The Genetic Program of Hematopoietic Stem Cells

Robert L. Phillips,[1] Robin E. Ernst,[1] Brian Brunk,[2] Natalia Ivanova,[1] Mark A. Mahan,[1] Julia K. Deanehan,[1] Kateri A. Moore,[1] G. Christian Overton,[2] Ihor R. Lemischka[1]*

Blood cell production originates from a rare population of multipotent, self-renewing stem cells. A genome-wide gene expression analysis was performed in order to define regulatory pathways in stem cells as well as their global genetic program. Subtracted complementary DNA libraries from highly purified murine fetal liver stem cells were analyzed with bioinformatic and array hybridization strategies. A large percentage of the several thousand gene products that have been characterized correspond to previously undescribed molecules with properties suggestive of regulatory functions. The complete data, available in a biological process–oriented database, represent the molecular phenotype of the hematopoietic stem cell.

Single hematopoietic stem cells can give rise to at least eight distinct blood cell lineages and can maintain lifelong blood production in mice (*1*, *2*). Their hallmark property is the ability to strike a balance between self-renewal and a commitment to differentiation (*2*). The mechanisms that govern these stem cell fate decisions must be under tight yet flexible control. Despite extensive functional and physical stem cell characterization, almost nothing is known about the molecular nature of these regulatory mechanisms. Several molecules have been shown to play roles in early aspects of hematopoietic development, but it has not been possible to elucidate regulatory pathways that function at the level of self-renewing hematopoietic stem cells. Key aspects of stem cell regulation are likely to be emergent properties of interacting pathways and networks, the elucidation of which requires an extensive description of the molecular components available to the stem cell, that is, its genetic program. Herein, we describe a large number of gene products that represent such a program and extend the known properties of stem cells.

During mouse development, the fetal liver is the first tissue from which hematopoietic stem cells can be purified [(*3*, *4*); see (*5*) for stem cell transplantation data and a description of the hematopoietic hierarchy] and is the site where they expand in number under normal conditions (*6*). These stem cells are $Sca^{pos}AA4.1^{pos}Kit^{pos}Lin^{neg/lo}$ (henceforth referred to as "FLHSC") (*3*, *4*). The FLHSC and stem cell–depleted $AA4.1^{neg}$ populations represent the two endpoints of the fetal liver hematopoietic hierarchy. Sufficient numbers of cells were purified to construct non–polymerase chain reaction (PCR)–based cDNA libraries (*7*). A second FLHSC library was constructed with a PCR-based technique (*8*). Sequences present in $AA4.1^{neg}$ cells were subtracted from each of the FLHSC libraries to remove "housekeeping" gene products and to enrich for transcripts expressed in primitive cells (*9*). Our overall approach encompasses high-throughput sequence acquisition and bioinformatics, as well as high-density array, reverse transcriptase–PCR (RT-PCR), and other hybridization analyses. The primary data and the results of computational analyses reside in the Stem Cell Database (SCDb) (*5*).

The non-PCR–based subtracted stem cell library (designated SA) was the source for most analyses. The efficacy of the subtraction was verified by the absence of β-actin and the retention or enrichment of gene products such as *flk2/flt3* and *CD34* (*5*). 5′-end sequences ob-

[1]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. [2]Computational Biology and Informatics Laboratory, University of Pennsylvania, Philadelphia, PA 19104, USA.

*To whom correspondence should be addressed. E-mail: ilemischka@molbio.princeton.edu

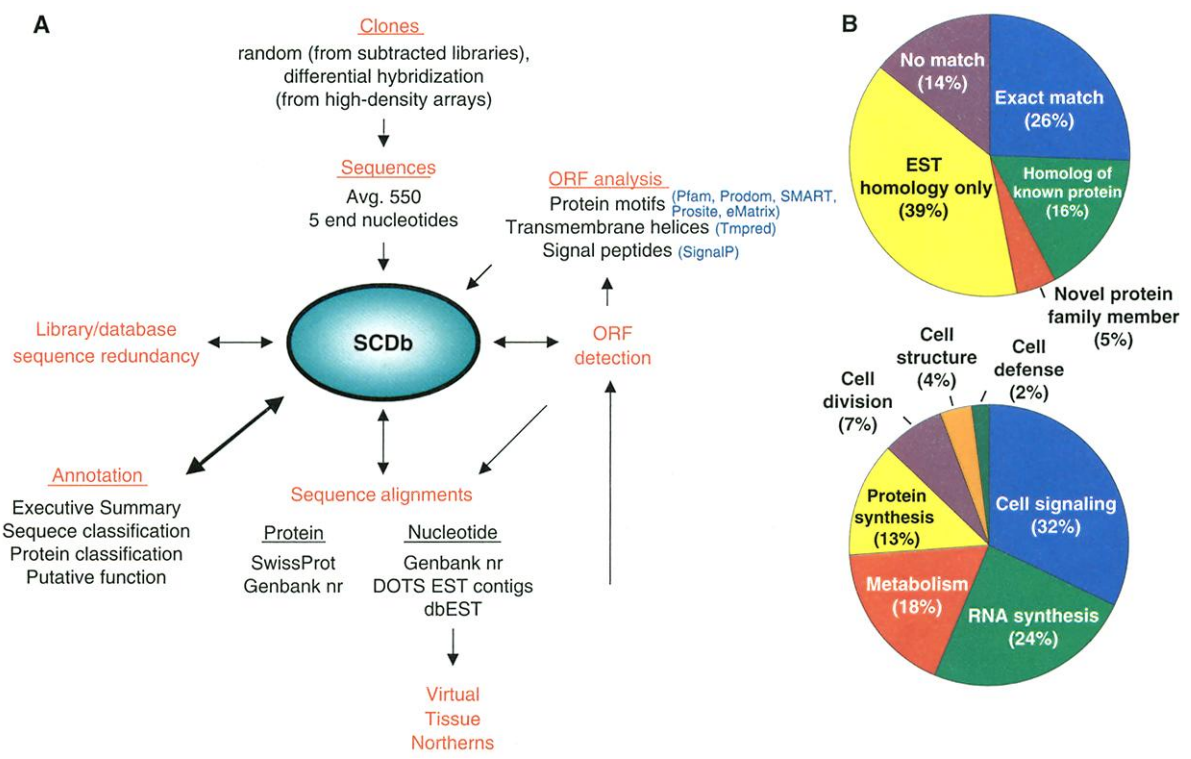


Fig. 1. (**A**) Automated bioinformatics analysis includes queries of six public databases and SCDb itself. Novel open reading frames (ORFs) are examined for protein motifs, transmembrane helices, and signal sequences. These data, along with the results of virtual Northerns, nearest neighbor analysis, and PubMed queries, are incorporated into an executive summary for each entry by the annotator. (**B**) Categorization of informative sequences by homology (top) or, if known, by protein type (bottom). See (*5*) for a detailed breakdown of protein families and access to sequences.

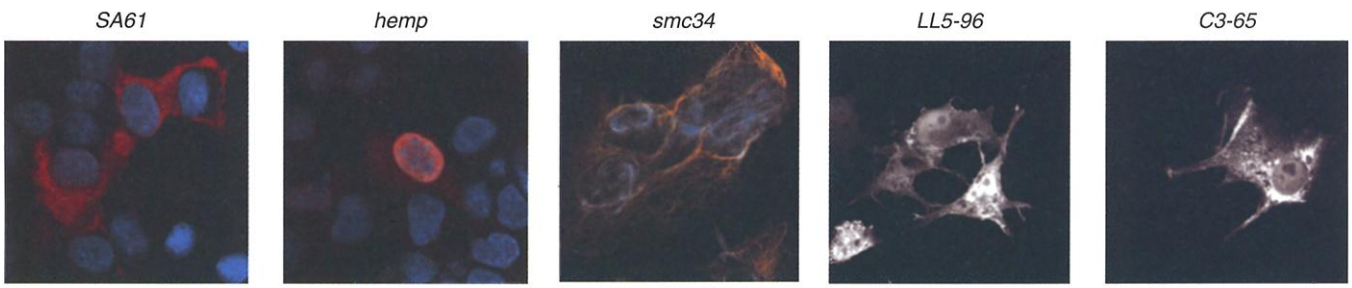
Table 1. Examples of known and novel sequences that have been placed into functional categories. Complete lists for each category can be found in SCDb (*5*). Asterisks denote sequences conserved in both *Drosophila* and *C. elegans*. HLH, helix loop helix; CD, cytoplasmic domain; LDL, low density lipid; TM, transmembrane; EGF, epidermal growth factor; IL, interleukin; GTP, guanosine triphosphate; ATP, adenosine triphosphate.

Category	Name	Description
Transcription factors and chromatin binding proteins	ALL-1	Related to *trithorax*
	AML-1/CBF	Essential for hematopoiesis
	Dnmt-3b	De novo DNA methyltransferase active in early development
	Evi-1	Oncogene; inhibition of Smad3 by Evi-1 may lead to leukemia
	macroH2A1.2	Role in maintenance of chromatin structure
(novel molecules)	C1-52	Related to Trithorax; HLH motif, strong basic region
	LL1-46	Full-length human homolog contains an HLH domain
	LL2in10006	SNF2 and helicase domains; apparent novel helicase
	LL2in10261	7 C2H2 zinc fingers
	LL2in11547	4 C2H2 zinc fingers; GLI/Kruppel family
	LL2in12114	5 C2H2 zinc fingers; GLI/Kruppel family
	LL2in13359	5 C2H2 zinc fingers; GLI/Kruppel family
	LL2in14463	Probably a novel histone deacetylase
	LL2in14423	Contains Kruppel-associated box
	SA49P01/*hemp*	zinc fingers and *mbt* repeats; Polycomb family
Surface molecules	Flk-2	Receptor tyrosine kinase; ligand is an important hematopoietic growth factor
	Smoothened	Receptor for Sonic Hedgehog
	Hem-1	TM4 protein; hematopoietic-specific; essential for oogenesis in *Drosophila*
	CD34	Stem cell antigen; adhesion molecule
	CD27	Receptor for CD70
(novel molecules)	LL2-79	LDL receptor domain; possibly a novel LDL receptor
	LL2in10147	Contains signal peptide, four potential transmembrane (TM) regions
	LL2in10500	Three TM domains apparent; related to ion channels
	LL2in10503	Novel plexin; possible semaphorin receptor; contains two Sema, two plexin domains
	LL2in10512	Rhodopsin family of seven TM receptors
	LL2in10832	Lectin C-type and EGF domains; plexin repeats; single TM
	LL2in11301	Signal, single TM, several EGF domains
	LL2in12756	Contains signal peptide, single TM helix
	LL2in14525	Secretin family of seven TM receptors
	LL2in14567	Seven TM domains; rhodopsin family signature
Secreted proteins	Angiopoietin	Ligand for TIE2
	IL-12 p35	Hematopoietic growth factor
	MIP2	Chemokine
	IL-16	Ligand for CD4; immune cell chemoattractant
	MIP-related protein 2	Chemokine
(novel molecules)	B3-68	Signal peptide; 198 amino acid (aa) open reading frame (ORF)
	C3-46	Signal peptide; apparent complete 226 aa ORF
	LL2in10328	Signal peptide, no TM detected
	LL2in10592	Strong signal; very hydrophilic molecule
	LL2in10705	Signal peptide; similar to *Xenopus* cement gland protein
	LL2in10995	Strong signal peptide; complete ORF
	LL2in11931	Signal peptide, no TM detected
	LL2in12074	Signal peptide, no TM detected
	LL2in12593	Signal peptide, no TM detected
	LL2in13297	Signal peptide; apparent complete 218 aa ORF
Signaling molecules	Dishevelled-1	Essential for signaling in the Wnt pathway
	Manic Fringe	Modulates activation of Notch signaling
	Ski	TGF-β pathway; arrests differentiation of hematopoietic cells in *Gallus*
	DOKL	DOK family; modulates signaling of Abl
	p56Dok-2	RasGAP-interacting tyrosine kinase substrate
(novel molecules)	C2-96	GTP-binding protein related to p21-ras
	C3-65	Novel protein methyltransferase
	LL2in10610	Pleckstrin homology; ArfGAP domains, ankyrin repeats
	LL2in10659	Ser/Thr kinase domain; related to NIK
	LL2in10688	SOCS family of inhibitors of cytokine signaling
	LL2in11043	Contains RhoGAP and Pleckstrin homology domains
	LL2in11898	Three Pleckstrin homology domains, RhoGAP domain
	LL2in12025	Probably a novel AAA ATPase
	LL2in12547	Ras-related protein
	LL2in12652	Contains a FYVE zinc finger
C. elegans orthologs	LL2-57	57% ident., 73% sim. to C14B1.5
	LL2grid-56	63% ident., 81% sim. to F16A11.2
	LL2in10841*	90% ident., 92% sim. to ZK287.5
	LL2in11115	77% ident., 92% sim. to U67954.1; contains SH3 domain
	LL2in11516	59% ident., 75% sim. to C09F5.2; TM4 protein
	LL2in11541	42% ident., 68% sim. to D2023.6; putative ABC transporter

Table 1 (*continued*).

Category	Name	Description
	LL2in12203	65% ident., 85% sim. to F49E10.1
	LL2in13408	53% ident., 67% sim. to ZC518.3; putative transcriptional repressor
	LL2in14227*	52% ident., 70% sim. to B0523.5; homolog of *Drosophila flightless-I*
	SA80*	63% ident., 73% sim. to U40030.1
Drosophila orthologs	C1-41	68% ident., 75% sim. to *abnormal wing discs* (nuc. diphosphate kinase)
	LL2-35	39% ident., 56% sim. to *germ cell-less*
	LL3-86	61% ident., 78% sim. to *kelch*
	LL2in10019	45% ident., 57% sim. to *kraken*
	LL2in10030	51% ident., 69% sim. to *muscleblind*
	LL2in10686	54% ident., 67% sim. to transcriptional regulator *dre4*
	LL2in10841*	97% ident., 100% sim. to ORF from cosmid 115C2.11
	LL2in11384	81% ident., 95% sim. to *pecanex*
	LL2in13158	42% ident., 63% sim. to Septin 1 (also *peanut*)
	LL2in14227*	53% ident., 75% sim. to *flightless-I*

tained from 5735 clones (representing 2119 nonredundant gene products) were characterized by the bioinformatic analyses shown in Fig. 1A. The sequences were compared (using the BLAST algorithm) to seven databases: SwissProt, GenBank protein and nucleotide collections, expressed sequence tags (ESTs), murine and human EST contigs (*10*), and SCDb itself (a measure of internal redundancy). Statistical analysis and limited empirical data suggest that these sequences represent a substantial portion (conservatively, 50 to 55%) of the library complexity (*5*). The nonredundant sequences were categorized by homology as depicted in Fig. 1B (upper panel). Novel sequences were extended with ESTs, and the predicted amino acid sequences were evaluated for motifs, hydrophobicity, and the presence of signal peptides (*5*). A functional categorization of informative protein sequences is shown in Fig. 1B (lower panel). Each entry in SCDb is accompanied by an "executive summary"—a distillation of predicted structural properties, physiological roles, tissue distribution, and oth-



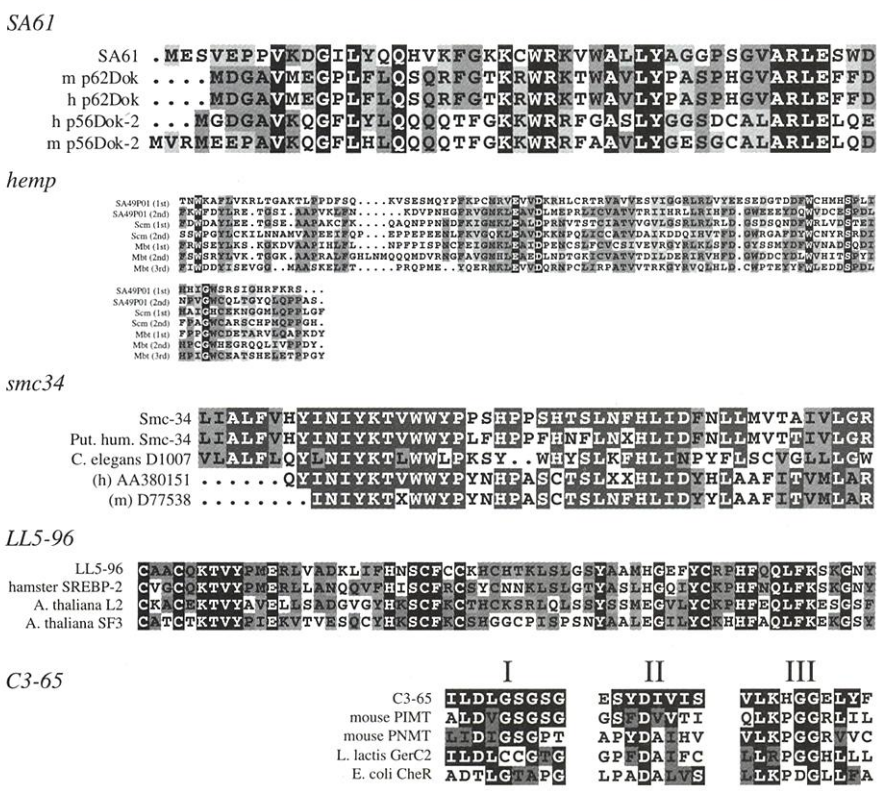


Fig. 2. Alignments of important motifs (right) and subcellular localization (above) for five proteins identified in hematopoietic stem cells. An alignment of SA61 with the two closely related molecules p62^{Dok} and p56^{Dok-2} establishes it as a member of the Dok family of tyrosine kinase substrates. Subcellular localization studies show accumulation of SA61-GFP fusion protein in the cytoplasm. Alignment of the *mbt* repeats of Hemp with those of the Polycomb-family transcriptional repressor proteins Sex Comb on Midleg (Scm) and Malignant Brain Tumor (Mbt) indicate that *hemp* is a novel Polycomb Group member. FLAG epitope-tagged Hemp protein localizes to the nuclei of transfected cells (red nuclei). A region of the eight-transmembrane protein SMC34 beginning in the first transmembrane domain is highly conserved in an ORF from *Caenorhabditis elegans* as well as mouse and human homologs of an apparently related protein. Tagged SMC34 protein (red) shows accumulation in the ER and overlap (yellow) with ER-staining calnexin antibody (green). Clone LL5-96 contains a single LIM domain that aligns with those of several transcription factors. An LL5-96–green fluorescent protein–(LL5-96-GFP) fusion protein preferentially localizes to Cos-1 nuclei. Clone C3-65 shows homology to prokaryotic and mammalian protein methyltransferases in three canonical methyltransferase domains. A C3-65-GFP fusion protein localizes to Cos-1 cytoplasm. Nuclei in the second, third, and fourth photographs were counterstained with Hoechst 33342 (blue).

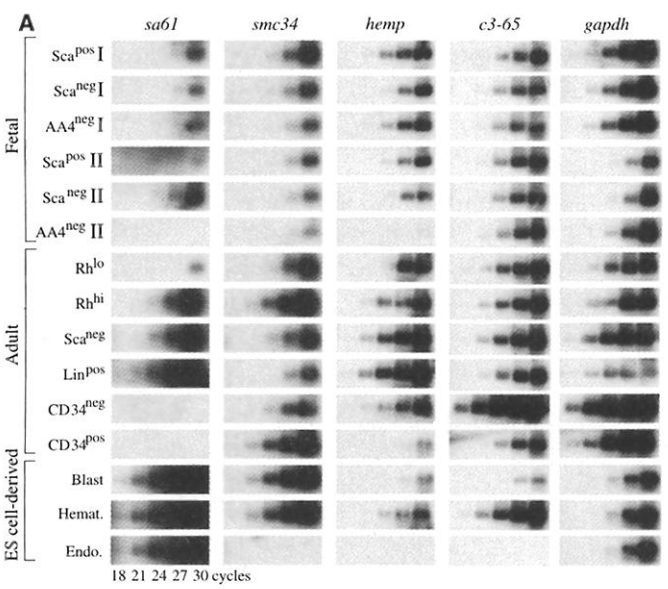
er features that may suggest functions in stem cell biology.

In-depth analyses of the entire subtracted FLHSC sequence set have revealed interesting molecules and molecular relationships. At least 161 transcription factors, 174 cell-surface or membrane-associated molecules, 28 secreted proteins, and 147 signaling molecules have been identified. Many of these are previously undescribed or are identified in stem cells for the first time. Examples of sequences in these categories are presented in Table 1 [for complete lists, alignments, and peptide motif analyses, see (*5*)]. Full-length sequences for five molecules with characteristics suggestive of regulatory roles were obtained (*5*). Sequence alignments and Flag epitope-tag subcellular localization data are presented in Fig. 2. Clone SA61 (Fig. 2, first panel), recently published as DOKL (*11*), has been suggested to function as a modulator of Abl signaling. SA49P01 or Hemp-1 (hematopoietic expressed mammalian polycomb; Fig. 2, second panel) contains a Cys2-Cys2 zinc finger domain, as well as two Lethal-3 Malignant Brain Tumor [*l(3)mbt*] repeats, and is a member of the Polycomb group of transcriptional repressors. The predicted SMC34 protein contains eight transmembrane domains and is localized to the endoplasmic reticulum (ER) (Fig. 2, third panel). LL5-96 (Fig. 2, fourth panel) encodes a LIM-domain protein. Clone C3-65 (Fig. 2, fifth panel) encodes a putative protein methyltransferase.

The overall developmental similarities of these different cell populations suggest that functionally important regulatory molecules should be expressed in multiple sources of stem cells. A comparison of fetal and adult populations should also uncover molecules whose expression is not simply a function of proliferative status. Accordingly, the expression of four of these transcripts was analyzed in fetal liver cells, different compartments of the adult bone marrow hematopoietic hierarchy (Fig. 3A), and embryonic stem (ES) cell–derived hemangioblasts, as well as their hematopoietic and endothelial progeny (*12*). The expression complexity in the hematopoietic hierarchy was further demonstrated by hybridization of several molecules to pools of cDNA populations representative of single progenitor cells with defined differentiation abilities (Fig. 3B) (*13*). The expression differences among these very closely related progenitor cells suggest a high degree of precision in the transcriptional control mechanisms functioning at distinct stages of the hematopoietic hierarchy.

In order to more extensively explore the molecular similarities between fetal and adult stem cells, we used nylon membrane arrays containing over 18,000 SA clones, including the entire sequenced set. Adult bone marrow stem cells are Rhodamine-123^{lo}Sca^{pos}Kit^{pos}-Lin^{neg} (Rh^{lo}), whereas mature cells are Lin^{pos} (*14*). Bidirectionally subtracted (*15*) probe populations from fetal liver or bone marrow were each hybridized to the SA arrays. Table 2 depicts differentially hybridizing genes from the fetal liver and adult bone marrow ranked by differential hybridization intensity. Four genes, three of which were previously undescribed, were found in both screens. We have shown that one of these, CD27, is a marker for a subset of purified bone marrow stem cells (*16*). Clone LL2in20044 contains coiled-coil and COOH-terminal transmembrane domains and is homologous to many members of the syntaxin and SNARE vesicle docking protein families [see (*5*) for additional hybridization data].

A comparison of the fetal and adult screens also reveals molecules that appear more predominant in either fetal or adult hematopoietic stem cells. The top two differential transcripts in the fetal liver screen are *macroH2A1.2* and *dnmt3b*. The variant histone macroH2A1.2 may affect developmental changes in chromatin structure, and has been shown to associate with inactive chromatin on the X chromosome, as well as certain cell type–specific non-X chromosome sites (*17*–*19*). The cytosine methyltransferase Dnmt-3b is essential for mammalian development, as is Dnmt-3a, also identified in SCDb (*20*, *21*). Both of these chromatin-modifying proteins are likely to play roles in stem cell biology. All clones hybridizing with the adult probe population are likely to be expressed in both fetal and adult stem cells. One of these is the homeotic transcriptional repressor TGIF, which is a corepressor [along with the histone deacetylase HD-1 and c-Ski (*22*), and in competition with p300/CBP] of the transforming growth factor–β (TGF-β)–mediated Smad2-



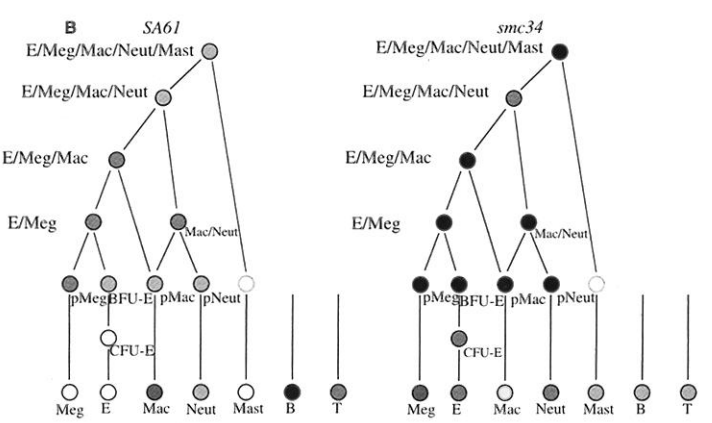


Fig. 3. (**A**) Expression of four transcripts in several different hematopoietic subpopulations determined by semi-quantitative PCR. Rows 1 through 6, cDNA populations I and II were made from independently isolated fetal liver cells as described (*4*): AA4.1^{pos}Lin$^{neg/lo}$Sca-1^{pos}c-Kit^{pos} (designated Sca^{pos}); AA4.1^{pos}Lin$^{neg/lo}$Sca-1^{neg}c-Kit^{pos} (Sca^{neg}); cells negative for AA4.1 by immunopanning (AA4.1^{neg}). Rows 7 through 10, PCR-amplified cDNA from adult bone marrow: Sca^{pos}Kit^{pos}Lin^{neg}Rh123^{low} (Rh^{lo}); Sca^{pos}Kit^{pos}Lin^{neg}Rh123^{high} (Rh^{hi}); Sca^{neg}Kit^{pos}Lin^{neg} (Sca^{neg}). Rows 11 through 12, Sca^{pos}Kit^{pos}Lin^{neg}CD34$^{neg/low}$ [CD34^{neg}; see (*30*)]; Sca^{pos}Kit^{pos}Lin^{neg}CD34^{pos} (CD34^{pos}). Rows 13 through 15, ES cell-derived populations (a gift from G. Keller): Blast, hemangioblasts; Hemat., hematopoietic embryoid bodies derived from blast cells; Endo., endothelial embryoid bodies derived from blasts (*12*). PCR cycles went from 18 cycles to 30 in three-cycle increments. Thus, each lane represents approximately an eightfold increase in PCR product over the previous one. Products were Southern-blotted and hybridized with specific cDNA probes. (**B**) Blots of cDNA from pools of individual hematopoietic progenitors (a gift of N. Iscove, Toronto, Canada) were used to examine expression of two transcripts along the hematopoietic hierarchy. The relative darkness of each circle represents the hybridization intensity of the corresponding slot on the blot. See (*5*) for actual blot images. The labels designate the measured colony-forming potential of each cell type. E, erythroid; Meg, megakaryocyte; Mac, macrophage; Neut, neutrophil; Mast, mast cell; B, B cell; T, T cell; BFU-E and CFU-E, erythroid blast-forming and colony-forming units, respectively.

Smad4 pathway. These regulators—TGIF, HD-1, Ski, and p300/CBP (as well as the Ski-interacting protein and *Drosophila* Bx42 homolog Skip)—have all been identified in the SA library, although TGIF is the most highly represented. The expression of these specific molecules in stem cells further supports a role for TGF-β in hematopoietic control (*23*). The p300/CBP protein is known to acetylate Histone H2A (*24*) and thus may also be involved in functional regulation of macroH2A1.2. The predicted protein encoded by clone LL2in14463, which contains a sequence signature indicative of a histone deacetylase, may also interact with these and other chromatin modifiers.

Other gene products identified in our studies can be placed into developmental pathways. For example, the importance of the Notch pathway in hematopoietic development (*25*) is underscored by the identification of Notch-1, Manic Fringe, nuclear factor kappa B (NFκB), and Dishevelled-1 in SCDb. Other molecules represented in SCDb include members of the Fos serum response pathway (c-Fos, Serum Response Factor, SAP1a, Ets-related protein) and the associated Ets family members FLI-1, FLI-1–associated protein, and PU.1. Among the many cell adhesion proteins uncovered in these studies are Semaphorin B and Neuropilin-1, as well as two apparently novel members of the plexin family. Neuropilin-1 and some members of the plexin family are known to bind

Table 2. Examples of the known and novel genes that show differential hybridization with intensity ratios greater than five. Upper, FLHSC versus AA4.1^{neg} populations; Lower, Rh^{lo} versus Lin^{pos} populations. Mean ratio is the average hybridization intensity ratio observed between the two probes. Count is the number of times a particular sequence was observed in the same screen. More data from the analyses as well as array images are presented in (*5*). The asterisks denote those transcripts that are highly differentially expressed in both screens.

FLHSC versus AA4.1^{neg}			
Sequence homology	Mean ratio	Count	Description
		Known	
Histone macroH2A1.2	32.1	5	Variant histone of unknown function
Dnmt-3b	21.8	17	DNA cytosine-5-methyltransferase
CD27*	18.7	17	Cell surface receptor for CD70
Sortilin	14.2	1	Neurotensin receptor
Nore1	11.6	7	Putative Ras effector molecule
Sm-B antigen	10.1	1	Small nuclear RNA protein
DSIP-immunoreactive pept.	9.1	1	Potential transcriptional coactivator
Homer-3	7.8	1	CC domain; crosslinks metabotropic GluRs
NFkappaB1	6.8	1	Transcription factor subunit
PEST P.I.P.	6.5	6	Actin-interacting; similar to Cdc15p
CD34	6.3	31	Antigen found on stem cells
LSM-1	6.0	1	Surface antigen; lymphocyte marker
p450red	5.8	1	Cytochrome reductase
Testin	5.3	10	Triple LIM domain protein
Gi2 G-protein	5.3	1	Adenylate cyclase inhibitor
LEI	5.0	12	Serine protease inhibitor
		Novel	
LL2in20044*	23.7	1	Apparently related to Syntaxins and SNAREs
LL2in40047	22.2	1	>163 aa ORF; no motifs/homologies
LL2in12491	9.5	1	TBC motif for GTPase activators
LL2in14967	6.1	7	Contains IBR domain, B-box zinc finger
LL2in13218*	6.0	8	878 bp extension, no ORF detected
LL2in40003	5.6	1	No motifs or protein homologies detected
LL2in40009*	5.2	7	1202-bp extension, no ORF detected
LL2in11975	5.1	2	May contain signal sequence, single tm helix

Rh^{lo} versus Lin^{pos}			
Sequence homology	Mean ratio	Count	Description
		Known	
Ligatin	12.1	1	Trafficking receptor for phosphoglycoproteins
CD69	10.3	1	Cell surface antigen
CD27*	8.5	17	Cell surface receptor for CD70
NSD-1	6.6	1	Nuclear receptor cofactor
Angiopoietin	6.1	2	Ligand for TIE2 receptor
TGIF	5.3	24	Homeobox transcription factor
		Novel	
LL2in13610	22.0	1	2-4 transmembrane domains predicted
LL2in12438	17.4	1	No homologies or motifs
LL2in10940	11.2	1	KIAA0161 ORF; B-Box zinc finger
LL2in11441	10.7	1	Two transmembrane domains
LL2in10016	10.1	1	No motifs/homologies
LL2in20044*	8.0	1	Related to Syntaxins and SNAREs
LL2in13620	7.9	4	140 aa ORF; signal sequence; possibly secreted
LL2in40009*	7.3	10	1202-bp extension, no ORF detected
LL2in11078	6.6	1	No motifs/homologies
LL2in13218*	6.2	2	878-bp extension, no ORF detected
LL2in10007	6.0	6	No motifs/homologies
LL2in13877	5.6	5	No motifs/homologies
LL2in14841	5.1	2	No motifs/homologies

semaphorins (*26*, *27*). Given the presence of all three types of molecules in the SA library, it is likely that semaphorins play a role in stem cell adhesion and homing behavior.

Our studies identify numerous individual candidate regulatory molecules, but they also pave the way for more global approaches to stem cell biology. In particular, the production of stem cell microarrays will permit the analysis of fluctuations in the genetic program as a function of permutations in self-renewal, commitment, or other stem cell properties (*3*, *4*, *28*). The large collection of gene products will also facilitate proteomic strategies to uncover protein interaction networks (*29*). We anticipate that the SCDb will be a resource for the stem cell community and will foster the collaborative and consortial interactions necessary for global approaches to important biological questions.

References and Notes

1. C. T. Jordan and I. R. Lemischka, *Genes Dev.* **4**, 220 (1990).
2. S. J. Morrison, N. M. Shah, D. J. Anderson, *Cell* **88**, 287 (1997).
3. K. A. Moore, H. Ema, I. R. Lemischka, *Blood* **89**, 4337 (1997).
4. C. T. Jordan, J. P. McKearn, I. R. Lemischka, *Cell* **61**, 953 (1990).
5. See http://stemcell.princeton.edu/ for additional experimental results and complete protocols and datasets.
6. H. Ema and H. Nakauchi, *Blood* **95**, 2284 (2000).
7. U. Gubler and B. J. Hoffman, *Gene* **25**, 263 (1983).
8. A. Chenchik *et al.*, *Biotechniques* **21**, 526 (1996).
9. J. L. Rubenstein *et al.*, *Nucleic Acids Res.* **18**, 4833 (1990).
10. From the Database of Transcribed Sequences project; see www.cbil.upenn.edu/DOTS/.
11. F. Cong, B. Yuan, S. P. Goff, *Mol. Cell. Biol.* **19**, 8314 (1999).
12. K. Choi, M. Kennedy, A. Kazarov, J. C. Papadimitriou, G. Keller, *Development* **125**, 725 (1998).
13. E. B. Voura, F. Billia, N. N. Iscove, R. G. Hawley, *Exp. Hematol.* **25**, 1172 (1997).
14. G. J. Spangrude and G. R. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7433 (1990).
15. L. Diatchenko, S. Lukyanov, Y. F. Lau, P. D. Siebert, *Methods Enzymol.* **303**, 349 (1999).
16. A. Wiesmann *et al.*, *Immunity* **12**, 193 (2000).
17. J. R. Pehrson, C. Costanzi, C. Dharia, *J. Cell. Biochem.* **65**, 107 (1997).
18. C. Costanzi and J. R. Pehrson, *Nature* **393**, 599 (1998).
19. J. E. Mermoud, C. Costanzi, J. R. Pehrson, N. Brockdorff, *J. Cell Biol.* **147**, 1399 (1999).
20. M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **99**, 247 (1999).
21. R. S. Hansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14412 (1999).
22. D. Wotton, R. S. Lo, S. Lee, J. Massagué, *Cell* **97**, 29 (1999).
23. O. G. Ottmann and L. M. Pelus, *J. Immunol.* **140**, 2661 (1988).
24. V. V. Ogryzko, R. L. Schiltz, V. Russanova, B. H. Howard, Y. Nakatani, *Cell* **87**, 953 (1996).
25. L. A. Milner and A. Bigas, *Blood* **93**, 2431 (1999).
26. M. L. Winburg *et al.*, *Cell* **95**, 903 (1998).
27. T. Takahashi *et al.*, *Cell* **99**, 59 (1999).
28. P. O. Brown and D. Botstein, *Nature Genet.* **21**, 33 (1999).
29. T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143 (2000).
30. M. Osawa, K. Hanada, H. Hamada, H. Nakauchi, *Science* **273**, 242 (1996).

31. We thank A. Beavis and J. Zawadzki for expert flow cytometry and J. Goodhouse for expert confocal microscopy. We thank J. Shepard, C. Wesley, A. Nixon, S. Sykes, W. Perry, and R. Wager for valuable contributions to these studies. We thank P. Sharp, T. Shenk, S. Tilghman, L. Enquist, and C. Jordan for critical reviews of the manuscript. Sequencing support was provided by Imclone Systems, Inc. This work was supported by NIH grants R01-DK42989 (I.R.L.) and R01-RR04026 (G.C.O.).

14 February 2000; accepted 5 April 2000

Role of 4.5*S* RNA in Assembly of the Bacterial Signal Recognition Particle with Its Receptor

Paul Peluso,[1] Daniel Herschlag,[2] Silke Nock,[1] Douglas M. Freymann,* Arthur E. Johnson,[3] Peter Walter[1]†

The mechanism by which a signal recognition particle (SRP) and its receptor mediate protein targeting to the endoplasmic reticulum or to the bacterial plasma membrane is evolutionarily conserved. In *Escherichia coli*, this reaction is mediated by the Ffh/4.5*S* RNA ribonucleoprotein complex (Ffh/4.5*S* RNP; the SRP) and the FtsY protein (the SRP receptor). We have quantified the effects of 4.5*S* RNA on Ffh-FtsY complex formation by monitoring changes in tryptophan fluorescence. Surprisingly, 4.5*S* RNA facilitates both assembly and disassembly of the Ffh-FtsY complex to a similar extent. These results provide an example of an RNA molecule facilitating protein-protein interactions in a catalytic fashion.

Ffh and FtsY are both guanosine triphosphatases (GTPases) (*1*–*5*) that interact with each other in a GTP-dependent manner and reciprocally stimulate each other's GTPase activity (*6*, *7*). The GTPase domains of Ffh and FtsY define them as members of a GTPase subfamily with unique properties (*8*–*11*). 4.5*S* RNA enhances association of Ffh and FtsY, which suggested a role for the RNA in stabilizing the complex (*6*). To analyze the role of 4.5*S* RNA in this reaction in more detail, we took advantage of the fact that FtsY, but not Ffh, contains tryptophan residues. This allowed us to monitor the interaction of Ffh and FtsY spectroscopically (Fig. 1). Recently, a similar assay was independently developed by Jagath and co-workers (*12*). In our studies we used a NH_2 terminally truncated version of FtsY (residues 48 to 494) that was previously shown to interact with Ffh in a manner indistinguishable from that of full-length FtsY (*5*).

Incubation of FtsY with Ffh/4.5*S* RNP in the presence of the nonhydrolyzable GTP analog GppNHp (5′-guanylylimidodiphosphate) shifted the tryptophan fluorescence emission maximum by ~10 nm and doubled the fluorescence intensity (Fig. 1A). This is consistent with burial of one or both of the tryptophans in a more hydrophobic environment upon formation of a Ffh-FtsY complex. These fluorescence changes occurred only in the presence of GppNHp and not in the presence of guanosine diphosphate (GDP) (Fig. 1B), consistent with the GppNHp dependence for complex formation determined by affinity chromatography (*6*).

Complex formation and stimulation of GTPase activity were previously shown to be dependent on the presence of 4.5*S* RNA. We were therefore surprised to observe that in the absence of 4.5*S* RNA, addition of GppNHp resulted in an increase and shift in fluorescence that was indistinguishable from that observed with the Ffh/4.5*S* RNP (Fig. 1, C and D). To understand the origin of this paradox and the role of 4.5*S* RNA in complex formation, we carried out a kinetic and thermodynamic analysis of the reaction.

We compared the kinetics of association between Ffh and FtsY in the absence and presence of 4.5*S* RNA (Fig. 2, A and B). The association with FtsY, monitored by fluorescence, was faster for Ffh/4.5*S* RNP than for the same concentration of Ffh by a factor of more than 100. Analogous determinations at a series of Ffh and Ffh/4.5*S* RNP concentrations gave second-order rate constants for

[1]Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. [2]Department of Biochemistry, Stanford University, Stanford, CA 94305, USA. [3]System Health Science Center, Texas A&M University, College Station, TX 77843, USA.

*Present address: Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, IL 60611, USA.

†To whom correspondence should be addressed. E-mail: walter@cgl.ucsf.edu