© 2014. Published by The Company of Biologists Ltd | Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

RESEARCH ARTICLE

STEM CELLS AND REGENERATION

A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification

Jane Gilmour[1], Salam A. Assi[1,2], Ulrike Jaegle[3], Divine Kulu[3], Harmen van de Werken[3], Deborah Clarke[4], David R. Westhead[2], Sjaak Philipsen[3,*] and Constanze Bonifer[1,*]

ABSTRACT

Mammalian development is regulated by the interplay of tissue-specific and ubiquitously expressed transcription factors, such as Sp1. *Sp1* knockout mice die *in utero* with multiple phenotypic aberrations, but the underlying molecular mechanism of this differentiation failure has been elusive. Here, we have used conditional knockout mice as well as the differentiation of mouse ES cells as a model with which to address this issue. To this end, we examined differentiation potential, global gene expression patterns and Sp1 target regions in Sp1 wild-type and Sp1-deficient cells representing different stages of hematopoiesis. $Sp1^{-/-}$ cells progress through most embryonic stages of blood cell development but cannot complete terminal differentiation. This failure to fully differentiate is not seen when Sp1 is knocked out at later developmental stages. For most Sp1 target and non-target genes, gene expression is unaffected by Sp1 inactivation. However, Cdx genes and multiple Hox genes are stage-specific targets of Sp1 and are downregulated at an early stage. As a consequence, expression of genes involved in hematopoietic specification is progressively deregulated. Our work demonstrates that the early absence of active Sp1 sets a cascade in motion that culminates in a failure of terminal hematopoietic differentiation and emphasizes the role of ubiquitously expressed transcription factors for tissue-specific gene regulation. In addition, our global side-by-side analysis of the response of the transcriptional network to perturbation sheds a new light on the regulatory hierarchy of hematopoietic specification.

KEY WORDS: Sp1 transcription factor, Hematopoiesis, Transcriptional network, Mouse

INTRODUCTION

Changes in gene expression programs during cell differentiation are regulated by the interplay of ubiquitous and tissue-specific transcription factors, and the epigenetic regulatory machinery. A large number of tissue-specific transcription factors have been described in whose absence a specific cell lineage is not, or is less efficiently, formed. In the hematopoietic system, this includes factors such as GATA1 and PU.1, which are crucial regulators of erythropoiesis or myelopoiesis, respectively (Orkin and Zon, 2008). The effect of removing tissue-specific regulators for a specific pathway leads to dramatic alterations in gene expression and often a block in differentiation at the point where this factor is crucially needed. The reason for this could be that such factors nucleate shifts in transcription factor assemblies to alter cell fates (Heinz et al., 2010; Lichtinger et al., 2012). However, many widely expressed genes are regulated by factors that are also more or less ubiquitously expressed and that cooperate with tissue-specific factors to activate specific gene expression programs. An interesting observation is that knockout of such genes can have surprisingly tissue-specific effects. This is true, for example, for the individual knockouts of NF1 family members (Gronostajski, 2000). A reason for this phenomenon could be compensation by other family members. However, the knockout of factors such as Sp1 or Oct1 arrests development at an early developmental stage (Marin et al., 1997; Sebastiano et al., 2010). We currently know very little of the molecular mechanisms underlying the impact of ubiquitously expressed transcription factors on tissue-specific gene expression. The Sp transcription factor family is a branch of the Krüppel family of zinc-finger proteins (Philipsen and Suske, 1999) with Sp1 as the founding family member. Sp1 is ubiquitously expressed and binds to a GC-rich consensus sequence that is found at most housekeeping genes regulated by CpG island promoters, but also in many tissue-specific genes (Wierstra, 2008). The effect of rendering Sp1 inactive in mice is dramatic. Sp1-deficient embryos die very early in development with variable phenotypes that range from developmental arrest as an amorphous cell mass to retarded embryos with a number of recognizable structures (Marin et al., 1997). Such a pleiotropic phenotype that affects multiple cell types is seen in a number of knockout mice carrying mutations in global regulators, but to actually investigate the molecular mechanisms that underlie how such factors are involved in the regulation of multiple genes in specific tissues is an extremely difficult task. The binding of ubiquitously expressed factors may be to a large extent invariant and the effect of crippling their activity on development may be dictated by tissue-specific factors or by changes in their binding during development. Moreover, the impact of a lack of factor activity on gene expression control during development may be progressive and may only accumulate over many different developmental stages, making it difficult to pinpoint the actual cause of the developmental defect. To date, neither of these scenarios has been investigated.

Here, we address these issues by differentiating mouse embryonic stem cells that carry a DNA binding-deficient Sp1 variant into hematopoietic cells. We purified differentiating cells at successive stages of hematopoietic development and measured global gene expression profiles, as well as differentiation capacity. We also determined the direct target genes of Sp1 using ChIP sequencing. Our data show that $Sp1^{-/-}$ cells are capable of progressing through all early embryonic stages of blood cell development up to the progenitor stage, but are then unable to progress further. This failure

[1]School of Cancer Sciences, Institute of Biomedical Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK. [2]Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK. [3]Department of Cell Biology, Erasmus MC, Rotterdam 3015 CN, The Netherlands. [4]Section of Experimental Haematology, Leeds Institute of Molecular Medicine, University of Leeds, Leeds LS9 7TS, UK.

*Authors for correspondence (j.philipsen@erasmusmc.nl; c.bonifer@bham.ac.uk)

Received 15 November 2013; Accepted 16 April 2014

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

of terminal differentiation is not seen when Sp1 is knocked out at later developmental stages. We demonstrate that the underlying mechanism of this inability to complete differentiation is a progressive deregulation of gene expression over multiple cell generations, with multiple developmental pathways involved in hematopoietic stem cell specification and myeloid differentiation being affected. All four Hox gene clusters, as well as their upstream regulators, the Cdx genes, are targets of Sp1 at an early, but not at a later, differentiation stage and the regulation of a subset of these genes is affected by Sp1 inactivation, providing a molecular explanation for the multiple developmental defects in Sp1-deficient mice.

RESULTS

The absence of Sp1 DNA binding activity affects multiple hematopoietic lineages

In the past decade, a number of attempts have been made to dissect the molecular mechanism of the developmental arrest caused by lack of Sp1 DNA-binding activity, using conditional knockout mice and CRE-recombinase enzyme expressed from different types of tissue-specific promoters. Although such experiments confirmed the severe defects in mice where Sp1 activity was removed in all tissues, other phenotypes were surprisingly mild, if at all visible (D. I. Kulu, PhD Thesis, Erasmus University, Rotterdam, The Netherlands, 2013). This indicates that the timing of the knockout is of essence and that cells have to undergo a number of differentiation stages for it to be visible. Remarkably, ES cells carrying two copies of the mutant Sp1 allele expressing a truncated protein lacking the entire DNA-binding domain ($Sp1^{-/-}$; Fig. 1A, supplementary material Fig. S1A) were indistinguishable from wild-type cells in culture, and contributed efficiently to chimeras until E9.5, after which contribution of the mutant cells rapidly declined to undetectable levels (Marin et al., 1997). This opened the possibility of using the differentiation of such cells *in vitro* to obtain molecular insights into the molecular mechanisms of differentiation perturbed by the lack of Sp1 activity. We first tested whether $Sp1^{-/-}$ ES cells were capable of differentiating into hematopoietic cells by performing macrophage release assays where ES cells are cultured in the presence of interleukin 3 and colony-stimulating factor 1 (Csf1). This drives the formation of embryoid bodies (EBs) containing macroscopically

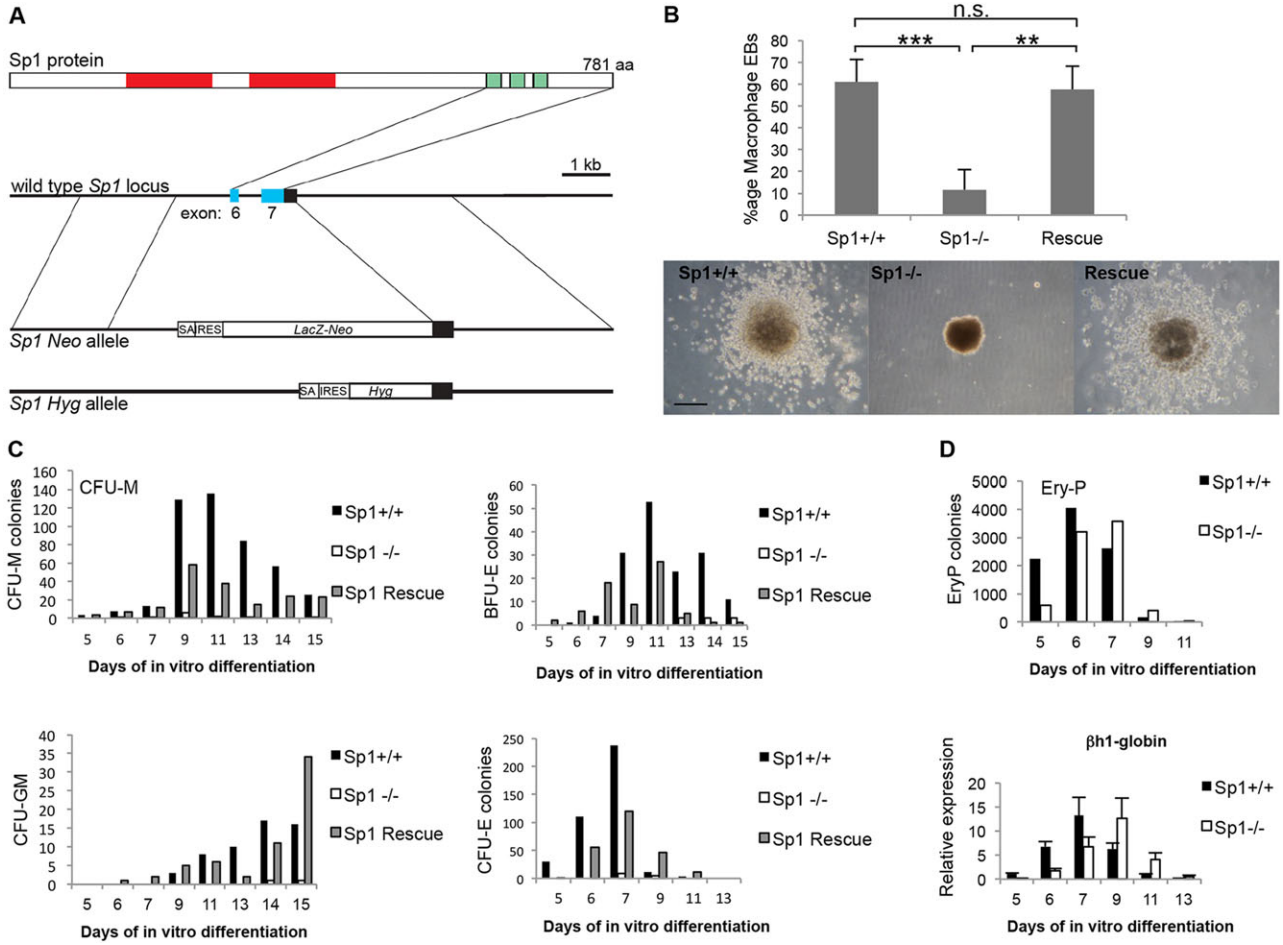


Fig. 1. Absence of Sp1 binding affects the developmental potential of multiple hematopoietic lineages. (A) The Sp1 deletion mutant. (B) Macrophage release assay. Embryoid bodies were allowed to form in methylcellulose under macrophage-promoting conditions. $Sp1^{-/-}$ cells show a reduced capacity to form macrophage-releasing EBs, whereas this capacity is restored in Sp1-rescue cells. n=3, error bars represent s.d., $^{**}P$<0.01, $^{***}P$<0.005. (C) Colony assays demonstrating that the re-expression of Sp1 rescues hematopoietic development. Embryoid bodies in methylcellulose were dispersed at different time-points and re-plated in methylcellulose under hematopoietic colony-forming conditions. $Sp1^{-/-}$ cells show reduced colony forming capacity in all lineages but especially to CFU-M and CFU-GM. A representative graph out of three independent experiments is shown for each colony type. (D) Top: Ery-P colony assay. Embryoid bodies in methylcellulose were dispersed at different time-points and re-plated in methylcellulose supplemented with erythropoietin. Graph shows combined data from two independent experiments performed in duplicate. Bottom: gene expression analysis showing expression of embryonic βh1-globin during EB differentiation, error bars indicate s.e.m. (n=3).

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

visible blood islands that release macrophages at later stages of differentiation (Faust et al., 1994; Clarke et al., 2000). *Sp1*$^{-/-}$ cells had a greatly reduced ability to form blood islands and macrophages in embryoid bodies compared with wild-type cells (Fig. 1B). Moreover, gene expression analysis with RNA prepared from developing EBs showed reduced levels of mRNA for genes important for myelopoiesis, such as *Spi1* (previously *Sfpi1*, *PU.1*), *Cebpa* and *Csf1r* (supplementary material Fig. S1B). Other hematopoietic lineages, such as erythroid cells, were also affected, as shown by colony assays demonstrating a near complete lack of colony-forming ability (Fig. 1C). This impediment of differentiation was not due to a proliferative defect, as shown by CFSE assays (supplementary material Fig. S1C). We used colony assays to show that mutant phenotypes were a direct result of Sp1 deficiency and not clonal variation of ES cells. Expression of Sp1 cDNA in the same *Sp1*$^{-/-}$ clone rescued both macrophage development and colony-forming ability (Fig. 1B,C). However, primitive erythropoiesis producing nucleated erythrocytes occurred at wild-type levels (Fig. 1D and supplementary material Fig. S1D). In addition, embryonic globin was expressed, but was up- and downregulated with delayed kinetics (Fig. 1D and supplementary material Fig. S1D), indicating that this developmental pathway was largely independent of Sp1.

Hematopoietic development in Sp1-deficient cells is progressively impaired

In both ES cells and in the whole organism, hematopoietic cells originate from mesodermal cells that form a precursor with hematopoietic, cardiac and endothelial potential: the hemangioblast (Fehling et al., 2003). This precursor cell type differentiates into specialized adherent endothelial cells forming a hemogenic endothelium (HE), which gives rise to the actual hematopoietic progenitor cells via a two-step process called the endothelial-hematopoietic transition (Chen et al., 2009; Lancrin et al., 2009)

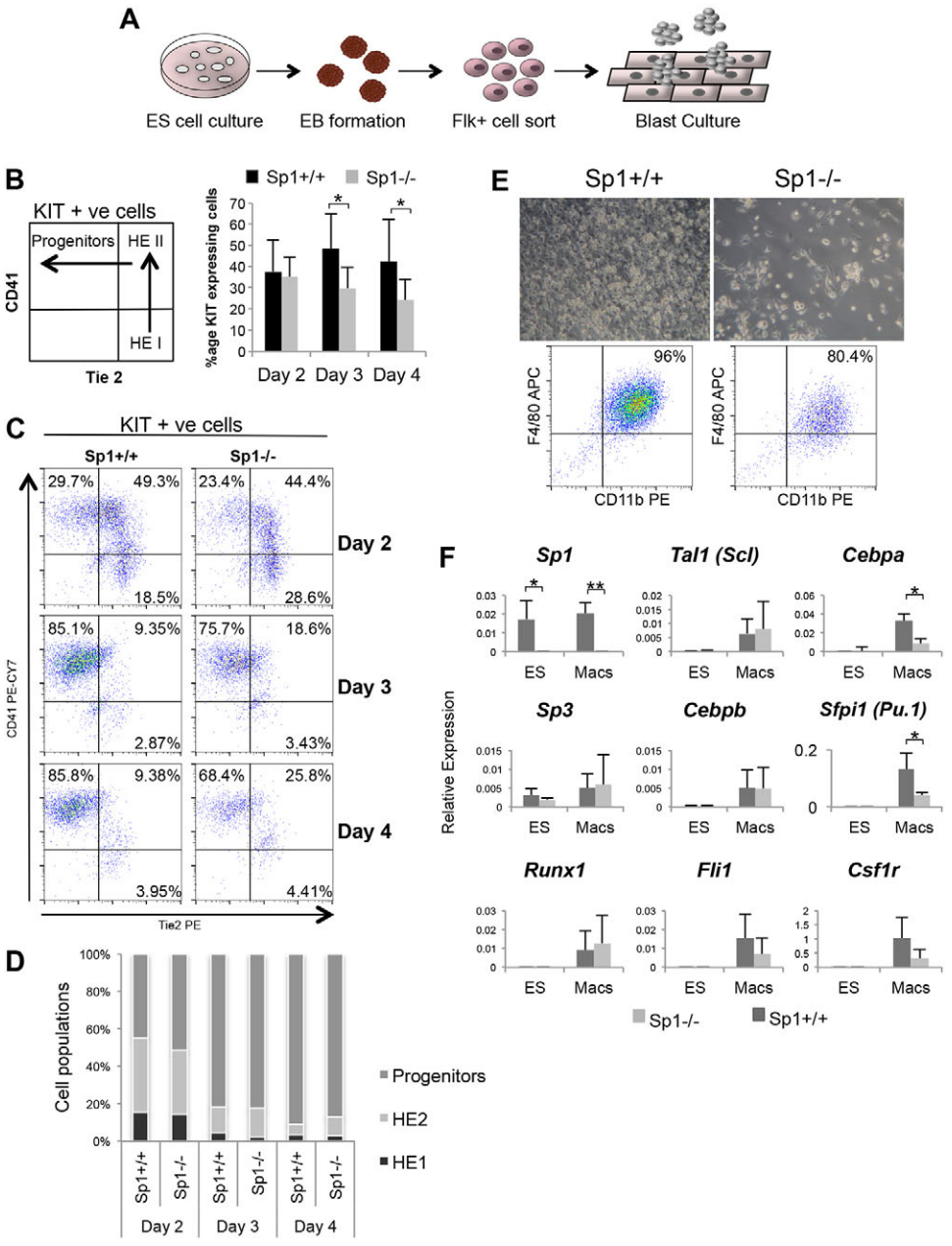


Fig. 2. Abolition of Sp1 DNA binding has a progressively deleterious effect on hematopoietic development. (A) The hematopoietic differentiation assay used to generate cell populations for microarrays and ChIP-seq. (B) Left: graph depicting FACS analysis of cell populations throughout blast culture differentiation. Cells were stained with antibodies to KIT, CD41 and Tie2. Right: average proportion of KIT-expressing cells as measured by FACS analysis at day 2, 3 and 4 of blast culture (for details, see supplementary material Fig. S2D) (n=4, error bars represent s.d., *P<0.05). (C) Representative FACS analysis demonstrating that HE1, HE2 and progenitors are formed in both differentiating wild-type and *Sp1*$^{-/-}$ cells, albeit with a slightly reduced frequency. (D) Graphical depiction of C (n=4). (E) Differentiation of progenitors to macrophages is dramatically reduced in Sp1 KO cells. Progenitors were seeded under macrophage-promoting conditions. Macrophage markers F4/80 and CD11b were assayed by FACS analysis after 7+ days of differentiation. (F) Gene expression analysis measuring the expression of the indicated hematopoietic regulator genes using RNA prepared from wild-type and *Sp1*$^{-/-}$ ES and macrophage cells (n=4, error bars represent s.d., *P<0.05, **P<0.01.).

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

(Fig. 2A). Morphological inspection of $Sp1^{+/+}$ and $Sp1^{-/-}$ cultures demonstrated they were undistinguishable (supplementary material Fig. S2A) and contained similar numbers of cells (supplementary material Fig. S2B). We also employed a blast colony assay (Kennedy et al., 1997) to show that the development of the earliest stage of hematopoietic development, the hemangioblast, was not disturbed and similar numbers of colonies were formed (supplementary material Fig. S2C). To examine in detail how and at which developmental stage the lack of Sp1 affected hematopoietic development, we measured levels of the stem cell factor receptor KIT, the endothelial marker TIE2 and the integrin CD41 on the cell surface during differentiation by flow cytometry (Fig. 2B-D, supplementary material Fig. S2D). The latter, together with CD45 is a marker of definitive hematopoietic progenitor cells (Mikkola et al., 2003). These experiments show that the proportion of KIT-positive cells and KIT expression levels were progressively reduced during differentiation of $Sp1^{-/-}$ cells. Interestingly, the proportion of CD45-positive cells was not altered, indicating that HE and progenitor cells were still formed albeit at slightly reduced numbers (supplementary material Fig. S2E). $Sp1^{-/-}$ progenitors were impaired in colony formation (supplementary material Fig. S2F) and in macrophage development, with only a few adherent cells expressing macrophage surface markers being formed (Fig. 2E). However, the phenotype of these cells was aberrant as the expression of important myeloid regulators in these cells was strongly reduced (Fig. 2F).

We next tested whether we could recapitulate this defect in myeloid maturation by removing Sp1 activity at an early myeloid progenitor stage. To this end, we generated mice carrying two floxed Sp1 alleles on an $Sp3^{+/-}$ heterozygous background. These mice were crossed with lysozyme-Cre mice, which excise the conditional allele at the common myeloid progenitor stage when lysozyme expression is upregulated (Tagoh et al., 2002). We prepared bone marrow cells from these mice and tested their ability to form colonies and generate macrophages. Supplementary material Fig. S2G shows that, although myeloid colony-forming activity in such progenitors was slightly reduced, they were still capable of efficiently producing macrophages even with the additional reduction of Sp3 levels (supplementary material Fig. S2H) and in spite of high levels of CRE activity (supplementary material Fig. S2I). Taken together, these experiments show: (1) that $Sp1^{-/-}$ cells are able to progress through multiple differentiation steps but are unable to complete terminal differentiation from myeloid progenitors; and (2) that the effect of Sp1 deficiency is cumulative rather than progenitor type specific.

Gene expression in differentiating Sp1-deficient cells becomes progressively deregulated

We next examined how Sp1 deficiency deregulated gene expression. To this end, we used the blast culture system to produce pure populations of differentiating cells by cell sorting as previously described (Lancrin et al., 2009; Lichtinger et al., 2012). We isolated $Flk1^{+}$ cells containing hemangioblasts, plated those cells into blast culture and purified two successive stages of hemogenic endothelium development before and after the EHT as well as floating progenitor cells (supplementary material Fig. S3A) using cell sorting as outlined in Fig. 2B. We then prepared RNA from these cells and measured global mRNA expression by microarray analysis (supplementary material Tables S1 and S2). Several hundreds of genes change their expression at each differentiation stage in wild-type cells and this was also true for $Sp1^{-/-}$ cells (Fig. 3A) whereas the majority of genes in wild-type and $Sp1^{-/-}$ cells in the different cell populations showed highly similar gene expression profiles (supplementary material Fig. S3B). This was confirmed by Pearson correlation analysis of biological replicates (supplementary material Fig. S3C), indicating a high level of correlation of gene expression profiles between $Sp1^{-/-}$ and wild-type cells at the start of

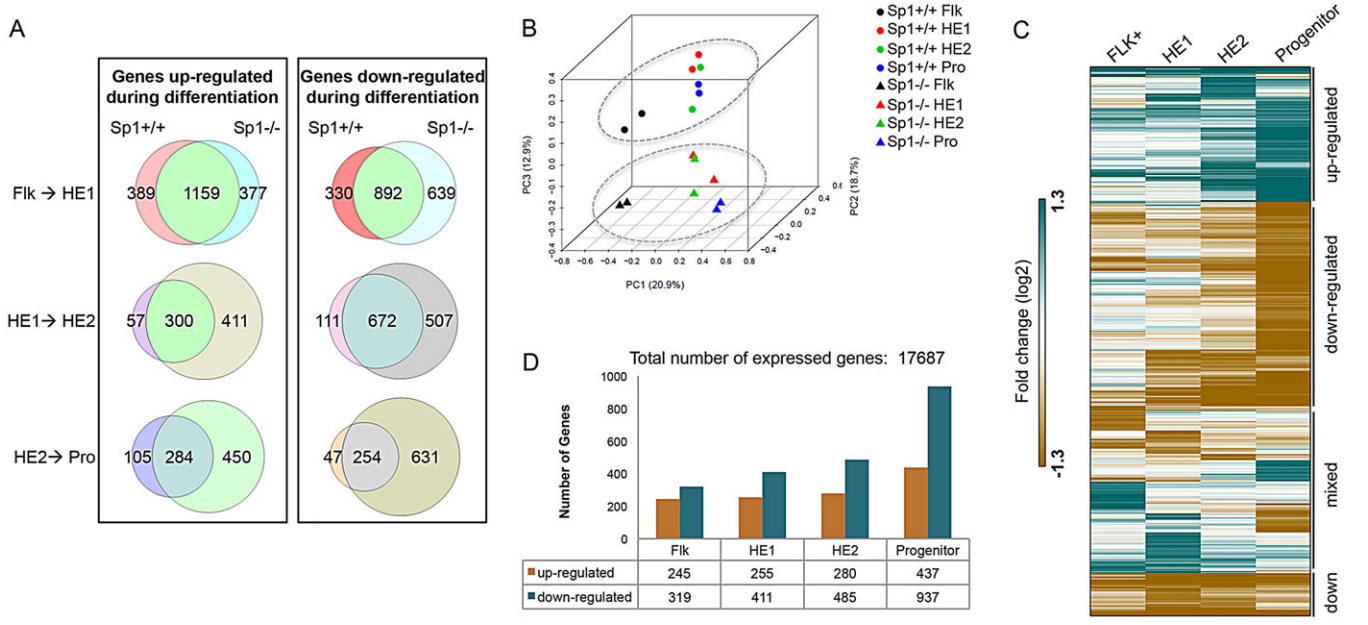


Fig. 3. Gene expression becomes progressively deregulated during hematopoiesis. (A) Venn diagrams showing the overlap in genes either up- or downregulated during transition from one cell population to the next during differentiation in $Sp1^{-/-}$ cells relative to wild type. (B) Principal component analysis using the three largest principal components. Although the first two components separate the cell populations, i.e. $Flk1^{+}$, HE1, HE2 and progenitors, the third component separates the wild-type and $Sp1^{-/-}$ samples, and is indicated by a gray ellipse. (C) Hierarchical clustering of gene expression fold changes for genes that are differentially expressed in $Sp1^{-/-}$ cells compared with wild type. (D) Graph indicating the numbers of up- and downregulated genes, as defined in C in $Sp1^{-/-}$ cell populations compared with the same wild-type cell population, indicating that the highest number of deregulated genes is found in the progenitor cell population.

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

blast culture in Flk1^{+} cells. However, gene expression patterns progressively deviated during differentiation (compare yellow frames, supplementary material Fig. S3C). This deviation occurred on a background of highly similar gene expression in each cell type, as demonstrated by principal component analysis (Fig. 3B) and analysis of fold-change in gene expression during differentiation (Fig. 3C). Differentiation-directed gene expression changed for most genes as in wild-type cells, indicating little variation between ES cell clones (see also supplementary material Fig. S3C), but a subset of genes was progressively deregulated and wild-type and *Sp1*$^{-/-}$ cells could be clearly distinguished. In total, 2220 genes changed expression more than twofold (up or down) as a result of Sp1 deficiency, with 564 deregulated genes in Flk1^{+} cells and 1374 deregulated genes in progenitors (Fig. 3D).

Sp1 deficiency affects multiple developmental pathways in a cell type-specific fashion

In order to identify which developmental pathways were deregulated in *Sp1*$^{-/-}$ differentiating cells, we clustered gene expression changes according to patterns of expression throughout differentiation (Fig. 4A, supplementary material Fig. S4A,B and Table S3). This analysis yielded 23 clusters and identified a large number of genes whose expression was unchanged, but also many genes whose expression pattern during differentiation was altered at specific time points of differentiation. Manual validation of expression changes for selected genes by real-time PCR is shown in supplementary material Fig. S4C. In all cell types, most genes were downregulated. Deregulation of gene expression by Sp1 deficiency affected multiple differentiation pathways and gene sets, some of which were common between cell types (Cluster 1, 2), but others were unique for each cell type, as demonstrated by clustering of their GO-terms according to *P* values (Fig. 4B and supplementary material Table S5). Fig. 4C lists selected genes deregulated in *Sp1*$^{-/-}$ cells, many of which are known developmental regulators. At early stages of differentiation, expression of genes encoding the caudal related transcription factors *Cdx1* and *Cdx2*, which are known to regulate early hematopoiesis (Wang et al., 2008; Lengerke and Daley, 2012), was reduced. In addition, multiple Hox genes were downregulated. In progenitor cells, Hox gene regulation was also affected but in addition we observed a strong downregulation of multiple genes known to impact on early hematopoietic stem cell specification, including genes in the BMP and Wnt pathways such as *Bmp4*, *Lef1* and *Wnt9a* (Lengerke et al., 2008; Nostro et al., 2008), and also genes that regulate stem cell function (*Myb* and *Fli1*) and myelopoiesis (*Cebpe*) (Verbeek et al., 2001; Mucenski et al., 1991; Spyropoulos et al., 2000). In addition, the expression of *Kit* was downregulated in *Sp1*$^{-/-}$ progenitor cells, explaining the reduction of KIT protein on their surface. An interesting finding was that upregulated genes in progenitor cells contained many genes involved in heme biosynthesis and erythroid biology (including *Klf1*, *Gata1* and adult globin genes), in spite of the fact that these cells are unable to form definitive erythroid colonies.

Sp1 binds to different targets during hematopoietic specification

We next performed ChIP-sequencing in wild-type Flk1^{+} cells and progenitor cells to test which of the deregulated genes were direct

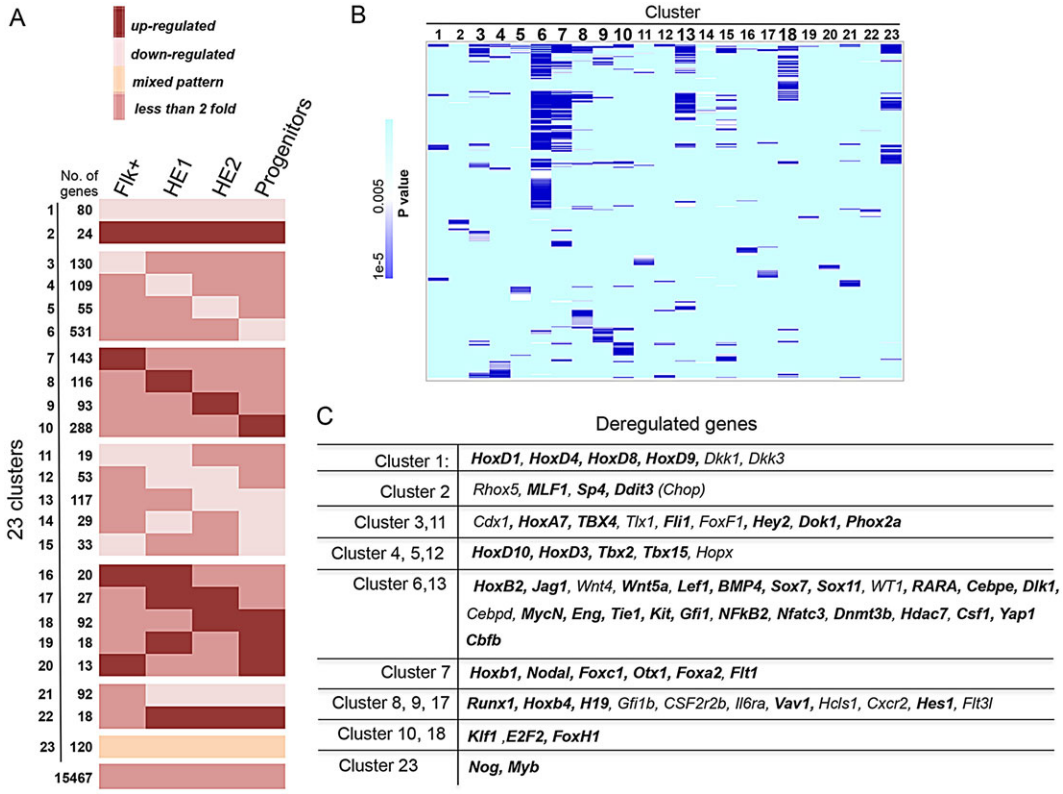
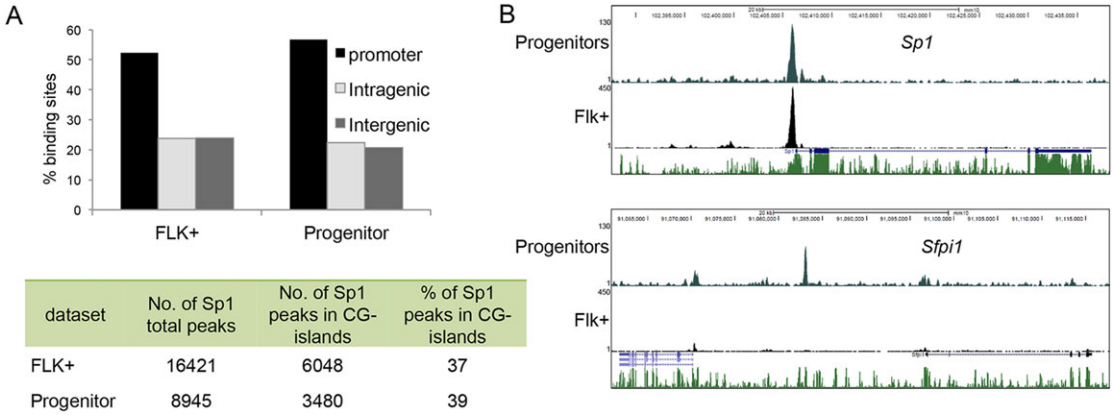


Fig. 4. Analysis of gene expression changes throughout differentiation. (A) Clustering of genes more than twofold up- or downregulated by Sp1 inactivation during differentiation. Up- and downregulated genes could be separated into 23 clusters according to changes in expression levels during differentiation, as defined in supplementary material Fig. S4A,B. The progenitor population shows the highest number of deregulated genes. (B) Clustering of GO terms after gene ontology analysis of the up- and downregulated genes in *Sp1*$^{-/-}$ cells relating to the clustering depicted in A, showing that the different clusters identified are enriched for different patterns of gene ontology terms. (C) List of known developmental regulator genes within the different clusters that are deregulated during the differentiation of *Sp1*$^{-/-}$ cells. Direct target genes of Sp1 at any developmental stage are indicated in bold.

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

targets of Sp1 and to examine whether Sp1 shifted position during differentiation and targeted different genes. Sp1 bound to a large number of sites (supplementary material Fig. S5A), most of which were promoters (Fig. 5A). As predicted from *in vitro* experiments, more than one-third of Sp1-binding sites were located in CG islands (Fig. 5A, bottom panel). The comparison of binding sites between Flk1$^+$ cells and progenitors demonstrated that Flk1$^+$ cells contained twice as many binding sites as progenitor cells and the expression level of Sp1 in the two cell types differed slightly (supplementary material Fig. S5B). There was a large overlap between binding sites (supplementary material Fig. S5C), and most of these sites (58%) coincided with CG islands, indicating that Sp1 is indeed involved in the regulation of ubiquitously expressed genes. By contrast, the Flk1$^+$ and the progenitor-specific peaks contained only 6% and 18% CG islands, respectively. However, there were also many cell type-specific sites, even when only peaks with high tag count were compared (supplementary material Fig. S5D). Such differential binding is exemplified in Fig. 5B, which shows a screenshot of the *Sp1* locus itself that was bound in both Flk1$^+$ cells (upper panel) and progenitors compared with the *Spi1* (*PU.1*) locus, which was only bound by Sp1 in progenitors (lower panel). Other examples were the four Hox gene clusters (A-D), which contained a large number of binding sites across each cluster and were specifically bound in Flk1$^+$ cells (supplementary material Fig. S5F), indicating that Sp1 cooperates with different factors at tissue-specific genes. This notion was confirmed by analyzing enriched motifs surrounding Sp1 peaks (Fig. 5C, supplementary material Fig. S5E). At promoters, Sp1 colocalized with binding motifs of widely expressed factors such as NFY and CREB in both Flk1$^+$ cells and progenitors. However, at distal elements it colocalized with different types of motifs, reflecting the activity of tissue-specific distal elements, with GATA motifs being prominent in Flk1$^+$ cells, and highlighting the important role of GATA2 at this stage (Lugus et al., 2007). Sp1 also colocalized with RUNX/ETS/GATA motifs, reflecting the importance of these factors in driving hematopoietic development (Wilson et al., 2010). Approximately half of the promoter sites contained Sp1 consensus motifs, but only a small fraction of the distal sites, indicating the possibility that such peaks originate from the association of Sp1 with upstream factors (Fig. 5D).

The integration of gene expression and ChIP-seq data showed that the majority of Sp1 binding sites were found at genes that were unaffected by the absence of Sp1 activity (supplementary material Fig. S6A). However, in Flk1+ cells more than two-thirds of the deregulated genes were direct Sp1 targets (Fig. 6A, supplementary material Fig. S6A) but this was true for only one-third of the



dataset	No. of Sp1 total peaks	No. of Sp1 peaks in CG-islands	% of Sp1 peaks in CG-islands
FLK+	16421	6048	37
Progenitor	8945	3480	39

C

Motifs associated with Sp1 in Flk+ve cells

Promoter peaks			Distal peaks		
Motif	Match	Log P value	Motif	Match	Log P value
	Sp-family	-1542.7		GATA	-1593.5
	NFY	-1406.37		GATA	-1585.4
	NFY	-1111.5		ETS	-634.4
	ETS	-1002.9		Sp - family	-634.4
	CREB	-444.1			

Motifs associated with Sp1 in Progenitor cells

Promoter peaks			Distal peaks		
Motif	Match	Log P value	Motif	Match	Log P value
	NFY	-1106.9		ETS	-1593.5
	Sp-family	-1017.8		GATA	-372.2
	ETS	-912.8		RUNX	-270.5
	CREB	-341.4		AP1	-135.3
				Sp-family	-78.6

D

Percentage motifs in peaks

Promoter	Distal	Promoter	Distal	
46	13	45	10	Sp1
39	21	41	16	NFY
45	48	49	64	ETS
12	48	13	40	GATA
13	19	14	31	RUNX
6	13	7	19	AP1
FLK+		Progenitor		

Fig. 5. ChIP-seq analysis of Sp1 binding in wild-type ES, Flk1$^+$ and progenitor cells. (A) Graph showing peak annotations, indicating that the majority of peaks in Flk1$^+$ and progenitor cells are in promoter regions. Table shows the percentage of Sp1-binding sites that are within CpG islands for each data set. (B) Screenshots showing examples of Sp1 binding at target loci. Shown are the *Sp1* locus, where Sp1 binds the promoter in both cell populations, and the *Spi1* (*PU.1*) locus where Sp1 binding increases as differentiation progresses. (C) Sp1 colocalizes with motifs for ubiquitously expressed factors at promoters and with motifs for tissue-specific factors at distal elements. Analysis of enriched binding motifs associated with regions of Sp1 binding in each of the cell populations in promoter and distal sites. This shows that motifs associated with the promoter remain relatively consistent, whereas motifs associated with distal binding sites change as differentiation progresses. (D) Table showing percentage of promoter and distal peaks that have the indicated motifs in Flk1$^+$ and progenitor populations ±200 bp from peak center.

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

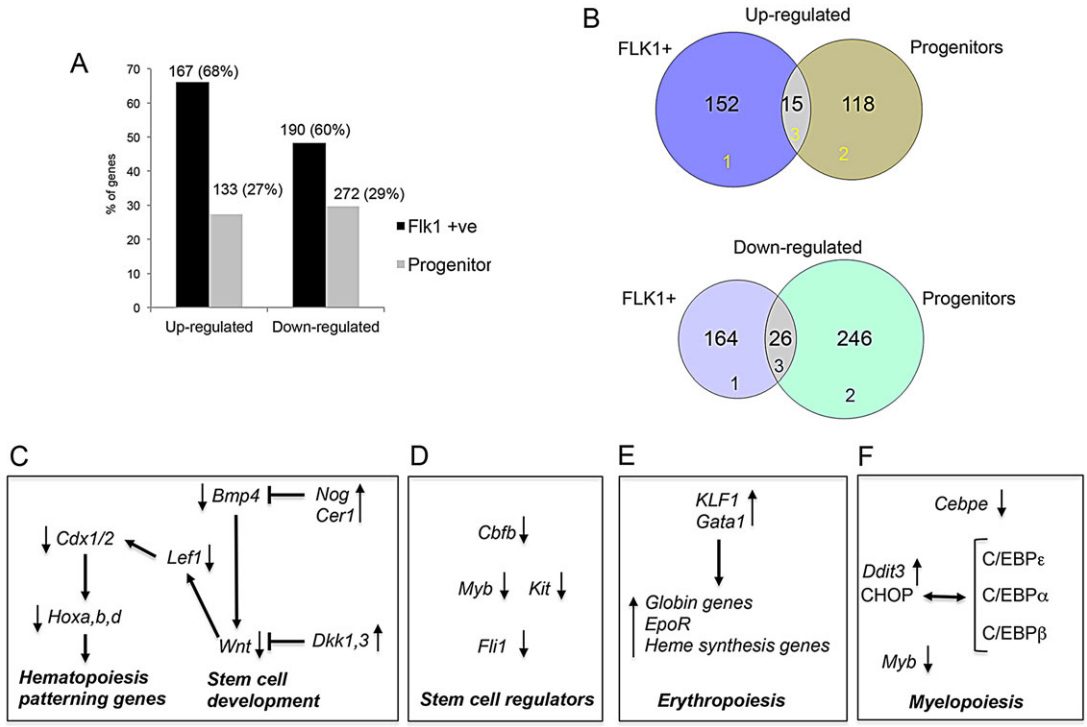


Fig. 6. Integration of ChIP-seq and microarray data. (A) Graph showing the percentage of genes that are up- or downregulated in *Sp1*$^{-/-}$ cells in Flk1^{+} and progenitor cells and that are Sp1 target genes. A hyper-geometric distribution was used to calculate the significance of enrichment of direct Sp1 target genes in Flk1+ and progenitor cells. Enrichment of both up- and downregulated Sp1 target genes was found to be significant, with *P*-values of 4.7e–133 and 3.6e–208 in Flk1^{+} cells and 4.3e–18 and 9.3e–46 in progenitor cells, respectively. (B) Venn diagrams showing the overlap between Sp1 target genes from the ChIP-Seq versus differentially expressed genes in Flk1+ and progenitor cells (upregulated, upper panel; downregulated, lower panel). (C) Schematic outline of the interaction between genes and factors during hematopoietic specification from mesoderm depicting the downregulation of the BMP/Wnt pathway after Sp1 inactivation. (D) Downregulation of stem cell regulators after Sp1 inactivation. (E) Upregulation of regulators of erythropoiesis and globin expression and heme biosynthesis after Sp1 inactivation. (F) Downregulation of myeloid regulators and upregulation of CHOP, which inhibits C/EBP activity after Sp1 inactivation.

deregulated genes in progenitor cells. There was little overlap between the two gene sets, again showing that Sp1 deficiency affected different gene sets depending on the cellular context (Fig. 6B, see also supplementary material Fig. S6B). A summary of all affected pathways in the different populations is depicted in supplementary material Fig. S7A-E. We found that deregulated Sp1 target genes were involved in multiple pathways, including metabolism, focal adhesion and signaling processes. Furthermore, genes involved in early hematopoietic specification and stem cell development pathways were also affected, and are highlighted in more detail in Fig. 6 and supplementary material Fig. S7.

Our analysis uncovered a clear hierarchy of progressive deregulation events impacting on hematopoiesis. The four Hox clusters A-D and *Cdx1* were major Sp1 targets in Flk1+ cells (supplementary material Table S4), but not in progenitor cells (Fig. 4C, supplementary material Fig. S5F), indicating that Sp1 activity is directly required for correct activity of early regulators of hematopoietic specification at this specific developmental stage (Fig. 6C). However, Sp1 did not only act as an activator, but also as repressor as many targets were upregulated in *Sp1*$^{-/-}$ cells (supplementary material Table S4). Two of the most interesting of these genes were *Nog* and *Cer1*, which encode inhibitors of BMP signaling (Smith et al., 1993) with the latter being upregulated 11-fold (supplementary material Table S1), reinforcing the idea that the BMP4/WNT signaling pathway is impaired in the absence of Sp1 DNA-binding activity. This pathway was also affected in progenitor cells (Fig. 4, cluster 6, 13), with a further reduction of *Bmp4*, *Lef1* and *Cdx1/3* expression (supplementary material Table S4).

In progenitors, downregulated Sp1 targets also contained a large number of genes known to be crucial for correct stem cell formation and myeloid differentiation (Fig. 4, supplementary material Table S4). This included transcription factor genes *Myb*, *Fli1*, *Cebpe* and *Cbfb*. *Cbfb* encodes the binding partner (CBFβ) for the master regulator of stem cell development RUNX1 (North et al., 1999) (Fig. 6D, supplementary material Table S4). Again, also in this cell type, Sp1 appeared to be able to act as a repressor, as evidenced by the upregulation of a number of genes, including the WNT inhibitors *Dkk1* and *Dkk3*, as well as *Klf1* and *Ddit3* (Fig. 6E,F). *Klf1*, which encodes the transcription factor KLF1 (EKLF), is an important activator of erythroid genes and controls multiple aspects of terminal erythropoiesis (Tallack et al., 2010; Pilon et al., 2011). In addition to KLF1, another important erythroid regulator, GATA1, was also upregulated in *Sp1*$^{-/-}$ progenitors, which together may explain the upregulation of genes involved in heme biosynthesis as well as fetal and adult globin genes. *Ddit3* was upregulated more than fourfold (supplementary material Table S2) and encodes the C/EBP family member CHOP, which has been shown to repress the transcriptional activity of other C/EBP proteins (Ron and Habener, 1992). C/EBP family members, in particular C/EBP α and ε are essential for myelopoiesis beyond the GMP stage (Friedman, 2007), which may provide one explanation why this differentiation pathway is blocked later than the multipotent progenitor stage.

Collectively, we conclude that the initial downregulation of the Sp1 target loci Cdx and Hox is followed by progressively deregulated expression of genes involved in the control of hematopoiesis over

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

multiple differentiation stages. These early deregulation events set a cascade in motion that culminates in a catastrophic failure of terminal differentiation.

DISCUSSION

In this study, we perturbed hematopoietic development by inactivating a single, ubiquitously expressed transcription factor (Sp1) and analyzed the effects of this perturbation on the dynamics of differential gene expression and cell differentiation. In addition to dissecting a so far unknown role of Sp1 in blood cell development, our study uncovered a number of novel aspects of the hierarchical control of blood cell development and homeostatic mechanisms maintaining cellular function.

Sp1 binds to developmental regulator genes in a cell-type specific fashion

Fifty percent of Sp1-binding sites reside in promoters, one-third of which are CG islands, many of which are shared between the two cell types analyzed. Sp1 differentially occupied distal cis-regulatory elements in Flk1^{+} cells and progenitors, and colocalized with different sets of motifs for cell-type specific transcription factors. This cell-type specific occupancy profile was particularly apparent in Flk1^{+} cells with the Hox genes being a prominent example. Many such genes are expressed throughout differentiation, suggesting that at the progenitor stage they are regulated by a set of factors that do not depend on Sp1. The difference in the occupancy profile is reflected in the sensitivity of target genes to Sp1 depletion. The vast majority of genes was not influenced by the absence of Sp1 during cell differentiation, indicating that robust mechanisms exist that compensate for the absence of this protein, even in the face of a changing cis-regulatory environment. Compensating factors are likely to be other Sp-family members such as Sp3 as indicated by the finding that mice heterozygous for either *Sp1* or *Sp3* are viable, but compound heterozygous are not (Kruger et al., 2007). However, the absence of Sp1 strongly affected the expression of distinct subsets of cell type-specific genes, demonstrating that, for such genes, Sp1 cooperates with cell type-specific factors and is a crucial component of stable transcription factor complexes regulating such genes.

The deregulation of gene expression patterns during ES cell-derived hematopoietic differentiation starts with a small number of genes in Flk1^{+} cells, most of which are direct Sp1 targets and this number increases as differentiation progresses, although the proportion of direct targets goes down. Remarkably, although many genes are deregulated in differentiating *Sp1*$^{-/-}$ cells and their number increases at later stages of differentiation, many others are not and, up to a point, cell identity is preserved as measured by the ability of these cells to give rise to specific progeny and expressing the majority of genes correctly. Albeit with reduced efficiency, Flk1^{+} cells still differentiate into hemogenic endothelium cells, which are able to undergo the endothelial-hematopoietic transition and thereafter express first CD41 and then CD45, indicating that powerful feedback mechanisms must exist that preserve cellular function. So far the nature of these mechanisms is unknown, but they are likely to involve post-transcriptional regulation mechanisms by, for example, miRNAs, which could buffer the effect of shifts in the transcriptional network on the proteome. Our data clearly show that the deregulation of gene expression after Sp1 inactivation is cumulative and progressive, providing a molecular explanation for: (1) the heterogeneity of the Sp1 knockout phenotype as individual animals will differentially accumulate defects, depending on how well their transcriptional network can compensate; and (2) the lack of effect of knocking out Sp1 at later stages of development.

Multiple pathways driving hematopoietic development are impaired in Sp1-deficient cells

Our perturbation study also provides profound insights into the hierarchical control of multipotent hematopoietic progenitor cell specification and defines several interlinked pathways involved in the control of this process in an unbiased fashion. This discussion will not be able to cover all aspects of deregulation, but will highlight a few important pathways. The first steps towards hematopoiesis involve the interplay of BMP and WNT signaling, and the activation of Cdx and Hox genes in cells of mesodermal origin (Lengerke et al., 2008; Nostro et al., 2008; Lengerke and Daley, 2012). Our data show that Sp1 is crucially involved in activating these pathways after the hemangioblast stage (Fig. 6C), as both *Cdx1* and some of its downstream Hox target genes are downregulated in the absence of Sp1. Interestingly, the BMP and WNT inhibitors Noggin/Cerberus and Dickkopf (*Nog*, *Cer1*, *Dkk1*, *Dkk3*) are upregulated, suggesting that they may be part of a negative-feedback loop within this pathway that enforces hematopoietic specification. Although all Hox gene clusters are Sp1 targets, not all of the individual genes responded equally to Sp1 knockout. Of the genes in the HoxA cluster, only *HoxA7* was downregulated more than twofold in Flk1+ cells, which raises the possibility that at this stage this gene may contribute to the defects in hematopoiesis and also to the hematopoietic defects observed in mice lacking the entire *HoxA* cluster (Di-Poi et al., 2010). Interestingly, the *HoxB* cluster responded differentially, with *HoxB4-HoxB8* being downregulated and *HoxB1* being upregulated in Flk1^{+} cells, consistent with the role of these genes in blood cell development (Bijl et al., 2006) and also confirming that HOXB3 is a repressor of *HoxB1* (Wong et al., 2011). The *HoxC* cluster was mostly unaffected by Sp1 knockout. The most dramatic effect of Sp1 inactivation on expression was seen with the *HoxD* cluster, where basically all genes except *HoxD13* were strongly downregulated at all developmental stages. So far, a role of HoxD proteins in hematopoiesis has not been described (Alharbi et al., 2013), although a leukemogenic version of *HoxD13* was discovered (Pineault et al., 2003). However, Abd-related *Hox* proteins are able to form heterodimers with another protein, MEIS1 (Shen et al., 1997), which also plays an important role in hematopoiesis (Hisa et al., 2004) and cooperates with Hox genes in leukemogenesis (Argiropoulos et al., 2007). *Meis1* was 1.8-fold downregulated after Sp1 knockdown in Flk1^{+} and progenitor cells (supplementary material Table S1), raising the possibility of a combined negative effect on hematopoietic development.

The downregulation of genes involved in hematopoietic specification has a significant knock-on effect on gene expression at the following differentiation stages and affects a number of well-known hematopoietic regulators. Levels of *Myb*, a gene that is absolutely essential for definitive hematopoiesis (Mucenski et al., 1991) and whose expression levels have to be strictly controlled (Taketani et al., 2002), are reduced throughout differentiation. This is also true for levels of *Fli1*, which is essential for fetal liver hematopoiesis (Spyropoulos et al., 2000) and is required for early priming events in blood cell development (Lichtinger et al., 2012). *Runx1* is prematurely upregulated at the first stage of the hemogenic endothelium (CD41^{-} cells), together with RUNX1-responsive targets *Gfi1* and *Spi1* (supplementary material Table S1) (Lancrin et al., 2012; Lichtinger et al., 2012) which is one of the first signs that

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

hematopoietic differentiation is going astray. A reason for this could be the fact that *HoxA3*, which is specifically expressed at this stage and has been shown to be required for the repression of *Runx1* in the hemogenic endothelium (Iacovino et al., 2011) is 1.6-fold downregulated (supplementary material Table S1). At the next stage, after the endothelial-hematopoietic transition (CD41+ cells), the expression of the erythroid regulators *Klf1* and *Gata1*, as well as expression of embryonic globin genes, increases almost fourfold, whereas expression of genes specifying the myeloid lineage, such as *Spi1* and *Cebpe*, is reduced (although normally expression would increase) (Lichtinger et al., 2012) (supplementary material Table S1). This indicates that developing cells pass through the endothelial-hematopoietic transition with an inherent bias for erythropoiesis and that the timing of RUNX1 expression is of the essence for correct myelopoiesis. This bias is further reinforced by the upregulation of *Ddit3* and is even more pronounced at the progenitor stage, explaining the inability of progenitors to form macrophages.

Our data demonstrate that primitive erythropoiesis is not affected by the lack of Sp1, but is delayed. In addition, the kinetics of globin regulation is altered as embryonic and adult globin genes and the erythropoietin receptor are upregulated at the same time, together with genes coding for multiple components of the heme biosynthetic pathway. However, definitive cells expressing adult globin genes are not functional erythroblasts, as they are unable to form erythroid colonies. This is consistent with the notion that, together with Sp3, Sp1 is required for the regulation of fetal liver erythropoiesis (Kruger et al., 2007).

In summary, our approach of following hematopoietic differentiation of $Sp1^{-/-}$ ES cells combined with the integration of global RNA expression and ChIP data turned out to be highly useful for dissecting a so far unexplained knockout phenotype. Our strategy may serve as an example on how to delineate molecular explanations for knockout phenotypes of other global regulators that also may be the result of cumulative deregulation events. So far we have focused on pathways known to be involved in hematopoietic specification to build the hierarchies outlined in the discussion and to validate our approach, thereby also discovering novel regulatory interactions. However, it should be noted that additional pathways are affected by Sp1 deficiency (supplementary material Fig. S7). This suggests that our approach could also be used to predict novel regulators of other developmental pathways whose individual role can then be tested experimentally.

MATERIALS AND METHODS

ES cell maintenance

$Sp1^{+/+}$, $Sp1^{-/-}$ and Sp1-rescue ES cells (Marin et al., 1997) were maintained on primary mouse embryonic fibroblasts in ES maintenance media: DMEM [high glucose from powder (Sigma D5648)], supplemented with 15% FCS, 1 mM sodium pyruvate, 100 units/ml penicillin and 100 µg/ml streptomycin, 1 mM glutamine, 0.15 mM MTG, 25 mM HEPES buffer, 10^3 U/ml ESGRO (Millipore) and 1× non-essential amino acids (Sigma). Prior to differentiation, the ES cells were grown without feeder cells for two passages on gelatinized tissue culture-treated plates.

ES cell differentiation in the blast culture system

In vitro differentiation of ES cells was performed essentially as described previously (Lichtinger et al., 2012). A more-detailed description can be found in supplementary Materials and methods.

Macrophage differentiation and colony assays

Macrophage release assays

ES cells were trypsinized and allowed to form embryoid bodies by plating in base methylcellulose (Stem Cell Technologies M3134) supplemented with 10% FCS, 100 units/ml Penicillin and 100 µg/ml Streptomycin, 1 mM glutamine, 0.15 mM MTG, 10 µg/ml insulin (Sigma), 10% M-CSF conditioned media, 5% IL-3 conditioned media, 10 ng/ml recombinant mouse M-CSF (R&D Systems), 100 units/ml IL-1 (Peprotech) at a cell concentration previously determined to give similar numbers of EB. After at least 14 days EB were counted and assessed for the number of EB which were surrounded by a halo of macrophages.

Macrophage differentiation from progenitors

Macrophages were generated by plating floating progenitors onto low adherence plates at $0.5\text{-}1\times10^6$ cells per 3 cm dish in IMDM supplemented with 10% FCS, 100 units/ml Penicillin and 100 µg/ml Streptomycin, 1 mM glutamine, 0.15 mM MTG, 10% M-CSF conditioned media, 5% IL3 conditioned media and 10 ng/ml murine M-CSF (R&D Systems). Macrophages were usually harvested after 7 days by first gently washing the cells with PBS to remove any non-adherent cells and then trypsinizing the adherent layer. Cell purity was checked by FACS for expression of F4/80 (eBioscience 17-4801) and CD11b (eBioscience 12-0112). RNA was prepared by lysing the macrophages in TRIzol.

Methylcellulose assays

Methylcellulose assays and RNA preparation from embryoid bodies were performed essentially as described previously (Clarke et al., 2000). For a complete description see supplementary Materials and methods.

Chromatin immunoprecipitation

Chromatin immunoprecipitation and library preparation were performed essentially as described previously (Lichtinger et al., 2012). For a complete description, see supplementary Materials and methods.

mRNA expression analysis

RNA was extracted from cells using TRIzol (Invitrogen) according to manufacturer's instruction. First-strand cDNA synthesis was carried out using Superscript II (Invitrogen) according to manufacturer's instructions using 250-500 ng of RNA. Real-time PCR was carried out using ABI SYBR green master mix with 5 µl of diluted cDNA and 0.25 µM primer per 15 µl reaction on an ABI 7900HT machine. Analysis was carried out on samples measured in duplicate. See supplementary material Table S6 for primer sequences. For microarray analysis of sorted populations, a further clean up step was performed using RNeasy Minelute columns (Qiagen).

Microarrays

Each source sample RNA (25 ng) was labeled with Cy3 dye as per the protocol detailed in the Low Input Quick Amp Labeling Kit (Agilent Technologies – 5190-2305). A specific activity of greater than 6.0 was confirmed by measurement of 260 nm and 550 nm wavelengths with a NanoDrop ND-1000 Spectrophotometer. Labeled RNA (600 ng) was hybridized for 16 h to Agilent SurePrint G3 Mouse 8×60 K microarray slides. After hybridization, slides were washed as per protocol and scanned with a High Resolution C Scanner (Agilent Technologies), using a scan resolution of 3 µm. Feature extraction was performed using Agilent Feature Extraction Software, with no background subtraction.

Analysis of mice

Bone marrow samples from wild-type Sp1cko/cko::Sp3cko/wt mice and from LysM-cre::Sp1cko/cko::Sp3cko/wt::YFP/wt mice were subjected to clean up using a MACS dead cell removal kit (Miltenyi Biotec). For CFU-C assays, cells were plated at 2×10^4 cells/ml in Methocult M3434 in duplicate 3 cm dishes and colonies were scored after 10 days. For CFU-Macrophage (CFU-M) assays, cells were plated at 2×10^4 cells/ml in M3134 methylcellulose (Stem Cell Technologies), supplemented with 10% FCS, 1% BSA, 100 units/ml penicillin, 100 µg/ml streptomycin, 1 mM glutamine, 0.2 mg/ml human transferrin, 10^{-4} M β2-Me, 10 µg/ml insulin (Sigma), 50 ng/ml M-CSF in duplicate 3 cm dishes and macrophage colonies scored after 8 days. Cells were also cultured for expansion of hematopoietic progenitor cells in liquid culture and subsequent differentiation to macrophages. BM cells were cultured in

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

progenitor expansion media [IMDM supplemented with 10% horse serum (Gibco), 100 units/ml penicillin and 100 µg/ml streptomycin, 1 mM glutamine, 20 ng/ml SCF, 10 ng/ml Flt3 ligand and 25 ng/ml TPO] at 1×10^6 cells/ml. After 2 weeks in culture, progenitors were seeded in macrophage differentiation media as described above.

Data analysis

Comparison of *Sp1*$^{+/+}$, *Sp1*$^{-/-}$ and Sp1-rescue samples was performed using Student's *t*-test. A detailed description of data analyses and more detailed methods can be found in supplementary materials. Data can be accessed at Gene Expression Omnibus (GEO) under the accession number GSE52499.

Acknowledgements

We thank Mengchu Wu at the Sequencing Facility at the Cancer Sciences Institute of Singapore (CSI), National University of Singapore, for expert service.

Competing interests

The authors declare no competing financial interests.

Author contributions

J.G., U.J., D.C. and D.K. performed experiments, S.A.A., H.v.d.W. and D.R.W. analyzed the data, S.P. and C.B. conceived the study and together with J.G. designed the experiments and wrote the manuscript.

Funding

This research was funded by grants from the Medical Research Council [G0901579] and the Biotechnology and Biological Sciences Research Council [BB/H008217/1] to C.B. and D.R.W., as well as the Landsteiner Foundation for Blood Transfusion Research [LSBR 1040], the Netherlands Scientific Organization [NWO DN 82-301, ZonMw 40-00812-98-08032 and 40-00812-98-12128] and the European Union fp7-health-2012 collaborative project THALAMOSS (306201) to S.P. Deposited in PMC for immediate release.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.106054/-/DC1

References

Alharbi, R. A., Pettengell, R., Pandha, H. S. and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000-1008.

Argiropoulos, B., Yung, E. and Humphries, R. K. (2007). Unraveling the crucial roles of Meis1 in leukemogenesis and normal hematopoiesis. *Genes Dev.* **21**, 2845-2849.

Bijl, J., Thompson, A., Ramirez-Solis, R., Krosl, J., Grier, D. G., Lawrence, H. J. and Sauvageau, G. (2006). Analysis of HSC activity and compensatory Hox gene expression profile in Hoxb cluster mutant fetal liver cells. *Blood* **108**, 116-122.

Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887-891.

Clarke, D., Vegiopoulos, A., Crawford, A., Mucenski, M., Bonifer, C. and Frampton, J. (2000). In vitro differentiation of c-myb(-/-) ES cells reveals that the colony forming capacity of unilineage macrophage precursors and myeloid progenitor commitment are c-Myb independent. *Oncogene* **19**, 3343-3351.

Di-Poï, N., Koch, U., Radtke, F. and Duboule, D. (2010). Additive and global functions of HoxA cluster genes in mesoderm derivatives. *Dev. Biol.* **341**, 488-498.

Faust, N., Bonifer, C., Wiles, M. V. and Sippel, A. E. (1994). An in vitro differentiation system for the examination of transgene activation in mouse macrophages. *DNA Cell Biol.* **13**, 901-907.

Fehling, H. J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G. and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130**, 4217-4227.

Friedman, A. D. (2007). Transcriptional control of granulocyte and monocyte development. *Oncogene* **26**, 6816-6828.

Gronostajski, R. M. (2000). Roles of the NFI/CTF gene family in transcription and development. *Gene* **249**, 31-45.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589.

Hisa, T., Spence, S. E., Rachel, R. A., Fujita, M., Nakamura, T., Ward, J. M., Devor-Henneman, D. E., Saiki, Y., Kutsuna, H., Tessarollo, L. et al. (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J.* **23**, 450-459.

Iacovino, M., Chong, D., Szatmari, I., Hartweck, L., Rux, D., Caprioli, A., Cleaver, O. and Kyba, M. (2011). HoxA3 is an apical regulator of haemogenic endothelium. *Nat. Cell Biol.* **13**, 72-78.

Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N. and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. *Nature* **386**, 488-493.

Krüger, I., Vollmer, M., Simmons, D. G., Elsässer, H.-P., Philipsen, S. and Suske, G. (2007). Sp1/Sp3 compound heterozygous mice are not viable: impaired erythropoiesis and severe placental defects. *Dev. Dyn.* **236**, 2235-2244.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V. and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892-895.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, O., Wilson, N. K., Moroy, T., Bonifer, C. et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314-322.

Lengerke, C. and Daley, G. Q. (2012). Caudal genes in blood development and leukemia. *Ann. N. Y. Acad. Sci.* **1266**, 47-54.

Lengerke, C., Schmitt, S., Bowman, T. V., Jang, I. H., Maouche-Chretien, L., McKinney-Freeman, S., Davidson, A. J., Hammerschmidt, M., Rentzsch, F., Green, J. B. A. et al. (2008). BMP and Wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. *Cell Stem Cell* **2**, 72-82.

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S. A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M. S., Wu, M. et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* **31**, 4318-4333.

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.

Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997). Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89**, 619-628.

Mikkola, H. K. A., Fujiwara, Y., Schlaeger, T. M., Traver, D. and Orkin, S. H. (2003). Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood* **101**, 508-516.

Mucenski, M. L., McLain, K., Kier, A. B., Swerdlow, S. H., Schreiner, C. M., Miller, T. A., Pietryga, D. W., Scott, W. J., Jr and Potter, S. S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell* **65**, 677-689.

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-Padilla, M. and Speck, N. A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* **126**, 2563-2575.

Nostro, M. C., Cheng, X., Keller, G. M. and Gadue, P. (2008). Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. *Cell Stem Cell* **2**, 60-71.

Orkin, S. H. and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644.

Philipsen, S. and Suske, G. (1999). A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res.* **27**, 2991-3000.

Pilon, A. M., Ajay, S. S., Kumar, S. A., Steiner, L. A., Cherukuri, P. F., Wincovitch, S., Anderson, S. M., Mullikin, J. C., Gallagher, P. G., Hardison, R. C. et al. (2011). Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* **118**, e139-e148.

Pineault, N., Buske, C., Feuring-Buske, M., Abramovich, C., Rosten, P., Hogge, D. E., Aplan, P. D. and Humphries, R. K. (2003). Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. *Blood* **101**, 4529-4538.

Ron, D. and Habener, J. F. (1992). CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. *Genes Dev.* **6**, 439-453.

Sebastiano, V., Dalvai, M., Gentile, L., Schubart, K., Sutter, J., Wu, G.-M., Tapia, N., Esch, D., Ju, J.-Y., Hubner, K. et al. (2010). Oct1 regulates trophoblast development during early mouse embryogenesis. *Development* **137**, 3551-3560.

Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M. and Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol. Cell. Biol.* **17**, 6448-6458.

Smith, W. C., Knecht, A. K., Wu, M. and Harland, R. M. (1993). Secreted noggin protein mimics the Spemann organizer in dorsalizing Xenopus mesoderm. *Nature* **361**, 547-549.

Spyropoulos, D. D., Pharr, P. N., Lavenburg, K. R., Jackers, P., Papas, T. S., Ogawa, M. and Watson, D. K. (2000). Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor. *Mol. Cell. Biol.* **20**, 5643-5652.

Tagoh, H., Himes, R., Clarke, D., Leenen, P. J. M., Riggs, A. D., Hume, D. and Bonifer, C. (2002). Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes Dev.* **16**, 1721-1737.

Development (2014) 141, 2391-2401 doi:10.1242/dev.106054

Taketani, T., Taki, T., Shibuya, N., Ito, E., Kitazawa, J., Terui, K. and Hayashi, Y. (2002). The HOXD11 gene is fused to the NUP98 gene in acute myeloid leukemia with t(2;11) (q31;p15). *Cancer Res.* **62**, 33-37.

Tallack, M. R., Whitington, T., Yuen, W. S., Wainwright, E. N., Keys, J. R., Gardiner, B. B., Nourbakhsh, E., Cloonan, N., Grimmond, S. M., Bailey, T. L. et al. (2010). A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res.* **20**, 1052-1063.

Verbeek, W., Wächter, M., Lekstrom-Himes, J. and Koeffler, H. P. (2001). C/EBPepsilon -/- mice: increased rate of myeloid proliferation and apoptosis. *Leukemia* **15**, 103-111.

Wang, Y., Yabuuchi, A., McKinney-Freeman, S., Ducharme, D. M. K., Ray, M. K., Chawengsaksophak, K., Archer, T. K. and Daley, G. Q. (2008). Cdx gene deficiency compromises embryonic hematopoiesis in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7756-7761.

Wierstra, I. (2008). Sp1: emerging roles–beyond constitutive activation of TATA-less housekeeping genes. *Biochem. Biophy. Res. Commun.* **372**, 1-13.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E. et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532-544.

Wong, E. Y. M., Wang, X. A., Mak, S. S., Sae-Pang, J. J., Ling, K. W., Fritzsch, B. and Sham, M. H. (2011). Hoxb3 negatively regulates Hoxb1 expression in mouse hindbrain patterning. *Dev. Biol.* **352**, 382-392.

SUPPLEMENTARY MATERIAL

Supplementary Methods

ES cell differentiation

ES cells were trypsinised and transferred as a single cell suspension to 15 cm low adherence bacteriological plates (Sterilin) at a concentration of 2.5 X 10^4 cells/ml in IVD media. IVD media - IMDM supplemented with 15% FCS, 100 units/ml Penicillin and 100 µg/ml Streptomycin, 1 mM glutamine, 0.15 mM MTG, 0.18 mg/ml Human transferrin (Roche 652202) and 50 µg/ml Ascorbic acid. After 3.25-3.75 days embryoid bodies were collected, briefly digested in 1x trypsin/EDTA and gently dissociated. To obtain a single cell suspension cells were passed through a cell strainer and resuspended in IMDM + 20% FCS. Flk1+ve (CD309) cells were isolated using a biotinylated Flk1 antibody (eBioscience 13-5821) used at 5 µl per 10^7 cells for 15 minutes on ice, followed by 2x wash with MACS buffer (PBS + 0.5% BSA and 2 mM EDTA). Cells bound by the antibody were isolated using MACS anti-biotin beads and MACS LS columns (Miltenyi Biotec) according to the manufacturer's instructions. Isolated Flk1+ve cells were plated in blast media at a concentration of 9-11X10^3 cells per cm^2 on gelatinized tissue culture treated dishes. Blast media – IMDM supplemented with 10% FCS, 100 units/ml Penicillin and 100 µg/ml Streptomycin, 1 mM glutamine, 0.45 mM MTG, 0.18 mg/ml Human transferrin, 25 µg/ml Ascorbic acid, 20% D4T conditioned media, 5 µg/L mVEGF (Peprotech), 10 µg/L mIL-6 (Peprotech). At Day 2, 3 and 4 surface marker expression was checked by FACS analysis using antibodies to KIT (BD Pharmingen 553356), CD41 (e-bioscience 25-0411) and Tie2 (e-bioscience 12-5987). Cells were also stained with KIT (as above or ebioscience 11-1171-85) in combination with CD34 (553733), CD45 (ebioscience 17-0451-82) or Flk1 (ebioscience 12-5821-83). Sorted populations were prepared at Day 2 of blast culture. Floating and adherent cells were harvested and combined before staining with KIT, CD41 and Tie2 antibodies in MACS buffer. After washing, cells were separated on a Moflow cell sorter into HE1 (KIT+ve, CD41-ve, Tie2+ve), HE2 (KIT+ve, CD41+ve, Tie2+ve) and progenitor (KIT+ve, CD41+ve, Tie2-ve) cell populations according to surface marker expression.

Colony assays:

ES cells were trypsinised and allowed to form embryoid bodies by plating in base

methylcellulose (Stem Cell Technologies M3134) supplemented with 10% FCS, 100 units/ml Penicillin and 100 μg/ml Streptomycin, 1 mM glutamine, 4.5 x 10^{-4}M mM MTG, 10 μg/ml insulin (Sigma), 5% IL-3 conditioned media, 10% M-CSF conditioned media, 100 units/ml IL-1 (Peprotech) at a cell concentration previously determined to give similar numbers of EB. EB were harvested at a range of time-points for CFU-C hematopoietic colony assays using Methocult® M3434 complete methylcellulose (Stem Cell Technologies) or for RNA isolation using TRIzol®. Assays were set up by washing out the methylcellulose with PBS, dispersing the EB with collagenase and plating cells at 10^5/ml in M3434 methylcellulose in duplicate 3cm bacteriological grade dishes. Dishes were scored from Day 3 for CFU-E and from Day 8 for BFU-E, CFU-M and CFU-GM.

Ery-P assays were based on those performed by Sturgeon *et al* and Sroczynska *et al* (Sturgeon et al., 2012) (Sroczynska et al., 2009). Assays were set up using EB at a range of time-points during differentiation as described above and dispersed cells were plated into methylcellulose supplemented with 10% Platelet Derived Serum (Antech), 5% PFHM (Invitrogen), 100 units/ml Penicillin and 100 μg/ml Streptomycin, 2 mM glutamine, 0.18 mg/ml Transferrin, 50 μg/ml ascorbic acid, 4.5 x 10^{-4}M MTG and 2 U/ml Erythropoietin (R&D Systems) at 10^5 cells/ml in duplicate 3 cm bacteriological grade dishes. Ery-P colonies were counted on Day 5 after plating. Staining of nucleated erythroblasts: Ery-P colonies were washed off the dishes with PBS and dispersed. Single cells were cyto-spun onto glass slides, fixed with methanol and stained with Accustain Wright-Giemsa stain (Sigma).

Blast colony assays were performed by seeding 10,000 Flk1+ve sorted cells in base methylcellulose supplemented with 10% FCS, 100 units/ml Penicillin and 100 μg/ml Streptomycin, 25% D4T conditioned medium, 0.18 mg/ml Transferrin, 25 μg/ml ascorbic acid, 4.5 x 10^{-4}M MTG, 5ng/ml VEGF and 5ng/ml IL-6 in duplicate 3cm bacteriological grade dishes. Colonies were counted after 8 days in culture (Kennedy et al., 1997).

Chromatin immunoprecipitation:

Flk1+ve sorted cells and KIT+ve sorted floating progenitors from Day 3 blast culture were used for ChIP-seq analysis. Cells were crosslinked for 10 min at room temperature with 1% formaldehyde (Thermo Scientific) and quenched with 1/10th volume 2 M glycine. Nuclei were prepared essentially as described in Lefevre *et al* 2003, sonicated using a Bioruptor water bath in immunoprecipitation buffer I (25 mM Tris 1 M pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 1% TritonX-100 and 0.25 % SDS). After centrifugation the sheared 0.5-2 kb

chromatin fragments (1-2 x 10^6 cells) were diluted with 2 volumes immunoprecipitation buffer II (25 mM Tris pH 8.0, 150mM NaCl, 2 mM EDTA pH 8.0, 1% TritonX-100, 7.5% glycerol) and precipitation was carried out for 2-3 hours at 4°C using 2 µg anti-Sp1 antibody (Santa Cruz sc-17824X) coupled to 15 µl Protein-G dynabeads (Invitrogen). Beads were washed with low salt buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 1% TritonX-100, 0.1% SDS), high salt buffer (20 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% TritonX-100, 0.1% SDS), LiCl buffer (10 mM Tris pH 8.0, 250 mM lithium chloride, 1 mM EDTA pH 8.0, 0.5% NP40, 0.5% sodium-deoxycholate) and TE pH 8.0 containing 50 mM sodium chloride. The immune complexes were eluted in 100 µl elution buffer (100 mM $NaHCO_3$, 1% SDS) and after adding 4 µl 5M sodium chloride and proteinase K, the crosslinks were reversed at 65°C overnight. DNA was extracted using the Ampure PCR purification kit (Beckman Coulter) according to manufacturer's instructions and analysed by qRT-PCR. Libraries of DNA fragments from chromatin immunoprecipitation were prepared from approximately 10 ng of DNA. Firstly, overhangs were repaired by treatment of sample material with T4 DNA polymerase, T4 PNK and Klenow DNA polymerase (all enzymes obtained from New England Biolabs, UK) in a reaction also containing 50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM Dithiothreitol, 0.4 mM dNTPs and 1 mM ATP. Samples were purified after each step using Ampure PCR purification kit (Beckman Coulter). 'A' bases were added to 3' ends of fragments using Klenow Fragment (3´- 5´ exo-), allowing for subsequent ligation of adapter oligonucleotides (Illumina part #1000521) using Quick T4 DNA ligase. After a further Ampure clean up to remove excess adaptors, fragments were amplified in a PCR reaction using adapter-specific primers (5'-CAAGCAGAAGACGGCATACGAGCTCTTCCGATC*T and 5'-AATGATACGGCGACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T). The libraries were purified and adapter dimers removed by running the PCR products on 2% agarose gels and excising gel slices corresponding to fragments approximately 200-300 bp in size, which were then extracted using the Qiagen gel extraction kit. Libraries were validated using quantitative PCR for known targets, and quality assessed by running 1 µl of each sample on an Agilent Technologies 2100 Bioanalyser. Once prepared, DNA libraries were subjected to massively parallel DNA sequencing on an Illumina Genome Analyzer.

SDS-PAGE and Western Blotting

Protein extracts were separated on 10 % SDS-PAGE gels and western blots prepared by wet

transfer onto nitrocellulose membrane. Blots were blocked with 5 % milk powder in 0.05 % TBS-Tween and incubated with anti-Sp1 (Millipore 07-645), anti-Sp3 (Santa Cruz sc-644) and anti-GAPDH (abcam ab-8245) antibodies. Proteins were visualised using Pierce SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific).

CFSE Assay

Single cell suspensions of ES cells at a concentration of 2 x10^6 cells/ml were labelled with 1μM CFSE in PBS for 10 min at room temperature. The CFSE staining was quenched with media and cells washed thoroughly to remove excess CFSE. Cells were seeded on gelatinised 12 well TC plates at a concentration of 4 x10^4 cells per well in ES cell maintenance media. FACS was performed every 12hrs and unstained cells were used as a negative control.

Data Analysis

Analysis of ChIP-sequencing data

The raw sequence data in fastq format returned by the Illumina Pipeline was aligned to the mm10 mouse genome build using BWA (Li and Durbin, 2010). The reads in the resulting alignment files in sam format were used to generate density maps using bed-tools (Quinlan and Hall, 2010) and data was displayed using the UCSC Genome Browser (Kent et al., 2002).

Regions of enrichment (peaks) of ChIP data were identified using MACS (Zhang et al., 2008) and cisGenome (Ji et al., 2008) software. The resulting peaks common for the two peak calling methods were considered for further analysis. Peak overlaps, gene annotations and CpG island measurements were performed using in-house scripts. Peaks were allocated to genes if located in either their promoters or within the region of 500 bp downstream and 2000 bp upstream of the transcription start sites (TSS), as intragenic if not in the promoter but within the gene body region, or if intergenic, to the nearest gene located within 100 kb. CpG island coordinates were downloaded from the UCSC genome browser and the number of peaks in CpG islands was calculated if the peak summit lies within the CpG island start and end coordinates.

A number of tools are designed for testing for differential binding sites; here we used MAnorm a Bioconductor R package (Shao et al., 2012). We found that the 10577 peaks unique for FLK1+ cells were statistically significant differential binding sites at a cut-off of a p-value of >= 0.5 and that 8136 (77%) unique peaks were statistically significant at a cut-off of a p-value of >= 0.1. We also found that the 3099 peaks unique for progenitor were statistically significant differential binding sites at a cut-off of a p-value of >= 0.5 and that 2368 (76%) progenitor unique peaks are statistically significant at a cut-off of a p-value of >= 0.1. Moreover 87% of the FLK1+ peaks were differential binding sites by comparing the 10577 unique FLK1+ peaks to the FLK1+ total peaks generated by MACS when using progenitors as a control sample. 77% of unique FLK1+ peaks were statistically significantly differential binding sites at a MACS FDR cut-off of 7.

De novo motif analysis was performed on promoters and non-promoter (distal) peaks separately using HOMER (Heinz et al., 2010). Motif lengths of 6, 8, 10, 12 and 14 bp were identified in within ± 100 bp from the peak summit and a random background sequence option was used. The motif matrices generated by HOMER were scanned against JASPAR with the use of STAMP to identify similarity to known transcription factor binding sites (Mahony and Benos, 2007). The top enriched motifs with a significant log p value score were recorded. The annotatePeaks function in HOMER was used to find occurrences of motifs in peaks and distribution of motif density around the peak summit. In this case we used the discovered motif position weight matrices (PWM) with the most significant log p value.

Analysis of microarray data

The microarray gene expression scanned images were analysed with Feature Extraction Software 10.7.1.1 (Agilent) (protocol GE1_107_Sep09, Grid: 028005_D_F_20100614 and platform Agilent SurePrint G3 Mouse GE 8x60K). The raw data output by Feature Extraction Software was analysed using the LIMMA R package (Smyth et al., 2005)with quantile normalisation and background subtraction with the normexp method (Ritchie et al., 2007) using an offset value of 16. Contrast matrix and eBays function were used and p value <= 0.01 was applied. Only genes with a minimum log2 intensity value equal to or greater than 6 in at least one array were selected as expressed genes. Genes that changed expression at least two fold up or down with respect to $Sp1^{-/-}$ were considered as differentially expressed.

The Principal Component Analysis (PCA) was carried out on the average value of the probe set intensity within each experiment and was calculated based on a pair-wise Pearson

correlation coefficient matrix using R (R Core Team, 2013).

Clustering of gene expression was carried out on signal intensity for all expressed genes and on fold-changes for genes associated with at least a two-fold change. Hierarchical clustering was used with Euclidean distance and average linkage clustering. Heatmaps were generated using Mev from TM4 microarray software suite (Saeed et al., 2006). We then clustered/grouped gene expression fold changes according to patterns of expression throughout differentiation (Fig. 4A). This analysis yielded 23 clusters and identified a large number of genes whose expression was unchanged. Signal intensity and fold changes of each cluster/group individually were hierarchically clustered and box plotted respectively. (Figs S4A, B)

Gene ontology (GO) analysis was performed using bingo (Maere et al., 2005) and David online tool at david.abcc.ncifcrf.gov (Huang da et al., 2009). Non redundant GO terms were filtered using REVIGO online tools at (http://revigo.irb.hr) with simRel as a similarity measure and a medium allowed similarity. KEGG Pathway network analysis was performed using ClueGO tools (Bindea et al. 2009) with kappa score = 0.3.

Supplementary tables:

Table 1: List of differentially regulated genes (wt versus Sp1-/-)

Table 2: List of genes in 23 clusters and their GO terms

Table 3: List of Sp1 target genes in Flk1+ cells and progenitors and their GO terms

Table 4: List of Sp1 bound genes that change expression and their GO terms.

Table 5: Hierarchical ranking of GO terms for the selected clusters according to pValue.

Table 6: Primer List for RT-PCR

Gene	Forward Primer	Reverse Primer
Sp1	TCATATTGTGGGAAGCGCTTT	CAGGGCAGGCAAATTTCTTCT
Sp3	CGACAGTCCTGCAGATATTAGGATC	AGGTCATTGGTGTTCAGTGTAGAGTC
Sp1(human)	GCGAGAGGCCATTTATGTGT	GGCCTCCCTTCTTATTCTGG
Runx1	GCAGGCAACGATGAAAACTACTC	CAAACCTGAGGTCGTTGAATCTC

Tal1 (Scl)	CAACAACAACCGGGTGAAGA	ATTCTGCTGCCGCCATCGTT
Sfpi1(Pu.1)	CCATAGCGATCACTACTGGGATTT	TGTGAAGTGGTTCTCAGGGAAGT
Cebpα	GCAGGAGGAAGATACAGGAAGCT	ACACCTAAGTCCCTCCCCTCTAAA
Cebpβ	GTTTCGGGACTTGATGCAATC	CGCAGGAACATCTTTAAGTGAT
Fli1	TCGTGAGGACTGGTCTGTATGG	GCTGTTGTCGCACCTCAGTTAC
Csf1R	CTTTGGTCTGGGCAAAGAAGAT	CAGGGCCTCCTTCTCATCAG
Cdx1	CACAGAGCGGCAGGTAAAGA	GGCCAGCATTAGTAGGGCAT
Cdx2	AGGAGGAAAAGTGAGCTGGC	TTGGCTCTGCGGTTCTGAAA
BMP4	GAGCCAACACTGTGAGGAGT	ATACGGTGGAAGCCCTGTTC
Wnt4	GCCATCGAGGAGTGCCAATA	GCCACACCTGCTGAAGAGAT
HoxA7	AGCCAGTTTCCGCATCTACC	CCTTCTCCAGTTCCAGCGTC
HoxB4	ACGGCCTACACTCGCCAG	GCGGTTGTAGTGAAACTCCTTCTC
HoxD1	CCACAGCACTTTCGAGTGGA	TGGTGCTGAAATTTGTGCGG
Lef1	ACGACAAGGCCAGAGAACAC	CATGTACGGGTCGCTGTTCA
βh1-globin	GGGTTAAGAACATGGACAACCTC	GGGTGAATTCCTTGGCAAAATGA
Hba-X	ATCATCATGTCCATGTGGGAGA	GGAAGTAGGTCTTCGTCTGGG
Hbb-Y	TGAACTGCACTGTGACAAGC	TGCCGAAGTGACTAGCCAAA

Supplementary References:

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-89.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.

Ji, H., Jiang, H., Ma, W., Johnson, D. S., Myers, R. M. and Wong, W. H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**, 1293-300.

Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N. and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. *Nature* **386**, 488-93.

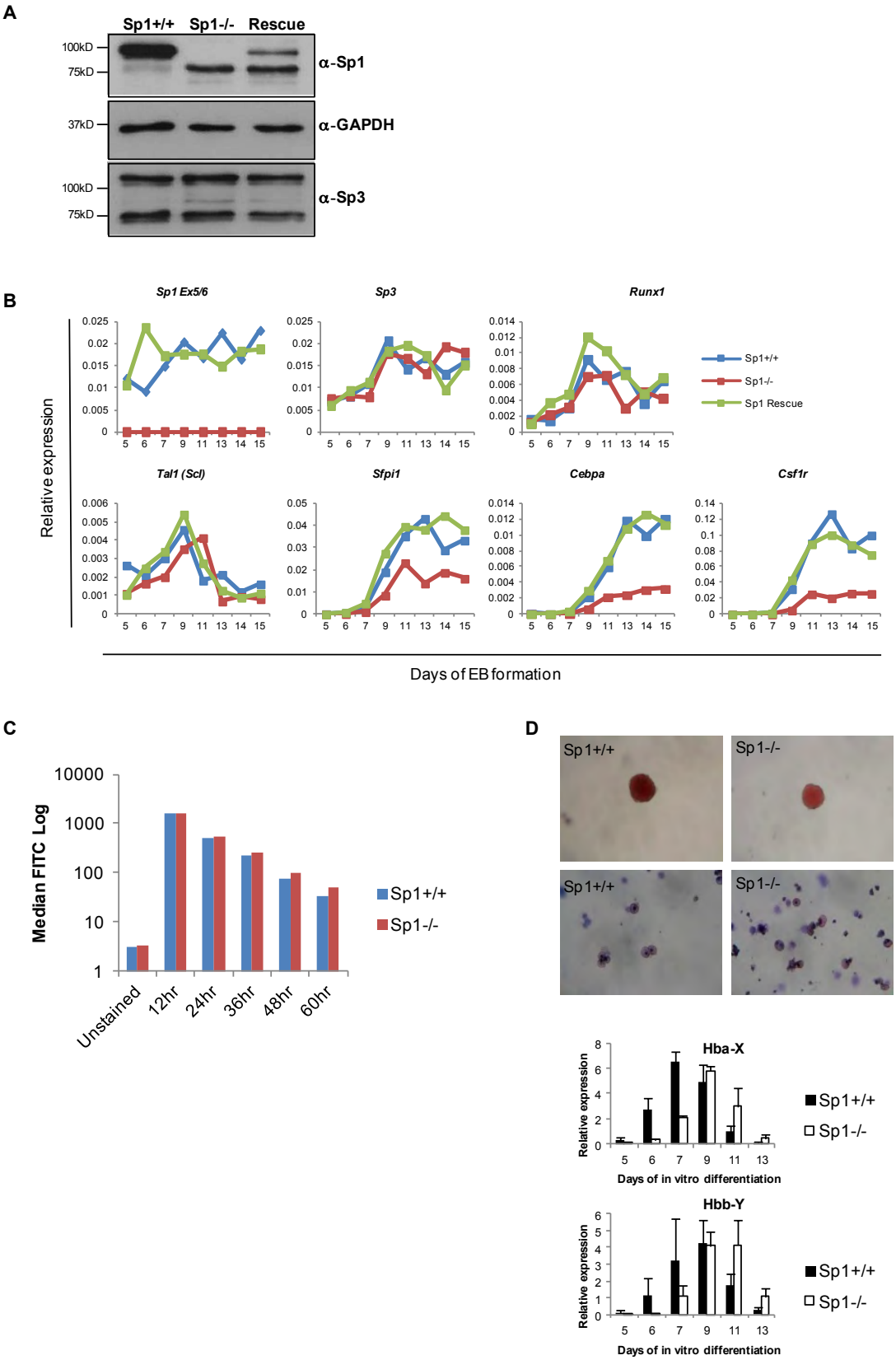
Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002). The human genome browser at UCSC. *Genome research* **12**, 996-1006.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95.

Maere, S., Heymans, K. and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-9.
Mahony, S. and Benos, P. V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic acids research* **35**, W253-8.
Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2.
Ritchie, M. E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. and Smyth, G. K. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23**, 2700-7.
Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., Li, J., Thiagarajan, M., White, J. A. and Quackenbush, J. (2006). TM4 microarray software suite. *Methods Enzymol* **411**, 134-93.
Shao, Z., Zhang, Y., Yuan, G. C., Orkin, S. H. and Waxman, D. J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* **13**, R16.
Smyth, G. K., Michaud, J. and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-75.
Sroczynska, P., Lancrin, C., Pearson, S., Kouskoff, V. and Lacaud, G. (2009). In vitro differentiation of mouse embryonic stem cells as a model of early hematopoietic development. *Methods Mol Biol* **538**, 317-34.
Sturgeon, C. M., Chicha, L., Ditadi, A., Zhou, Q., McGrath, K. E., Palis, J., Hammond, S. M., Wang, S., Olson, E. N. and Keller, G. (2012). Primitive erythropoiesis is regulated by miR-126 via nonhematopoietic Vcam-1+ cells. *Dev Cell* **23**, 45-57.
Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.

Supplementary Figures:

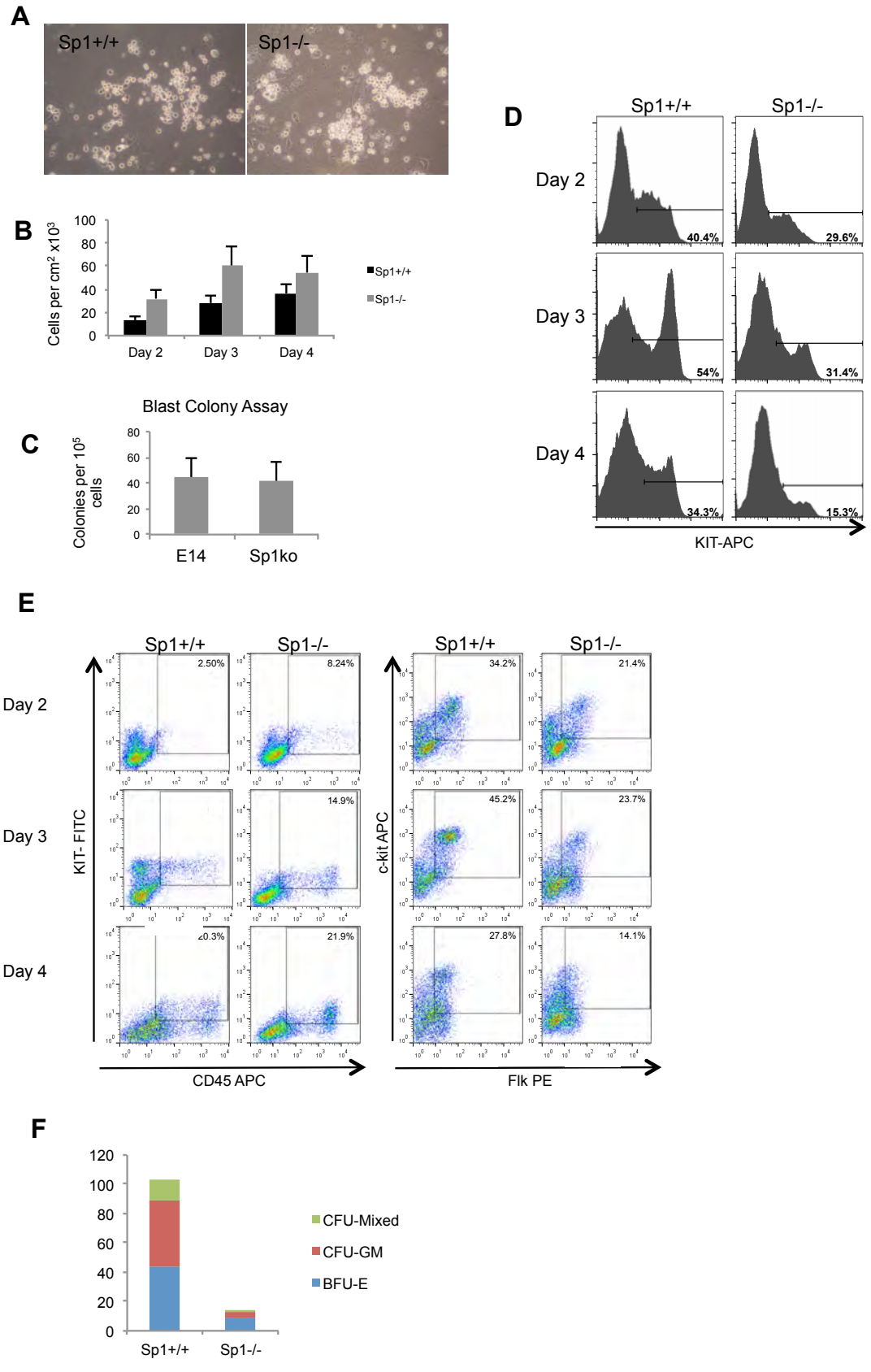
Supplementary Figure 1.



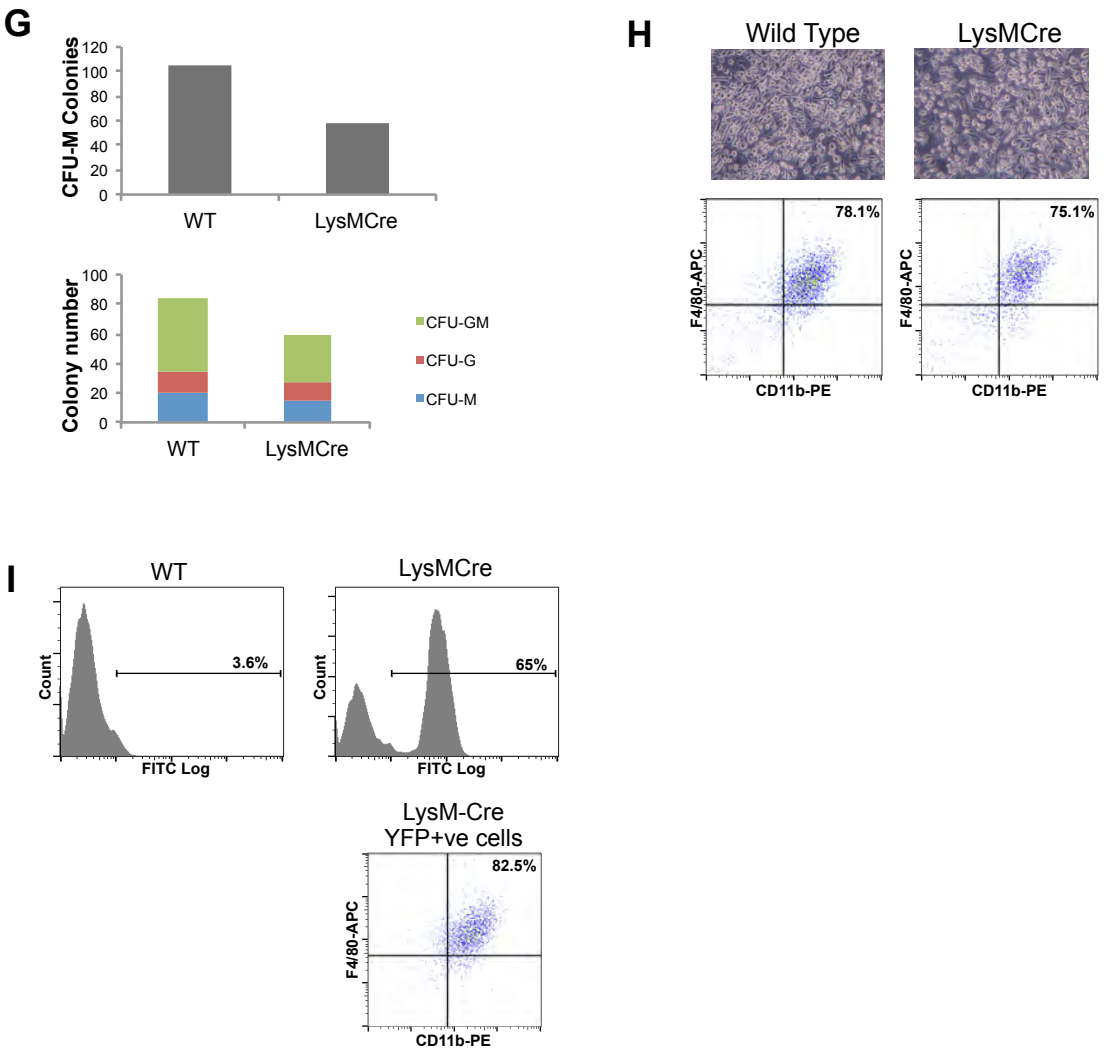
Supplementary Figure 1

A. Western blot showing the expression levels of WT Sp1 and the deletion mutant in Sp1+/+ wild type, Sp1-/- and Sp1 rescue ES cells. This shows that the Sp1-/- cells do not express the full length Sp1 as they lack the DNA binding domain. Sp1 rescue cells express both the truncated version and the full length Sp1. **B**. Gene expression analysis showing the expression of hematopoietic regulator genes from RNA prepared from embryoid bodies grown in methylcellulose. RNA was taken from EBs used to plate out for the hematopoietic colony assays at various time-points. **C.** CFSE proliferation analysis. Wild-type and Sp1-/- ES cells were labelled with CFSE, plated onto gelatine coated plates and CFSE measured by FACS every 12hrs. **D**. Upper panels: Representative images of individual Ery-P colonies for Sp1+/+ and Sp1-/-, similar morphology was observed. Middle panels: Wright-Giemsa staining of dispersed Ery-P colonies showing nucleated erythroblasts in both Sp1+/+ and Sp1-/-. Lower panels: Gene expression analysis of embryonic globins Hba-X and Hbb-Y. Sp1-/- samples show a delay in expression of embryonic globins but these reach similar levels to Sp1+/+.

Supplementary Figure 2.



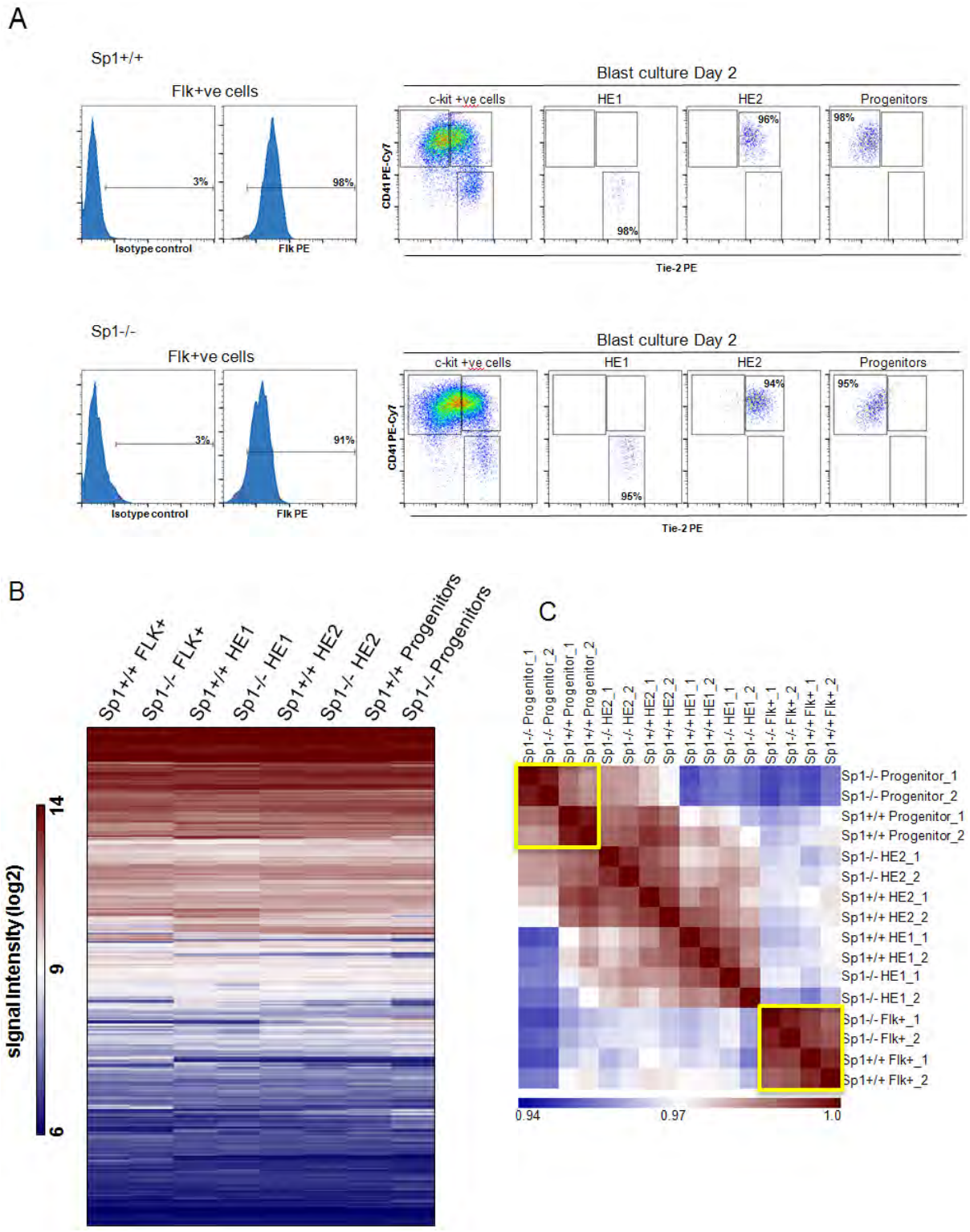
Supplementary Figure 2. cont'd



Supplementary Figure 2.

A. Representative images of Day 4 blast cultures of WT and $Sp1^{-/-}$ cells demonstrating that cultures appear morphologically normal. **B.** Cell counts of Sp1+/+ and Sp1-/- cells at days 2, 3 and 4 of blast culture. N=4, no significant difference between Sp1+/+ and Sp1-/- cell counts was observed. **C.** Blast colony assays of Sp1+/+ and Sp1-/- Flk1+ve cells. N=4, no significant difference between Sp1+/+ and Sp1-/- colony formation was observed. **D.** Representative FACS analysis of KIT expression from blast culture differentiation indicates lower levels of KIT expression and a lower proportion of KIT positive cells in $Sp1^{-/-}$ cells (refers to graph in Fig 2B). **E.** Representative FACS analysis of Sp1+/+ and Sp1-/- blast culture populations demonstrating staining of KIT

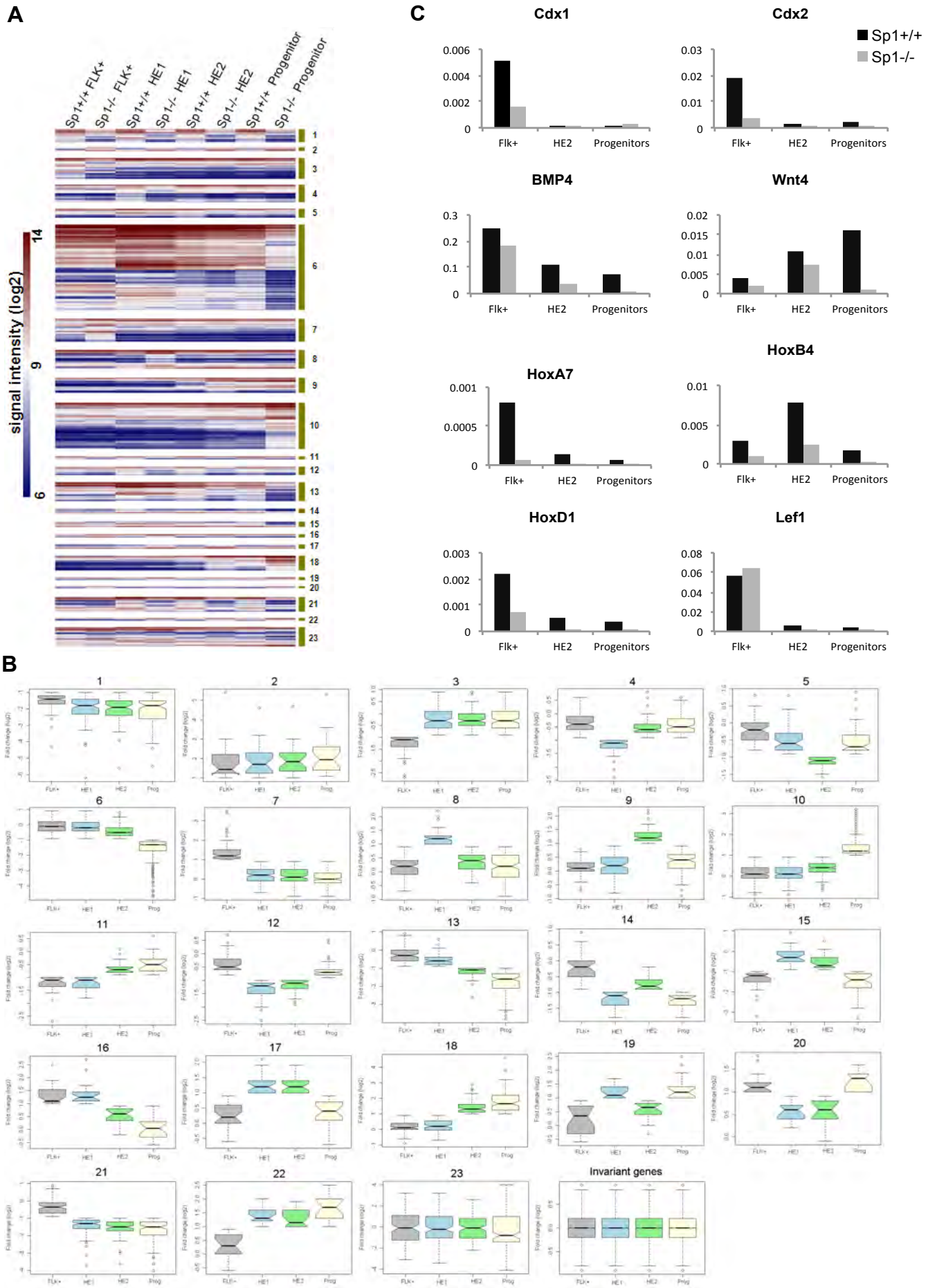
in combination with CD45 (left) and Flk1 (right) at days 2, 3 and 4 of differentiation. **F**. Colony assays from Day 3 progenitors. Colonies were scored from Day 8, and numbers of all colonies were shown to be reduced in the Sp1-/- samples. **G.** Sp1 deletion at the progenitor stage only slightly reduces myeloid colony formation. Colony assays from cells from Sp1fl/fl/Sp3+/fl x LysMCre and WT bone marrow samples. Dead cell removal was performed on frozen bone marrow cells from WT and conditional deleted mice. Cells were plated out for CFU-M (top) and CFU-C (bottom) colony assays. **H.** Sp1 deletion at the progenitor stage does not affect macrophage differentiation. Progenitors were expanded from Sp1fl/fl/Sp3+/fl x LysMCre and WT bone marrow samples and then differentiated to macrophages in liquid culture. Photographs and FACS analysis show that macrophages differentiate normally from LysMCre Sp1-/- cells. **I.** Conditional deletion is effective in macrophages. To ensure that LysMCre has been activated in macrophages, cells were gated on the YFP+ve cells in the LysMCre Sp1-/- which account for approximately 65% of the cells. These were also positive for F4/80 and CD11b markers of macrophage differentiation. This suggests that Cre-mediated deletion of Sp1 binding in developing adult macrophages does not have a major impact on macrophage differentiation.



Supplementary Figure 3.

A. Sorted populations used for expression microarray analysis. Flk1+ve cells were sorted from embryoid bodies cultured in IVD medium for 3.25-3.75 days using MACS columns. FACS histograms show Flk1 staining in Flk1+ve and –ve cell populations. More than 90% of the +ve population stained positive for Flk1. Cells were harvested at Day 2 of blast culture

differentiation, stained with anti KIT, Tie2 and CD41 antibodies and sorted into HE1 (KIT+ve, Tie2+ve, CD41-ve), HE2 (KIT+ve, Tie2+ve, CD41+ve) and progenitor (KIT+ve, Tie2-ve, CD41+ve) cell populations on a Moflow FACS sorter. The sorted populations were more than 95% pure. **B**. Microarray expression analysis of RNA prepared from purified Flk1+ cells, the first and second stage of the hemogenic endothelium and from progenitors (for a detailed description see Figure 2A). Hierarchical clustering of gene expression signals for all genes that are expressed in Sp1+/+ and Sp1-/- cells in the four cell populations. Technical replicates were averaged. **C.** Pearson correlation analysis of the microarray signals obtained with the sorted populations. Duplicates have a strong similarity and cell populations cluster together but show less similarity as differentiation progresses.


A
Sp1+/+ FLK+
Sp1-/- FLK+
Sp1+/+ HE1
Sp1-/- HE1
Sp1+/+ HE2
Sp1-/- HE2
Sp1+/+ Progenitor
Sp1-/- Progenitor
signal intensity (log2)
14
9
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
C
Sp1+/+
Sp1-/-
Cdx1
Cdx2
BMP4
Wnt4
HoxB4
HoxA7
HoxD1
Lef1
Flk+
HE2
Progenitors
B
Invariant genes

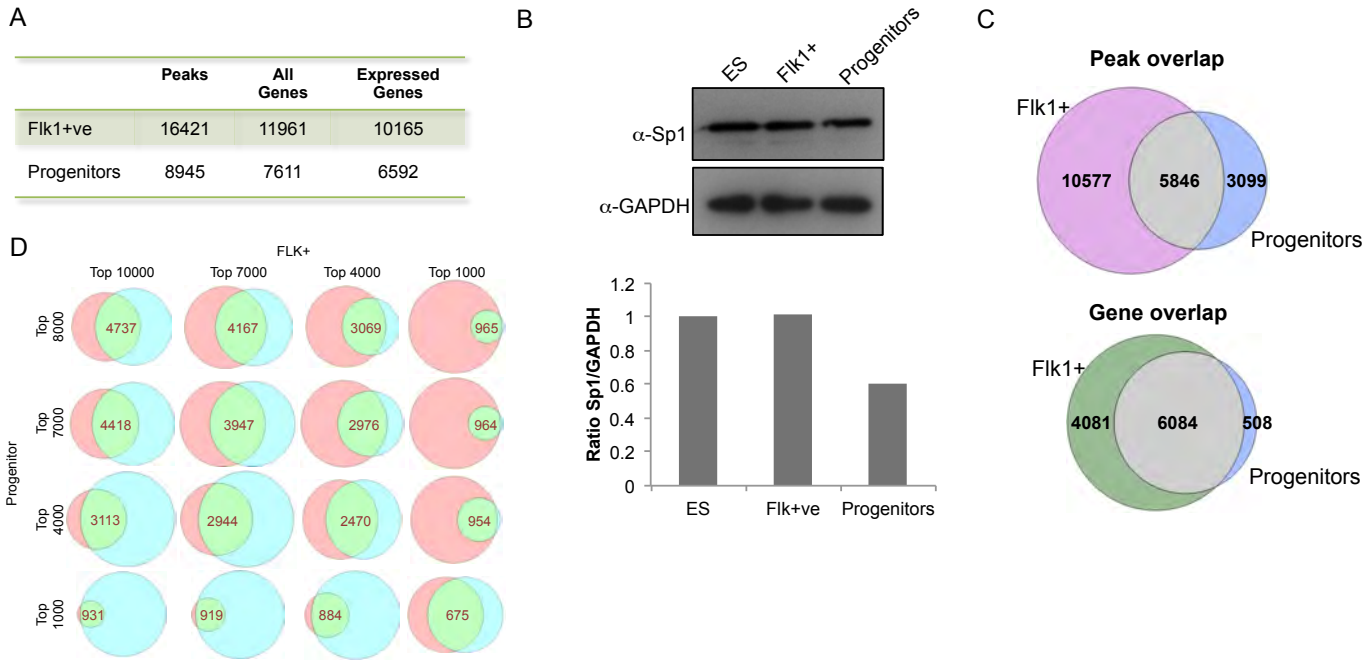
Supplementary Figure 4.

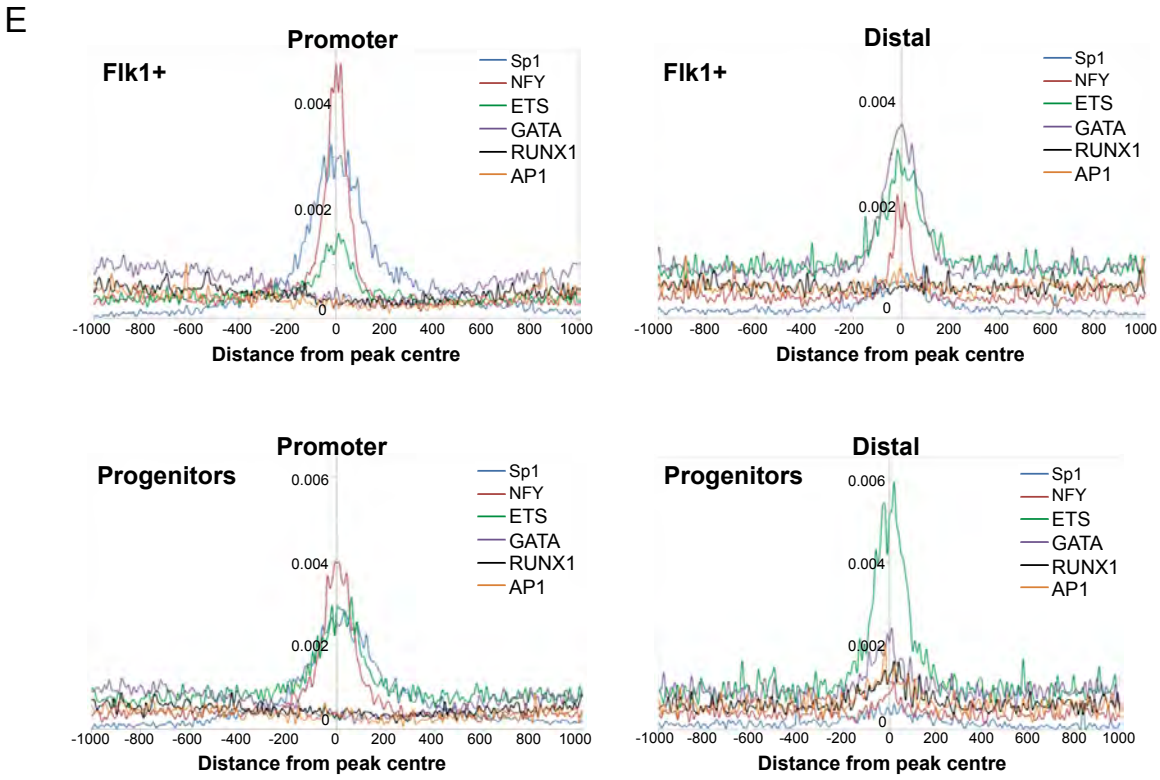
A. Hierarchical clustering of gene expression signal intensity defining the 23 clusters for each of the populations as depicted in the heat-map. **B**. Box-plots depicting gene expression fold-change of genes within the 23 clusters demonstrating that clustering is valid. **C.** Validation of microarray gene expression was performed for selected genes using RT-PCR.

Supplementary Figure 5

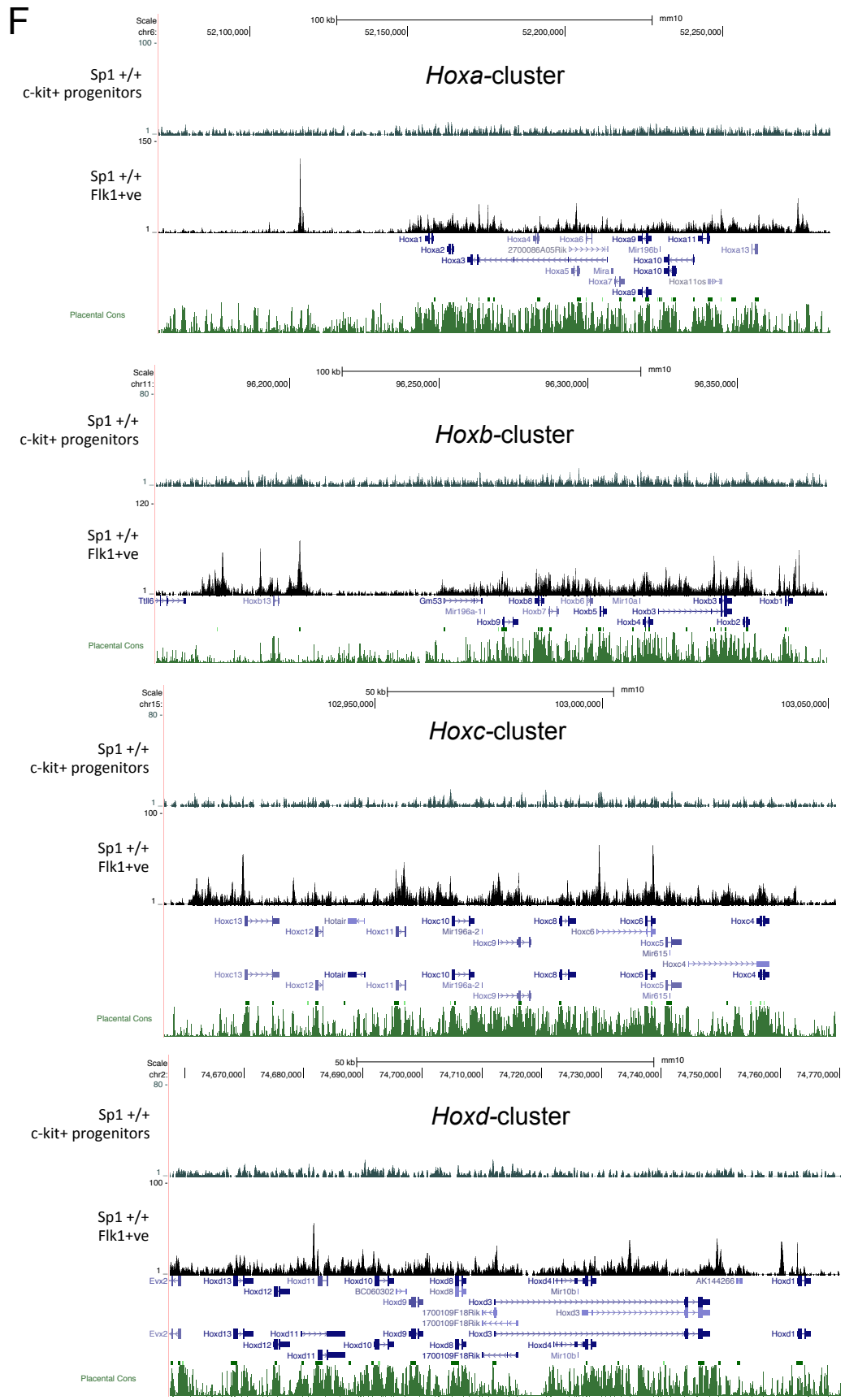
A

	Peaks	All Genes	Expressed Genes
Flk1+ve	16421	11961	10165
Progenitors	8945	7611	6592





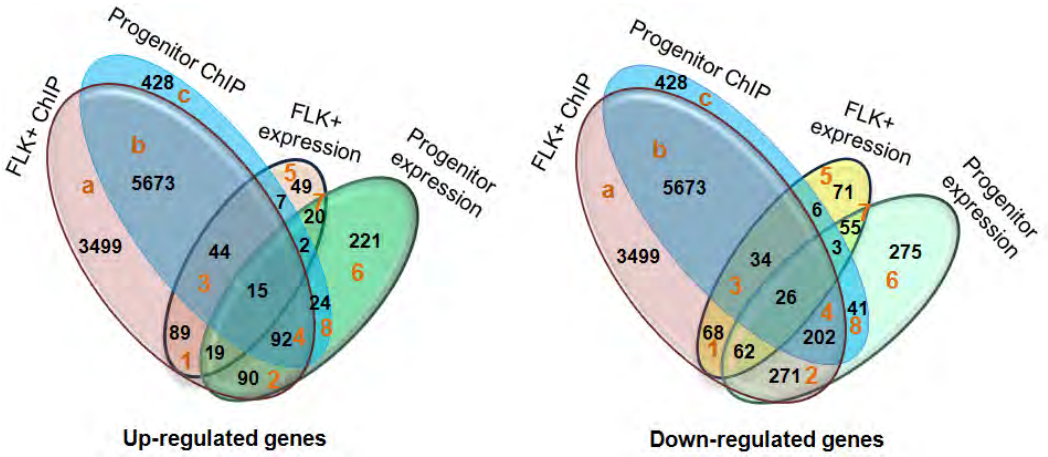
Supplementary Figure 5 ct'd



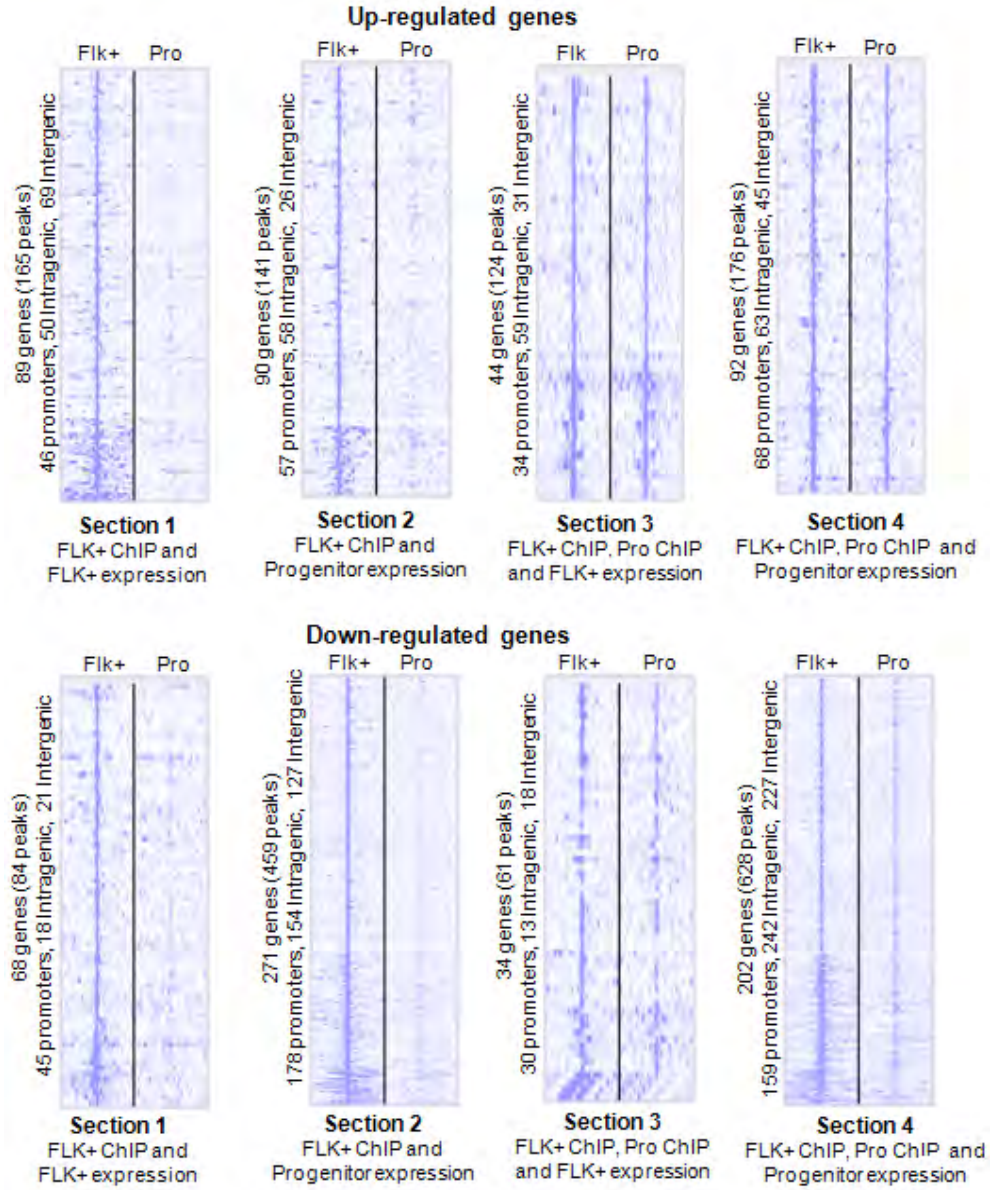
Supplementary Figure 5.

A. Table showing the numbers of peaks and genes for each cell population. **B**. Western blot showing Sp1 expression in ES, Flk1+ and Progenitor cells. Levels of Sp1 were quantified using Bio-Rad Quantity One Software and normalised to GAPDH. Graph shows an average of two experiments. **C**. Venn diagram showing the overlap in Sp1 peaks (left panel) and genes (right panel) between the Flk1+ve and progenitor populations. **D**. Sp1 binds to an overlapping, but not identical set of targets in Flk1+ve and progenitor cells; Venn diagrams showing the overlap in the binding sites for the top hits in Flk1+ve and progenitors ranked in order of increasing tag-counts. **E**. Motif distribution around the Sp1 binding sites show the positions of the other TF binding motifs that co-localise with Sp1, indicating that most motifs are located within 400 base-pairs of the Sp1 binding site. **F**. Screenshots showing Sp1 binding at the *HoxA, B, C, D* gene clusters.

A

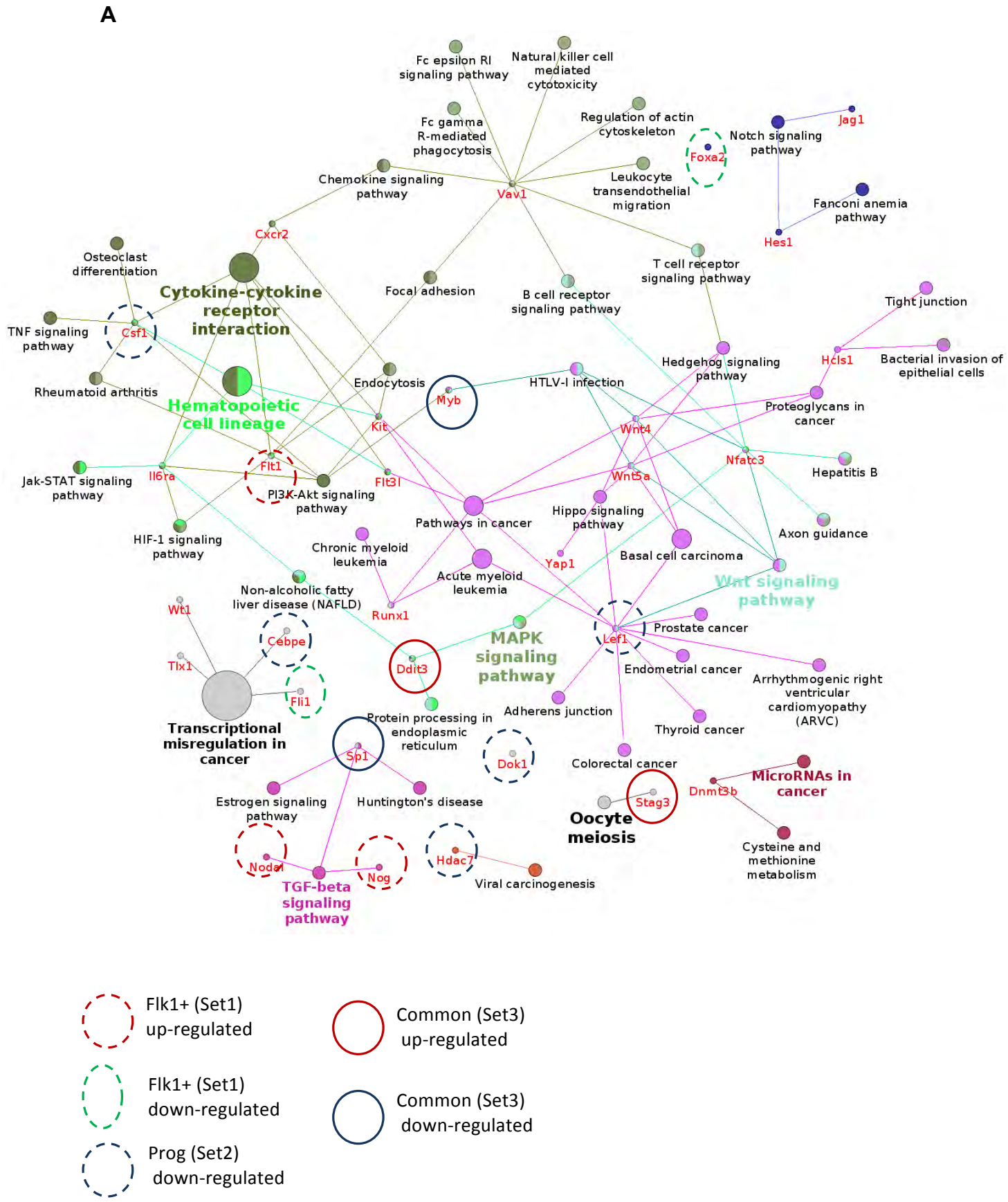

Progenitor ChIP
FLK+ ChIP
FLK+ expression
Progenitor expression
428
c
b
a
5673
3499
5
49
7
1
20
2
44
3
15
24
89
19
92
4
8
1
90
2
221
6
Up-regulated genes
Progenitor ChIP
FLK+ ChIP
FLK+ expression
Progenitor expression
428
c
b
a
5673
3499
5
71
6
1
55
3
34
3
26
4
41
68
1
62
202
8
271
2
275
6
Down-regulated genes

B


Up-regulated genes
Flk+
Pro
89 genes (165 peaks)
46 promoters, 50 Intragenic, 69 Intergenic
Section 1
FLK+ ChIP and
FLK+ expression
90 genes (141 peaks)
57 promoters, 58 Intragenic, 26 Intergenic
Section 2
FLK+ ChIP and
Progenitor expression
Flk
44 genes (124 peaks)
34 promoters, 59 Intragenic, 31 Intergenic
Section 3
FLK+ ChIP, Pro ChIP
and FLK+ expression
92 genes (176 peaks)
68 promoters, 63 Intragenic, 45 Intergenic
Section 4
FLK+ ChIP, Pro ChIP and
Progenitor expression
Down-regulated genes
68 genes (84 peaks)
45 promoters, 18 Intragenic, 21 Intergenic
271 genes (459 peaks)
178 promoters, 154 Intragenic, 127 Intergenic
34 genes (61 peaks)
30 promoters, 13 Intragenic, 18 Intergenic
202 genes (628 peaks)
159 promoters, 242 Intragenic, 227 Intergenic

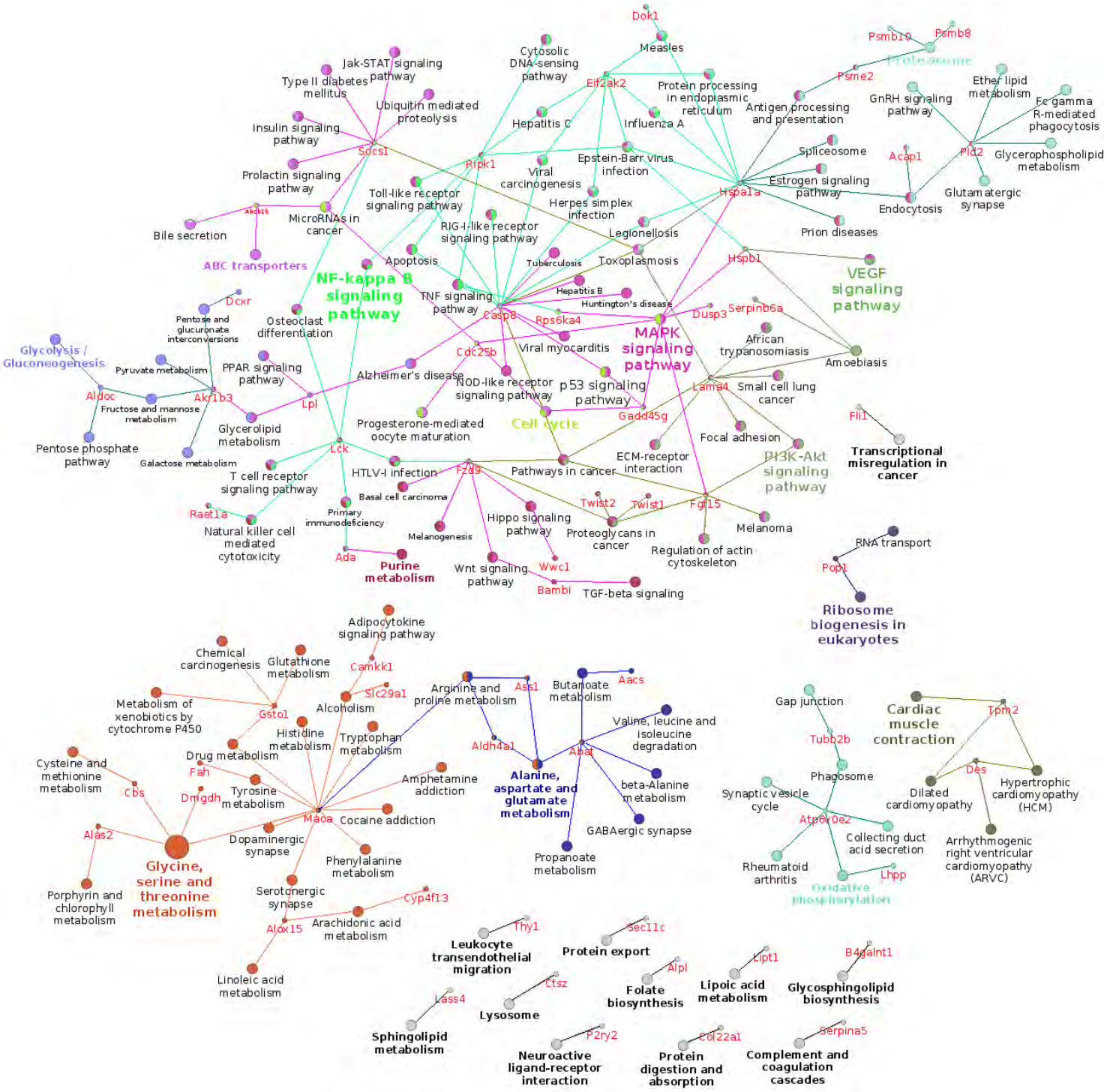
Supplementary Figure 6.

A. Venn diagram showing the integration of ChIP-seq and gene expression data in Flk1+ and progenitor cells. The number of genes in sections a, b and c shows the number of genes that are neither up nor down regulated which are the same numbers in the right and left panel. **B**. Heatmaps depicting the distribution of ChIP-seq signals found in genes responsive to Sp1 knock-out around the Sp1 binding sites in Flk1+ cells and progenitors. The labelling on the left indicates the number of genes and the position of the peaks (promoter, intragenic, intergenic)



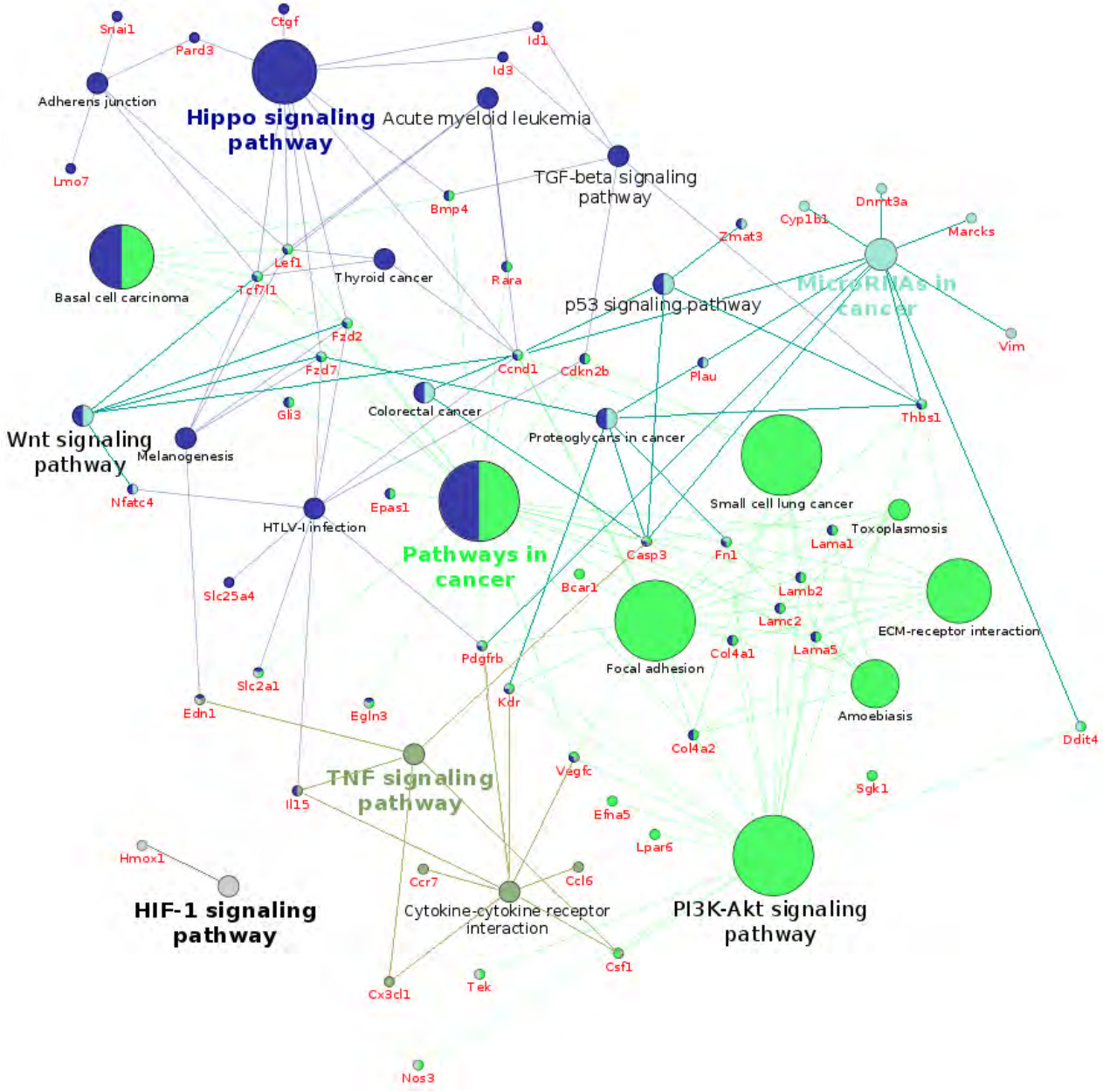
KEGG pathway analysis of deregulated genes in Sp1-/- Flk1+ cells and progenitors

B



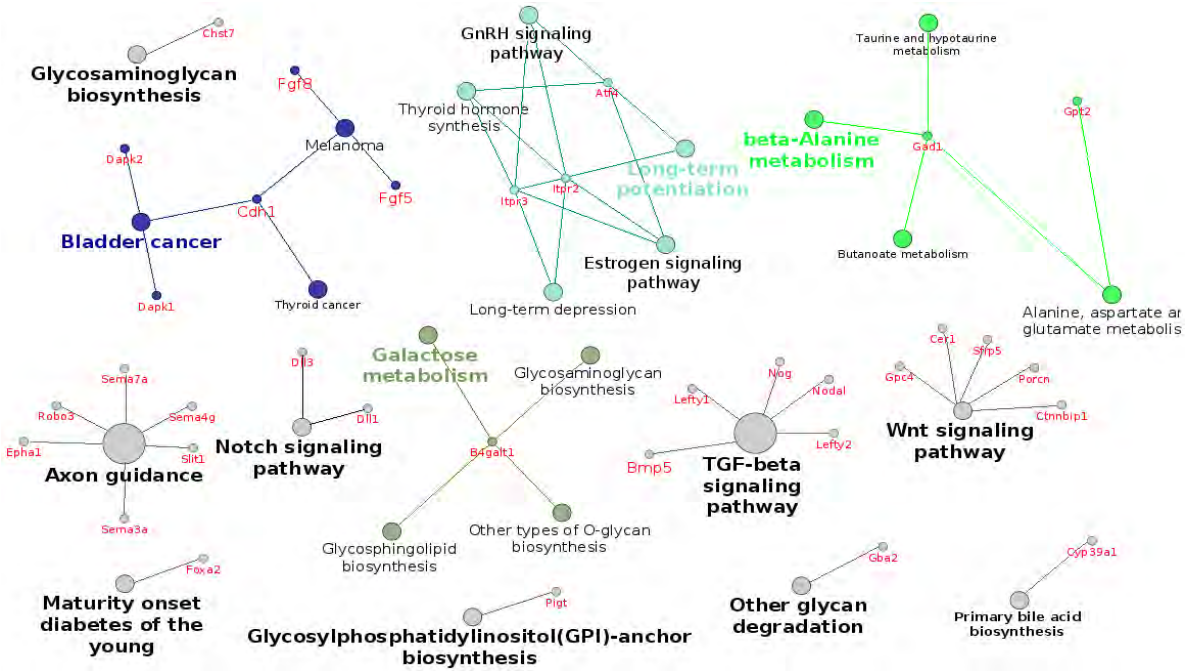
KEGG pathway analysis of all Sp1 target genes down-regulated in Flk1+ cells

C



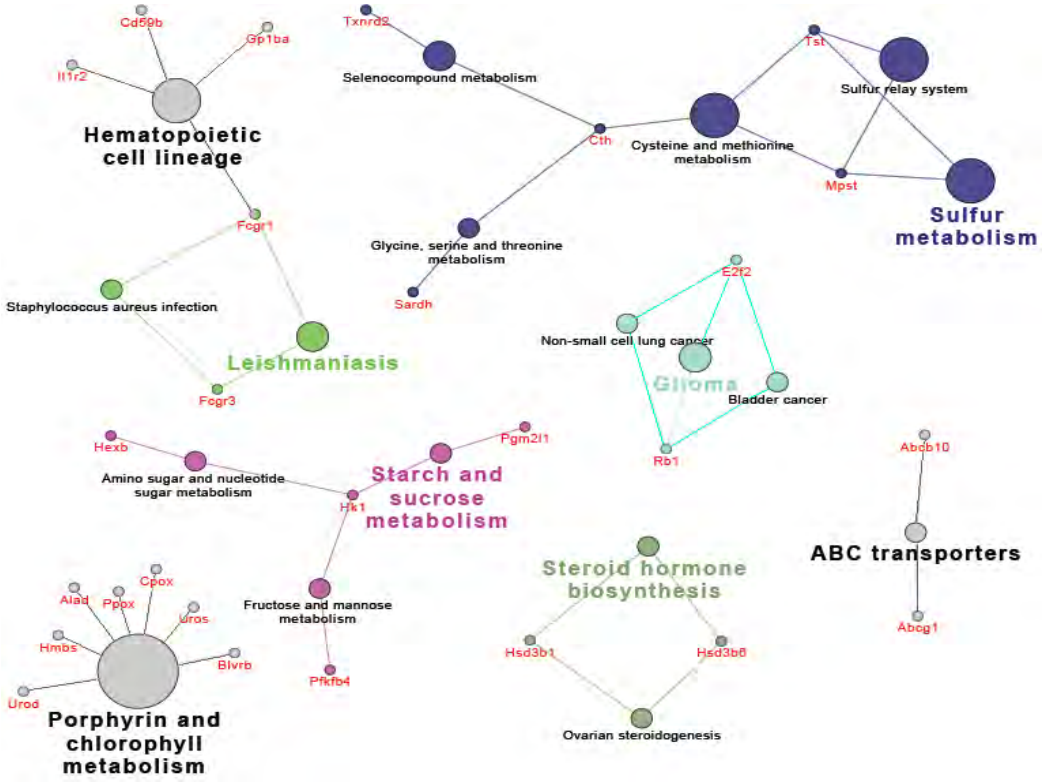
KEGG pathway analysis of all Sp1 target genes down-regulated in progenitor cells

D



KEGG pathway analysis of all target genes up-regulated in Flk1+ cells

E



KEGG pathway analysis of all target genes up-regulated in progenitor cells

Supplementary Figure 7. KEGG pathway analysis: functionally grouped KEGG pathway term networks using kappa statistics implemented by ClueGO to link the terms in the network. The right-sided enrichment (depletion) test based on the hyper-geometric distribution is used for terms and groups. The groups are created by iterative merging of initially defined groups based on the kappa score threshold. The relationship between the selected terms is defined based on their shared genes and the final groups are randomly coloured where functional groups represented by their most significant term. One, two or more colours represents that a gene/term is a member of one, two or more groups respectively. The size of the nodes reflects the enrichment significance of the terms. The network is automatically laid out using the layout algorithm supported by Cytoscape. **A**: KEGG pathway analysis of selected deregulated genes in Sp1-/- Flk1+ cells and progenitors that are shown in Figure 4C. **B**: KEGG pathway analysis of all Sp1 target genes down-regulated in Flk1+ cells. **C**: KEGG pathway analysis of all Sp1 target genes down-regulated in progenitor cells. **D**: KEGG pathway analysis of all target genes up-regulated in Flk1+ cells. **E**: KEGG pathway analysis of all target genes up-regulated in progenitor cells.

Download Table S1

Download Table S2

Download Table S3

Download Table S4

Download Table S5