The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity

John E. Pimanda[†], Ian J. Donaldson[†], Marella F. T. R. de Bruijn[‡], Sarah Kinston[†], Kathy Knezevic[†], Liz Huckle[§], Sandie Piltz[†], Josette-Renée Landry[†], Anthony R. Green[†], David Tannahill[§], and Berthold Göttgens[†¶]

[†]Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 2XY, United Kingdom; [‡]Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom; and [§]The Wellcome Trust Sanger Institute, Cambridge CB10 1SA, United Kingdom

Edited by Eric N. Olson, University of Texas Southwestern Medical Center, Dallas, TX, and approved November 1, 2006 (received for review August 21, 2006)

Hematopoietic stem cell (HSC) development is regulated by several signaling pathways and a number of key transcription factors, which include Scl/Tal1, Runx1, and members of the Smad family. However, it remains unclear how these various determinants interact. Using a genome-wide computational screen based on the well characterized *Scl* +19 HSC enhancer, we have identified a related Smad6 enhancer that also targets expression to blood and endothelial cells in transgenic mice. Smad6, Bmp4, and Runx1 transcripts are concentrated along the ventral aspect of the E10.5 dorsal aorta in the aorta–gonad–mesonephros region from which HSCs originate. Moreover, Smad6, an inhibitor of Bmp4 signaling, binds and inhibits Runx1 activity, whereas Smad1, a positive mediator of Bmp4 signaling, transactivates the *Runx1* promoter. Taken together, our results integrate three key determinants of HSC development; the Scl transcriptional network, Runx1 activity, and the Bmp4/Smad signaling pathway.

hematopoiesis | SMAD6 | hematopoietic stem cell | aorta–gonad–mesonephros | bioinformatics

The coordinated expression of genes lies at the heart of metazoan development, with complex gene-regulatory networks governing the spatial variation and temporal sequence with which genes are expressed (reviewed in ref. 1). Transcription factors and the cis-regulatory sequences to which they bind form the building blocks of gene-regulatory networks (2). Recognizing the components and hierarchy of gene-regulatory networks and their interactions with cell-signaling pathways not only provides insights into biology but also is fundamental to understanding how deregulation of networks contributes to pathology.

In the postgenomics era, classical methods to screen for potential gene-regulatory regions, such as DNase I hypersensitivity mapping by Southern blotting of selected loci, are being superseded by new genome-wide techniques such as ChIP coupled to genomic tiling arrays (ChIP/chip) (3)or high-throughput sequencing of sequence tags indicating DNase I hypersensitivity (4). Although these large-scale screening techniques may rapidly identify the positions of large numbers of candidate regulatory elements, they do not accurately predict their function, i.e., whether a given element is an enhancer or silencer, nor the pattern of activity in cell types other than those analyzed. Moreover, the large amounts of biological material required prevent use of these techniques in the study of early developmental programs and adult stem cell systems. Consequently, *in silico* approaches are increasingly used to perform genome-wide identification of candidate regulatory elements. Transcription factors often act as components of multiprotein complexes binding to clusters of binding sites. Well characterized regulatory modules with known combinations of transcription factor-binding sites can greatly facilitate the *in silico* prediction of other cis-regulatory modules that may be important in regulating the same biological process. This method is proving to be a powerful tool in assembling gene-regulatory codes and has been applied in yeast, flies, and, more recently, in mammals to identify genes and their regulatory modules, which are involved in specific developmental processes (5–7).

Hematopoeitic stem cell (HSC) specification and subsequent differentiation into its many lineages is one of the best understood vertebrate developmental systems (8), yet little is known about the transcriptional control of HSCs *per se*. The SCL transcription factor is a key regulator of embryonic HSC development. We have shown that the *Scl* +19 element (elements numbered by their distance in kb from the transcriptional start site) directs reporter gene expression in transgenic mice to HSCs and endothelium (9, 10). We have also shown that the activity of this element depends on two conserved Ets-binding sites and one conserved GATA-binding site (Fig. 1*A*). To identify functionally related regulatory elements, we have catalogued conserved ETS and GATA-binding sites in the human genome and identified 67 clusters with two conserved ETS- and one conserved GATA-binding site, which also conformed to the orientation and spacing constraints of the Scl +19 element. One cluster each was situated in the *Fli1* and *PRH* gene loci, respectively. Based on their known link with blood development, the clusters were selected for functional validation and shown to have similar *in vivo* activity to the *Scl* +19 enhancer (6).

However, the role of the remaining elements, as regulators of blood development, were unknown. Therefore, we performed a computational analysis incorporating features of the previously identified *Fli1* +12 and *PRH* +1 enhancers into a set of new search parameters and validated all predicted enhancers by using a series of biological filters. This process led to the identification of Smad6 as a potential regulator of early blood and endothelial development, and subsequent experiments allowed us to integrate three key determinants of HSC development; the Scl network, Runx1 activity, and the Bmp4/Smad signaling pathway.

Results

Identification of cis-Regulatory Elements Active During Blood and Endothelial Development. The spacing and orientation constraints of the *Scl* +19 enhancer (Fig. 1*A*) were used as a template for the initial screen for conserved ETS- and GATA-binding sites (6). To limit the number of elements for transgenic *in vivo* analysis, we used the prior validation of *FLI1* +12 and *PRH* +1 elements as bona fide hematopoietic cis-regulatory modules and introduced a more stringent set of filtration parameters (Fig. 1*B*). As each of the core ETS- and GATA-binding motifs of these elements was located within a larger block of sequence identity, we extended the 4-bp conserved motifs into 5-bp conserved motifs. The base pairs flanking the core

Author contributions: J.E.P. and B.G. designed research; J.E.P., I.J.D., S.K., K.K., L.H., S.P., J.-R.L., and D.T. performed research; J.E.P. and M.F.T.R.d.B. contributed new reagents/analytic tools; J.E.P., M.F.T.R.d.B., D.T., and B.G. analyzed data; and J.E.P., A.R.G., and B.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: AGM, aorta–gonad–mesonephros; E, embryonic day; HSC, hematopoietic stem cell.

¶To whom correspondence should be addressed. E-mail: bg200@cam.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0607196104/DC1.

www.pnas.org/cgi/doi/10.1073/pnas.0607196104

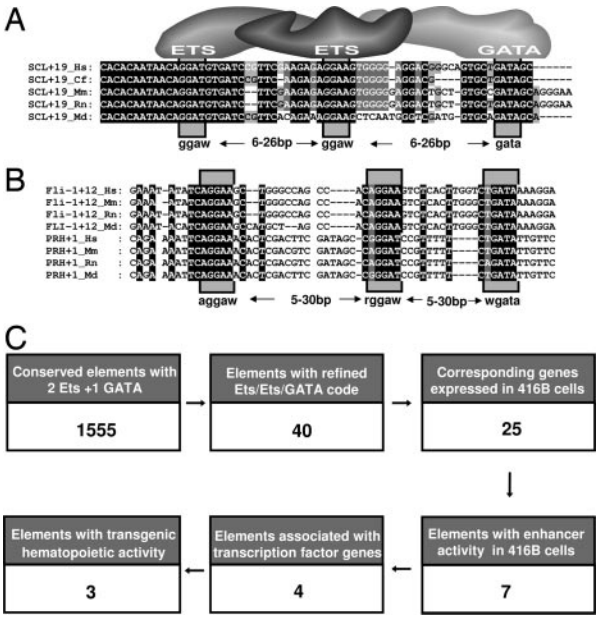

Fig. 1. *In silico* identification of candidate cis-regulatory modules controlling blood and endothelial development. (*A*) Schematic diagram of a five-way sequence alignment of the SCL + 19 E/E/G (E/E/G) binding site cluster bound by a trimeric protein complex. The binding site consensus sequence, ggaw*6–26bp*ggaw*6–26bp*gata was used as a template for the initial genome-wide computational screen for conserved E/E/G motifs. Hs, human; Cf, dog; Mm, mouse; Rn, rat; Md, opossum. (*B*) The hematopoietic enhancer elements, *FLI1* +12 and *PRH* +1 were identified by using the SCL + 19 enhancer template in *A*. These elements share the extended binding-site consensus sequence, aggaw*5–30bp*rggaw*5–30bp*wgata. (*C*) Stepwise filtering of candidate cis-regulatory modules.

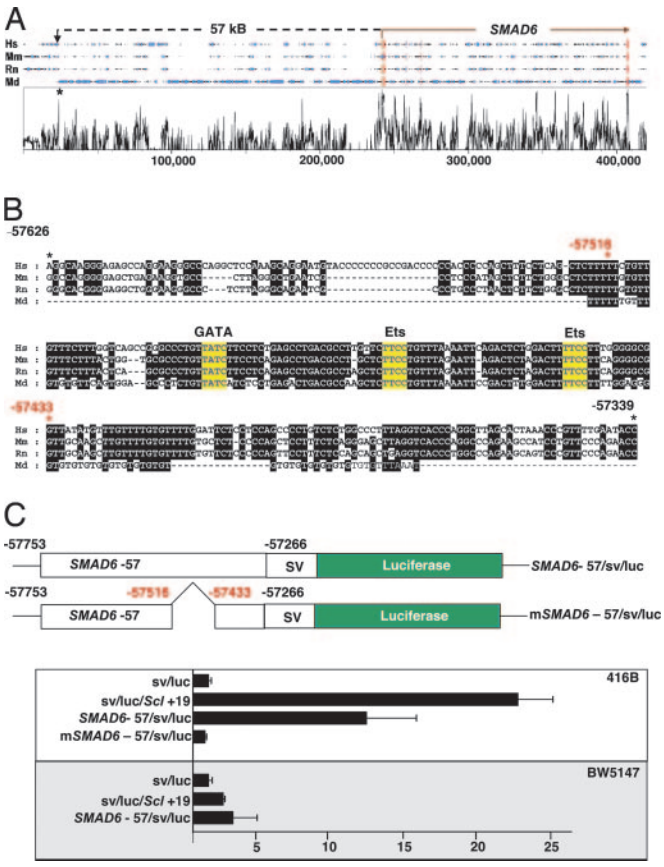

Fig. 2. The SMAD6 E/E/G cluster is located within a highly conserved region 57 kb upstream of exon 1 and is required for its activity in cell lines. (*A*) Mammalian *SMAD6* loci aligned by using multilagan with sequences from Hs, human; Mm, mouse; Rn, rat; and Md, opossum. Shown are coding exons (red rectangles), untranslated regions (beige), and repetitive elements (blue). The base pair numbering includes gaps introduced by the alignment program. The asterisk marks the E/E/G homology peak, located ≈57 kb upstream of exon 1 of human *SMAD6*. (*B*) Nucleotide sequence alignment of the −57 homology peak, with the conserved E/E/G binding sites shaded in yellow. Base pair numbering is from exon 1 of *hSMAD6*. (*C*) A 488-bp fragment of human DNA incorporating the *SMAD6* −57 homology peak was subcloned into a SV/luciferase reporter vector (*SMAD6* −57/SV/*luc*) and tested for enhancer activity along with a deletion (−57516 to −57433) mutant lacking the E/E/G cluster (m*SMAD6* −57/SV/*luc*) and the *Scl* +19 stem cell enhancer (SV/*luc*/*Scl* +19).

motifs were therefore incorporated into the original *Scl* +19 orientation and constraints filter, i.e., *a*ggaw*5–30bp**r*ggaw*5–30bp**w*gata (compare Fig. 1*B* with 1*A*). Using this refined filter, we identified 40 elements, which were localized to 39 human genes with 38 known mouse homologues [Fig. 1*C*; and see supporting information (SI)].

Transcriptional activity of the *Scl* +19, *FLI1* +12, and *PRH* +1 elements in stable transfection assays in 416B murine hematopoietic progenitor cells predicted subsequent *in vivo* hematopoietic activity in transgenic mice (6). Therefore, to prioritize candidate elements for transgenic analysis, we used the transcriptional activity of the *in silico*-predicted elements in stable transfection analysis in 416B cells as a predictor of *in vivo* hematopoietic activity. We first determined the expression profile of the 38 mouse genes corresponding to the 40 elements. Of these 38 genes, 25 were expressed in 416B cells (Fig. 1*C*; and see SI). We successfully subcloned 21 elements corresponding to these 25 genes into luciferase reporter constructs and analyzed their enhancer activity in stable transfection assays in 416B cells (see SI). Four elements could not be subcloned despite repeated attempts (CMC2, KDGG, S15B, and Y140). However, none of these genes are thought to function as transcription factors, and their regulatory elements are therefore unlikely to be components of transcriptional networks. Of the 21 tested elements, 7 had ≈10-fold or greater activity than the control SV/luciferase vector (Fig. 1*C*; and see SI). Of the seven active elements, four were transcription factors, i.e., FLI1, PRH, SMAD6, and MEIS1 (Fig. 1*C*). *SMAD6* −57 and *MEIS1* +140, the two newly identified elements, were tested in F_0 transgenic analyses. The *MEIS1* +140 element targeted reporter gene activity to the neuroepithelium but not blood or endothelium (see SI). The *SMAD6* −57 element, however, was active in blood and endothelium and was studied further.

Characterization of the *SMAD6* −57 Enhancer. Long-range comparative genomic sequence analysis demonstrated that the *SMAD6* locus contains multiple peaks of noncoding sequence homology (Fig. 2*A*). A single peak located 57 kb upstream of exon 1 of human *SMAD6* contained the Ets/Ets/GATA (E/E/G) motif (marked with an asterisk in Fig. 2*A*). This element was ≈96 kb downstream of the next most proximal human gene. The E/E/G motif was located in a block of sequence highly conserved in human/mouse/rat/opossum genomes (Fig. 2*B*).

To further assess enhancer activity of the *SMAD6* −57 element, the 488-bp fragment was analyzed in stable transfection assays and compared with the *Scl* +19 enhancer. A 604-bp fragment of the *Scl* +19 enhancer (SV/*luc*/*Scl* +19) increased mean luciferase activity by ≈24-fold in 416B progenitors but was inactive in T cell line BW5147 (Fig. 2*C*). Similarly, the *SMAD6* −57/SV/*luc* construct increased luciferase activity by ≈12-fold in 416B cells yet displayed only weak activity in BW5147 cells. Importantly, deletion of the E/E/G core binding motif (m *SMAD6* −57/SV/*luc*), abolished activity of the enhancer in 416B cells (Fig. 2*C*).

The *Scl* +19, *FLI1* +12 and *PRH* +1 enhancers target reporter

gene expression to blood progenitors, endothelium, and endocardium (6, 9, 10). To compare the biological activity of the *SMAD6* −57 element with the above, we generated transgenic mice carrying a *lacZ* reporter driven by the SV40 minimal promoter and the *SMAD6* −57 element (*SMAD6* −57/SV/*LacZ* in Fig. 3*A*). One of three *SMAD6* −57/SV/*LacZ* E11.5 F_0 transgenic embryos and both *SMAD6* −57/SV/*LacZ* transgenic lines (L7814 and L7876) showed strong β-galactosidase expression in the heart and vasculature (Fig. 3 *A* and *B*; and see SI). In addition, both *SMAD6* −57/SV/*LacZ* lines showed expression in the fetal liver, the major hematopoietic organ at embryonic day (E) 11.5 and the brain (Fig. 3*B*).

A detailed developmental analysis of *SMAD6* −57/SV/*LacZ* (Line 7814) was performed (Fig. 3*A*). Whole-mount analysis of gastrulating E7.5 embryos showed expression of the transgene in the posterior primitive streak (Fig. 3*A* compare region marked by arrow in *Aii* with *Aiii*). This finding is consistent with the pattern of expression of the transgene in the heart, vasculature, and hematopoietic organs during later stages of development (Fig. 3 *Aii–Avi*). To evaluate the role of the E/E/G motif in directing expression of the transgene to these tissues, we analyzed E11.5 F_0 embryos generated with a mutant *SMAD6* −57 construct lacking the E/E/G-binding sites, i.e., m *SMAD6* −57/SV/*LacZ* (mut) (Fig. 3*Ai*). Three of eight transgenic embryos showed no staining, whereas five showed scant nonspecific staining in adventitia but no staining in endothelium, heart, fetal liver, or brain (Fig. 3*Avii*).

The pattern of expression generated by the −57 enhancer was compared with the expression of Smad6 as assessed by *in situ* hybridization (Fig. 3*C*). As with the transgene, prominent Smad6 expression was observed in the heart valves, endocardium, blood vessels, fetal liver, and brain (compare corresponding panels in Fig. 3*C* with 3*B*). This finding is consistent with reports that targeted insertion of a *LacZ* reporter into the *Smad6* locus resulted in expression in the heart and blood vessels and that abnormalities of the cardiac valves and outflow tract septation defects were observed in *Smad6*-null mice. The *Scl* +19 enhancer in *Scl* 6E5/lacZ/3′enh transgenic mice (Line 2269) targets hematopoietic stem and progenitor cells in the bone marrow (11). The *SMAD6* −57 enhancer targets a similar proportion of cells (≈8%) in the bone marrow of adult mice (Line 7814) (Fig. 3*D*). Furthermore, cells targeted by the *SMAD6* −57/SV/*LacZ* transgene express ≈4-fold more endogenous Smad6 by quantitative RT-PCR than transgene-negative cells. This difference is ≈30-fold in favor of $LacZ^+$ cells in the c-*Kit*$^+$ fraction (data not shown). HSCs are highly purified as $CD150^+/CD48^-/CD41^-$ cells (12), and cells targeted by both the *Scl* +19 and *SMAD6* −57 enhancers are enriched in cells with this phenotype. Moreover, cells targeted by both transgenes show a relative paucity of lineage committed $CD150^-/CD48^-/CD41^-$ cells. These data suggest that both transgenes preferentially target blood stem and progenitor cells in the bone marrow.

The *Smad6* −57 Enhancer Is Bound by Fli-1, Elf, Erg, and GATA2. To integrate the new *SMAD6* enhancer into the SCL regulatory network, it was important to identify transcription factors that bind the enhancer *in vivo*. Studies have shown that the *Scl* +19 enhancer is bound by GATA2 and the Ets factors Fli1, Elf, and Erg (10), and our transgenic analysis had shown that these sites within the *SMAD6* enhancer were important (see above). We therefore performed quantitative ChIP assays in 416B cells with antibodies to a panel of five Ets factors and GATA2, all of which are expressed in this cell line (Fig. 4*A*). With the exception of Erg, these Ets and GATA factors were also expressed in NIH 3T3 fibroblasts, and ChIP assays were therefore performed in NIH 3T3 cells to compare transcription factor binding at the *SMAD6* enhancer in a nonhematopoietic cell line. ChIP assays were also performed by using an antibody that binds to acetylated histone H3 to assess chromatin accessibility to transcription factor binding.

Immunoprecipitated chromatin samples were analyzed by quantitative real-time PCR with the levels of enrichment normalized to

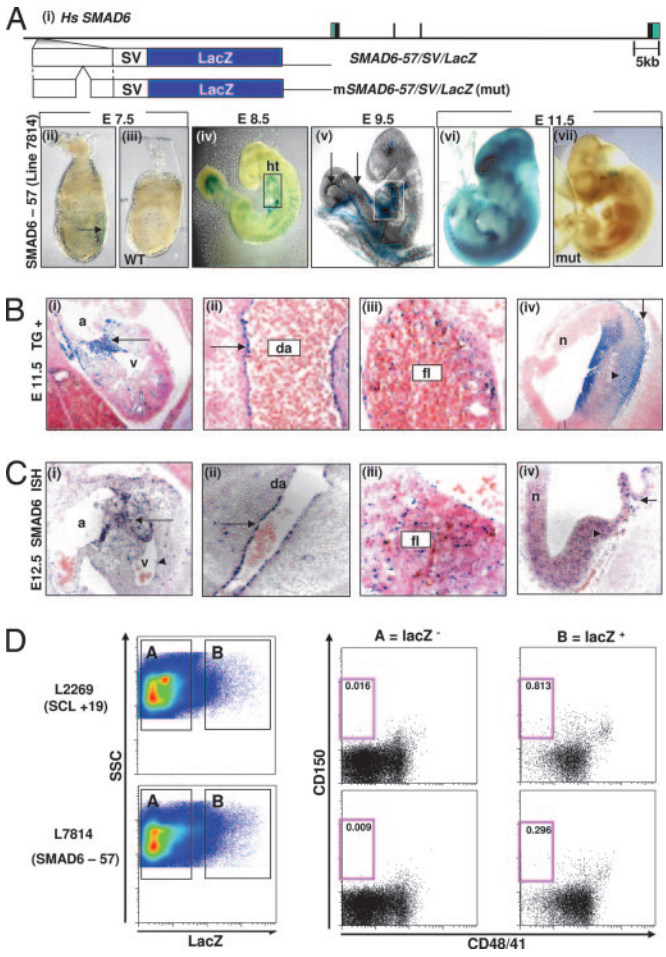

Fig. 3. The E/E/G cluster in the *SMAD6* −57 enhancer directs reporter activity to blood, blood vessels, heart, and brain. (*Ai*) Schematic diagram of the human *SMAD6* locus. A fragment of DNA corresponding to the *SMAD6* −57 region and a deletion mutant lacking the E/E/G motif were subcloned into the SV/LacZ reporter vector and used to generate transgenic mice. (*Aii–Avii*) E7.5-E11.5 X-Gal-stained whole-mount embryos from a *SMAD6* −57/SV/LacZ transgenic line (7814) (*Aii*) E7.5 transgenic embryo showing staining within the primitive streak (arrow). (*Aiii*) E7.5 WT embryo with no staining. (*Aiv*) E8.5 transgenic embryo showing staining of the primitive heart (boxed). (*Av*) E9.5 transgenic embryo showing staining of the cardiac chambers (boxed) and blood vessels (arrows). (*Avi*) E11.5 transgenic embryo showing widespread endothelial/hematopoietic staining resulting in a generalized blue color. (*Avii*) E11.5 mutant transgenic embryo appears pale by contrast, owing to a lack of endothelial/hematopoietic staining. (*Bi–Biv*) Histological sections of a E11.5 X-Gal-stained transgenic embryo. (*Bi*) Heart, showing staining of the endocardium and endocardial cushions (arrow). (*Bii*) Dorsal aorta, showing staining of the endothelium (arrow). (*Biii*) Fetal liver, showing staining of blood (round) and endothelial (flat) cells. (*Biv*) Brain, showing staining of blood vessels (arrow) and brain (arrowhead). (*Ci–Civ*) ISH for Smad6 RNA expression in E12.5 embryos. (*Ci*) Heart, showing Smad6 expression in the cardiac cushions (arrow) and the endocardium (arrowhead). (*Cii*) Dorsal aorta, showing Smad6 expression in the aortic wall (arrow). (*Ciii*) Fetal liver, showing Smad6-positive cells scattered through the parenchyma. (*Civ*) Smad6 expression in the brain (arrowhead) and surrounding blood vessels (arrow). (*D*) Flow cytometry of bone marrow from *Scl* 6E5/lacZ/3′enh (L2269) and *SMAD6* −57/SV/LacZ (L7814) transgenic mice. The *Scl* +19 and *SMAD6* −57 transgenes target LacZ expression to a similar proportion of bone marrow cells. The LacZ-positive cells in L7814 (box B) express ≈4-fold more Smad6 by quantitative RT-PCR than LacZ-negative cells (box A). Cells targeted by the *Scl* +19 and *SMAD6* −57 transgenes (box B) are enriched in $CD150^+/48^-/41^-$ (1:2.2 = HSC) cells and are relatively deficient in $CD150^-/48^-/41^-$ (lineage committed nonproliferating) cells. a, atrium; da, dorsal aorta; fl, fetal liver; h, heart; isv, intersomitic vessel; n, neural; s, somite; v, ventricle.

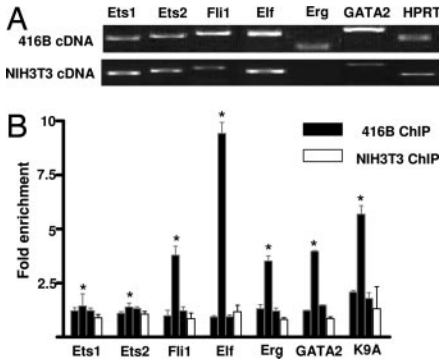

Fig. 4. The *Smad6* −57 enhancer in 416B cells is bound *in vivo* by Fli-1, Elf, Erg, and GATA2. (*A*) RT-PCR expression profile of selected ETS and GATA transcription factors in 416B hematopoietic cells and NIH 3T3 fibroblasts. Erg is not expressed in NIH 3T3 fibroblasts. (*B*) ChIP assays were performed on 416B hematopoietic progenitors (filled bars) and NIH 3T3 fibroblasts (open bars) and analyzed by real-time PCR. The levels of enrichment at *Smad6* −57 (filled bars with asterisk) and flanking regions *Smad6* −58 and −56 (filled bars without asterisk) were normalized to control IgG and plotted as fold increase over enrichments at a control region.

that obtained with a control rabbit antibody and plotted as fold increase over that measured at a control region (the α-fetoprotein promoter) (Fig. 4*B*). The *Smad6* −57 enhancer has an active chromatin mark (H3K9A) in 416B cells but not NIH 3T3 cells. In 416B cells, there was specific enrichment of Fli1, Elf, GATA-2, and Erg at the *Smad6* −57 enhancer (Fig. 4*B*, filled bars with asterisk) but not at −58 or −56 regions flanking the enhancer (Fig. 4*B*, filled bars without asterisk) or in NIH 3T3 cells (Fig. 4*B*, open bars). These data demonstrate that Fli1, Elf, Erg, and GATA2 are recruited to the *Smad6* −57 enhancer as they are to the *Scl* +19, *Fli1* +12, and *Prh* +1 enhancers. Taken together, our results link the *Scl* +19, *Fli1* +12, *Prh* +1, and *Smad6* −57 enhancers in a transcriptional network that regulates early embryonic blood and endothelial development.

Smad6 and Smad1 Regulate Runx1 Activity. The SMADs are the intracellular mediators of the TGF-β superfamily of ligands, which include the TGF-β isoforms, BMPs, and other related factors. SMAD6 is an inhibitory SMAD that modulates BMP signaling at multiple levels, and SMAD1 is a receptor SMAD that transduces BMP signals to the nucleus. BMP4 signals are also modulated by the transcriptional up-regulation of SMAD6 by SMAD1 (13). Consistent with the polarized expression of BMP4 in the human aorta–gonads–mesonephros (AGM) (14), we show that Bmp4 transcripts were concentrated along the ventral surface of the hemogenic dorsal aorta in E10.5 mouse embryos (Fig. 5*Ai*). Definitive hematopoiesis is established in the embryo at E10.5 and is marked by the emergence of HSCs in the AGM. Given that Smad6 is expressed in the dorsal aorta in the AGM at E12.5 (Fig. 3*Cii*), we determined its expression profile at E10.5. Smad6, like Bmp4, is concentrated along the hemogenic ventral aspect of the AGM (Fig. 5*Aii*), reminiscent of the expression pattern of Runx1 (compare Fig. 5 *Ai* and *Aii* with *Aiii*), a transcription factor required for the development of definitive HSCs in the AGM (15, 16).

Members of the RUNX family of transcription factors have been identified as key targets of TGF-β superfamily signaling and have been shown to bind the receptor SMADs, SMAD1/5 and SMAD2/3, which modulate BMP and TGF-β signaling, respectively (17). Furthermore, Smad6 had been shown to bind Runx2 (a bone-specific transcription factor) and, by acting as an adaptor, mediate Smurf1 (a E3 ubiquitin ligase) -induced proteosomal degradation of Runx2 (18). Given the temporal and spatial overlap of Smad6 with Runx1 expression in the AGM region, we investigated a possible role for Smad6 in modulating Runx1 activity. Cos-7 cells were therefore transfected with Myc-Runx1 and Flag-Smad6 expression plasmids. After immunoprecipitation using a mouse anti-Flag antibody, complex formation of Smad6 with Runx1 was demonstrated by Western blot analysis using a rabbit anti-Myc antibody (Fig. 5*Bi*). This interaction was confirmed by GST pull-down assays using *in vitro*-translated Smad6 that was shown to bind to the GST-Runx1 fusion protein, but not GST alone (Fig. 5*Bii*, compare lanes 2 and 3).

To test the effect of Smad6 expression on Runx1, we constructed a reporter plasmid responsive to Runx1 activity (Fig. 5*Ci*). The Runx1 P1 promoter contains three consensus binding sites for RUNX (ACCACA), and its activity was increased ≈2.5-fold by cotransfecting a Runx1 expression plasmid (Fig. 5*Cii*), consistent with a predicted autoregulatory loop (19). Cotransfection with either Smad6 or Smurf1 alone did not result in a significant reduction of Runx1-mediated luciferase activity, but cotransfection with Smad6 and Smurf1 reduced Runx1 mediated luciferase activity to baseline (Fig. 5*Cii*). Similar results were obtained by using an alternate Runx1 reporter, pBXH2-LTR-luc, which uses the U3 region of the long terminal repeat of the BXH2 retrovirus (see SI). To correlate this reduction in luciferase activity with loss of Runx1 protein, Cos-7 cells were also transfected with Myc-Runx1 alone or in combination with Smad6 and/or Smurf1 and immunoblotted for Myc-Runx1. GFP expression was used to control for transfection efficiency and protein loading. Runx1 levels were reduced by ≈1/3 when cotransfected with both Smad6 and Smurf1 (Fig. 5*Ciii*). Therefore, Smad6 interacts with Runx1 and, in conjunction with Smurf1, is able to inhibit Runx1 activity.

The 416B cell line, which expresses Smad6, also expresses Runx1 and Smad1 RNA (data not shown). The Runx1 P1 promoter has, in addition to Runx-binding sites, several conserved Smad-binding sites [including an overlapping Runx/Smad-binding site shown in Fig. 5*C* (1)]. We performed ChIP assays in 416B cells and demonstrated that histone acetylation (H3K9A) at the Runx1 P1 promoter was ≈50-fold higher than the control region (promoter of α-fetoprotein, a nonexpressed gene) (Fig. 5*D*). Runx1 and Smad1 were also enriched relative to the control region by ≈9- and ≈5-fold, respectively. The functional significance of this binding was investigated by transfecting Cos-7 cells with a Smad1 expression plasmid, with and without a Runx1 expression plasmid (Fig. 5*E*). Smad1 was able to transactivate the Runx1 P1 reporter and, when coexpressed with Runx1, was able to enhance Runx1-mediated luciferase activity (Fig. 5*E*). These data provide a direct link between Smad1, a transducer of Bmp4 signaling, and Runx1 activity.

Taken together, our results suggest a model in which ETS and GATA factors directly regulate the expression of key hematopoietic transcription factors, such as Scl, and indirectly regulate others, such as Runx1, by a common association with the Smad signaling pathway (Fig. 6).

Discussion

Prediction of Mammalian Hematopoietic Enhancer Elements. We show in this article that an understanding of combinatorial binding mechanisms can be used to successfully identify novel regulatory elements with similar function. The predictive power of our approach relies on the prior identification of tissue-specific combinations of transcription factor-binding sites. Our analysis was limited to the well characterized regulatory code of the SCL +19 enhancer but would be applicable to combinatorial regulatory codes of other key hematopoietic genes and, indeed, genes expressed in other tissues.

Bmp4 Signaling and Runx1 Activity in the AGM Region. Runx1 is required for definitive hematopoiesis (20, 21), where it is thought to regulate the emergence of HSCs in the AGM region (15, 16). The level of Runx1 expression is important; germ-line deletion of both

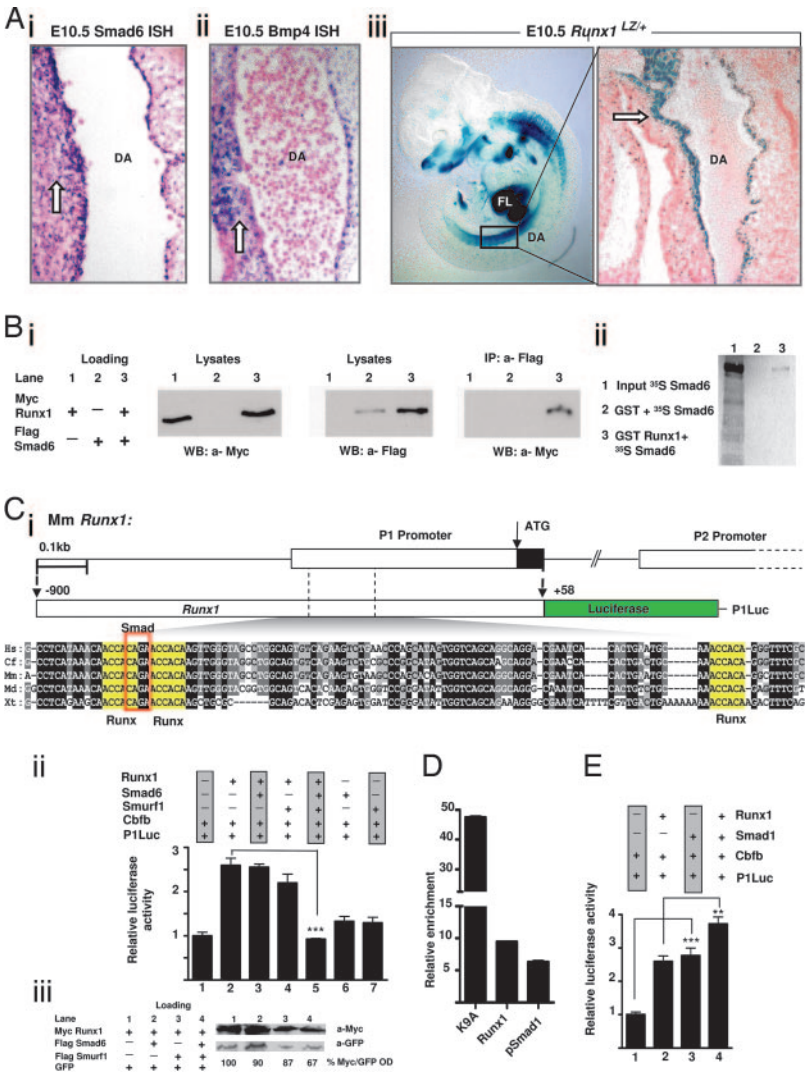

Fig. 5. Integration of Bmp4, Smad6, and Runx1 activity. (*A*) Bmp4 and Smad6 expression in the dorsal aorta at E10.5 matches the expression of Runx1. (*Ai*) ISH for Bmp4 RNA. Dorsal aorta of a E10.5 embryo, showing subendothelial expression of Bmp4 concentrated along the ventral aspect (arrow) in the AGM. (*Aii*) ISH for Smad6 RNA, showing endothelial and subendothelial expression of Smad6 concentrated along the ventral floor of the dorsal aorta. (*Aiii*) A whole-mount X-Gal-stained and cleared, E10.5 Runx1$^{LZ/+}$ embryo, showing prominent LacZ expression in the fetal liver (fl) and dorsal aorta (da) in the AGM. Higher magnification of *Inset*, showing prominent LacZ expression along the ventral surface of the dorsal aorta (arrow). (*B*) Smad6 forms a stable protein complex with Runx1. (*Bi*) Cos-7 cells were transfected with Flag-Smad6 and Myc-Runx1 expression plasmids. Cell lysates were immunoprecipitated (IP) with a mouse anti-Flag antibody and resolved by Western blot by using a rabbit anti-Myc antibody. Runx1 coprecipitated with Smad6 (see IP: α-Flag, lane 3). (*Bii*) *In vitro*-translated Smad6 (lane 1, input) was bound specifically by GST–Runx1 (lane 3) but not GST alone (lane 2) in pull-down experiments. (*C*) Smad6 mediates a reduction in Runx1 levels. (*Ci*) Runx1 transcripts originate from alternate promoters (P1 and P2). The P1 promoter has several conserved Runx- and Smad-binding sites. Three Runx-binding sites (yellow) and a Smad box are present within an ≈120-bp region. Hs, human; Cf, dog; Mm, mouse; Md, opossum; Xt, frog (*Cii*) A fragment of the P1 promoter was subcloned into a promoterless luciferase reporter vector (P1Luc) and used to monitor Runx1 promoter activity. Transfection with *Runx1* resulted in ≈2.5-fold increase in luciferase activity. Cotransfection with *Smad6* and *Smurf1*, however, reduced Runx1-mediated luciferase activity to baseline. (*Ciii*) The reduction in Runx1 activity in *Cii* correlates with a reduction in Runx1 protein expression. Myc-Runx1 band ODs are normalized to their respective GFP ODs and are reported as a percentage of the Myc-Runx1/GFP OD in lane 1. (*D*) The *Runx1* P1 promoter is in an open chromatin configuration (enrichment of K9A at histone H3) in 416B cells and is bound by Runx1 and P-Smad1 *in vivo*. (*E*) *Smad1* transactivates the *Runx1* P1 promoter.

Runx1 alleles results in the absence of HSCs, whereas haploinsufficiency results in early emergence of HSCs in the AGM (15, 16, 22).

Bmp4 induces the formation of mesoderm and hematopoietic precursors, and inhibition of Bmp4 signaling impairs ventral mesoderm formation and commitment of blood and endothelial precursors (23). In addition, BMP4 has also been shown to enhance stem cell activity during *in vitro* culture of human cord blood cells (24) and hematopoietic differentiation of human ES cells (25). Early hematopoietic expression of Scl and Runx1 in embryos requires Bmp4 signaling (26), yet the mechanisms by which these genes respond to Bmp4 are obscure. Our data show that Bmp4 and Runx1 are coexpressed in the AGM at E10.5 during mouse embryonic development and that P-Smad1 (the transducer of Bmp4 activity) is bound *in vivo* to the Runx1 promoter in a hematopoietic progenitor cell line and transactivates the Runx1 promoter.

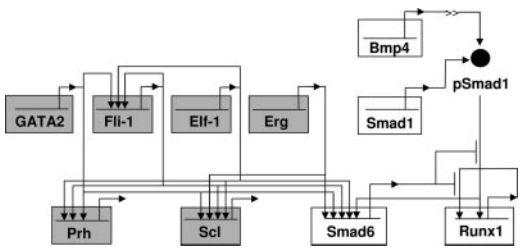

Fig. 6. Integration of Runx1 and the BMP signaling pathway into the SCL transcriptional network. A schematic diagram of an emerging hematopoietic transcriptional network operating within the embryo during the establishment of definitive hematopoiesis in the AGM.

The control of Bmp dose is vital for normal development of the various mesodermal compartments and also for normal hematopoiesis (27). However, although Smads in general play an essential role in hematopoiesis (reviewed in ref. 28), the specific role of SMAD6 in hematopoiesis is unknown. Smad6 mutant mice are viable, with cardiovascular defects (29), but because only the C-terminal MH2 domain was disrupted, and the N-terminal region is required for full Smad6 activity, these mice may not represent the phenotype of a true null allele (30). Although the level of Runx1 activity in the AGM is important in setting the tempo of HSC emergence (15), little is known about mechanisms by which Runx1 activity is regulated other than that Runx1 probably promotes its own expression (15). Runx1 expression in the AGM region is rapidly down-regulated after E12, whereas Smad6 and other components of the SCL network continue to be expressed, raising the possibility that Smad6 is involved in down-regulating Runx1. Our data show that Smad6 forms a complex with Runx1 and modulates its activity, at least in part, by promoting Smurf1-mediated protein degradation. Another effect of SMAD6 modulation of Runx1

activity could be the inhibition of Runx1: R-SMAD-mediated transcriptional regulatory activity (17). Furthermore, Bmp4 upregulates Runx1 expression by an input (P-Smad1 activity), which, itself, is modulated by Smad6 (31). Together, our data suggest that Smad6 establishes a rheostat that fine-tunes BMP signaling and, by inhibiting Runx1 activity, is likely to control the emergence of HSCs in the AGM.

Interestingly, the other inhibitory SMAD, SMAD7, which can inhibit both TGFβ and BMP signaling, when overexpressed in human umbilical cord blood severe combined immunodeficient repopulating cells, resulted in the *in vivo* expansion of hematopoietic progenitors and increased myeloid differentiation at the expense of lymphoid commitment (32). This phenotype could at least partly be the result of SMAD7 inhibiting Runx1 activity, because conditional deletion of Runx1 in the adult hematopoietic compartment was associated with a myeloproliferative phenotype also characterized by an expansion of early progenitors as well as a defect in the lymphoid compartment (33). This study establishes the first direct link between the SCL transcriptional network, Runx1 activity, and the Bmp signaling pathway.

Materials and Methods

***In Silico* Identification of Clusters of Conserved Binding Sites.** Conserved transcription factor-binding sites in these genomes were identified by using TFBSsearch (34). Binding site clusters were identified as described (35). Only clusters situated in introns and the flanking 100 kb of sequence 5′ and 3′ of Ensembl gene annotation were considered.

Cell Transfection and Analysis. Candidate enhancer sequences were PCR amplified from human genomic DNA and subcloned into the reporter plasmid pGL2-promoter (Promega, Madison, WI), electroporated into 416B cells with pGKneo, and tested in stable transfection assays as described (10). Cos-7 cell transfections were performed by using a Calcium Phosphate-mediated Profection system (Promega). The DNA content transfected in each well was kept constant by transfecting empty vector, pCDNA3. Either pEFBOS-LacZ or piGFP was used to control for transfection efficiency. Western blot and immunoprecipitation were performed 48 h after transfection. Cell lysis, immunoprecipitation, and Western blotting analysis were performed as described (18). Band density was quantified by using a Bio-Rad GS-710 densitometer and QuantityOne software.

GST Pull-Down Assay. GST and GST-Runx1 were expressed, by using pGEX and pGEX-MmRunx1 plasmids, respectively, in Rosetta, competent cells (Novagen) cultured overnight at 30°C, and GST pull-down assays performed by using standard protocols. MmSmad6 protein was prepared by *in vitro* transcription/translation using the TNT system (Promega).

Transgenic Analysis of Candidate Enhancers. Transgenic embryos were generated and stained with X-Gal as detailed in ref. 9). The $Runx1^{LZ/+}$ embryos were from a line generated by North *et al.* (15) and were cleared with Glycerol/1% KOH (36). For *in situ* RNA hybridization, digoxigenin-labeled RNA probes were hybridized by using the Ventana Discovery platform (Tucson, AZ). Smad6 in sorted cell populations was measured by quantitative SYBR green real-time PCR (MX 3000; Stratagene, La Jolla, CA) and normalized to β-actin expression. See SI for details on flow cytometry.

ChIP Assay. ChIP assays were performed in 416B and NIH 3T3 cells as described (37). Enrichment was measured by real-time PCR using SYBR green (Stratagene). The levels of enrichment with specific antibodies (see SI) were normalized to that obtained with a control rabbit antibody and were calculated as a fold increase over that measured at a control region (α-fetoprotein promoter).

We thank D. Chen and R. Shen for Smad6 and Smurf1 plasmids (University of Rochester Medical Center, Rochester, NY); Y. Ito and M. Osato for Runx1,Cbfβ, and *p*-BXH2 -LTR-luc plasmids (Institute of Molecular and Cell Biology, Singapore); P. ten Dijke and E. Pardali for Smad1 and Smad4 plasmids (Leids Universitair Medisch Centrum, Leiden, The Netherlands); I. Kitabayashi for the GST-Runx1 plasmid (National Cancer Center Research Institute, Tokyo, Japan); A. Lacy-Hulbert for the piGFP plasmid, (Massachusetts Institute of Technology, Cambridge, MA); D. R. Chen and A. Kolb-Kokochinski for ISH (Sanger Institute). This work was supported by grants from the Biotechnology and Biological Sciences Research Council, IBM, the Leukaemia Research Fund, the Wellcome Trust, the Medical Research Council, Cambridge–MIT Institute, and a C. J. Martin/R. G. Menzies Fellowship from the National Health and Medical Research Council (Australia) (to J.E.P.).

1. Michelson AM (2002) *Proc Natl Acad Sci USA* 99:546–548.
2. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, *et al.* (2002) *Science* 295:1669–1678.
3. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, *et al.* (2004) *Science* 303:1378–1381.
4. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, *et al.* (2006) *Genome Res* 16:123–131.
5. Markstein M, Markstein P, Markstein V, Levine MS (2002) *Proc Natl Acad Sci USA* 99:763–768.
6. Donaldson IJ, Chapman M, Kinston S, Landry JR, Knezevic K, Piltz S, Buckley N, Green AR, Göttgens B (2005) *Hum Mol Genet* 14:595–601.
7. Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E, Taipale J (2006) *Cell* 124:47–59.
8. Göttgens B, Green A (2004) *Transcriptional Regulation of Hematopoietic Stem Cells* (Elsevier Academic, Amsterdam).
9. Sanchez M, Göttgens B, Sinclair AM, Stanley M, Begley CG, Hunter S, Green AR (1999) *Development (Cambridge, UK)* 126:3891–3904.
10. Göttgens B, Nastos A, Kinston S, Piltz S, Delabesse EC, Stanley M, Sanchez MJ, Ciau-Uitz A, Patient R, Green AR (2002) *EMBO J* 21:3039–3050.
11. Sanchez MJ, Bockamp EO, Miller J, Gambardella L, Green AR (2001) *Development (Cambridge, UK)* 128:4815–4827.
12. Kiel MJ, Yilmaz OH, Iwashita T, Terhorst C, Morrison SJ (2005) *Cell* 121:1109–1121.
13. Ishida W, Hamamoto T, Kusanagi K, Yagi K, Kawabata M, Takehara K, Sampath TK, Kato M, Miyazono K (2000) *J Biol Chem* 275:6075–6079.
14. Marshall CJ, Kinnon C, Thrasher AJ (2000) *Blood* 96:1591–1593.
15. North T, Gu TL, Stacy T, Wang Q, Howard L, Binder M, Marin-Padilla M, Speck NA (1999) *Development (Cambridge, UK)* 126:2563–2575.
16. Cai Z, de Bruijn M, Ma X, Dortland B, Luteijn T, Downing RJ, Dzierzak E (2000) *Immunity* 13:423–431.
17. Pardali E, Xie XQ, Tsapogas P, Itoh S, Arvanitidis K, Heldin CH, ten Dijke P, Grundstrom T, Sideras P (2000) *J Biol Chem* 275:3552–3560.
18. Shen R, Chen M, Wang YJ, Kaneki H, Xing L, O'Keefe RJ, Chen D (2006) *J Biol Chem* 281:3569–3576.
19. Levanon D, Groner Y (2004) *Oncogene* 23:4211–4219.
20. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR (1996) *Cell* 84:321–330.
21. Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA (1996) *Proc Natl Acad Sci USA* 93:3444–3449.
22. North TE, de Bruijn MF, Stacy T, Talebian L, Lind E, Robin C, Binder M, Dzierzak E, Speck NA (2002) *Immunity* 16:661–672.
23. Maeno M, Mead PE, Kelley C, Xu RH, Kung HF, Suzuki A, Ueno N, Zon LI (1996) *Blood* 88:1965–1972.
24. Bhatia M, Bonnet D, Wu D, Murdoch B, Wrana J, Gallacher L, Dick JE (1999) *J Exp Med* 189:1139–1148.
25. Chadwick K, Wang L, Li L, Menendez P, Murdoch B, Rouleau A, Bhatia M (2003) *Blood* 102:906–915.
26. Mead PE, Kelley CM, Hahn PS, Piedad O, Zon LI (1998) *Development (Cambridge, UK)* 125:2611–2620.
27. Gupta S, Zhu H, Zon LI, Evans T (2006) *Development (Cambridge, UK)* 133:2177–2187.
28. Larsson J, Karlsson S (2005) *Oncogene* 24:5676–5692.
29. Galvin KM, Donovan MJ, Lynch CA, Meyer RI, Paul RJ, Lorenz JN, Fairchild-Huntress V, Dixon KL, Dunmore JH, Gimbrone MA, Jr, *et al.*(2000) *Nat Genet* 24:171–174.
30. Nakayama T, Berg LK, Christian JL (2001) *Mech Dev* 100:251–262.
31. Hata A, Lagna G, Massague J, Hemmati-Brivanlou A (1998) *Genes Dev* 12:186–197.
32. Chadwick K, Shojaei F, Gallacher L, Bhatia M (2005) *Blood* 105:1905–1915.
33. Growney JD, Shigematsu H, Li Z, Lee BH, Adelsperger J, Rowan R, Curley DP, Kutok JL, Akashi K, Williams IR, *et al.* (2005) *Blood* 106:494–504.
34. Chapman MA, Donaldson IJ, Gilbert J, Grafham D, Rogers J, Green AR, Göttgens B (2004) *Genome Res* 14:313–318.
35. Donaldson IJ, Chapman M, Göttgens B (2005) *Bioinformatics* 21:3058–3059.
36. Schatz O, Golenser E, Ben-Arie N (2005) *BioTechniques* 39:650, 652, 654 passim.
37. Forsberg EC, Downs KM, Bresnick EH (2000) *Blood* 96:334–339.