Control of developmentally primed erythroid genes by combinatorial co-repressor actions

Ralph Stadhouders1,∗, Alba Cico2,∗, Tharshana Stephen2, Supat Thongjuea3,4, Petros Kolovos1, H. Irem Baymaz1, Xiao Yu1, Jeroen Demmers5, Karel Bezsarostí5, Alex Maas1, Vilma Barroca6, Christel Kockx7, Zeliha Ozgur7, Wilfred van IJcken7, Marie-Laure Arcangeli8, Charlotte Andrieu-Soler2, Boris Lenhard9, Frank Grosveld1,10,** & Eric Soler1,2,10,11,**

How transcription factors (TFs) cooperate within large protein complexes to allow rapid modulation of gene expression during development is still largely unknown. Here we show that the key haematopoietic LIM-domain-binding protein-1 (LDB1) TF complex contains several activator and repressor components that together maintain an erythroid-specific gene expression programme primed for rapid activation until differentiation is induced. A combination of proteomics, functional genomics and in vivo studies presented here identifies known and novel co-repressors, most notably the ETO2 and IRF2BP2 proteins, involved in maintaining this primed state. The ETO2–IRF2BP2 axis, interacting with the NCOR1/SMRT co-repressor complex, suppresses the expression of the vast majority of archetypical erythroid genes and pathways until its decommissioning at the onset of terminal erythroid differentiation. Our experiments demonstrate that multimeric regulatory complexes feature a dynamic interplay between activating and repressing components that determines lineage-specific gene expression and cellular differentiation.

DOI: 10.1038/ncomms9893

1 Department of Cell Biology, Erasmus Medical Center, 3015CN Rotterdam, The Netherlands. 2 Inserm UMR967, CEA/DSV/IRCM, Laboratory of Molecular Hematopoiesis, Université Paris-Saclay, 92265 Fontenay-aux-Roses, France. 3 Computational Biology Unit, Bergen Center for Computational Science, N-5008 Bergen, Norway. 4 MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. 5 Department of Proteomics, Erasmus Medical Center, 3015CN Rotterdam, The Netherlands. 6 CEA/DSV/IRCM/SCSR, Université Paris-Saclay, 92265 Fontenay-aux-Roses, France. 7 Center for Biomics, Erasmus Medical Center, 3015CN Rotterdam, The Netherlands. 8 Inserm UMR967, CEA/DSV/IRCM, Laboratory of Hematopoietic and Leukemic Stem cells, Université Paris-Saclay, 92265 Fontenay-aux-Roses, France. 9 Department of Molecular Sciences, Faculty of Medicine, MRC Clinical Sciences Centre, Institute of Clinical Sciences, Imperial College London, London W12 ONN, UK. 10 Cancer Genomics Center, Erasmus Medical Center, 3015CN Rotterdam, The Netherlands. 11 Laboratory of Excellence GR-Ex, 75015 Paris, France. ∗ These authors contributed equally to this work. ** These authors jointly supervised this work. † Present address: Centre for Genomic Regulation (CRG), Gene regulation, Stem cells and Cancer programme, 08003 Barcelona, Spain. Correspondence and requests for materials should be addressed to R.S. (email: Ralph.Stadhouders@crg.eu) or to E.S. (email: eric.soler@cea.fr).
Hematopoietic development relies on the stepwise activation and repression of lineage-specific gene expression programmes. This process is regulated by sets of conserved transcription factors (TFs) acting in a combinatorial and/or antagonistic pattern to establish cellular identity through tight control of gene regulatory networks. Exactly how TFs and the cofactors they recruit cooperate within large protein complexes to rapidly modulate gene expression during differentiation is still not completely understood. We set out to address this issue using a well-characterized erythroid differentiation system driven by a multimeric TF complex nucleated by the haematopoietic master regulators LIM-domain-binding protein 1 (LDB1), GATA-binding protein 1 (GATA1), T-cell acute lymphocytic leukaemia protein 1 (TAL1), LIM domain-only 2 and eight-twenty-one 2 (ETO2)—hereafter referred to as the LDB1 complex. The LDB1 complex plays a pivotal role in promoting differentiation of the erythroid and megakaryocytic lineages. It was previously shown to bind the regulatory regions of developmentally regulated erythroid genes, which are rapidly induced by the LDB1 complex upon terminal erythroid differentiation. Despite being already present by the LDB1 complex in immature progenitors, premature full activation of these erythroid genes is prevented by the LDB1-complex member ETO2 (also referred to as the myeloid-transforming gene on chromosome 16 or MTG16), a transcriptional co-repressor. ETO2 belongs to a family of transcriptional repressors known as the ETO family, which further consists of the founder member ETO (or MTG8) and the myeloid translocation gene, related-1 (MTGR1) proteins. ETO2 plays key roles in the maintenance of haematopoietic stem cells, the development of the lymphoid system and regulating erythroid differentiation. The vast majority (88%) is already bound by an ETO2-containing LDB1 complex in progenitors (see ref. 5). Analysis of glioma-wide RNA-Seq data indicates that >95% of the primed genes show no overt signs of paused RNAPII accumulating at the promoter before induction, ruling out RNAPII pause-release as a general mechanism of primed gene activation (Supplementary Fig. 1b). In accordance with low-level expression and strong TF binding, the active H3K4Me2 chromatin mark is already deposited at the regulatory elements of primed erythroid genes in progenitors (Supplementary Fig. 1c). Previously published ChIP-Seq data obtained from primary erythroid progenitors further confirms this finding (illustrated in Supplementary Fig. 6e).

We conclude that primed archetypical erythroid genes in progenitor cells are uniformly defined by: (1) low expression levels that are strongly induced upon differentiation; (2) occupancy of the TFs responsible for their later induction; (3) an active chromatin environment at the regulatory elements controlling their expression and (4) not having substantial amounts of paused RNAPII at their promoters. Thus, primed erythroid genes are subjected to active transcriptional repression involving ETO2 that prevents their premature activation.

Identification of ETO2 protein partners in erythroid cells. We next employed a proteomics approach to characterize the molecular determinants of ETO2’s repressive activity. An epitope-tagged form of ETO2 (ETO2-V5-Bio) was stably expressed in MEL cells and used in single-step protein complex capture experiments. The affinity tag contains a Bio peptide sequence that is efficiently biotinylated by the bacterial BirA enzyme, resulting in the biotinylation of ETO2-V5-Bio (Fig. 1a—full-size images of all western blots shown can be found in Supplementary Figs 10 and 11). The C-terminal tag fused to ETO2 does not interfere with its functions, since ETO2-V5-Bio shows (i) proper intracellular localization (Supplementary Fig. 2a), (ii) the ability to interact with its known binding partner LDB1 (ref. 7) (Fig. 1b) and (iii) binding to known genomic target sites (Supplementary Fig. 2c). Thus, tag addition does not affect ETO2 in its ability to form complexes. A streptavidin pull-down was carried out and co-purified proteins were identified by mass spectrometry (liquid chromatography–tandem mass spectrometry (LC–MS)/MS) (Fig. 1c). In addition to...
known components of the LDB1 complex (for example, TAL1, the E proteins E2A and HEB, single-stranded DNA-binding protein (SSBP) 2/3/4)\(^7\), we also detect additional interactions with the LSD1/Co-REST repressor complex, the haematopoietic TF GFI1B and the transcriptional repressor IRF2BP2. MS analysis of endogenously precipitated ETO2 protein complexes (immunoprecipitation (IP)–MS) and individual co-IP experiments in MEL cells confirm the endogenous interaction of ETO2 with these factors (Fig. 1d; Supplementary Fig. 2). Whereas the ability of ETO2 to interact with GFI1B was reported previously\(^3\), and the LSD1 protein was found to be associated with the LDB1 complex (including ETO2) in erythroid cells\(^2\), the involvement of IRF2BP2 in these complexes has not been reported yet. We therefore set out to investigate this interaction in more detail.

**ETO2 interacts with IRF2BP2 via a unique N-terminal domain.**

IRF2BP2 is a highly conserved zinc-finger/RING-finger protein belonging to a family of three evolutionary conserved factors (IRF2BP1, IRF2BP2 and IRF2BPL) sharing high sequence homology. IRF2BP1 and IRF2BP2 were originally identified as interacting partners of IRF2, mediating its ability to repress in vitro reporter expression\(^2\). Recently, several other studies reported a repressive role for IRF2BP2 in complex with nuclear factor of activated T-cells 1 (NFAT1)\(^2\), p53 (ref. 25) or enhanced

---

**Figure 1 | Identification of ETO2-binding partners in erythroid progenitor cells.** (a) Schematic of the ETO2 protein, its 4 Nervy homology regions (NHR1-4) and the C-terminal V5-Bio tag (top). Fusion protein expression and proper tag function in MEL cells were validated by WB analysis. MEL cells expressing only the BirA enzyme were used as a control. (b) Efficient streptavidin IP of ETO2-V5-Bio in MEL cells. Interaction of ETO2-V5-Bio with LDB1 (a known binding partner) was used for validation. (c) ETO2-V5-Bio-interacting proteins identified by LC-MS/MS in MEL cells. Only proteins pulled down in two independent experiments and with low background scores are shown. (d) Co-IP validations of selected ETO2-V5-Bio-interacting proteins in MEL cells using an endogenous ETO2 antibody. Species-matched IgG was used to control for aspecific binding. Full-size images of all western blots shown can be found in Supplementary Fig. 10. Strept-HRP, streptavidin-HRP; Sup, supernatant; endog., endogenous; WB, western blot; IP, immunoprecipitation.
at puberty 1 (EAP1)\(^2\), although an activating role for IRF2BP2 in regulating \(\text{VEGFA}\) expression has been described as well\(^2\). To map the domains mediating the interaction between ETO2 and IRF2BP2, a series of deletion mutants was generated and used in co-IP experiments. ETO2 contains four highly conserved domains (neuronal homology regions 1–4) shared with the other members of the ETO family (ETO and MTGR1)\(^2\), but also two unique sequences at its N terminus not shared with ETO/MTGR1, which we termed Unique Sequence 1 and 2 (US1/2) (Fig. 2a). As shown in Fig. 2b, ETO2 interacts with IRF2BP2 via its US2 domain, suggesting that ETO2 is the only protein from the ETO family able to bind IRF2BP2. Using a similar strategy, we find that the IRF2BP2 RING-finger domain mediates the interaction with ETO2 (Fig. 2c). RING-finger domains are characteristic of E3 ubiquitin ligases catalysing the ubiquitination of target proteins, which often leads to protein degradation\(^2\). Since ETO2 interacts with the RING-finger domain of IRF2BP2, we tested whether ETO2 stability could be affected by this interaction. Increasing amounts of IRF2BP2 were co-expressed together with ETO2 in HEK 293T cells, and ETO2 protein levels were monitored by western blot analysis. As shown in Supplementary Fig. 3, even when expressed in large excess, IRF2BP2 does not significantly affect ETO2 protein levels under these conditions.

**IRF2BP2 enhances ETO2-mediated transcriptional repression.** The functional role of the ETO2–IRF2BP2 interaction was first investigated in vitro using luciferase reporter assays. ETO2 was fused to a Gal4 DNA-binding domain and co-expressed in HEK 293T cells together with a luciferase reporter plasmid containing

---

**Figure 2 | ETO2 and IRF2BP2 interact via their US2 and RING domains, respectively, to cooperatively repress reporter gene activity.** (a) Schematic of the ETO2 and IRF2BP2 proteins and known functional domains. First and last amino-acid positions of known functional domains are indicated by numbers. Highlighted are two unique N-terminal amino-acid sequences (US1 and US2) only present in ETO2. (b) ETO2 interaction domain mapping using a collection of Flag-tagged deletion mutants that were overexpressed in HEK 293T cells together with V5-IRF2BP2. Bands representing the ETO2 mutant proteins are marked by an asterisk. (c) An HA-tagged IRF2BP2 lacking the C-terminal RING-finger domain (HA-deltaRING) was used in co-IP experiments with Flag-ETO2. (d) Luciferase reporter assay to test repression of a Gal4-responsive promoter (coupled to a firefly luciferase gene) by ETO2 and its interacting partners IRF2BP2 and LSD1. Fusion to a Gal4 DNA-binding domain (Gal4-ETO2) was used to target ETO2 to the promoter. Different combinations of Gal4-ETO2 and IRF2BP2, deltaRING and LSD1 were co-transfected and firefly luciferase expression was measured after 48 h. Co-transfection with equal amounts of a Renilla luciferase expression plasmid was used for normalization. Bars represent average values of at least three independent transfection experiments; error bars denote s.d. Full-size images of all WBs shown can be found in Supplementary Fig. 11. WB, western blot; IP, immunoprecipitation.
Gal4-responsive elements. As previously reported, ETO2 induces a 20–30-fold repression of luciferase activity (Fig. 2d)35. Co-expression of IRF2BP2 further increases ETO2-mediated transcriptional repression in a dose-dependent manner. This effect is not seen when using a RING-finger deletion mutant of IRF2BP2 (IRF2BP2deltaRING) unable to interact with ETO2. Importantly, the ETO2-interacting partner LSD1 (Fig. 1c,d), a known transcriptional repressor, does not significantly enhance ETO2-mediated repression (Fig. 2d) despite its ability to interact with ETO2 in HEK cells (Supplementary Fig. 3b).

Genome-wide analysis of ETO2 and IRF2BP2 chromatin binding. We next performed ChIP-Seq experiments to determine whether IRF2BP2 is enriched at critical regulatory sites occupied by ETO2. In erythroid progenitors, IRF2BP2-binding sites occur at numerous cis-regulatory regions of late erythroid genes controlled by ETO2 and LDB1. For example, IRF2BP2 and ETO2 show co-occupancy on the Gypa, Slc22a4, Epb4.2, Als2 and Slc4a1 genes, as well as the α- and β-globin clusters (see Fig. 3a for examples). These genes are critical markers of mature erythroid cells and reside in a primed state in erythroid progenitors such as MEL cells5,6 (Supplementary Fig. 1). This suggests that IRF2BP2 might cooperate with ETO2 to maintain these erythroid genes in a primed state, suppressing the actions of the other LDB1-complex members required for their rapid activation upon terminal differentiation35,36. A genome-wide comparison of ETO2- and IRF2BP2-binding patterns revealed that 61% of ETO2-binding sites are also occupied by IRF2BP2 (Fig. 3b). However, many genomic locations are bound by IRF2BP2 in the absence of ETO2 and vice versa (Fig. 3c), indicating that both proteins are also involved in different regulatory complexes. Analysing peak distribution relative to transcription start sites (TSSs) (using Genomic Regions Enrichment of Annotations Tool (GREAT)31, see Methods) reveals that both common and ETO2-specific binding occurs predominantly (>80%) distal (>5 kb) from a TSS (Supplementary Fig. 4), in agreement with the binding distribution of the LDB1 complex5. However, IRF2BP2-only peaks show significantly more (42%) proximal promoter (<5 kb) of the TSS) binding. Interestingly, the LSD1 and GFI1B repressors are also found enriched at ETO2/IRF2BP2-binding sites, overlapping with the positioning of the LDB1 complex (Supplementary Fig. 5).

We tried to substantiate these observations on ETO2 and IRF2BP2 chromatin occupancy by performing a Gene Ontology term analysis on putative target genes assigned to the different binding site subsets using GREAT. This confirms a strong enrichment for erythroid functions among the common target genes (Fig. 3b). ETO2-specific target genes show some enrichment for common blood-cell-related functions, as well as for several housekeeping processes. Intriguingly, IRF2BP2-specific target genes show strong associations with biological processes and functions involved in survival, apoptosis and cancer (Fig. 3b).

A de novo DNA motif search performed on ETO2- and IRF2BP2-occupied genomic binding sites reveals an enrichment of several different TF-binding motifs. ETS (E-twenty-six, for example, the Friend leukaemia integration 1 (FLI1)/ETS-related gene (ERG) TFs) and G/C-rich (early growth response (EGR) family) TF motifs were detected in all three categories, while Runt-related TF (RUNX) and TCF motifs show enrichment at sites occupied only by IRF2BP2 (Fig. 3d). Strikingly, nearly all ETO2-only and ETO2–IRF2BP2 shared sites contain the typical LDB1-complex signature represented by a composite E-box/GATA motif (CTGN(6–8)WGATAR)33,35, while this motif is completely absent from the IRF2BP2-only-binding sites (Fig. 3d). This shows that in MEL cells, the IRF2BP2-only-binding sites are LDB1-complex independent. However, IRF2BP2-only sites did show some enrichment for GATA motifs, indicating that a small subset of these sites could be GATA1-targeted. Interestingly, IRF2BP2-only sites are specifically enriched for several other TF motifs, including CTCF-binding factor (CTCF), Specificity protein/Krüppel-like factor (SP/KLF) and leucine zipper (that is, activating protein-1 family) TF motifs.

Ir2bp2 expression during erythroid differentiation. It is well established that ETO2 expression levels diminish as erythroid progenitors undergo terminal differentiation32,34. Furthermore, in a G1E-ER model system of erythroid differentiation, expression of Cbfa2t3 (encoding ETO2) was repressed upon GATA1-driven erythroid maturation33,34,35. These and other observations34 suggest that Cbfa2t3 expression is regulated by the ETO2-containing LDB1 complex, which involves an ETO2 negative autoregulatory loop. To gain more insight into the regulation of Ir2bp2 during erythropoiesis, we examined its expression levels during mouse FL erythropoiesis. RNA-Seq analysis of fluorescence-activated cell sorting (FACS)-purified populations of developing erythroid cells (the same populations as shown in Fig. 7i,j, see the Methods section for more details) indicates that Ir2bp2 expression is reduced upon differentiation (Fig. 4a). These observations are further validated by quantitative PCR (qPCR) experiments (Supplementary Figs 8 and 9). A similar trend was observed by others using various in vivo and in vitro model systems for erythroid development35,36 (Fig. 4b,c). As was reported for Cbfa2t3, Ir2bp2 expression is lost upon GATA1-driven erythroid maturation in a G1E-ER model system (Fig. 4c).

In addition, using genome-wide data sets previously generated by our laboratory5 and the Encyclopedia Of DNA Elements (ENCODE) consortium36, we identify two putative enhancer elements within the Ir2bp2 locus bound by the ETO2/IRF2BP2-containing LDB1 complex (Fig. 4d–f). When G1E-ER cells are differentiated by translocation of GATA1 into the nucleus, the TAL1 activator5,32,33,37 is displaced from these putative regulatory elements (Fig. 4f), along with a loss of Ir2bp2 expression (Fig. 4c) and RNAIP1 occupancy of the locus (Fig. 4f). Collectively, these data indicate that during erythroid differentiation Ir2bp2 expression is repressed in a GATA1-dependent manner. We speculate that, similar to events observed at the Cbfa2t3 locus34, Ir2bp2 regulation involves negative auto-regulation by ETO2/IRF2BP2.
An IRF2BP2–ETO2 axis imposes transcriptional repression. We next tried to address the functional roles played by ETO2 and IRF2BP2 in erythroid cells. Short hairpin RNA (shRNA)-mediated knockdowns (KDs) of Cbfa2t3 and Irf2bp2 were performed in MEL cells, after which the expression of several ETO2-LDB1 target genes was measured. As shown in Fig. 5, depleting
ETO2 (Fig. 5a) or IRF2BP2 (Fig. 5b) results in increased *Alas2, Epb4.2, Gypa* and *Slc22a4* expression levels, establishing the repressive roles of ETO2 and IRF2BP2 in regulating primed archetypical erythroid genes. This result also corroborates that ETO2 and IRF2BP2 form a functional erythroid co-repressor complex. In marked contrast, when performing the same experiments for LSD1 (encoded by the *Kdm1a* gene, Fig. 5c), which co-occupies the same genes (Supplementary Fig. 5a), either very minor changes (*Alas2, Gypa, Slc22a4*) or decreased expression (*Epb4.2*) is observed. This result, together with the data derived from the reporter assays (Fig. 2d) suggest that LSD1 does not mediate transcriptional repression by ETO2 and might even play an opposite role. To more comprehensively identify genes controlled by ETO2 and IRF2BP2, transcriptome analyses were carried out by RNA-Seq after ETO2 and IRF2BP2 depletion in MEL cells. Differentially expressed genes were also compared with the ones obtained after LSD1 depletion. Strikingly, we observe a high degree of correlation when comparing genes significantly misregulated after ETO2 or IRF2BP2 KD (Fig. 5d,e), showing that genes controlled by ETO2 are also regulated by IRF2BP2, both in a positive and negative manner. Conversely, comparison of genes misregulated in both the *Cbfa2t3* (ETO2) and *Kdm1a* (LSD1) KD shows an inverse trend, as genes repressed by ETO2 are activated by LSD1 and vice versa (Fig. 5d,f). In addition, the KD of another ETO2-interacting repressor *Gfi1b* (encoding GFI1B), which is known to interact with both ETO2 and LSD1, results in a very similar profile of differentially expressed genes when compared with the *Cbfa2t3* and *Irf2bp2* KD results (Fig. 5d). This suggests that ETO2, IRF2BP2 and GFI1B negatively regulate a set of common genes and form a repressive complex in erythroid progenitor cells. Finally, we compared misregulated genes from...
Top 142 misregulated genes after Cbfa2t3 RNAi

MEL (ETO2-KD)

IRF2BP2 KD log2 FC

ETO2 KD log2 FC

LSD1 KD log2 FC

MEL (IRF2BP2-KD)

MEL (ETO2–/–)

MEL (IRF2BP2–/–)

MEL (WT)

MEL (ETO2–/–)

MEL (IRF2BP2–/–)
the ETO2-, IRF2BP2- and LSD1-depletion experiments to the gene expression changes obtained after MEL cell differentiation (Fig. 5g). The emerging correlations confirm the results presented in Fig. 5a–c: genes derepressed upon ETO2/IRF2BP2 depletion are upregulated during erythroid differentiation (including many of the primed terminal erythroid differentiation genes, Fig. 5g), while the opposite trend emerges for LSD1.

We next conducted rescue experiments in which wild-type Cbfa2t3/Ifn2bp2 or loss-of-interaction mutant complementary DNAs (cDNAs) were transfected into ETO2/IRF2BP2-depleted MEL cells (Fig. 5h). While the introduction of wild-type cDNA restores erythroid gene repression (Fig. 5h, orange bars), mutant cDNAs encoding ETO2/IRF2BP2 proteins no longer able to interact with each other (see Fig. 2) cannot or only partially induce transcriptional repression (Fig. 5h, grey bars). These experiments confirm that in erythroid progenitors ETO2 and IRF2BP2 function as transcriptional repressors in a highly cooperative manner. Moreover, IRF2BP2 ChIP experiments in an ETO2-deficient MEL cell line (generated using CRISPR/Cas9 technology38, see the Methods section) reveal a loss of chromatin binding in the absence of ETO2 (Fig. 5i). The reciprocal experiment—ETO2 ChIPs in an IRF2BP2−/− MEL cell line—shows that ETO2 binding to the genome does not strictly rely on IRF2BP2 (Fig. 5i). Of note, the core LDB1 complex remains bound at the examined regulatory sites in the absence of ETO2 or IRF2BP2—ruling out the possibility of a global disruption of LDB1-complex recruitment in these cells (Supplementary Fig. 5b).

IRF2BP2 and ETO2 repress essential erythroid pathways. To obtain functional insight into the genes affected in the Cbfa2t3 and Irf2bp2 KD experiments, we applied Ingenuity Pathway Analysis on the misregulated genes to link the transcriptional regulatory activities of ETO2 and IRF2BP2 to biological functions. In MEL cells, 2,625 genes are found differentially expressed upon IRF2BP2 depletion, and 724 upon ETO2 depletion. Combining these data sets, 58% of the ETO2 misregulated genes (420) are also found affected in the IRF2BP2 data set (Fig. 6a). Approximately 55% of the commonly misregulated genes are found upregulated and therefore appear to be repressed by ETO2/IRF2BP2. These 234 genes are highly enriched for erythroid functions (Fig. 6b).

In fact, 71% of the genes coding for the major components of the haem biosynthesis pathway are bound (as determined by GREAT analysis, see Methods) by ETO2 and IRF2BP2 (Fig. 6c; left graph). Furthermore, over 77% of the erythrocyte-specific membrane structural components and ion transporters are also targeted by the ETO2/IRF2BP2 complex (Fig. 6c; right graph).

In correspondence with this binding pattern, almost all of the above mentioned erythroid genes are misregulated upon ETO2 and/or IRF2BP2 depletion (100% of the haem biosynthesis genes and 78% of the erythrocyte membrane proteins are affected in at least one KD, see Fig. 6c), with a strong preference for derepression. In agreement with their co-occupancy by both proteins, many genes are upregulated after either Cbfa2t3 or Irf2bp2 KD (71% of haem biosynthesis genes and 36% of erythrocyte membrane protein genes, see Fig. 6c). In addition, α- and β-globin gene activation was observed in both KD experiments (Irf2bp2 KD— Hba-a1/2: 8.2-fold up, Hbb-b1/2: 4.0-fold up; Cbfa2t3 KD— Hba-a1/2: 3.9-fold up, Hbb-b1/2: 1.7-fold up). In general, 30–40% of the genes misregulated upon either Cbfa2t3 or Irf2bp2 KD are also bound by the corresponding factor (Fig. 6d), despite the inevitable presence of indirectly affected genes within our sets of putative targets. Strikingly, 70% of the genes bound and regulated by both ETO2 and IRF2BP2 in erythroid progenitors are found to be repressed, and these genes again exhibit a significant enrichment for erythroid functions (Fig. 6d). Together, these observations strongly indicate that the ETO2/IRF2BP2 complex controls the expression of key genes critical for erythroid cell identity and function.

ETO2 and IRF2BP2 also modulate a set of 186 genes that are downregulated upon factor depletion (Fig. 6b), of which 28% were also co-occupied. This suggests that ETO2/IRF2BP2-containing complexes also function in gene activation. Over-represented among these are genes known to play a role in blood cell activation, proliferation and cell death (Fig. 6b). Such pathways are known to be suppressed upon erythroid differentiation and might (in part) be activated by ETO2/IRF2BP2 in progenitor cells39. Surprisingly, a large fraction of the over-represented processes are related to leukocyte and lymphocyte biology (Fig. 6b).

IRF2BP2 interacts with NCOR co-repressor proteins. Although our data strongly suggest a repressor function for IRF2BP2 in erythroid gene regulation, how IRF2BP2 achieves gene repression is still unclear. We therefore purified endogenous IRF2BP2-containing protein complexes from MEL cells and identified the interacting proteins by IP-MS. As shown in Fig. 7, we retrieve known interacting proteins such as the other IRF2BP family members23 and several LDB1-complex members. In addition, IRF2BP2 also interacts with proteins involved in the cell cycle and transcriptional regulation (Fig. 7b). Among the latter group are several protein complexes known to mediate transcriptional repression. Prominent among these is the nuclear receptor co-repressor/silencing mediator for retinoid and thyroid receptors (NCOR/SMRT) co-repressor complex. Key components of this
complex are the NCOR1 and 2 proteins (the latter is also known as SMRT), and their repressive actions have been well documented. Intriguingly, Ncor1−/− mice die in utero due to abnormal erythropoiesis. To test whether NCOR proteins are indeed recruited to the regulatory elements of ETO2/IRF2BP2 target genes in erythroid progenitors, we performed NCOR1 ChIP-Seq in MEL cells. This reveals a significant overlap between NCOR1- and ETO2/IRF2BP2-binding sites (1,164 sites, Fig. 7c,d). In accordance with a possible cooperative relationship between these proteins, we find that these co-occupied sites include >64% of the primed erythroid-specific genes involved in haem biosynthesis and red cell membrane function (Fig. 7c,d). Furthermore, the ETO2–IRF2BP2–NCOR1 triad occupies key regulatory elements within the α- and β-globin loci (Supplementary Fig. 5c).

We next sought to perturb NCOR1 activity to assess its contribution to ETO2/IRF2BP2 target gene repression in erythroid progenitors. Multiple attempts at depleting NCOR1 levels in MEL cells using RNA interference (RNAi) (via both lentiviral shRNA delivery and short interfering RNA (siRNA)
transfections) failed to provide consistent effects on erythroid gene expression between individual shRNAs/siRNAs (n = 3–4 for each, Supplementary Fig. 6). We believe that the numerous Ncor1 isoforms expressed in MEL cells (as assessed by RNA-Seq, Supplementary Fig. 6) could contribute to the inconsistent effects seen with the different RNAi constructs, as alternative splicing has been reported to generate many functionally distinct NCOR1 isoforms40,42. In an attempt to inhibit NCOR1 protein function in a different manner, we decided to target the histone deacetylase 3 (HDAC3) that mediates the repressive actions of NCOR1 (reviewed in ref. 40). Apicidin is a small-molecule HDAC inhibitor reported to have specificity for HDAC3 (ref. 43), and we
IRF2BP2-deficient mice show abnormal FL erythropoiesis. Next, we interrogated IRF2BP2 function in vivo. For this purpose, we used an IRF2BP2-deficient mouse model generated by a genetrap strategy (Fig. 7e). The genetrap vector (containing a strong splice acceptor) was retrovirally inserted in the Irf2bp2 intron and results in a complete disruption of full-length messenger RNA (mRNA) production (Fig. 7e–g). Animals homozygous for the Irf2bp2 genetrap allele (hereafter referred to as Irf2bp2trp/trp mice) are rarely obtained and did not survive past 4 weeks of age, displaying severe growth retardation. In fact, although Irf2bp2trp/trp embryos appear to develop normally up to E18.5 and are obtained at the expected Mendelian ratio, live births are very rare (<5% of the expected number). This indicates that Irf2bp2trp/trp mice die either late during gestation or immediately after birth. Given that Irf2bp2 expression is particularly high in the developing mouse lungs, skeletal muscle and heart at E17.5 (ref. 27), defects in these tissues might underlie the observed lethality. To determine whether definitive erythropoiesis is affected in these mice, we collected E13.5 FL tissue from litters and NCOR1-binding sites in MEL cells. Note the significant co-localization of all three factors on the chromatin (1,164 sites), which included the mature erythrocytes is impaired in the absence of IRF2BP2. These data indicate that IRF2BP2 is important for effective FL erythropoiesis, as the output of mature erythrocytes is impaired in the absence of a functional Irf2bp2 allele.

We further characterized terminal differentiation in Irf2bp2trp/trp FLs by separating the Ter119+ population based on its forward scatter (FSC) profile (Fig. 7). As erythroid differentiation is paralleled by a reduction in cell size, this analysis visualizes a terminal differentiation gradient ranging from large and nucleated cells (high FSC) to small, enucleated cells (low FSC). Early enucleating cells (medium FSC) are more abundantly present in IRF2BP2-deficient FLs, while the percentage of small and enucleated erythrocytes is reduced (Fig. 7i).

FL erythropoiesis at earlier developmental time points (E11.5 and E12.5) also shows altered proportions of the different developing erythroid populations in Irf2bp2trp/trp embryos—in particular, within the mature Ter119+ compartment (Supplementary Fig. 7). Defects observed at E12.5 are similar in nature to the E13.5 phenotype, while at E11.5 erythropoiesis seems less severely affected: Irf2bp2trp/trp FLs exhibit a modest decrease in pro-erythroblast abundance and an increased presence of small terminally differentiated red cells. Moreover, gene expression analysis of selected ETO2/IRF2BP2 target genes (as identified in MEL cells) in sorted FL erythroid populations reveals an upregulation of these late erythroid markers in Irf2bp2trp/trp red cell precursors, as well as a downregulation of Myb expression—a marker for early erythroid progenitors (Supplementary Figs 8 and 9). Derepression of primed late erythroid genes is most striking at E11.5 and diminished at E13.5. Intriguingly, we observe a significant upregulation of the related Irf2bp1 gene in Irf2bp2trp/trp early erythroid precursors (Supplementary Figs 8 and 9). This phenomenon was most prominent at E13.5 and could indicate the existence of a compensatory mechanism involving IRF2BP1.

Combined, these observations point at disturbed erythroid differentiation kinetics and transcriptional regulation in the absence of IRF2BP2, confirming the notion that IRF2BP2 is important for effective erythropoiesis in vivo.

Discussion

Developmental processes are coordinated by spatio-temporal changes in gene expression laid down by the combinatorial actions of TFs and the cofactors they recruit. Exactly how TFs in
large multimeric complexes cooperate to create a regulatory environment that allows for rapid modulation of gene expression programmes is under intense investigation. Here we address the observation of a master haematopoietic TF complex, containing key factors required for the activation of a tissue-specific gene expression programme, that binds its target genes but maintains them in a developmental stage-specific primed state. Previous studies have shown that the activating LDB1 TF complex is already recruited to genes of the late erythroid-specific transcriptome in erythroid progenitors, before their full activation13,24. We have further confirmed and characterized the transcriptional and epigenetic status of these primed erythroid genes, revealing that they reside in active chromatin and are not controlled through an RNAPII pause–release mechanism13. One particular complex member, the ETO2 co-repressor, was found to mediate priming by repressing LDB1-complex target gene expression13,24. ETO2-mediated repression remains poorly understood, although the GFI1B TF, HDACs and the SIN3 transcription regulator family member A (Sin3A) repressor protein have been implicated (either directly or via their interaction with TAL1)13,30,47. We set out to further investigate the molecular mechanisms used by ETO2 to suppress terminal erythroid gene expression in progenitor cells.

A proteomics approach was first used to catalogue ETO2-interacting proteins in MEL erythroid progenitors (Fig. 1), identifying several repressor candidates known to bind ETO2 or other LDB1-complex members (for example, GFI1B3 and LSD1 (ref. 22)). Interestingly, we also detect the IRF2BP2 co-repressor in our interaction screen. Follow-up experiments firmly establish a cooperative role for ETO2 and IRF2BP2 in maintaining the late erythroid transcriptional programme in a primed state in undifferentiated progenitors: (1) IRF2BP2 strongly enhances ETO2-mediated repression in vitro, which is fully dependent on the ETO2–IRF2BP2 interaction (Fig. 2); (2) ETO2 and IRF2BP2 chromatin occupancy shows extensive genome-wide co-localization at genes involved in red blood cell development and function (Fig. 3); (3) similar to Gfi1b (ETO2), Irf2bp2 expression is reduced upon erythroid differentiation, concomitant with the upregulation of its erythroid target genes (Fig. 4); (4) depletion of ETO2 or IRF2BP2 leads to overlapping effects on gene expression, in particular the strong derepression of the late erythroid-specific transcriptome (Fig. 5); (5) ETO2 and IRF2BP2 mutant proteins unable to interact with each other show impaired induction of erythroid gene repression (Fig. 5); (6) ETO2 and IRF2BP2 bind the regulatory regions of >70% of the critical haem biosynthesis and erythrocyte membrane genes, the majority of which are repressed by both factors (Fig. 6).

Our biochemical analyses of IRF2BP2 protein complexes in MEL cells reveal the presence of NCOR/SMRT co-repressor complex members (Fig. 7b). In accordance, a key component of this complex, NCOR1, shows extensive genomic co-occupancy with IRF2BP2 and the ETO2/LDB1 complex in erythroid progenitors (Fig. 7d). Among these co-occupied sites, we find the vast majority of ETO2/IRF2BP2-repressed erythroid genes. Inhibition of NCOR1 complex activity by the apicidin inhibitor results in the upregulation of genes repressed by ETO2 and IRF2BP2 (Supplementary Fig. 6). On the basis of these data, we propose that IRF2BP2 confers repression upon ETO2/LDB1-complex target genes in part via its interaction with the NCOR/SMRT co-repressor complex. In accordance with our hypothesis, NCOR1-deficient mice showed abnormal FL erythropoiesis and developed severe anaemia during mid-gestation41.

We have also investigated the role of other ETO2-interacting putative repressor proteins. Although we could not detect Sin3A in our ETO2 IPs from erythroid cell lines, we did find the GFI1B TF and the LSD1 lysine demethylase, both of which have been implicated in the repression of LDB1-complex target genes3,22,48. Both proteins colocalize with the ETO2-containing LDB1 complex on the erythroid progenitor genome (Supplementary Fig. 5). In discordance with the findings of Hu et al.,23 we find no evidence for LSD1-mediated repression of the erythroid-specific Ebp4.2 gene (Fig. 5). In fact, we observe the opposite effect of LSD1 depletion on the late red cell transcriptome when compared with the Cba2t3/Irf2bp2 KDs (Fig. 5), similar to the loss of erythroid marker expression and differentiation upon LSD1 KD reported by Saleque et al.49. We conclude that LSD1, as part of the LDB1 complex, in general fulfills an activating role in erythroid differentiation (that is, possibly through controlling H3K4 methylation status50). In contrast, GFI1B, a DNA-binding repressor previously found to be required for terminal erythroid differentiation31, appears to repress LDB1-complex target genes in a similar manner as ETO2/IRF2BP2 (Fig. 5d). As was reported for ETO2, interactions between GFI1B and the activating LDB1-complex member TAL1 were strongly diminished upon terminal erythroid differentiation3. Cooperation of GFI1B with ETO2 and IRF2BP2 seems a plausible scenario warranting further investigation.

Intriguingly, IRF2BP2 binds many genomic regions independent of ETO2 and the LDB1 complex (Fig. 3). IRF2BP2-only sites are generally located closer to TSSs and therefore appear more frequently involved in short-range or promoter-based gene regulation as compared with LDB1-complex-associated IRF2BP2 sites (Supplementary Fig. 4). Furthermore, IRF2BP2 depletion affected the expression of numerous genes in an ETO2-independent fashion (Fig. 6). These observations suggest that IRF2BP2 plays additional roles in erythroid progenitors, independent of ETO2 and the LDB1 complex. In such cases, targeting of IRF2BP2 to the DNA could be mediated by ETS, TCF and EGR/SP family TFs, as binding motifs for these factors are strongly enriched at sites only bound by IRF2BP2 (Fig. 3c,d). Surprisingly, we did not detect a significant enrichment of IRF-binding motifs at these regions, nor did we find IRF TFs interacting with IRF2BP2 in our MS experiments. IRF2BP2 was originally identified as an IRF2-interacting factor in a yeast two-hybrid screen23. An IRF2–IRF2BP2 complex was recently detected in the K562 human erythroleukaemia cell line52, and IRF2 is expressed in MEL and primary murine erythroid progenitor cells (>5.0 RPKM as measured by RNA-Seq). Whether this discrepancy reflects a species–specific difference or differences in experimental systems is unclear. Nevertheless, our combined analysis of IRF2BP2-binding sites and protein partners does provide preliminary insight into the ETO2/LDB1-independent functions of IRF2BP2. Genes bound only by IRF2BP2 are significantly enriched for functions related to proliferation and apoptosis (Fig. 3b), and the cell-cycle regulator cyclin-dependent kinase 11B interacts with IRF2BP2 (Fig. 7b). Interestingly, several studies have implicated IRF2BP2 in the regulation of cell survival25,26,33.

In agreement with our experiments in MEL cells, IRF2BP2 also appears to be important for erythropoiesis and erythroid gene regulation in vivo. Perinatal lethality of IRF2BP2-deficient mice precluded the analysis of adult erythropoiesis in our Irf2bp2 genetrap model. However, analysis of mid-gestation-definitive FL erythropoiesis in these mice shows that IRF2BP2 is required for an effective output of terminal erythroid differentiation (Fig. 7i,j; Supplementary Fig. 7b). The exact nature of this defect remains to be determined, but our experiments indicate the presence of a partial differentiation block at the erythroblast stage, before enucleation (Fig. 7i,j). Alternatively, the observed erythroblast expansion could be a consequence of accelerated progenitor differentiation or represent a compensatory mechanism, which could also explain the partially exhausted progenitor.
component observed at E13.5 (Fig. 7i) and the premature erythroid differentiation at E11.5 as compared with E13.5 (Supplementary Figs 8 and 9). As IRF2BP2 and IRF2BPL deficient progenitors that is significantly more pronounced at E13.5 (Supplementary Figs 8 and 9). Intriguingly, we observe a strong upregulation of ETO1 expression during in vivo erythroid maturation—in particular at E13.5 (Supplementary Figs 8 and 9). As IRF2BP2 and IRF2BPL reside in the same protein complexes (Fig. 7b), we hypothesize that elevated IRF2BPL levels might compensate for a loss of IRF2BP2 in vivo. Future investigations will reveal whether the other IRF2BP family members indeed play a role in erythropoiesis and in the regulation of primed erythroid gene expression. While this paper was under review, IRF2BP2 was shown to be important for macrophage-mediated inflammation, suggesting that IRF2BP2 may have a multifaceted role in blood cell development and function.

In conclusion, we show that the control of developmentally primed erythroid genes depends on the cooperative actions of ETO2 and its novel binding partner IRF2BP2. Repression by the ETO2–IRF2BP2 axis is lost during erythroid differentiation, resulting in the full activation of the late erythroid-specific transcriptome by the LDB1 complex. These results provide new insight into the control of lineage-specific transcriptional programmes, as they suggest that an intricate balance between the activating and repressive components of a TF complex underlies the implementation of lineage-specific gene expression. Furthermore, using an IRF2BP2-deficient mouse model, we confirm the relevance of a functional Irf2bp2 allele for effective erythropoiesis in vivo.

Methods

Cell culture and Irf2bp2 genetrap animals. MEL (C8 clone56) and HEK 293T (ATCC) cells were maintained in DMEM containing 10% FCS and penicillin/streptomycin. ETO2-V5-Bio MEL cells expressing BirA were generated and maintained as described previously21. Apicidin (Calbiochem, 178276) was reconstituted in dimethylsulfoxide to obtain a 10 mM stock. MEL cells were treated by adding Apicidin directly to the medium to a final concentration of 100 nM. Myb (C21/6) C57/B6 ES cells were provided by the Texas A&M Institute for Genomic Medicine (College Station, TX) through the insertion of a genetrap construct in the first intron of the Irf2bp2 gene (clone ISt11591C1). Genetrap location was verified using standard PCR and sequencing methods. Mouse ES cells were injected into blastocysts and implanted into pseudopregnant albino foster mice. The location was verified using standard PCR and sequencing methods. Mouse ES cells were obtained by reconstituting the Ldb1 knockout allele using CRISPR/Cas9 (see Fig. 2). For rescue experiments, MEL cells transduced with shRNAs against Irf2bp2 (ETO2) or Irf2bp2deltaRING-deletion mutant (see Fig. 2) in the pcDNA3.1 expression plasmid was a kind gift from Dr Jan van der Knaap (Erasmus MC). A Renilla reporter plasmid (pUC57-Ren/Ren) and the pCMV–β-geo reporter plasmid were obtained using GeneCellin (BioCell Challenge). RNA extractions and gene expression analysis were performed by rapidly washing the beads containing bound ETO2 directly to the medium to a final concentration of 100 nM KCl and boiled for 5 min at 95 °C. Extracts were treated with 1 U Benzonase nuclease (Millipore). Protein extracts were incubated with the anti-Flag, anti-V5 or anti-HA antibodies cocktail (Roche). Extracts were treated with 1 U Benzonase nuclease (Millipore). Protein extracts were incubated with the anti-Flag, anti-V5 or anti-HA antibodies overnight at 4 °C, followed by addition of protein A or G Sepharose bead slurry (50 μl slurry per IP; Millipore) and incubation at 4 °C for 1 h. Beads were pelleted, washed three times in Heng 100 buffer (10 mM HEPES KOH pH 7.5, 150 mM NaCl, 1% Triton X-100) and boiled in Laemmli buffer before being subjected to western blot analysis. Proteins were loaded on a 4–12% acrylamide gel and lanes were cut for LC–MS/MS analysis as described above. For MS analysis of ETO2- and IRF2BP2-interacting proteins, two independent biological replicates (for both experimental and control samples) were analysed to ensure reproducibility and specific binding partner identification. The following antibodies were used: ETO2 G-20 (Santa Cruz, sc9741), an IRF2BP2 rat monoclonal KT139 (clone 10G3, produced by Absea Antibodies, Beijing), GFIIB B-7 (Santa Cruz, sc8559), LDL1 N-18 (Santa Cruz, sc11198), LSD1 (Abcam, ab17721), RUNX1 H-65 (Santa Cruz, sc26879); E2A V-18 (sc-349), HEB A 20 (sc-357), SSBP3 (Abcam, ab83815), V5 (Invitrogen, R960-25), Flag M2 (Sigma, F1804) and haemagglutinin (HA) (Sigma, H6908).

Co-IPs and luciferase assays in HEK 293T cells. HEK 293T cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions. For ETO2–IRF2BP2 interaction domain mapping, we constructed a series of Flag-tagged ETO2–deletion mutants, V5-IRF2BP2 and the HA-IRF2BP2deltaRING-deletion mutant (Fig. 2) in the pCMV–β-geo expression vector (Invitrogen). HEK 293T cells were lyzed 48 h post transfection in whole-cell lysis buffer (20 mM HEPES KOH pH 7.5, 150 mM KCl, 1% Triton, 2.5 mM EDTA, 5 mM dithiothreitol, 0.1% Triton X-100 (Sigma) and protease inhibitor cocktail (Roche)). Extracts were treated with 1 U Benzonase nuclease (Millipore). Protein extracts were incubated with the anti-Flag, anti-V5 or anti-HA antibodies overnight at 4 °C, followed by addition of protein A or G Sepharose bead slurry (50 μl slurry per IP; Millipore) and incubation at 4 °C for 1 h. Beads were pelleted, washed three times in lysis buffer and boiled for 5 min at 95 °C in Laemmli buffer before being subjected to western blot analysis. Full-length Kdm1a (LSD1) cDNA was cloned in mosPDNA3.1 for luciferase assay experiments. The co-IP was performed by fusing full-length Cbfα23 cDNA sequence to a Gal4 DNA-binding domain in pcDNA3.1. The Gal4-responsive firefly luciferase plasmid was a kind gift from Dr Jan van der Knaap (Erasmus MC). A Renilla luciferase expressing vector (pRL-TK, Promega) was co-transfected and used for normalisation. A firefly luciferase assay was performed using the DualLuciferase Reporter Assay System (Promega) according to the manufacturer’s instructions.44 For rescue experiments, MEL cells transfected with shRNAs against cbfα23 (ETO2) or Irf2bp2 were transfected with expression vectors containing shRNA-immune cDNAs encoding ETO2, IRF2BP2 or mutant versions of these proteins using GeneCellin (BioCellChallenge). RNA extractions and gene expression analysis by qPCR were carried out 48 h after transfection.

ChIP and ChIP-Seq experiments. Protocols for the preparation of chromatin from MEL cells, IP and sample preparation for Illumina sequencing have been previously described in great detail52. For NCCR1 CHIP-Seq, 105 MEL cells were cultured...
crosslinked with 2 mM d Princeton glutaraldehyde (Thermo Fisher Scientific) and 1% formaldehyde as previously described. Antibodies used for ChIP (10 μg per precipitation) were chosen to those used for IP (details above), except 1:100 GTF (D-19 Santa Cruz, sc8599). Reads were mapped against NCBI build 37.1 of the mouse genome (mm9) using Bowtie (version 2.0.0)89. Uniquely mapped reads were extended to 200 bp in the 3’ direction and were transformed into a genome-wide read density (rppm) using custom R scripts. MACS (version 1.4.2)90 and UCAT (version 0.80)91 and in-house peak-calling software (available from the Github repository, https://github.com/supatt/ChIPseqTools.git) with default parameters were used to computationally identify binding sites. We combined binding sites identified by all three methods to define consensus-binding regions using GenomeRanges. Genomic Regions Predicted by each program were filtered on each genomic location. We assessed value ‘1’ to the genomic locations that overlapped with the predicted binding sites, and we assigned ‘0’ to the non-overlapping locations. We then generated the whole-genome coverage vectors from all binding regions of all methods and summed up these coverage vectors. We selected only genomic regions that had a summed up coverage value ≥ 2 and a minimum region length ≥ 150 bp as candidate consensus-binding regions. Consensus-binding regions were given P values based on a negative binomial distribution as modified from PeakSeg82 and assigned P values were adjusted using the Benjamini–Hochberg method. Candidate binding sites were then selected for the downstream analysis based on the following criteria: read counts ≥ 10 reads, fold changes (FCs) ≥ 2 compared with immunoprecipitation (IP) control and adjusted P value < 0.01. To classify co-binding patterns, ETO2- and IRF2BP2-binding sites were combined using GenomicRanges (using the ‘findOverlaps’ function with the minimum overlap regions = 250 bp from the peak centre). Binding signal coverage for each site was then normalized to equal levels of background signal in both Cbfa2t3 and IgG control experiments (normalization method was modified from PeakSeg82). Normalized coverage for the IP control experiment was subtracted from the normalized coverage for the Antibody experiment. We next retrieved the subtracted coverage within ± 0.5 kb relative to the centre of each binding site and calculated the standard z-scores in each sub-window (25 bp). The matrix of standard z-scores for each individual binding site was then subjected to K-means clustering (K = 3) clustering. Clustering analysis results were visualized with Java Treeview83. After K-means clustering, we selected representative binding sites for each co-binding pattern (1,760 for ETO2-only, 2,730 for ETO2/IRF2BP2 and 5 random sets of 4,000 IRF2BP2-only). We retrieved repeat-masked 200-bp DNA sequences centred on each binding site and removed the binding sites that contain ≥ 1 million cells were sorted for RNA extraction.

**Flow cytometry.** E11.5–13.5 embryos were harvested and dissected to collect the Fl. Hundred microlitre of PBS containing 1 million single cells obtained from whole blood of E13.5 FLs were stained with 10 μM Calcein-PE (553673) antibodies (BD Pharmingen). Hoechst was used as a viability dye (Sigma), and cells positive for Hoechst staining were excluded from further analysis (>70% of the total cell population consisted of viable cells). Flowcytometry analysis was performed using a BD LSRFortessa flowcytometry (BD Biosciences), collecting a minimum of 10,000 events per sample. FACS sorting was performed with a BD FACSAria III (BD Biosciences) using the above described staining protocol. At least 1 million cells were selected for RNA extraction.

**Immunofluorescence.** MEL cells were fixed on poly-prep glass slides (Sigma) and fixed in 4% paraformaldehyde for 15 min at room temperature. Cells were permeabilized with 0.1% Triton X-100, blocked with 0.5% bovine serum albumin/0.15% glycine in PBS and incubated overnight with ETO2 or VS antibodies at 4°C. After a 2-h incubation with appropriate secondary antibodies at room temperature, coverslips were mounted on glass slides with Vectashield (DAP, Vector Laboratories).

**RNA sequencing data sets.** The following publicly available data sets were used: LDB1, GATA1 and ETO2 ChIP-Seq data (MEL, SRA ERA000161 (ref. 5)); RNA-Seq data (MEL/G1E/G1E-ER, ENCODE Penn State University; available at the UCSC Genome Browser (mouse genome, mm9)); p300 ChIP-Seq data (MEL, ENCODE Stanford/Yale; available at the UCSC Genome Browser (mouse genome, mm9)); H3K27Ac ChIP-Seq data (MEL/LF14.5/Brain, ENCODE Ludwig Institute for Cancer Research; available at the UCSC Genome Browser (mouse genome, mm9)); GATA1, TAL1 and RNAIP ChiP-Seq data (MEL/G1E/G1E-ER and MEL/G1E/G1E-ER, ENCODE Penn State University). Database online25).

**References**

1. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
2. Love, P. E., Warzecha, C. & Li, L. Ldb1 complexes: the new master regulators of erythroid gene transcription. *EMBO J.* **25**, 10235–10250 (2006).
3. Schuh, A. H. et al. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol. Cell. Biol.* **25**, 10235–10250 (2005).
4. Goardon, N. et al. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J.* **25**, 357–366 (2006).
5. Soler, E. et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* **24**, 277–289 (2010).
6. Li, L. et al. Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood* **121**, 4575–4585 (2013).
7. Meier, N. et al. Novel binding partners of Ldb1 are required for haematopoietic development. *Development* **133**, 4913–4923 (2006).
8. Kiefer, C. M. et al. Distinct Ldb1/NLL complexes orchestrate gamma-globin repression and reactivation through ETO2 in human adult erythroid cells. *Blood* **118**, 6200–6208 (2011).
ARTICLE

9. Fischer, M. A., Moreno-Miralles, I., Hunt, A., Chyla, B. J. & Hiebert, S. W. Myeloid translocation gene 16 is required for maintenance of haematopoietic stem cell quiescence. EMBO J. 31, 1494–1505 (2012).

10. Hunt, A., Fischer, M. A., Engel, M. E. & Hiebert, S. W. Mtg16/Eto2 contributes to murine T-cell development. Mol. Cell. Biol. 31, 2544–2551 (2011).

11. Chyla, B. J. et al. Deletion of Mtg16, a target of t(6;21), alters hematopoietic progenitor cell proliferation and lineage allocation. Mol. Cell. Biol. 28, 6234–6247 (2008).

12. Thollnier, C. et al. Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. J. Exp. Med. 209, 2017–2031 (2012).

13. Gruber, T. A. et al. An Inv(16)(p13.3q24.3)-encoded CBFAML1-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell 22, 683–697 (2012).

14. Micci, F. et al. Translocation t(1;16)(p31;q24) rearranging CBFAML1 is specific for acute erythroleukemia. Leukemia 25, 1510–1511 (2012).

15. Barrett, C. W. et al. Kaiso directs the transcriptional coresspr MTG16 to the Kaiso binding site in target promoters. PLoS ONE 7, e51205 (2012).

16. Hiebert, S. W., Fischer, M. A., Moreno-Miralles, I., Hunt, A., Chyla, B. J. & Hiebert, S. W. Gene induction and repression during terminal erythropoiesis. Mol. Cell. Biol. 28, 6759–6772 (2008).

17. Enver, T. & Greaves, M. Loops, lineage, and leukemia. Cell 94, 9–12 (1998).

18. Pimkin, M. et al. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. Genome Res. 24, 1932–1944 (2014).

19. Fischer, M. A., Moreno-Miralles, I., Hunt, A., Chyla, B. J. & Hiebert, S. W. An Inv(16)(p13.3q24.3)-encoded CBFAML1-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell 22, 683–697 (2012).

20. Wong, P. et al. Identification of a novel transcription complex that selectively modulates apoptosis of breast cancer cells through regulation of FASTKD2. J. Biol. Chem. 286, 49988–49999 (2011).

21. Koeppel, M. et al. Global regulation of erythroid gene expression by a novel transcription complex that selectively modulates apoptosis of breast cancer cells through regulation of FASTKD2. J. Biol. Chem. 286, 49988–49999 (2011).

22. Mottis, A., Mouchiroud, L. & Auwerx, J. Emerging roles of the coexpressors NCoR1 and SMRT in homeostasis. Genes Dev. 27, 819–835 (2013).

23. Jepsen, K. et al. Combinatorial roles of the nuclear receptor coexpressor in transcription and development. Cell 102, 753–763 (2000).

24. Goodson, M. L., Mengeling, B. J., Jonas, B. A. & Privalsky, M. L. Alternative mRNA splicing of coexpressors generates variants that play opposing roles in adipocyte differentiation. J. Biol. Chem. 286, 40982–40999 (2011).

25. Kingsley, P. D. et al. Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature 446, 882–887 (2007).

26. Chen, H. H. et al. A novel transcription complex that selectively modulates apoptosis of breast cancer cells through regulation of FASTKD2. J. Biol. Chem. 286, 49988–49999 (2011).

27. Deisseroth, A. & Hendrick, D. Human alpha-globin gene expression following ablation of the negative control. Bioinformatics 27, 6479–6483 (2011).

28. Portales-Casamar, E. et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. 38, D105–D110 (2010).

29. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. Genome Biol. 11, R36 (2010).

30. Kim, Y. et al. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 6, 369–77 (2009).

31. Palii, C. G. et al. Genome-wide identification of TAL1's functional targets: an integrated encyclopedia of DNA binding in hematopoietic progenitors. J. Exp. Med. 215, 1375–1389 (2008).

32. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization on microarray expression data. Bioinformatics 14, 787–794 (1998).

33. Rozowsky, J. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R37 (2008).

34. Saldanha, A. J. Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246–3248 (2004).

35. Portales-Casamar, E. et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. 38, D105–D110 (2010).

36. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. Genome Biol. 8, R24 (2007).

37. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).

38. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and genes fusions. Genome Biol. 14, R36 (2013).

39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 530 (2014).

40. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian genome-wide translational activity by RNA-Seq. Nat. Methods 6, 219–222 (2009).

41. Saeed, A. I. et al. TMA microarray software suite. Methods Enzymol. 411, 134–193 (2006).

Acknowledgements

We thank members of the Soler and Grosveld laboratories for helpful discussions. We thank Annouk van Oosten for constructing the VS-IRF2BP2 expression plasmid and Dr Stephen Goodfellow (St George’s Hospital Medical School, London, UK) for providing human IRF2BP2 cDNA, Dr Jan van der Kraap (Eramus MC, Rotterdam, the Netherlands).
for the Gal4-responsive luciferase reporter construct and Erasmus MC Experimental Animal Facility personnel for animal care and handling. We also thank Dr Paul-Henri Romeo (INSERM, Paris, France) for providing murine Eto2 cDNA, and Erasmus MC Biomics personnel for excellent technical assistance with Illumina sequencing and data analysis. R.S. is supported by the Royal Netherlands Academy of Arts and Sciences (KNAW; ‘Academy Assistant’ fellowship), an EMBO Long-term Fellowship (ALTF 1201–2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014, proposal #654933). C.A.–S. is supported by a Marie Curie European Reintegration Grant (FP7-PEOPLE-2010-RG). A.C. is supported by a CEA/IRTELIS fellowship. P.K. is supported by grants from EpigenSys/ERASysBio+/FP7 (NL: NWO, UK: BSR, D: BMBF), the Bluescript EU Integrated Project and the Netherlands Genomics Initiative (MEC Booster grant). F.G. is supported by a KNAW Academy Professorship, the Cancer Genomics Center (NGI, NL), the NIRM (NL) consortium and the SyBioSS EU Consortium. E.S. is supported by grants from the FSER (Schlumberger Foundation for Education and Research), ARC foundation (‘projet ARC’ no. SFI20121205625) and the Atip-Avenir program.

Author contributions
R.S., F.G. and E.S. conceived and designed the experiments. R.S., A.C., T.S., V.B., P.K., H.I.B., X.Y., C.A.–S. and E.S. performed the experiments. R.S., S.T., A.C., T.S., V.B., M.–L.A., B.L. and E.S. analysed the data. J.D. and K.B. performed proteomics experiments. C.K., Z.O. and W.v.I. performed high-throughput sequencing experiments. A.M. performed ES cell injections into blastocysts to generate If/Ifp2 genetrap chimeric mice. R.S., S.T., F.G. and E.S. wrote the paper.

Additional information
Accession codes: Nucleotide sequences for ChIP-Seq and RNA-Seq data sets have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE59859. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium/PRIDE partner repository under accession code PXD001892.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Stadhouders, R. et al. Control of developmentally primed erythroid genes by combinatorial corepressor actions. Nat. Commun. 6:8893 doi: 10.1038/ncomms9893 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/