

# The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation

Eric Soler,<sup>1,7,9</sup> Charlotte Andrieu-Soler,<sup>1,7</sup> Ernie de Boer,<sup>1,7</sup> Jan Christian Bryne,<sup>2,8</sup> Supat Thongjuea,<sup>2,8</sup> Ralph Stadhouders,<sup>1</sup> Robert-Jan Palstra,<sup>1</sup> Mary Stevens,<sup>1</sup> Christel Kockx,<sup>3</sup> Wilfred van IJcken,<sup>3</sup> Jun Hou,<sup>1</sup> Christine Steinhoff,<sup>4</sup> Erikjan Rijkers,<sup>5</sup> Boris Lenhard,<sup>2,11</sup> and Frank Grosveld<sup>1,6,10</sup>

<sup>1</sup>Department of Cell Biology, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands; <sup>2</sup>Computational Biology Unit-Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway; <sup>3</sup>Biomics Department, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands; <sup>4</sup>Department of Computational Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; <sup>5</sup>Department of Biochemistry, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands; <sup>6</sup>Center for Biomedical Genetics and Cancer Genomics Center, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands

One of the complexes formed by the hematopoietic transcription factor Gata1 is a complex with the Ldb1 (LIM domain-binding protein 1) and Tal1 proteins. It is known to be important for the development and differentiation of the erythroid cell lineage and is thought to be implicated in long-range interactions. Here, the dynamics of the composition of the complex—in particular, the binding of the negative regulators Eto2 and Mtgr1—are studied, in the context of their genome-wide targets. This shows that the complex acts almost exclusively as an activator, binding a very specific combination of sequences, with a positioning relative to transcription start site, depending on the type of the core promoter. The activation is accompanied by a net decrease in the relative binding of Eto2 and Mtgr1. A Chromosome Conformation Capture sequencing (3C-seq) assay also shows that the binding of the Ldb1 complex marks genomic interaction sites *in vivo*. This establishes the Ldb1 complex as a positive regulator of the final steps of erythroid differentiation that acts through the shedding of negative regulators and the active interaction between regulatory sequences.

[*Keywords:* ChIP sequencing; transcription factor complexes; development; differentiation; erythropoiesis; long-range interactions]

Supplemental material is available at <http://www.genesdev.org>.

Received August 5, 2009; revised version accepted December 4, 2009.

Hematopoietic stem cell differentiation to the erythroid lineage involves coordinated changes in transcription, often by functionally conserved genes such as *Gata2*, *Tal1*, *Lmo2*, *Gata1*, and *Runx1* (Cantor and Orkin 2001). These factors can form different complexes, but also interact with each other in a transcription complex known as the Ldb1 complex. The nuclear protein Ldb1 (LIM domain-binding protein 1) has no DNA-binding or enzymatic activities. Its main functional domains are the LIM interaction domain (LID) in the C-terminal part of the

protein, interacting with LIM-only protein (LMO) and LIM homeodomain (LIM-HD) factors; the dimerization domain in the N-terminal part of the protein; the Ldb1/Chip conserved domain (LCCD) domain, which interacts with the Ssbp proteins; and the NLS (nuclear localization signal) (Jurata and Gill 1997; Breen et al. 1998; Matthews and Visvader 2003; Xu et al. 2007). Studies in *Xenopus* have shown that LIM and dimerization domains are important for its function *in vivo* (Breen et al. 1998).

In murine erythroid cells, Ldb1 was originally described as part of a complex composed of the transcription factors Tal1, Gata1, and E2A, and the non-DNA-binding Lmo2 functions as a bridging molecule, together with the assistance of Ldb1, between the DNA-binding factors (Wadman et al. 1997). The complex binds to a GATA-E-box motif and is thought to bind to a number of genes that are up-regulated in erythroid differentiation (Brand et al. 2004), including the  $\beta$ -globin locus control region (LCR) and the  $\beta$ -globin gene promoter, where it is thought to be

<sup>7</sup>These authors contributed equally to this work.

<sup>8</sup>These authors contributed equally to this work.

Corresponding authors.

<sup>9</sup>E-MAIL [e.soler@erasmusmc.nl](mailto:e.soler@erasmusmc.nl); FAX 31-10-7044743.

<sup>10</sup>E-MAIL [f.grosveld@erasmusmc.nl](mailto:f.grosveld@erasmusmc.nl); FAX 31-10-7044743.

<sup>11</sup>Corresponding author for informatics.

E-MAIL [boris.lenhard@bccs.uib.no](mailto:boris.lenhard@bccs.uib.no); FAX 47-555-84295.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.551810>. Freely available online through the *Genes & Development* Open Access option.

important for looping of the LCR to the promoter (Song et al. 2007).

A recent proteomics study (Meier et al. 2006) in mouse erythroleukaemic (MEL) cells expanded the number of proteins present in the Ldb1/Lmo2/Tal1/E2A/Gata-1 complex to include HEB, Lmo4, and Lyl1 (closely related to Tal1), and a number of proteins (Ssbp1–4) important for the stability of the Ldb1 protein (Xu et al. 2007). In the proerythroblast-proliferating state, this complex was found to interact with another complex consisting of Gata1/Tal1/E47/HEB/Mtgr1/Eto2 as well as with the cell cycle regulator Cdk9 and E2-2, with the equilibrium favoring the larger multiprotein complex (Meier et al. 2006). Upon differentiation and cessation of proliferation of the MEL cells, Cdk9 and E2-2 are no longer part of the larger complex, which may favor the dissociation into the two smaller complexes and allow the dimerization of the smaller complex to form loops in the chromatin. The process coincides with a decrease in the level of the suppressive factors Eto2 and Mtgr1 (Davis et al. 2003; Schuh et al. 2005), and an increase in the level of another Lim-only factor, Lmo4 (Grutz et al. 1998; Kenny et al. 1998).

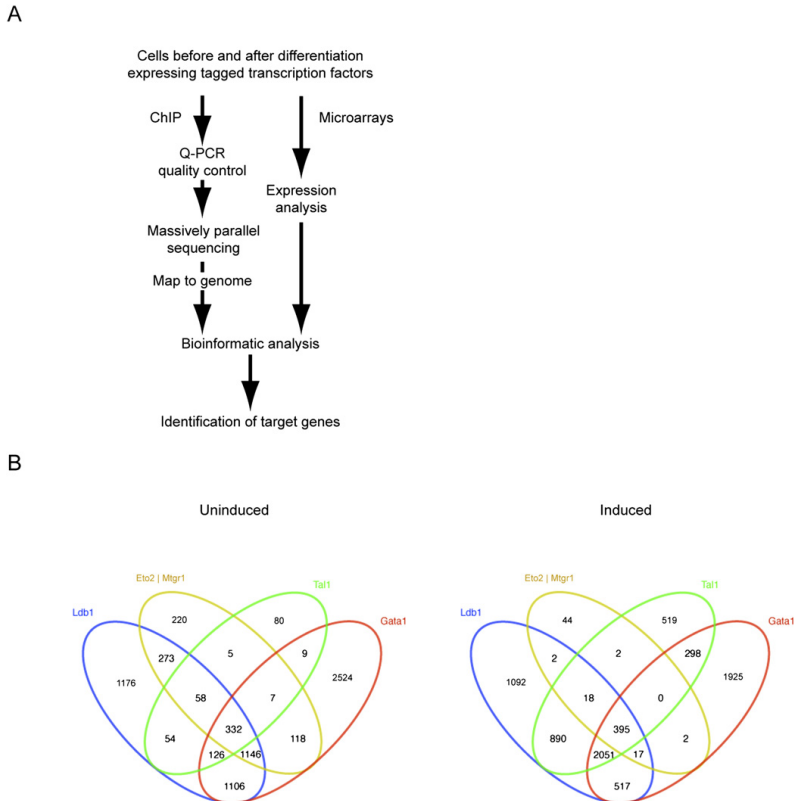
Absence of Ldb1 results in early death of the embryo between embryonic day 9.5 (E9.5) and E10.5 with a series of developmental defects, including truncation of the anterior head structure, no heart formation, axis duplication, and absence of hematopoiesis (Mukhopadhyay et al. 2003; M Mylona, JC Bryne, W van IJcken, and F Grosveld, unpubl.). The latter defect resembles the knockout phenotypes of the hematopoietic transcription factors Gata2/Gata1, Lmo2, and Tal1 (Warren et al. 1994; Robb et al. 1995; Shivdasani et al. 1995; Porcher et al. 1996). The early phenotype of these knockouts is explained by the observation that Ldb1, Cdk9, E2A, Lmo2, Gata1, and Eto2 are expressed in the PSp/AGM region of the 9.5-d-post-coitum (dpc) mouse embryo that will later give rise to hematopoietic stem cells (Meier et al. 2006). Differentiation of embryonic stem cells *in vitro* shows that the number of blast colony-forming cells, the *in vitro* equivalent of the hemangioblast, is severely reduced, and they have lost the potential to grow and differentiate (M Mylona, JC Bryne, W van IJcken, and F Grosveld, unpubl.).

Although a number of target genes of the Ldb1 complex have been described (Anguita et al. 2004; Brand et al. 2004; Lahlil et al. 2004; Tripic et al. 2009) its targets are largely unknown, although a subset of these were reported very recently (Wilson et al. 2009). Here, we report a dynamic picture of the last steps of erythroid differentiation in a model system, through the analysis of the target genes of the Ldb1 complex, by integrating its genome-wide target-binding sequences with the changing gene expression profiles of C88 MEL cells before and after differentiation. This analysis shows a changing binding site profile for the Ldb1 complex during differentiation, for the first time reveals the differences between different classes of core promoters with respect to the preferred relative position of activating complexes, and shows that binding sites of the Ldb1 complex mark positions of long-range interactions.

## Results

### *Analysis of binding sites*

The basic scheme (Fig. 1A) for the analysis of the Ldb1 complex entails the introduction of a tag into a number of the transcription factors that have been identified previously (Meier et al. 2006). The biotinylation tag, which is biotinylated by the bacterial biotin ligase BirA (Cronan and Reed 2000; de Boer et al. 2003), and the viral V5 tag (Southern et al. 1991) are both insensitive to formaldehyde fixation and, in particular, the biotin tag is excellent in chromatin immunoprecipitations (ChIPs) followed by sequencing (ChIP-seq) (e.g., see Ku et al. 2008; Kolodziej et al. 2009; Soler et al. 2009). The tagged genes are subsequently introduced by stable transfection into C88 MEL cells, and only clones that show expression levels lower than or similar to the endogenous factors (both before and after differentiation of the C88 cells) are used for further analysis to avoid possible phenotypic or other changes due to the overexpression of one of the factors. None of the transfected cell lines showed an aberrant phenotype, and all were fully capable of differentiation (data not shown). Moreover, the tagged proteins are regulated the same as the endogenous factors (Supplemental Fig. 1). After a particular cell line was grown, the ChIP-seq protocol was carried out as described (Soler et al. 2009) for Ldb1, Gata1, Tal1 (Sc1), Mtgr1, and Eto2. Typically, the ChIPs of known binding sites show an enrichment of 10-fold to 30-fold for either V5 or the biotin tag, whereas sequences that do not bind the complex are not enriched (Supplemental Fig. 2A). The precipitated material was sequenced using the standard Illumina/Solexa protocol, the sequences were mapped back to the genome using the Eland software (Illumina), and peaks were visualized using in-house-developed software (Solex) (Soler et al. 2009; HWJ Rijkers, unpubl.) or the ChIP-seq Visualization Browser based on G-Browse (<http://tracc.genereg.net>). Between 10 and 60 million sequences were mapped back to the genome for each of the factors. For example, at the comparable noise level (maximal empirical *P*-value of 5%), we were able to detect 5205 Gata1-binding sites, 4173 Tal1-binding sites, and 4982 Ldb1 sites in the induced state in the entire genome (Fig. 1B). The number of actual binding sites is certainly higher than that, as confirmed by a much higher than random co-occurrence of these factors at sites that fall below the 10% false discovery rate (FDR) threshold. We validated the MEL cell data by carrying out the same analysis in primary erythroid cells derived from mouse E13.5 fetal liver, which shows an overlap of 83% of the binding sites (Supplemental Fig. 3). The significance of the data is immediately apparent when the transcription factor-binding sites are compared with the control ChIP-seq (Supplemental Fig. 2B). The analysis of the correlation of binding sites between the factors shows that the large majority of strong Ldb1-binding locations after differentiation is also bound by Gata1 and Tal1, establishing the Ldb1 complex as the most common way of Ldb1-mediated regulation of terminal erythropoiesis (Figs. 1B, 5 [below]). On the other hand, as expected, each of the factors can form



**Figure 1.** Analysis of transcription factor-binding sites by ChIP-seq using Solexa Genome Analyzer II. (A) Description of the work flow used for transcription factor analysis. (B) Numbers of different complexes of the five measured factors, estimated using conservative criteria (empirical  $P$ -value of 0.05 on the experiment with the weakest signal). The factors were considered part of the same complex if the distance between centers of their binding peaks were within the resolution of the method (200 bp).

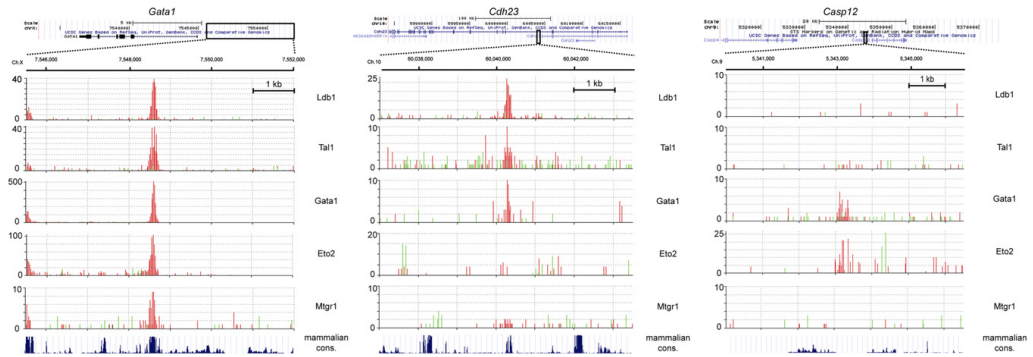
complexes with factors other than the ones examined here (Meier et al. 2006). For example, a subset of strong Gata1 sites shows little or no concomitant Tal1 and Ldb1 binding, revealing the existence and positions of different types of Gata1 complexes in the same cell system. Indeed, Gata1 is known to form a number of complexes without Tal1 or the other members of the Ldb1 complex (Hong et al. 2005; Rodriguez et al. 2005). However, even between the different members of the complete complex there are a number of different target sites bound by smaller complexes. For example, the  $-3.5$ -kb enhancer of the *Gata1* gene binds the entire complex (Fig. 2), whereas the regulatory sequence of the *Cdh23* gene binds an Ldb1 complex without Eto2, while the *Casp12* gene binds an Eto2 complex but not Ldb1.

The list of binding sites contains many of the known target genes (Wilson et al. 2009), but also many more novel ones. In a number of cases, however, it is not clear what the target gene is that may be regulated by the Ldb1 complex;

in particular, in those situations where binding sites are found at a large distance from any gene (see below).

#### Expression analysis to identify target genes

Two experiments were carried out in order to identify genome-wide the target genes of the Ldb1 complex. First, the binding sites of the complex were also determined after differentiation of the cells, and the gene expression patterns were determined before and after differentiation. The cells used above were differentiated, and the binding sites of the factors were determined as above (<http://tracc.genereg.net/download>). Strikingly, most binding sites were still binding the same complexes, but the ratio of binding of the different factors was changed. Most importantly, the number of binding sites and the peak height of the suppressor factors Eto2 and Mtgr1 decreased after differentiation relative to the other factors, indicating that the Ldb1 complex changes toward an activating state during differentiation.



**Figure 2.** Differences in transcription factor complex composition at selected loci. The number of overlapping sequence reads originating from the different ChIP-seq experiments were plotted relative to chromosomal position using Solex (red bars). Signals obtained from the control ChIP-seq experiment are also shown (green bars). Chromosomal positions as well as sequence conservation (mammalian cons.) are indicated *above* and *below* ChIP-seq plots, respectively. The figure shows clustering of all five transcription factors at the *Gata1* hematopoietic enhancer (*left* panel), Ldb1 complex without Eto2 at the *Cdh23* locus (*middle* panel), and Eto2 complex without Ldb1 at the *Casp12* gene promoter (*right* panel).

Figure 3 shows examples of the binding sites of a number of target genes before and after differentiation.

All of these genes—*Gypa* (glycophorin A), *Epb4.2* [Band 4.2], *Alas2*, and *Slc22a4*—are induced late during erythroid differentiation and show a difference in the ratio of the binding of the different components of the Ldb1 complex. In particular, the level of the negative regulators Eto2 and Mtrg1 (Davis et al. 2003) decreases during differentiation (Meier et al. 2006), whereas levels of Ldb1 and Tal1 increase, which is reflected in a change in the relative ratio of Eto2/Mtrg1 to Ldb1, or Tal1 binding to the regulatory sequences of these genes (Fig. 3E,F). This observation is supported by the genome-wide data analysis; at the same stringency threshold (empirical *P*-value), there is a significant drop in the Eto2- and Mtrg1-level content of Ldb1 complexes upon the induction of differentiation.

We next determined the changes in gene expression pattern before and after differentiation by microarrays to allow the identification of target genes of the Ldb1 complex (Fig. 4; Supplemental Table 1; Supplemental Fig. 4). This showed that many genes in the different pathways—such as heme biosynthesis, cell cycle, apoptosis, and gas transport—that are active in terminal erythroid differentiation are target genes of the Ldb1 complex (e.g., see Supplemental Table 2). This suggests that the function of Eto2/Mtrg1 is to suppress target genes before differentiation. This is confirmed by knocking down Eto2, which shows that direct targets of the complex are derepressed in both MEL cells and primary erythroid cells (Supplemental Fig. 5) without differentiation of the cells.

#### Motif analysis of transcription factor-binding sites

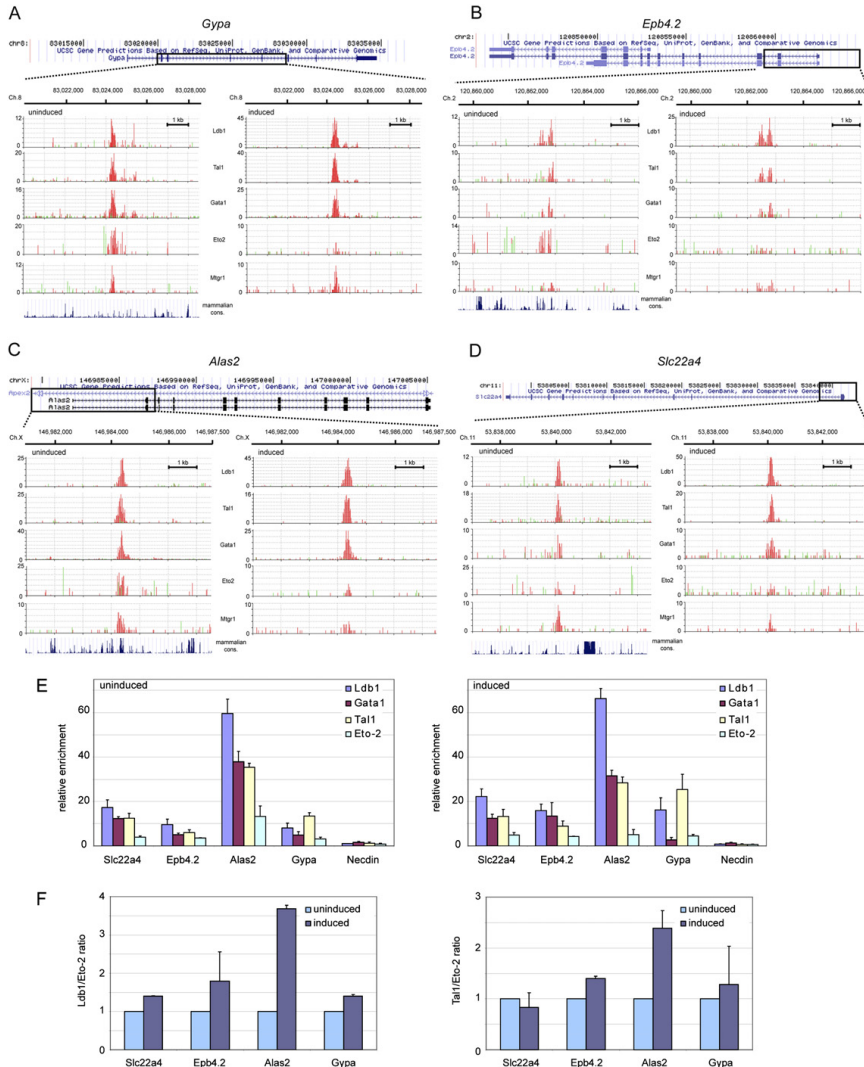
It is clear that the top of any of the peaks of binding sites points directly at the recognition sequence of the binding site. This provides a consensus sequence of the binding sites as they are used in vivo. The Gata1 site strongly

matches the previously reported consensus WGATAR (Supplemental Fig. 6), and is much better-defined than, e.g., DNA SELEX-derived Gata1 profiles (e.g., MA0035 from JASPAR database) (Bryne et al. 2008). The Tal1-binding sites give a different picture: While the published E-box motif consensus is CANNTG (Murre et al. 1991), our motif discovery on the bound sites using MEME (Bailey et al. 2009) revealed a preference for TG dinucleotide upstream of WGATAR, with a preferred consensus (C)TGN<sub>7-8</sub>WGATAR. Within the first 1284 top sites co-occupied by Gata1 and Tal1, 1045 (81%) show the presence of a TG dinucleotide 7–8 bases upstream of the WGATAR motif (TGN<sub>7-8</sub>WGATAR), among which 671 sites contain a C upstream of the TG (CTGN<sub>7-8</sub>WGATAR). The remaining 239 sites show an overrepresentation of E-box motif CWGCTG (with a positional overrepresentation on the 5' side of the GATA sequence) (Supplemental Fig. 6). The separation of Gata1-binding sequences into those that bind Tal1 and those that do not, followed by de novo motif discovery on each set, revealed that this upstream element is strongly overrepresented in Tal1-positive sites only (Supplemental Fig. 6).

The other surprising observation is that, often, more than one consensus binding site is observed at any given position. For example, one of the binding sites of the *Epb4.2* (Band 4.2) gene shows a double site (Supplemental Fig. 7). Multiple sites may serve to increase the efficiency of complex recruitment and/or may allow simultaneous interactions with several distant sites.

#### Correlation of transcription factor-binding sites with gene expression and promoter type

The genes that were up-regulated or down-regulated were subsequently compared with the positions of experimentally detected binding sites to determine whether they are direct targets of the Ldb1 complex. Supplemental Figure 8



**Figure 3.** Dynamics of the Ldb1 complex components during the course of erythroid differentiation. ChIP-seq data showing binding of Ldb1, Tal1, Gata1, Eto2, and Mtg1 to the *Gypa*, *Epb4.2*, *Alas2*, and *Slc22a4* genes. (A–D) Binding profiles before (uninduced) and after (induced) differentiation. (E) ChIP analysis on nontransfected MEL cells showing enrichments of endogenous factors before (uninduced) and after (induced) differentiation. Data are represented as mean  $\pm$  SEM. (F) Ratios of Ldb1 and Tal1 enrichments over Eto2 showing increases of Ldb1 and/or Tal1 after differentiation (ratios before differentiation are set to 1).

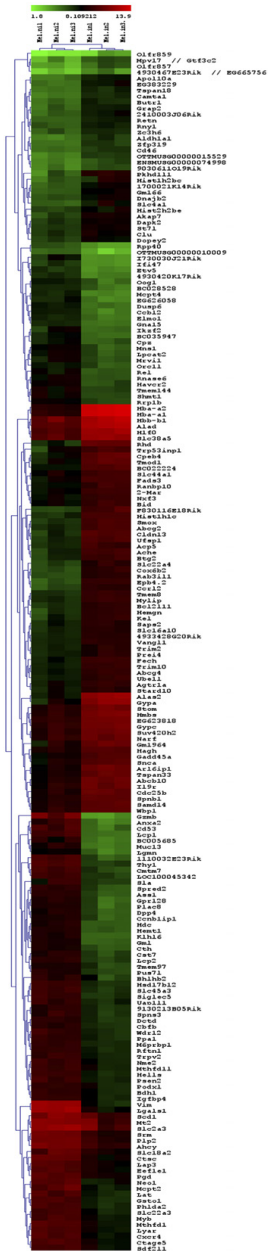
shows the example of the top 25 up-regulated or down-regulated genes and the target binding sites of the complex within 100 kb of the transcription start site (TSS). It is clear that most of these up-regulated genes are direct targets of the complex (84% have at least one strong Ldb1 complex-binding site within 50 kb of the TSS). On the other hand, most of the 25 most strongly down-regulated

genes (72%) show no evidence of binding of the Ldb1/Tal1/Gata1 complex, although there are exceptions (Supplemental Fig. 8). Including the top 50 up-regulated and down-regulated genes shows even more dramatic differences, with 96% of the up-regulated genes having binding of the complex versus only 20% for the down-regulated ones, with up-regulated ones typically sporting stronger

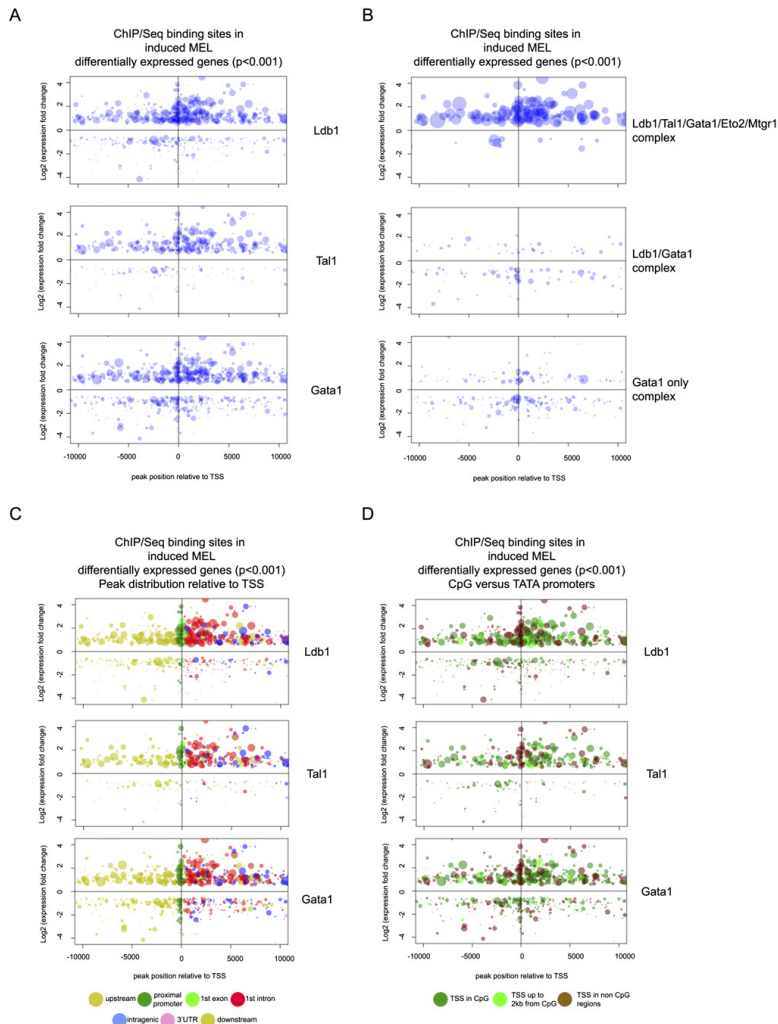
binding and greater proximity to the TSS (data not shown). Further analysis of these data by separating the binding sites of all the genes that were significantly up-regulated or down-regulated ( $P < 0.001$ ) showed that the strong

binding sites and the vast majority of binding sites of the Ldb1 complex are on up-regulated genes, but not down-regulated genes ( $P = 0$ ,  $\chi^2$  test) (Fig. 5A,B; Supplemental Fig. 8): Indeed, the significantly down-regulated genes are largely devoid of either Ldb1 or Tal1 binding. In contrast, Gata1 (which is also part of a number of repressive complexes) (Hong et al. 2005; Rodriguez et al. 2005) or a subcomplex of Gata1 with Ldb1 (lacking Tal1 and Eto2/Mtgr1) also shows frequent binding to the regulatory regions of repressed genes (Fig. 5B). Thus, the Ldb1 complexes can be subdivided into two main entities: an Ldb1/Gata1/Tal1/Eto2/Mtgr1-activating complex, and an Ldb1/Gata1 with both activating and repressing activities (Fig. 5B; Supplemental Fig. 8).

The analysis also showed that the Ldb1 complex binds in different positions relative to the TSS of the gene, from the proximal promoter region to tens of kilobases away. In a number of recent studies, we and others investigated differences in responsiveness to long-range regulation between different classes of core promoters (Engstrom et al. 2007; Megraw et al. 2009). We therefore also asked the following questions: (1) Where are the binding sites located primarily relative to the TSS? (2) Are there any differences in their distribution between genes with different types of core promoters? Although there were strong complex-binding sites in "textbook positions" upstream of the gene, the highest frequency of sites was found to map to the first intron of genes (Fig. 5C). However, when we classified the corresponding promoters into the CpG and non-CpG classes, a striking difference emerged: The Ldb1 complex-binding sites in proximal promoters of regulatory genes were found almost exclusively in non-CpG island promoters (red bubbles in Fig. 5D; Supplemental Fig. 9). Non-CpG promoters show an overrepresentation of TATA boxes and are associated predominantly with specific expression in terminally differentiated tissues (Carninci et al. 2006). On the other hand, the up-regulated genes with CpG promoters (green bubbles in Fig. 5D) showed Ldb1 complex binding predominantly downstream from the promoter, most in their first intron (Fig. 5C,D; Supplemental Figs. 8, 9). CpG island promoters most often lack TATA boxes, and are associated with either housekeeping genes or developmentally regulated genes. In our recent study (Akalin et al. 2009), a subset of CpG island promoters was shown to be the type that is most responsive to long-range regulation. It is interesting, however, that many of the most up-regulated



**Figure 4.** Hierarchical clustering of normalized expression levels. The figure summarizes the average of normalized expression levels of gene-specific probes (annotated with the same gene name after normalization) of all samples under study. Genes that showed a significant difference ( $P$ -value of Student's  $t$ -test,  $<0.05$ ) were selected. Genes with no annotation were removed, and the 100 most up-regulated and down-regulated genes were biclustered (euclidean distance, average linkage). The resulting heat map of absolute  $\log_2$ -transformed expression levels for uninduced and induced samples is shown. The color coding is a double gradient, where the high-end genes are shown in red and the low-end genes are shown in green.



**Figure 5.** Bubble plot representation of transcription factor-binding sites around differentially regulated genes. The bubble plots encode up to four quantitative parameters per one ChIP signal: distance from the promoter of a gene ( $X$ -axis), log<sub>2</sub> of fold change in expression of that gene between the induced and uninduced states ( $Y$ -axis), the intensity of the binding signal/peak (bubble diameter), and, optionally, different categorical information (bubble color) (*A*) Bubble plots showing peak size and distance to promoter of the closest gene and fold change of that gene between MEL-induced and uninduced cells. Only significantly differentially expressed genes are shown. There is a clear bias for the elements of the complex to associate with genes up-regulated upon induction ( $P \sim 0$ , Wilcoxon test). (*B*) Plot for Gata1 split by the co-occurrence with other factors at the same location. The Ldb1 complex associates with up-regulated genes, while Gata1/Ldb1 without Tal1 and Eto2/Mtgr1 clusters close to promoters of both up-regulated and down-regulated genes. (*C*) Same as *A*, colored by the region of/around the reference gene. The first exon and first intron are marked separately from the rest of the internal introns and exons—the former since it is often part of the promoter, and the latter to show that it is a preferred binding region. (*D*) Same as *A*, with bubbles/peaks colored by the promoter type of the associated genes. The distribution shows that the non-CpG promoters of up-regulated genes have clear preference for complex binding close to the TSS, and that CpG island promoters respond to regulation from larger distances, most often from the first intron.

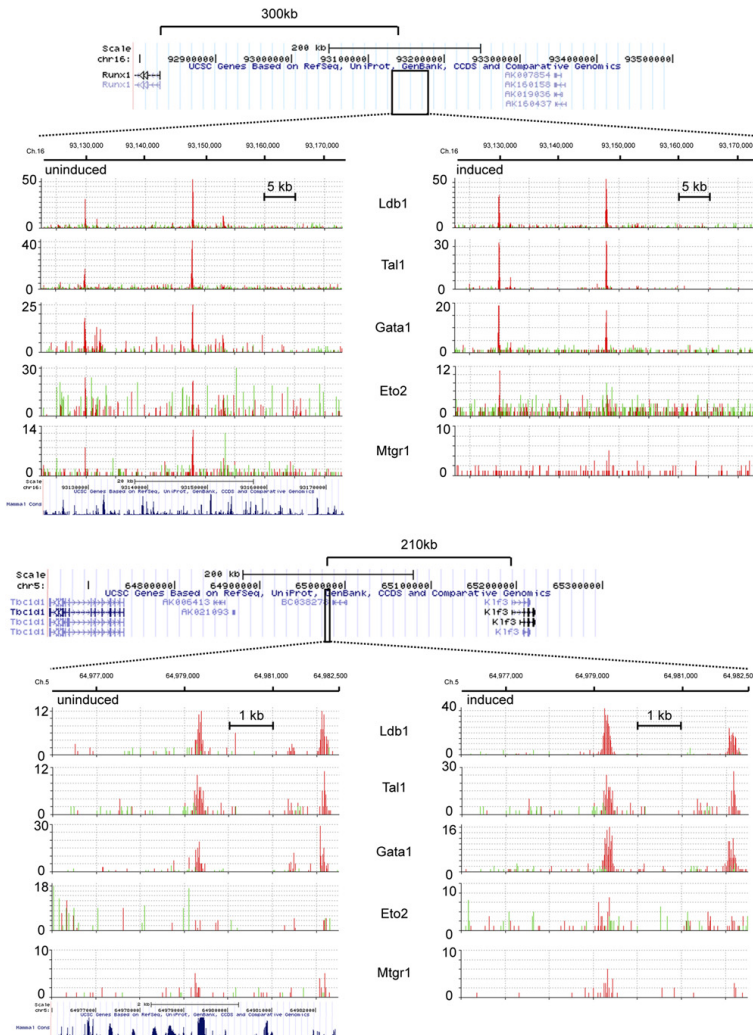


genes often show more than one strong binding site for the Ldb1 complex (Supplemental Fig. 8). The significance of these multiple sites is unknown at present.

*Binding sites in large intergenic regions*

In the plots shown in Figure 5, the binding sites are within a relatively short distance from the genes, and the relevant gene can be identified easily. However, there is a class of binding sites that are located in areas that are totally devoid of genes for hundreds of kilobases. Figure 6

shows two such cases. The first is a binding site that is present in an area of chromosome 16 without any (known) genes within a few hundred kilobases. The closest gene, AK007854, is >150 kb away, but this is unlikely to be a target gene, as we could not detect up-regulation upon differentiation (data not shown). The closest gene that is up-regulated in the differentiated cells is *Runx1* (log fold change 0.77;  $P = 0.005$ ) situated 300 kb away from the binding sites. Interestingly, these binding sites are the only sites within 250 kb of DNA. Both sites have never



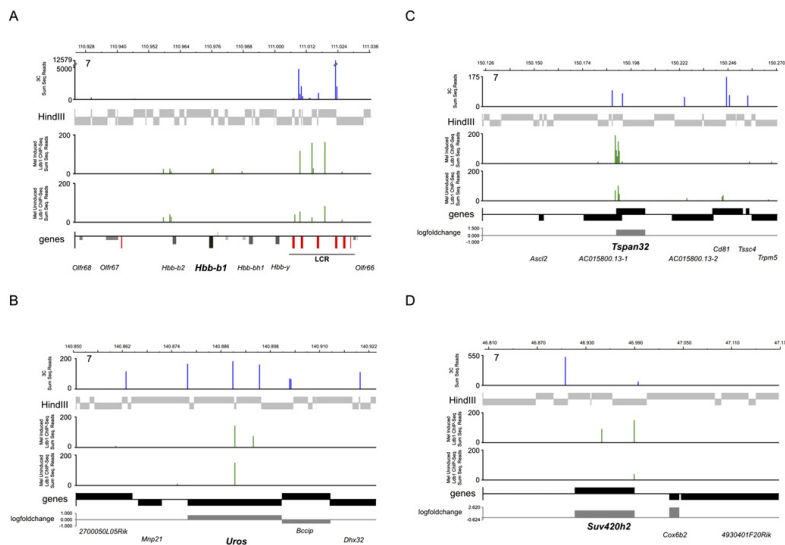
**Figure 6.** Binding sites in large intergenic regions. The top panel shows two binding sites at ~300 kb upstream of the *Runx1* gene in a gene-poor region. The bottom panel shows two binding sites at ~210 kb from the *Klf3* gene. The binding profiles of the different transcription factors in uninduced and induced MEL cells are shown.



been mapped before as regulatory sites of the *Runx1* gene, and both show a relative decrease of *Eto2* and *Mtgr1* consistent with up-regulation. The second example is similar: The site is several hundred kilobases away from the closest up-regulated gene, *Klf3*, whereas a number of other genes are much closer to the binding site. *Klf3* is an antagonist of *Klf1* activated by the same *Klf1*, a late erythroid activator that is required for the activation of many erythroid genes during the last steps of differentiation. *Klf3* is a developmental regulatory gene that is itself regulated by long-range interactions in what is known as a genomic regulatory block (GRB) (Kikuta et al. 2007). GRB target genes are often flanked by large intergenic regions spanned by highly conserved and other noncoding elements that act as long-range enhancers (Sandelin et al. 2004; Pennacchio et al. 2006; Kikuta et al. 2007). Other large intergenic regions with *Ldb1* complex-binding sites are found around genes such as *Ets2*, *Max*, *Mef2c*, or *Pim1*, all known to be involved in the regulation of hematopoiesis. These and similar data for many other genes (data not shown) suggest that the *Ldb1* complex is involved in long-range interactions. We therefore tested this possibility for chromosome 7 (see below), which has been analyzed previously by Chromosome Conformation Capture-on-Chip (4C) (Simonis et al. 2006).

#### Interactions between regulatory regions

The *Ldb1* complex is thought to be important for long-range interactions. It was clear from the ChIP-seq data that most binding sites of the *Ldb1* complex were with up-regulated genes (see above), and that the data identified a number of regions that are thought to interact with each other. The best-documented example is the  $\beta$ -globin locus. The hypersensitive sites 1–4 (HS1–4) of the LCR of the  $\beta$ -globin locus all bind the *Ldb1* complex before differentiation (Fig. 7), and the level of binding to HS2–4 increases upon differentiation when the  $\beta$ -globin gene is expressed at high levels. We therefore carried out a Chromosome Conformation Capture sequencing (3C-seq) experiment focusing on chromosome 7, using the  $\beta$ -globin gene ( $\beta$ -major promoter) as the viewpoint to determine whether binding sites of the *Ldb1* complex would identify long-range interactions between genes and regulatory regions. In this experiment, 12.5-dpc fetal liver and fetal brain tissues were used, and interacting sequences were identified as *HindIII* fragments using Illumina paired-end sequencing. Peaks on a genomic location were considered to be fetal liver-specific when the 3C-seq read counts in the fetal brain samples were <10% of the sequence reads in the fetal liver samples. The interacting sequences were



**Figure 7.** Long-range interactions between *ldb1* complex-bound regulatory regions. A selection of long-range interactions between the *Ldb1* complex-bound  $\beta$ -major promoter and other *ldb1*-bound regulatory regions identified in 12.5-dpc fetal livers. Coordinates in megabases (Mb) on chromosome 7 are indicated at the top of each panel. The sum of the 3C-seq reads are indicated in blue. The *HindIII* fragments analyzed by 3C-seq are indicated by light-gray bars. The sum of the ChIP-seq reads for *Ldb1* are indicated in green; induced C88 MEL cells are on top, and uninduced C88 MEL cells are below. Genes are indicated by black bars, while the log fold change in gene expression between uninduced and induced C88 MEL cells is indicated in gray. (A) Long-range interactions between the  $\beta$ -major promoter and the  $\beta$ -globin LCR (mainly HS2 and HS4). The red bars indicate the approximate position of hypersensitive sites. (B,C) 3C-seq identifies the known long-range interactions between the  $\beta$ -major promoter and the *Uros* gene (B) and *Kcnq1* region (more specifically, *Tspan32*) (C). Both genes are bound by the *Ldb1* complex. (D) A previously unknown long-range interaction between the  $\beta$ -major promoter and the *Suv420h2* gene is identified, which correlates with *Ldb1* binding.

subsequently compared with the Ldb1-binding sites before and after differentiation. The results coincide with detection of the *ldb1* complex at the promoter of the  $\beta$ -globin gene (Fig. 7A) and the presence of an interaction signal between the LCR and the promoter, in line with previous data (Tolhuis et al. 2002; Palstra et al. 2003; Song et al. 2007). This suggests that the Ldb1 complex forms loops within larger loops that are mediated by other factors, such as CTCF/cohesin (Splinter et al. 2006; S Krpic, M Eijpe, and F Grosveld, unpubl.). Other interactions of the  $\beta$ -major promoter with up-regulated genes are also visible; for example, the known interaction with *Uros* (Osborne et al. 2004; Simonis et al. 2006). It shows one binding site of the Ldb1 complex before differentiation and an additional site after differentiation that are matched by interactions seen by 3C-seq (Fig. 7B). Interestingly, the resolution of these interactions is very defined as opposed to a broad peak of interactions by regular 4C (Simonis et al. 2006). For example, the known interaction in the region of the *Kcnq1* gene maps specifically to the *Tspan32* gene. It shows binding before differentiation that increases after differentiation and is matched by interactions on the same HindIII fragment (Fig. 7C). New interactions are also detected. For example, the *Suv420h2* gene shows a specific interaction peak that coincides with its binding site for the Ldb1 complex. Interestingly, in addition to *Suv420h2*, its neighboring gene, *Cox6b2*, is also highly induced upon differentiation (Fig. 7D).

## Discussion

### ChIP-seq

Here we determined and analyzed the genome-wide binding sites of the Ldb1 complex (Wadman et al. 1997; Meier et al. 2006) before and after differentiation of MEL cells—a model system for the differentiation of proerythroblast cells to fully differentiated hemoglobinised cells. The binding sites were matched with the gene expression profiles before and after differentiation to determine the target genes of the complex during differentiation. The binding sites were obtained by using a tagging approach, with the biotin tag generally yielding the best results. The precipitate can be washed more stringently due to the high-affinity binding, resulting in lower background signals. The use of antibodies directed against a particular transcription factor usually results in a much lower signal-to-background ratio, and such antibodies were used only to confirm that the nontagged endogenous factor binds the same sequence as found with the tagged factor (e.g., see Fig. 3). The resolution and sequence coverage of the experiment enables practically certain identification of thousands of binding sites (Supplemental Figs. 2B, 3), and further increases in sequencing power and the development of more advanced error models and processing algorithms are likely to push the confidence limits much further very soon. The very high mapping scores at a small number of genomic locations are easily filtered out using the control experiments. They can be recognized (or

filtered out) in the Solex program or the nascent Bioconductor (<http://www.bioconductor.org>) pipeline as perfect blocks of sequences rather than a distribution of sequences around a recognition site. It is also apparent from the data that it is not easy to standardize the sequence results and compare them across the different factors. Clearly, a factor like Eto2 that does not bind DNA directly but is only bound to a DNA-binding factor via bridging proteins such as Ldb1 is less efficiently cross-linked to DNA with formaldehyde, and reproducibly has a lower signal-to-background ratio than a factor like Gata1 or Tall that binds DNA directly and is cross-linked more efficiently. Thus, greater sequencing depth is required to obtain the same level of confidence for Eto2 than is needed for Gata1 or Tall. Finally, it should be noted that none of the methods—i.e., tagging the 5' or 3' end or using antibodies directly—will guarantee detecting all of the binding sites. All may bring about structural changes in the complex or not be freely accessible at all sites.

### Ldb1 complexes and associated binding sites

A number of different complexes have been identified for Gata1 (Rodriguez et al. 2005), and a number of subcomplexes have been identified for Ldb1 in erythroid cells (Meier et al. 2006). Both of these observations made by transcription factor pull-downs and mass spectrometry are confirmed by the ChIP-seq data (e.g., see Figs. 2, 5). Most importantly, the Ldb1 complex changes during differentiation; in particular, the relative level of the negative regulators Eto2/Mtgr1 is decreased (Meier et al. 2006). This is confirmed by the decrease of Eto2 binding in the ChIP-seq (Fig. 1; Supplemental Fig. 2), and suggests that the full complex containing Eto2/Mtgr1 is bound to genes that are poised to be expressed during the last steps of differentiation. This is indeed observed when the binding site analysis is combined with the gene expression analysis, which shows that the vast majority of the genes are up-regulated by the Ldb1 complex, many of which are related to heme synthesis and red cell membrane structure. At present, it is not clear how the complex activates gene expression, but it is very likely that at least part of the effect is mediated through its binding of Cdk9 before differentiation and is largely lost during differentiation (Meier et al. 2006). A recent study has shown that Cdk9 binds the entire mediator complex (C Bezstarosty, A Ghamari, F Grosveld, and J Demmers, in prep.), which may be an intermediate step before Cdk9 phosphorylates the Ser2 residue of the C-terminal domain (CTD) tail of RNA polymerase II, and thus allows transcriptional elongation through the gene. Thus, one of the main functions of the Ldb1 complex may be the "delivery" of Cdk9 to the basic transcription machinery, which would fit well with the results described for the function of the LCR in the  $\beta$ -globin locus (Sawado et al. 2003; see below).

The peaks of the binding activities directly identify the DNA-binding motif, which in the case of the full complex is near enough to the published sequence for the Gata1-binding motif, but very different for the E-box motif bound

by Tal1. It shows a specific motif, (C)TGN<sub>7-8</sub>WGATAR, or a C(A/T)GCTG motif different from the published E-box sequence (CANNTG). In contrast, Gata1 also binds many negatively regulated genes, and these do not show the E-box motif. It is also worth noting that many binding sites occur as multiple binding sites (Supplemental Figs. 6, 7); presumably, this increases the efficiency of binding the complex, although it is equally possible that this allows interactions with more than one distant regulatory site at the same time (see below).

The analysis of the position of the binding sites is even more interesting. First, the complex binds most often downstream from the TSS, and is very frequently found in the first intron of a gene (Fig. 5). Second, there is a clear difference in the binding pattern of the Ldb1 complex to CpG-rich promoters, as found in the majority of genes and non-CpG promoters that encompass TATA-box promoters. In CpG promoters, most commonly the binding sites are found between 1 and 3 kb downstream from the TSS, whereas in the non-CpG and/or TATA-box-containing promoters, most binding sites are found in the promoter, with the exception of the Gata1-binding sites, which show a second peak downstream from the TSS. The latter belong to other Gata1 complexes (E deBoer, C Andrieu-Soler, E Soler, JC Bryne, S Thongjuea, M Stevens, C Kockx, Z Ozgur, W van IJcken, B Lenhard et al., in prep.). Since it has been reported that transcription of TATA-less genes is independent of Cdk9 (Montanuy et al. 2008), we compared the sensitivity of each category of promoters to Cdk9 inhibition by 5,6-dichloro-1-b-d-ribofuranosylbenzimidazole (DRB). This showed both CG-rich and non-CG-rich promoters to be sensitive to DRB, and only a modest difference in the average response of the CG (39% of normal) and non-CG (30% of normal) promoters was observed (data not shown). Similarly, when TATA-box-containing promoters versus non-TATA-box promoters were compared, the values were 40% and 34%, respectively. We conclude that both types of promoters are dependent on cdk9, and that the dependency is very similar for the different type of promoters.

#### *The Ldb1 complex and long-range interactions*

Ldb1 was originally described as a factor as important for long-range interactions in *Drosophila* as Chip (Morcillo et al. 1997), a protein that can interact with the insulator protein Su(Hw) (Torigoi et al. 2000). It was subsequently shown to be involved in at least one long-range interaction in erythroid cells—namely, between the LCR and the  $\beta$ -globin gene (Song et al. 2007)—by showing that Ldb1 can be detected at the  $\beta$ -globin promoter after differentiation in the absence of a binding site of the Ldb1 complex in the promoter of the  $\beta$ -globin gene, although it should be noted that, in the human  $\beta$ -globin cluster, loops are still formed between the LCR and the  $\beta$ -globin gene in the absence of the promoter, suggesting that additional sequences and interactions are important (Patrinos et al. 2004). The Chip-seq data show that binding sites of the Ldb1 complex are often at long distances from a gene that it (possibly) up-regulates (Fig. 6). When the ChIP-seq data

are combined with 3C-seq data focusing on chromosome 7 that contains the  $\beta$ -globin locus, it is evident that the Ldb1 complex binds to the HS2,3 and HS4 of the LCR; this binding appears to increase upon differentiation, and is accompanied by the appearance of binding of the Ldb1 complex at the  $\beta$ -globin gene (Hbb-b1) (Fig. 7). This confirms the published data (Song et al. 2007), and suggests, in combination with our previous data on the interactions within the locus (Palstra et al. 2003), that large loops are formed independently of the Ldb1 complex before differentiation that does not involve part of the LCR (HS2 and HS3), but that loops are formed within these loops dependent on the Ldb1 complex. It is presently not clear which biochemical property of the complex is involved in this process—most likely at least its dimerization domain—but it is less clear why this would only be the case after differentiation, and which change in the complex (or another interacting molecule[s]) is responsible for loop formation. Interestingly, further analysis of a number of known and novel interactions on chromosome 7 using the  $\beta$ -globin gene as the viewpoint suggest that the Ldb1 complex may be important for the interactions taking place between different loci on the same chromosome. Although it does not constitute definitive proof, we therefore analyzed whether the number of Ldb1-binding sites in HindIII fragments interacting with the  $\beta$ -globin promoter is higher than would be expected on a random basis. There are 910 positive HindIII fragments on chromosome 7, covering 3.6 Mb (2.36%) of the total chromosome length of 153 Mb. Assuming that Ldb1 binds randomly, this would mean that a random location on chromosome 7 has a 2.36% chance of being in a HindIII fragment that interacts with the  $\beta$ -globin promoter. There are 272 significant Ldb1-binding sites on chromosome 7. Any random 272 locations would have 6.42 locations that overlap with a HindIII fragment interacting with the  $\beta$ -globin promoter ( $2.36\% \times 272$ ). We observe 31 such Ldb1 sites and 241 that do not interact. The probability of finding 31 versus the randomly expected 6.42 is  $5.3 \times e^{-13}$  (Poisson distribution), and hence is highly significant.

At present, it is not clear what the significance of these interactions is within a living cell, as they are of a completely different scale than the interactions seen within a locus and take place at a much lower frequency, as demonstrated by the different scales in Figure 7. Clearly, the interactions within a locus are of crucial importance for the proper activation of genes, and it is clear that the Ldb1 complex plays a primary role in this interaction process.

#### **Materials and methods**

##### *Plasmid construction*

Expression vectors for bacterial biotin ligase BirA and bio tagging were as described (de Boer et al. 2003). Bio-tagged Ldb1 was as described (Meier et al. 2006). V5-Bio or V5-Flag-Bio constructs (Kolodziej et al. 2009; Soler et al. 2009) were used to tag Eto2, Mtrg1, and Gata1 cDNA at the C-terminal end of the protein. All constructs were verified by DNA sequencing and transfected into MEL cells to obtain stable clones expressing the tagged protein. Only clones expressing the tagged protein at low levels (<50%

increase in expression) were used for analysis, and all of these showed the normal induction of erythroid terminal differentiation.

#### Cell culture

MEL cell lines were maintained in DMEM supplemented with 10% fetal calf serum and penicillin/streptomycin. MEL cells in the log phase of growth were induced to differentiate with 2% dimethylsulfoxide (DMSO) for 4 d.

#### ChIP and ChIP-seq assays

Preparation of cross-linked chromatin for ChIP and ChIP-seq ( $1 \times 10^7$  cells per ChIP;  $1 \times 10^8$  cells per ChIP-seq), sonication to 200–800-base-pair (bp) fragments, immunoprecipitations, and DNA purification were as described (Rodriguez et al. 2005; Kolodziej et al. 2009; Soler et al. 2009). Detailed materials and methods are described in the Supplemental Material.

#### Microarray and statistical analysis

The GeneChip Mouse Gene 1.0 ST array oligonucleotide microarray (Affymetrix) was used. Detailed materials and methods and statistical analysis are described in the Supplemental Material.

#### 3C sequencing

The 3C library was prepared as described previously (Simonis et al. 2006). The library was sequenced using the Illumina paired-end protocol, and data were visualized using Signal map (Nimblegen) (see the Supplemental Material).

#### Acknowledgments

We are grateful to Sjaak Philipsen and Wouter de Laat for their advice, to Xianjun Dong for creating Supplemental Figure 8, to Marieke van Lindern and Rastislav Horos for their help with primary erythroid cell culture and lentiviral transductions, and to Zeliha Ozgur for microarray processing. We thank Mirjam van den Hout for Illumina GAP analyses. This work was supported by the NIH and the Cells into Organs and EuTRACC consortium. E.S and C.A.-S. were supported by Marie Curie fellowships. E.R. (HWJR) was supported by NBIC (NL), J.C.B. and S.T. were supported by EuTRACC, and B.L. was supported by the grants from Norwegian Research Council (YFF) and Bergen Research Foundation. R.J.P. is supported by Netherlands Organisation for Scientific Research (NWO) grant number 700.57.408.

#### References

Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B. 2009. Transcriptional features of genomic regulatory blocks. *Genome Biol* **10**: R38. doi: 10.1186/gb-2009-10-4-r38.

Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WC, Higgs DR. 2004. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J* **23**: 2841–2852.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208. doi: 10.1093/nar/gkp335.

Brand M, Ranish JA, Kummer NT, Hamilton J, Igarashi K, Francastel C, Chi TH, Crabtree GR, Aebersold R, Groudine M. 2004. Dynamic changes in transcription factor complexes

during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11**: 73–80.

Breen JJ, Agulnick AD, Westphal H, Dawid IB. 1998. Interactions between LIM domains and the LIM domain-binding protein Ldb1. *J Biol Chem* **273**: 4712–4717.

Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. 2008. JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. *Nucleic Acids Res* **36**: D102–D106. doi: 10.1093/nar/gkm955.

Cantor AB, Orkin SH. 2001. Hematopoietic development: A balancing act. *Curr Opin Genet Dev* **11**: 513–519.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Sempere CA, Taylor MS, Engstrom PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–635.

Cronan JE Jr, Reed KE. 2000. Biotinylation of proteins in vivo: A useful posttranslational modification for protein analysis. *Methods Enzymol* **326**: 440–458.

Davis JN, McGhee L, Meyers S. 2003. The ETO (MTG8) gene family. *Gene* **303**: 1–10.

de Boer E, Rodriguez P, Bonte E, Krijgsveld J, Katsantoni E, Heck A, Grosveld F, Strouboulis J. 2003. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci* **100**: 7480–7485.

Engstrom PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B. 2007. Genomic regulatory blocks underlie extensive micro-synteny conservation in insects. *Genome Res* **17**: 1898–1908.

Grutz G, Forster A, Rabbitts TH. 1998. Identification of the LMO4 gene encoding an interaction partner of the LIM-binding protein LDB1/NLI1: A candidate for displacement by LMO proteins in T cell acute leukaemia. *Oncogene* **17**: 2799–2803.

Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J* **24**: 2367–2378.

Jurata LW, Gill GN. 1997. Functional analysis of the nuclear LIM domain interactor NLI. *Mol Cell Biol* **17**: 5688–5698.

Kenny DA, Jurata LW, Saga Y, Gill GN. 1998. Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. *Proc Natl Acad Sci* **95**: 11257–11262.

Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, et al. 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* **17**: 545–555.

Kolodziej KE, Pourfarzad F, de Boer E, Krpic S, Grosveld F, Strouboulis J. 2009. Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol Biol* **10**: 6.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**: e1000242. doi: 10.1371/journal.pgen.1000242.

Lahlil R, Lecuyer E, Herblot S, Hoang T. 2004. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol* **24**: 1439–1452.

Matthews JM, Visvader JE. 2003. LIM-domain-binding protein 1: A multifunctional cofactor that interacts with diverse proteins. *EMBO Rep* **4**: 1132–1137.

Megraw M, Pereira F, Jensen ST, Ohler U, Hatzigeorgiou AG. 2009. A transcription factor affinity-based code for mammalian transcription initiation. *Genome Res* **19**: 644–656.

- Meier N, Krpic S, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F. 2006. Novel binding partners of Ldb1 are required for haematopoietic development. *Development* **133**: 4913–4923.
- Montanuy I, Torremocha R, Hernandez-Munain C, Sune C. 2008. Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9. *J Biol Chem* **283**: 7368–7378.
- Morcillo P, Rosen C, Baylies MK, Dorsett D. 1997. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. *Genes & Dev* **11**: 2729–2740.
- Mukhopadhyay M, Teufel A, Yamashita T, Agulnick AD, Chen L, Downs KM, Schindler A, Grinberg A, Huang SP, Dorward D, et al. 2003. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**: 495–505.
- Murre C, Voronova A, Baltimore D. 1991. B-cell- and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. *Mol Cell Biol* **11**: 1156–1160.
- Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36**: 1065–1071.
- Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. 2003. The  $\beta$ -globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**: 190–194.
- Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, Grosveld FG. 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes & Dev* **18**: 1495–1509.
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499–502.
- Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin S. 1996. The T cell leukemia oncoprotein SCL/Tal1 is essential for development of all hematopoietic lineages. *Cell* **86**: 47–57.
- Robb L, Lyons I, Li R, Hartley L, Kontgen F, Harvey RP, Metcalf D, Begley C. 1995. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci* **92**: 7075–7079.
- Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J. 2005. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **24**: 2354–2366.
- Sandelin A, Bailey P, Bruce S, Engstrom PG, Klos JM, Wasserman WW, Ericson J, Lenhard B. 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**: 99. doi: 10.1186/1471-2164-5-99.
- Sawado T, Halow J, Bender MA, Groudine M. 2003. The  $\beta$ -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes & Dev* **17**: 1009–1018.
- Schuh AH, Tipping AJ, Clark AJ, Hamlett J, Guyot B, Iborra FJ, Rodriguez P, Strouboulis J, Enver T, Vyas P, et al. 2005. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol* **25**: 10235–10250.
- Shivdasani RA, Mayer EL, Orkin SH. 1995. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**: 432–434.
- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348–1354.
- Soler E, Andrieu-Soler C, de Boer E, Rijkers HWJ, Bryne JC, Thongjuea S, Demmers J, van IJcken W, Grosveld F. 2009. A systems approach to analyze transcription factors in mammalian cells. *Methods* (in press).
- Song SH, Hou C, Dean A. 2007. A positive role for NLI/Ldb1 in long-range  $\beta$ -globin locus control region function. *Mol Cell* **28**: 810–822.
- Southern JA, Young DF, Heaney F, Baumgartner WK, Randall RE. 1991. Identification of an epitope on the P and V proteins of simian virus 5 that distinguishes between two isolates with different biological characteristics. *J Gen Virol* **72**: 1551–1557.
- Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the  $\beta$ -globin locus. *Genes & Dev* **20**: 2349–2354.
- Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active  $\beta$ -globin locus. *Mol Cell* **10**: 1453–1465.
- Torigoi E, Bennani-Baiti IM, Rosen C, Gonzalez K, Morcillo P, Ptashne M, Dorsett D. 2000. Chip interacts with diverse homeodomain proteins and potentiates bicoid activity in vivo. *Proc Natl Acad Sci* **97**: 2686–2691.
- Tripic T, Deng W, Cheng Y, Zhang Y, Vakoc CR, Gregory GD, Hardison RC, Blobel GA. 2009. SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113**: 2191–2201.
- Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA1 and Ldb1/NLI proteins. *EMBO J* **16**: 3145–3157.
- Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts T. 1994. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* **78**: 45–57.
- Wilson NK, Miranda-Saavedra D, Kinston S, Bonadies N, Foster SD, Calero-Nieto F, Dawson MA, Donaldson RJ, Dumon S, Frampton J, et al. 2009. The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood* **113**: 5456–5465.
- Xu Z, Meng X, Cai Y, Liang H, Nagarajan L, Brandt SJ. 2007. Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins. *Genes & Dev* **21**: 942–955.

**Erratum**

**Genes & Development 24:** 277–289 (2010)

**The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation**

Eric Soler, Charlotte Andrieu-Soler, Ernie de Boer, Jan Christian Bryne, Supat Thongjuea, Ralph Stadhouders, Robert-Jan Palstra, Mary Stevens, Christel Kockx, Wilfred van IJcken, Jun Hou, Christine Steinhoff, Erikjan Rijkers, Boris Lenhard, and Frank Grosveld

Due to an editing error, an incorrect affiliation was given for Boris Lenhard. The correct affiliation is as follows:

Computational Biology Unit-Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway

We apologize for any confusion caused by this error.

Please note that this has been corrected in the online version of this paper.