Development 135, 2071-2082 (2008) doi:10.1242/dev.018200

Activation of Eklf expression during hematopoiesis by Gata2 and Smad5 prior to erythroid commitment

Felix Lohmann and James J. Bieker*

The hierarchical progression of stem and progenitor cells to their more-committed progeny is mediated through cell-to-cell signaling pathways and intracellular transcription factor activity. However, the mechanisms that govern the genetic networks underlying lineage fate decisions and differentiation programs remain poorly understood. Here we show how integration of Bmp4 signaling and Gata factor activity controls the progression of hematopoiesis, as exemplified by the regulation of *Eklf* during establishment of the erythroid lineage. Utilizing transgenic reporter assays in differentiating mouse embryonic stem cells as well as in the murine fetal liver, we demonstrate that Eklf expression is initiated prior to erythroid commitment during hematopoiesis. Applying phylogenetic footprinting and in vivo binding studies in combination with newly developed loss-of-function technology in embryoid bodies, we find that Gata2 and Smad5 cooperate to induce *Eklf* in a progenitor population, followed by a switch to Gata1-controlled regulation of *Eklf* transcription upon erythroid commitment. This stage- and lineage-dependent control of Eklf expression defines a novel role for Eklf as a regulator of lineage fate decisions during hematopoiesis.

KEY WORDS: Megakaryocytic-erythroid progenitor, Lineage fate decision, ES cell in vitro differentiation, Eklf (Klf1)

INTRODUCTION

During hematopoiesis, a small pool of uncommitted stem and progenitor cells gives rise to large numbers of terminally differentiated cells of the various lymphoid and myeloid lineages through a process of proliferation and gradual differentiation. This process is characterized by the hierarchical progression of cells from a state of high plasticity to successively more-restricted states of lineage commitment and maturation. The elucidation of the signaling pathways and transcription factor networks that regulate these cell fate decisions and establish lineage-specific differentiation programs is crucial to the understanding of blood cell development in particular and of stem cell biology in general (Orkin and Zon, 2008).

In lower vertebrates and mammals alike, signaling by Bmp4, a member of the Tgfβ-like group of growth factors that signal via the Smad protein family of intracellular effectors, is essential for the establishment of hematopoiesis shortly after gastrulation (Gupta et al., 2006; Maeno et al., 1996; Winnier et al., 1995). In the mouse embryo, Bmp4 induces a mesodermal cell population within the primitive streak at embryonic day (E) 7.0 to give rise to a common progenitor for hematopoietic and endothelial lineages termed the hemangioblast. This progenitor subsequently migrates into the extraembryonic yolk sac where it differentiates into the hematopoietic, endothelial and vascular smooth muscle cell lineages that will form the cardiovascular system of the embryo (Huber et al., 2004).

Of particular importance to the understanding of the mechanisms that underlie these early events of hematopoietic commitment has been the application of embryonic stem (ES) cell in vitro differentiation technology (Keller, 2005). With regard to yolk sac hematopoiesis, ES cell differentiation studies have shown that Bmp4 signaling is required throughout embryoid body (EB) development, from the initial establishment of hematopoietic fate from mesoderm (Park et al., 2004) to its subsequent progression to erythroid commitment (Adelman et al., 2002). In addition to Bmp4 signaling, members of the GATA-motif-binding family of zinc-finger transcription factors play key regulatory roles in hematopoietic development. Gata2 expression occurs early in hematopoiesis and is crucial for the proliferation and expansion of hematopoietic stem cells and progenitors (Tsai et al., 1994), whereas *Gata1* is activated subsequently and is essential for the establishment of erythroid commitment and differentiation (Pevny et al., 1991; Weiss et al., 1994). Importantly, *Gata2* is a direct Bmp4 target gene (Lugus et al., 2007), and Smad-dependent Bmp4 signaling is necessary and sufficient to induce the expression of Gata1 alongside that of erythroid Krüppel-like factor (Eklf; Klf1) (Adelman et al., 2002), another key regulator of erythropoiesis.

Eklf was originally identified as an erythroid-specific transcription factor that functions as an essential regulator of β-like globin switching in red cells (Miller and Bieker, 1993; Nuez et al., 1995; Perkins et al., 1995). However, in contrast to this well-defined role, Eklf must fulfill additional functions during hematopoiesis as it is expressed much earlier, prior to erythroid differentiation. In the developing mouse embryo, *Eklf* message is first detected in situ in the extraembryonic mesoderm of the yolk sac as early as at the neural plate stage, by E7.5 (Southwood et al., 1996), which coincides with the presence of primitive erythroid progenitors (Palis et al., 1999). Similarly, during the differentiation of mouse ES cells in vitro, Eklf expression is activated prior to terminal erythroid differentiation (Adelman et al., 2002). These observations raise the possibility that Eklf plays a role prior to erythroid commitment and lead us to ask where exactly Eklf fits into the transcription factor hierarchy that establishes the hematopoietic program.

In order to address these issues, we have developed novel reporter gene and RNAi-based loss-of-function assays for use during ES cell in vitro differentiation, as well as a transgenic mouse model to characterize the onset of Eklf expression during hematopoiesis. We show that *Eklf* is activated in a progenitor population of erythroid-megakaryocytic potential prior to erythroid commitment and delineate the transcriptional mechanisms that govern its differential

Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, Box 1020, 1 Gustave Levy Place, New York, NY 10029, USA.

*Author for correspondence (e-mail: james.bieker@mssm.edu)

Accepted 3 April 2008

expression in the two compartments. Based on our findings, a model integrating Bmp4- and Gata factor-mediated transcriptional control of hematopoiesis is presented.

MATERIALS AND METHODS

Generation of plox reporter and shRNA plasmids

The Peklf-GFP reporter construct was generated by cloning the *Sac*I to *Nco*I (–950 to the ATG) murine *Eklf* fragment (Chen et al., 1998) into the *Eco*RI site of the plox-plasmid (Kyba et al., 2002). GFP (pEGFP-C_1, Clontech) was inserted downstream of Peklf, replacing the *Eklf* with the GFP translation start codon (ATG), placing GFP upstream of the plox-plasmid *Not*I site, followed by poly(A). The Peklf-intron-GFP construct was generated by inserting murine *Eklf* exon 1 and intron 1 into Peklf-GFP upstream of GFP, mutating *Eklf* exon 1 ATG to GTA to make the GFP start codon the first ATG present in the resulting spliced mRNA. All mutations and deletions were introduced using QuikChange (Stratagene); primer sequences are available upon request.

The plox-GFP-Intron-miR plasmid was generated by inserting the miR-30 backbone (*Sal*I to *Pme*I) of pSM2 (Open Biosystems) into the β-globin intron of pSG5 (Stratagene) at *Mfe*I. The resulting Intron-miR sequence was placed between the GFP and poly(A) of pEGFP-C_1, subsequently inserting GFP-Intron-miR-poly(A) into the plox-plasmid. shRNAs sequences were generated by PCR as described previously (Paddison et al., 2004) and cloned into the miR-30 backbone (*Xho*I to *Eco*RI) of plox-GFP-Intron-miR. shRNA sequences against NM_008541 (shRNA-1 is HP_248638; shRNA-2 is HP_257681) were obtained from RNAiCodex (http://codex.cshl.edu/scripts/newmain.pl). The non-silencing shRNA-control was purchased from Open Biosystems (RHS1703).

ES cell culture and in vitro differentiation

Ainv18 mouse ES cells (Kyba et al., 2002) were maintained and differentiated according to established protocols (Choi et al., 2005; Kennedy and Keller, 2003) and have been described previously (Manwani et al., 2007; Zafonte et al., 2007). Each ES cell clone was generated by targeting the respective plox construct to the Ainv18 homing site by co-electroporation with a Cre recombinase expression plasmid and subsequent antibiotic selection. Site-specific integration of the plox construct was checked by PCR on genomic DNA. Retention of hematopoietic differentiation capacity as measured by Flk1 expression levels on EB day 4 was verified for each clone. All reporter assay experiments were performed in the absence of doxycycline (dox) (Sigma); shRNA expression was induced by adding of 1 μg/ml dox.

Flow cytometry and colony assays

Single-cell suspensions of murine fetal livers or EBs were prepared and stained using the following antibodies: CD71-PE, Ter119-APC (eBiosciences), Flk1-PE, Ter119-PE, CD41-APC, Gr1-APC, Mac1-APC, CD42d-APC (BD Biosciences). Flow cytometry data were analyzed using FlowJo (TreeStar). Sorted cells were cultured in MethoCult M3234 (StemCell Technologies) supplemented with cytokines according to manufacturer's instructions. Cytospins were fixed with May-Grünwald Solution (Fluka) and stained with Giemsa Solution (Sigma), and pictures were taken using an Axioplan2 microscope (Zeiss) with a 100× or 63× oil-immersion objective.

Phylogenetic sequence comparison and regulatory site prediction

Genomic sequences of mammalian *Eklf* or *Gata1* loci were obtained from the ENSEMBL Genome Browser between February and August, 2005. Each alignment and the subsequent identification of conserved blocks of sequence homology and putative transcription factor binding sites were performed according to Chapman et al. (Chapman et al., 2004). Smad binding motif consensus sequences: 5′-AGAC (and complement 5′-GTCT) (Shi et al., 1998; Zawel et al., 1998); 5′-GTCTT (Lee et al., 2004); 5′-CAGC (Korchynskyi and ten Dijke, 2002); 5′-GCTG (Benchabane and Wrana, 2003); 5′-TGGAGC (Hata et al., 2000); 5′-TGGACC (Oren et al., 2005). Sp1/E2F/Smad binding motif: 5′-CCGCCC (Chen et al., 2002; Frederick et al., 2004; Yagi et al., 2002).

Chromatin immunoprecipitation (ChIP) assay

ChIP experiments were performed as previously described for G1E-ER-Gata1 cells (Im et al., 2005) or EBs (Lugus et al., 2007). G1E-ER-Gata1 cells were treated with 0.1 μM 17β-estradiol (Sigma) for 14 hours to induce ER-Gata1 nuclear translocation prior to cross-linking. Anti-Gata2 and anti-Gata1 rabbit polyclonal antibodies were a gift from Emery H. Bresnick (University of Wisconsin Medical School, Madison, WI). Normal rabbit IgG (Upstate) was used as a negative control. Primers were designed using Primer Express software (Applied Biosystems). Primer specificity was verified by showing that each primer pair generated a single amplicon and dissociation curve; primer sequences are available upon request.

Gene expression analysis by qRT-PCR

Purified total RNA (1 μg) was reverse-transcribed using the ImProm-II RT Kit (Promega). *Eklf*, *Gata1*, *Gata2*, *GFP*, globin βH1, *Smad1*, *Smad5*, *Gapdh* mRNA or 18S rRNA expression was detected using pre-made TaqMan Gene Expression Assays (Applied Biosystems) according to the manufacturer's protocol. Gene expression levels were calculated using a relative standard curve made from serially diluted cDNA samples, followed by normalization to 18S rRNA or *Gapdh* mRNA. For each clone and time point, three independent RNA samples derived from three independent EB differentiations performed in parallel were used to calculate the arithmetic mean (average) and s.d. *P*-values were computed using Student's *t*-test.

Protein expression analysis by western blot

Primary antibodies: Eklf (Southwood et al., 1996), Gata1 (N-6, Santa Cruz Biotechnology), Gata2 (H-116, Santa Cruz Biotechnology), Hsp90 (H-114, Santa Cruz Biotechnology), GFP (JL-8, Clontech). Secondary antibodies: goat anti-rabbit IgG(H+L)-HRP (Southern Biotechnologies); goat anti-mouse IgG(H+L)-HRP and goat anti-rat IgG(H+L)-HRP (Pierce).

RESULTS

Eklf expression is initiated prior to erythroid commitment in the fetal liver

To characterize the expression pattern of Eklf during hematopoiesis, we developed a reporter expression cassette for use in a transgenic mouse model. To this end, a cDNA encoding green fluorescent protein (GFP) was placed under the control of the previously described 950 bp mouse *Eklf* promoter (Peklf) (Chen et al., 1998), which had been shown to direct erythroid-specific transgene expression in developing mouse embryos (Xue et al., 2004). However, to avoid the pitfalls usually associated with the random integration of a transgene, such as position-effect variegation, we took advantage of a site-specific integration mechanism featured in the previously engineered Ainv18 ES cell line (Kyba et al., 2002). Using Cre-mediated recombination, we targeted a single copy of the Peklf-GFP construct in a uni-directional fashion to a loxP homing site in a genomic locus of open chromatin conformation and selected transgenic ES cell clones based on antibiotic resistance (Fig. 1A). After establishing that the *tet* operon (tetOP) contained in Ainv18 ES cells does not interfere with the transcriptional control conferred by Peklf (see Fig. S1 in the supplementary material), we created a Peklf-GFP 'reporter mouse' by injecting ES cells of one Peklf-GFP Ainv18 clone into mouse embryos at the blastocyst stage and breeding chimeric animals to homozygosity.

Next we investigated the pattern of Eklf expression in vivo by analyzing Peklf-GFP transgene activity during fetal liver hematopoiesis, as it is the site of blood cell production between E11 and E16 in the developing mouse embryo. Flow cytometric analysis of fetal livers from Peklf-GFP embryos revealed that ~36% of all cells were GFP^+ at E13.5 (Fig. 1B). To correlate GFP expression with the status of erythroid lineage commitment in fetal liver cells, we determined the transferrin receptor (CD71; Tfrc – Mouse Genome Informatics) versus Ter119 (Ly76) cell-surface marker expression profile of fetal liver cells as an indicator of hematopoietic differentiation, as previously described (Zhang et al., 2003). At E13.5, we found ~61% of all fetal liver cells in the committed erythroid compartment, as indicated by their $CD71^{hi}$/$Ter119^-$ to

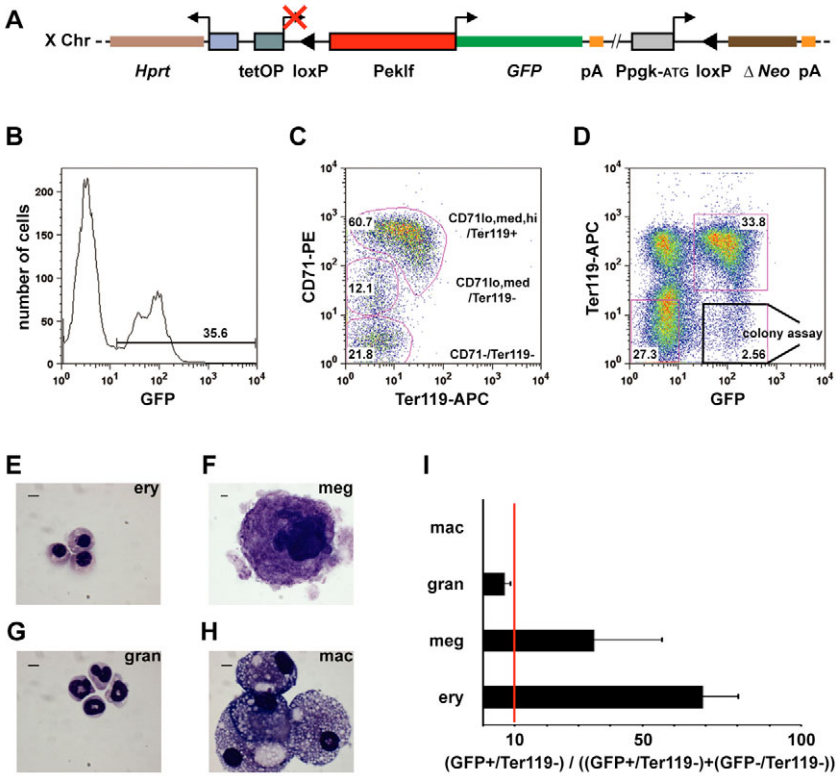

Fig. 1. EKLF expression is initiated prior to erythroid commitment in the mouse fetal liver. (**A**) Schematic of the Peklf-GFP transgene after site-specific, uni-directional integration into the Ainv18 homing site, yielding the Peklf-GFP ES cell clone. (**B-D**) Flow cytometric analysis of GFP (B) and CD71/Ter119 expression (C) in fetal liver cells derived from E13.5 Peklf-GFP embryos, as quantified in Table 1. GFP$^+$/Ter119$^-$ cells selected for colony assays are marked in D. (**E-H**) Erythroid (CFU-E and BFU-E), megakaryocyte, granulocyte and macrophage progenitor-derived colonies from GFP$^+$/Ter119$^-$ cells were quantified and compared with those derived from GFP$^-$/Ter119$^-$ cells. Cells were morphologically assessed by cytospin and May-Grünwald/Giemsa staining. (E) Erythroid cells; (F) megakaryocyte; (G) granulocytes; (H) macrophages. Scale bars: 5 μm. (**I**) Enrichment is defined as a colony number greater than 10% of total colonies (red vertical line) because ~10% of Ter119$^-$ sorted fetal liver cells at E13.5 are GFP$^+$. The arithmetic mean of three independent experiments ±s.d. is shown.

CD71^{lo}/Ter119$^+$ expression profile (Fig. 1C, Table 1). Most GFP$^+$ cells were part of this population as they also expressed the Ter119 marker (~34%, Fig. 1D) and were thus identified as erythroid cells. By contrast, ~12% of fetal liver cells were characterized as progenitors by their low to medium CD71 levels in combination with a lack of Ter119 expression (CD71$^{lo/med}$/Ter119$^-$, Fig. 1C, Table 1). Interestingly, a small fraction of GFP$^+$ cells (~2-3% of total, Fig. 1D; 17% of progenitors, Table 1) were found in this progenitor compartment (CD71$^{lo/med}$/Ter119$^-$) of the murine fetal liver.

Therefore, we next tested directly the progenitor activity of the GFP$^+$/Ter119$^-$ cells in hematopoietic colony assays in culture to investigate whether the Peklf-GFP transgene is indeed active prior to erythroid commitment. Peklf-GFP fetal liver cells were obtained from mouse embryos at E13.5 and separated from differentiating erythroid cells (Ter119$^+$). The remaining pool of fetal liver cells (Ter119$^-$) was sorted into Peklf-GFP-expressing (GFP$^+$/Ter119$^-$, 2.56%) and non-expressing (GFP$^-$/Ter119$^-$, 27.3%) populations (Fig. 1D), followed by seeding in semisolid media in the presence of appropriate growth factors.

The sorted GFP$^+$/Ter119$^-$ population gave rise to a substantial enrichment of erythroid progenitor-derived colonies (colony-forming unit erythroid, CFU-E; burst-forming unit erythroid, BFU-E) in culture (Fig. 1E,I). Most interestingly, it was also enriched for megakaryocytic progenitor-derived colonies, which were present in significant numbers (Fig. 1F,I), but not for granulocytic (Fig. 1G,I) or macrophage colonies (that are readily detectable in assays seeded with GFP$^-$/Ter119$^-$ cells, Fig. 1H). Our data clearly indicate that the *Eklf* promoter (Peklf) activates transcription in progenitor cells well before the onset of erythroid differentiation. More specifically, Peklf-GFP expression preferentially marks progenitors that will give rise to erythroid or megakaryocytic colonies.

According to current models of hematopoiesis, the erythroid and megakaryocytic lineages are derived from a common, bipotential progenitor (megakaryocytic-erythroid progenitor, MEP), which in turn stems from a progenitor common to all myeloid cells (Adolfsson et al., 2005; Forsberg et al., 2006). Therefore, our findings strongly suggest that Eklf is expressed at least as early as in a progenitor population of erythroid-megakaryocytic potential.

Eklf expression is initiated prior to erythroid commitment in differentiating EBs

To better define the onset and pattern of Eklf expression during hematopoiesis in the context of stage- and lineage-specific cell-surface markers and genes, we next employed the differentiation of

Table 1. The Peklf-GFP transgene is expressed in fetal liver progenitors at E13.5

	Cell surface marker	% of total FL	% GFP$^+$ of total FL	% GFP$^+$ within population
CD71$^-$	Ter119$^-$	21.8	0.1	0.41
CD71$^{lo/med}$	Ter119$^-$	12.1	2.1	17.1
CD71$^{lo/med/hi}$	Ter119$^+$	60.7	34.2	56.4

Flow cytometric analysis of GFP expression in relation to lineage commitment and differentiation status of fetal liver (FL) cells in Peklf-GFP embryos at E13.5 based on the CD71/Ter119 cell-surface marker expression profile according to Zhang et al. (Zhang et al., 2003): progenitors and proerythroblasts (CD71$^{lo/med}$/Ter119$^-$); proerythroblasts and early basophilic erythroblasts (CD71^{hi}/Ter119$^{-/lo}$); early and late basophilic erythroblasts (CD71^{hi}/Ter119^{hi}); chromatophilic and orthochromatophilic erythroblasts (CD71^{med}/Ter119^{hi}); late orthochromatophilic erythroblasts and reticulocytes (CD71^{lo}/Ter119^{hi}). The results of one representative experiment of three are shown.

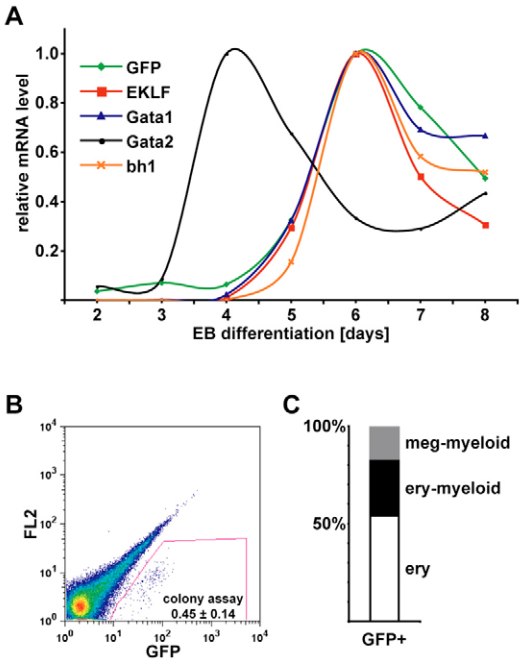

Fig. 2. EKLF expression is initiated prior to erythroid commitment during EB differentiation. (**A**) GFP expression in relation to hematopoietic marker gene expression between days 2 and 8 of mouse EB differentiation. Levels of *GFP*, *Eklf*, *Gata1*, *Gata2* and globin βH1 (bh1) mRNA in samples harvested at the indicated day after EB differentiation were monitored by qRT-PCR and normalized to 18S rRNA. Values are presented relative to the maximum expression level per gene, which was set to 1. (**B**) Flow cytometric analysis of GFP expression at day 5.5 of differentiation in EBs derived from Peklf-GFP ES cells. (**C**) Percentages of progenitor-derived colonies from sorted GFP$^+$ EB cells in one representative experiment.

the Peklf-GFP Ainv18 ES cell clone to EBs as an in vitro assay of hematopoietic development. Throughout an 8-day ES cell differentiation time-course, *GFP* transcription under control of Peklf closely mirrored that of the endogenous *Eklf* gene with regard to onset and levels of expression (Fig. 2A). Expression of Eklf and GFP was initiated at or near day 4 during the stage of EB development that is characterized by the presence of mesoderm-derived hematopoietic progenitor cells as indicated by a peak in Gata2 expression (Fig. 2A) and high levels of the receptor tyrosine kinase Flk1 (Kdr – Mouse Genome Informatics) (Chung et al., 2002) (Table 2). Interestingly, the activation of *Eklf* (and *GFP*) expression occurred simultaneously to that of Gata1 at day 4, and all three were expressed in virtually the same pattern thereafter (Fig. 2A). Following the initial activation of expression, Eklf and GFP levels began to rise rapidly around day 5 of EB development. By day 6, Eklf and GFP expression peaked, whereas the levels of the progenitor markers Gata2 and Flk1 decreased, which coincided with the appearance of committed erythroid cells in EBs as indicated by the expression of the red cell-restricted cell-surface marker Ter119 (Table 2) and embryonic β-like globin (globin βH1; Hbb-bh1 – Mouse Genome Informatics) (Fig. 2A). Subsequently, mature myeloid cells emerged by day 8 of EB development (see Fig. S2 in the supplementary material). At that point, GFP expression among these lineage-committed cells was restricted to erythroid cells and did not occur in cells of the granulocyte-macrophage or megakaryocytic lineages (see Fig. S2 in the supplementary material).

As these findings imply that Eklf is first expressed in a progenitor population during hematopoietic EB differentiation, analogous to its expression pattern during fetal liver hematopoiesis, we next isolated GFP$^+$ cells from EBs at day 5.5 and subjected them to hematopoietic colony assays in culture (Fig. 2B). As expected, GFP$^+$ cells gave rise to colonies of mixed erythroid-myeloid lineages as well as megakaryocytic-myeloid lineages in addition to erythroid colonies (Fig. 2C), demonstrating that Peklf-regulated transcription of *GFP* is activated prior to erythroid commitment in EBs.

Therefore, the expression of Eklf (GFP) during the hematopoietic differentiation of EBs can be divided into two phases: the onset of expression in a progenitor population between days 4 and 6 of EB development, followed by erythroid-restricted expression after day 6.

Comparative phylogenetic sequence analysis of the *Eklf* locus

To elucidate the mechanism regulating Eklf expression in a progenitor population, we sought to identify genomic regulatory regions that control *Eklf* transcription at this early stage in hematopoiesis, in contrast to previous studies, which had been focused on the regulation of Eklf expression in differentiating erythroid cells.

In order to predict putative transcription factor binding sites based on cross-species sequence conservation (Chapman et al., 2004), we generated a comparative, multi-sequence alignment of the *Eklf* locus from different mammalian species. This approach has been demonstrated to be superior in identifying functional cis-regulatory elements to pairwise alignments, which had previously been performed for the *Eklf* locus (Chen et al., 1998).

Fig. 3A shows a five-species alignment across the entire genomic *Eklf* locus between the two neighboring genes. Importantly, aside from exons, only two domains of non-coding sequence display a significant level of conservation between all five species across the 17 kb of the alignment: a stretch of ~1000 bp with three peaks of conservation directly upstream of the *Eklf* transcription start site, and a shorter peak region within the first intron. The former corresponds to the 950 bp region used in the Peklf-GFP construct described above, which harbors an upstream enhancer, two erythroid hypersensitive sites (EHS1 and EHS2) (Chen et al., 1998) and a proximal promoter (Crossley et al., 1994). The latter consists of a

Table 2. Erythroid commitment occurs by day 6 of EB development

	% positive cells				
Cell surface marker	EB day 4	EB day 5	EB day 6	EB day 7	EB day 8
Flk1	46.7±3.3	36.1±0.2	ND	ND	ND
Ter119	ND	0.1±0.1	7.1±0.5	9.3±1.1	12.1±0.7

Flow cytometric analysis of Flk1 and Ter119 cell-surface marker expression in EBs between days 4 and 8 of differentiation (arithmetic mean of three independent experiments ±s.d.).
ND, not determined.

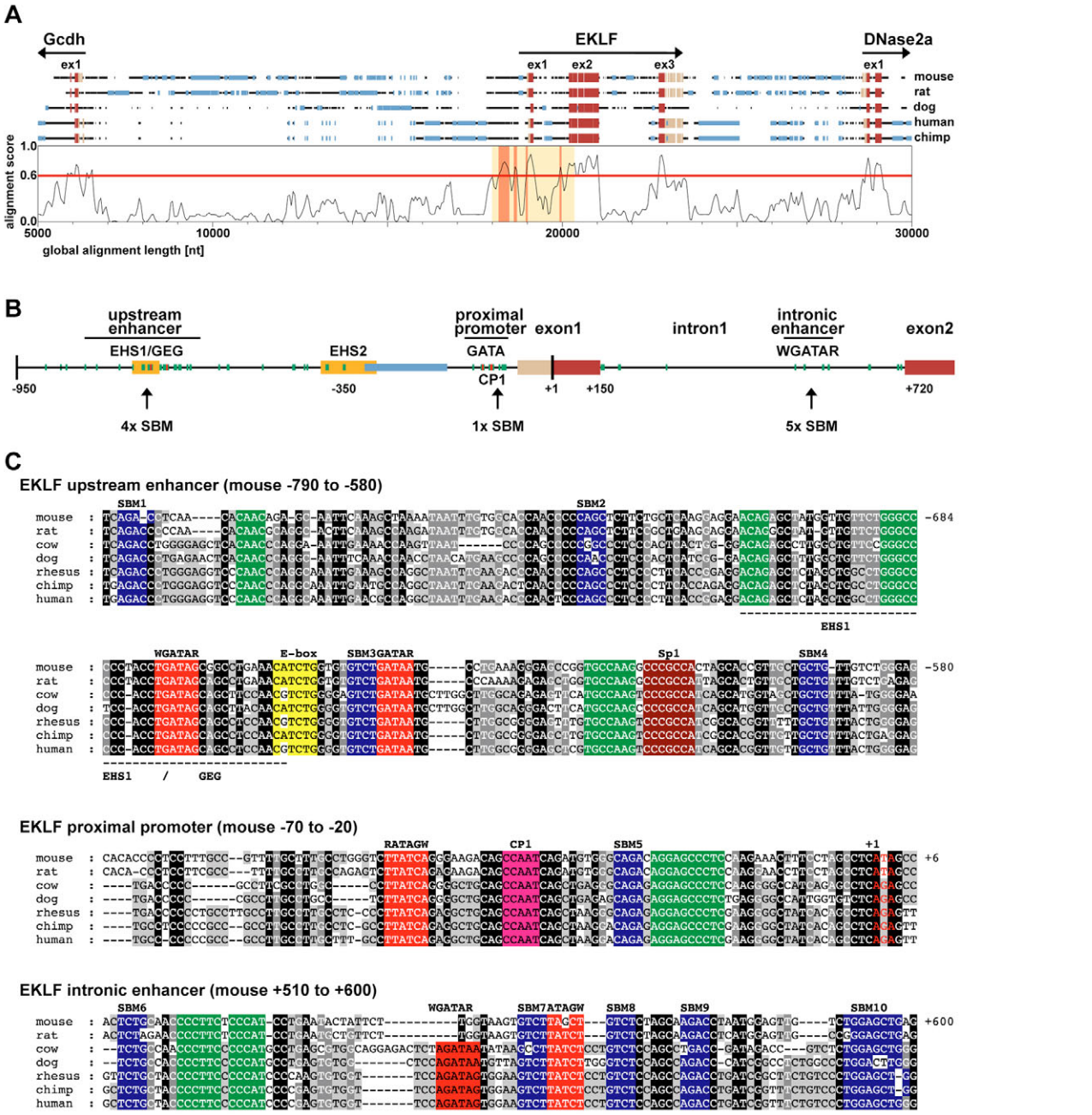

Fig. 3. All three cis-regulatory elements of the *Eklf* locus contain highly conserved Smad binding motifs adjacent to Gata sites. (**A**) Alignment of five mammalian *Eklf* genomic loci encompassing 17 kb (~30 kb global alignment length) surrounding the *Eklf* transcription unit in between the two nearest neighboring genes, *Gcdh* and *Dnase2a*. Each sequence is represented as a black line (breaks demarcate gaps in the alignment), repeats or low complexity DNA are represented as blue bars, untranslated regions (UTR) as light tan bars and exons as brown bars. The degree of sequence conservation between species is expressed as an alignment score (*y*-axis) per nucleotide position (*x*-axis; global length). Peaks of conservation above 0.6 (red line) within non-coding regions are shaded in orange. The 950 bp region upstream of *Eklf* exon 1, exon 1 itself, and intron 1 are highlighted in light yellow. (**B**) Enlarged view of the aligned mouse *Eklf* sequence from –950 bp to the beginning of exon 2. The three conserved *Eklf* cis-regulatory elements are indicated (upstream enhancer, proximal promoter, intronic enhancer) together with blocks of four or more nucleotides of perfect homology (green bars), erythroid hypersensitive sites (EHS1 and EHS2, dark yellow bars) and previously known Gata and Cp1 sites (small red bars). Arrows indicate positions of newly identified Smad binding motifs (SBM) and WGATAR motifs. (**C**) Detail of seven-species alignment of all three *Eklf* cis-regulatory regions (upstream enhancer, proximal promoter, intronic enhancer) at the single nucleotide level. Newly identified Smad binding motifs (SBM1 through SBM10) are highlighted (blue), as are unassigned conserved blocks of perfect homology (green), WGATAR motifs (red), E-box motifs (yellow), Cp1 (pink) and putative Sp1 (brown) sites. The EHS1/GEG region within the upstream enhancer is underlined.

previously unreported sequence element just upstream of the splice branch site preceding the second exon, which we designated as a putative intronic enhancer.

Fig. 3B provides a more detailed view of the aligned mouse sequence from the first repeat upstream of the 950 bp mark to the beginning of exon 2. Sequence blocks of four or more nucleotides

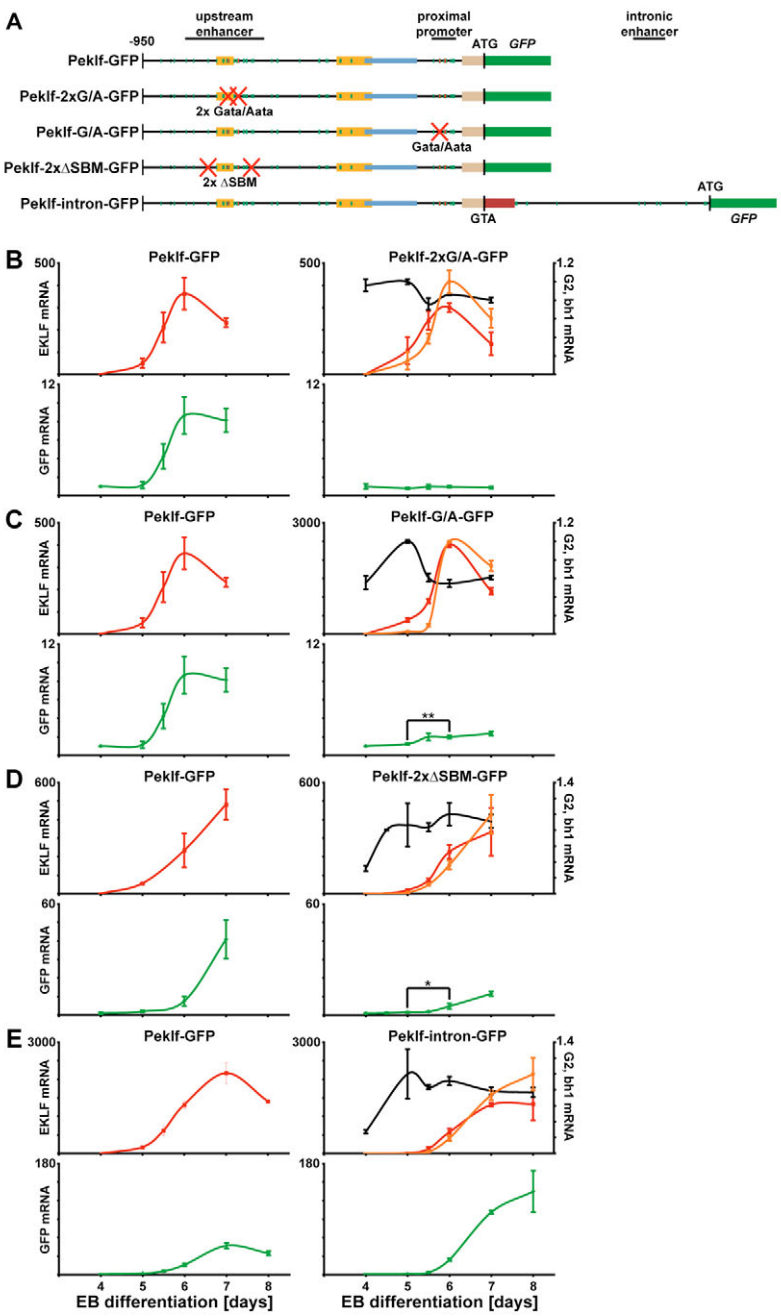

Fig. 4. Gata factor and Smad binding motifs regulate *Eklf* expression in vivo. (**A**) Schematic of the Peklf-GFP reporter construct and its four derivatives, which were targeted to the Ainv18 ES cell line, that carry two GATA motif mutations (2xG/A) in the upstream enhancer (Peklf-2xG/A-GFP), or one GATA motif mutation in the proximal promoter (Peklf-G/A-GFP), two Smad binding motif deletions (2xΔSBM) in the upstream enhancer (Peklf-2xΔSBM-GFP), or an insertion of *Eklf* intron 1 (Peklf-intron-GFP). Positions of *Eklf* cis-regulatory elements and translation start codons (ATG) are indicated and color-coded as in Fig. 3. (**B-E**) qRT-PCR analysis of GFP transgene expression in the four derivative Ainv18 ES cell reporter clones between days 4 and 7 or 8 of EB differentiation. Expression levels were normalized to 18S rRNA or *Gapdh*. For *Eklf* and *GFP*, normalized levels are presented relative to a value of 1 on EB day 4, whereas for *Gata2* (G2) and globin βH1 (bh1) the maximum expression level per gene is set to 1. Arithmetic mean ±s.d.; *, 0.01<*P*<0.05, *n*=3; **, 0.001<*P*<0.01, *n*=3. (Left) Endogenous *Eklf* (red) and *GFP* (green) expression from the Peklf-GFP clone, differentiated in parallel in each experiment, is shown for cross-comparison. (Right) Endogenous *Eklf* (red) and *GFP* (green) expression from each mutant clone (Peklf-2xG/A-GFP, Peklf-G/A-GFP, Peklf-2xΔSBM-GFP, Peklf-intron-GFP) is shown along with that of *Gata2* (black) and globin βH1 (orange). Within each panel, high levels in Gata2 expression are indicative of a progenitor population in differentiating EBs prior to day 5.5, after which the onset of globin βH1 expression is indicative of erythroid commitment (expression patterns as in Fig. 2A).

perfectly conserved between all five species cluster predominantly within the three alignment score peak regions. Most surprisingly, we found the majority of perfectly conserved sequence blocks within the *Eklf* alignment to represent consensus motifs for Smad binding sites. In total, we identified ten potential Smad binding sites across the entire length of the alignment (see Materials and methods for a list of binding motif consensus sequences and references). Notably, these conserved Smad sites are not found in a random distribution throughout the locus, but are confined to the upstream enhancer, the proximal promoter and the intronic enhancer of the *Eklf* gene. In addition, we found blocks of perfect conservation within each of the three cis-elements that do not match any known consensus motifs but which are mostly GC-rich in nature, a feature that has been described to serve Smad binding as well (Ishida et al., 2000).

Fig. 3C shows the three *Eklf* regulatory regions at the nucleotide level in an alignment expanded to seven mammalian species. Surprisingly, all conserved Smad sites, numbered Smad binding motif 1 (SBM1) through SBM10, are in the vicinity of previously reported transcription factor binding sites. These include: two conserved Gata factor binding sites surrounding a bHLH factor binding site [GATA/E-box/GATA (GEG) motif] as well as a putative Sp1 site within the upstream enhancer (Anderson et al., 1998); and a Gata factor site next to a Cp1 (Nfya – Mouse Genome Informatics) binding site within the proximal promoter (Crossley et al., 1994). In addition, we found two previously unreported WGATAR motifs in the putative intronic enhancer in an arrangement similar to that of the upstream enhancer, albeit with a lesser degree of homology as only one of the intronic GATA motifs is conserved between all seven species.

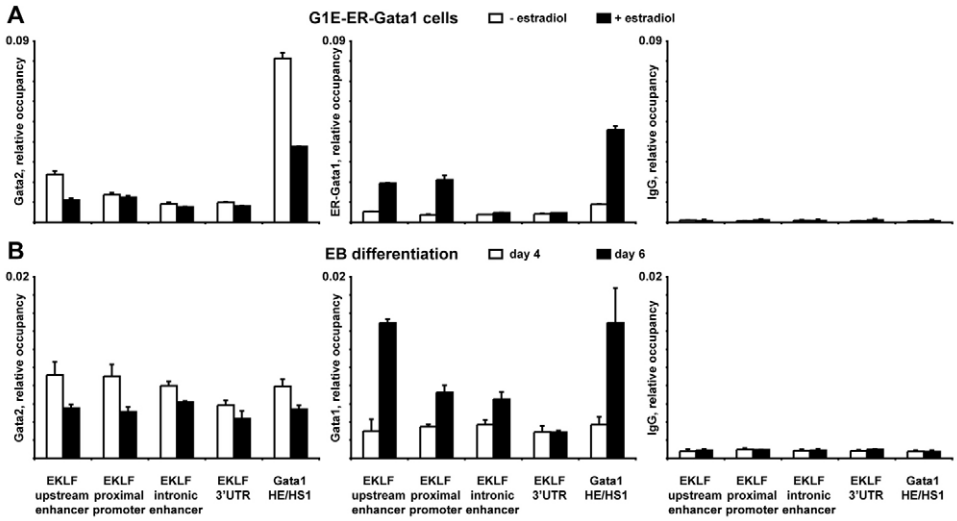

Fig. 5. Gata2 to Gata1 switch in occupancy at *Eklf* cis-regulatory regions during G1E-ER-Gata1 and EB differentiation. Quantitative ChIP analysis of Gata2 and Gata1 binding to cis-regulatory regions of the *Eklf* or *Gata1* locus as indicated. Occupancy values recorded at a non-conserved GATA motif in the *Eklf* 3′UTR serve as baseline signal. (**A**) G1E-ER-Gata1 cells before (white) and after (black) 14 hours of estradiol treatment. (**B**) Differentiating EBs before (day 4, white) or after (day 6, black) erythroid commitment. Normal rabbit immunoglobulin (IgG) ChIP serves as negative control. Arithmetic mean of three independent experiments ±s.d. is shown.

Most importantly, however, all three regulatory regions of the *Eklf* gene display a similar layout, in which Smad sites are clustered around one or two Gata factor binding motifs. The upstream enhancer contains four Smad sites and a putative Sp1/E2F/Smad site surrounding the GEG motif, whereas the proximal promoter contains one Smad site next to the Gata/Cp1 sites. Similarly, the five Smad sites in the intronic enhancer are arranged around the one perfectly conserved Gata site.

Taken together, the high number of conserved Smad binding sites located exclusively within the three cis-regulatory regions of the *Eklf* locus strongly suggest that Smad proteins play a functional role in the transcriptional control of *Eklf* expression in response to Bmp4 signaling. Interestingly, we found such a clustering of conserved Smad binding motifs around a GATA motif within a known enhancer of the *Gata1* gene as well (for a phylogenetic alignment and detailed description of *Gata1* cis-regulatory regions, see Fig. S3 in the supplementary material).

Gata factor and Smad binding motifs regulate EKLF expression

To examine the contribution of the identified cis-regulatory elements and individual transcription factor binding sites to the control of *Eklf* transcription during hematopoiesis, we expanded the ES cell in vitro differentiation reporter assay by generating a set of new GFP transgene reporter constructs that incorporated insertions, deletions or point mutations into *Eklf* cis-regulatory elements based on the alignment in Fig. 3. In total, we created four new Ainv18 ES cell clones as shown in Fig. 4A.

The introduction of a point mutation in each of the two Gata binding sites of the upstream *Eklf* enhancer (Peklf-2xG/A-GFP) completely abolished GFP transgene transcription between days 4 and 7 of EB development (Fig. 4B). By contrast, a small but significant rise in GFP transgene expression levels occurred between days 4 and 5.5 when a point mutation was introduced into the Gata site at the proximal *Eklf* promoter (Peklf-G/A-GFP; Fig. 4C). Only subsequently, upon erythroid commitment after day 5.5, did this Gata site mutation result in a failure to increase GFP transgene levels, whereas mRNA levels of the endogenous *Eklf* gene continued to grow.

Next we tested the potential contribution of the newly identified Smad binding motifs found within the three *Eklf* cis-regulatory regions to the control of GFP transgene expression. As the onset of Eklf expression at day 4 of EB differentiation requires functional Gata binding sites within the upstream *Eklf* enhancer, which also contains Smad binding motifs, we hypothesized that Smad-mediated control of *Eklf* transcription was most likely to occur at the upstream enhancer. To test this, we deleted the two Smad binding motifs in the upstream enhancer that directly surround the GEG motif (Peklf-2xΔSBM-GFP), without altering the GEG motif itself so as to avoid interfering with Gata factor binding. The deletion of the two Smad binding sites abolished the onset of GFP transgene transcription between days 4 and 5.5 of EB development (Fig. 4D). Thereafter however, transgene expression was activated as GFP levels began to rise between days 5.5 and 7. Importantly, the GFP expression pattern resulting from the deletion of Smad sites in the upstream enhancer differed from the pattern produced by the mutation of GATA motifs in the upstream enhancer, arguing that the deletion of Smad sites in the upstream enhancer does not simply impact Gata binding in this region.

As Smads and Gata factors act on the upstream enhancer, we reasoned that a similar mode of transcriptional control might occur at the intronic *Eklf* enhancer, given that both regulatory regions display a similar layout of Gata sites surrounded by Smad binding motifs. Inclusion of the intronic *Eklf* enhancer (Peklf-intron-GFP) increased the maximal level of GFP expression about threefold as compared with levels recorded with Peklf-GFP alone, without altering the overall pattern of transgene expression (Fig. 4E). Furthermore, the intronic *Eklf* enhancer extended the rise of transgene expression until day 8 of EB development, in contrast to expression of the endogenous *Eklf*, which plateaus at day 7. This argues that although the newly identified, highly conserved region within the first intron of *Eklf* is not required for transgene expression, it indeed acts as an enhancer of *Eklf* transcription.

In summary, our EB reporter assay results in the context of a transgene integrated at a homing site suggest that Gata factor and Smad binding at the *Eklf* upstream enhancer are required for the onset of *Eklf* transcription in a progenitor population. By contrast, upon erythroid commitment, Gata factor-mediated control of transcription at the upstream enhancer and the proximal promoter is sufficient for the maintenance of *Eklf* expression. In addition, the highly conserved region within the first *Eklf* intron that was identified through the phylogenetic alignment acts as an enhancer of *Eklf* transcription throughout hematopoiesis.

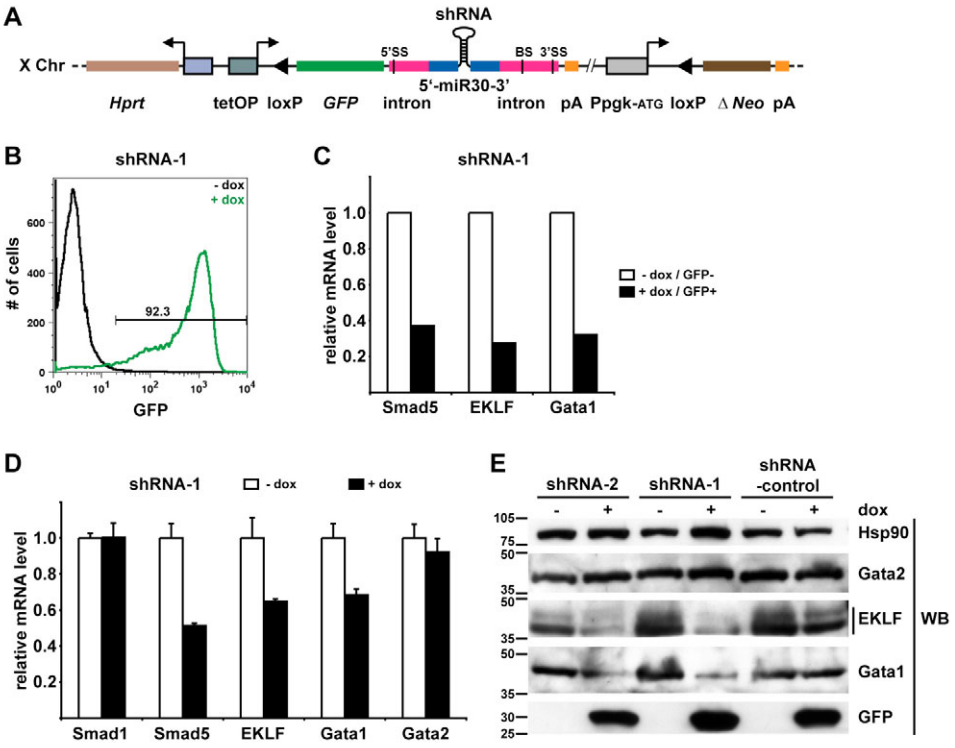

Fig. 6. *Smad5* knockdown in mouse EBs results in reduced Eklf and Gata1 expression. (**A**) Schematic of the GFP-Intron-miR transgene after site-specific, uni-directional integration into the Ainv18 homing site, including the location of the miR-30 backbone and shRNA insertions into the GFP intron. (**B**) Flow cytometric analysis of GFP expression in the Smad5 shRNA-1 Ainv18 ES cell clone in doxycycline treated (24 hours, +dox) or untreated (–dox) EBs harvested at day 5 of differentiation. (**C**) Analysis of *Smad5*, *Eklf* and *Gata1* expression by qRT-PCR in the Smad5 shRNA-1 clone. Untreated (–dox) or doxycycline-treated (24 hours; +dox) cells were harvested at day 5 of EB differentiation. To enrich for transgene-expressing populations, cells from untreated EBs were FACS sorted for GFP-negative cells (–dox/GFP$^-$), whereas cells from dox-treated EBs were FACS sorted for GFP expression (+dox/GFP$^+$) prior to being monitored for RNA expression. Expression levels were normalized to *Gapdh* and untreated samples were set to 1. (**D**) Analysis of *Smad1*, *Smad5*, *Eklf*, *Gata1* and *Gata2* expression by qRT-PCR in the shRNA-1 clone in cells harvested from EBs at day 5 of differentiation without FACS sorting. RNA was isolated from untreated (–dox) or doxycycline-treated (24 hours, +dox) cells as indicated. Expression levels were normalized to *Gapdh* and each untreated sample was set to 1. Arithmetic mean ±s.d. (**E**) Western blot analysis of Gata2, Eklf, Gata1 and GFP protein expression in shRNA-1, shRNA-2 and shRNA-control Ainv18 ES cell clones harvested at day 6 of EB differentiation. Lysates were prepared from unsorted EBs that had been treated with (+dox) or without (–dox) doxycycline for 48 hours. Hsp90 expression levels were used as a loading control. Molecular mass (kDa) markers for each blot are shown to the left.

Gata2 binds to the *Eklf* locus in undifferentiated G1E-ER-Gata1 cells and at the progenitor stage of hematopoietic EB differentiation

Having established that Gata factor binding to the upstream *Eklf* enhancer is required for the onset of transgene expression in EBs at day 4, we next asked which Gata factor regulates Eklf expression at the progenitor stage. Gata2 is an obvious candidate in this regard, as its functional role in hematopoietic progenitors is well documented and we find *Eklf* transcription initiated in EBs at a time when high levels of Gata2, but not of Gata1, are present. To test this, we performed quantitative chromatin immunoprecipitation (ChIP) assays in G1E-ER-Gata1 cells and in differentiating EBs.

The G1E-ER-Gata1 cell line is used as a tool to study erythroid differentiation in vitro as it has been engineered to progress from an undifferentiated to a more differentiated state in a estradiol-dependent manner mediated by Gata1 target gene regulation (Grass et al., 2003; Weiss et al., 1997; Welch et al., 2004), which coincides with a rise in Eklf expression levels (see Fig. S4 in the supplementary material) (see Im et al., 2005).

In undifferentiated G1E-ER-Gata1 cells, Gata2 occupied the *Eklf* upstream enhancer (Fig. 5A). Gata2 occupancy levels at the upstream enhancer were twofold enriched compared with those at the proximal promoter and the intronic enhancer, which did not display any significant Gata2 occupancy as compared with a non-conserved GATA motif in the 3′UTR of the murine *Eklf* locus that served as a negative control. Upon differentiation of G1E-ER-Gata1 cells, Gata2 occupancy levels at the *Eklf* upstream enhancer dropped, which, in turn, coincided with an increase of ER-Gata1 fusion protein binding to the *Eklf* upstream enhancer and to the proximal promoter (Fig. 5A). This switch in Gata factor occupancy argues that ER-Gata1 replaces Gata2 at the upstream enhancer while also binding to the proximal promoter during the induced erythroid differentiation of G1E-ER-Gata1 cells, which correlates with an increase in Eklf expression.

To assess Gata factor-mediated control of Eklf expression under conditions that more closely resemble the endogenous state than does a cell line, we next examined Gata occupancy patterns at the *Eklf* locus during hematopoiesis in differentiating EBs. At day 4 of EB development, an enrichment of Gata2 occupancy was detected at the upstream enhancer and the proximal promoter of the *Eklf* locus, and to a lesser degree at the intronic enhancer (Fig. 5B). Therefore, the onset of Eklf expression at the progenitor stage coincides with Gata2 occupancy at all three cis-regulatory sites of the *Eklf* gene. [Similar to Lugus et al., we detected an overall background signal with the anti-Gata2 ChIP in EBs that was higher than in G1E-ER-Gata1 cells, most likely owing to the cellular heterogeneity of EBs (Lugus et al., 2007).] At day 6 of EB differentiation, Gata2 occupancy levels decreased. By contrast, Gata1 binding occurred at all three conserved *Eklf* cis-regulatory regions, especially at the upstream enhancer (Fig. 5B). Thus, a Gata switch at the *Eklf* locus dependent on the progression of hematopoietic development is observed in EBs. This in vivo occupancy profile of Gata factors suggests a mechanism of transcriptional control by which Gata2 binding mediates the initiation of Eklf expression during hematopoiesis prior to erythroid

commitment, whereas Gata1 replaces Gata2 at the *Eklf* locus during erythroid differentiation, which results in an increase of Eklf expression.

Importantly, a Gata factor switch upon erythroid commitment, similar to the one observed at the *Eklf* locus between days 4 and 6 of EB differentiation, was also detected at a cis-regulatory region of the *Gata1* gene (Fig. 5B), analogous to results previously described in G1E-ER-Gata1 cells (Pal et al., 2004) and reproduced here (Fig. 5A). This regulatory region, termed hypersensitive site 1 (HS1) within the hematopoietic enhancer (HE) of the *Gata1* (G1) gene (G1HE/HS1, see Fig. S3 in the supplementary material), is known to be required for Gata1 expression in a progenitor population (Vyas et al., 1999). Thus, the initial onset of *Eklf* and *Gata1* transcription does not only occur at the same time during the progenitor stage in EBs, but is also regulated in a similar fashion, dependent on Gata2 binding to specific enhancer elements.

The knockdown of Smad5 in EBs results in reduced Eklf and Gata1 expression

As Bmp4 signaling via the Smad pathway is necessary and sufficient to induce Eklf and Gata1 expression at day 4 of EB development (Adelman et al., 2002), we hypothesized that the onset of Eklf expression in a progenitor population prior to erythroid commitment is regulated directly by Smad proteins, in light of the fact that Smad binding sites in the *Eklf* upstream enhancer are required for GFP transgene expression between days 4 and 5.5 of EB development. The most likely Bmp4 effector in this regard is Smad5, as it has been implicated in promoting the establishment of erythroid fate during hematopoiesis (Fuchs et al., 2002; Liu et al., 2003; McReynolds et al., 2007). By contrast, Smad1 acts earlier in hematopoiesis, at the hemangioblast stage (Zafonte et al., 2007).

As no anti-Smad5 antibodies functional in ChIP assays have been reported to date, we developed a novel loss-of-function assay for the Ainv18 ES cell line that allowed us to specifically knockdown Smad5 levels in an inducible, RNAi-mediated manner during EB development so as to test the impact of Smad5 activity on Eklf expression. Utilizing the *tet* operon of the Ainv18 ES cell line (Kyba et al., 2002), we coupled the transcription of a specific, short hairpin RNA (shRNA) within a microRNA (miR) backbone to that of a *GFP* cDNA in a doxycycline (dox)-dependent fashion (Fig. 6A). We generated three different expression cassettes, each of which was stably inserted into the plox homing site of the Ainv18 ES cell line via Cre-mediated recombination. Two of these encoded shRNAs targeting a unique sequence within the *Smad5* mRNA (shRNA-1 and shRNA-2), whereas a third encoded a no-target control shRNA. Fig. 6B shows the tight transcriptional control imparted by the *tet* operon in response to dox treatment as exemplified by GFP expression levels in the case of the shRNA-1 Ainv18 ES cell clone.

As we hypothesized that Smad5-mediated transcriptional control occurs at the onset of *Eklf* expression prior to erythroid commitment, we induced the knockdown of *Smad5* mRNA at day 4 of EB development by treating the shRNA-1 Ainv18 ES cell clone with dox for 24 hours. To quantify the maximal RNAi effect, we sorted EBs at day 5 into dox-treated, GFP-expressing cells (+dox/GFP$^+$) or untreated, GFP-negative cells (–dox/GFP$^-$), and isolated RNA for qRT-PCR analysis. In the shRNA-1-expressing population (+dox/GFP$^+$), the *Smad5* mRNA level was knocked down to ~40% of the level seen in the non-expressing population (–dox/GFP$^-$), demonstrating that the transgenic RNAi assay is functional (Fig. 6C). More importantly, the dox-induced knockdown of *Smad5* coincided with a reduction of *Eklf* and *Gata1* mRNA to ~25% and ~30%, respectively, of the levels observed in untreated cells (–dox/GFP$^-$), suggesting that Smad5 regulates both genes.

Given the strong knockdown effect observed in the shRNA-1 Ainv18 clone upon dox treatment, we reasoned that a cell-sorting step based on GFP expression should not be required to detect the substantial drop in Smad5, Eklf and Gata1 levels. Instead, we used unsorted, whole EBs for a more extended loss-of-function analysis. Analogous to the effect described above, a 24 hour dox treatment of developing EBs from day 4 to 5 resulted in a 50% reduction of *Smad5* mRNA levels as compared with EBs not treated with dox (Fig. 6D). Moreover, the knockdown of *Smad5* correlated once again with a significant drop in *Eklf* and *Gata1* levels, to ~65% and ~70% of those observed in untreated EBs, respectively. Importantly, shRNA-1 targets *Smad5* mRNA specifically, as the levels of *Smad1* mRNA were unperturbed, despite the high degree of sequence homology between the two mRNAs. In addition, *Gata2* mRNA levels were not affected by the dox treatment, demonstrating that the negative effect observed on *Eklf* and *Gata1* transcription between EB day 4 and 5 is in fact a direct and specific consequence of the reduction of *Smad5* mRNA levels.

To further corroborate this point, we next included the shRNA-2 and the shRNA-control clones in our experiments. A western blot analysis of whole-cell lysates made from EBs on day 6 of differentiation after 48 hours of dox treatment revealed that Eklf and Gata1 protein levels are strongly reduced in cells carrying either shRNA-1 or shRNA-2, but not in cells carrying the shRNA-control (Fig. 6E). By contrast, Gata2 protein levels did not display any significant variability between dox-treated and untreated cells in any of the three different clones. Importantly, the observed RNAi effect is due to the specific knockdown of *Smad5*, as two independent shRNAs complementary to different regions of the target mRNA produced the same result, in contrast to the no-target control shRNA that did not alter the expression of any of the genes tested. We conclude that Smad5 protein is required at day 4 of EB development for the correct onset of *Eklf* and *Gata1* transcription.

DISCUSSION

In this report, we demonstrate that the initial activation of *Eklf* during mammalian hematopoietic development occurs prior to erythroid commitment and is likely to be within a progenitor population of erythroid-megakaryocytic potential. Utilizing a bioinformatics-based phylogenetic footprinting approach, we identify three evolutionarily conserved cis-regulatory elements that are critical to the control of Eklf expression throughout hematopoiesis and which share a similar architecture of Smad binding consensus motifs clustered around Gata factor binding sites. Applying novel reporter assay and loss-of-function technology to the ES in vitro differentiation system in combination with in vivo binding studies, we show that the onset of Eklf expression at the progenitor stage is dependent on Gata2 and Smad5, whereas the maintenance of Eklf expression in committed erythroid cells is regulated by Gata1. As the activation of *Eklf* requires Bmp4 signaling (Adelman et al., 2002), we propose a two-tiered, stage- and lineage-dependent model of *Eklf* regulation during hematopoiesis, as described below.

Model of stage- and lineage-dependent activation of *Eklf*

Early in hematopoiesis, Gata2 and Smad5 activate *Eklf* in a cooperative fashion. This integration of the Bmp4 signaling pathway with the Gata2 target gene network is achieved at the upstream

enhancer of *Eklf*, where functional Gata and Smad binding sites are found in close proximity to one another. It results in the low-level expression of *Eklf*. Upon erythroid lineage commitment, Gata1 replaces Gata2 and binds to the *Eklf* upstream enhancer and proximal promoter, regulating *Eklf* expression at high levels throughout erythroid differentiation in a Smad-independent manner. At this stage, Gata1 nucleates an Scl (Tal1 – Mouse Genome Informatics)-containing protein complex occupying the GEG motif at the upstream enhancer (Rodriguez et al., 2005), similar to the GATA/E-box-bound complexes regulating the erythroid-specific expression of other Gata1 target genes (Anguita et al., 2004). In parallel, Gata1 binding occurs at the *Eklf* proximal promoter next to a CCAAT-box, possibly in a complex with Cp1 or a C/EBP family member (Crossley et al., 1994; Gordon et al., 2005), while occupancy changes from Gata2 to Gata1 at the intronic enhancer as well. Such 'Gata-switches' at the same GATA motif have also been shown to occur during the regulation of Gata1 (Pal et al., 2004) (see also below), α-globin (Anguita et al., 2004), Scl (Lugus et al., 2007), Gata2 (Grass et al., 2003; Grass et al., 2006; Kobayashi-Osaki et al., 2005) and Kit expression (Jing et al., 2008), resembling the two-tiered regulatory mechanism described here for *Eklf*. Thus, the layout of transcription factor binding sites contained in the *Eklf* cis-regulatory regions enables a response of dual specificity that is dependent on the stage of hematopoiesis, which in turn results in a transcriptional output of varying degree in a lineage-specific manner.

Gata-Smad cooperation

Parallels to the Gata2-Smad5 cooperation in our model of *Eklf* activation can be found in the regulation of cardiac-specific genes during embryonic heart development (reviewed by Peterkin et al., 2005). Similar to the architecture observed in the cis-regulatory regions of *Eklf*, neighboring Smad and Gata factor binding sites are found in enhancers of *Nkx2.5*, which is the earliest known marker of cardiogenesis, and the presence of both types of consensus motif is required for conveying transcriptional control in response to BMP signaling (Brown et al., 2004; Lee et al., 2004).

Given the requirement of Bmp4 signaling for the establishment and progression of (yolk sac) hematopoiesis as well as the prominence of Gata2-controlled gene expression in hematopoietic progenitors, we propose that the functional cooperation of (BMP-activated) Smads and hematopoietic Gata factors underlies target gene regulation in hematopoietic development and lineage fate decisions, as exemplified here for the case of *Eklf*.

Model of *Gata1* activation at the progenitor stage

In support of the above notion, we show that the onset of Gata1 expression is dependent on Smad5 as well. We identify a cluster of conserved Smad binding motifs next to a known Gata site in the HE/HS1 region of the *Gata1* locus, which we find to be occupied by Gata2 at the progenitor stage during EB differentiation. This *Gata1* enhancer region is required for transgene expression in both the erythroid and megakaryocytic lineage (Vyas et al., 1999). Thus, as the regulation of Gata1 expression is stage- and lineage-dependent, we speculate that the onset of *Gata1* transcription in a progenitor population of erythroid-megakaryocytic potential is mediated via Gata2-Smad5 cooperation at the G1HE/HS1 element, followed by Gata2 replacement and Gata1 autoregulation upon erythroid commitment in a Smad-independent manner.

These models of *Eklf* and *Gata1* activation argue that a layout featuring GATA motifs in combination with Smad binding sites denotes a cis-regulatory element that is utilized at the progenitor stage and could explain the block in erythroid differentiation that is observed upon perturbation or lack of Smad5 expression during hematopoiesis (Fuchs et al., 2002; Liu et al., 2003; McReynolds et al., 2007) or in response to stress erythropoiesis (Porayette and Paulson, 2008).

A novel role for Eklf in erythroid lineage commitment

The addition of *Eklf* to the growing list of Bmp4-regulated genes illustrates that the Bmp4 signaling pathway and Smad activity are required throughout hematopoiesis, as opposed to being necessary only for the initial establishment of hematopoietic fate from mesoderm during development (Lugus et al., 2007; Pimanda et al., 2007; Zafonte et al., 2007). As Gata factors regulate a multitude of genes central to the hematopoietic program, the delineation of mechanisms that integrate the cross-talk between Bmp4 signaling and the Gata factor-controlled gene network is essential to an understanding of hematopoiesis (Loose et al., 2007). Of particular interest in this regard (and for stem cell biology in general) are the mechanisms underlying lineage fate decisions and the accompanying progression from a highly proliferative progenitor state to that of lineage commitment and differentiation.

According to models of lineage fate decisions from a (bipotential) progenitor cell, commitment to one specific lineage over another is established through a cross-antagonistic mechanism of opposing transcription factors (Cantor and Orkin, 2002). Quantitatively, the uncommitted progenitor state is characterized by the co-expression of such antagonistic transcription factors at a low level. However, this priming state is disrupted following a rise in the transcription levels and thereby dominance of one regulatory factor over the other, which subsequently leads to lineage commitment and ultimately lineage differentiation (Huang et al., 2007; Roeder and Glauche, 2006).

Here, we describe how the stage- and lineage-dependent integration of Gata2 activity and Bmp4/Smad5 signaling versus Gata1-anchored complex binding to the same *Eklf* cis-regulatory element translates into such a two-tiered transcriptional profile between the megakaryocytic-erythroid progenitor and differentiating erythroid cells. As the expression of Eklf prior to erythroid commitment has recently been corroborated (Bottardi et al., 2006; Frontelo et al., 2007), studies conducted in parallel in our laboratory show that Eklf indeed plays a directive role in erythroid versus megakaryocytic development in accordance with the antagonistic model described above (Frontelo et al., 2007; Siatecka et al., 2007). Thus, a new role for Eklf as a regulator of lineage fate decisions during hematopoiesis is defined, the misregulation of which could potentially underlie disease mechanisms.

A new 'toolbox' for the ES cell in vitro differentiation system

Expanding on the gain-of-function assay previously described for the Ainv18 ES cell line (Kyba et al., 2002; Manwani et al., 2007; Willey et al., 2006), we provide novel transgenic reporter and loss-of-function assays for use in ES cells and EBs. In combination with in vivo binding studies performed during the in vitro differentiation of ES cells, we are able to delineate the mechanisms of *Eklf* transcriptional regulation at a higher resolution than had previously been achieved, either in transgenic mouse models of Eklf expression (Anderson et al., 2000; Xue et al., 2004) or in gene ablation studies of Gata factors (Pevny et al., 1991; Tsai et al., 1994) and Smad5 (Liu et al., 2003). Together with the rapidly expanding number of protocols available for the differentiation of ES cells in culture, these assays provide a powerful 'toolbox' for the study of genetic

interactions that govern early mouse development and lineage decision processes, similar to techniques established for the fish and frog model systems.

We thank Ian Donaldson for help with SynPlot; Michael Kyba for kindly providing the Ainv18 ES cell clone; Hogune Im, Jeffrey Grass and Emery Bresnick for training, advice and for the gift of Gata1 and Gata2 antibodies; Li Xue for animal husbandry; Paul Gadue, Jesse Lugus and Kyunghee Choi for reagents and advice; Saghi Ghaffari for help in the preparation of the manuscript; and Pilar Frontelo, Deepa Manwani, Miroslawa Siatecka and Tanushri Sengupta for advice throughout the project. This work was supported by NIH PHS grant R01 DK48721 to J.J.B. The Flow Cytometry, Microscopy and Quantitative PCR Shared Research Facilities are supported by MSSM. The Mouse Genetics Shared Research Facility is partially funded by NCI grant R24 CA88302.

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/135/12/2071/DC1

References

Adelman, C. A., Chattopadhyay, S. and Bieker, J. J. (2002). The BMP/BMPR/Smad pathway directs expression of the erythroid-specific EKLF and GATA1 transcription factors during embryoid body differentiation in serum-free media. *Development* **129**, 539-549.

Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O. J., Thoren, L. A. et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295-306.

Anderson, K. P., Crable, S. C. and Lingrel, J. B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J. Biol. Chem.* **273**, 14347-14354.

Anderson, K. P., Crable, S. C. and Lingrel, J. B. (2000). The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. *Blood* **95**, 1652-1655.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G. and Higgs, D. R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J.* **23**, 2841-2852.

Benchabane, H. and Wrana, J. L. (2003). GATA- and Smad1-dependent enhancers in the Smad7 gene differentially interpret bone morphogenetic protein concentrations. *Mol. Cell. Biol.* **23**, 6646-6661.

Bottardi, S., Ross, J., Pierre-Charles, N., Blank, V. and Milot, E. (2006). Lineage-specific activators affect beta-globin locus chromatin in multipotent hematopoietic progenitors. *EMBO J.* **25**, 3586-3595.

Brown, C. O., 3rd, Chi, X., Garcia-Gras, E., Shirai, M., Feng, X. H. and Schwartz, R. J. (2004). The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *J. Biol. Chem.* **279**, 10659-10669.

Cantor, A. B. and Orkin, S. H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-3376.

Chapman, M. A., Donaldson, I. J., Gilbert, J., Grafham, D., Rogers, J., Green, A. R. and Gottgens, B. (2004). Analysis of multiple genomic sequence alignments: a web resource, online tools, and lessons learned from analysis of mammalian SCL loci. *Genome Res.* **14**, 313-318.

Chen, C. R., Kang, Y., Siegel, P. M. and Massague, J. (2002). E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. *Cell* **110**, 19-32.

Chen, X., Reitman, M. and Bieker, J. J. (1998). Chromatin structure and transcriptional control elements of the erythroid Kruppel-like factor (EKLF) gene. *J. Biol. Chem.* **273**, 25031-25040.

Choi, K., Chung, Y. S. and Zhang, W. J. (2005). Hematopoietic and endothelial development of mouse embryonic stem cells in culture. *Methods Mol. Med.* **105**, 359-368.

Chung, Y. S., Zhang, W. J., Arentson, E., Kingsley, P. D., Palis, J. and Choi, K. (2002). Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression. *Development* **129**, 5511-5520.

Crossley, M., Tsang, A. P., Bieker, J. J. and Orkin, S. H. (1994). Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J. Biol. Chem.* **269**, 15440-15444.

Forsberg, E. C., Serwold, T., Kogan, S., Weissman, I. L. and Passegue, E. (2006). New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. *Cell* **126**, 415-426.

Frederick, J. P., Liberati, N. T., Waddell, D. S., Shi, Y. and Wang, X. F. (2004). Transforming growth factor beta-mediated transcriptional repression of c-myc is dependent on direct binding of Smad3 to a novel repressive Smad binding element. *Mol. Cell. Biol.* **24**, 2546-2559.

Frontelo, P., Manwani, D., Galdass, M., Karsunky, H., Lohmann, F., Gallagher, P. G. and Bieker, J. J. (2007). Novel role for EKLF in megakaryocyte lineage commitment. *Blood* **110**, 3871-3880.

Fuchs, O., Simakova, O., Klener, P., Cmejlova, J., Zivny, J., Zavadil, J. and Stopka, T. (2002). Inhibition of Smad5 in human hematopoietic progenitors blocks erythroid differentiation induced by BMP4. *Blood Cells Mol. Dis.* **28**, 221-233.

Gordon, C. T., Fox, V. J., Najdovska, S. and Perkins, A. C. (2005). C/EBPdelta and C/EBPgamma bind the CCAAT-box in the human beta-globin promoter and modulate the activity of the CACC-box binding protein, EKLF. *Biochim. Biophys. Acta* **1729**, 74-80.

Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J. and Bresnick, E. H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc. Natl. Acad. Sci. USA* **100**, 8811-8816.

Grass, J. A., Jing, H., Kim, S. I., Martowicz, M. L., Pal, S., Blobel, G. A. and Bresnick, E. H. (2006). Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. *Mol. Cell. Biol.* **26**, 7056-7067.

Gupta, S., Zhu, H., Zon, L. I. and Evans, T. (2006). BMP signaling restricts hemato-vascular development from lateral mesoderm during somitogenesis. *Development* **133**, 2177-2187.

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A. and Massague, J. (2000). OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. *Cell* **100**, 229-240.

Huang, S., Guo, Y. P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695-713.

Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. and Keller, G. (2004). Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625-630.

Im, H., Grass, J. A., Johnson, K. D., Kim, S.-I., Boyer, M. E., Imbalzano, A. N., Bieker, J. J. and Bresnick, E. H. (2005). Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc. Natl. Acad. Sci. USA* **102**, 17065-17070.

Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., Sampath, T. K., Kato, M. and Miyazono, K. (2000). Smad6 is a Smad1/5-induced smad inhibitor. Characterization of bone morphogenetic protein-responsive element in the mouse Smad6 promoter. *J. Biol. Chem.* **275**, 6075-6079.

Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X. and Blobel, G. A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell* **29**, 232-242.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129-1155.

Kennedy, M. and Keller, G. M. (2003). Hematopoietic commitment of ES cells in culture. *Methods Enzymol.* **365**, 39-59.

Kobayashi-Osaki, M., Ohneda, O., Suzuki, N., Minegishi, N., Yokomizo, T., Takahashi, S., Lim, K. C., Engel, J. D. and Yamamoto, M. (2005). GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene. *Mol. Cell. Biol.* **25**, 7005-7020.

Korchynskyi, O. and ten Dijke, P. (2002). Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. *J. Biol. Chem.* **277**, 4883-4891.

Kyba, M., Perlingeiro, R. C. and Daley, G. Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29-37.

Lee, K. H., Evans, S., Ruan, T. Y. and Lassar, A. B. (2004). SMAD-mediated modulation of YY1 activity regulates the BMP response and cardiac-specific expression of a GATA4/5/6-dependent chick Nkx2.5 enhancer. *Development* **131**, 4709-4723.

Liu, B., Sun, Y., Jiang, F., Zhang, S., Wu, Y., Lan, Y., Yang, X. and Mao, N. (2003). Disruption of Smad5 gene leads to enhanced proliferation of high-proliferative potential precursors during embryonic hematopoiesis. *Blood* **101**, 124-133.

Loose, M., Swiers, G. and Patient, R. (2007). Transcriptional networks regulating hematopoietic cell fate decisions. *Curr. Opin. Hematol.* **14**, 307-314.

Lugus, J. J., Chung, Y. S., Mills, J. C., Kim, S. I., Grass, J. A., Kyba, M., Doherty, J. M., Bresnick, E. H. and Choi, K. (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. *Development* **134**, 393-405.

Maeno, M., Mead, P. E., Kelley, C., Xu, R. H., Kung, H. F., Suzuki, A., Ueno, N. and Zon, L. I. (1996). The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in Xenopus laevis. *Blood* **88**, 1965-1972.

Manwani, D., Galdass, M. and Bieker, J. J. (2007). Altered regulation of beta-like globin genes by a redesigned erythroid transcription factor. *Exp. Hematol.* **35**, 39-47.

McReynolds, L. J., Gupta, S., Figueroa, M. E., Mullins, M. C. and Evans, T. (2007). Smad1 and Smad5 differentially regulate embryonic hematopoiesis. *Blood* **110**, 3881-3890.

Miller, I. J. and Bieker, J. J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol. Cell. Biol.* **13**, 2776-2786.
Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**, 316-318.
Oren, T., Torregroza, I. and Evans, T. (2005). An Oct-1 binding site mediates activation of the gata2 promoter by BMP signaling. *Nucleic Acids Res.* **33**, 4357-4367.
Orkin, S. H. and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644.
Paddison, P. J., Cleary, M., Silva, J. M., Chang, K., Sheth, N., Sachidanandam, R. and Hannon, G. J. (2004). Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nat. Methods* **1**, 163-167.
Pal, S., Cantor, A. B., Johnson, K. D., Moran, T. B., Boyer, M. E., Orkin, S. H. and Bresnick, E. H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc. Natl. Acad. Sci. USA* **101**, 980-985.
Palis, J., Robertson, S., Kennedy, M., Wall, C. and Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073-5084.
Park, C., Afrikanova, I., Chung, Y. S., Zhang, W. J., Arentson, E., Fong, G., Rosendahl, A. and Choi, K. (2004). A hierarchical order of factors in the generation of FLK1- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells. *Development* **131**, 2749-2762.
Perkins, A. C., Sharpe, A. H. and Orkin, S. H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375**, 318-322.
Peterkin, T., Gibson, A., Loose, M. and Patient, R. (2005). The roles of GATA-4, -5 and -6 in vertebrate heart development. *Semin. Cell Dev. Biol.* **16**, 83-94.
Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H. and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-260.
Pimanda, J. E., Donaldson, I. J., de Bruijn, M. F., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J. R., Green, A. R., Tannahill, D. et al. (2007). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc. Natl. Acad. Sci. USA* **104**, 840-845.
Porayette, P. and Paulson, R. F. (2008). BMP4/Smad5 dependent stress erythropoiesis is required for the expansion of erythroid progenitors during fetal development. *Dev. Biol.* (in press).
Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* **24**, 2354-2366.
Roeder, I. and Glauche, I. (2006). Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. *J. Theor. Biol.* **241**, 852-865.
Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massague, J. and Pavletich, N. P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* **94**, 585-594.
Siatecka, M., Xue, L. and Bieker, J. J. (2007). Sumoylation of EKLF promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Mol. Cell. Biol.* **27**, 8547-8560.
Southwood, C. M., Downs, K. M. and Bieker, J. J. (1996). Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev. Dyn.* **206**, 248-259.
Tsai, F.-Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W. and Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221.
Vyas, P., McDevitt, M. A., Cantor, A. B., Katz, S. G., Fujiwara, Y. and Orkin, S. H. (1999). Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development* **126**, 2799-2811.
Weiss, M. J., Keller, G. and Orkin, S. H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8**, 1184-1197.
Weiss, M. J., Yu, C. and Orkin, S. H. (1997). Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **17**, 1642-1651.
Welch, J. J., Watts, J. A., Vakoc, C. R., Yao, Y., Wang, H., Hardison, R. C., Blobel, G. A., Chodosh, L. A. and Weiss, M. J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136-3147.
Willey, S., Ayuso-Sacido, A., Zhang, H., Fraser, S. T., Sahr, K. E., Adlam, M. J., Kyba, M., Daley, G. Q., Keller, G. and Baron, M. H. (2006). Acceleration of mesoderm development and expansion of hematopoietic progenitors in differentiating ES cells by the mouse Mix-like homeodomain transcription factor. *Blood* **107**, 3122-3130.
Winnier, G., Blessing, M., Labosky, P. A. and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105-2116.
Xue, L., Chen, X., Chang, Y. and Bieker, J. J. (2004). Regulatory elements of the EKLF gene that direct erythroid cell-specific expression during mammalian development. *Blood* **103**, 4078-4083.
Yagi, K., Furuhashi, M., Aoki, H., Goto, D., Kuwano, H., Sugamura, K., Miyazono, K. and Kato, M. (2002). c-myc is a downstream target of the Smad pathway. *J. Biol. Chem.* **277**, 854-861.
Zafonte, B. T., Liu, S., Lynch-Kattman, M., Torregroza, I., Benvenuto, L., Kennedy, M., Keller, G. and Evans, T. (2007). Smad1 expands the hemangioblast population within a limited developmental window. *Blood* **109**, 516-523.
Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B. and Kern, S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol. Cell* **1**, 611-617.
Zhang, J., Socolovsky, M., Gross, A. W. and Lodish, H. F. (2003). Role of Ras signaling in erythroid differentiation of mouse fetal liver cells: functional analysis by a flow cytometry-based novel culture system. *Blood* **102**, 3938-3946.