation of the cells with high-doses of only 3 cytokines at low density in vitro vs. low-level stimulation by many more cytokines at much higher cell densities in vivo would be expected to result in culture-induced gene expression program collapse. Similarly, comparing culture-derived Lin⁻ vs. Lin⁺ cells (B4 and B5) reveals 2.8 and 4-fold as many genes up-regulated vs. down-regulated respectively. This finding is consistent with previous studies suggesting that undifferentiated stem and progenitor populations are “lineage primed”, expressing low levels of many lineage associated genes which are turned off upon lineage commitment and differentiation (Bruno et al., 2004).

By examination of Supplementary Tables S3 through S14, it is evident that biological process changes inferred from gene expression \( (P \leq 10^{-4}) \) are consistent with what would be expected from the baseline comparisons. For both fresh (Table S3) and culture-derived (Table S9) cells, Lin⁻ populations are consistently enriched in cell cycle control genes \( (P = 2.35 \times 10^{-6} \text{ and } 3.44 \times 10^{-8} \text{ respectively}) \), and Lin⁺ cells in immunity and defence (i.e. white blood cells) and porphyrin metabolism (i.e. red blood cells) genes \( (P = 2.72 \times 10^{-25} \text{ and } 8.20 \times 10^{-10} \text{ respectively}) \). This is expected, as a defining feature of blood progenitor populations is tight cell cycle control machinery (Passegue et al., 2005), and Lin⁺ cells by definition contain mature blood cell populations. Lin⁻ (Table S5), Lin⁺ (Table S8), and combined cell population comparisons (Table S14) reveal culture-derived cells up-regulate cell cycle genes \( (P = 6.75 \times 10^{-13}, 9.40 \times 10^{-7}, \text{ and } 1.11 \times 10^{-11} \text{ respectively}) \), and down-regulate signal transduction genes \( (P = 7.35 \times 10^{-7}, 1.03 \times 10^{-10}, \text{ and } 6.05 \times 10^{-17} \text{ respectively}) \). Again, these findings are expected as cells in culture are actively proliferating in response to high-dose cytokine stimulation, while cells in vivo (under non-regenerative conditions) are largely quiescent in response to a microenvironment consisting of low-level stimulation by many autocrine and paracrine ligands.

Results from the experimental sample comparisons (C1-C6) were grouped into 3 classes; HSC-expansion correlated, HSC-depletion correlated, and non-correlated. In addition to analyzing differentially expressed genes for biological process and pathway enrichment, we also filtered genes lists for known ligands and receptors to identify autocrine and paracrine factors mediating
endogenous pathway activation. Results for Biological process, pathway, secreted factor, and receptor expression changes are displayed in Supplementary Tables S15 through S25. While the \( P \)-values are lower in comparison to the baseline comparisons (as the populations are more similar), a number of interesting features emerge from analysis of these tables (\( P \leq 10^{-3} \)).

HSC-expansion is consistently associated with blood clotting (Table S15) and blood coagulation pathways (Table S18) (i.e. megakaryocyte development) while HSC-depletion is associated with macrophage-mediated immunity (Table S17) (i.e. monocyte development), consistent with compositional analysis presented in Figure 1c. HSC-expansion correlates with 5HT2, \( \alpha \)-adrenergic, Metabotropic glutamate, and Muscarinic acetylcholine receptor signalling, (Table S18) which constitute GPCRs for the neurotransmitters serotonin, adrenaline/noradrenaline, glutamate, and acetylcholine respectively. This is consistent with reports suggesting a role for neurological signals in regulating stem cell fate \textit{in vivo} (Spiegel et al., 2008), but novel in that these signals are apparently being activated endogenously by hematopoietic cells, rather than by niche-infiltrating neurons. HSC-expansion is also associated with cell adhesion (Table S15), and specifically up-regulation of ICAM2, ICAM3, SELL, and SELP genes (Table S21), which may be simply related to megakaryocyte enrichment. HSC-depletion correlates (though at a lower level of significance, \( P < 0.005 \)) with TGF-\( \beta \) signalling (Table S20), corresponding to up-regulation of the TGF-\( \beta \) ligands TGFB2 and FST, and down-regulation of the TGF-\( \beta \) inhibitor LTBP1 (Table S23 and S21). Analogously, HSC expansion correlates with activation of VEGF signalling (Table S18) and up-regulation of VEGFC (Table S21). In summary, this analysis identified pathways, cell populations, and endogenous secreted factors associated with stem cell expansion and depletion in culture.

### 3.4.4 Inter-cellular network reconstruction

We next sought to systematically reconstruct the inter-cellular signalling networks established in culture by integrating our gene expression profiles with publicly available datasets. Based on expression patterns of a manually curated set of 230 secreted ligand-receptor pairs (Supplementary Table S26), we defined “autocrine” and “paracrine” signalling interactions between Lin\(^-\) and Lin\(^+\) cell populations for each time point / culture condition, identifying a set of
57 putative signalling interactions active in at least one condition analysed (Figure 3-4a,b). Including secreted factors that were differentially expressed and/or consistently detected in conditioned media, we produced a non-redundant list of 74 inter-cellular signalling molecules for further analysis (Supplementary Table S27).
**Figure 3-4. Reconstruction of inter-cellular network dynamics from gene expression data.**

(a) Schematic depiction of methodology: based on Boolean expression patterns of secreted proteins and associated receptors, Lin⁻ autocrine (−/−), Lin⁺ autocrine (+/+), Lin⁻ to Lin⁺ paracrine (−/+), and Lin⁺ to Lin⁻ paracrine (+/−) signalling loops are defined for the four culture time points / conditions (d4, d8SE, d8NN, and d12).

(b) Boolean mapping and hierarchical clustering of 58 signalling loops scored positive in at least 1 of the conditions (red = present, blue = absent).

(c) Dynamic representation of signalling between culture-derived Lin⁻ progenitors (LIN⁻) and differentiated megakaryocytes (MEG), erythrocytes (E), and monocytes (MONO) mediated via secreted proteins which were both differentially expressed and scored as active signalling loops, colour coded according to differential expression patterns.

To interrogate how these 74 molecules (as well as the three exogenously provided ligands KITL, FLT3L, and THPO) mediate interactions between the Lin⁻ and heterogeneous Lin⁺ cells produced in culture (see Figure 3-5), we mapped the Lin⁺ ligand / receptor expression patterns onto the megakaryoblast, erythroblast, and monoblast expression profiles published in (Ferrari et al., 2007). This data set was chosen as the experimental protocols were very similar to ours (liquid culture of UCB progenitors), as reflected in the gene set Activity scores. This information was converted to a directed graph representing a set of theoretical cell-cytokine-cell interactions, with vertices (cells and ligands) colour-coded according to correlation with HSC-expansion / depletion (Figure 3-6). The result is a densely connected network, each cell population both secreting and stimulated by multiple ligands.

Limiting our analysis to those ligands which are both differentially expressed and scored as positive in our intercellular signalling loops suggests that complex and dynamic interactions between positive and negative regulators may tip the balance between stem cell supportive vs. non-supportive conditions (Figure 3-4c). In addition to the dynamic changes in gene expression, other cell culture variables will globally affect the network – the concentration of endogenous ligands would increase with increasing cell density and culture duration, resulting in the functional dominance of endogenous compared to exogenously supplied factors.
Figure 3-5. Emergent cell population heterogeneity in vitro.

Typical flow cytometry profiles of an in-vitro derived cell population after 8 days of culture. Lin- progenitors proliferate and differentiate to produce CD133+CD38- progenitors, as well as differentiated (CD133-) erythrocytes (CD71+CD235a+), megakaryocytes (CD41+), granulocytes (CD33+CD15+), and monocytes (CD33+CD14+).

Dimensionality of the network may also be reduced by focusing on families of ligands – Figure 3-6 b and c depict two sub-networks extracted from the total network; TGF-β, and receptor tyrosine kinase (RTK) signalling growth factor (EGF, PDGF, and VEGF) modules respectively. While all populations secrete TGF-β ligands, megakaryocytes also secrete the TGF-β inhibitor LTBP1, as well as all three of the expansion-associated RTK growth factors. It is also notable that all three RTK pathways are known to display cross-talk and functional synergistic effects. Examination of select sub-networks hence provides some mechanistic detail behind the putative cell population-ligand functional interactions.
a

HSC Expansion correlated
Non-correlated
HSC Depletion correlated
Non-differentially expressed
Ligand / receptor expression
**Figure 3-6. Reconstructed inter-cellular signalling network.**
Signalling between culture-derived Lin- progenitors (LIN-) and differentiated megakaryocytes (MEG), erythrocytes (E), and monocytes (MONO) mediated via 77 secreted proteins, colour coded according to differential expression patterns. (a) Total network. (b) TGF-β sub-network. (c) VEGF / EGF / PDGF sub-network. Functional interactions between secreted proteins are indicated via grey connections in the two sub-networks; LTBP1 inhibits TGF-β ligands from activating the corresponding receptors (b), and PDGF is known to display intracellular synergism with both EGF and VEGF signalling (c).

### 3.4.5 Inter-cellular network validation

We next sought to functionally test the correlative data derived from the gene expression studies through a series of cell culture experiments. Cultures were designed to replicate the conditions under which the putative regulatory factors were identified, thereby capturing both direct and indirect effects. UCB Lin’ cells were cultured in serum-free cytokine-supplemented media for 8-days as before, with or without saturating amounts of exogenous putative regulatory ligands, and assayed for total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) population-fold expansions. Ligands for validation were selected based on the level of expression (PM/MM > 100), signalling loop activity, and commercial availability. We tested 18 of the potential regulatory factors; expansion-correlated ligands EGF, PDGF-ββ, IL-16, SPARC, and VEGF-A; depletion-correlated ligands CCL2, CCL3, CCL4, CXCL10, FST, TNF-SF9, and TGF-β2; non-correlated ligands CCL5, CXCL8, HGF, and THBS1; as well as 2 proteins consistently detected in conditioned media but not differentially expressed, ADIPOQ and CXCL7.

Total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) population expansions in comparison to control cultures are shown in **Figure 3-7a**. TGF-β2 was the only factor found to significantly affect TNC and CFC output (negatively), consistent with previous studies (Cashman et al., 1990). Remarkably, at the level of LTCIC population outputs, 3/5 predicted stimulatory ligands (EGF, PDGFB, and VEGF) displayed significant positive effects, 5/7 predicted inhibitory factors (CCL3, CCL4, CXCL10, TNF-SF9, and TGF-β2) displayed negative effects, while only 1/5 non-correlated ligands (CXCL7) displayed a significant effect. *P*-values
(based on hypergeometric distributions) for this distribution of stimulatory, inhibitory, and null effects are 0.0042, 0.039, and 0.041 respectively, hence patterns of differential expression predict, with statistical significance, functional effects on primitive progenitor (LTCIC) output in culture. We also tested the functional activities of serotonin (5HT1) and the TGF-βRI antagonist SB-505124. Consistent with predictions, serotonin stimulation specifically enhanced LTCIC output, while inhibition of endogenous TGF-β signalling enhanced CFC but inhibited LTCIC expansions (Figure 3-8). The effects however are not statistically significant ($P = 0.1$ to 0.2).

While it was not practically feasible to perform *in vivo* repopulation assays (SRC) for this screen, experiments are currently underway to test a subset of molecules for effects on SRC expansion. It is however notable that our previous work has identified LTCIC readouts as strongly correlative to SRC expansion in this system (Kirouac et al., 2009a).

We similarly wished to test the functional effects of the Lin$^+$ sub-populations on primitive progenitor (LTCIC) expansion, to see if there is a similar correlation between cell-type specific gene expression patterns and functional activity. Thus, we developed a novel bio-assay to test direct functional interactions between mature cell populations and stem cells, schematically depicted in Figure 3-9 and described in the Methods and Materials. As shown in Figure 3-7b, in comparison to control cultures monocytes (and erythrocytes) have a significant inhibitory effect, while megakaryocytes have a significant stimulatory effect on LTCIC expansion *in vitro*. We further calculated the Activity scores for the stimulatory, inhibitory, and non-functional ligand gene sets in available expression profiles; CD14$^+$ monocytes, CD41$^+$ megakaryocytes, and CD235a$^+$ erythrocytes from (Ferrari et al., 2007). Figure 3-7c shows that stimulatory ligands are enriched in megakaryocytes and depleted in monocytes, while inhibitory ligands display the inverse expression pattern. This set of experiments therefore confirms predictions from the gene expression analysis; production of megakaryocytes enhances, while production of monocytes inhibits stem cell expansion *in vitro* in a non-autonomous manner via cell population-specific secretome profiles.
Figure 3-7. Functional validation of cell population and secreted protein effects on progenitor expansion in vitro.
(a) 8-day fold-expansions of total cells (TNC), progenitors (CFC), and primitive progenitors (LTCIC) for cultures supplemented with stem cell expansion (green), depletion (red), and non-correlated (yellow) ligands in comparison to control cultures. Error bars = std, n = 4 to 8. (b) 7-day Fold LTCIC expansion from an enriched primitive cell population (LinRho−CD34+CD38−) co-cultured with in vitro-generated monocytes (CD14+), granulocytes (CD15+), megakaryocytes (CD41+), and erythrocytes (CD235a+) in comparison to control cultures. Error bars = 95% CI, n = 3. (c) Activity scores for secreted protein gene sets correlated with stem cell expansion (green), depletion (red), and non-correlated (yellow), for megakaryocyte, erythrocyte, and monocyte cell population expression profiles from Ferrrari et al. (2006).

Figure 3-8. Functional effects of serotonin (5HT1) and the TGF-β inhibitor SB505124 on culture output.

8-day fold expansion of total cells (TNC), progenitors (CFC), and primitive progenitors (LTCIC) from liquid cultures supplemented with serotonin (5HT1) (a) or SB505124 (b) compared to control cultures. Based on the relative effects on TNC, CFC, and LTCIC growth, serotonin signalling displays SF4-type activity, and TGF-β inhibition displays SF3-type activity. While results are not statistically significant (p ≥ 0.1) performing more replicate experiments would enhance the statistical power and may add confidence to the results. Error bars = std, n = 3 (a) and n = 7 (b).
Figure 3-9. Schematic representation of co-culture bioassay workflow.
Representative FACS sorting strategy and colony readouts. Blue arrows indicate work flow.
3.4.6 Model simulations enable functional classification of endogenous regulatory ligands

Our mathematical model of hematopoiesis is a useful tool for quantifying the relationship between system parameters and cell fate outcomes (Kirouac et al., 2009a), thus providing a theoretical framework capable of classifying experimentally identified regulatory ligands into distinct functional categories. To interrogate the mechanisms of cytokine activity, we simulated 8-day UCB Lin⁻ cell cultures with a range of exogenously supplemented soluble regulatory factors - proliferation and self-renewal inhibitors (SF1 and SF2) and stimulators (SF3 and SF4) (Figure 3-10a). Figure 3-10b depicts theoretical dose-response relationships between SF concentrations (normalized to ED₅₀ values) and 8-day fold expansions of the three cell populations (derived from phase portraits presented in Figure 3-11). Cell population expansion data presented in Figure 3-7a is overlaid for direct comparison between theoretical and experimental outputs. Examining the relative dose-response sensitivities, TGF-β2 displays SF1-type effects (inhibition of all three cell population expansions), the chemokines CCL3, CCL4, CXCL10 and TNFSF9 display SF2-type effects (targeted inhibition of LTCIC output), and EGF, PDGFB, VEGFA (and possibly serotonin) display SF4-type effects (targeted stimulation of LTCIC output). None of the ligands displayed SF3-type effects, consistent with computational analyses indicating that deletion of the SF3 feedback loop is dispensable for model performance (Kirouac et al., 2009a). Though some cytokines identified as active signalling loops such as IL6, M-CSF (CSF1) and GM-CSF (CSF2), are known to display such activities, the endogenous effects may be obscured by high levels of exogenously provided proliferative signals. It is notable that the antagonist SB-505124 displays SF3-type effects, consistent with the classification of TGF-β2 as an endogenous proliferation inhibitor.
Figure 3-10. Simulated activities of theoretical proliferation and self-renewal regulatory factors functionally classify experimentally identified ligands.

(a) Schematic diagram representing theoretical interactions between endogenous secreted factors and cellular kinetic parameters. The probability of HSC self-renewal and the proliferation rate of HSCs and progenitors are regulated by the balance of endogenous inhibitory (SF2, SF1) and stimulatory (SF4, SF3) signals. (b) Simulated dose-response relationships between theoretical factor concentrations (expressed as ED50 values) and 8-day fold total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) expansions, normalized to control culture output. Based on the relative effects on TNC, CFC, and LTCIC output, TGFB2 displays SF1-type activity, CCL3, CCL4, CXCL10, and TNFSF9 display SF2-type activity, and EGF PDGF, and VEGF display SF4-type activity (indicated by dashed lines).
3.4.7 Intra-cellular molecular network integration

Stem and progenitor cells exposed to combinatorial extracellular signals must propagate this information through intracellular molecular networks, and respond appropriately by modifying cell fate decisions (Figure 3-12a). While many genes have been demonstrated to significantly stimulate or repress blood stem cell self-renewal divisions in isolation, how these are integrated within the cell’s molecular circuitry remains to be resolved. To understand how the experimentally identified positive and negative regulatory signals are integrated at the intra-cellular level, we constructed a literature-curated blood stem cell self-renewal network as a platform for further analysis. Based on an extensive literature survey of PubMed, we compiled a list of 112 genes shown to modulate HSC self-renewal (in both human and/or mouse), as
measured by *in vivo* repopulating assays (see **Supplementary Table S28**). This set of genes was then searched against the i2D protein-protein interaction database (Brown and Jurisica, 2005) for binding partners, producing a densely connected network of 2131 vertices and 5490 (non-unique) edges (**Figure 3-13a**). The network was then filtered for active genes, consistently scored Present in culture-generated, self-renewing (d4 and d8-S/E) Lin⁻ cells (n = 1728).

To understand how the various ligands signal through this network, we compiled gene sets from PANTHER and KEGG databases comprising the experimentally defined stimulatory pathways [EGF (n = 150), PDGF (n = 189), and VEGF (n = 80)], and the experimental and literature defined inhibitory signalling pathways [chemokine (n = 193), FasL (n = 37), Ras (n = 92), and Insulin-IGF-MAPK (n = 50)]. From these, we identified 24 signalling molecules shared by all stimulatory pathways, and 15 signalling molecules shared by at least 3 of the inhibitory pathways. 16/24 stimulatory and 12/15 inhibitory common molecules were represented in the “active” self-renewal network, of which 5 molecules were common to both sets (**Supplementary Table S29**). A sub-network based on the first neighbours of the common signalling molecules consisting of 55 vertices and 138 (unique) interactions is shown in **Figure 3-12b**, capturing 41/107 self-renewal-associated represented in i2D, a level of enrichment corresponding to $P < 10^{-8}$.

Performing the same analysis on the common signalling molecules shared by the 5 neurological pathways associated with HSC expansion produces a sub-network which completely overlaps with the VEGF/EGF/PDGF sub-network (**Figure 3-13c**), thus linking neurological signalling pathways (and the functional effects of serotonin) with stimulatory growth factor signalling and self-renewal. Analysis of the TGFB signalling pathway components resulted in a very large sub-network due to an inherent bias in that a number of TGFB signalling components were used as baits in construction of the HSC self-renewal network.

To independently test both the utility, and *in vivo* biological relevance of the self-renewal network as an analytical platform, we queried for the presence of self-renewal effectors identified from a recent *in vivo* screen. Guy Sauvageau’s group has defined 16 novel nuclear
factors with HOXB4-like stimulatory activities on murine HSC self-renewal (Deneault et al., 2009). We searched 15/16 genes with human orthologs against the active network, identifying 6 (40%) (Figure 3-13d), a level of enrichment corresponding to $P < 10^{-6}$.

This analysis therefore demonstrates that integrating lists of literature-curated genes with protein interaction and pathway databases (i2D, PANTHER, and KEGG) can yield novel insights into the molecular mechanisms governing incompletely defined biological processes – in this case how a novel set of signalling molecules can modulate the self-renewal machinery of human blood stem cells. In particular, we have identified key molecular control points as targets for manipulation of self-renewal. However, the networks do not contain information on the directionality or functional significance of protein interactions. It is notable that most of the common signalling molecules interact with multiple functionally annotated genes, and vice versa, with no statistical bias between stimulatory/inhibitory pathways and genes. Due to this complexity, it is not possible to predict the functional effects of perturbing individual genes.
Figure 3-12. Integration of endogenous regulatory signals in the HSC intra-cellular self-renewal network.

(a) Endogenous secreted stimulators (VEGF, EGF, PDGF) and inhibitors (CCL3, CCL4, CXCL10, TNFSF9, and TGFβ2) activate cell surface receptors on HSCs, inducing signal transduction events which are coherently processed by the intra-cellular network to modulate rates of self-renewal vs. differentiation.  

(b) Common signal transduction molecules shared by stimulatory pathways (left; green box), inhibitory pathways (right; red box), and both (centre) are densely connected to known self-renewal effector genes. Physical protein-protein interactions from stimulatory and inhibitory pathways are represented as green and red edges respectively, while internal interactions are represented as blue edges.  

(c) Five small molecule antagonists, described in the table with targets indicated on the network, were tested for functional effects on 8-day fold expansions of total cells (TNC), progenitors (LTCIC), and primitive progenitors (LTCIC) with respect to control cultures.  

(d) To classify the functional activities of the molecules, culture simulations were run over a feasible range of HSC self-renewal probabilities and proliferation rates. Based on the relative effects on TNC, CFC, and LTCIC output, inhibition of PI3-K, PLC, and MEK-1 reduces self-renewal, while inhibition of Raf and Akt reduces proliferation.
Figure 3-13. Reconstructed intra-cellular self-renewal signalling network.
(a) A curated list of 112 genes known to effect HSC self-renewal was searched against the i2D protein interaction database for binding partners, resulting in a densely connected network of 2131 vertices and 5431 (non-unique) edges. For clearer visualization of the individual genes, the network was filtered to display only direct physical interactions between the self-renewal effectors, resulting in 104 genes connected through 180 (unique) edges (b). Sub-networks constructed from the first-neighbours shared neurotransmitter signalling molecules (c) and self-renewal enhancing nuclear factors reported in Deneault et al. (2009) (d) active in culture are highly enriched for self-renewal-associated genes.

3.4.8 Targeted intra-cellular molecular network perturbation

We finally wished to test our computational predictions for conserved intra-cellular signalling molecules involved in self-renewal divisions. We obtained 5 small molecule antagonists targeting 10 (non-unique) proteins represented in the self-renewal sub-network, as indicated in Figure 3-12b and described in the adjacent table. Liquid cultures were supplemented with the 5 molecules, and total cell (TNC), progenitor (CFC), and primitive progenitor (LTCIC) expansions were compared to control values. All 5 antagonists displayed cell population-specific inhibitory effects to varying degrees (Figure 3-12c). As the targets may be involved in multiple cellular processes, we utilized our mathematical model to de-convolute the functional effects on proliferation (and survival) vs. self-renewal. We simulated liquid progenitor cultures over a range of HSC self-renewal probabilities (0 to 100%) and proliferation rates (0 to 4 day⁻¹), and compared TNC, CFC, and LTCIC outputs to our control results (Figure 3-12d). Note that inhibiting self-renewal has a more pronounced effect on progenitors (CFC and LTCIC) compared to total cell (TNC) output, while inhibiting proliferation has a very similar effect on all three cell populations. By overlaying the experimental data from Figure 3-12b it is apparent that inhibition of PI3-K, PLC, and MEK signalling is consistent with specific targeting of self-renewal, while inhibition of Raf and Akt signalling is consistent with specific targeting of proliferation (or possibly survival as the two parameters are lumped). We are currently broadening our small molecule screens, including antagonists of JNK, PKC, and STAT signalling, in an attempt to identify small molecules which stimulate as well as repress self-
renewal. While still underway, this work indicates that the endogenous secreted regulators identified appear to be acting directly on primitive progenitors and stem cells, and highlights the ability of our computational methods to identify molecular targets from publicly available databases.

### 3.5 DISCUSSION

To date, dynamic interactions between hierarchically related cell populations have not been experimentally interrogated. Using *in vitro* hematopoiesis as a model system, we have integrated genome-scale molecular profiling technologies, biological interaction databases, extensive literature curation, and mathematical modelling to identify the soluble factor-mediated intercellular networks regulating blood stem cell fate decisions. These can broadly be reduced to an antagonistic positive-negative feedback circuit wherein progenitor expansion is modulated by a balance of megakaryocyte-derived stimulatory factors (EGF, PDGF, VEGF, and possibly serotonin) vs. monocyte-derived inhibitory factors (CCL3, CCL4, CXCL10, TGFβ2, and TNFSF9) (*Figure 3-14*). This complex milieu of endogenous regulatory signals is integrated and processed within a core intra-cellular signalling network, resulting in modulation of cell-level kinetic parameters (proliferation rates and self-renewal probabilities).

![Figure 3-14. Schematic summary of experimental findings.](image)
In addition to exogenous growth factors (FLT3LG, KITL, and THPO in our cultures), stem cell output is regulated by soluble factor mediated cell-cell interactions. We define an antagonistic axis of inter-cellular communication, wherein culture-derived monocytes secrete high levels of factors (CCL3, CCL4, CXCL10, TGFB2, and TNSFS9) which inhibit stem cell self-renewal, while culture-derived megakaryocytes secrete high levels of factors (EGF, PDGF, VEGF) which stimulate stem cell self-renewal, functioning as coupled positive and negative feedback circuits.

Importantly, we did not set out to measure gene expression profiles of HSCs, as has been done extensively by others (Joshi and Enver, 2003), but rather heterogeneous populations (Lin⁻ and Lin⁺) comprising the HSC cellular microenvironment. Our results suggest new directions to modulate the stem cell niche in vitro and in vivo. It has previously been shown that long-term repopulation activity (LT-SRC) is confined to VEGFR1⁺ sub-population of CD34⁺ cells, and that VEGFR1 antibody administration, while having no effect on steady state hematopoiesis, suppresses HSC cycling and hematopoietic regeneration following 5-FU treatment (Hattori et al., 2002). Similarly, VEGFR2 antibody administration or conditional gene knock-out inhibits hematopoietic regeneration following sub-lethal irradiation (Hooper et al., 2009). Combined with the fact that VEGFA is up-regulated in the bone marrow micro-environment during HSC regeneration, and conditional knock-out of this gene leads to hematopoietic failure (Gerber et al., 2002), there is significant evidence that VEGF signalling is required for inducing HSC expansion in vivo. In fact, the stimulatory effects THPO, widely used in HSC expansion media formulations and necessary for self-renewal in vivo (Qian et al., 2007), is mediated in part through the induction of VEGFA secretion (Kirito et al., 2005).

PDGF, in combination with other stimulatory cytokines, has been shown to enhance the expansion of hematopoietic stem and progenitors in vitro (Su et al., 2002), and endogenous PDGF signalling, in combination with VEGF, is required for maintenance of long-term bone marrow cultures (Duhrsen et al., 2001). However, there exists no published data on the effects of EGF signalling on hematopoietic stem and progenitor cell fate in vitro or in vivo. It is notable that while there exists little data on the effects of these three ligands on HSC fate decisions, these are all potent mitogens for other diverse cell types, are known to play roles in the maintenance
and growth of various neoplasms, and induce downstream signal transduction events (MAPK/ERK, PLC, and PI3K pathways) in common with the potent self renewal stimulatory factor KITL (Kent et al., 2008a). It is therefore not completely unexpected that these ligands function as endogenous positive regulators of stem cell expansion. However, it is still not conclusive whether the functional effects of these ligands on primitive progenitor growth are mediated directly or indirectly via secondary cell populations.

The bone marrow microenvironment is heavily innervated, and it has recently become apparent that complex cross-talk between nervous, bone, blood cells regulates stem cell fate in vivo (Spiegel et al., 2008). Consistent with our results, the neurotransmitters dopamine and serotonin have both been shown enhance HSC proliferation and engraftment (Spiegel et al., 2007; Yang et al., 2007). We thus extend these findings to suggest that these stimulatory signals are endogenously produced by blood cells, and induce downstream signalling events which converge on the core self-renewal machinery.

There exists a wealth of literature on the potent inhibitory effects of TGF-β ligands on HSCs cycling in vitro and in vivo (Larsson and Karlsson, 2005), however the effects on self-renewal are less clear due to the experimental difficulties in uncoupling the two processes. While TNFSF2 (TNF-α) is a well established inhibitor of HSC self-renewal (Dybedal et al., 2001), there exists no published data on the effects of other TNF family members, including TNFSF9. Many chemokines have been screened for effects on myeloid colony growth in vitro (Broxmeyer and Kim, 1999). Consistent with our findings, inhibitory effects were reported for CCL3 (MIP-1α), while IL8, CCL5 (RANTES), and CXCL7 (NAP-2) are reported to have no effects on colony growth. We have additionally identified CCL4 (MIP-1β), and CXCL10 (IP-10) as specific inhibitors of LTCIC expansion, indicating effects on self-renewal.

Our results provide insight into direct vs. indirect effects of exogenous regulators. It is very difficult (if not impossible) to differentiate between direct vs. indirect effects when relying on the
use of retrospective colony assays as functional readouts, as these measure proliferation and
differentiation of heterogeneous, evolving cell populations. For example, PDGF is known to act
in a positive feedback signal during megakaryopoiesis, secreted by platelets and stimulating the
proliferation of megakaryocyte progenitors (Su et al., 2001). The stimulatory effects of PDGF
may thus be mediated indirectly through enhanced megakaryocyte production and associated
growth factor secretion [such as VEGF (Mohle et al., 1997)], or via IL-1β induction from
monocytes (Yan et al., 1993). The effects of exogenous ligands are thus mediated through
complex and dynamic interactions with various cell populations and endogenous secreted ligands
(Koller et al., 1995). In fact, the results of recent screen for nuclear factors stimulating HSC self-
renewal found that 4/10 identified potent regulators acted through non-cell autonomous
mechanisms (Deneault et al., 2009). This is a commonly neglected to complexity of genetic
studies – inter-cellular signalling between hematopoietic cell sub-populations obscures direct vs.
indirect effects. To illustrate the complexity of the intra-cellular signalling networks established
in vitro, we reconstructed a functional interaction network between the 77 secreted ligands listed
in Supplementary Table S27, defining a densely connected network integrating various
stimulatory, inhibitory, and exogenously supplemented ligands (Figure 3-15).
Figure 3-15. Functional Cytokine Interaction Network.

The integrated directed-undirected graph depicts functional and physical interactions between The 77 ligands used to define the inter-cellular network, identified from literature searches of PubMed and the Cytokines Online Pathfinder Encyclopaedia (COPE) as occurring in the hematopoietic cell populations generated \textit{in vitro}. The top ranking module as defined using the Cytoscape MCODE algorithm consists of densely inter-connected stimulatory, inhibitory, and exogenously supplemented ligands.
One of the most fundamental questions in cell biology is how are cells capable of converting complex information about their microenvironment into logical cell fate decisions? The functional response of a cell population to extracellular signals is highly context-dependant, as cytokines display non-linear combinatorial interactions in vitro (Natarajan et al., 2006) and in vivo (Roeder et al., 1998b). However, the combinatorial possibilities of external signals generally converge on a limited set of coherent cellular responses. It is empirically established that the induction of self-renewal divisions requires signalling through multiple pathways, and pathways associated with self-renewal such as Wnt, Notch, and Prostaglandin are tightly integrated (Duncan et al., 2005; Goessling et al., 2009). Indeed, enforced expression of HOXB4 (a potent self-renewal inducing transcription factor) modulates Wnt, Notch, Hedgehog, FGF, TGF-β, GPCR, TNF-α, cell cycle, and apoptosis-associated genes, which essentially encompasses every signalling pathway implicated in HSC self-renewal (Schiedlmeier et al., 2007). This lead us to ask whether the various pathways regulating self-renewal divisions converge on a limited set of key nodes, and whether we could utilize available databases to identify these.

From this framework, we attempted to systematically integrate the wealth of published literature on the genetic regulation of HSC self-renewal using protein-protein interaction databases to construct an intra-cellular self-renewal signalling network. Using this network as a platform for further interrogation, we found that stimulatory and inhibitory signalling pathways converge on a core set of molecules integrated within this network, and targeted inhibition of these molecules can be used to specifically control self-renewal. This work further demonstrates that the classic notion of linear, independent signalling pathways is a gross oversimplification, as pathway “cross-talk” actually dominates the molecular response as signals are propagated from the cell surface to the nucleus.

We can speculate on the biological logic underlying the topology of the inter-cellular singling network controlling stem cell self-renewal. It may be related to cellular localization –
megakaryocytes are reported to play an integral part of the HSC niche via stimulating proliferation and differentiation of osteoblasts (Kacena et al., 2006), while monocytes, macrophages, granulocytes, and erythrocytes, comprising a large proportion of peripheral blood, may serve as sentinel markers of systemic cell density. Further studies are thus required to elucidate the full spectrum of cellular interactions active \textit{in vivo}, including lymphocyte and stromal cell interactions.

The most immediate practical utility of these findings lay in designing strategies for the expansion of HSCs for clinical use. An ideal culture media may consist of high-level supplementation with multiple stimulatory factors (EGF, PDGFB, VEGF, and/or serotonin) plus small molecule or antibody-based antagonists of the inhibitory factors (CCL3, CCL4, CXCL10, TGFB2, and TNFSF9). Combined with appropriate cell sub-population selection and media dilution rates, we thus have a foundation to systematically manipulate inter-cellular signalling dynamics and thereby control HSC self renewal \textit{in vitro}. While we have tested and validated a limited number of intra-cellular targets using small molecules, the network will serve as a predictive platform for future (high-throughput) screens for modulators of self-renewal. Due to the high dimensional complexity and redundancy of extracellular signals, we predict that targeting key intracellular control points will prove to be a better strategy for inducing robust HSC expansion.

These findings also have broader implications for stem cell biology, emphasizing the importance of interpreting cell fate decisions in context of interconnected, hierarchical tissue networks. Regulatory effects are often mediated indirectly, and conversely, cell fate decisions can be systematically manipulated via non-autonomous parameters. This work further proves the utility of complementing experimental data with computational analyses to provide deeper insight into mechanisms underlying complex cellular systems.
4 CONCLUSIONS & FUTURE WORK
4.1 Summary of results

Communication networks between cells, tissues, and organ systems are necessary for homeostasis in multicellular organisms. Inter-cellular communication networks are particularly relevant in stem cell biology, as stem cell fate decisions (self-renewal, proliferation, lineage specification) are tightly regulated based on physiological demand and responsive to external perturbations. Hematopoiesis, the process by which blood cells develop, serves as a prototype for other stem cell systems. Motivated by the fact that hematopoietic stem cell transplantation is a curative therapy for a number of hematopoietic and immunological diseases, herein we explore the behaviour of inter-cellular regulatory networks as tools to regulate cell fate during in vitro human blood stem cell propagation.

We have integrated a diverse range of experimental and theoretical literature to develop for predictive purposes a novel mathematical model of hematopoiesis as described in Chapter 2. Our resulting model links functional cellular assays to specific model outputs, defines cell-level kinetic parameters such as cell cycle rates and self-renewal probabilities as functions of culture variables, and simulates feedback regulation using cell-cell interaction networks. Computational analyses of the system dynamics indicate that non-cell autonomous parameters (cell-cell interactions) are dominant factors controlling stem cell growth. As a result of negative feedback interactions, our analyses suggest an antagonistic relationship between mature and primitive cell compartments, a finding empirically supported by numerous in vitro and in vivo observations.

To investigate the practical utility of the model, we cultured human umbilical cord blood stem cells under a range of conditions and dynamic perturbations (media exchange and mature cell depletion), and compared experimentally observed cell population outputs to model simulations. Our model simulations quantitatively recapitulate experimental results; culturing enriched progenitor populations at low initial cell densities, with frequent media exchange, and with progenitor re-selection enhances stem and progenitor expansion as a consequence of reduced inhibitory feedback signalling. Using a protein array system, we identify a limited number of secreted molecules (TGF-β1, CCL4, CCL5, and CXCL8) displaying dynamic profiles consistent
with predicted inter-cellular regulators. Additionally, we show the variability in secretion rates observed for these putative regulators is sufficient to explain the experimentally observed distribution of cell population outputs.

Computational analyses of the model revealed that system dynamics are relatively robust to changes in kinetic parameters, but highly sensitive to topological alterations of the proposed cell-cell regulatory network. We extend this concept by demonstrating that deleting a single regulatory connection (negative feedback control of self-renewal) is necessary and sufficient to reproduce the characteristic features of in vitro leukemic transformation.

While our model is capable of predictively simulating cell population dynamics in vitro, the secreted molecules mediating the inter-cellular communication (SF1-4) represent theoretical factors, most likely classes of functionally related molecules. Additionally, the model sheds no light on the intra-cellular molecular networks processing these extra-cellular signals into cell fate decisions. This motivated the work presented in Chapter 3 with the ultimate goal of reverse engineering the inter-cellular signalling networks via a "bottom-up" high-throughput profiling strategy.

UCB progenitors and differentiated cells were transcriptionally profiled during culture under stem cell supportive vs. non-supportive conditions, in parallel with functional and phenotypic analyses, and sub-proteome profiling of conditioned media. Through the use of literature-derived cell population-characteristic gene sets, pathway, and biological process databases we identified signal transduction activity, cell populations, and endogenous secreted factors differentially associated with stem cell expansion and depletion in culture. Using a manually curated set of ligand-receptor pairs and integration with published datasets, we additionally used our gene expression data to reconstruct a network of cell-cytokine interactions between progenitors (Lin-) and differentiated populations (megakaryocytes, monocytes, and erythrocytes) generated in culture. This analysis indicates that dynamic interactions between positive and
negative regulators, in the context of tuneable cell culture parameters, may tip the balance between stem cell supportive vs. non-supportive conditions.

To test the functional significance of the correlative gene expression analyses, UCB progenitors were cultured in the presence of putative endogenous regulatory ligands, or in vitro-derived differentiated cells. We find that patterns of differential expression predict, with statistical significance, functional effects on primitive progenitor (Long Term Culture-Initiating Cells; LTCIC) output in culture. Specifically, production of megakaryocytes enhances, while production of monocytes inhibits stem cell expansion in vitro in a non-autonomous manner via cell population-specific secretome profiles. Using our recently published mathematical model of hematopoiesis to interrogate the mechanisms of cytokine activity, we were able to functionally classify endogenous ligands into distinct categories – self-renewal stimulators (EGF, PDGF, VEGF, serotonin), self-renewal inhibitors (CCL3, CCL4, CXCL10, TNFSF9), and proliferation inhibitors (TGFB2).

To understand how the experimentally identified positive and negative regulatory signals are integrated at the intra-cellular level, we constructed a blood stem cell self-renewal network, based on an extensive literature survey and protein interaction mapping, as a platform for further analysis. We found that the positive and negative regulators activate signal transduction cascades which converge on a limited common set of signalling molecules. By targeted inhibition of these molecular control points using small molecule antagonists, in conjunction with computational analysis of our model to de-convolute the functional effects, we define PI3-K, PLC, and MEK signalling as required for self-renewal, while Raf and Akt signalling as required for proliferation.

In summary, we have integrated genome-scale molecular profiling technologies, biological interaction databases, extensive literature curation, and mathematical modelling to identify the soluble factor-mediated inter-cellular networks regulating blood stem cell fate decisions. These
can broadly be reduced to an antagonistic positive-negative feedback circuit wherein stem cell supportive environments are regulated by a balance of megakaryocyte-derived stimulatory factors vs. monocyte-derived inhibitory factors. This complex milieu of endogenous regulatory signals is integrated and processed within a core intra-cellular signalling network, resulting in the modulation of cell-level kinetic parameters (proliferation rates and self-renewal probabilities). This represents, to our knowledge, the first attempt to comprehensively elucidate non-autonomous signals balancing stem cell homeostasis and regeneration.

4.2 Future computational work

4.2.1 Evaluation of model assumptions

Numerous simplifying assumptions must be made during the process of model development, in order to produce a tractable system of equations. It is generally a useful activity to systematically evaluate the effect each underlying assumption has on the model results in order to quantitatively test model robustness or sensitivity to the underlying assumptions. As a starting point, herein we consider the simplifying assumption of complete symmetric differentiation divisions in vivo \(X_{i+1} = 2X_i\). This assumption was required to estimate the total number of compartments, and assign compartment – functional assay relationships in lieu of experimental data on the probabilities of self-renewal in vivo. The total number of cells generated, on average, from a stem cell is given by the equation:

\[
X_T = 2^n \times (1 - f_1) \times (1 - f_2) \times ... (1 - f_n) \tag{1}
\]

Which can be expressed as:

\[
X_T = 2^n \prod_{i=1}^{n} (1 - f_i) \tag{2}
\]

Hence the total number of compartments \(n\) required to generate \(X_T\) total cells from a stem cell is given by:

\[
n = \frac{\ln(X_T) - \ln(\prod_{i=1}^{n} (1 - f_i))}{\ln(2)} \tag{3}
\]
Thus posing a potentially intractable problem, as the self-renewal probabilities ($f_i$) for each compartment cannot be measured, and are in fact dynamic parameters regulated by exogenous and endogenous signals. Assuming that the cellular hierarchy of freshly isolated umbilical cord blood cells results from symmetrically organized differentiation divisions ($f_i = 0$ for $i = [1, 2, ..., n]$), equation (3) then reduces to:

$$n = \frac{\ln(X_T)}{\ln(2)} \quad (4)$$

The number of compartments ($n$) comprising the functional readouts listed in Table 2-1 [total cells (TNC) undifferentiated (Lin-) cells, progenitors (CFC), primitive progenitors (LTCIC) and stem cells (SRC)] can then be calculated based on their respective frequencies in freshly isolated umbilical cord blood. Based on this assumption, we used in vitro culture data and the reverse engineering strategy described in chapter 2 via equations (2-18) through (2-24) and associated text to estimate the probability of self-renewal ($f_i$) for each compartment ($i$) in our in vitro culture system using:

$$f_i = f_{MAX} \exp\left[-\frac{(1-i)^2}{2D_{SR}^2}\right] \quad (5)$$

Where $f_{MAX}$ and $D_{SR}$ were calculated to be $0.63 \pm 0.09$ and $2.0 \pm 0.7$ respectively, results specific to the culture conditions described.

The parameter estimates and resulting model simulations may, however be dependent upon the initial assumption of complete symmetric differentiation in the freshly isolated cells. To explore how this assumption affects our results, we alternatively assumed that the probabilities of self-renewal for freshly isolated cells could be variable and specified using equation (5); simulations were then run with the maximal probability of self-renewal ($f_{MAX}$) ranging from 0 to 99% (Figure 4-1a). Modifying the self-renewal divisions in this manner has the effect of adding additional compartments ($n_x$) at the base of the hierarchy:

$$n_x = -\frac{\ln(\prod_{i=1}^{N}(1-f_i))}{\ln(2)} \quad (6)$$
The additional compartments \( (n_x) \) required to account for the cellular composition of umbilical cord blood as a function of the maximum probability of self-renewal is depicted in Figure 4-1b, corresponding to \( n = 20 \) for \( f_{\text{MAX}} = 0 \), to \( n = 32 \) for \( f_{\text{MAX}} = 99\% \).

To test how the additional compartments (resulting from changes in \( f_{\text{MAX}} \)) affect our results, we simulated 8-day un-manipulated (NS/NE) cultures of Lin\(^-\) umbilical cord blood cells for initial assumptions of \( f_{\text{MAX}} \) from 0 to 99\%. As the model parameters depicted in Table 2-2 are dependent upon the initial assumption, a generalized pattern search algorithm (psearchtool) was used to adjust parameter values within 95\% confidence intervals based on the objective function defined in equation (2-24). The resulting proportion of differentiated (\%Lin\(^+\)) cells, and fold-expansion of total cells (TNC), progenitors (CFC), and primitive progenitors (LTCIC) are depicted in Figures 4-1c to d. The experimental means ± standard deviations are overlaid on the figures as solid and dashed lines respectively, allowing visual comparison of simulated to experimental results.
Figure 4-1. Effect on model simulations of varying the assumed probability of self-renewal in fresh umbilical cord blood cells.
(a) The probability of self-renewal \((f)\) as a function of compartment number \((i)\) for \(f_{MAX}\) varied from 0% to 99%. (b) Additional compartments \(\left(n_x\right)\) required to account for the cellular composition of umbilical cord blood as a function of the maximum probability of self-renewal \(f_{MAX}\) assumed for freshly isolated cells. Proportion of differentiated \((%Lin^+\)) cells \((c)\), and fold expansions of total cells \((TNC)\) \((d)\), progenitors \((CFC)\) \((e)\), and primitive progenitors \((LTCIC)\) \((f)\) from 8-day culture simulations for initial assumptions of \(f_{MAX}\) from 0 to 99% (points) vs. experimental data ± standard deviation (lines).

Importantly, this analysis demonstrates that model outputs generated based on initial assumptions of \(f_{MAX}\) from 0 to 80% (corresponding to \(n = 20\) to 26) are equally capable of recapitulating the experimental data, however model outputs derived from assumptions of \(f_{MAX} \geq 90\%\) and (corresponding to \(n \geq 28\)) are incapable of recapitulating our experimental outputs for the later cell compartments \((Lin^+, TNC,\) and \(CFC)\). The fidelity of the \(LTCIC\) results lies in their relative isolation from the later cell compartments; additional early stage divisions essentially increase the delay time for differentiation, thereby reducing the growth of later compartments at day-8 of culture; longer term (greater than 12 day) simulations demonstrate compartment size effects on LTC-IC and SRC (not shown). Thus, assumptions of the in vivo probability of self-renewal \((f_{MAX})\) for stem cells, required as an initial step for defining the compartment-functional assay relationships, do not significantly affect resulting simulations for physiologically reasonable self-renewal probabilities ranging of 0 to 80% (Ando et al., 2006; Ema et al., 2000; Glimm and Eaves, 1999). As this parameter is experimentally hidden, the simplest assumption of 0% is acceptable.

Similar analysis may be performed on other simplifying assumptions, for example the use of Hill-type functions to quantify dose response relationships between kinetic parameters and secreted factors, and Gaussian functions to quantify relationships between kinetic parameters and compartment number. These functions were chosen for their mathematical simplicity and malleability rather than biological evidence, however other functions (i.e. sigmoidal, exponential, log-normal, etc...) may better reproduce the experimental data.
4.2.2 Evaluation of alternative models

The inter-cellular signalling network topology presented in chapter 2 is capable of capturing much of the system dynamics, however it is most likely a simplification of the actual complexity. While we have tested the model performance against a limited number (9) of systematically altered network topologies, extending this approach should yield more realistic and better performing models, as has been demonstrated for models of signal transduction (Kuepfer et al., 2007) and fetal erythropoiesis (Socolovsky et al., 2007). The general approach is as follows:

1. Alter the model structure.
2. Estimate model parameters using global optimization methods based on available experimental data (Banga, 2008).
3. Compare alternative models outputs using a quantifiable metric such as the Akaike Information Criterion (AIC) (Landaw and DiStefano, 1984), and perhaps including more complex measures such as system robustness (Kitano, 2004).

In terms of the network topology, the inclusion of autocrine stimulatory and inhibitory signalling loops, both at the level of stem cells and more differentiated progeny, as well as feed-forward loops should eventually be tested due to the biological relevance (Torok-Storb, 1988). However, an intrinsic balance exists between producing a more biologically realistic model, and computational limits on complexity. Considering the 20 compartments in the hierarchy, systematically testing the presence vs. absence of stimulatory and inhibitory connections between each of the compartments (let alone secreted factor regulatory interactions) would result in $41^{20} \approx 1.8 \times 10^{32}$ alternate topologies. The combinatorial explosion of possibilities must therefore be limited by experimental evidence and reasonable assumptions (i.e. functionally grouping compartments). That being said, there is still room for exploring additional topologies as experimental data sets on the system dynamics continued to grow. The number of free model parameters is limited to the number of experimental readouts to avoid over-fitting the system – hence the number of free parameters (and resulting model complexity) may increase as the experimental data training set expands (Ashyraliyev et al., 2009; Doyle and Stelling, 2006).
Ultimately, model development is an iterative process consisting of (1) verification (determining network connectivity); (2) calibration (estimate free parameters); and (3) validation (evaluate model performance against new experimental data) (Aldridge et al., 2006; Gadkar et al., 2005). Systematically testing the combinatorial possibilities of model topologies in an efficient manner will require parallel, high performance computing (Ballarini et al., 2009). Such analysis is in fact ideally suited to distributed computing, as parameter estimation algorithms and performance metrics are performed on each of the alternative model topologies independently, followed by model ranking and selection.

A biologically relevant additional complexity to consider is implementation of regulatory protein secretion rates and/or dose response characteristics as gradient functions of differentiation status. Rather than simply assigning secretion and response functions to (somewhat) arbitrarily defined compartments, perhaps simple Gaussian, exponential, or Hill-type functions be used to define more biologically relevant characteristics, as was implemented for self-renewal probabilities ($f_{\text{MAX}}$) and proliferation rates ($u_{\text{MAX}}$) as functions of compartment number ($n$).

A consideration for future analysis is the implementation of (at least partially) stochastic and/or discrete rather than purely deterministic, continuous terms. Cells (and cell fate decisions) are ultimately discrete entities, regulated by stochastic biochemical processes. For large cell numbers with limited compositional heterogeneity, continuous differential equation-based models are appropriate simplifications. However, when dealing with limited cell numbers and/or rare cell populations (such as HSCs), stochastic and discrete effects may play important roles in system dynamics (Aldridge et al., 2006). For these reasons, many groups have developed stochastic models of haematopoiesis – defining cell compartment transitions via probability distributions rather than kinetic rate constants (Loeffler and Roeder, 2004). Stochastic effects may in fact play a role in the wide sample-to-sample availability in culture outputs, and other features insufficiently accounted for by the model in its current form. Importantly, hybrid stochastic/deterministic switching algorithms are available which automatically alternate
between computational methods based on a pre-set concentration threshold (i.e. 20 cells) (Hoops et al., 2006). Other discrete modelling approaches based on cellular interactions within populations, such as agent-based (Chavali et al., 2008), game theoretic (Nowak and Sigmund, 2004), or hybrid models (Walker et al., 2008) may yield important information on cell culture dynamics which cannot be accounted for with differential equations. Ultimately, it is very difficult to predict which alternate modelling approaches are most relevant, and practically impossible to predict what can be learned about the system properties from the different approaches without going forward and carrying out the work.

For modelling in general to be appreciated by a wider, non-mathematically inclined audience, and for the utility of this particular model to be fully exploited, it will be necessary to make models accessible in a user-friendly format. The current industry standard computing language is Systems Biology Markup Language (SBML) (Hucka et al., 2003) – while developed specifically for biochemical reaction networks, it could easily be appropriated to represent inter-cellular networks. Many free, user-friendly software tools are available to construct (Kitano et al., 2005; Rodriguez et al., 2007) and analyze SBML-based models (Evans et al., 2008; Zi and Klipp, 2006; Zi et al., 2008), hence conversion of the current MATLAB code to this format would allow one to exploit these accessible tools. Importantly, toolboxes are available which interconvert MATLAB and SBML code (Keating et al., 2006; Schmidt et al., 2007), however, SBML files are required as the original source.

The BioModels database is emerging as the industry standard biological model repository (Le Novere et al., 2006) (http://www.ebi.ac.uk/biomodels-main/) - uploading an SBML-formatted file would make the model freely accessible for use by others. While still in its relevant infancy (containing 217 models as of July 2009), the database may soon play a role similar to that of Genbank (Benson et al., 2009) (http://www.ncbi.nlm.nih.gov/) for sequence data, or Gene Expression Omnibus (Barrett et al., 2007) (http://www.ncbi.nlm.nih.gov/geo/) for gene expression data, as a standard requirement for publication and searchable resource for the dissemination of biological knowledge. For more customized and widely accessible sharing, a Graphical User Interface (GUI) could be created to run the model simulations. Implementing
adjustable model parameters and culture variables as dials and switches would allow users to intuitively explore culture dynamics \textit{in silico} without requiring extensive mathematical training.

\section*{4.3 Future model applications}

The utility of the model ultimately lies in optimizing culture operating conditions. As non-stem cell autonomous parameters (cell-cell interactions) were identified as key regulators of stem and progenitor cell fate, the model could be used as a tool to develop media dilution strategies control cell-cell signalling during culture. Media dilution strategies can be broadly classified into two categories; (1) Fed-batch, wherein fresh media is fed into the culture (either continuously or periodically) without removal of culture media, thereby increasing the culture volume over time, and (2) Perfusion, wherein an equivalent volume of culture media is removed per volume fresh media added, thereby maintaining the culture at a constant volume. Furthermore, the media dilution rate \([D = \text{flow rate} (F) / \text{volume} (V)]\) may be implemented along fixed- or adaptive-trajectories. For fixed-trajectories, the dilution rate may be constant [expressed as a fraction of the initial culture volume; \(D(t) = F/V_0\)] or proportional, adjusted in proportion to the current culture volume [\(D(t) \sim V(t)\)].

To directly compare the 3 non-adaptive feeding strategies described (fed-batch with constant dilution, fed-batch with proportional dilution, and perfusion) 8-day cultures of progenitor enriched (Lin') umbilical cord blood cells were simulated under each condition. We then compared the \textit{in silico} growth rate of the measurable cell compartments comprising the hematopoietic hierarchy as a function of dilution rate (Figure 4-2). Dilution rates were directly compared by normalizing to the total volume of media required \((V_T)\) throughout the 8-day cultures, initiated at 1 ml. The growth rate response of total nucleated cells \([TNC (a)]\), progenitors \([CFC (b)]\), primitive progenitors \([LTCIC (c)]\), and stem cell \([SRC (d)]\) is consistent – population expansions are enhanced at increasing dilution rates, and fed-batch operation with constant media dilution is the most effective strategy in all cases. The mechanism underlying the culture performance is apparent from the final (day-8) media concentrations of the soluble proliferation \([SF1 (e)]\) and self-renewal \([SF2 (e)]\) inhibitors. Increasing dilution rates reduces the
concentration of endogenous inhibitory factors, and fed-batch culture operation at constant dilution rate results in lower concentrations of the inhibitors compared to the two other feeding strategies, particularly with respect to the self-renewal inhibitor \( SF_2 \) for total media volumes > 20 ml (equivalent to \( D > 2.5 \text{ day}^{-1} \)). It is notable that empirically, fed-batch (rather than perfusion or static culture) culture operation is widely accepted as the optimal used feeding strategy for biopharmaceutical production (Birch and Racher, 2006).
Figure 4-2. Batch-Fed & Perfusion culture outputs as a function of dilution rate.

8-day Fold expansions of TNC (a), CFC (b), LTCIC (c), SRC (d), and final (day-8) media concentrations of theoretical proliferation inhibitor $SF1$ (e) and self-renewal inhibitor $SF2$ (f) as functions of media dilution rate, normalized to total volume ($V_T$) of media required.
Building upon predictions that fed-batch media delivery strategies outperform perfusion, the model may further be utilized to design adaptive-feedback controlled fed-batch culture strategies. During adaptive feedback control, key culture parameters are maintained at a desired level by continuously adjusting external culture variables (Portner et al., 2004). In this case the external culture variable is the fresh media dilution rate, and the culture parameters to be controlled could be the concentration of the proliferation inhibitor \( SF1 \), the self-renewal inhibitor \( SF2 \), or cell population density, schematically represented in Figure 4-3.

**Figure 4-3. Adaptive feedback-controlled bioreactor.**

Fresh media is supplied to the cell culture through a computer-controlled pump. The media feeding rate is continuously adjusted as a function of culture composition, such as the concentration of an endogenously secreted molecule, or cell population density.

To compare the effectiveness of feedback control based on concentrations of the theoretical inhibitors \( SF1, SF2 \), or total cell \( TNC \) density *in silico*, simulated culture results for these three control strategies are depicted in Figure 4-4. Simple proportional feedback control algorithms were utilized, wherein:

\[
\text{IF } x \leq x^* : \quad D = 0 \\
\text{ELSE: } \quad D = K_C (x - x^*)
\]  

(7)
Where $D = \text{dilution rate (day}^{-1})$, $K_C = \text{Proportional gain}$, $x = \text{control variable}$, and $x^* = \text{control variable set point}$. Dynamic concentration profiles for $SF1$ ($a_1$), $SF2$ ($b_1$), and $TNC$ density ($c_1$) over a range of controller set-points (blue = highest set-point, to red = lowest set point) demonstrate that the proportional control algorithms implemented work reasonably well in maintaining the cultures at steady state with respect to the control objectives. Results may be improved by incorporating integral and/or derivative error terms (Oggunaike and Ray, 1994).

The corresponding 8-day cell population expansions for the three respective control strategies ($a_1i$, $b_1i$, and $c_1i$) as a function of the controller set-points predict that maintaining lower set points for all three control variables (not-surprisingly) results in improved culture outputs, and that controlling cultures based on cell density is in fact the most effective strategy.

Measuring cell densities and specific protein concentrations could be conducted relatively easily off-line using current flow cytometric-based liquid chip technology (Morgan et al., 2004). The sample volumes ($< 100\mu\text{L}$), and manual preparation required would limit sampling frequency at most a few times/day. Based on the cell culture kinetics (i.e. average TNC doubling time $\leq 1$ day), such frequencies would be sufficient to implement open-loop feedback optimal (OLFO) algorithms to control the media dilution trajectory (Becker et al., 2007). Future implementation of advanced optical sensor technologies in development may circumvent the need for culture sampling and allow for real-time culture monitoring and control (Ulber et al., 2003). Additionally, the use of multiplex measurements (i.e. secretome profiles) may allow for more complex, multivariate process control techniques (Junker and Wang, 2006). Ultimately, results from bioreactor optimization and control experiments can be utilized as the data training set for the next generation of model verification – validation – calibration studies.
Figure 4-4. Adaptive feedback controlled fed-batch cultures.

Media dilution rates \( (D) \) were adapted to maintain concentrations of theoretical proliferation \([SF1](a)\), self-renewal \([SF2](b)\) inhibitors, or total cell density \((c)\). Phase portraits (i) depict dynamic profiles of the control variables over a range of settings, and the corresponding fold-expansion of the various cell populations as a function of controller setting are shown in (ii).
4.4 Future inter-cellular network analysis work

A limitation of the reconstructed inter-cellular network in its current form is the granularity - Lin⁻ and Lin⁺ cells both comprise functionally heterogeneous populations. While no definitive markers exist for very primitive blood progenitors, particularly in culture, sub-fractionating the Lin⁻ population via CD133, ALDH, and CD38 expression (Hess et al., 2006; Terstappen et al., 1991) could further elucidate another level of inter-cellular communication. While a bioinformatic approach was used to extract the megakaryocyte (CD41⁺), erythrocyte (CD235a⁺), and monocyte (CD14⁺) expression profiles from the heterogeneous Lin⁺ populations, this does not fully substitute for experimentally generating the data. In particular, granulocytes (CD15⁺) constitute a significant fraction of the Lin⁺ population, but expression profiles for in vitro generated cells are not publicly available, thus leaving a gap in the resulting network. Repeating the intra-cellular network reconstruction methodology using more purified populations would refine the network structure, perhaps identifying more candidate regulators and better elucidation of direct vs. indirect signalling mechanisms.

The methodology could be further applied to other hematopoietic stem and progenitor culture systems as a standard platform for assessing conserved features of intra-cellular signalling in supportive vs. non-supportive micro-environments. For example, intra-cellular crosstalk between stromal cells, progenitors, and differentiated populations may underlie functional differences between supportive vs. non-supportive stromal cell lines (Li et al., 2004; Wineman et al., 1996), rather than some unique, as yet undefined molecular signature (Hackney et al., 2002; Punzel et al., 1999; Wagner et al., 2005b). Indeed, it is known that soluble factor cross-talk between blood progenitors and stoma is required for supportive activity (Gupta et al., 1998; Koller et al., 1999), indicating our methodology may be an ideal platform to define this elusive cross-talk. This approach could be further extended to analyze the in vivo stem cell niche. Much progress has been made recently in identifying the cellular composition of the HSC microenvironment in vivo (Hooper et al., 2009; Naveiras et al., 2009; Sacchetti et al., 2007), and methods are now available for the experimental generation of ectopic HSC niches (Chan et al., 2009). Systematically reconstructing the interactions between progenitors, differentiated cells, and various microenvironmental components (osteoblasts, vascular endothelial cells, adipocytes)
in response to dynamic perturbations (i.e. bleeding, irradiation, HSC transplantation) would shed light on the complex, and elusive signals regulating stem cell fate *in vivo*.

Additionally, the ligand-receptor compendium used to construct the cell-cytokine networks ([Supplemental Table S26](#)) is incomplete. The list of ligands likely represents only a fraction of the soluble signalling molecules present in the human genome (Grimmond *et al.*, 2003), of which the majority signal through as yet undefined receptors (Ben-Shlomo *et al.*, 2007). The list should therefore be updated regularly in light of new information, and additional complexities, such as ligand-receptor affinities could be incorporated.

### 4.5 Future intra-cellular network analysis work

The intra-cellular networks presented in Chapter 2 should be treated as a work in progress – rather than a systematic *a priori* reconstructions, these are literature-curated and thus incomplete and subject to extensive bias. The 112 self-renewal regulators selected to construct the network represent what is known from literature as of June 2009, rather than the complete biological set of regulatory genes. The list is in fact quite dynamic, as new genetic studies are continually reported, and older reports occasionally disputed by new evidence. The WNT gene family represents a prime example of conflicting reports due to differing experimental methods and context (Mikels and Nusse, 2006). An additional complexity arises due to the fact that many genes in this list may represent indirect or non-autonomous regulators, as it is often not possible to discriminate between autonomous and non-autonomous effects (Deneault *et al.*, 2009).

The protein interaction databases used to construct the blood stem cell self-renewal network are known to be riddled with false positives and negatives, as evidenced by the small degree of overlap between independent datasets (Futschik *et al.*, 2007; Ramirez *et al.*, 2007). While not explicitly considered in our analysis, confidence scores can be assigned to interactions based on their reoccurrence in independent datasets (Missiuro *et al.*, 2009). The resulting self-renewal network ($V, E = 2130, 5490$) displays a scale-free degree distribution, consistent with other
protein-protein interaction networks (Rual et al., 2005). However it has been questioned whether this distribution arises as a result of selective sampling rather than representing true biological network topology (Barabasi and Albert, 1999; Han et al., 2005). By examination of the top 10 highest connected nodes (Table 4-1) it is apparent that these represent some of the most extensively studied genes in the human genome. While it is clear that such “hub” proteins are functionally and evolutionary significant (Batada et al., 2006; Jeong et al., 2001), the bias in connectivity toward extensively studied (and well funded) genes cannot be denied. For example, Jeff Wrana’s “LUMINER” screen for interaction partners with TGF-β pathway components is included in the i2D database (Barrios-Rodiles et al., 2005), accounting for the particularly high connectivity TGFBR1 and SMAD4.

Table 4-1. 10 most highly connected nodes in the blood stem cell self-renewal network

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
<th>SYMBOL</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>7157</td>
<td>TUMOR PROTEIN P53</td>
<td>TP53</td>
<td>550</td>
</tr>
<tr>
<td>7046</td>
<td>TRANSFORMING GROWTH FACTOR, BETA RECEPTOR 1</td>
<td>TGFBR1</td>
<td>512</td>
</tr>
<tr>
<td>4089</td>
<td>SMAD, MOTHERS AGAINST DPP HOMOLOG 4</td>
<td>SMAD4</td>
<td>428</td>
</tr>
<tr>
<td>4609</td>
<td>V-MYC MYELOCYTOMATOSIS VIRAL ONCOGENE</td>
<td>MYC</td>
<td>334</td>
</tr>
<tr>
<td>1387</td>
<td>CREB BINDING PROTEIN</td>
<td>CREBBP</td>
<td>297</td>
</tr>
<tr>
<td>5925</td>
<td>RETINOBLASTOMA 1</td>
<td>RB1</td>
<td>274</td>
</tr>
<tr>
<td>1499</td>
<td>CATENIN (CADHERIN-ASSOCIATED PROTEIN), BETA 1</td>
<td>CTNNB1</td>
<td>245</td>
</tr>
<tr>
<td>1026</td>
<td>CYCLIN-DEPENDENT KINASE INHIBITOR 1A (P21, CIP1)</td>
<td>CDKN1A</td>
<td>176</td>
</tr>
<tr>
<td>5781</td>
<td>PROTEIN TYROSINE PHOSPHATASE, TYPE 11</td>
<td>PTPN11</td>
<td>166</td>
</tr>
<tr>
<td>2932</td>
<td>GLYCOGEN SYNTHASE KINASE 3 BETA</td>
<td>GSK3B</td>
<td>162</td>
</tr>
</tbody>
</table>

The pathway databases currently available (PANTHER, KEGG) are similarly incomplete and biased (Bader et al., 2006). In particular, chemokine signal transduction pathways remain poorly defined—we have shown that CCL3, CCL4, and CXCL10 display similar inhibitory effects on stem cell self-renewal, while CCL5 and CXCL8 display no functional effects. However, signalling through distinct chemokine family members is lumped together as “chemokine signalling”. The same holds true for TGFB, WNT, and Interleukin signalling pathways – while different family members display distinct functional and signal transduction activities, these nuances are not reflected in the databases. Moreover, many signal transduction pathways are simply not represented in any databases, such as TNFSF9, which we have identified as a potent
self-renewal inhibitor. Analysis is thus limited to extensively studied, “canonical” signalling pathways.

A more fundamental issue relates to the fact that such canonical signalling pathways represent only a fraction of the actual biochemical response to ligand stimulation. High-throughput, dynamic profiling studies indicate that the traditional notion of linear signal transduction cascades are gross over-simplifications (Janes et al., 2005; Miller-Jensen et al., 2007), and pathway “cross-talk” actually represents the majority of biochemical events (Li et al., 2008b). While systematic methods are being developed to reconstruct signal transduction cascades \textit{a priori} from dynamic molecular profiling data (Nelander et al., 2008; Saez-Rodriguez et al., 2008), validated databases derived from such studies are far from being widely available.

The rapid development of molecular interaction and pathway databases poses an issue in itself – as results are dependent upon databases used, data must be continually re-analyzed in light of new information. While updates are generally minor and will have little impact on downstream analysis, occasionally new information arises which will substantially affect results derived from the bioinformatics. To systematically address this issue, bioinformatic methods must be fully algorithmitized to allow for automatic updating, an otherwise Sisyphean task.

4.6 Future validation studies

The primary motivation behind this work has been to translate biological knowledge about cellular regulatory networks into optimized culture conditions for the \textit{in vitro} expansion of HSCs. We have identified a number of endogenous self-renewal stimulators and inhibitors which may be exogenously supplied or blocked, and defined culture variables (low cell density, repeated Lin\textsuperscript{-} selection, and media dilution) which enhance primitive progenitor expansion through modulating inter-cellular feedback. However, these various factors have not been systematically integrated to define an optimized culture process. Due to the dynamic complexity and feedback regulatory mechanisms involved, it is unlikely that a single factor will possess the
ability to induce sustained self-renewal divisions \textit{in vitro}. Targeting conserved intra-cellular proteins wherein stimulatory and/or inhibitory signals converge thus may be a better strategy to overcome molecular redundancy and feedback than focusing on extra-cellular signals. We have demonstrated that this approach can be utilized in principle, to inhibit self-renewal. More systematic screens using small molecule agonists/antagonist targeting the conserved molecules described in Chapter 3 should identify a subset stimulating self-renewal.

A major limitation to the implementation of high-throughput screens is the time consuming and laborious functional assays required to read-out stem cell activity. The LTC-IC assay, as implemented for screening ligand and mature cell co-culture effects, is practically limited to measuring at most 10-20 conditions per experiment. The long time frame required (8 weeks in total) and intrinsic variability also severely limits the use in iterative screening or process development cycles. While no phenotypic markers currently available exclusively define primitive progenitors, we have recently demonstrated that CD133$^+$CD38$^-$ phenotype tracks with both primitive progenitor (LTCIC) and stem cell (SRC) activity during \textit{in vitro} culture (Ito et al., 2009). Fold expansion of the CD133$^+$CD38$^-$ population may therefore serve as a surrogate readout of stem cell expansion. Importantly, this could be implemented using 96 or 386-well culture plates, and automatically quantified using flow cytometry. Initial validation and optimization studies could be conducted using VEGF and CCL3 as respective positive and negative controls

Due to the complexity of the system, it is highly unlikely that linearly independent relationships will exist between culture variables. Factorial design of experiments (DOE) using stimulatory ligands and small molecule agonists/antagonists in combination will therefore be required for media supplement optimization (Lim et al., 2007), as is standard practice in the biopharmaceutical industry (Moran et al., 2000). After an optimized media supplement, or limited set of supplements are defined, the next step will be to experimentally test fed-batch culture strategies using the new media. Using an automated media delivery system (Csaszar et al., 2009) it will be possible to conduct a limited number (≤10) of fed-batch cultures in parallel, to experimentally test the effects of dilution rate and/or different feeding strategies on culture
output, and possibly relationships between media supplement and dilution strategies, using
LTCIC expansion as a primary culture readout.

The gold standard assay for human HSC activity is limiting dilution SRC assays (either using
NOD/Scid, NOD/Scidβ2m<sup>-/-</sup>, or NOD/Scid/γ<sub>c</sub>-/- mice). Due to the large numbers of mice and
cells required, these in vivo assays are inappropriate for initial screening studies. However, once
an optimized culture (media and feeding strategy) is devised, quantifying SRC expansion will be
necessary as a pre-clinical predictor of transplantable human HSC content.

4.7 Broader implications and significance

One of the most fundamental questions in biology is “how do biological systems respond
appropriately to complex and dynamic environments?” This is relevant at all scales of life, from
cellular biochemistry, to tissues, organ systems, organisms, and even whole ecosystems. It is
therefore often appropriate to consider biological networks as information processing systems.
While the philosophical foundations for such an approach was originally formulated as
Cybernetics and General Systems Theory over 40 years ago (von Bertalanffy, 1968), the
necessary experimental and computational methods required have only recently become
available. As a testament to the information processing capabilities of biological systems, the
most advanced optimization algorithms in use today – genetic algorithms, ant colony and particle
swarm intelligence, and artificial immune systems - are based directly on concepts derived from
biology (Larranaga et al., 2006).

This work represents an attempt at systematically utilizing a biological knowledgebase – from
both published literature and molecular databases – to elucidate the regulatory networks
underlying observable features of in vitro hematopoiesis. Experimental and computational
exploration of the system has highlighted the importance of inter-cellular feedback in regulating
cell fate decisions. Inter-cellular communication is a paradoxically essential yet often neglected
feature of cellular systems. Cell populations are commonly viewed as homogenous, autonomous
units responding directly to external stimuli. A more biologically relevant, yet complicated picture arises if one envisions cell populations as networks, with population-level responses to external stimuli mediated through alterations of network structure and dynamics. This work highlights the importance of considering such nuances when interpreting biological data. On a more practical level, the bioinformatic methods developed to reconstruct and analyze both inter- and intra-cellular networks are applicable to many other systems involving complex inter-cellular communication. Specific examples include, but are not limited to, other adult stem cell systems, disease pathology, developmental processes, and tissue engineering.
## 5 SUPPLEMENTARY TABLES

### Supplementary Table S1. Published gene sets used for compositional analysis

<table>
<thead>
<tr>
<th>CELL POPULATION</th>
<th>NO. GENES</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryonic Stem Cells</td>
<td>109</td>
<td>(Assou et al., 2007)</td>
</tr>
<tr>
<td>MPB CD34+ G0</td>
<td>84</td>
<td>(Graham et al., 2007)</td>
</tr>
<tr>
<td>MPB CD34+ M/G2/S'</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+ Fresh</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+ 7d Cultured</td>
<td>20</td>
<td>(Li et al., 2006)</td>
</tr>
<tr>
<td>UCB CD34+CD38+ Fresh</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+CD38- Fresh</td>
<td>39</td>
<td>(Wagner et al., 2004)</td>
</tr>
<tr>
<td>UCB CD34+CD38- 7d Cultured SDF</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+CD38- 7d Cultured FDF</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>UCB &amp; MPB CD133+</td>
<td>135</td>
<td>(Toren et al., 2005)</td>
</tr>
<tr>
<td>UCB CD34+ culture-derived Erythroblasts (glyA+)</td>
<td>269</td>
<td>(Ferrari et al., 2007)</td>
</tr>
<tr>
<td>UCB CD34+ culture-derived Monoblasts (CD14+)</td>
<td>277</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+ culture-derived Myeloblasts (CD14+)</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>UCB CD34+ culture-derived Megakaryoblasts (CD41+)</td>
<td>214</td>
<td></td>
</tr>
<tr>
<td>BM CD34+ culture-derived Erythroblasts (CD71+)</td>
<td>39</td>
<td>(Komor et al., 2005)</td>
</tr>
<tr>
<td>BM CD34+ culture-derived Granuloblasts (CD15+)</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>BM CD34+ culture-derived Megakaryoblasts (CD61+)</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>murine HSC</td>
<td>227 / 374</td>
<td>(Chambers et al., 2007)</td>
</tr>
<tr>
<td>murine B cells</td>
<td>51 / 100</td>
<td></td>
</tr>
<tr>
<td>murine T cells</td>
<td>23 / 42</td>
<td></td>
</tr>
<tr>
<td>murine NK cells</td>
<td>50 / 103</td>
<td></td>
</tr>
<tr>
<td>murine Monocytes</td>
<td>69 / 108</td>
<td></td>
</tr>
<tr>
<td>murine Granulocytes</td>
<td>64 / 101</td>
<td></td>
</tr>
<tr>
<td>murine Erythrocytes</td>
<td>182 / 349</td>
<td></td>
</tr>
<tr>
<td>murine Differentiation common</td>
<td>20 / 44</td>
<td></td>
</tr>
<tr>
<td>murine Myeloid common</td>
<td>101 / 179</td>
<td></td>
</tr>
<tr>
<td>murine Lymphoid common</td>
<td>11 / 20</td>
<td></td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Bone marrow</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>379</td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Fetal Brain</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Fetal liver</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Fetal lung</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>211</td>
<td>(Ge et al., 2005)</td>
</tr>
<tr>
<td>Lung</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Ovary</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Pancreas</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Placenta</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Prostate</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Salivary gland</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Skeletal muscle</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Small intestine</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Spleen</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Stomach</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Tissue</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Testis</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>Thymus</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Thyroid</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Trachea</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Uterus</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

MPB = mobilized peripheral blood; UCB = umbilical cord blood; BM = bone marrow

### Supplementary Table S2. Pairwise sample comparisons

<table>
<thead>
<tr>
<th>NAME</th>
<th>COMPARISON</th>
<th>P-VALUE</th>
<th>FDR</th>
<th>NO. UP</th>
<th>NO. DOWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (B1)</td>
<td>d0 (Lin^- vs. Lin^+)</td>
<td>0.3</td>
<td>10</td>
<td>1252</td>
<td>1292</td>
</tr>
<tr>
<td>C2 (B2)</td>
<td>Lin^- (d4 vs. d0)</td>
<td>0.3</td>
<td>3</td>
<td>996</td>
<td>2306</td>
</tr>
<tr>
<td>C3 (B3)</td>
<td>Lin^- (d4 vs. d0)</td>
<td>0.5</td>
<td>2</td>
<td>1288</td>
<td>2781</td>
</tr>
<tr>
<td>C4 (B4)</td>
<td>d4 (Lin^- vs. Lin^+)</td>
<td>0.1</td>
<td>12</td>
<td>538</td>
<td>191</td>
</tr>
<tr>
<td>C5 (B5)</td>
<td>Total (Lin^- vs. Lin^+)</td>
<td>0.5</td>
<td>2</td>
<td>1708</td>
<td>418</td>
</tr>
<tr>
<td>C12 (B6)</td>
<td>Total (cultured vs. d0)</td>
<td>0.4</td>
<td>9</td>
<td>1481</td>
<td>2582</td>
</tr>
<tr>
<td>C6 (C1)</td>
<td>Lin^- d8 (SE vs. NSNE)</td>
<td>0.1</td>
<td>100</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>C7 (C2)</td>
<td>Lin^- d8 (SE vs. NSNE)</td>
<td>0.2</td>
<td>44</td>
<td>422</td>
<td>204</td>
</tr>
<tr>
<td>C8 (C3)</td>
<td>Lin^- (d8NSNE vs. d4)</td>
<td>0.3</td>
<td>21</td>
<td>403</td>
<td>431</td>
</tr>
<tr>
<td>C9 (C4)</td>
<td>Lin^- (d8NSNE vs. d4)</td>
<td>0.3</td>
<td>26</td>
<td>355</td>
<td>521</td>
</tr>
<tr>
<td>C10 (C5)</td>
<td>Lin^- (d12 vs. d8SE)</td>
<td>0.3</td>
<td>54</td>
<td>330</td>
<td>320</td>
</tr>
<tr>
<td>C11 (C6)</td>
<td>Lin^- (d12 vs. d8SE)</td>
<td>0.3</td>
<td>22</td>
<td>576</td>
<td>474</td>
</tr>
</tbody>
</table>

### Supplementary Table S3. B1 - d0 (Lin^- vs. Lin^+) - Lin^- Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell cycle</td>
<td>8.75E-08</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>1.61E-07</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>2.35E-06</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>9.85E-05</td>
</tr>
</tbody>
</table>

### Supplementary Table S4. B1 - d0 (Lin^- vs. Lin^+) - Lin^+ Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunity and defense</td>
<td>2.72E-25</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>8.45E-22</td>
</tr>
<tr>
<td>Intracellular signaling cascade</td>
<td>7.35E-12</td>
</tr>
<tr>
<td>Cytokine and chemokine mediated signaling pathway</td>
<td>2.36E-10</td>
</tr>
<tr>
<td>Cell surface receptor mediated signal transduction</td>
<td>2.15E-09</td>
</tr>
<tr>
<td>T-cell mediated immunity</td>
<td>2.66E-07</td>
</tr>
<tr>
<td>JAK-STAT cascade</td>
<td>7.40E-07</td>
</tr>
<tr>
<td>B-cell- and antibody-mediated immunity</td>
<td>2.45E-06</td>
</tr>
<tr>
<td>Calcium mediated signaling</td>
<td>5.60E-06</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>6.20E-06</td>
</tr>
<tr>
<td>Cell motility</td>
<td>7.00E-06</td>
</tr>
<tr>
<td>Macrophage-mediated immunity</td>
<td>9.15E-06</td>
</tr>
<tr>
<td>Cell communication</td>
<td>1.52E-05</td>
</tr>
<tr>
<td>Natural killer cell mediated immunity</td>
<td>2.06E-05</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>2.24E-05</td>
</tr>
<tr>
<td>Protein biosynthesis</td>
<td>3.94E-05</td>
</tr>
</tbody>
</table>

### Supplementary Table S5. B2 - Lin^- (d4 vs. d0) - d4 Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell cycle</td>
<td>6.75E-13</td>
</tr>
</tbody>
</table>
### Supplementary Table S6. B2 - Lin⁻ (d4 vs. d0) - d0 Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>1.92E-07</td>
</tr>
<tr>
<td>Intracellular signaling cascade</td>
<td>8.15E-07</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>6.55E-05</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td></td>
</tr>
<tr>
<td>Cell cycle control</td>
<td></td>
</tr>
<tr>
<td>Mitosis</td>
<td></td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td></td>
</tr>
</tbody>
</table>

### Supplementary Table S7. B3 - Lin⁺ (d4 vs. d0) - d4 Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal transduction</td>
<td>3.06E-17</td>
</tr>
<tr>
<td>Immunity and defense</td>
<td>1.51E-14</td>
</tr>
<tr>
<td>Intracellular signaling cascade</td>
<td>3.40E-11</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>3.28E-08</td>
</tr>
<tr>
<td>Cell surface receptor mediated signal transduction</td>
<td>2.95E-07</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>4.79E-07</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>6.65E-07</td>
</tr>
<tr>
<td>Nucleoside, nucleotide and nucleic acid metabolism</td>
<td>8.45E-07</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>8.60E-07</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>9.40E-07</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>2.77E-06</td>
</tr>
<tr>
<td>Protein modification</td>
<td>5.15E-06</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>1.13E-05</td>
</tr>
<tr>
<td>Oncogenesis</td>
<td>2.16E-05</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>2.61E-05</td>
</tr>
<tr>
<td>B-cell- and antibody-mediated immunity</td>
<td>7.25E-05</td>
</tr>
</tbody>
</table>

### Supplementary Table S8. B3 - Lin⁺ (d4 vs. d0) - d0 Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunity and defense</td>
<td>1.65E-18</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>1.03E-12</td>
</tr>
<tr>
<td>Intracellular signaling cascade</td>
<td>1.41E-10</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>7.30E-10</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>1.54E-09</td>
</tr>
<tr>
<td>Protein modification</td>
<td>1.95E-09</td>
</tr>
<tr>
<td>mRNA transcription</td>
<td>2.78E-09</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>2.30E-08</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>2.50E-08</td>
</tr>
<tr>
<td>mRNA transcription regulation</td>
<td>4.34E-08</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>5.70E-08</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>1.06E-07</td>
</tr>
<tr>
<td>Chemosensory perception</td>
<td>2.48E-07</td>
</tr>
<tr>
<td>T-cell mediated immunity</td>
<td>3.50E-07</td>
</tr>
<tr>
<td>Olfaction</td>
<td>6.45E-07</td>
</tr>
<tr>
<td>BIOLOGICAL PROCESS</td>
<td>P-VALUE</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Nucleoside, nucleotide and nucleic acid metabolism</td>
<td>1.36E-06</td>
</tr>
<tr>
<td>Cell motility</td>
<td>2.55E-06</td>
</tr>
<tr>
<td>Protein metabolism and modification</td>
<td>5.70E-06</td>
</tr>
<tr>
<td>Sensory perception</td>
<td>7.65E-06</td>
</tr>
<tr>
<td>Mesoderm development</td>
<td>1.35E-05</td>
</tr>
<tr>
<td>B-cell- and antibody-mediated immunity</td>
<td>2.79E-05</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>2.92E-05</td>
</tr>
<tr>
<td>Oncogenesis</td>
<td>3.98E-05</td>
</tr>
<tr>
<td>Hematopoiesis</td>
<td>4.57E-05</td>
</tr>
</tbody>
</table>

**Supplementary Table S9. B4 - d4 (Lin\(^-\) vs. Lin\(^+\)) - Lin\(^-\) Enriched**

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular signaling cascade</td>
<td>3.50E-08</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>2.61E-07</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>1.69E-05</td>
</tr>
<tr>
<td>Protein modification</td>
<td>2.11E-05</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>4.14E-05</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>6.60E-05</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>7.10E-05</td>
</tr>
<tr>
<td>Mesoderm development</td>
<td>7.25E-05</td>
</tr>
</tbody>
</table>

**Supplementary Table S10. B4 - d4 (Lin\(^-\) vs. Lin\(^+\)) - Lin\(^+\) Enriched**

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunity and defense</td>
<td>1.00E-04</td>
</tr>
</tbody>
</table>

**Supplementary Table S11. B5 - Total (Lin\(^-\) vs. Lin\(^+\)) - Lin\(^-\) Enriched**

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular signaling cascade</td>
<td>1.73E-21</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>2.67E-17</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>4.42E-14</td>
</tr>
<tr>
<td>Protein modification</td>
<td>4.11E-12</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>5.60E-10</td>
</tr>
<tr>
<td>Intracellular protein traffic</td>
<td>3.99E-09</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>5.80E-09</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>3.44E-08</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>3.56E-08</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>4.72E-07</td>
</tr>
<tr>
<td>Oncogenes</td>
<td>9.45E-07</td>
</tr>
<tr>
<td>General vesicle transport</td>
<td>2.48E-06</td>
</tr>
<tr>
<td>Mesoderm development</td>
<td>3.01E-06</td>
</tr>
<tr>
<td>MAPKKK cascade</td>
<td>4.17E-06</td>
</tr>
<tr>
<td>Protein biosynthesis</td>
<td>4.56E-06</td>
</tr>
<tr>
<td>mRNA transcription</td>
<td>5.25E-06</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>1.73E-05</td>
</tr>
<tr>
<td>Cell surface receptor mediated signal transduction</td>
<td>3.52E-05</td>
</tr>
<tr>
<td>mRNA transcription regulation</td>
<td>3.76E-05</td>
</tr>
<tr>
<td>Receptor protein tyrosine kinase signaling pathway</td>
<td>4.16E-05</td>
</tr>
<tr>
<td>Cell motility</td>
<td>5.60E-05</td>
</tr>
<tr>
<td>Nucleoside, nucleotide and nucleic acid metabolism</td>
<td>8.20E-05</td>
</tr>
</tbody>
</table>
### Supplementary Table S12. B5 - Total (Lin\(^-\) vs. Lin\(^+\)) - Lin\(^+\) Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porphyrin metabolism</td>
<td>8.20E-10</td>
</tr>
<tr>
<td>Coenzyme and prosthetic group metabolism</td>
<td>1.12E-05</td>
</tr>
</tbody>
</table>

### Supplementary Table S13. B6 - Total (cultured vs. d0) - Culture enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Cycle</td>
<td>1.11E-11</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>1.91E-09</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>1.27E-06</td>
</tr>
<tr>
<td>Mitosis</td>
<td>6.75E-06</td>
</tr>
<tr>
<td>Cell structure</td>
<td>7.30E-06</td>
</tr>
<tr>
<td>Transport</td>
<td>1.01E-05</td>
</tr>
<tr>
<td>Cell motility</td>
<td>2.00E-05</td>
</tr>
<tr>
<td>Porphyrin metabolism</td>
<td>2.23E-05</td>
</tr>
</tbody>
</table>

### Supplementary Table S14. B6 - Total (cultured vs. d0) - d0 Enriched

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunity and defense</td>
<td>2.19E-18</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>6.05E-17</td>
</tr>
<tr>
<td>Intracellular signaling cascade</td>
<td>7.55E-17</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>1.29E-13</td>
</tr>
<tr>
<td>mRNA transcription</td>
<td>4.05E-12</td>
</tr>
<tr>
<td>Nucleoside, nucleotide and nucleic acid metabolism</td>
<td>1.71E-11</td>
</tr>
<tr>
<td>Protein modification</td>
<td>3.66E-11</td>
</tr>
<tr>
<td>mRNA transcription regulation</td>
<td>4.77E-11</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>1.03E-10</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>2.39E-09</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>5.85E-09</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>2.30E-08</td>
</tr>
<tr>
<td>MAPK/KK cascade</td>
<td>1.19E-07</td>
</tr>
<tr>
<td>T-cell mediated immunity</td>
<td>1.25E-07</td>
</tr>
<tr>
<td>JNK cascade</td>
<td>1.48E-07</td>
</tr>
<tr>
<td>Protein metabolism and modification</td>
<td>4.63E-07</td>
</tr>
<tr>
<td>Stress response</td>
<td>6.95E-07</td>
</tr>
<tr>
<td>Induction of apoptosis</td>
<td>1.28E-06</td>
</tr>
<tr>
<td>Hematopoiesis</td>
<td>2.38E-06</td>
</tr>
<tr>
<td>Chemosensory perception</td>
<td>5.50E-06</td>
</tr>
<tr>
<td>Sensory perception</td>
<td>7.40E-06</td>
</tr>
<tr>
<td>Oncogenesis</td>
<td>9.55E-06</td>
</tr>
<tr>
<td>Olfaction</td>
<td>1.26E-05</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>1.66E-05</td>
</tr>
<tr>
<td>Mesoderm development</td>
<td>2.44E-05</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>2.74E-05</td>
</tr>
<tr>
<td>JAK-STAT cascade</td>
<td>4.30E-05</td>
</tr>
<tr>
<td>B-cell- and antibody-mediated immunity</td>
<td>4.39E-05</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>7.60E-05</td>
</tr>
<tr>
<td>Intracellular protein traffic</td>
<td>9.45E-05</td>
</tr>
</tbody>
</table>
### Supplementary Table S15. HSC Expansion-Correlated Biological Processes

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>C</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cell- and antibody-mediated immunity</td>
<td>C5</td>
<td>5.85E-05</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>2.13E-04</td>
</tr>
<tr>
<td>Blood clotting</td>
<td>C2</td>
<td>1.26E-06</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>4.06E-05</td>
</tr>
<tr>
<td>Calcium mediated signaling</td>
<td>C2</td>
<td>7.55E-05</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>8.95E-05</td>
</tr>
<tr>
<td>Cell motility</td>
<td>C4</td>
<td>2.31E-05</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>1.23E-09</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>9.50E-06</td>
</tr>
<tr>
<td>Cell structure and motility</td>
<td>C4</td>
<td>1.61E-05</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>1.28E-08</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>3.02E-05</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>C2</td>
<td>8.60E-05</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>8.30E-06</td>
</tr>
<tr>
<td>Lipid, fatty acid and steroid metabolism</td>
<td>C5</td>
<td>2.21E-04</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>2.91E-04</td>
</tr>
<tr>
<td>MHCI-mediated immunity</td>
<td>C2</td>
<td>1.41E-04</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>8.55E-07</td>
</tr>
<tr>
<td>Phospholipid metabolism</td>
<td>C5</td>
<td>1.73E-05</td>
</tr>
<tr>
<td>Protein metabolism and modification</td>
<td>C5</td>
<td>1.56E-05</td>
</tr>
<tr>
<td>T-cell mediated immunity</td>
<td>C5</td>
<td>3.48E-04</td>
</tr>
</tbody>
</table>

### Supplementary Table S16. Non-Correlated Biological Processes

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>C</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>C3</td>
<td>5.85E-05</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>4.14E-04</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>C3</td>
<td>7.45E-05</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>2.93E-04</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>3.51E-04</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>2.14E-06</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>4.17E-04</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>C4</td>
<td>4.25E-06</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>3.42E-04</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>1.15E-06</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>1.04E-04</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>1.91E-05</td>
</tr>
<tr>
<td>Cell proliferation and differentiation</td>
<td>C4</td>
<td>2.85E-04</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>3.85E-04</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>1.68E-05</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>1.71E-04</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>2.22E-05</td>
</tr>
<tr>
<td>Developmental processes</td>
<td>C4</td>
<td>3.64E-04</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>3.03E-04</td>
</tr>
<tr>
<td>Immunity and defense</td>
<td>C4</td>
<td>7.30E-06</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>5.65E-09</td>
</tr>
</tbody>
</table>
### Intracellular signaling cascade

<table>
<thead>
<tr>
<th>Process</th>
<th>C</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA transcription</td>
<td>C3</td>
<td>3.58E-05</td>
</tr>
<tr>
<td>mRNA transcription regulation</td>
<td>C3</td>
<td>1.83E-04</td>
</tr>
<tr>
<td>mRNA transcription termination</td>
<td>C3</td>
<td>1.83E-04</td>
</tr>
<tr>
<td>Nucleoside, nucleotide and nucleic acid metabolism</td>
<td>C3</td>
<td>3.65E-05</td>
</tr>
<tr>
<td>Protein modification</td>
<td>C3</td>
<td>2.04E-06</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>C3</td>
<td>2.16E-05</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>C4</td>
<td>1.53E-06</td>
</tr>
</tbody>
</table>

### Supplementary Table S17. Depletion-Correlated Biological Processes

<table>
<thead>
<tr>
<th>BIOLOGICAL PROCESS</th>
<th>C</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome segregation</td>
<td>C2</td>
<td>8.20E-05</td>
</tr>
<tr>
<td>Induction of apoptosis</td>
<td>C3</td>
<td>4.46E-04</td>
</tr>
<tr>
<td>Intracellular protein traffic</td>
<td>C5</td>
<td>2.33E-04</td>
</tr>
<tr>
<td>JNK cascade</td>
<td>C4</td>
<td>4.74E-05</td>
</tr>
<tr>
<td>Macrophage-mediated immunity</td>
<td>C4</td>
<td>5.10E-04</td>
</tr>
<tr>
<td>MAPKKK cascade</td>
<td>C4</td>
<td>3.72E-06</td>
</tr>
<tr>
<td>Mesoderm development</td>
<td>C3</td>
<td>7.95E-06</td>
</tr>
<tr>
<td>Mitosis</td>
<td>C3</td>
<td>4.42E-05</td>
</tr>
<tr>
<td>Oncogenesis</td>
<td>C3</td>
<td>1.56E-04</td>
</tr>
<tr>
<td>Other polysaccharide metabolism</td>
<td>C5</td>
<td>1.06E-04</td>
</tr>
<tr>
<td>RNA catabolism</td>
<td>C4</td>
<td>4.55E-04</td>
</tr>
</tbody>
</table>

### Supplementary Table S18. HSC Expansion-Correlated Pathways

<table>
<thead>
<tr>
<th>PATHWAY</th>
<th>C</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HT2 type receptor mediated signaling pathway</td>
<td>C6</td>
<td>4.64E-04</td>
</tr>
<tr>
<td>Alpha adrenergic receptor signaling pathway</td>
<td>C2</td>
<td>2.22E-05</td>
</tr>
<tr>
<td>Axon guidance mediated by netrin</td>
<td>C5</td>
<td>6.14E-04</td>
</tr>
<tr>
<td>PATHWAY</td>
<td>C6</td>
<td>C5</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Blood coagulation</td>
<td>7.11E-11</td>
<td>7.41E-09</td>
</tr>
<tr>
<td>Endothelin signaling pathway</td>
<td>1.07E-05</td>
<td></td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway</td>
<td>4.99E-05</td>
<td></td>
</tr>
<tr>
<td>Histamine H1 receptor mediated signaling pathway</td>
<td>3.43E-04</td>
<td></td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group I pathway</td>
<td>3.60E-05</td>
<td></td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor 1 and 3 signaling pathway</td>
<td>2.29E-04</td>
<td></td>
</tr>
<tr>
<td>p53 pathway feedback loops 2</td>
<td>3.17E-05</td>
<td></td>
</tr>
<tr>
<td>PI3 kinase pathway</td>
<td>2.98E-04</td>
<td></td>
</tr>
<tr>
<td>VEGF signaling pathway</td>
<td>3.28E-04</td>
<td></td>
</tr>
</tbody>
</table>
| Supplementary Table S19. Non-Correlated Pathways
<table>
<thead>
<tr>
<th>PATHWAY</th>
<th>C</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiogenesis</td>
<td>C5</td>
<td>3.10E-04</td>
</tr>
<tr>
<td>B cell activation</td>
<td>C6</td>
<td>3.84E-04</td>
</tr>
<tr>
<td>EGF receptor signaling pathway</td>
<td>C5</td>
<td>2.04E-04</td>
</tr>
<tr>
<td>FGF signaling pathway</td>
<td>C6</td>
<td>2.95E-04</td>
</tr>
<tr>
<td>Inflammation mediated by chemokine and cytokine signalling</td>
<td>C4</td>
<td>3.12E-05</td>
</tr>
<tr>
<td>Integrin signalling pathway</td>
<td>C3</td>
<td>1.38E-05</td>
</tr>
<tr>
<td>PDGF signalling pathway</td>
<td>C5</td>
<td>2.91E-04</td>
</tr>
<tr>
<td>T cell activation</td>
<td>C2</td>
<td>7.29E-09</td>
</tr>
</tbody>
</table>
| Supplementary Table S20. Depletion-Correlated Pathways
<table>
<thead>
<tr>
<th>PATHWAY</th>
<th>C</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis signaling pathway</td>
<td>C4</td>
<td>4.02E-04</td>
</tr>
<tr>
<td>Circadian clock system</td>
<td>C3</td>
<td>6.93E-04</td>
</tr>
<tr>
<td>Insulin/IGF pathway-MAP kinase cascade</td>
<td>C5</td>
<td>1.08E-03</td>
</tr>
<tr>
<td>Pathway</td>
<td>C</td>
<td>p-value</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>Interleukin signaling pathway</td>
<td>C4</td>
<td>8.67E-04</td>
</tr>
<tr>
<td>Ras Pathway</td>
<td>C4</td>
<td>4.44E-05</td>
</tr>
<tr>
<td>TGF-beta signaling pathway</td>
<td>C3</td>
<td>1.61E-03</td>
</tr>
<tr>
<td>Toll receptor signaling pathway</td>
<td>C3</td>
<td>1.11E-03</td>
</tr>
</tbody>
</table>

Supplementary Table S21.  HSC Expansion-Correlated Ligands (16)

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>C</th>
<th>FOLD X</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2920</td>
<td>CXCL2</td>
<td>C2</td>
<td>2.09</td>
<td>1.55 - 3.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4</td>
<td>2.11</td>
<td>1.53 - 3.23</td>
</tr>
<tr>
<td>6372</td>
<td>CXCL6</td>
<td>C6</td>
<td>3.7</td>
<td>1.46 - 10.05</td>
</tr>
<tr>
<td>1950</td>
<td>EGF</td>
<td>C2</td>
<td>3.36</td>
<td>2.16 - 4.74</td>
</tr>
<tr>
<td>3384</td>
<td>ICAM2</td>
<td>C6</td>
<td>3.46</td>
<td>1.97 - 5.93</td>
</tr>
<tr>
<td>3385</td>
<td>ICAM3</td>
<td>C6</td>
<td>1.85</td>
<td>1.51 - 2.3</td>
</tr>
<tr>
<td>3603</td>
<td>IL16</td>
<td>C6</td>
<td>1.95</td>
<td>1.2 - 3.31</td>
</tr>
<tr>
<td>4052</td>
<td>LTBP1</td>
<td>C2</td>
<td>2.86</td>
<td>1.92 - 5.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4</td>
<td>1.85</td>
<td>1.27 - 3.41</td>
</tr>
<tr>
<td>3084</td>
<td>NRG1</td>
<td>C5</td>
<td>5.93</td>
<td>1.7 - 1E+08</td>
</tr>
<tr>
<td>5155</td>
<td>PDGFB</td>
<td>C2</td>
<td>5.53</td>
<td>3.5 - 11.56</td>
</tr>
<tr>
<td>6402</td>
<td>SELL</td>
<td>C5</td>
<td>2.88</td>
<td>1.31 - 7.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C6</td>
<td>4.24</td>
<td>2.55 - 8.09</td>
</tr>
<tr>
<td>6403</td>
<td>SELP</td>
<td>C6</td>
<td>2.29</td>
<td>1.34 - 7.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>2.83</td>
<td>1.92 - 5.26</td>
</tr>
<tr>
<td>6678</td>
<td>SPARC</td>
<td>C6</td>
<td>2.36</td>
<td>1.23 - 5.51</td>
</tr>
<tr>
<td>7078</td>
<td>TIMP3</td>
<td>C4</td>
<td>1.97</td>
<td>1.37 - 2.81</td>
</tr>
<tr>
<td>7112</td>
<td>TMPO</td>
<td>C4</td>
<td>2.06</td>
<td>1.45 - 3.21</td>
</tr>
<tr>
<td>7424</td>
<td>VEGFC</td>
<td>C6</td>
<td>2.35</td>
<td>1.61 - 4.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>2.67</td>
<td>1.98 - 3.96</td>
</tr>
<tr>
<td>7481</td>
<td>WNT11</td>
<td>C2</td>
<td>6.6</td>
<td>2.67 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4</td>
<td>9.55</td>
<td>2.92 - 1E+08</td>
</tr>
</tbody>
</table>

Supplementary Table S22.  Non-Correlated Ligands (6)

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>C</th>
<th>FOLD X</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6352</td>
<td>CCL5</td>
<td>C3</td>
<td>7.5</td>
<td>1.55 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>4.05</td>
<td>2.09 - 9.19</td>
</tr>
<tr>
<td>2921</td>
<td>CXCL3</td>
<td>C3</td>
<td>3.74</td>
<td>1.47 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>1.87</td>
<td>1.23 - 3.43</td>
</tr>
<tr>
<td>1839</td>
<td>HBEGF</td>
<td>C3</td>
<td>4.61</td>
<td>2.08 - 18.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>1.89</td>
<td>1.33 - 3.19</td>
</tr>
<tr>
<td>3082</td>
<td>HGF</td>
<td>C4</td>
<td>5.12</td>
<td>1.83 - 15.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3</td>
<td>2.77</td>
<td>2.01 - 3.92</td>
</tr>
<tr>
<td>3576</td>
<td>IL8</td>
<td>C3</td>
<td>8.04</td>
<td>4.18 - 30.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4</td>
<td>2.81</td>
<td>1.87 - 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C6</td>
<td>1.83</td>
<td>1.31 - 2.72</td>
</tr>
<tr>
<td>7057</td>
<td>THSB1</td>
<td>C3</td>
<td>4.17</td>
<td>1.41 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>2.51</td>
<td>1.73 - 3.94</td>
</tr>
</tbody>
</table>

Supplementary Table S23.  HSC Depletion-Correlated Ligands (11)

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>C</th>
<th>FOLD X</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6347</td>
<td>CCL2</td>
<td>C4</td>
<td>17.69</td>
<td>7.23 - 1E+08</td>
</tr>
</tbody>
</table>
### Supplementary Table S24. HSC Expansion-Correlated Receptors (11)

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>C</th>
<th>FOLDX</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>ACVR1B</td>
<td>C3</td>
<td>1.39</td>
<td>1.23 - 1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>1.96</td>
<td>1.31 - 3.7</td>
</tr>
<tr>
<td>1236</td>
<td>CCR7</td>
<td>C3</td>
<td>1.75</td>
<td>1.25 - 2.72</td>
</tr>
<tr>
<td>2260</td>
<td>FGFR1</td>
<td>C5</td>
<td>1.91</td>
<td>1.32 - 3.05</td>
</tr>
<tr>
<td>3480</td>
<td>IGF1R</td>
<td>C3</td>
<td>4.22</td>
<td>2.75 - 8.24</td>
</tr>
<tr>
<td>3561</td>
<td>IL2RG</td>
<td>C3</td>
<td>1.66</td>
<td>1.27 - 2.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5</td>
<td>2.14</td>
<td>1.71 - 2.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C6</td>
<td>2.61</td>
<td>1.67 - 5.17</td>
</tr>
<tr>
<td>3572</td>
<td>IL6ST</td>
<td>C2</td>
<td>2.32</td>
<td>1.49 - 4.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C6</td>
<td>3.11</td>
<td>2.17 - 4.84</td>
</tr>
<tr>
<td>4352</td>
<td>MPL</td>
<td>C4</td>
<td>2.72</td>
<td>1.82 - 5.15</td>
</tr>
<tr>
<td>4914</td>
<td>NTRK1</td>
<td>C3</td>
<td>5.08</td>
<td>2 - 98.57</td>
</tr>
<tr>
<td>7010</td>
<td>TEK</td>
<td>C4</td>
<td>5.94</td>
<td>1.6 - 3.96</td>
</tr>
<tr>
<td>27242</td>
<td>TNFRSF21</td>
<td>C5</td>
<td>2.26</td>
<td>1.61 - 3.49</td>
</tr>
<tr>
<td>8718</td>
<td>TNFRSF25</td>
<td>C4</td>
<td>2.04</td>
<td>1.34 - 3.68</td>
</tr>
</tbody>
</table>

### Supplementary Table S25. HSC Depletion-Correlated Receptors (12)

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>C</th>
<th>FOLDX</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1524</td>
<td>CX3CR1</td>
<td>C5</td>
<td>1.77</td>
<td>1.21 - 2.56</td>
</tr>
<tr>
<td>1436</td>
<td>CSF1R</td>
<td>C6</td>
<td>1.78</td>
<td>1.3 - 2.56</td>
</tr>
<tr>
<td>2057</td>
<td>EPOR</td>
<td>C6</td>
<td>1.97</td>
<td>1.48 - 2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3</td>
<td>1.75</td>
<td>1.55 - 1.98</td>
</tr>
<tr>
<td>2261</td>
<td>FGFR3</td>
<td>C6</td>
<td>2.78</td>
<td>2.05 - 3.89</td>
</tr>
<tr>
<td>7855</td>
<td>FZD5</td>
<td>C6</td>
<td>2.3</td>
<td>1.44 - 3.73</td>
</tr>
<tr>
<td>3482</td>
<td>IGF2R</td>
<td>C6</td>
<td>1.53</td>
<td>1.3 - 1.79</td>
</tr>
<tr>
<td>7850</td>
<td>IL1R2</td>
<td>C4</td>
<td>4.95</td>
<td>2.31 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3</td>
<td>5.94</td>
<td>2.06 - 1E+08</td>
</tr>
<tr>
<td>3597</td>
<td>IL13RA1</td>
<td>C6</td>
<td>2.53</td>
<td>1.64 - 1E+08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5</td>
<td>2.57</td>
<td>1.76 - 3.42</td>
</tr>
</tbody>
</table>
### Supplementary Table S26. Ligand-Receptor Interactions

<table>
<thead>
<tr>
<th>LIGAND</th>
<th>RECEPTOR(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adiponectin, (ADIPOQ)</td>
<td>ADIPOR 1/2</td>
</tr>
<tr>
<td>agouti related protein homolog (AGRP)</td>
<td>Attractin (ATRN)</td>
</tr>
<tr>
<td>amphiregulin (AREG)</td>
<td>EGFR</td>
</tr>
<tr>
<td>angiopoietin-like 1 (ANGPTL1)</td>
<td>TEK</td>
</tr>
<tr>
<td>angiopoietin-like 2 (ANGPTL2)</td>
<td>TEK</td>
</tr>
<tr>
<td>angiopoietin-like 3 (ANGPTL3)</td>
<td>TEK</td>
</tr>
<tr>
<td>angiopoietin-like 4 (ANGPTL4)</td>
<td>TEK</td>
</tr>
<tr>
<td>angiopoietin-like 6 (ANGPTL5)</td>
<td>TEK</td>
</tr>
<tr>
<td>angiopoietin-like 7 (ANGPTL6)</td>
<td>TEK</td>
</tr>
<tr>
<td>Betacellulin (BTC)</td>
<td>EGF-R</td>
</tr>
<tr>
<td>bone morphogenetic protein 1 (BMP1)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 10 (BMP10)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 15 (BMP15)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 2 (BMP2)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 3 (BMP3)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 4 (BMP4)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 5 (BMP5)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>Bone morphogenetic protein 6 (BMP6)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 7 (BMP7)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 8a (BMP8A)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>bone morphogenetic protein 8b (BMP8B)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>brain-derived neurotrophic factor (BDNF)</td>
<td>BMPR 1A/1B/2</td>
</tr>
<tr>
<td>cardiotrophin 1 (CTF1)</td>
<td>LIF-R</td>
</tr>
<tr>
<td>XCL1 / XCL2 (I-309)</td>
<td>CCR8</td>
</tr>
<tr>
<td>CCL11 (Eotaxin)</td>
<td>CCR3</td>
</tr>
<tr>
<td>CCL13 (MCP-4)</td>
<td>CCR1</td>
</tr>
<tr>
<td>CCL14 / CCL15 (MIP-5 / MIP-1delta)</td>
<td>CCR 1/10</td>
</tr>
<tr>
<td>CCL16 (HCC-4 / LCC-1)</td>
<td>CCR 1/2</td>
</tr>
<tr>
<td>CCL17 (TARC / ABCD-2 / Dendrokine)</td>
<td>CCR4</td>
</tr>
<tr>
<td>CCL18 (pulmonary and activation-regulated / PARC)</td>
<td>???</td>
</tr>
<tr>
<td>CCL19 (MIP-3beta / Exodus-3)</td>
<td>CCR7</td>
</tr>
<tr>
<td>CCL2 (MCP-1 / GDCF)</td>
<td>CCR2</td>
</tr>
<tr>
<td>CCL20 (MIP-3alpha / Exodus-1)</td>
<td>CCR6</td>
</tr>
<tr>
<td>CCL21 (6CKine)</td>
<td>CCR7</td>
</tr>
<tr>
<td>CCL22 (ABCD-1 / MDC)</td>
<td>CCR4</td>
</tr>
<tr>
<td>CCL23 (Myeloid progenitor inhibitory factor-1 / MPIF-1)</td>
<td>CCR1</td>
</tr>
<tr>
<td>CCL24 (MPIF-2 / Eotaxin-2 / Ck-beta-6)</td>
<td>CCR3</td>
</tr>
<tr>
<td>CCL25 (TECK)</td>
<td>CCR9</td>
</tr>
<tr>
<td>CCL26 (MIP-4alpha / Eotaxin-3)</td>
<td>CCR3</td>
</tr>
<tr>
<td>CCL27 (ALP / CTACK)</td>
<td>CCR10</td>
</tr>
<tr>
<td>CCL28 (MEC / CCK1)</td>
<td>CCR 3/10</td>
</tr>
<tr>
<td>CCL3 (MIP-1alpha)</td>
<td>CCR 1/5</td>
</tr>
<tr>
<td>CCL4 (MIP-1beta)</td>
<td>CCR5</td>
</tr>
<tr>
<td>CCL5 (RANTES)</td>
<td>CCR 1/5</td>
</tr>
<tr>
<td>CCL7 (MCP-3)</td>
<td>CCR 1/5</td>
</tr>
<tr>
<td>CCL8 (MCP-2)</td>
<td>CCR 2/3/5</td>
</tr>
<tr>
<td>CX3CL1 (Fractalkine)</td>
<td>CX3CR1</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Receptor Name</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>CXCL1 (melanoma growth stimulating activity, alpha / GRO-a)</td>
<td>IL8RB (CXCR2)</td>
</tr>
<tr>
<td>CXCL10 (IP-10)</td>
<td>CXCR3</td>
</tr>
<tr>
<td>CXCL11 (I-TAC)</td>
<td>CXCR3</td>
</tr>
<tr>
<td>CXCL12 (stromal cell-derived factor 1)</td>
<td>CXCR3</td>
</tr>
<tr>
<td>CXCL13 (B-cell chemoattractant)</td>
<td>CXCR5</td>
</tr>
<tr>
<td>CXCL14 (MIP-2gamma / BRAK)</td>
<td>???</td>
</tr>
<tr>
<td>CXCL16 (SR-PSOX)</td>
<td>CXCR6</td>
</tr>
<tr>
<td>CXCL2 (GRO-b)</td>
<td>IL8RB</td>
</tr>
<tr>
<td>CXCL3 (GRO-g)</td>
<td>IL8RB</td>
</tr>
<tr>
<td>CXCL5 (ENA-78)</td>
<td>CXCR2</td>
</tr>
<tr>
<td>CXCL6 (granulocyte chemotactic protein 2)</td>
<td>IL8R A/B</td>
</tr>
<tr>
<td>CXCL7 (pro-platelet basic protein / NAP-2)</td>
<td>IL8RB</td>
</tr>
<tr>
<td>CXCL9 (MIG)</td>
<td>CXCR3</td>
</tr>
<tr>
<td>ciliary neurotrophic factor (CNF)</td>
<td>gp130</td>
</tr>
<tr>
<td>colony stimulating factor 1 (MCSF)</td>
<td>CSF1R</td>
</tr>
<tr>
<td>colony stimulating factor 2 (GMCSF)</td>
<td>CSF2R A/B</td>
</tr>
<tr>
<td>colony stimulating factor 3 (GCSF)</td>
<td>CSF3R / SMO</td>
</tr>
<tr>
<td>Desert Hedgehog homolog (DHH)</td>
<td>PTCH 1/2 / SMO</td>
</tr>
<tr>
<td>Delta-like 1 (DLL1)</td>
<td>NOTCH 1-4</td>
</tr>
<tr>
<td>Delta-like 3 (DLL2)</td>
<td>NOTCH 1-4</td>
</tr>
<tr>
<td>Endoglin (ENG)</td>
<td>xxx</td>
</tr>
<tr>
<td>endothelin 1 (EDN1)</td>
<td>EDNRA</td>
</tr>
<tr>
<td>endothelin 2 (EDN2)</td>
<td>EDNRA</td>
</tr>
<tr>
<td>endothelin 3 (EDN3)</td>
<td>EDNRB</td>
</tr>
<tr>
<td>epidermal growth factor (EGF)</td>
<td>EGF</td>
</tr>
<tr>
<td>Erythropoietin (EPO)</td>
<td>EPOR</td>
</tr>
<tr>
<td>Fas ligand (FASL / TNF superfamily member 6)</td>
<td>FAS</td>
</tr>
<tr>
<td>fibroblast growth factor 1 (FGF1)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 10 (FGF10)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 11 (FGF11)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 12 (FGF12)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 13 (FGF13)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 14 (FGF14)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 17 (FGF17)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 18 (FGF18)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 19 (FGF19)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 2 (FGF2)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 4 (FGF4)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 6 (FGF6)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 7 (FGF7)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 8 (FGF8)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fibroblast growth factor 9 (FGF9)</td>
<td>FGFR1-4</td>
</tr>
<tr>
<td>fms-related tyrosine kinase 3 ligand (FLT3LG)</td>
<td>FLT3</td>
</tr>
<tr>
<td>Follistatin (FST)</td>
<td>???</td>
</tr>
<tr>
<td>Galactosidase, beta 1 (GLB1)</td>
<td>???</td>
</tr>
<tr>
<td>glial cell derived neurotrophic factor (GDNF)</td>
<td>GFRA1</td>
</tr>
<tr>
<td>growth differentiation factor 15 (GDF15)</td>
<td>???</td>
</tr>
<tr>
<td>heparin-binding EGF-like growth factor (HBEGF)</td>
<td>???</td>
</tr>
<tr>
<td>hepatocyte growth factor (HGF)</td>
<td>met</td>
</tr>
<tr>
<td>hypocretin (orexin) neuropeptide precursor (HCRT)</td>
<td>HCRTR 1/2</td>
</tr>
<tr>
<td>Indian Hedgehog homolog (IHH)</td>
<td>PTCH 1/2 / SMO</td>
</tr>
<tr>
<td>inhibin, alpha (INHA / activin A)</td>
<td>ACVR 1/1B/2/2B</td>
</tr>
<tr>
<td>inhibin, beta A (INHBA / activin AB alpha polypeptide)</td>
<td>ACVR 1/1B/2/2B</td>
</tr>
<tr>
<td>inhibin, beta B (INHBB / activin AB beta polypeptide)</td>
<td>ACVR 1/1B/2/2B</td>
</tr>
<tr>
<td>insulin-like growth factor 1 (IGF1)</td>
<td>IGF1R</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Gene Name</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Insulin-like growth factor 2 (IGF1)</td>
<td>IGF2R</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 1 (IGFBP1)</td>
<td>xxx</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 2 (IGFBP2)</td>
<td>xxx</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 3 (IGFBP3)</td>
<td>xxx</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 4 (IGFBP4)</td>
<td>xxx</td>
</tr>
<tr>
<td>Interleukin receptor antagonist (IL1RN)</td>
<td>IL1R 1/2</td>
</tr>
<tr>
<td>Interleukin 1, alpha (IL1A)</td>
<td>IL1RA</td>
</tr>
<tr>
<td>Interleukin 1, beta (IL1B)</td>
<td>IL1R1/2</td>
</tr>
<tr>
<td>Interleukin 2 (IL2)</td>
<td>IL2R A/B/G</td>
</tr>
<tr>
<td>Interleukin 3 (IL3)</td>
<td>IL3RA</td>
</tr>
<tr>
<td>Interleukin 4 (IL4)</td>
<td>IL4R</td>
</tr>
<tr>
<td>Interleukin 5 (IL5)</td>
<td>IL5RA</td>
</tr>
<tr>
<td>Interleukin 6 (IL6)</td>
<td>IL6R / IL6RST</td>
</tr>
<tr>
<td>Interleukin 7 (IL7)</td>
<td>IL7R</td>
</tr>
<tr>
<td>Interleukin 8 (IL8)</td>
<td>IL8RB</td>
</tr>
<tr>
<td>Interleukin 9 (IL9)</td>
<td>IL9R / IL2RG</td>
</tr>
<tr>
<td>Jagged 1 (JAG1)</td>
<td>NOTCH 1-4</td>
</tr>
<tr>
<td>Jagged 2 (JAG2)</td>
<td>NOTCH 1-4</td>
</tr>
<tr>
<td>KIT ligand (KITLG / stem cell factor)</td>
<td>KIT</td>
</tr>
<tr>
<td>Latent transforming growth factor beta binding protein 1 (LTBP1)</td>
<td>xxx</td>
</tr>
<tr>
<td>Leptin (LEP)</td>
<td>LEPR</td>
</tr>
<tr>
<td>Leukemia inhibitory factor (LIF)</td>
<td>LIFR</td>
</tr>
<tr>
<td>Leukocyte cell-derived chemotaxin 2 (LECT2)</td>
<td>???</td>
</tr>
<tr>
<td>Lipoprotein, Lp(a) (LPA)</td>
<td>???</td>
</tr>
<tr>
<td>Lymphotoxin alpha (TNF superfamily, member 1 / TNFSF1 / TNF-beta)</td>
<td>TNFRSF1 A/B</td>
</tr>
<tr>
<td>Macrophage migration inhibitory factor (MIF / glycosylation-inhibiting factor)</td>
<td>CD74</td>
</tr>
<tr>
<td>Macrophage stimulating 1 ( MST1 / hepatocyte growth factor-like / HLP)</td>
<td>RON</td>
</tr>
<tr>
<td>Matrix metalloproteinase 1 (MMP1)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 10 (MMP10)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 11 (MMP11)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 12 (MMP12)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 13 (MMP13)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 14 (MMP14)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 19 (MMP19)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 20 (MMP20)</td>
<td>xxx</td>
</tr>
<tr>
<td>Matrix metalloproteinase 7 (MMP7)</td>
<td>xxx</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Gene Symbol(s)</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>matrix metalloproteinase 8 (MMP8)</td>
<td>xxx</td>
</tr>
<tr>
<td>matrix metalloproteinase 9 (MMP9)</td>
<td>xxx</td>
</tr>
<tr>
<td>nerve growth factor, beta polypeptide (NGF)</td>
<td>NTRK1 / NGFR</td>
</tr>
<tr>
<td>neuregulin 1 (NRG1 / glial growth factor 1)</td>
<td>???</td>
</tr>
<tr>
<td>neuregulin 2 (NRG2)</td>
<td>???</td>
</tr>
<tr>
<td>neuregulin 3 (NRG3)</td>
<td>???</td>
</tr>
<tr>
<td>neuregulin 4 (NRG4)</td>
<td>???</td>
</tr>
<tr>
<td>neuritin 1 (NRN1)</td>
<td>???</td>
</tr>
<tr>
<td>neurotrophin 3 (NTF3)</td>
<td>NTRK 2/3 / NGFR</td>
</tr>
<tr>
<td>neurotrophin 5 (NTF5)</td>
<td>NTRK 1/2 / NGFR</td>
</tr>
<tr>
<td>Notch homolog 1 (NOTCH1)</td>
<td>JAG 1/2</td>
</tr>
<tr>
<td>Notch homolog 2 (NOTCH2)</td>
<td>JAG 1/2</td>
</tr>
<tr>
<td>Notch homolog 3 (NOTCH3)</td>
<td>JAG 1/2</td>
</tr>
<tr>
<td>Notch homolog 4 (NOTCH4)</td>
<td>JAG 1/2</td>
</tr>
<tr>
<td>oncostatin M (OSM)</td>
<td>OSMR</td>
</tr>
<tr>
<td>placental growth factor (PGF)</td>
<td>Fhl1</td>
</tr>
<tr>
<td>Plasminogen (PLG)</td>
<td>???</td>
</tr>
<tr>
<td>plasminogen activator, urokinase (PLAU)</td>
<td>PLAU</td>
</tr>
<tr>
<td>Platelet-derived growth factor alpha polypeptide (PDGFB)</td>
<td>PDGFRB</td>
</tr>
<tr>
<td>platelet-derived growth factor beta polypeptide (PDGFA)</td>
<td>PDGFRB</td>
</tr>
<tr>
<td>secreted protein, acidic, cysteine-rich (SPARC / osteonectin)</td>
<td>???</td>
</tr>
<tr>
<td>selectin E (SELE / endothelial adhesion molecule 1)</td>
<td>xxx</td>
</tr>
<tr>
<td>selectin L (SELL / lymphocyte adhesion molecule 1)</td>
<td>xxx</td>
</tr>
<tr>
<td>selectin P (SELP / granule membrane protein 140kDa, antigen CD62)</td>
<td>xxx</td>
</tr>
<tr>
<td>Sonic Hedgehog homolog (SHH)</td>
<td>PTCH 1/2 / SMO</td>
</tr>
<tr>
<td>teratocarcinoma-derived growth factor 1 (TDGF1 / Cripto-1)</td>
<td>EGFR</td>
</tr>
<tr>
<td>thrombopoietin (THPO / mpl ligand)</td>
<td>MPL</td>
</tr>
<tr>
<td>Thrombospondin 1 (TSP1)</td>
<td>???</td>
</tr>
<tr>
<td>thymic stromal lymphopoietin (TSLP)</td>
<td>CRLF2</td>
</tr>
<tr>
<td>Thymopoietin (TMPO)</td>
<td>???</td>
</tr>
<tr>
<td>tissue inhibitor of metalloproteinase 1 (TIMP1)</td>
<td>xxx</td>
</tr>
<tr>
<td>tissue inhibitor of metalloproteinase 2 (TIMP2)</td>
<td>xxx</td>
</tr>
<tr>
<td>tissue inhibitor of metalloproteinase 3 (TIMP3)</td>
<td>xxx</td>
</tr>
<tr>
<td>tissue inhibitor of metalloproteinase 4 (TIMP4)</td>
<td>xxx</td>
</tr>
<tr>
<td>transforming growth factor, alpha (TGFα)</td>
<td>EGFR</td>
</tr>
<tr>
<td>transforming growth factor, beta 1 (TGFβ1)</td>
<td>TGFBR 1/2/3</td>
</tr>
<tr>
<td>Transforming growth factor, beta 2 (TGFβ2)</td>
<td>TGFBR 1/2/3</td>
</tr>
<tr>
<td>transforming growth factor, beta 3 (TGFβ3)</td>
<td>TGFBR 1/2/3</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 10 (TNFSF10 / TRAIL)</td>
<td>TNFRSF10A</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 11 (TNFSF11 / RANKL)</td>
<td>TNFRSF11A</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 12 (TNFSF12 / TWEAK)</td>
<td>???</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 13 (TNFSF13 / APRIL)</td>
<td>TNFRSF17</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 13b (TNFSF13B / BAFF)</td>
<td>TNFRSF 17/13b</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 14 (TNFSF14 / LIGHT)</td>
<td>TNFRSF14</td>
</tr>
<tr>
<td>tumor necrosis factor (ligand) superfamily, member 15 (TNFSF15 / VEGI)</td>
<td>TNFRSF 21/25 / CCR9</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 18 (TNFSF18 / GITRL)</td>
<td>TNFRSF18</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 4 (TNFSF4 / OX40L)</td>
<td>TNFRSF4</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 7 (TNFSF7 / CD70 / CD27L)</td>
<td>TNFRSF7</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 8 (TNFSF8 / CD30L)</td>
<td>TNFRSF8</td>
</tr>
<tr>
<td>Tumor necrosis factor (ligand) superfamily, member 9 (TNFSF9 / 4-1BBL)</td>
<td>4-1BB</td>
</tr>
<tr>
<td>vascular cell adhesion molecule 1 (VCAM1)</td>
<td>ITGB1</td>
</tr>
<tr>
<td>vascular endothelial growth factor (VEGFA)</td>
<td>FLT1 / KDR</td>
</tr>
<tr>
<td>vascular endothelial growth factor B (VEGFB)</td>
<td>FLT1</td>
</tr>
<tr>
<td>vascular endothelial growth factor C (VEGFC)</td>
<td>FLT4</td>
</tr>
<tr>
<td>vascular endothelial growth factor D (VEGF-D)</td>
<td>KDR / FLT4</td>
</tr>
</tbody>
</table>
wingless-type MMTV integration site family, member 1 (WNT1) | FZD 1-10
wingless-type MMTV integration site family, member 2 (WNT2) | FZD 1-10
wingless-type MMTV integration site family, member 3 (WNT3) | FZD 1-10
wingless-type MMTV integration site family, member 4 (WNT4) | FZD 1-10
wingless-type MMTV integration site family, member 5A (WNT5A) | FZD 1-10
wingless-type MMTV integration site family, member 5B (WNT5B) | FZD 1-10
wingless-type MMTV integration site family, member 6 (WNT6) | FZD 1-10
wingless-type MMTV integration site family, member 7A (WNT7A) | FZD 1-10
wingless-type MMTV integration site family, member 7B (WNT7B) | FZD 1-10
wingless-type MMTV integration site family, member 8A (WNT8A) | FZD 1-10
wingless-type MMTV integration site family, member 8B (WNT8B) | FZD 1-10
wingless-type MMTV integration site family, member 9A (WNT9A) | FZD 1-10
wingless-type MMTV integration site family, member 9B (WNT9B) | FZD 1-10
wingless-type MMTV integration site family, member 10A (WNT10A) | FZD 1-10
wingless-type MMTV integration site family, member 11 (WNT11) | FZD 1-10
wingless-type MMTV integration site family, member 16 (WNT16) | FZD 1-10

Legend: ??? = undefined receptor(s); xxx = non-signalling

### Supplementary Table S 27. Secreted ligand summary

<table>
<thead>
<tr>
<th>GENE</th>
<th>DIFFERENTIAL EXPRESSION</th>
<th>NETWORK</th>
<th>PROTEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADIPOQ</td>
<td>-</td>
<td>+</td>
<td>RayBio</td>
</tr>
<tr>
<td>ANGPTL6</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>AREG</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>BDNF</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>BMP1</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CCL2</td>
<td>HSC Depletion</td>
<td>-</td>
<td>Luminex</td>
</tr>
<tr>
<td>CCL3</td>
<td>HSC Depletion</td>
<td>+</td>
<td>RayBio / Luminex</td>
</tr>
<tr>
<td>CCL4</td>
<td>HSC Depletion</td>
<td>+</td>
<td>RayBio / Luminex</td>
</tr>
<tr>
<td>CCL5</td>
<td>Non-correlated</td>
<td>+</td>
<td>RayBio / Luminex</td>
</tr>
<tr>
<td>CXCL1</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CXCL10</td>
<td>HSC Depletion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CXCL12</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CXCL2</td>
<td>HSC Expansion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CXCL3</td>
<td>Non-correlated</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CXCL6</td>
<td>HSC Expansion</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CXCL7</td>
<td>-</td>
<td>+</td>
<td>RayBio</td>
</tr>
<tr>
<td>DLL1</td>
<td>HSC Depletion</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>EGF</td>
<td>HSC Depletion</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>EPO</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FASL</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FGF11</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FGF12</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FGF18</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FGF4</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>FST</td>
<td>HSC Depletion</td>
<td>???</td>
<td>-</td>
</tr>
<tr>
<td>GMCSF</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>HBEGF</td>
<td>Non-correlated</td>
<td>???</td>
<td>-</td>
</tr>
<tr>
<td>HGF</td>
<td>Non-correlated</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IHH</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IL13</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IL15</td>
<td>HSC Depletion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ID</td>
<td>SYMBOL</td>
<td>ALIAS</td>
<td>LOCAL.</td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>9370</td>
<td>ADIPOQ</td>
<td>Adiponectin</td>
<td>Ligand</td>
</tr>
<tr>
<td>284</td>
<td>ANGPT1</td>
<td>-</td>
<td>Ligand</td>
</tr>
<tr>
<td>23452</td>
<td>ANGPTL2</td>
<td>-</td>
<td>Ligand</td>
</tr>
<tr>
<td>27329</td>
<td>ANGPTL3</td>
<td>-</td>
<td>Ligand</td>
</tr>
<tr>
<td>253935</td>
<td>ANGPTL5</td>
<td>-</td>
<td>Ligand</td>
</tr>
<tr>
<td>472</td>
<td>ATM</td>
<td>-</td>
<td>Intracellular</td>
</tr>
</tbody>
</table>

Legend: + = Present; - = Absent; ??? = undefined receptor(s); xxx = non-signalling
<table>
<thead>
<tr>
<th>Gene</th>
<th>Alternate Name</th>
<th>Location</th>
<th>Direction 1</th>
<th>Direction 2</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIRC5</td>
<td>Survivin</td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Fukuda et al., 2002; Leung et al., 2007)</td>
</tr>
<tr>
<td>BMI1</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Lessard and Sauvageau, 2003; Park et al., 2003)</td>
</tr>
<tr>
<td>BMP4</td>
<td></td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Bhatia et al., 1999)</td>
</tr>
<tr>
<td>CCL3</td>
<td>MIP-1alpha</td>
<td>Ligand</td>
<td>-</td>
<td>0</td>
<td>(Mayani et al., 1995)</td>
</tr>
<tr>
<td>CCND1</td>
<td>CyclinD1</td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Kozar et al., 2004)</td>
</tr>
<tr>
<td>CCND2</td>
<td>CyclinD2</td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Kozar et al., 2004)</td>
</tr>
<tr>
<td>CCND3</td>
<td>CyclinD3</td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Kozar et al., 2004)</td>
</tr>
<tr>
<td>CD34</td>
<td></td>
<td>Receptor</td>
<td>+</td>
<td>0</td>
<td>(Bhatia et al., 1997; Salati et al., 2008)</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>p21Cip1/Waf1</td>
<td>Intracellular</td>
<td>+</td>
<td>-</td>
<td>(Cheng et al., 2000)</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>p27Kip1</td>
<td>Intracellular</td>
<td>-</td>
<td>-</td>
<td>(Walkley et al., 2005)</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>p16Ink4/P19Arf</td>
<td>Intracellular</td>
<td>-</td>
<td>-</td>
<td>(Janzen et al., 2006; Stepanova and Sorrentino, 2005)</td>
</tr>
<tr>
<td>CDKN2C</td>
<td>p18Ink4c</td>
<td>Intracellular</td>
<td>-</td>
<td>-</td>
<td>(Yuan et al., 2004)</td>
</tr>
<tr>
<td>CDX4</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Bansal et al., 2006)</td>
</tr>
<tr>
<td>CEBPA</td>
<td>C/EBP-α</td>
<td>Intracellular</td>
<td>-</td>
<td>-</td>
<td>(Heath et al., 2004)</td>
</tr>
<tr>
<td>CREBBP</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Rebel et al., 2002)</td>
</tr>
<tr>
<td>CSF2</td>
<td>GM-CSF</td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Yao et al., 2004)</td>
</tr>
<tr>
<td>CSF3</td>
<td>G-CSF</td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Kobari et al., 2000)</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>β-catenin</td>
<td>Intracellular</td>
<td>-</td>
<td>+</td>
<td>(Kirstetter et al., 2006; Scheller et al., 2006)</td>
</tr>
<tr>
<td>CXCL12</td>
<td>SDF-1</td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Lataille et al., 2000)</td>
</tr>
<tr>
<td>CXCR4</td>
<td>CXCR-4</td>
<td>Receptor</td>
<td>+</td>
<td>+</td>
<td>(Kahn et al., 2004)</td>
</tr>
<tr>
<td>DLL4</td>
<td>Delta-4</td>
<td>Ligand</td>
<td>+</td>
<td>-</td>
<td>(Lauret et al., 2004)</td>
</tr>
<tr>
<td>ELF4</td>
<td>MEF</td>
<td>Intracellular</td>
<td>-</td>
<td>+</td>
<td>(Lacorazza et al., 2006)</td>
</tr>
<tr>
<td>ENG</td>
<td>Endoglin</td>
<td>Receptor</td>
<td>+</td>
<td>+</td>
<td>(Chen et al., 2002)</td>
</tr>
<tr>
<td>EZH2</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>0</td>
<td>(Kamminga et al., 2006)</td>
</tr>
<tr>
<td>FASLG</td>
<td>TNFSF6</td>
<td>Ligand</td>
<td>-</td>
<td>-</td>
<td>(Bryder et al., 2001)</td>
</tr>
<tr>
<td>FG1</td>
<td></td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(de Haan et al., 2003)</td>
</tr>
<tr>
<td>FG2</td>
<td></td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Yeoh et al., 2006)</td>
</tr>
<tr>
<td>FLT3</td>
<td></td>
<td>Receptor</td>
<td>+</td>
<td>+</td>
<td>(Chung et al., 2005)</td>
</tr>
<tr>
<td>FLT3LG</td>
<td>FL</td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Chung et al., 2005)</td>
</tr>
<tr>
<td>FOXO1</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>-</td>
<td>(Tothova et al., 2007)</td>
</tr>
<tr>
<td>FOXO3</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>-</td>
<td>(Tothova et al., 2007)</td>
</tr>
<tr>
<td>FOXO4</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>-</td>
<td>(Tothova et al., 2007)</td>
</tr>
<tr>
<td>HOX3</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Chung et al., 2005)</td>
</tr>
<tr>
<td>GATA2</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Tothova et al., 2007)</td>
</tr>
<tr>
<td>GF11</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Hock et al., 2004)</td>
</tr>
<tr>
<td>GSK3B</td>
<td>GSK-3β</td>
<td>Intracellular</td>
<td>-</td>
<td>-</td>
<td>(Trowbridge et al., 2006)</td>
</tr>
<tr>
<td>HMGB3</td>
<td></td>
<td>Intracellular</td>
<td>-</td>
<td>0</td>
<td>(Nemeth et al., 2006)</td>
</tr>
<tr>
<td>HOX7</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Wang et al., 2007)</td>
</tr>
<tr>
<td>HOX8</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Lawrence et al., 2005; Lawrence et al., 1997)</td>
</tr>
<tr>
<td>HOX10</td>
<td></td>
<td>Intracellular</td>
<td>+</td>
<td>+</td>
<td>(Antonchuk et al., 2002; Krosl et al., 2003a)</td>
</tr>
<tr>
<td>IGF2</td>
<td></td>
<td>Ligand</td>
<td>+</td>
<td>+</td>
<td>(Fischbach et al., 2005)</td>
</tr>
<tr>
<td>IFNG</td>
<td>IFN-γ</td>
<td>Ligand</td>
<td>-</td>
<td>-</td>
<td>(Yang et al., 2005)</td>
</tr>
</tbody>
</table>

---

182
<table>
<thead>
<tr>
<th>Code</th>
<th>Gene</th>
<th>Type</th>
<th>Location</th>
<th>Expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3485</td>
<td>IGFBP2</td>
<td>Ligand</td>
<td>+ +</td>
<td>5</td>
<td>(Zhang et al., 2008a)</td>
</tr>
<tr>
<td>3549</td>
<td>IHH</td>
<td>Indian Hedgehog</td>
<td>Ligand</td>
<td>+ +</td>
<td>4 (Kobune et al., 2004)</td>
</tr>
<tr>
<td>3562</td>
<td>IL3</td>
<td>-</td>
<td>Ligand</td>
<td>+ +</td>
<td>7 (Bryder and Jacobsen, 2000)</td>
</tr>
<tr>
<td>3569</td>
<td>IL6</td>
<td>-</td>
<td>Ligand</td>
<td>+ +</td>
<td>31 (Üeda et al., 2000)</td>
</tr>
<tr>
<td>3572</td>
<td>IL6ST</td>
<td>gp130</td>
<td>Receptor</td>
<td>+ +</td>
<td>74 (Audet et al., 2001; Jenkins et al., 2005)</td>
</tr>
<tr>
<td>3624</td>
<td>INHBA</td>
<td>Activin A</td>
<td>Ligand</td>
<td>- -</td>
<td>23 (Utsugisawa et al., 2006)</td>
</tr>
<tr>
<td>3635</td>
<td>INPP5D</td>
<td>SHIP</td>
<td>Intracellular</td>
<td>+ -</td>
<td>29 (Desponts et al., 2006)</td>
</tr>
<tr>
<td>34561</td>
<td>IRGM</td>
<td>Lrg-47</td>
<td>Intracellular</td>
<td>+ -</td>
<td>0 (Feng et al., 2008)</td>
</tr>
<tr>
<td>3676</td>
<td>ITGA4</td>
<td>VLA-4</td>
<td>Receptor</td>
<td>+ -</td>
<td>21 (Priestley et al., 2006)</td>
</tr>
<tr>
<td>182</td>
<td>JAG1</td>
<td>-</td>
<td>Ligand</td>
<td>+ +</td>
<td>14 (Karanu et al., 2000)</td>
</tr>
<tr>
<td>3717</td>
<td>JAK2</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>126 (Zhao et al., 2002)</td>
</tr>
<tr>
<td>3726</td>
<td>JUNB</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>51 (Santaguida et al., 2009)</td>
</tr>
<tr>
<td>3815</td>
<td>KIT</td>
<td>c-kit</td>
<td>Receptor</td>
<td>+ +</td>
<td>80 (Zhao et al., 2002)</td>
</tr>
<tr>
<td>4254</td>
<td>KITLG</td>
<td>SCF</td>
<td>Ligand</td>
<td>+ +</td>
<td>17 (Kent et al., 2008b)</td>
</tr>
<tr>
<td>8379</td>
<td>MAD1L1</td>
<td>Mad-1</td>
<td>Intracellular</td>
<td>- -</td>
<td>43 (Walkley et al., 2005)</td>
</tr>
<tr>
<td>4170</td>
<td>MCL1</td>
<td>-</td>
<td>Intracellular</td>
<td>+ 0</td>
<td>46 (Opferman et al., 2005)</td>
</tr>
<tr>
<td>4211</td>
<td>MEIS1</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>83 (Wang et al., 2006a)</td>
</tr>
<tr>
<td>4297</td>
<td>MLL</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>35 (McMahon et al., 2007)</td>
</tr>
<tr>
<td>4318</td>
<td>MMP9</td>
<td>-</td>
<td>Ligand</td>
<td>- +</td>
<td>37 (Heissig et al., 2002)</td>
</tr>
<tr>
<td>4352</td>
<td>MPL</td>
<td>c-mpl</td>
<td>Receptor</td>
<td>+ +</td>
<td>9 (Yoshiiara et al., 2007)</td>
</tr>
<tr>
<td>4602</td>
<td>MYB</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>25 (Lei et al., 2005)</td>
</tr>
<tr>
<td>4609</td>
<td>MYC</td>
<td>c-Myc</td>
<td>Intracellular</td>
<td>+ +</td>
<td>334 (Laurenti et al., 2008; Satoh et al., 2004)</td>
</tr>
<tr>
<td>4613</td>
<td>MYCN</td>
<td>n-Myc</td>
<td>Intracellular</td>
<td>+ +</td>
<td>16 (Laurenti et al., 2008)</td>
</tr>
<tr>
<td>4851</td>
<td>NOTCH1</td>
<td>-</td>
<td>Ligand</td>
<td>+ -</td>
<td>103 (Lauret et al., 2004; Varnum-Finney et al., 2000)</td>
</tr>
<tr>
<td>4856</td>
<td>NOV</td>
<td>-</td>
<td>Ligand</td>
<td>+ -</td>
<td>14 (Gupta et al., 2007)</td>
</tr>
<tr>
<td>5087</td>
<td>PBX1</td>
<td>-</td>
<td>Intracellular</td>
<td>- -</td>
<td>50 (Cellot et al., 2007; Krosli et al., 2003b)</td>
</tr>
<tr>
<td>84108</td>
<td>PCGF2</td>
<td>Mel-18</td>
<td>Intracellular</td>
<td>- -</td>
<td>6 (Kajiume et al., 2004)</td>
</tr>
<tr>
<td>1911</td>
<td>PHC1</td>
<td>Rae28</td>
<td>Intracellular</td>
<td>+ +</td>
<td>18 (Ohta et al., 2002)</td>
</tr>
<tr>
<td>8842</td>
<td>PROM1</td>
<td>CD133</td>
<td>Receptor</td>
<td>+ 0</td>
<td>1 (Hess et al., 2006)</td>
</tr>
<tr>
<td>5728</td>
<td>PTEN</td>
<td>-</td>
<td>Intracellular</td>
<td>+ -</td>
<td>46 (Yilmaz et al., 2006; Zhang et al., 2006b)</td>
</tr>
<tr>
<td>9536</td>
<td>PTGES</td>
<td>Prostaglandin E2</td>
<td>Ligand</td>
<td>+ +</td>
<td>2 (North et al., 2007)</td>
</tr>
<tr>
<td>5741</td>
<td>PTH</td>
<td>Parathyroid Hormone</td>
<td>Ligand</td>
<td>+ +</td>
<td>11 (Calvi et al., 2003)</td>
</tr>
<tr>
<td>5781</td>
<td>PTPN11</td>
<td>Shp-2</td>
<td>Intracellular</td>
<td>+ -</td>
<td>166 (Chan et al., 2006)</td>
</tr>
<tr>
<td>5879</td>
<td>RAC1</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>62 (Gu et al., 2003)</td>
</tr>
<tr>
<td>5880</td>
<td>RAC2</td>
<td>-</td>
<td>Intracellular</td>
<td>+ +</td>
<td>129 (Gu et al., 2003)</td>
</tr>
<tr>
<td>5925</td>
<td>RB1</td>
<td>pRb</td>
<td>Intracellular</td>
<td>+ -</td>
<td>274 (Viatour et al., 2008)</td>
</tr>
<tr>
<td>5933</td>
<td>RBL1</td>
<td>p107</td>
<td>Intracellular</td>
<td>+ -</td>
<td>62 (Viatour et al., 2008)</td>
</tr>
<tr>
<td>5934</td>
<td>RBL2</td>
<td>p130</td>
<td>Intracellular</td>
<td>+ -</td>
<td>58 (Viatour et al., 2008)</td>
</tr>
<tr>
<td>861</td>
<td>RUNX1</td>
<td>AML-1</td>
<td>Intracellular</td>
<td>+ +</td>
<td>58 (Mulloy et al., 2003)</td>
</tr>
<tr>
<td>10019</td>
<td>SH2B3</td>
<td>Lnk</td>
<td>Intracellular</td>
<td>- -</td>
<td>16 (Ema et al., 2005; Velazquez et al., 2002)</td>
</tr>
<tr>
<td>6469</td>
<td>SHH</td>
<td>Sonic Hedgehog</td>
<td>Ligand</td>
<td>+ +</td>
<td>9 (Bhardwaj et al., 2001)</td>
</tr>
<tr>
<td>4089</td>
<td>SMAD4</td>
<td>-</td>
<td>Intracellular</td>
<td>+ 0</td>
<td>428 (Karlsson et al., 2007)</td>
</tr>
</tbody>
</table>
### Supplementary Table S29. Common stimulatory and inhibitory signalling molecules integrated into the HSC self-renewal network

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>SYMBOL</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STIMULATORY COMMON</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>673</td>
<td>BRAF</td>
<td>V-RAF MURINE SARCOMA VIRAL ONCOGENE HOMOLOG B1</td>
</tr>
<tr>
<td>5290</td>
<td>PIK3CA</td>
<td>PHOSPHOINOSITIDE-3-KINASE, CATALYTIC, ALPHA POLYPEPTIDE</td>
</tr>
<tr>
<td>5293</td>
<td>PIK3CD</td>
<td>PHOSPHOINOSITIDE-3-KINASE, CATALYTIC, DELTA POLYPEPTIDE</td>
</tr>
<tr>
<td>5294</td>
<td>PIK3CG</td>
<td>PHOSPHOINOSITIDE-3-KINASE, CATALYTIC, GAMMA POLYPEPTIDE</td>
</tr>
<tr>
<td>5335</td>
<td>PLCG1</td>
<td>PHOSPHOLIPASE C, GAMMA 1</td>
</tr>
<tr>
<td>5336</td>
<td>PLCG2</td>
<td>PHOSPHOLIPASE C, GAMMA 2 (PHOSPHATIDYLINOSITOL-SPECIFIC)</td>
</tr>
<tr>
<td>5578</td>
<td>PRKCA</td>
<td>PROTEIN KINASE C, ALPHA</td>
</tr>
<tr>
<td>5580</td>
<td>PRKCD</td>
<td>PROTEIN KINASE C, DELTA</td>
</tr>
<tr>
<td>5583</td>
<td>PRKCH</td>
<td>PROTEIN KINASE C, ETA</td>
</tr>
<tr>
<td>5584</td>
<td>PRKCI</td>
<td>PROTEIN KINASE C, IOTA</td>
</tr>
<tr>
<td><strong>INHIBITORY COMMON</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2885</td>
<td>GRB2</td>
<td>GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2</td>
</tr>
<tr>
<td>5604</td>
<td>MAP2K1</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE KINASE 1</td>
</tr>
<tr>
<td>5602</td>
<td>MAPK10</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE 10</td>
</tr>
<tr>
<td>5599</td>
<td>MAPK8</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE 8</td>
</tr>
<tr>
<td>5601</td>
<td>MAPK9</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE 9</td>
</tr>
<tr>
<td>6654</td>
<td>SOS1</td>
<td>SON OF SEVENLESS HOMOLOG 1 (DROSOPHILA)</td>
</tr>
<tr>
<td>6772</td>
<td>STAT1</td>
<td>SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 1, 91KDA</td>
</tr>
<tr>
<td><strong>COMMON TO STIMULATORY AND INHIBITORY SIGNALLING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>207</td>
<td>AKT1</td>
<td>V-AKT MURINE THYMOMA VIRAL ONCOGENE HOMOLOG 1</td>
</tr>
<tr>
<td>208</td>
<td>AKT2</td>
<td>V-AKT MURINE THYMOMA VIRAL ONCOGENE HOMOLOG 2</td>
</tr>
<tr>
<td>5594</td>
<td>MAPK1</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE 1</td>
</tr>
</tbody>
</table>

*SR* = effect on self-renewal; + stimulatory, - inhibitory

*Prolif* = effect on HSC proliferation; + stimulatory, - inhibitory, 0 no-effect

*K(i2D)* = protein interaction partners defined in the i2D database
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5595</td>
<td>MAPK3</td>
<td>MITOGEN-ACTIVATED PROTEIN KINASE 3</td>
</tr>
<tr>
<td>5894</td>
<td>RAF1</td>
<td>V-RAF-1 MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1</td>
</tr>
</tbody>
</table>
6 REFERENCES


Lajtha, L. G. (1971). Kinetics of the haemopoietic stem cells. Haematologia (Budap)


Punzel, M., Gupta, P., Roodell, M., Mortari, F., and Verfaillie, C. M. (1999). Factor(s) secreted by AFT024 fetal liver cells following stimulation with human cytokines are important for human LTC-IC growth. Leukemia 13, 1079-1084.


Wang, J. C., Doedens, M., and Dick, J. E. (1997). Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. Blood 89, 3919-3924.


7 Appendices
7.1 MATLAB Code

7.1.1 Cellular Balance ODEs

function [dX] = HSCgrowth20(t,X,P);
% hematopoietic cell growth rates, compartment model
% call syntax: dX = HSCgrowth(t,X)
% input variables to function: f_i, u_i
% inputs: X = [X(1);X(2);......X(18);X(19)]
% P = [g1;g2;r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
% output: dX = [dX(1);dX(2);......dX(20);dX(21)]

% set parameters
gr1 = P(1)*(t.^5)./(2^5 + t.^5);
gr2 = P(2)*(t.^5)./(2^5 + t.^5);
sr1 = P(3);
sr2 = P(4);
k1 = P(5);
k2 = P(6);
fmax = P(7);
dSR = P(8);
mmax = P(9);
dGR = P(10);
Ls2 = P(11);
ks2 = P(12);
sr3 = P(13);
k3 = P(14);
sr4 = P(15);
k4 = P(16);

% self-renewal probabilities
f(1) = fmax*(exp(-(0^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(2) = fmax*(exp(-(1^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(3) = fmax*(exp(-(2^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(4) = fmax*(exp(-(3^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(5) = fmax*(exp(-(4^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(6) = fmax*(exp(-(5^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
f(7) = fmax*(exp(-(6^2)/(2*dSR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23]^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);

% growth & differentiation rates
u(1) = gr1*(exp(-(1-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(2) = gr1*(exp(-(2-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(3) = gr1*(exp(-(3-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(4) = gr1*(exp(-(4-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
u(5) = gr1*(exp(-(5-nmax)^2)/(2*dGR^2)))*(1+X(25).^k4)/(1 +X(23).^k2 +X(25).^k4);
\[ u(13) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(14) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(15) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(16) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(17) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(18) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(19) = gr2 \circ (1/(1+X(22).^k1)); \]
\[ u(20) = gr2 \circ (1/(1+X(22).^k1)); \]

% differential equations of cell compartments
\[ dX(1) = (2*f(1)-1).*u(1).*X(1); \]
\[ dX(2) = 2*u(1).*(1-f(1)).*X(1) + (2*f(2)-1).*u(2).*X(2); \]
\[ dX(3) = 2*u(2).*(1-f(2)).*X(2) + (2*f(3)-1).*u(3).*X(3); \]
\[ dX(4) = 2*u(3).*(1-f(3)).*X(3) + (2*f(4)-1).*u(4).*X(4); \]
\[ dX(5) = 2*u(4).*(1-f(4)).*X(4) + (2*f(5)-1).*u(5).*X(5); \]
\[ dX(6) = 2*u(5).*(1-f(5)).*X(5) + (2*f(6)-1).*u(6).*X(6); \]
\[ dX(7) = 2*u(6).*(1-f(6)).*X(6) + (2*f(7)-1).*u(7).*X(7); \]
\[ dX(8) = 2*u(7).*(1-f(7)).*X(7) + (2*f(8)-1).*u(8).*X(8); \]
\[ dX(9) = 2*u(8).*(1-f(8)).*X(8) + (2*f(9)-1).*u(9).*X(9); \]
\[ dX(10) = 2*u(9).*(1-f(9)).*X(9) + (2*f(10)-1).*u(10).*X(10); \]
\[ dX(11) = 2*u(10).*(1-f(10)).*X(10) + (2*f(11)-1).*u(11).*X(11); \]
\[ dX(12) = 2*u(11).*(1-f(11)).*X(11) + (2*f(12)-1).*u(12).*X(12); \]
\[ dX(13) = 2*u(12).*(1-f(12)).*X(12) + (2*f(13)-1).*u(13).*X(13); \]
\[ dX(14) = 2*u(13).*(1-f(13)).*X(13) + (2*f(14)-1).*u(14).*X(14); \]
\[ dX(15) = 2*u(14).*(1-f(14)).*X(14) + (2*f(15)-1).*u(15).*X(15); \]
\[ dX(16) = 2*u(15).*(1-f(15)).*X(15) + (2*f(16)-1).*u(16).*X(16); \]
\[ dX(17) = 2*u(16).*(1-f(16)).*X(16) + (2*f(17)-1).*u(17).*X(17); \]
\[ dX(18) = 2*u(17).*(1-f(17)).*X(17) + (2*f(18)-1).*u(18).*X(18); \]
\[ dX(19) = 2*u(18).*(1-f(18)).*X(18) + (2*f(19)-1).*u(19).*X(19); \]
\[ dX(20) = 2*u(19).*(1-f(19)).*X(19) + (2*f(20)-1).*u(20).*X(20); \]
\[ dX(21) = 0; \]

% differential equations of secreted factors
% Interacting FB loops (Wichman and Loeffler model)

% SF1 secretion rate ~ lin+  F1 (lin+) inhibits lin- u (~k1) & induces SF2 secretion (~ks, Ls)
\[ dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)); \]

% SF2 secretion rate ~ SF1 lin-  F2 (lin-) inhibits HSC f (~k2)
\[ dX(23) = sr2*((X(22).^ks2)/(Ls2^ks2+(X(22).^ks2))).*(X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13)); \]

% SF3 secretion rate ~ lin+  F3 (lin+) induces u (~k3)
\[ dX(24) = sr3.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)); \]

% SF4 secretion rate ~ lin+  F4 (lin+) induces f (~k4)
\[ dX(25) = sr4.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)); \]

% Output vector
\[ dX = [dX(1);dX(2);dX(3);dX(4);dX(5);dX(6);dX(7);dX(8);dX(9);dX(10);dX(11);dX(12);dX(13);dX(14);dX(15);dX(16);dX(17);dX(18);dX(19);dX(20);dX(21);dX(22);dX(23);dX(24);dX(25)]; \]

7.1.2 Cell balance solver; 8days NSNE condition

function [X,OUTnsne] = HSCgrowthsolverNSNE20(Xi,P)
%ODE solver for HSCgrowth; 21 compartment unilineage model using bioprocess
% in cell selection every 4d
% culture time, sampling rate must be set internally
% input variables: Xi = [Xi(1);Xi(2);Xi(3);......Xi(20);Xi(21);F1;F2;F3;F4]
% P = [gr1;gr2;sv22;sv23;k22;k23] parameter vector
% output variables: X = [X(1);X(2);....X(25)] output cell population
% OUTnn = [%lin+ cellX CFCX LTCICX LTHSCX]
% Call syntax: Bioprocess8d(Xi,P)
% make OUTs a globally accessible variable for use in optimization function
% global OUTsne global OUTnsnet global t global LIN
global TNC
global CFC
global LTCIC
global SRC

%Integrate function from [0 4] called "a"
[t,Xa] = ode23(@(t,X)HSCgrowth20(t,X,P), [0:0.5:4], X0);

%cell populations a
X1a = Xa(:,1);
X2a = Xa(:,2);  %X2 = ST-SRC
X3a = Xa(:,3);  %X3 = MPP
X4a = Xa(:,4);
X5a = Xa(:,5);
X6a = Xa(:,6);
X7a = Xa(:,7);
X8a = Xa(:,8);  %LTC-IC
X9a = Xa(:,9);
X10a = Xa(:,10);
X11a = Xa(:,11);  %CFC
X12a = Xa(:,12);
X13a = Xa(:,13);  %lin-
X14a = Xa(:,14);
X15a = Xa(:,15);
X16a = Xa(:,16);
X17a = Xa(:,17);
X18a = Xa(:,18);
X19a = Xa(:,19);
X20a = Xa(:,20);  %mature cells
X21a = Xa(:,21);

%secreted factors
X22a = Xa(:,22);
X23a = Xa(:,23);
X24a = Xa(:,24);
X25a = Xa(:,25);

ra = size(Xa);  %determines dimensions of Xa [lenght width]
ka = ra(1,1);  %ka = length of Xa (total numer of t iterations)

Xbi =
1.0*[X1a(ka);X2a(ka);X3a(ka);X4a(ka);X5a(ka);X6a(ka);X7a(ka);X8a(ka);X9a(ka);X10a(ka);X11a(ka);X12a(ka);X13a(ka);X14a(ka);X15a(ka);X16a(ka);X17a(ka);X18a(ka);X19a(ka);X20a(ka);X21a(ka);X22a(ka);X23a(ka);X24a(ka);X25a(ka)];

%Integrate after selection
[tb,Xb] = ode23(@(t,X)HSCgrowth20(t,X,P),[4:0.5:8], Xbi);

%cell populations b
X1b = Xb(:,1);
X2b = Xb(:,2);  %X2 = ST-SRC
X3b = Xb(:,3);  %X3 = MPP
X4b = Xb(:,4);
X5b = Xb(:,5);
X6b = Xb(:,6);
X7b = Xb(:,7);
X8b = Xb(:,8);
X9b = Xb(:,9);
X10b = Xb(:,10);
X11b = Xb(:,11);
X12b = Xb(:,12);
X13b = Xb(:,13);
X14b = Xb(:,14);
X15b = Xb(:,15);
X16b = Xb(:,16);
X16b = Xb(:,16);
X17b = Xb(:,17);
X18b = Xb(:,18);
X19b = Xb(:,19);
X20b = Xb(:,20);
\[ X_{21b} = X_{b(:,21)}; \]

%secreted factors
\[ X_{22b} = X_{b(:,22)}; \]
\[ X_{23b} = X_{b(:,23)}; \]
\[ X_{24b} = X_{b(:,24)}; \]
\[ X_{25b} = X_{b(:,25)}; \]

%total populations
\[ X_1 = [X_{1a};X_{1b}]; \]
\[ X_2 = [X_{2a};X_{2b}]; \]
\[ X_3 = [X_{3a};X_{3b}]; \]
\[ X_4 = [X_{4a};X_{4b}]; \]
\[ X_5 = [X_{5a};X_{5b}]; \]
\[ X_6 = [X_{6a};X_{6b}]; \]
\[ X_7 = [X_{7a};X_{7b}]; \]
\[ X_8 = [X_{8a};X_{8b}]; \]
\[ X_9 = [X_{9a};X_{9b}]; \]
\[ X_{10} = [X_{10a};X_{10b}]; \]
\[ X_{11} = [X_{11a};X_{11b}]; \]
\[ X_{12} = [X_{12a};X_{12b}]; \]
\[ X_{13} = [X_{13a};X_{13b}]; \]
\[ X_{14} = [X_{14a};X_{14b}]; \]
\[ X_{15} = [X_{15a};X_{15b}]; \]
\[ X_{16} = [X_{16a};X_{16b}]; \]
\[ X_{17} = [X_{17a};X_{17b}]; \]
\[ X_{18} = [X_{18a};X_{18b}]; \]
\[ X_{19} = [X_{19a};X_{19b}]; \]
\[ X_{20} = [X_{20a};X_{20b}]; \]
\[ X_{21} = [X_{21a};X_{21b}]; \]
\[ X_{22} = [X_{22a};X_{22b}]; \]
\[ X_{23} = [X_{23a};X_{23b}]; \]
\[ X_{24} = [X_{24a};X_{24b}]; \]
\[ X_{25} = [X_{25a};X_{25b}]; \]

\[ t = [ta;tb]; \]

%total populations
\[ X_{T1} = X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}+X_{13}; \]
\[ X_{T2} = X_{14}+X_{15}+X_{16}+X_{17}+X_{18}+X_{19}+X_{20}; \]
\[ TNC = X_{T1}+X_{T2}; \]
\[ CFC = X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}; \]
\[ LTCIC = X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8; \]
\[ LTSRC = X_1; \]
\[ STSRC = X_2+X_3; \]
\[ LIN = 100.*((X_{T2}./(X_{T1}+X_{T2}))); \]
\[ SR = X_{20}; \]
\[ GR = X_{21}; \]

%secreted factor concentrations
\[ F_1 = X_{22}; \]
\[ F_2 = X_{23}; \]
\[ F_3 = X_{24}; \]
\[ F_4 = X_{25}; \]

%fold expansion of cell populations
\[ r = \text{size}(X_1); \]
\[ k = r(1,1); \]
\[ \text{lin} = X_{T2}(k)/\text{TNC}(k)*100; \]
\[ \text{TNCX} = (X_1(k)+X_2(k)+X_3(k)+X_4(k)+X_5(k)+X_6(k)+X_7(k)+X_8(k)+X_9(k)+X_{10}(k)+X_{11}(k)+X_{12}(k)+X_{13}(k)+X_{14}(k)+X_{15}(k)+X_{16}(k)+X_{17}(k)+X_{18}(k)+X_{19}(k)+X_{20}(k))/(X_1(1)+X_2(1)+X_3(1)+X_4(1)+X_5(1)+X_6(1)+X_7(1)+X_8(1)+X_9(1)+X_{10}(1)+X_{11}(1)+X_{12}(1)+X_{13}(1)+X_{14}(1)+X_{15}(1)+X_{16}(1)+X_{17}(1)+X_{18}(1)+X_{19}(1)+X_{20}(1)); \]
\[ \text{CFCX} = (X_1(k)+X_2(k)+X_3(k)+X_4(k)+X_5(k)+X_6(k)+X_7(k)+X_8(k)+X_9(k)+X_{10}(k)+X_{11}(k))/(X_1(1)+X_2(1)+X_3(1)+X_4(1)+X_5(1)+X_6(1)+X_7(1)+X_8(1)+X_{10}(1)+X_{11}(1)+X_{12}(1)+X_{13}(1)+X_{14}(1)+X_{15}(1)+X_{16}(1)+X_{17}(1)+X_{18}(1)+X_{19}(1)+X_{20}(1)); \]
\[ \text{LTCICX} = (X_1(k)+X_2(k)+X_3(k)+X_4(k)+X_5(k)+X_6(k)+X_7(k)+X_8(k))/(X_1(1)+X_2(1)+X_3(1)+X_4(1)+X_5(1)+X_6(1)+X_7(1)+X_8(1)+X_{10}(1)+X_{11}(1)+X_{12}(1)+X_{13}(1)+X_{14}(1)+X_{15}(1)+X_{16}(1)+X_{17}(1)+X_{18}(1)+X_{19}(1)+X_{20}(1)); \]
\[ \text{LTHSCX} = X_1(k)/(X_1(1)); \]
7.1.3 Cell Balance Solver; 8 days, NS/E condition

function [X,OUTnsne] = HSCgrowthsolverNSE20(Xi,P)
%ODE solver for HSCgrowth no lin+ removal, ME at d4; 20 compartment model
%culture time, sampling rate must be set internally
%input variables:  
%Xi = [Xi(1);Xi(2);Xi(3);......Xi(20);Xi(21);F1;F2;F3]
%Call syntax: HSCgrowthsolverNSE(Xi,P)
%Same as HSCgrowthsolverNSNE20(Xi,P), replace reselection with:
Xbi =
1.0*[X1a(ka);X2a(ka);X3a(ka);X4a(ka);X5a(ka);X6a(ka);X7a(ka);X8a(ka);X9a(ka);X10a(ka);X11a(ka);X12a(ka);X13a(ka);X14a(ka);X15a(ka);X16a(ka);X17a(ka);X18a(ka);X19a(ka);X20a(ka);0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0]
%set input for 4-8 integration

7.1.4 Cell Balance Solver; 8 days, S/NE condition

function [X,OUTsne] = HSCgrowthsolverSNE20(Xi,P)
%ODE solver for HSCgrowth with lin+ removal at d4, no ME; 20 compartment model
%culture time, sampling rate must be set internally
%input variables:  
%Xi = [Xi(1);Xi(2);Xi(3);......Xi(20);Xi(21);F1;F2;F3;F(1)]
%Call syntax: HSCgrowthsolverNSE(Xi,P)
%Same as HSCgrowthsolverNSNE20(Xi,P), replace reselection with:
Xbi =
1.0*[X1a(ka);X2a(ka);X3a(ka);X4a(ka);X5a(ka);X6a(ka);X7a(ka);X8a(ka);X9a(ka);X10a(ka);X11a(ka);X12a(ka);X13a(ka);X14a(ka);X15a(ka);X16a(ka);X17a(ka);X18a(ka);X19a(ka);X20a(ka);0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0]
%set input for 4-8 integration

7.1.5 Cell Balance Solver; 8 days, S/E condition

function [X,OUTsne] = HSCgrowthsolverSE20(Xi,P)
%ODE solver for HSCgrowth with lin+ removal at d4, no ME; 20 compartment model
%culture time, sampling rate must be set internally
%input variables:  
%Xi = [Xi(1);Xi(2);Xi(3);......Xi(20);Xi(21);F1;F2;F3;F(1)]
%Call syntax: HSCgrowthsolverNSE(Xi,P)
%Same as HSCgrowthsolverNSNE20(Xi,P), replace reselection with:
Xbi =
1.0*[X1a(ka);X2a(ka);X3a(ka);X4a(ka);X5a(ka);X6a(ka);X7a(ka);X8a(ka);X9a(ka);X10a(ka);X11a(ka);X12a(ka);X13a(ka);X14a(ka);X15a(ka);X16a(ka);X17a(ka);X18a(ka);X19a(ka);X20a(ka);0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0]
%set input for 4-8 integration

7.1.6 Objective function for Parameter estimation – NS/NE, NS/E, S/NE, & S/E data

function [SS] = Sumsquares20(P)
%runs HSCgrowthsolver and Bioprocess8d and compares simulated TNC, CFC, 
%LTCIC, SRC expansions and %lin+ to GJM thesis data
%input variable: P = estimated parameter vector
% P = [gr1,gr2,sv22,sv23,k22,k23]
%Call syntax: Sumsquares(P)

%make OUTnn, OUTse a globally accessible variables
global OUTnsne
global OUTnse
global OUTsne
global OUTse

%input lin- cell population
A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0;0];

%run functions
HSCgrowthsolverNSNE20(A,P);
HSCgrowthsolverNSE20(A,P);
HSCgrowthsolverSNE20(A,P);
HSCgrowthsolverSE20(A,P);

%Calculate Sum of Squares differential
SS = 1*(30-OUTnsne(1))^2/30^2 + 1*((23-OUTnsne(2))^2)/23^2 + 1*((17-OUTnsne(3))^2)/17^2 + 1*((6.6-OUTnsne(4))^2)/6.6^2 + 1*((1-OUTnsne(5))^2)/1^2 + ... + 1*((33-OUTse(2))^2)/33^2 + 1*((21-OUTse(3))^2)/21^2 + 1*((14.6-OUTse(4))^2)/14.6^2 + 1*((5-OUTse(5))^2)/5^2;

7.1.7 Objective Function; 8-week culture

function [SS] = Sumsquares56(P)
%runs Bioprocess56d and compares simulated TNC, CFC, and SRC expansions to JED data
%input variable: P = estimated parameter vector
% P = [gr1,gr2,sv22,sv23,k22,k23]
%Call syntax: Sumsquares(P)

%make OUTnn, OUTse a globally accessible variables
global OUT56

%input lin- cell population
A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0;0];

%Calculate Sum of Squares differential
SS = ((6-OUT56(5,1))^2/6^2 + (25-OUT56(5,2))^2/25^2 + ... + (4-OUT56(3,1))^2/4^2 + (0-OUT56(3,2))^2/0^2 + (0-OUT56(3,3))^2/0^2 + ... + (1-OUT56(1,1))^2/1^2 + (0-OUT56(1,2))^2/0^2 + (0-OUT56(1,3))^2/0^2 + (0-OUT56(1,4))^2/0^2 + (0-OUT56(1,5))^2/0^2 + (0-OUT56(1,6))^2/0^2 + (0-OUT56(1,7))^2/0^2 + (0-OUT56(1,8))^2/0^2;

7.1.8 Random Parameter Space Search

function [OUTnsne] = HSCRandP
%Randomaly searches parameter space
%Runs HSCgrowthNSNE20 algorithm over parameters uniformly distributed within defined bounds
%call syntax: HSCRandP

A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0];
for i = [1:1000]
R = rand(1,16); % uniform [0 1] random number generator
P(1) = 1 + (10-1)*R(1);
P(2) = 0.1 + (1-0.1)*R(2);
P(3) = 1e-7 + (1e-5)*(100.^R(3)/100); % creates log-uniform distribution
P(4) = 1e-7 + (1e-5)*(100.^R(4)/100);
P(5) = 1 + (5-1)*R(5);
P(6) = 1 + (5-1)*R(6);
P(7) = 0.5 + (0.8-0.5)*R(7);
P(8) = 0.5 + (5-0.5)*R(5);
P(9) = 3 + (8-3)*R(9);
P(10) = 3 + (10-3)*R(10);
P(11) = 1e-4 + (1e-2 - 1e-4)*100.*R(11)/100;
P(12) = 0.1 + (5-0.1)*100.*R(12)/100;
P(13) = 1e-7 + (1e-5)*(100.^R(13)/100);
P(14) = 0.05 + (1-0.05)*R(14);
P(15) = 1e-7 + (1e-5)*(100.^R(15)/100);
P(16) = 0.05 + (1-0.05)*R(16);
end

% P = [P(1);P(2);P(3);P(4);P(5);P(6);P(7);P(8);P(9);P(10);P(11);P(12);P(13);P(14);P(15);P(16)];
global OUTnsne
HSCgrowthsolverNSNE20(A,P)
OUTnsne;
LIN(i,1) = OUTnsne(1);
TNCX(i,1) = OUTnsne(2);
CFCX(i,1) = OUTnsne(3);
LTCICX(i,1) = OUTnsne(4);
SRCX(i,1) = OUTnsne(5);
P1(i,1) = P(1);
P2(i,1) = P(2);
P3(i,1) = P(3);
P4(i,1) = P(4);
P5(i,1) = P(5);
P6(i,1) = P(6);
P7(i,1) = P(7);
P8(i,1) = P(8);
P9(i,1) = P(9);
P10(i,1) = P(10);
P11(i,1) = P(11);
P12(i,1) = P(12);
P13(i,1) = P(13);
P14(i,1) = P(14);
P15(i,1) = P(15);
P16(i,1) = P(16);
end

% write results to excel file
xlswrite('Outputdata.xls', LIN, 'Total', 'A2')
xlswrite('Outputdata.xls', TNCX, 'Total', 'B2')
xlswrite('Outputdata.xls', CFCX, 'Total', 'C2')
xlswrite('Outputdata.xls', LTCICX, 'Total', 'D2')
xlswrite('Outputdata.xls', SRCX, 'Total', 'E2')
xlswrite('Outputdata.xls', P1, 'Total', 'F2')
xlswrite('Outputdata.xls', P2, 'Total', 'G2')
xlswrite('Outputdata.xls', P3, 'Total', 'H2')
xlswrite('Outputdata.xls', P4, 'Total', 'I2')
xlswrite('Outputdata.xls', P5, 'Total', 'J2')
xlswrite('Outputdata.xls', P6, 'Total', 'K2')
xlswrite('Outputdata.xls', P7, 'Total', 'L2')
xlswrite('Outputdata.xls', P8, 'Total', 'M2')
xlswrite('Outputdata.xls', P9, 'Total', 'N2')
xlswrite('Outputdata.xls', P10, 'Total', 'O2')
xlswrite('Outputdata.xls', P11, 'Total', 'P2')
xlswrite('Outputdata.xls', P12, 'Total', 'Q2')
xlswrite('Outputdata.xls', P13, 'Total', 'R2')
xlswrite('Outputdata.xls', P14, 'Total', 'S2')
xlswrite('Outputdata.xls', P15, 'Total', 'T2')
7.1.9 Randomize parameters

function [OUTnsne] = HSCVarP(mu)
%Runs 8-day NS/NE culture simulations
%16 parameters normally distributed with CV = "mu" around Po vector input
%call syntax: HSCVarP(mu)

A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0];
Po = [4.999048088 0.224787811 1.86E-05 6.78E-05 0.465365035 0.222517874 0.8 2.718383148 5.821336742 3.25185841 0.057917522 7.85E-06 0.486402199 2.05E-06 0.35546875];

global OUTnsne

for i = [1:50]
R = 1+mu*randn(1,16);  %random vector, mean = 1, stdv = mu
% set all R(j) > 0
if R(1) < 0
    R(1) = 0
else R(1) = R(1)
end
if R(2) < 0
    R(2) = 0
else R(2) = R(2)
end
if R(3) < 0
    R(3) = 0
else R(3) = R(3)
end
if R(4) < 0
    R(4) = 0
else R(4) = R(4)
end
if R(5) < 0
    R(5) = 0
else R(5) = R(5)
end
if R(6) < 0
    R(6) = 0
else R(6) = R(6)
end
if R(7) < 0
    R(7) = 0
else R(7) = R(7)
end
if R(8) < 0
    R(8) = 0
else R(8) = R(8)
end
if R(9) < 0
    R(9) = 0
else R(9) = R(9)
end
if R(10) < 0
    R(10) = 0
else R(10) = R(10)
end
if R(11) < 0
    R(11) = 0
else R(11) = R(11)
end
if R(12) < 0
    R(12) = 0
else R(12) = R(12)
end
if R(13) < 0
    R(13) = 0
else R(13) = R(13)
end

% code to use these R values in simulation

end
else R(13) = R(13)\end{align*}
end
\begin{align*}
\text{if } R(14) < 0 \\
R(14) = 0 \end{align*}
\begin{align*}
\text{else } R(14) = R(14) \end{align*}
\begin{align*}
\text{end}
\begin{align*}
\text{if } R(15) < 0 \\
R(15) = 0 \end{align*}
\begin{align*}
\text{else } R(15) = R(15) \end{align*}
\begin{align*}
\text{end}
\begin{align*}
\text{if } R(16) < 0 \\
R(16) = 0 \end{align*}
\begin{align*}
\text{else } R(16) = R(16) \end{align*}
\begin{align*}
P(1) = P(1) \times R(1); \quad &\% gr1 \\
P(2) = P(2) \times R(2); \quad &\% gr2 \\
P(3) = P(3) \times R(3); \quad &\% sr1 \\
P(4) = P(4) \times R(4); \quad &\% sr2 \\
P(5) = P(5) \times R(5); \quad &\% k1 \\
P(6) = P(6) \times R(6); \quad &\% k2 \\
P(7) = P(7) \times R(7); \quad &\% fmax \\
P(8) = P(8) \times R(8); \quad &\% dSR \\
P(9) = P(9) \times R(9); \quad &\% max \\
P(10) = P(10) \times R(10); \quad &\% dGR \\
P(11) = P(11) \times R(11); \quad &\% Ls2 \\
P(12) = P(12) \times R(12); \quad &\% ks2 \\
P(13) = P(13) \times R(13); \quad &\% sr3 \\
P(14) = P(14) \times R(14); \quad &\% k3 \\
P(15) = P(15) \times R(15); \quad &\% sr4 \\
P(16) = P(16) \times R(16); \quad &\% k4
\end{align*}

7.1.10 Phase Portrait Generator

function \([\text{OUTnsnet}] = \text{PhasePort}(n, Popt)\)
\% PhasePort writes Phase portrait time courses to excel sheet:
\% "Output_data_PhasePort.xls"
\% Syntax: PhasePort(n, Popt) with \(n\) = number of phases, \(Popt = [P1; P2; \ldots; P16]\)
\% set \(\delta t = 0.5\) in HSCgrowthsolverNSNE19 (i.e. \(t = [0:0.5:8]\))

global OUTnsnet
A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0]; \% input cell population (lin-)
Plo = [1; 0.1; 1E-07; 1E-07; 0; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1];
Phi = [10; 1; 5E-05; 5E-05; 5E-05; 5E-05; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0];
\%
for \(k = [1:1:16];\)
\%
S = sparse(k,1,1,16,1); \quad \% sparse vector, all 0s except \(S(k) = 1\)
Px = Popt - S.*Popt; \quad \% creates Popt vector with Popt(k) = 0
for \(i = [1:n]\)
Pz = Px + S.*(Plo + ((deltaP.*i^2)/(n^2))); % creates Pz vector with all Ppot values, except P(k) = Plo + n*delta

HSCgrowthSolverNSNE20(A,Pz);

LIN(:,i) = OUTnsnet(:,2);
TNC(:,i) = (1/8191)*OUTnsnet(:,3);
CFC(:,i) = (1/2047)*OUTnsnet(:,4);
LTCIC(:,i) = (1/11)*OUTnsnet(:,5);
SRC(:,i) = (1/1)*OUTnsnet(:,6);
end

xlswrite('Outputdata_PhasePort.xls', LIN, k, 'B2')   %write matrix to file "outputdata.xls"
xlswrite('Outputdata_PhasePort.xls', TNC, k, 'B22')
xlswrite('Outputdata_PhasePort.xls', CFC, k, 'B42')
xlswrite('Outputdata_PhasePort.xls', LTCIC, k, 'B62')
xlswrite('Outputdata_PhasePort.xls', SRC, k, 'B82')

end

7.1.11 Local Parameter Sensitivity Analysis

function [LPSM] = LPSA(delta)
%LPSA performs LPSA output toward OUTnsne vectors to excel spreadsheet
%delta = fold change in parameter
%call syntax LPSA(delta)
global OUTnsne
A = [1;2;4;8;16;32;64;128;256;512;1024;2048;4096;0;0;0;0;0;0;0;0;0;0;0;0];   %input cell population (lin-)
Pin = [3.37;0.168;6.92E-06;6.51E-06;1.61;1.58;0.574;3.41;6.11;2.81;0.0099;4.68;8.47E-06;0.553451056;3.40E-06;0.905];
HSCgrowthSolverNSNE20(A,Pin);
OUTi = OUTnsne;                         %baseline culture output @ Pin
for k = [1:1:16];
    S = sparse(k,1,1,16,1);                 %sparse vector with all zeros except 1 at position k
    PS = Pin.*S;                            %sparse vector with all zeros except Pin(k) at position k
    Pk = Pin + (delta -1)*PS;                %Parameter vector with position k = delta*P(k)
    HSCgrowthSolverNSNE19(A,Pk);            %culture output with delta*parameter(k)
    OUTo = OUTnsne;
    LIN(k,1) = (OUTo(1)- OUTi(1))/(OUTi(1)*(delta-1));      %Sensitivity coefficient for each observable
    TNCX(k,1) = (OUTo(2)- OUTi(2))/(OUTi(2)*(delta-1));
    CFCX(k,1) = (OUTo(3)- OUTi(3))/(OUTi(3)*(delta-1));
    LTCICX(k,1) = (OUTo(4)- OUTi(4))/(OUTi(4)*(delta-1));
    SRCX(k,1) = (OUTo(5)- OUTi(5))/(OUTi(5)*(delta-1));
end
LPSM = [LIN TNCX CFCX LTCICX SRCX]          %Sensitivity coefficinets compiled into matrix
xlswrite('Outputdata.xls', LIN, 'LPSA', 'A2')   %write matrix to file "outputdata.xls"
xlswrite('Outputdata.xls', TNCX, 'LPSA', 'B2')
xlswrite('Outputdata.xls', CFCX, 'LPSA', 'C2')
xlswrite('Outputdata.xls', LTCICX, 'LPSA', 'D2')
xlswrite('Outputdata.xls', SRCX, 'LPSA', 'E2')

end

7.1.12 Bootstrap Algorithm

function[Xnsne,Xnse,Xsne,Xse] = Bootstrap
%Create synthetic dataset (n=50) for 8-day culture: NS/NE, NS/E, S/NE, S/E
%based on experimental avg and std
OUTnsne = [26 22.6 17.3 6.6 1.0];
STDnsne = [12 9.4 9.8 3.2 2.0];
OUTnse = [26 25.5 12.2 6.6 1.0];
STDnse = [12 9.7 0.9 2.0];

OUTsne = [26 17.4 10.2 3.7 1.0];
STDsne = [12 9.9 8.7 0.9 2.0];

OUTse = [26 32.4 20.7 14.6 4.0];
STDse = [12 13.5 12.8 2.7 2.0];

for i = [1:50];
    for k = [1:5];
        R1 = randn(1,5);
        R2 = randn(1,5);
        R3 = randn(1,5);
        R4 = randn(1,5);
        Xnsne(i,k) = OUTnsne(1,k) + R1(1,k).*STDnsne(1,k);
        Xnse(i,k) = OUTnse(1,k) + R2(1,k).*STDnse(1,k);
        Xsne(i,k) = OUTsne(1,k) + R3(1,k).*STDsne(1,k);
        Xse(i,k) = OUTse(1,k) + R4(1,k).*STDse(1,k);
    end
end

xlswrite('Bootstrap.xls', Xnsne, 'NSNE', 'B2')
xlswrite('Bootstrap.xls', Xnse, 'NSE', 'B2')
xlswrite('Bootstrap.xls', Xsne, 'SNE', 'B2')
xlswrite('Bootstrap.xls', Xse, 'SE', 'B2')

7.1.13 Fed Batch, constant dilution rate cell balance

function [dX] = HSCgrowthFedBatch(t,X,P,D);
    % hematopoietic cell growth rates, compartment model, fedbatch operation
    % Inputs:
    %       t = [t init: delta t: t final]
    %       X = [X(1);X(2);......X(18);X(19)]
    %       P = [gr1;gr2;r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
    % Outputs:
    %       dX = [dX(1);dX(2);dX(3);dX(4);dX(5);dX(6);dX(7);dX(8);dX(9);dX(10);dX(11);dX(12);dX(13);dX(14);dX(15);dX(16);dX(17);dX(18);dX(19);dX(20);dX(21)]

    % SF1 secretion rate ~ lin+          F1 (lin+) inhibits lin- u (~k1) & induces SF2 secretion (~ks, Ls)
    dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+D*t);

    % SF2 secretion rate ~ SF1.lin- F2 (lin-) inhibits HSPC f (~k2)
    dX(23) = sr2*((X(22).^ks2)/(Ls2^ks2+(X(22).^ks2))).*(X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13))/(1+D*t);

    % SF3 secretion rate ~ lin+          F3 (lin+) induces u (~k3)
    dX(24) = sr3.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+D*t);

    % SF4 secretion rate ~ lin+          F4 (lin+) induces f (~k4)
    dX(25) = sr4.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+D*t);

    % Output vector
    dX = [dX(1);dX(2);dX(3);dX(4);dX(5);dX(6);dX(7);dX(8);dX(9);dX(10);dX(11);dX(12);dX(13);dX(14);dX(15);dX(16);dX(17);dX(18);dX(19);dX(20);dX(21);dX(22);dX(23);dX(24);dX(25)];

7.1.14 Fed Batch, proportional dilution rate cell balance

function [dX] = HSCgrowthFedBatchVD(t,X,P,D);
    % hematopoietic cell growth rates, compartment model, fedbatch operation
    % Variable Dilution rate = D*V(t)
%call syntax: dX = HSCgrowthFedBatch(t,X,P,D)
%where
%Inputs:    t = [t_init: delta_t: t_final]
%           X = [X(1);X(2);......X(18);X(19)]
%           P = [gr1;gr2;1r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
%Output:    dX = [dX(1);dX(2);......dX(20);dX(21)]

%same as HSCgrowthFedBatch, except alter SF1-4 concentrations;

%dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(exp(D*t));
%SF1 secretion rate ~ lin+         F1 (lin+) inhibits lin- u (~k1) & induces SF2 secretion (~ks, Ls)
%SF2 secretion rate ~ SF1.lin-     F2 (lin-) inhibits HSPC f (~k2)
%dX(23) = sr2*((X(22).^ks2)/(Ls2^ks2+(X(22).^ks2))).*(X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13))/(exp(D*t));
%SF3 secretion rate ~ lin+         F3 (lin+) induces u (~k3)
%dX(24) = sr3.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(exp(D*t));
%SF4 secretion rate ~ lin+         F4 (lin+) induces f (~k4)
%dX(25) = sr4.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(exp(D*t));

7.1.15 Perfusion cell balance

function [dX] = HSCgrowthPerfuse(t,X,P,D);
%hematopoietic cell growth rates, compartment model
%call syntax: dX = HSCgrowthPerfuse(t,X,P,D)
%where
%Inputs:    t = [t_init: delta_t: t_final]
%           X = [X(1);X(2);......X(18);X(19)]
%           P = [gr1;gr2;1r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
%Output:    dX = [dX(1);dX(2);......dX(20);dX(21)]

%same as HSCgrowthFedBatch, except alter SF1-4 concentrations;

%dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)) - D*X(22);
%SF2 secretion rate ~ SF1.lin-     F2 (lin-) inhibits HSPC f (~k2)
%dX(23) = sr2*((X(22).^ks2)/(Ls2^ks2+(X(22).^ks2))).*(X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13)) - D*X(23);
%SF3 secretion rate ~ lin+         F3 (lin+) induces u (~k3)
%dX(24) = sr3.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)) - D*X(24);
%SF4 secretion rate ~ lin+         F4 (lin+) induces f (~k4)
%dX(25) = sr4.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20)) - D*X(25);

7.1.16 Feed-back controlled fed batch cell balance based on SF2 concentration

function [dX] = HSCgrowthFBControlSF(t,X,P,Kp,SF2max)
%hematopoietic cell growth rates, compartment model, Feedback control,
%Variable Dilution rate = D*V(t)
%call syntax: dX = HSCgrowthFBControlSF(t,X,P,Kp,SF2max)
%where
%Inputs:    t = [t_init: delta_t: t_final]
%           X = [X(1);X(2);......X(19);X(20)]
%           P = [gr1;gr2;1r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
%           Kp = Proportional controller gain
%           SF2max = max [SF2] setpoint
%Output:    dX = [dX(1);dX(2);......dX(20);dX(21)]

%same as HSCgrowthFedBatch, except utlize extra term X21 to alter SF1-4 concentrations;

Feed rate; dX21 = dV
%Set Proportional Controller Constant
%XT = X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13)+X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20);
%Feedback control on SF1 = X(22)
if X(22) <= SF1max
dX(21) = 0;
else
dx(21) = Kp*(X(22) - SF1max);
end

%differential equations of secreted factors
%Interacting FB loops (Wichman and Loeffler model)

%SF1 secretion rate ~ lin+   F1 (lin+) inhibits lin- u (~k1) & induces SF2 secretion (~ks, Ls)
dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+X(21));

%SF2 secretion rate ~ SF1.lin- F2 (lin-) inhibits HSPC f (~k2)
dX(23) = sr2*((X(22).^ks2)/(Ls2^ks2+(X(22).^ks2)).*(X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13))/(1+X(21));

%SF3 secretion rate ~ lin+   F3 (lin+) induces u (~k3)
dX(24) = sr3.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+X(21));

%SF4 secretion rate ~ lin+   F4 (lin+) induces f (~k4)
dX(25) = sr4.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+X(21));

7.1.17 **Feed-back controlled Fed Batch cell balance based on total cell density**

function [dx] = HSCgrowthFBControlX(t,X,P, Kp, Xmax)
%hematopoietic cell growth rates, compartment model, Feedback control,
%Variable Dilution rate = D*V(t)
%call syntax: dX = HSCgrowthFBControlSF(t,X,P,Kp,SF2max)
%where
%Inputs:    t = [t_init: delta_t: t_final]
%           X = [X(1);X(2);......X(19);X(20)]
%           P = [gr1;gr2;r1(sv22);r2(sv23);k1(k22);k2(k23);k3(k24)]
%           Kp = Proportional controller gain
%           Xmax = cell density set-point as a multiple of input (10^5)
%Output:    dX = [dX(1);dX(2);........dX(20);dX(21)]
%same as HSCgrowthFBControlSF except alter SF concentrations;
%Feed rate: dX21 = dV/dt
%Total cell number
XT = X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)+X(9)+X(10)+X(11)+X(12)+X(13)+X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20);
%total cell density
Xd = XT/(1 + X(21));
%Cell density set point related to Lin- input (sum(A)=8191)
Xset = Xmax*8191;
%FB control algorithm; density based
if Xd <= Xset
dx(21) = 0;
else
dx(21) = Kp*(Xd - Xset);
end
%differential equations of secreted factors
%Interacting FB loops (Wichman and Loeffler model)

%SF1 secretion rate ~ lin+   F1 (lin+) inhibits lin- u (~k1) & induces SF2 secretion (~ks, Ls)
dX(22) = sr1.*(X(14)+X(15)+X(16)+X(17)+X(18)+X(19)+X(20))/(1+X(21));

%SF2 secretion rate ~ SF1.lin- F2 (lin-) inhibits HSPC f (~k2)
dX(23) =
\[
sr2^* \left( \frac{X(22)^{ks2}}{Ls2^{ks2} + X(22)^{ks2}} \right)^{*} \left( X(1) + X(2) + X(3) + X(4) + X(5) + X(6) + X(7) + X(8) + X(9) + X(10) + X(11) + X(12) + X(13) \right)/(1 + X(21));
\]

\[
\%SF3 secretion rate \sim \text{lin}^+ \quad \text{F3} (\text{lin}^+) \text{ induces } u (-k3)
\]
\[
dX(24) = sr3^* \left( X(14) + X(15) + X(16) + X(17) + X(18) + X(19) + X(20) \right)/(1 + X(21));
\]

\[
\%SF4 secretion rate \sim \text{lin}^+ \quad \text{F4} (\text{lin}^+) \text{ induces } l (-k4)
\]
\[
dX(25) = sr4^* \left( X(14) + X(15) + X(16) + X(17) + X(18) + X(19) + X(20) \right)/(1 + X(21));
\]

### 7.2 Colony Pictures

![Figure 7-1. 8-day cultured cells at 4x (left) and 40x (right) magnification.](image1)

![Figure 7-2. BFU-E at 4x magnification.](image2)
Figure 7-3. CFU-E at 4x (left) and 40x (right) magnification.

Figure 7-4. CFU-G at 4x (left) and 40x (right) magnification.

Figure 7-5. CFU-M at 4x (left) and 40x (right) magnification.
Figure 7-6. CFU-Meg at 4x (left) and 40x (right) magnification.

Figure 7-7. CFU-MEGG at 4x magnification.

Figure 7-8. Mixed colonies at 4x magnification.
Figure 7-9. LTCIC wells at 4x (left) and 40x (right) magnification.

Figure 7-10. LTCIC bioassay at 4x magnification; negative (left) and positive (right).

Figure 7-11. LTCIC bioassay at 40x magnification; negative (left) and positive (right).
Copyright Acknowledgements

Portions of chapter 1 have been published in *Cell Stem Cell* (Kirouac and Zandstra, 2008) and *Current Opinion in Biotechnology* (Kirouac and Zandstra, 2006), co-authored by Peter Zandstra. Authorization to reproduce this work has been obtained from the publisher and co-author.

Portions of chapter 2 have been published in *Molecular Systems Biology* (Kirouac et al., 2009a), co-authored by Gerard Madlambayan, Mei Yu, Edward Sykes, Caryn Ito, and Peter Zandstra. Authorization to reproduce this work has been obtained from the publisher and co-authors.

Portions of chapter 3 have been submitted to *Cell* (Kirouac et al., 2009b), co-authored by Caryn Ito, Aline Roch, Mei Yu, Edward Sykes, Gary Bader, and Peter Zandstra. Authorization to reproduce this work has been obtained from the publisher and co-authors.

Figure 1-8 is reproduced from *Oncogene* (Larsson and Karlsson, 2005). Authorization to reproduce this image has been obtained from the publisher.