SUPPLEMENTARY METHODS AND MATERIALS

Functional assays used to measure stem and progenitor cells

In the absence of reliable phenotypic markers, functional assays are used to retrospectively identify and quantify hematopoietic stem and progenitor cells. In vitro, the colony forming cell (CFC) and long term culture-initiating cell (LTCIC) assays quantitatively measure cellular proliferation and differentiation, and output colony formation at 2 weeks and 7 weeks, for committed and primitive progenitors respectively (Eaves et al., 1992). The ability to reconstitute and sustain multi-lineage hematopoiesis *in vivo* is the gold standard, and most clinically relevant, stem cell assay. A mouse (or other animal) is given a sub-lethal dose of irradiation which partially ablates the bone marrow, followed by infusion of test cells. Six or more weeks later, mice are sacrificed and the marrow is analyzed for donor cell content and composition via flow cytometry. For human cells, the xenotransplant model using immune-deficient (NOD/SCID, NOD/SCID-$\beta2m^{-/-}$, or NOD/SCID-$\gamma_c^{-/-}$) mice is referred to as the *Scid*-mouse repopulating cell (SRC) assay. Murine cells are similarly assayed a congenic strain [the competitive repopulating cell assay (CRA)] (Coulombel, 2004). While exceedingly useful in the fundamental understanding of adult stem cell biology, the retrospective nature of these analyses is a significant limitation.

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 (described below), 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 S1A**, 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 4-parameter logistic curves (r^2 = 0.97, 0.60, 0.92, and 0.88 for *CD34*, *CD133*, *CD38*, and *ABC-B1*/Rho123 respectively), likely due to a detection limits and saturation effects at low and high gene expression levels respectively.

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 as described below. The majority of secreted proteins scored Present at the transcript level were not detected at protein level in conditioned media. 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 S1B**). 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.

Conditioned media proteome analysis – Raybiotech^{TM} antibody arrays

Conditioned media samples corresponding to each culture condition and time point used for microarray profiling (d4, d8-SE, d8-NSNE, d12, and additionally d16) were assayed in duplicate using the Raybio Human Cytokine Array C Series 2000 (Raybiotech Inc., Norcross, GA, USA) following the manufacturers instructions. The arrays contain antibody spots against a set of 120 cytokines, chemokines, proteases, and soluble receptors, functioning as a multiplex sandwich ELISA. 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.

Conditioned Media Proteome analysis – Luminex^{TM} Liquid Chips

Conditioned media samples were also assayed in triplicate 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 (Kirouac et al., 2009).

Conditioned media serotonin analysis – ELISA

To validate predictions from the pathway enrichment analysis of the GeneChip data, conditioned media samples were tested in triplicate for serotonin (5HT) using quantitative ELISA kits (ALPCO, Salem NH, USA) according to manufacturers' instructions.

Phenotypic analysis

Analysis of cell surface expression was accomplished by suspending $5x10^4$ 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). 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.

Description of microarray datasets

The microarray datasets used to extract the 55 cell type-characteristic gene sets (**Table S1**) 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 Hematopoetic fingerprints database 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) (Dennis et al., 2003), resulting in slightly greater than 50% mapping efficiency. These were compiled to produce a non-redundant list of 5390 Affymetrix probesets (genes), each assigned membership in ≥ 1 gene set. From this mapping, we computed the distribution of gene-gene set memberships, and pair-wise fractional gene overlap between gene sets (defined as the number of shared genes / total number of genes in each gene set).

Microarray data processing

Arrays were normalized via smoothing spline-least squares regression, artefacts removed, and expression indices (Probe Match/Mismatch; 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 $\geq 10\%$

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. 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.

Inter-cellular network reconstruction process

The following Boolean logic operations were performed on each ligand (L) receptor (R) pair:

TRUE *IF* L = P *AND* L (PM/MM) > 50 *AND* [R_1 OR R_2 OR….R_N = P]

Where R_1, R_2, …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).

An excel file is provided (**Supplementary excel worksheet**) which semi-automates the reconstruction process using Affymetrix HU133 expression profiles in assigning combinatorial ligand-receptor interactions. By pasting a matrix of expression indices for specific cell populations corresponding to the secreted factor (SF_A) and receptor (R_A) probeset IDs, subsequent worksheets calculate unique expression indices for each gene (SF_B and R_B), assign combinatorial secreted factor-receptor interactions (R_C and R_D) and call for the presence/absence (1/0) of secreted factor-receptor interactions based on user defined expression thresholds (SF_Call and R_Call).

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 erythrocblasts 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-SE population (our data) as well as the erythrocblast 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 represented as a directed graph.

Hypergeometric Z-score calculations

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-NSNE, d8-SE, 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 $r = 4$.

To 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 othologues; $N = 20177$, $R = 15$, $n = 1728$, and $r = 6$.

Characterization of cell cycle and apoptosis of CD34$^+$ and CD34$^-$ cells in culture

UCB Lin$^-$ cells were labelled with 10uM Carboxyfluorescein succinimidyl ester **(**CFSE; Invitrogen) for 10 min at 37˚C. Cells were then quenched on ice with HBSS + 20% FBS and washed 3 times. CSFE-stained cells were sorted for the middle 50% of the CFSE peak using a FACS ARIA sorter (BD Biosciences, Franklin Lakes, NJ, USA) to tighten the peak distribution. Supplements were added to baseline cell culture media, and cultures performed as described previously. Samples were taken for cell counting and flow cytometry analysis on day 2, 4, 6, and 8 of culture. CFSE labelled cells were stained for CD34-APC (Beckman Coulter), and unlabelled cells co-stained for CD34-APC and Annexin V-FITC (Biovision, Mountain View, CA, USA) as described above. Samples were analyzed on a FACS CANTO Instrument (BD Biosciences) and data analysis performed with FACSDiva software (BD Biosciences). For CFSE staining, up to 9 peaks (representing successive cell divisions) were discernible by day-8 of culture. Gates were set around peaks and the fraction of total cells within each gate quantified. The average generation number of the cultured cell populations was then calculated as the number of cell divisions × fraction of the population within the given division.

Model description

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. 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) as depicted in **Figure S4A**. The cellular growth rate for compartment *i* is given by the equation:

$$\frac{dX_i}{dt} = (1 - f_{i-1})u_{i-1}X_{i-1} + (2f_i - 1)u_i X_i \qquad \text{for } i = [1, 2, 3, \ldots., n] \qquad (1)$$

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^+.

The cell-level kinetic parameters u_i and f_i are not constant, but functions of differentiation status, time in culture, and secreted molecule-mediated inter-cellular networks. We used Gaussian-type functions to describe kinetic variables as functions of compartment number (*i*), a Hill-type function to introduce a lag-phase and therefore an explicit time (*t*) dependency, and coupled Hill-type functions to incorporate the effects of secreted regulatory factor concentrations (*SF1-4*), represented schematically in **Figure S4B**. The use of well mixed, liquid suspension cultures allows for the assumption of spatial homogeneity, thus all cells would be exposed to an identical microenvironment. The resulting master equations define u_i and f_i respectively as functions of compartment number (*i*), time (*t*) and secreted factor concentrations (*SF1-4*):

$$u_i = u_{MAX} \exp\left[\frac{-(i - n_{MAX})^2}{2D_{GR}{}^2}\right] \times \left(\frac{t^{kt}}{\tau_D{}^{kt} + t^{kt}}\right) \times \left(\frac{1 + [SF3]^{K3}}{1 + [SF1]^{K1} + [SF3]^{K3}}\right) \qquad (2)$$

$$f_i = f_{MAX} \exp\left[\frac{-(i - 1)^2}{2D_{SR}{}^2}\right] \times \left(\frac{1 + [SF4]^{K4}}{1 + [SF2]^{K2} + [SF4]^{K4}}\right) \qquad (3)$$

The secreted factors (*SF1-4*) thus producing an inter-cellular communication network, structured as a coupled positive-negative feedback control circuit between the Lin$^+$ and Lin$^-$ cell populations. Stem and progenitor cell population dynamics are thus dependent upon the composition and functional activity of differentiated cells. The model contains 16 free parameters, inaccessible to experimental measurement. These were therefore estimated using a hybrid genetic algorithm and data from (Madlambayan et al., 2005) as a training set. The resulting values are shown in the table below

P	*DESCRIPTION*	*UNITS*	*EST*
u_{MAX}	Maximum proliferation rate of lin$^-$ cells	day^{-1}	6.26×10^0
u_+	Maximum proliferation rate of lin$^+$ cells	day^{-1}	2.04×10^{-1}
n_{MAX}	Compartment with maximal proliferation	-	5.32×10^0
D_{GR}	Proliferative decay term	-	3.38×10^0
f_{MAX}	Self-renewal probability of LT-HSC	-	6.34×10^{-1}
D_{SR}	Self-renewal decay term	-	1.96×10^0
sr1	Secretion rate of *SF1*	pg/cell.day	2.37×10^{-5}
sr2	Secretion rate of *SF2*	pg/cell.day	2.93×10^{-5}
sr3	Secretion rate of *SF3*	pg/cell.day	5.96×10^{-6}
sr4	Secretion rate of *SF4*	pg/cell.day	5.30×10^{-6}
k1	Hill coefficient for *SF1* (Equation 14)	-	6.14×10^{-1}
k2	Hill coefficient for *SF2* (Equation 15)	-	5.55×10^{-1}
k3	Hill coefficient for *SF3* (Equation 14)	-	6.25×10^{-1}
k4	Hill coefficient for *SF4* (Equation 15)	-	5.33×10^{-1}
Ls	[*SF1*] inducing ½ maximal *SF2* secretion	pg/ml	9.15×10^{-1}
ks	Hill coefficient for *SF1* (Equation 9)	-	1.08×10^0

See Kirouac et al. (2009) for a more thorough description, analyses, and comparison to experimental data.

Model-based classification of ligand and small molecule functional activities

This is essentially a model discrimination problem, wherein rather than testing alternative model structures or parameter sets, we are tuning individual model variables [input *SF1-4* concentrations, or baseline self-renewal (f_{MAX}) and proliferation (u_{MAX}) rates] to fit experimental data (culture supplementation with ligands or small molecule supplementation). While the problem is not amenable to standard statistical tests of significance, we can use the well established model discrimination metric, the Akaike Information Criterion (*AIC*) to compare model variations (Kreutz and Timmer, 2009). As the number of parameters is conserved, the difference in AIC values used for comparing model variations (ΔAIC) is reduced to comparing the Weighted Residual Sum of Squares (*WRSS*), defined as:

$$WRSS = \sum_{i=1}^{N} w_i \left(\frac{E_i - S_i}{STD_i}\right) \qquad (4)$$

For N observables (in this case $N = 3$; the % change in TNC, CFC, and LTCIC expansion over 8-days in culture induced by media supplements), wherein E_i = experimental observation, S_i = simulated observation, and STD_i = standard deviation of experimental measurement, and w_i = weighting term, chosen as the $P\text{-value}_i^{-1}$, thus weighting by statistical significance (Landaw and DiStefano, 1984). To functionally classify ligands as specific model variables (*SF1-4*) we first simulated TNC, CFC, and LTCIC growth over 8-day cultures over a range of *SF1-4* input concentrations, thus producing 4 theoretical dose-response curves. We then calculated the *WRSS* for each experimentally tested ligand across each theoretical dose response (*SF1-4* = 0.1-10 × ED_{50}), identified the minimum value ($WRSS_{MIN}$), and divided this by the control WRSS ($SF1\text{-}4_{t=0} = 0$; $WRSS_{control}$), producing a term which linearly correlates with ΔAIC, defined as $\Delta WRSS$:

$$\Delta WRSS = \frac{WRSS_{MIN}}{WRSS_{control}} \qquad (5)$$

If $\Delta WRSS < 1$, assignment of the experimentally tested ligand as the given *SF* fits the data better than control, where minimum values represent the best classification.

Analogously, to classify the small molecule kinase inhibitors as stimulators vs. inhibitors of self-renewal or proliferation, TNC, CFC, and LTCIC growth was simulated over a range of self-renewal (f_{MAX}) and proliferation (u_{MAX}) rates. We then calculated the $\Delta WRSS$ values for each of the five kinase inhibitors against $|u_{MAX}|$ and $|f_{MAX}| > 10\%$ control.

Robustness of statistical gene set enrichment to PPI confidence filters

Many algorithms are available for connecting seed nodes (Huang and Fraenkel, 2009; Pinkert et al., 2010; Yosef et al., 2009)**,** however the utility of various alternatives against our approach in this case is not obvious. Our objective was to look for statistical enrichment of target genes (common signalling molecules) interacting with known HSC self-renewal modulators expressed in the cultured Lin$^-$ cells. The most straight forward approach was therefore to filter for expressed genes, and perform a search for interactions in a PPI meta-database (i2D). While PPI databases are notoriously noisy (Cusick et al., 2009), for our purpose we are not focused on validating specific interactions, but rather using the networks for statistical analysis which should be robust to a limited amount of false positives. To test the robustness of our statistical analysis against false positives, we decided to test the algorithm against a series of increasingly stringently filtered PPI networks. Intuitively, interactions represented in more than one database carry more confidence (Ramirez et al., 2007). We define the "edge weight" as the number of databases (represented in i2D) a given edge is represented in. As shown in **Figure S6A**, the distribution of edge weights for the "active self-renewal" PPI network follows an approximate scale-free distribution (almost 70% of edges found only once, while 1 edge is represented 30 times). We then filtered the network for edge weights of at least 1 (all edges), 2, 3, 4 and 5, and scored the resultant networks for enrichment of the target gene set using the Hypergeometric Z-Score. Networks filtered for edge weights of greater than 1, 2, and 3 are all significantly enriched for the common signalling molecules (*P-value* = 3.5×10^{-9}, 10^{-4}, and 10^{-6} respectively), while networks filtered on edge weights of 4 and up become increasingly small to detect statistically relevant numbers of target genes (predicted gene overlap < 1 for networks filtered on edges weights of 4 and above) (**Figure S6B**). This, while the liberal PPI networks used may have high error rates, the results of our analysis are robust to increasingly stringent confidence filters.

SUPPLEMENTARY REFERENCES

Assou, S., Le Carrour, T., Tondeur, S., Strom, S., Gabelle, A., Marty, S., Nadal, L., Pantesco, V., Reme, T., Hugnot, J.P*., et al.* (2007). A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells *25*, 961-973.

Chambers, S.M., Boles, N.C., Lin, K.Y., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C*., et al.* (2007). Hematopoietic Fingerprints: An Expression Database of Stem Cells and Their Progeny. Cell Stem Cell *1*, 578-591.

Coulombel, L. (2004). Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. Oncogene *23*, 7210-7222.

Cusick, M.E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A.R., Simonis, N., Rual, J.F., Borick, H., Braun, P., Dreze, M*., et al.* (2009). Literature-curated protein interaction datasets. Nat Methods *6*, 39-46.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Eaves, C.J., Sutherland, H.J., Udomsakdi, C., Lansdorp, P.M., Szilvassy, S.J., Fraser, C.C., Humphries, R.K., Barnett, M.J., Phillips, G.L., and Eaves, A.C. (1992). The human hematopoietic stem cell in vitro and in vivo. Blood Cells *18*, 301-307.

Ferrari, F., Bortoluzzi, S., Coppe, A., Basso, D., Bicciato, S., Zini, R., Gemelli, C., Danieli, G.A., and Ferrari, S. (2007). Genomic expression during human myelopoiesis. BMC Genomics *8*, 264.

Ge, X., Yamamoto, S., Tsutsumi, S., Midorikawa, Y., Ihara, S., Wang, S.M., and Aburatani, H. (2005). Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics *86*, 127-141.

Graham, S.M., Vass, J.K., Holyoake, T.L., and Graham, G.J. (2007). Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources. Stem Cells *25*, 3111-3120.

Huang, S.S., and Fraenkel, E. (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Sci Signal *2*, ra40.

Kirouac, D.C., Madlambayan, G.J., Yu, M., Sykes, E.A., Ito, C., and Zandstra, P.W. (2009). Cell-cell interaction networks regulate blood stem and progenitor cell fate. Mol Syst Biol *5*, 293.

Komor, M., Guller, S., Baldus, C.D., de Vos, S., Hoelzer, D., Ottmann, O.G., and Hofmann, W.K. (2005). Transcriptional profiling of human hematopoiesis during in vitro lineage-specific differentiation. Stem Cells *23*, 1154-1169.

Kreutz, C., and Timmer, J. (2009). Systems biology: experimental design. FEBS J *276*, 923-942.

Landaw, E.M., and DiStefano, J.J., 3rd (1984). Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. Am J Physiol *246*, R665-677.

Li, Q., Cai, H., Liu, Q., and Tan, W.S. (2006). Differential gene expression of human CD34+ hematopoietic stem and progenitor cells before and after culture. Biotechnol Lett *28*, 389-394.

Madlambayan, G.J., Rogers, I., Kirouac, D.C., Yamanaka, N., Mazurier, F., Doedens, M., Casper, R.F., Dick, J.E., and Zandstra, P.W. (2005). Dynamic changes in cellular and microenvironmental composition can be controlled to elicit in vitro human hematopoietic stem cell expansion. Exp Hematol *33*, 1229-1239.

Pinkert, S., Schultz, J., and Reichardt, J. (2010). Protein interaction networks--more than mere modules. PLoS Comput Biol *6*, e1000659.

Ramirez, F., Schlicker, A., Assenov, Y., Lengauer, T., and Albrecht, M. (2007). Computational analysis of human protein interaction networks. Proteomics *7*, 2541-2552.

Schadt, E.E., Li, C., Su, C., and Wong, W.H. (2000). Analyzing high-density oligonucleotide gene expression array data. J Cell Biochem *80*, 192-202.

Uchida, N., Combs, J., Chen, S., Zanjani, E., Hoffman, R., and Tsukamoto, A. (1996). Primitive human hematopoietic cells displaying differential efflux of the rhodamine 123 dye have distinct biological activities. Blood *88*, 1297-1305.

Wagner, W., Ansorge, A., Wirkner, U., Eckstein, V., Schwager, C., Blake, J., Miesala, K., Selig, J., Saffrich, R., Ansorge, W.*, et al.* (2004). Molecular evidence for stem cell function of the slow-dividing fraction among human hematopoietic progenitor cells by genome-wide analysis. Blood *104*, 675-686.

Yosef, N., Ungar, L., Zalckvar, E., Kimchi, A., Kupiec, M., Ruppin, E., and Sharan, R. (2009). Toward accurate reconstruction of functional protein networks. Mol Syst Biol *5*, 248.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. mRNA expression indices correlate with cell surface expression and secretion of proteins.

(A) 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. Four parameter logistic curves were fit to the individual data sets with r^2 values as indicated on the figure of 0.97, 0.60, 0.88, and 0.92 respectively. **(B)** 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. Z-scores of ± 2 are indicated by dashed lines, corresponding to enrichment / depletion *P*-values < 0.05.

Figure S2. Activity scores for 55 published gene sets averaged across 10 Experimental Samples.

Activity scores were computed for the 55 cell type-characteristic gene sets described in **Table S1** for each of the 10 profiles. Average Activity Scores across the 10 populations ± standard deviations are shown for each gene set.

Figure S3. Gene overlap between 55 published gene sets.

(A) Assignment of individual genes into 1 or more of the 55 characteristic gene sets. The distribution of gene set membership follows an approximate scale-free distribution; the majority of genes (87%) being gene set-specific and only 0.003% members of 4 different sets. **(B)** Pair-wise comparison of fractional gene overlap between all 55 gene sets, represented as a square matrix color coded from 0% to 100% overlap.

Figure S4. **Quantitative ELISA measurements of serotonin (5HT) in conditioned media.**

Error bars represent standard deviation (std), *n* = 3.

Figure S5. Functional effects of serotonin (5HT1), the TGF-β inhibitor SB505124, and select combinatorial ligand stimulation on culture output.

8-day fold expansion of total cells (TNC), progenitors (CFC), and primitive progenitors (LTCIC) from liquid cultures supplemented with serotonin (5HT) **(A)** or SB505124 **(B)** compared to control cultures. While results are not statistically significant ($p \geq 0.1$) performing more replicate experiments would enhance the statistical power and may add confidence to the results. **(C)** Select stimulations with multiple stimulatory (EGF, HGF, VEGF) and inhibitory (CCL3, CCL4, CXCL8, CXCL10) lignads reveal non-linear combinatorial effects, as the functional effects of single ligands are reduced by co-stimulation. Error bars = std, $n = 3$ **(A, C)** and $n = 7$ **(B)**.

Figure S6. Schematic representation of co-culture bioassay workflow.

Representative FACS sorting strategy and colony readouts. Blue arrows indicate work flow.

Figure S7. Tracking proliferation and apoptosis in CD34$^+$ cells throughout culture.

(A) Representative CFSE and AnnexinV fluorescence distributions for CD34$^+$ cells over 8 days in culture. **(B)** Quantification of the effects of TGFB2, CCL4, and VEGF stimulation on the average generation number (derived from CFSE plots; **i**) and % AnnexinV$^+$ cells (**ii**) at days 2, 4, and 6 in comparison to control culture.

Figure S8. Schematic representation of mathematical model structure and use in classifying endogenous ligand functional activities.

(A) The hematopoietic hierarchy is represented as a series of discrete cellular compartments. The number of cells in each compartment (X_i) is determined by the balance of cells entering due to differentiation from the previous compartment (X_{i-1}), cells leaving due to differentiation and cell death, and cell amplification due to self-renewal. **(B)** Compartment-specific proliferation rates and self-renewal probabilities are modulated by the balance of stimulatory (*SF3*, *SF4*) and inhibitory (*SF1*, *SF2*) regulatory factors secreted in lineage-specific patterns by differentiated cells, forming a coupled positive-negative feedback circuit. **(C)** Differential Weighted Residual Sum of Squares (ΔWRSS) for each secreted factor (*SF1-4*) supplement simulations vs. control for 10 endogenous ligands with significant effects, as well as the small molecules serotonin (5HT) and the TGF-βR inhibitor SB505124. Values below 1 fit the data better than control, where minimum values (indicated with asterisks) represent the best classification.

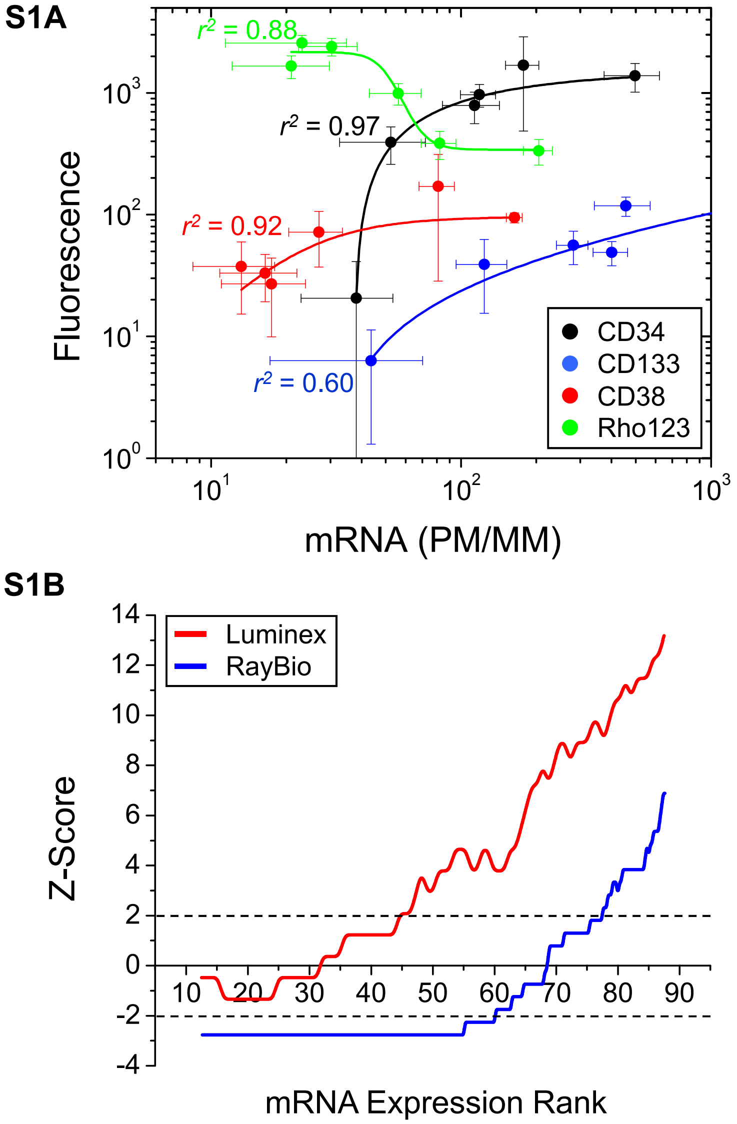
Figure S9. Reconstructed intra-cellular self-renewal signalling network.

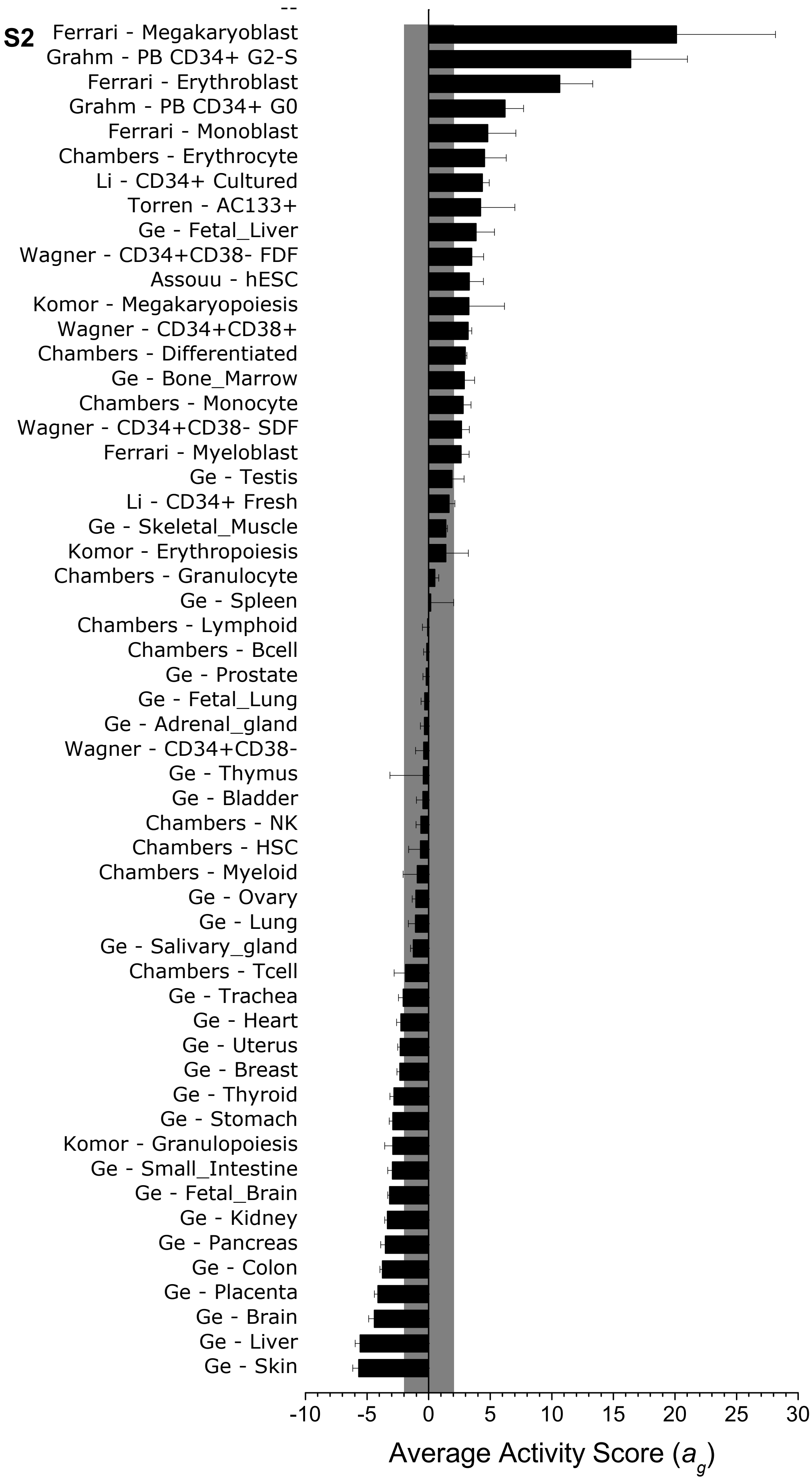
(A) A curated list of 112 genes known to effect HSC self-renewal was searched against the i2D protein interaction database for binding partners, resulting in a densely connected network of 2131 vertices and 5431 (non-unique) edges. For clearer visualization of the individual genes, the network was filtered 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.

Figure S10. Effect of "edge weight" filtering on statistical enrichment of gene sets.

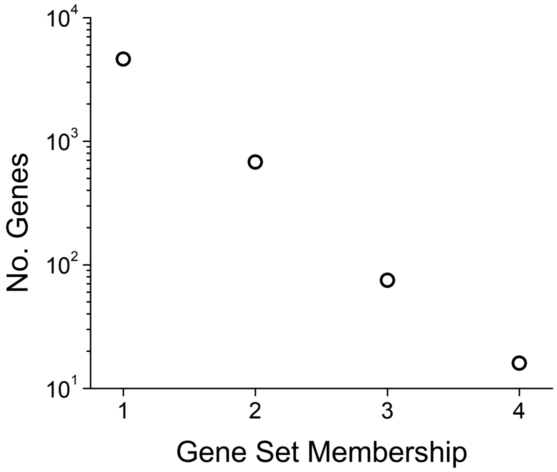
(A) Approximate scale-free distribution of edge weights (defined as independently reported interactions in the i2D database) in the intra-cellular self-renewal network. **(B)** Statistical enrichment of the common signal transduction molecules in networks filtered for edge weights of greater than 1 to 5. Z-Scores > 2 correspond to P-values < 0.05. Numbers indicate the number of common signal transduction molecules present in the various networks / expected by chance.

Figure S11. Classification of 5 kinase inhibitors functional activities via Differential Weighted Residual Sum of Squares (ΔWRSS). ΔWRSS for simulations of inhibition vs. stimulation of self renewal and proliferation, compared to by control for the 5 kinase inhibitor treatments. Values below 1 fit data better than control, where minimum values (indicated with asterisks) represent the best classification.


S1A
r^2 = 0.88
r^2 = 0.97
r^2 = 0.92
r^2 = 0.60
Fluorescence
10^0
10^1
10^2
10^3
10^1
10^2
10^3
mRNA (PM/MM)
● CD34
● CD133
● CD38
● Rho123
S1B
Luminex
RayBio
Z-Score
14
12
10
8
6
4
2
0
-2
-4
10
20
30
40
50
60
70
80
90
mRNA Expression Rank

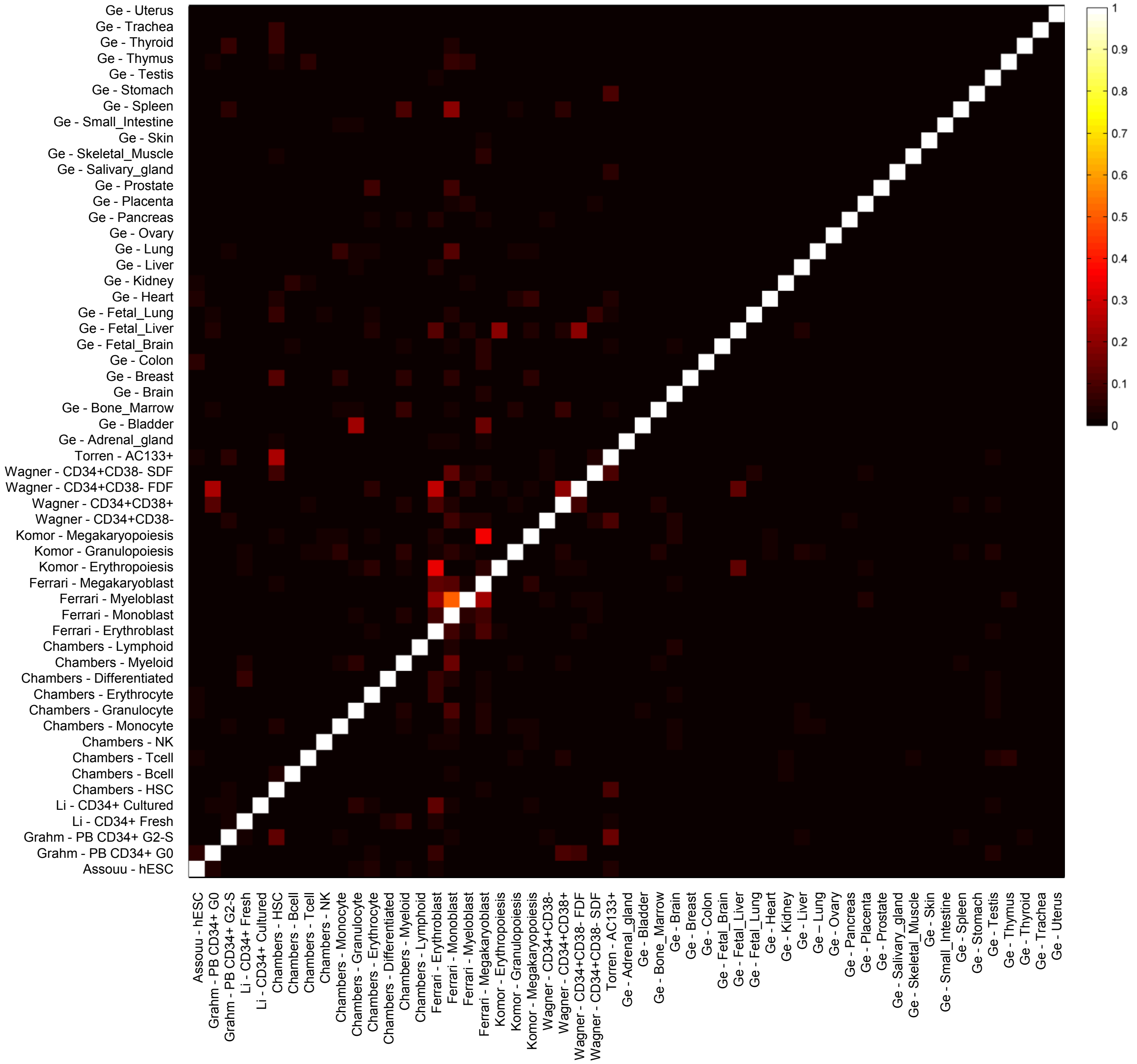

S2
--
Ferrari - Megakaryoblast
Grahm - PB CD34+ G2-S
Ferrari - Erythroblast
Grahm - PB CD34+ G0
Ferrari - Monoblast
Chambers - Erythrocyte
Li - CD34+ Cultured
Torren - AC133+
Ge - Fetal_Liver
Wagner - CD34+CD38- FDF
Assouu - hESC
Komor - Megakaryopoiesis
Wagner - CD34+CD38+
Chambers - Differentiated
Ge - Bone_Marrow
Chambers - Monocyte
Wagner - CD34+CD38- SDF
Ferrari - Myeloblast
Ge - Testis
Li - CD34+ Fresh
Ge - Skeletal_Muscle
Komor - Erythropoiesis
Chambers - Granulocyte
Ge - Spleen
Chambers - Lymphoid
Chambers - Bcell
Ge - Prostate
Ge - Fetal_Lung
Ge - Adrenal_gland
Wagner - CD34+CD38-
Ge - Thymus
Ge - Bladder
Chambers - NK
Chambers - HSC
Chambers - Myeloid
Ge - Ovary
Ge - Lung
Ge - Salivary_gland
Chambers - Tcell
Ge - Trachea
Ge - Heart
Ge - Uterus
Ge - Breast
Ge - Thyroid
Ge - Stomach
Komor - Granulopoiesis
Ge - Small_Intestine
Ge - Fetal_Brain
Ge - Kidney
Ge - Pancreas
Ge - Colon
Ge - Placenta
Ge - Brain
Ge - Liver
Ge - Skin
-10
-5
0
5
10
15
20
25
30
Average Activity Score (a_g)

S3A



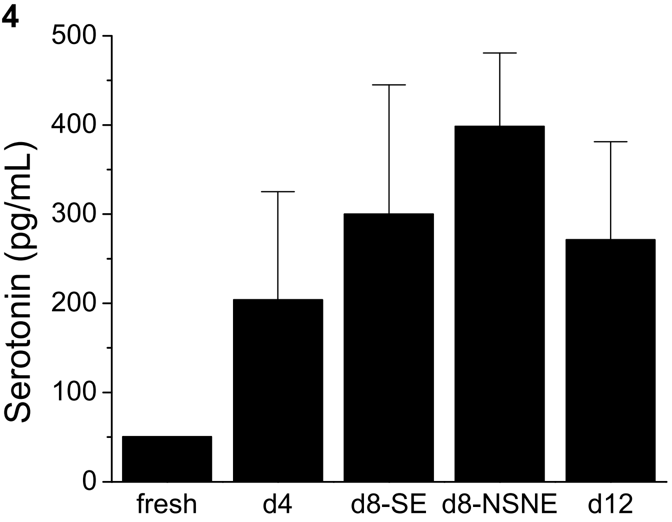
No. Genes
10^4
10^3
10^2
10^1
1
2
3
4
Gene Set Membership

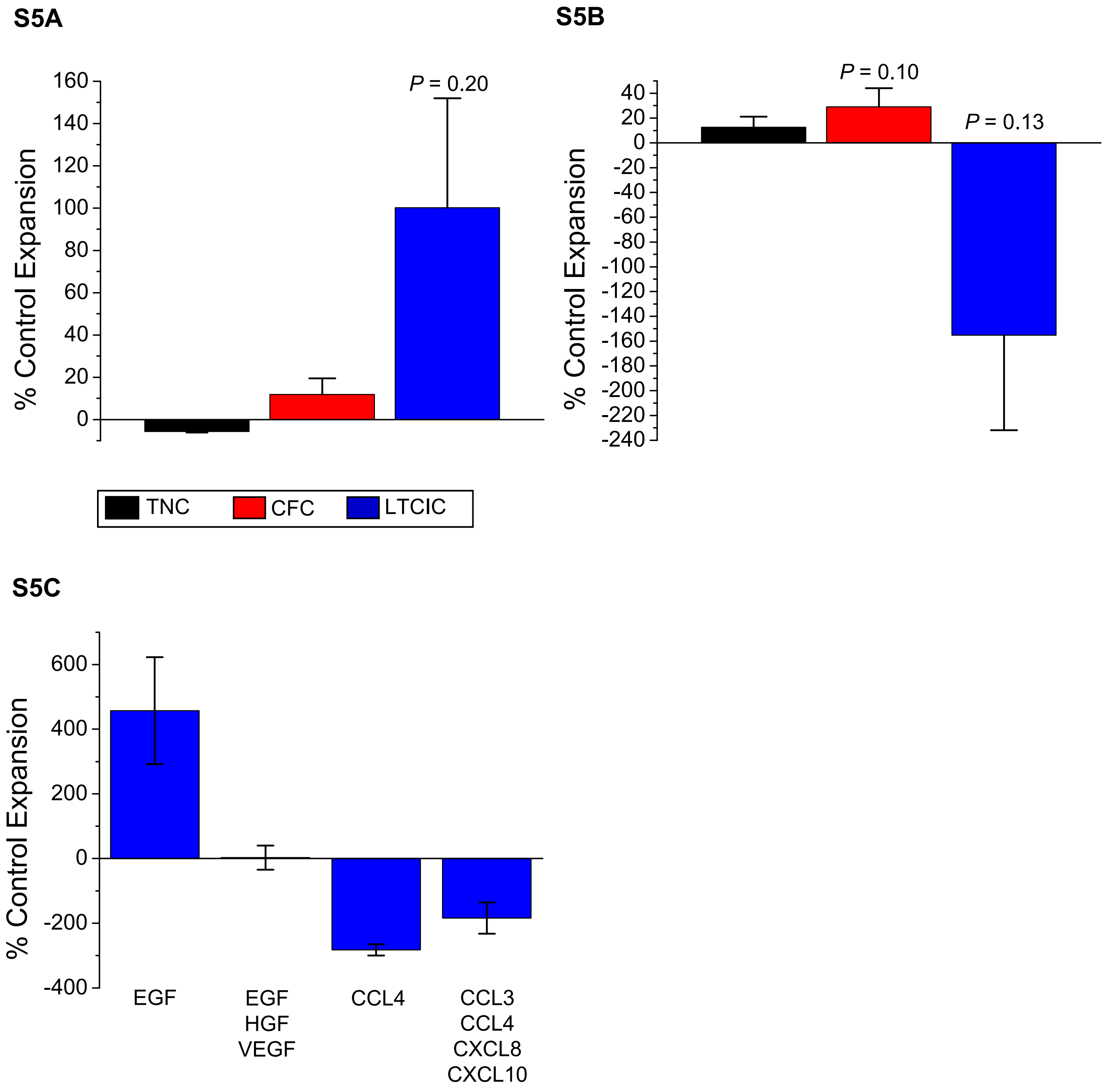
S3B



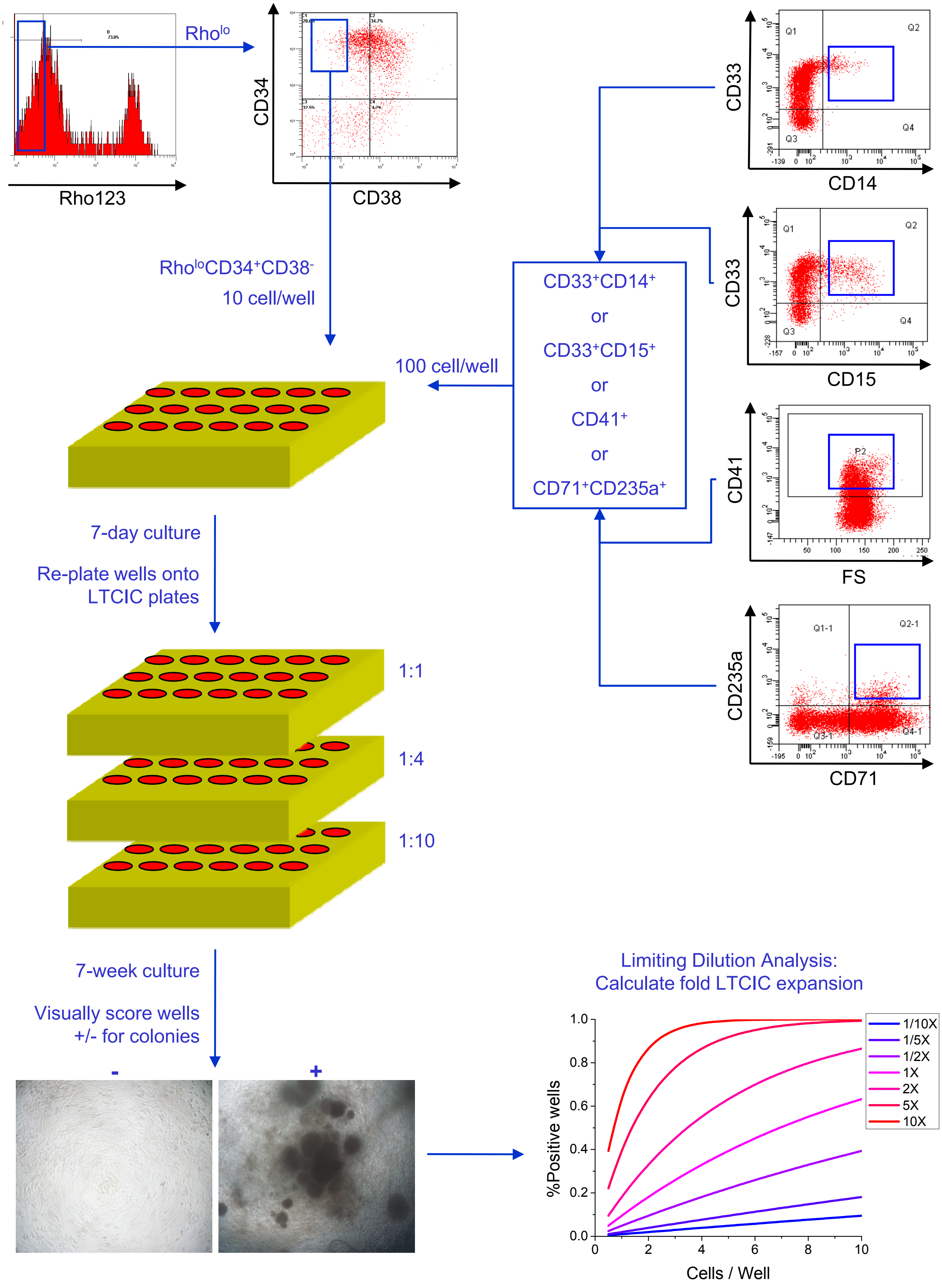
Ge - Uterus
Ge - Trachea
Ge - Thyroid
Ge - Thymus
Ge - Testis
Ge - Stomach
Ge - Spleen
Ge - Small_Intestine
Ge - Skin
Ge - Skeletal_Muscle
Ge - Salivary_gland
Ge - Prostate
Ge - Placenta
Ge - Pancreas
Ge - Ovary
Ge - Lung
Ge - Liver
Ge - Kidney
Ge - Heart
Ge - Fetal_Lung
Ge - Fetal_Liver
Ge - Fetal_Brain
Ge - Colon
Ge - Breast
Ge - Brain
Ge - Bone_Marrow
Ge - Bladder
Ge - Adrenal_gland
Torren - AC133+
Wagner - CD34+CD38- SDF
Wagner - CD34+CD38- FDF
Wagner - CD34+CD38+
Wagner - CD34+CD38-
Komor - Megakaryopoiesis
Komor - Granulopoiesis
Komor - Erythropoiesis
Ferrari - Megakaryoblast
Ferrari - Myeloblast
Ferrari - Monoblast
Ferrari - Erythroblast
Chambers - Lymphoid
Chambers - Myeloid
Chambers - Differentiated
Chambers - Erythrocyte
Chambers - Granulocyte
Chambers - Monocyte
Chambers - NK
Chambers - Tcell
Chambers - Bcell
Chambers - HSC
Li - CD34+ Cultured
Li - CD34+ Fresh
Grahm - PB CD34+ G2-S
Grahm - PB CD34+ G0
Assouu - hESC
Assouu - hESC
Grahm - PB CD34+ G0
Grahm - PB CD34+ G2-S
Li - CD34+ Fresh
Li - CD34+ Cultured
Chambers - HSC
Chambers - Bcell
Chambers - Tcell
Chambers - NK
Chambers - Monocyte
Chambers - Granulocyte
Chambers - Erythrocyte
Chambers - Differentiated
Chambers - Myeloid
Chambers - Lymphoid
Ferrari - Erythroblast
Ferrari - Monoblast
Ferrari - Myeloblast
Ferrari - Megakaryoblast
Komor - Erythropoiesis
Komor - Granulopoiesis
Komor - Megakaryopoiesis
Wagner - CD34+CD38-
Wagner - CD34+CD38+
Wagner - CD34+CD38- FDF
Wagner - CD34+CD38- SDF
Torren - AC133+
Ge - Adrenal_gland
Ge - Bladder
Ge - Bone_Marrow
Ge - Brain
Ge - Breast
Ge - Colon
Ge - Fetal_Brain
Ge - Fetal_Liver
Ge - Fetal_Lung
Ge - Heart
Ge - Kidney
Ge - Liver
Ge – Lung
Ge - Ovary
Ge - Pancreas
Ge - Placenta
Ge - Prostate
Ge - Salivary_gland
Ge - Skeletal_Muscle
Ge - Skin
Ge - Small_Intestine
Ge - Spleen
Ge - Stomach
Ge - Testis
Ge - Thymus
Ge - Thyroid
Ge - Trachea
Ge - Uterus
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

S4




S5A
% Control Expansion
160
140
120
100
80
60
40
20
0
P = 0.20
TNC
CFC
LTCIC
S5B
% Control Expansion
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
-200
-220
-240
P = 0.10
P = 0.13
S5C
% Control Expansion
600
400
200
0
-200
-400
EGF
EGF
HGF
VEGF
CCL4
CCL3
CCL4
CXCL8
CXCL10

S6



Rho^lo

Rho^loCD34^+CD38^-
10 cell/well

CD33^+CD14^+
or
CD33^+CD15^+
or
CD41^+
or
CD71^+CD235a^+

100 cell/well

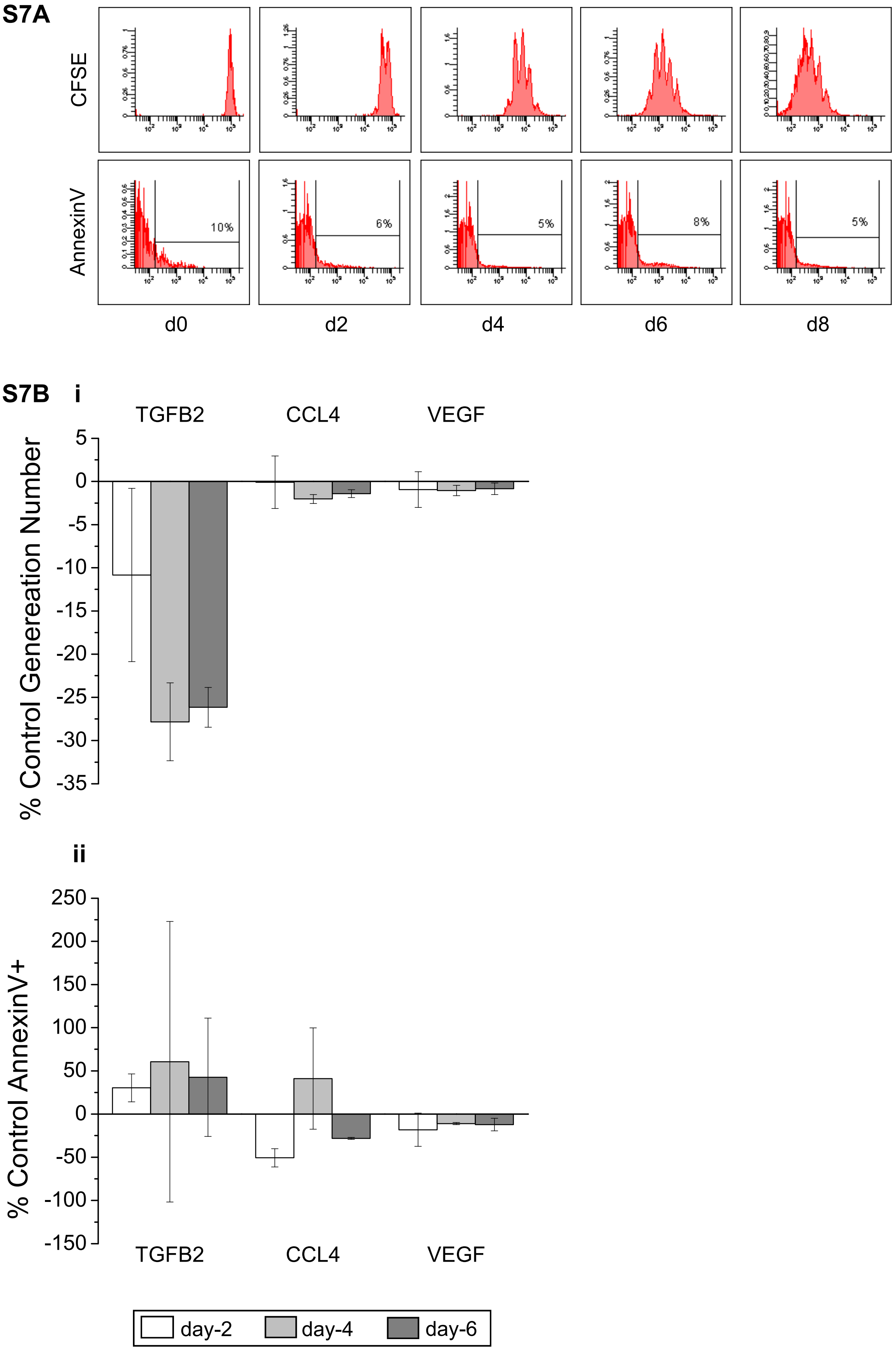
7-day culture

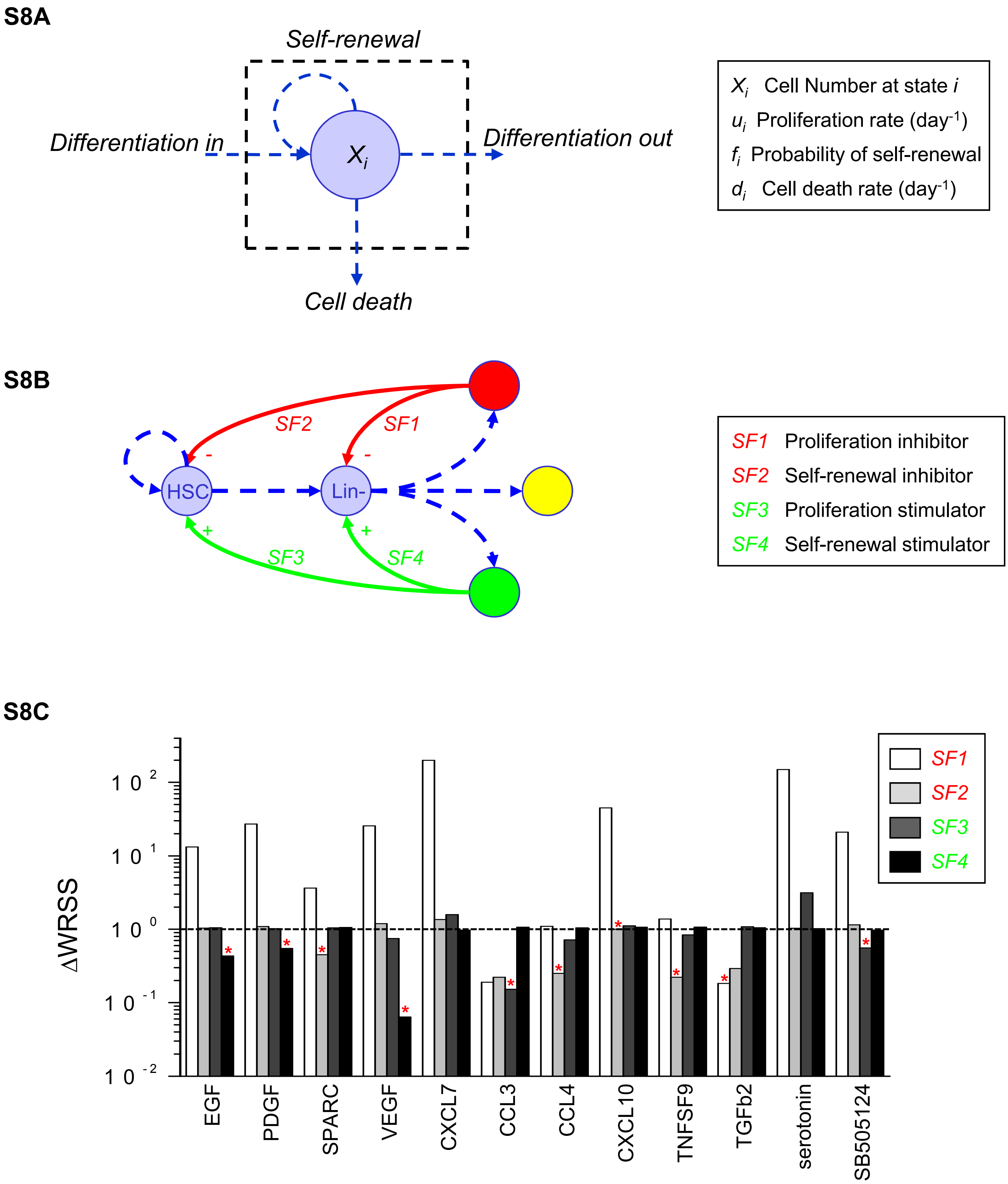
Re-plate wells onto LTCIC plates

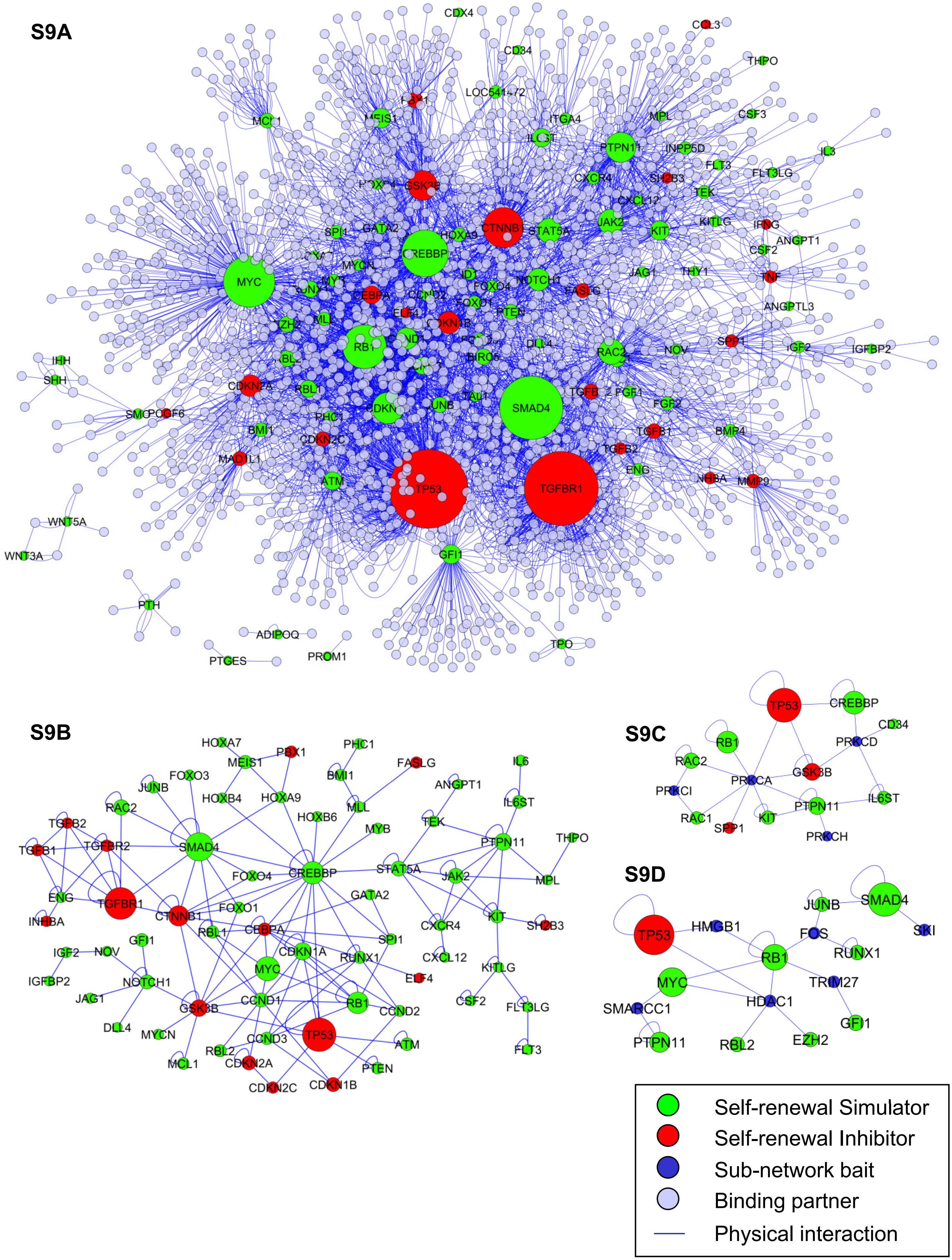
7-week culture

Visually score wells +/- for colonies

Limiting Dilution Analysis:
Calculate fold LTCIC expansion

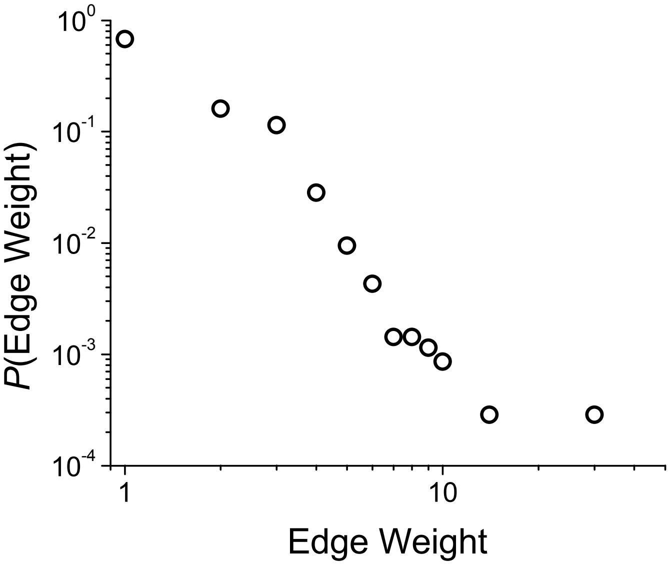

S7A
CFSE
AnnexinV
10%
6%
5%
8%
5%
d0
d2
d4
d6
d8
S7B
i
TGFB2
CCL4
VEGF
% Control Genereation Number
5
0
-5
-10
-15
-20
-25
-30
-35
ii
% Control AnnexinV+
250
200
150
100
50
0
-50
-100
-150
TGFB2
CCL4
VEGF
day-2
day-4
day-6


S8A
Self-renewal
Differentiation in
X_i
Differentiation out
Cell death
X_i Cell Number at state i
u_i Proliferation rate (day^{-1})
f_i Probability of self-renewal
d_i Cell death rate (day^{-1})
S8B
SF2
SF1
-
-
HSC
Lin-
+
+
SF3
SF4
SF1 Proliferation inhibitor
SF2 Self-renewal inhibitor
SF3 Proliferation stimulator
SF4 Self-renewal stimulator
S8C
ΔWRSS
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
EGF
PDGF
SPARC
VEGF
CXCL7
CCL3
CCL4
CXCL10
TNFSF9
TGFb2
serotonin
SB505124
*
SF1
SF2
SF3
SF4

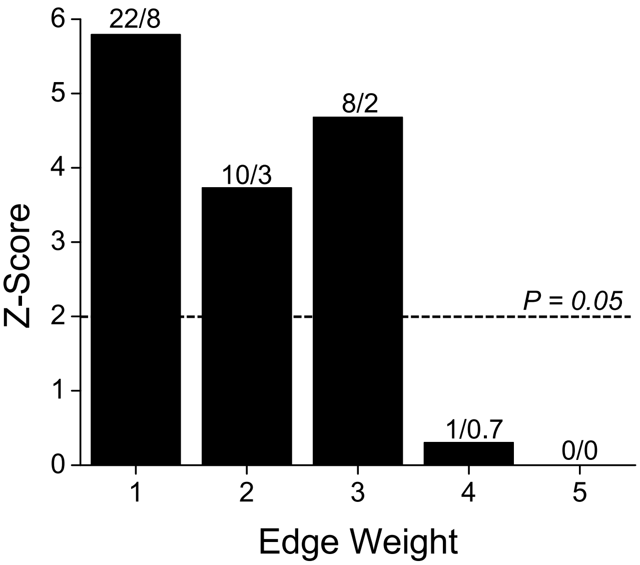


🟢	Self-renewal Simulator
🔴	Self-renewal Inhibitor
🔵	Sub-network bait
○	Binding partner
—	Physical interaction

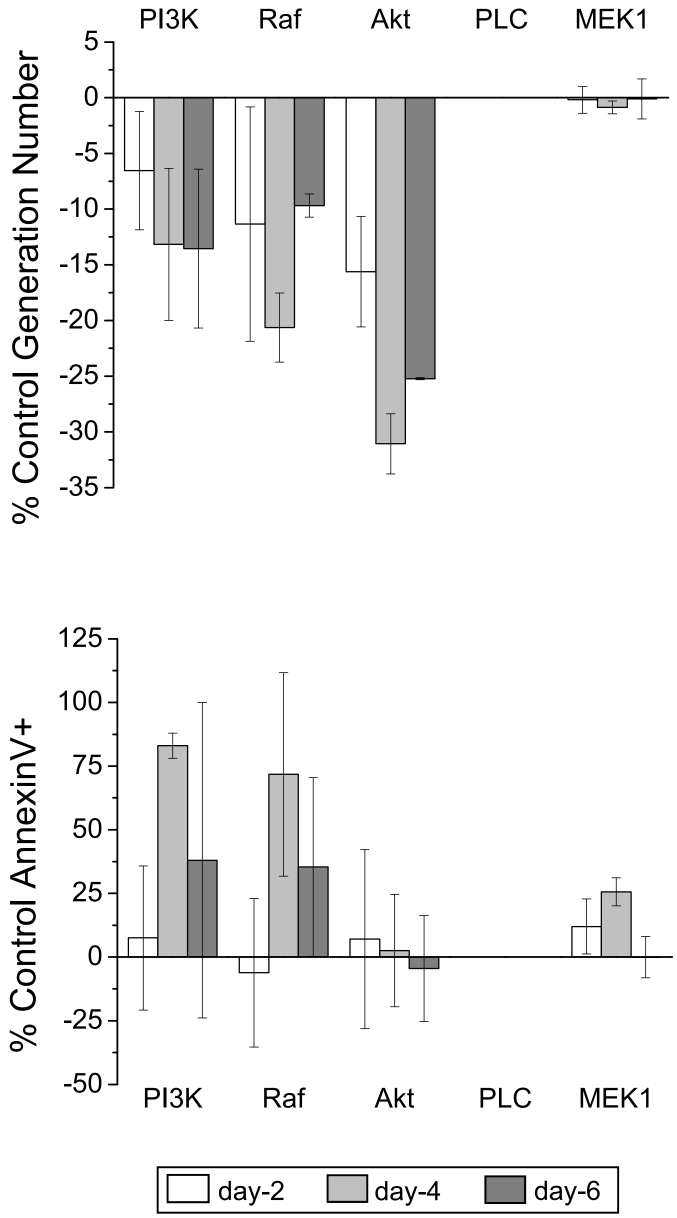
S10A


10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
P(Edge Weight)
1
10
Edge Weight

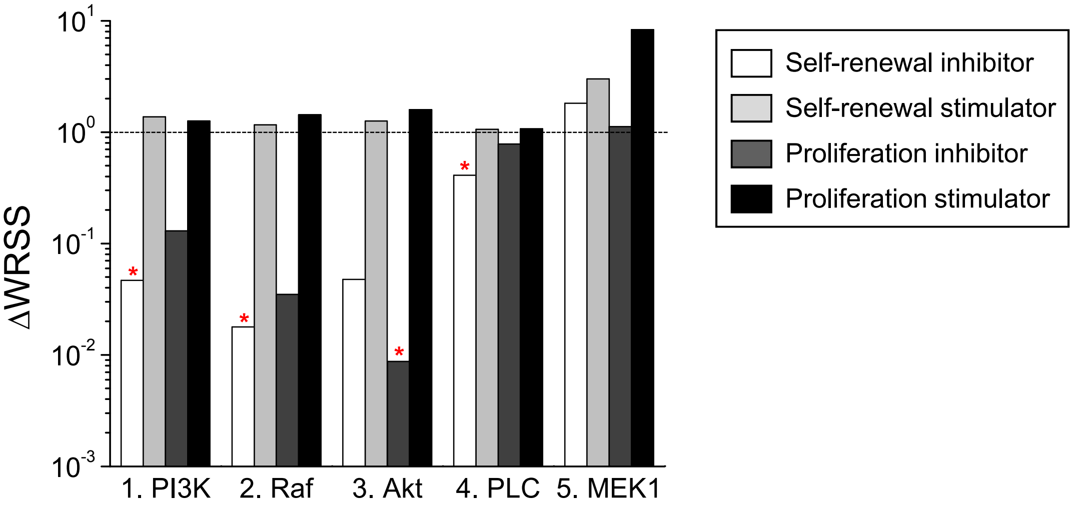
S10B


6
5
4
3
2
1
0
Z-Score
22/8
10/3
8/2
1/0.7
0/0
P = 0.05
1
2
3
4
5
Edge Weight

S11


PI3K
Raf
Akt
PLC
MEK1
% Control Generation Number
5
0
-5
-10
-15
-20
-25
-30
-35
% Control AnnexinV+
125
100
75
50
25
0
-25
-50
PI3K
Raf
Akt
PLC
MEK1
day-2
day-4
day-6

S11


10^1
10^0
10^-1
10^-2
10^-3
ΔWRSS
*
*
*
*
1. PI3K
2. Raf
3. Akt
4. PLC
5. MEK1
Self-renewal inhibitor
Self-renewal stimulator
Proliferation inhibitor
Proliferation stimulator