

Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells

Mira T. Kassouf,¹ Jim R. Hughes,¹ Stephen Taylor,² Simon J. McGowan,² Shamit Soneji,¹ Angela L. Green,¹ Paresh Vyas,¹ and Catherine Porcher^{1,3}

¹MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford University, Oxford OX3 9DS, United Kingdom; ²Computational Biology Research Group (CBRG), Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford University, Oxford OX3 9DS, United Kingdom

Coordination of cellular processes through the establishment of tissue-specific gene expression programs is essential for lineage maturation. The basic helix-loop-helix hemopoietic transcriptional regulator TAL1 (formerly SCL) is required for terminal differentiation of red blood cells. To gain insight into TAL1 function and mechanisms of action in erythropoiesis, we performed ChIP-sequencing and gene expression analyses from primary fetal liver erythroid cells. We show that TAL1 coordinates expression of genes in most known red cell-specific processes. The majority of TAL1's genomic targets require direct DNA-binding activity. However, one-fifth of TAL1's target sequences, mainly among those showing high affinity for TAL1, can recruit the factor independently of its DNA binding activity. An unbiased DNA motif search of sequences bound by TAL1 identified CAGNTG as TAL1-preferred E-box motif in erythroid cells. Novel motifs were also characterized that may help distinguish activated from repressed genes and suggest a new mechanism by which TAL1 may be recruited to DNA. Finally, analysis of recruitment of GATA1, a protein partner of TAL1, to sequences occupied by TAL1 suggests that TAL1's binding is necessary prior or simultaneous to that of GATA1. This work provides the framework to study regulatory networks leading to erythroid terminal maturation and to model mechanisms of action of tissue-specific transcription factors.

[Supplemental material is available online at <http://www.genome.org>. The ChIP-seq and expression array data from this study have been submitted to the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession nos. GSE18720 and GSE21877, respectively. Processed data is also available at <https://gbrowse.molbiol.ox.ac.uk/cgi-bin/gbrowse/SCLtargets/>.]

Understanding the mechanisms by which stem and multipotential progenitor cells progressively commit to uni-lineage programs of gene expression are key biological questions. Red blood cell production (erythropoiesis) has been extensively characterized, providing an ideal model to study cell differentiation. The red blood cell lineage is characterized by the maturation of erythroid precursors into terminally differentiated, enucleated erythrocytes. Progressive erythroid cellular maturation stages have been defined by morphological criteria (for review, see Klinken 2002) and expression of cell surface markers (Socolovsky et al. 2001). Early erythroid precursors express increasing levels of erythropoietin receptor (EPOR), which is required for terminal erythroid maturation (Lin et al. 1995; Wu et al. 1995). Cells of the next erythroid maturation stage, proerythroblasts, express high levels of the transferrin receptor TFRC (also known as CD71) and start to produce hemoglobin. The red cell-specific antigen Ter119 is expressed on all subsequent differentiating murine erythroid cells, while expression of TFRC decreases as maturation progresses. An important characteristic of erythrocytes is the unique composition of their membrane and cytoskeleton, which is required to confer high flexibility to mature red blood cells while maintaining their

transport and mechanical properties (Mohandas and Gallagher 2008). In mammals, definitive red cell production initially occurs in the fetal liver and then shifts to the bone marrow in adult life.

Tissue-specific transcriptional regulators play essential roles in establishing red cell-specific gene expression programs (Cantor and Orkin 2002). One example is TAL1 (formerly SCL), a basic helix-loop-helix (bHLH) transcription factor (Lecuyer and Hoang 2004). TAL1 is initially required for specification of hemopoietic cells during embryonic development (Shivdasani et al. 1995; Porcher et al. 1996; D'Souza et al. 2005; Patterson et al. 2005). Later in hemopoietic differentiation, continued TAL1 expression is critically required for erythroid maturation as lack of TAL1 leads to a block in erythropoiesis (Hall et al. 2003, 2005; Mikkola et al. 2003; Schlaeger et al. 2005; McCormack et al. 2006).

TAL1 functions as an obligate heterodimer. It interacts with the ubiquitously expressed bHLH E-proteins (or TCF3 [also known as E2A]) to bind to its DNA recognition motif, an E-box (CANNTG). In red cells, TAL1 is part of multiprotein complexes that include the LIM-only domain protein LMO2 and the LIM domain-binding protein LDB1. The TAL1/TCF3/LMO2/LDB1 complex recruits co-factors with activator or repressor functions, such as EP300, GF11B, CBFA2T3 (ETO2), and KDM1A (Huang et al. 1999, 2000; Schuh et al. 2005; Goardon et al. 2006; Hu et al. 2009). It can also bind other DNA-bound transcription factors such as the hemopoietic regulator GATA1 (to form the "pentameric complex"; Wadman et al. 1997) or the ubiquitously expressed protein SP1 (Lecuyer et al.

³Corresponding author.

E-mail catherine.porcher@imm.ox.ac.uk; fax 44-1865-222-500.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.104935.110>.

2002). Evolutionary conserved association of E-box and GATA motifs separated by 9 to 12 nucleotides [GATA(n_{9-12})CANNTG] has been reported in regulatory regions of many erythroid-specific genes (Anderson et al. 1998; Vyas et al. 1999; Lahlil et al. 2004).

To date, only a handful of red cell-specific genes have been reported as direct, functional TAL1 target genes. These include genes coding for the transcription factors erythroid Kruppel-like factor 1 (KLF1) and GATA1, the membrane proteins Band 4.2 (EPB4.2) and glycophorin A (GYPA), the cytokine receptor *Kit* as well as the alpha- and beta-globin gene clusters (*Hba* and *Hbb*, respectively) (Vyas et al. 1999; Anderson et al. 2000; Lecuyer et al. 2002; Xu et al. 2003; Anguita et al. 2004; Lahlil et al. 2004; Song et al. 2007; Kassouf et al. 2008; Manwani and Bieker 2008). TAL1 can be recruited to DNA either directly by its basic DNA-binding domain or independently of its DNA-binding activity. Extending our initial in vitro observations (Porcher et al. 1999), we recently reported on the DNA-binding independent functions of TAL1 in a knock-in mouse model expressing a DNA-binding mutant form of TAL1 (TAL1^{RER}) (Kassouf et al. 2008). We showed that direct DNA binding is dispensable for hemopoietic specification in early development. In contrast to conventional *Tal1*-null embryos that die from complete absence of embryonic blood at embryonic day (E)9.5 (Shivdasani et al. 1995), *Tal1*^{RER/RER} embryos survived at that stage. They die from day E14.5 onward with anemia at a time when massive expansion of erythroid cells occurs in the fetal liver to satisfy the oxygen transport requirements in the growing embryo (Kassouf et al. 2008). Though *Tal1*^{RER/RER} erythroid precursors are specified, terminal maturation of *Tal1*^{RER/RER} erythroid cells is grossly perturbed, with decreased numbers of TFRC⁺/Ter119⁺ red cells that fail to fully hemoglobinize and mature. Therefore, direct DNA binding by TAL1 is required for terminal erythroid maturation. At a molecular level, expression of many known TAL1 erythroid target genes was deregulated. We also reported that the DNA-binding mutation did not always fully prevent binding of TAL1 to its target loci in red cells (Kassouf et al. 2008), thereby supporting previous observations that TAL1 can be recruited to gene loci in absence of functional E-boxes (Vyas et al. 1999; Lecuyer et al. 2002).

Here, using chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-seq), we present a genome-wide characterization of the sequences bound by TAL1 in wild-type and *Tal1*^{RER/RER} early Ter119⁻ fetal liver erythroid precursors, coupled with gene expression analyses. The aims of the work are to characterize the breadth of red cell processes regulated by TAL1 that execute terminal differentiation, distinguish TAL1 direct and indirect DNA binding functions, and define DNA motifs underlying the sequences bound by TAL1. More generally, we hoped this would characterize networks and processes coordinated during lineage maturation and model the mechanisms of action of bHLH tissue-specific transcription factors.

Results and Discussion

Genome-wide mapping of sequences bound by TAL1

We performed anti-TAL1 ChIP assays from material isolated from immature, Ter119⁻ erythroid cell populations derived from day E12.5 wild-type (*Tal1*^{WT/WT}) fetal livers followed by high-throughput sequencing (ChIP-seq) (Fig. 1A). To compare TAL1's direct versus indirect DNA-binding activities, we also analyzed material isolated from *Tal1*^{RER/RER} fetal livers; importantly, expression levels of the TAL1 protein in Ter119⁻ *Tal1*^{RER/RER} cells

remained similar to those observed in wild-type cells (Kassouf et al. 2008). Approximately 4 million uniquely mapped reads were aligned for each sample (Fig. 1A) and displayed on the Generic Genome Browser (GBrowse). Using Cisgenome (Ji et al. 2008), 4364 peaks were identified from the *Tal1*^{WT/WT} sample and 694 peaks from the *Tal1*^{RER/RER} sample. The peaks were ranked by the number of reads mapped in a region, with peak 1 having the greatest number of reads. After appropriate quality filtering, peaks 1–2994 (*Tal1*^{WT/WT} sample) and 1–594 (*Tal1*^{RER/RER} sample) were retained for further analyses.

We then compared the genomic coordinates of the peaks detected in the *Tal1*^{WT/WT} and *Tal1*^{RER/RER} samples. Out of the 2994 peaks from *Tal1*^{WT/WT} cells, 2400 were not detected in the *Tal1*^{RER/RER} sample; this group was termed “WT only” (80.1% of total peaks; Fig. 1B, Supplemental Table 1). The remaining 594 peaks were identified in both samples. Thus, peaks detected in the mutant samples are a subset of the peaks present in the wild-type sample. A quantitative analysis revealed that the intensity of the peaks shared between *Tal1*^{WT/WT} and *Tal1*^{RER/RER} samples varied. This allowed us to divide the shared peaks into two categories, based on the ratio of their intensity in mutant (RER) versus wild-type (WT) samples (Fig. 1B). The first category represented peaks partially affected by the DNA-binding mutation (suggesting the occurrence of both direct and indirect TAL1 binding) and defined by RER/WT ratios between 0.1 and 0.8 (457 peaks representing 15.2% of total peaks; Fig. 1B; Supplemental Table 2). The second category represented peaks minimally affected by the DNA-binding mutation (suggesting that recruitment to DNA may occur mainly independently of direct TAL1 DNA-binding activity) and defined by RER/WT ratios between 0.8 and 1.8 (137 peaks representing 4.6% of total peaks; Fig. 1B; Supplemental Table 3). As an example, Figure 1C shows the peaks in the three different categories (WT only, 0.1–0.8, and 0.8–1.8) over two genomic regions in GBrowse.

The distribution of the peaks detected in the *Tal1*^{RER/RER} sample was compared with that of their corresponding wild-type peaks, according to their intensities. In Figure 1D, the same peak number is associated to a peak in the *Tal1*^{RER/RER} sample and its corresponding wild-type peak. All the peaks belonging to the category RER/WT 0.1–0.8 were found within the peak number range 1–1705. The category RER/WT 0.8–1.8 shows fewer peaks of high intensity and more peaks of medium to low intensity (peak numbers ~1500–2985). This suggested that, among the sequences with the strongest affinity for TAL1 (peaks one to ~1700), some might contain specific arrangements of *cis*-acting elements allowing for recruitment of TAL1 independently of its DNA-binding activity, albeit with less affinity than through direct binding. Sequences likely to recruit TAL1 through both direct and indirect mechanisms with similar efficiency show a lower affinity for TAL1.

This initial analysis provides genome-wide molecular evidence that direct DNA-binding activity is required to recruit TAL1 to ~80% of its red cell target sequences. Thus, it is not surprising that loss of direct TAL1 DNA binding causes a dramatic erythroid phenotype leading to embryonic lethality (Kassouf et al. 2008). As the peaks detected from the *Tal1*^{RER/RER} sample are a subset of the wild-type peaks, mutation of the DNA-binding domain of TAL1 did not create novel DNA-binding specificities but, instead, suggests that, at a subset of target sequences, TAL1 is likely to be normally recruited to DNA indirectly, as previously suggested by our group and others (Porcher et al. 1999; Lecuyer et al. 2002; Kassouf et al. 2008).

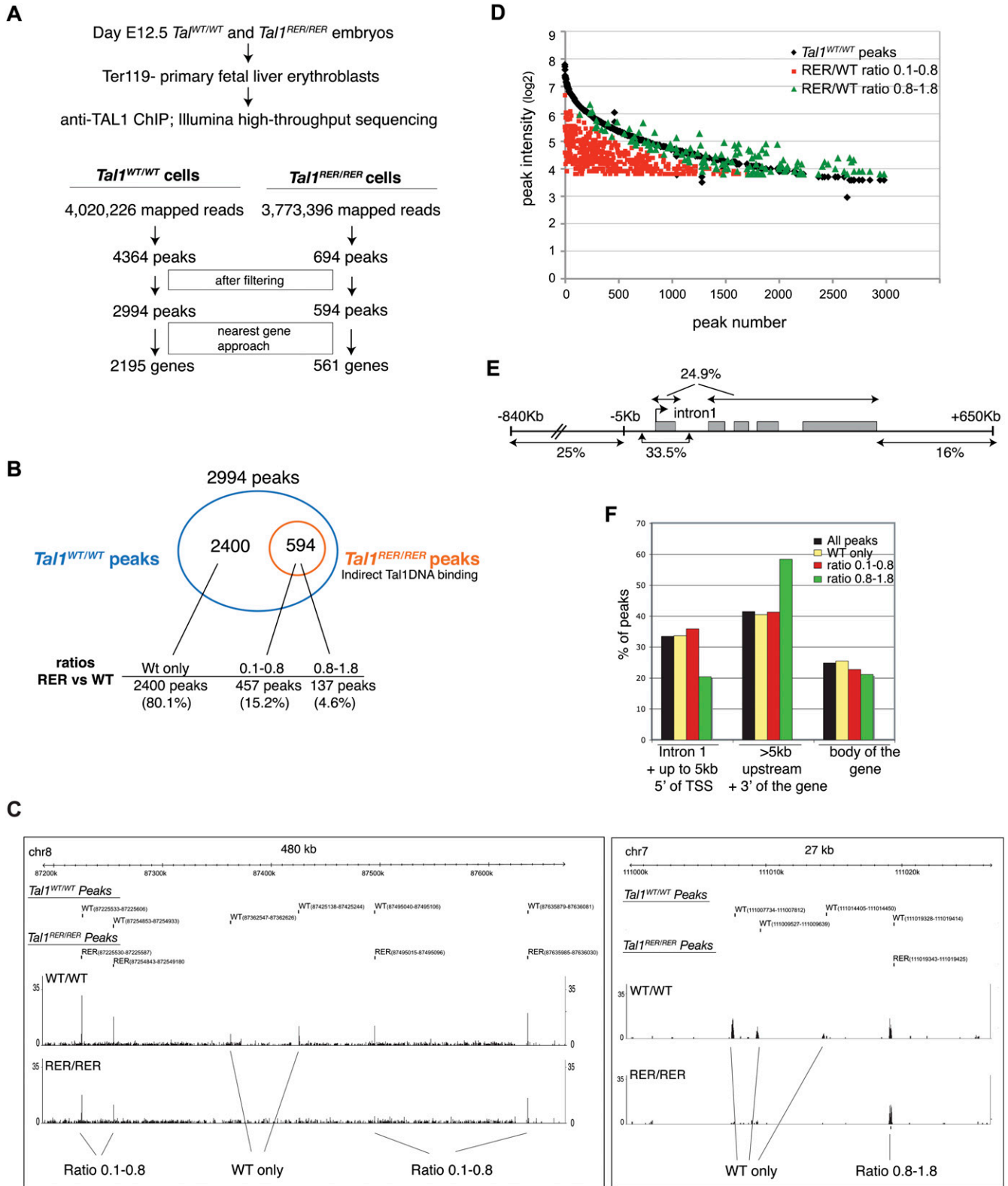


Figure 1. (Legend on next page)

Detection of known TALI binding sites and validation of new binding sites

Using the nearest gene approach, we identified 2195 genes associated with the 2994 peaks in the *Tal1*^{WT/WT} sample (Fig. 1A). TALI binding peaks were distributed between the proximal promoter and intragenic and distal binding sites (representing 33.5%, 24.9%, and 41% of total peaks, respectively; Fig. 1E; Supplemental Table 4), suggesting both short and long-range transcriptional control. Although peaks at proximal promoters are more likely to be functionally attributed to the relevant gene, our data identify potential active distal sites for investigation. Of note, we noticed an enrichment of peaks that are bound by the DNA-binding mutant form of TALI (RER/WT ratios 0.8–1.8) at distal *cis*-elements, at the expense of proximal regions (Fig. 1F; Supplemental Table 4). This suggested that, when TALI recruitment is largely independent of its direct DNA-binding activity, this is more likely to occur at distant enhancers, rather than promoters.

The list of TALI candidate target genes identified by ChIP-seq contains most, if not all, previously reported functional targets of TALI in red cells. For example, TALI was bound to the promoter region and the +40 downstream element in the *Tal1* locus itself and to the promoter and enhancer regions of *Epb4.2*, *Hba* and *Hbb* loci, *Klf1*, and *Gypa* (Fig. 2A, top panels, black tracks) (for review, see Ogilvy et al. 2007; Kassouf et al. 2008). New, previously unreported TALI binding sites were identified in the loci of known TALI target genes, such as the intragenic peaks detected in the *Gypa* gene (Fig. 2A). ChIP-seq also revealed binding to genomic sequences associated with numerous genes previously not reported as TALI targets. We selected six of these elements, two located in promoter regions (*Aqp9* and *Trim10*), three within intragenic sequences (*Alad*, *Prdx2*, and *Mme11*), and one upstream of the transcriptional unit (*Ssr1*), for validation purposes (Fig. 2A, bottom panels, black tracks).

ChIP-seq analysis of these same loci from material derived from the *Tal1*^{RER/RER} sample revealed either absence of the corresponding wild-type peaks (*Epb4.2*, *Gypa*, *Hba* HS-26, HS-21, HS-8 and *Hbb* HS1–3, *Aqp9*, and *Alad*), decreased binding (*Tal1*, *Prdx2*, *Trim10*, *Hba* HS-31), or unperturbed binding (*Hba* HS4, *Mme11*, and *Ssr1*) (Fig. 2A, red tracks).

Real-time PCR (qPCR) analysis of TALI-ChIP material confirmed binding of TALI on known and newly discovered genomic targets (Fig. 2B, black bars; data not shown). It also confirmed binding of the DNA-binding mutant form of TALI (Fig. 2B, red bars; data not shown). Importantly, this provided us with a means to validate the measurements of peak intensity, as RER/WT ratios calculated from ChIP-seq data correlated with those obtained from qPCR analyses. The “WT only” category represented by *Epb4.2*, *Gypa*, *Aqp9*, and *Alad* shows minimal TALI binding in *Tal1*^{RER/RER}

samples. *Tal1*, *Prdx2*, and *Trim10* from the RER/WT “0.1–0.8 ratio” category show a decrease in TALI enrichment by qPCR in *Tal1*^{RER/RER} samples. Finally, there is no statistically significant difference in enrichment on the *Mme11* and *Ssr1* loci when comparing *Tal1*^{WT/WT} and *Tal1*^{RER/RER} samples, thereby validating the category RER/WT 0.8–1.8.

In summary, all known erythroid TALI-bound sequences interrogated in our analysis were identified in our ChIP-seq experiment. This ChIP-seq analysis has yielded a high-resolution map of possibly all genomic sequences bound by TALI in erythroid precursors and comparison between *Tal1*^{WT/WT} and *Tal1*^{RER/RER} peaks gives a sense as to whether TALI is recruited directly or indirectly to any one binding site.

TALI candidate target genes are involved in regulatory and red cell-associated processes

Two-thousand-thirty-seven out of 2195 genes associated with sequences bound by wild-type TALI had Gene Ontology (GO) (Fig. 3A; Supplemental Tables 5,6). One-third of the genes are involved in transcription and signaling (16.8% and 15.8%, respectively). Close examination of the next set of GO categories (metabolism, transport, adhesion/migration, cytoskeleton, and redox processes, altogether accounting for 35.8% of genes associated with TALI-occupied DNA segments) highlighted genes coding for proteins involved, among other tissues, in red cell structures and functions, such as the heme pathway enzymes, ion/water channel proteins and solute carriers, proteins involved in membrane integrity, cell-cell interaction, and oxidative processes. The remaining categories comprised other general cellular processes, accounting for 15.8% of genes. This overview revealed the breadth of the transcriptional control exerted by TALI on red cell-specific processes.

To check whether the sequences that can be bound by the TALI^{RER} protein were associated with a functional subset of TALI's target genes, we have identified and compared high-level functional categories in both sets of genes using the Ingenuity Pathway Analysis software (Fig. 3B). The same significantly enriched categories were found in wild-type and *Tal1*^{RER/RER} samples ($1.1 \times 10^{-11} < P\text{-values} < 0.05$), indicating that TALI indirect DNA binding does not preferentially occur on specific subsets of targets. Cellular development, cell growth and proliferation, hematopoiesis, and cell death were among the most enriched categories.

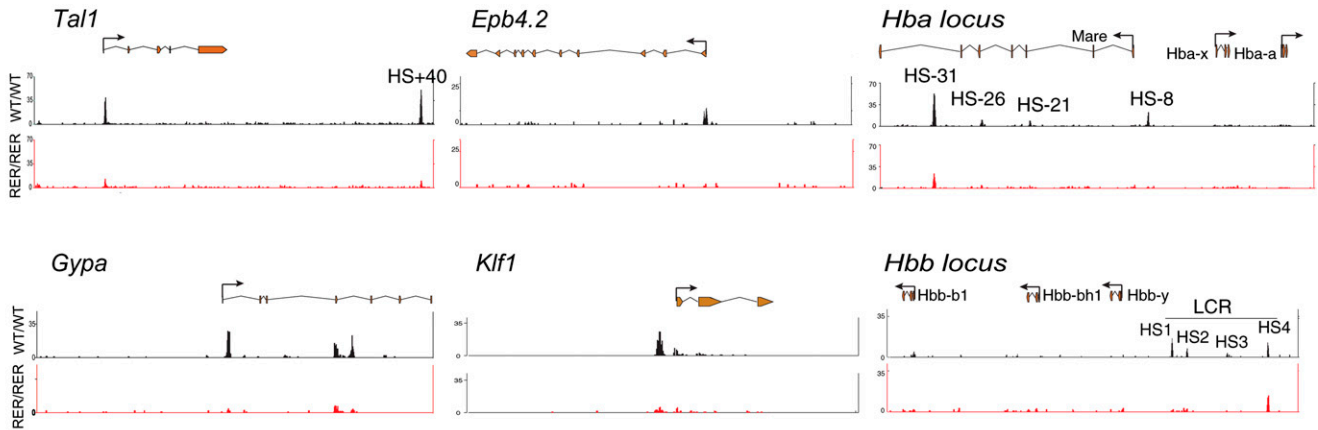
ChIP-seq data combined with gene expression analyses reveal functional locus occupancy by TALI

The impact of TALI direct DNA binding on gene expression was assessed in the same populations as those interrogated by ChIP-seq

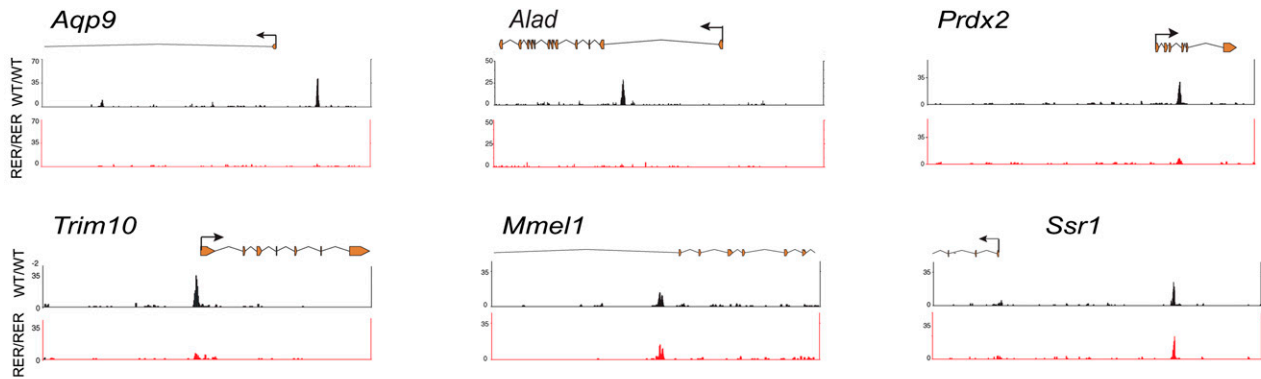
Figure 1. Detection of ChIP-seq peaks in *Tal1*^{WT/WT} and *Tal1*^{RER/RER} samples. (A) Outline of the experimental strategy. (B) Venn diagram showing that the peaks identified in material isolated from *Tal1*^{RER/RER} Ter119⁺ fetal liver cells (594 peaks, in orange circle) are a subset of the peaks identified in material isolated from *Tal1*^{WT/WT} cells (2994 peaks, in blue circle). Below, the peaks are divided into three categories: “WT only” when not detected in the *Tal1*^{RER/RER} sample; “0.1–0.8” or “0.8–1.8” according to the ratio of intensity between *Tal1*^{RER/RER} and the corresponding *Tal1*^{WT/WT} peaks. (C) TALI ChIP-seq peaks are displayed on two genomic loci (on chromosomes 8 and 7, top track) on GBrowse. For both sets of samples (*Tal1*^{WT/WT} and *Tal1*^{RER/RER}), the sequencing reads, identified as peaks, are mapped onto the chromosome view along with their coordinates and visualized along the sequence in GBrowse. The peaks exclusively detected from the wild-type sample (*Tal1*^{WT/WT} Peaks) are labeled “WT only.” All the peaks detected from the mutant population (*Tal1*^{RER/RER} Peaks) correspond to genomic locations also identified as peaks in the wild-type population. For those peaks, the ratio of intensity between wild-type and mutant samples is shown (RER/WT ratios 0.1–0.8 or 0.8–1.8). (D) The distribution of the 594 peaks detected in the *Tal1*^{RER/RER} sample (RER/WT ratios 0.1–0.8 and 0.8–1.8) is compared with that of their corresponding peaks (i.e., detected at the same position) in the *Tal1*^{WT/WT} sample, according to their intensities. The “WT only” peaks are not shown. (E) Genomic distribution in percentages of the *Tal1*^{WT/WT} peaks with respect to gene loci. In gray, exons; position of intron 1 is shown; thin lines on either side of the locus represent upstream and downstream flanking sequences; the arrow shows position of the transcription start site (TSS). (F) Distribution in percentages of the *Tal1*^{WT/WT} peaks as a whole (All peaks) and after fractionation according to their requirement for direct DNA-binding activity (WT only, ratios 0.1–0.8 or 0.8–1.8), with respect to the three main genomic locations, as indicated on the graph.

A

Known functional targets



New genomic targets



B

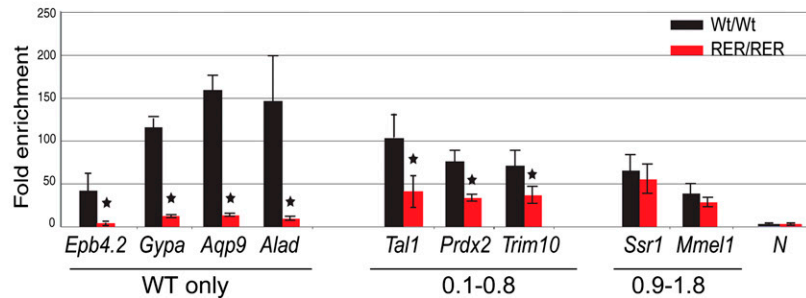


Figure 2. Profile of TAL1 binding on chosen loci. (A) Selected known functional or novel genomic targets of TAL1 are represented. For each locus are shown (from top to bottom): the RefSeq annotation of the gene or part of the gene (orange, exons; thin lines, introns; arrow, position of the TSS); the ChIP-seq profiles in Ter119⁻ populations from *Tal1*^{WT/WT} (black tracks) and *Tal1*^{RER/RER} (red tracks) fetal liver cells. (B) Real-time PCR analysis of anti-TAL1 ChIP on selected loci. Chromatin derived from Ter119⁻ populations from *Tal1*^{WT/WT} and *Tal1*^{RER/RER} fetal liver culture cells was immunoprecipitated using anti-TAL1 antibodies and the loci indicated on the graph analyzed by real-time PCR. The y-axis represents the enrichment over input DNA, normalized to a control sequence in the *Gapdh* gene. N, negative control. Error bars, ±1 SD, from at least three independent experiments (*P < 0.01). Below the graph are shown the categories the peaks belong to, as detected by ChIP-seq.

(*Tal1*^{WT/WT} and *Tal1*^{RER/RER} Ter119⁻ fetal liver erythroid populations) using microarray analysis (Fig. 3C). Five-hundred-eleven differentially expressed genes were identified (Supplemental Table 7), of which 248 (49%) and 263 (51%) were up- and down-regulated in mutant cells, respectively (Fig. 3D).

To identify direct targets of TAL1, we focused on the differentially expressed genes (511 genes) whose genomic loci contained sequences present in the ChIP-seq data set of 2195 genes. Eighty-three genes were in this intersection and likely represent direct TAL1 binding target genes (Fig. 3E; Supplemental Table 8).

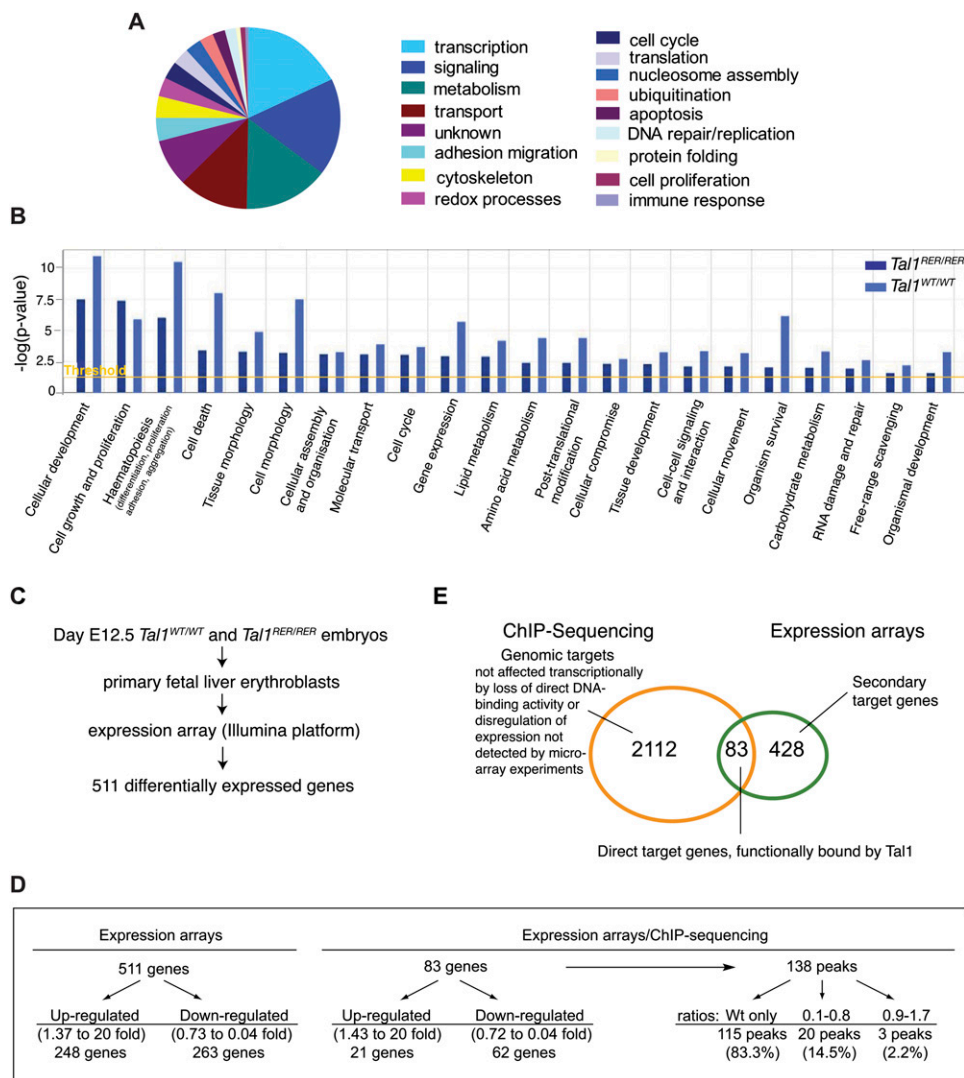


Figure 3. Combining ChIP-seq data with gene expression analyses. (A) Pie chart showing the distribution of the genes identified as candidate targets of TAL1 in $Tal1^{WT/WT}$ fetal liver cells according to their GO. (B) Highly enriched functional categories ($1.1 \times 10^{-11} < P\text{-values} < 0.05$) were identified in the gene sets characterized from $Tal1^{WT/WT}$ and $Tal1^{RER/RER}$ samples using Ingenuity software. The Ingenuity Knowledge Base served as background population and the Fisher's exact test was used. The threshold corresponds to a P -value of 0.05. (C) Microarray analysis: outline of the experimental strategy. (D) Characteristics of genes revealed by expression arrays (511) and those identified in the intersection with ChIP-seq data (83). See text for details. (E) Venn diagram showing the overlap between the genes detected by ChIP-seq and those revealed by expression array.

The remaining 428 differentially expressed genes could either represent secondary targets or be bound by TAL1 below the level of detection or at distal elements not identified by the nearest gene approach. Expression of the remaining TAL1-bound 2112 genes is not altered by loss of direct TAL1 DNA-binding activity (Fig. 3E). This may occur for a number of reasons. First, 19.1% of TAL1's genomic targets are bound by the DNA-binding mutant form of TAL1 (peak categories RER/WT 0.1–0.8 and 0.8–1.8) and may not be transcriptionally sensitive to the DNA-binding mutation. Second, redundancy between TAL1 and another hematopoietic-specific bHLH protein, such as LYL1, could explain the lack of perturbation of expression of some targets. *Lyl1* and *Tal1* show very similar patterns of expression in hematopoietic tissues, including erythroid cells (Visvader et al. 1991; Giroux et al. 2007); they were also recently found to have redundant functions in hematopoietic stem cells (Souroullas et al. 2009). Third, TAL1 might not exert

a function at all the sites it interacts with, or might bind genomic sites to prepare them for activation at later stages of differentiation, upon subsequent recruitment of transcriptional regulators. Finally, technical limitations associated with expression microarrays (such as probe design and hybridization conditions) could also lead to failure to detect differentially expressed genes and contribute to the underrepresentation of genes transcriptionally affected.

Further characterization of these 83 putative direct TAL1 target genes revealed that they were associated with 138 peaks of TAL1 binding in the ChIP-seq database (Fig. 3D; Supplemental Table 8). As expected, the large majority of these peaks (135/138) were affected in the $Tal1^{RER/RER}$ sample (Fig. 3D); they were either absent (115/138, 83%, WT only category), or reduced (20/138, 14.5%, 0.1–0.8 RER/WT ratios). The genomic distribution of these peaks was very similar to the whole set of TAL1 genomic targets (Supplemental Table 4, last column). Finally, the distribution of

these 83 differentially expressed direct TAL1 target genes in the GO categories was similar to the whole TAL1 gene list (Supplemental Table 9). In summary, the 83 genes presented features very similar to those of the overall set of genes bound by TAL1.

Strikingly, 62 out of 83 (75%) of the differentially expressed genes were down-regulated in *Tal1^{RE/RE}* erythroid populations compared with control cells, whereas 25% (21 out of 83) were up-regulated (Fig. 3D). The high proportion of down-regulated genes suggested that direct DNA binding is required for TAL1 to preferentially exert activator function in red cells. The fold change in expression did not exceed 2.5 for the majority of the genes (63 out of 83 genes; Supplemental Table 8). Only a few genes were dramatically down-regulated (*Lpl*, *Txn1l*, *Aqp9*, *Prdx2*, *Slc4a1*) or up-regulated (*Trib2*, *Dapp1*, *Hbb-b2*) (by 3.5–20-fold). There was no difference in the genomic localization of the peaks associated with activated or repressed genes, when all peaks (138) or only the nearest peak associated to each gene (83) were considered (data not shown).

An integrated approach identifies TAL1's core transcriptional red cell network

We set out to identify targets of TAL1 with functions in erythropoiesis that may contribute to the erythroid phenotype of the *Tal1^{RE/RE}* mouse when their expression is perturbed. From the list of genes identified through ChIP-seq, 80 loci were selected that (1) contained sequences bound by wild-type TAL1, (2) showed reduced or absence of binding by the DNA-binding mutant form of TAL1, and (3) have functions relevant or potentially relevant to red cell biology, based on previously reported functional in vitro and in vivo studies (Table 1). Importantly, when mouse models of these genes are available, the reported phenotype has similarities with the TAL1 DNA-binding mutant mice.

To this first level of analysis, we have incorporated our gene expression data. Out of these 80 genes, 23 showed deregulated expression (Table 1). Among them are some of the most transcriptionally affected genes (such as *Aqp9*, *Slc4a1*, *Prdx2*, *Lpl*, *Hbb-b2*, *Dapp1*). Confirming our hypothesis that microarray analyses might only detect a fraction of the genes transcriptionally affected, we also found, by qPCR, that expression levels of selected red cell-specific targets of TAL1, not present in the microarray list, were deregulated in *Tal1^{RE/RE}* cells. Three of these are included in Table 1 (*Tspan33*, *Trim10*, and *Ank1*).

These 80 genes can be divided into six main groups: organization of the cell membrane and cytoskeleton, signaling, redox processes, the heme biosynthetic pathway, as well as transcription and lipid metabolism. Figure 4 gives a schematic representation of the genes that, we believe, form the core of the TAL1 transcriptional red cell network and their cellular localization. Below, we describe some of these genes and associated processes (see Table 1 for references).

Membrane proteins and cytoskeleton organization

The red cell membrane is a highly specific structure characterized by a lipid bilayer anchored to a spectrin-based filamentous network of skeletal proteins. Numerous transmembrane proteins serve diverse functions such as transport and adhesion/migration (for review, see Mohandas and Gallagher 2008). Genes encoding 26 membrane-associated proteins are candidate targets of TAL1 (Table 1; Fig. 5; data not shown). They are involved in processes including transport, formation, and function of the erythroblastic islands (Manwani and Bieker 2008) and cytoskeleton organization.

Heme pathway, heme complex: Redox processes, hypoxia

TAL1's target genes control the complex series of enzymatic steps involved in the heme biosynthetic pathway. TAL1 binding was detected at the proximal promoter or intronic regions of all the genes encoding enzymes in the pathway (Supplemental Fig. S1). TAL1 could also be involved in the transport of heme through the mitochondria and in the control of oxygen homeostasis (*Abcb10* and *Egln3*; Table 1).

Signaling: Transcription, cell cycle, proliferation

TAL1 may also control expression of proteins involved in key signaling and regulatory pathways in red cell survival, proliferation, and terminal maturation (Table 1). Examples are TSPAN33, a newly described erythrocyte *trans*-membrane protein that belongs to a protein family believed to function as organizers of membrane microdomains and supramolecular signaling complexes, and the main red cell-specific cytokine receptor, EPOR, as well as a number of molecules involved in this pathway (FOXO3, PIM1, DYRK3, and DAPK2) (Fig. 5; Table 1).

Key transcriptional regulators of erythroid development are among the candidate targets of TAL1. To cite but a few, all members of the pentameric complex (including TAL1 itself), LYL1, NFE2, ZFPM1 (FOG1), FLI1, SFPI1 (PU.1), CBFA2T3 (ETO2), E2F2, E2F4, TRIM10, and HMGN are some of the critical regulators that orchestrate erythroid cell commitment, proliferation, and differentiation (Table 1; Supplemental Fig. 2). Interestingly, some of these transcription factors are also known partners of TAL1, such as E2A, LMO2, LDB1, GATA1, and CBFA2T3 (Wadman et al. 1997; Schuh et al. 2005), or proteins interacting with partners of TAL1, such as ZFPM1 (partner of GATA1) (Tsang et al. 1997). These interactions underlie the complexity of genetic regulatory networks and emphasize the importance of feed forward and autoregulatory transcriptional loops (Swiers et al. 2006; Fujiwara et al. 2009).

A recent study identified 139 putative targets of TAL1 in a hemopoietic progenitor cell line (Wilson et al. 2009). More than half of the genes (80/139) are also targets of TAL1 in erythroid cells, thereby highlighting a conserved role for TAL1 at certain loci across hematopoietic development. These include genes coding for cytoskeleton proteins (*Epb4.1*), for proteins located in the plasma membrane (*Flt1*), and for nuclear proteins (*Cbfa2t3*, *Cebpe*, *E2f2*, *Gata2*, *Gfi1b*, *Fli1*, *Hhex*, *Nfe2*, *Runx1*). In addition to binding events common to the progenitor cell line and the fetal liver Ter119⁺ erythroid cells, additional TAL1-bound genomic segments can be detected in some of these loci in red cells (*Cbfa2t3*, *Nfe2*, and *Tal1* loci; Supplemental Fig. 2, black track, stars; data not shown) suggesting that TAL1 exerts distinct transcriptional control over the same targets in distinct biological contexts.

Novel potential players in erythropoiesis

Finally, potentially new players in erythroid biology have been identified, thereby demonstrating the strength of our comprehensive approach (Table 1; Fig. 5). MICALL2 is a protein associated with the cytoskeleton that has been implicated in adhesion and repulsion mechanisms. In light of a possible involvement of TAL1 in the regulation of the function of the erythroblastic islands, investigation of the functional significance of this molecule and its regulation by TAL1 in erythropoiesis is of interest. The lipoprotein lipase *Lpl* is a functional target of TAL1 (its expression is 20-fold down-regulated in *Tal1^{RE/RE}* cells), suggesting that this protein is likely to play a major role in maintaining the specific phospholipid composition of the red cell membrane bilayer. Further supporting

Table 1. Genes bound by TAL1 with known erythroid functions and with functions potentially relevant in erythropoiesis

Gene	Peak nos.	Fold change	Official full name	Processes	References
A. Genes bound by TAL1 with known erythroid functions					
Membrane proteins: cytoskeleton organization					
<i>Aqp1</i>	507		aquaporin 1 (Colton blood group)	Transport	Yang et al. 2000; Blanc et al. 2009
<i>Aqp3</i>	1337		aquaporin 3 (Gill blood group)	Transport	Roudier et al. 2002
<i>Aqp9</i>	38, 885, 2497	0.105 ^a	aquaporin 9, water and glycerol channel	Transport	Liu et al. 2007
<i>Slc2a3</i>	1040	1.43 ^a	solute carrier family 2	Transport (solute carrier)	Johnstone et al. 1998
<i>Slc4a1</i>	505, 585	0.2 ^a	solute carrier family 4, (erythrocyte membrane protein)	Transport (solute carrier)	Peters et al. 1996; Paw et al. 2003
<i>Rhag</i>	549		Rh-associated glycoprotein	Transport (gas exchanger)	Van Kim et al. 2006; Goossens et al. 2010
<i>Abcb6</i>	2101		ATP-binding cassette, sub-family B (MDR/TAP), member 6	Phospholipid transport	Daleke 2008
<i>Abca1</i>	1756		ATP-binding cassette, sub-family A (ABC1), member 1	Phospholipid transport	Hamon et al. 2000
<i>Abcc1</i>	1581		ATP-binding cassette, sub-family C (CFTR/MRP)	Phospholipid transport	Kamp and Haest 1998
<i>Ermap</i>	723		erythroblast membrane-associated protein	Adhesion (blood group)	Ye et al. 2000
<i>Maca</i>	2386		macrophage erythroblast attacher	Adhesion (function in erythroblastic island)	Hanspal et al. 1998
<i>Icam4</i>	217		intercellular adhesion molecule 4	Adhesion (function in erythroblastic island)	Lee et al. 2006
<i>Itga4</i>	1665, 2311	0.52 ^a	integrin, alpha 4	Adhesion	Sathyanarayana et al. 2007
<i>Cd47</i>	146, 855, 2902	0.53 ^a	CD47 molecule	Adhesion	Oldenborg et al. 2002
<i>Lama5</i>	2185	0.66 ^b	laminin, alpha 5	Adhesion	Gauthier et al. 2005; Wautier et al. 2007
<i>Span1</i>	22, 1197		spectrin, alpha, erythrocytic 1 (elliptocytosis 2)	Adhesion, migration	Bodine et al. 1984; Wanderssee et al. 2000
<i>Gypa</i>	114		glycophorin A (MNS and Wright blood group)	Cytoskeleton organization	Arimitsu et al. 2003
<i>Gypc</i>	16		glycophorin C (Gerlich blood group)	Transport	Chang and Low 2001
<i>Epb4.1</i>	64		erythrocyte membrane protein band 4.1	Cytoskeleton stability	Shi et al. 1999; Salomao et al. 2008
<i>Epb4.9</i>	601		erythrocyte membrane protein band 4.9 (dema1in)	Cytoskeleton organization	Khanna et al. 2002; Chen et al. 2007
<i>Epb4.2</i>	1217		erythrocyte membrane protein band 4.2	Cytoskeleton organization	Peters et al. 1999
<i>Ank1</i>	385	0.6 ^b	ankyrin 1, erythrocytic	Cytoskeleton organization	Rank et al. 2009
<i>Add1</i>	469		adducin 1 (alpha)	Cytoskeleton organization	Robledo et al. 2008
<i>Add2</i>	864		adducin 2 (beta)	Cytoskeleton organization	Chen et al. 2007; Robledo et al. 2008
<i>Add3</i>	901		adducin 3 (gamma)	Cytoskeleton organization	Sahr et al. 2009
<i>Stom</i>	185		stomatrin	Cytoskeleton organization	Zhu et al. 1999
Signaling					
<i>Epor</i>	44	0.7 ^a	erythropoietin receptor	Signaling	Menon et al. 2006; Sathyanarayana et al. 2008
<i>Pim1</i>	138, 892, 1549, 2710		pim-1 oncogene	Signaling (anti-apoptotic)	Sathyanarayana et al. 2008
<i>Dapk2</i>	246,651, 717,870,1059		death-associated protein kinase 2	Signaling, apoptosis	Wojchowski et al. 2006
<i>PIK3r1</i>	211, 546, 560, 1844		phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	Signaling (EpoR pathway), apoptosis	Wojchowski et al. 2006
<i>Ryk</i>	1778	1.47 ^a	RYK receptor-like tyrosine kinase	Signaling	Koritschoner et al. 1999
<i>Tspan33</i>	2	0.5 ^b	tetraspanin 33	Signaling	Heikens et al. 2007; Charrin et al. 2009
<i>Cd59b</i>	46		CD59 molecule, complement regulatory protein	Signaling	Qin et al. 2003, 2009
<i>Dusp1</i>	1351, 2565	1.64 ^a	dual specificity phosphatase 1	Signaling (MAP kinase)	Liu et al. 2003, 2005

(continued)

Table 1. Continued

Gene	Peak nos.	Fold change	Official full name	Processes	References
Redox processes, hypoxia <i>Egln3</i>	1025, 2061, 2139	0.59 ^a	egl nine homolog 3 (<i>C. elegans</i>)	Oxidation reduction (hypoxia)	Appelhoff et al. 2004; Stockmann and Fandrey 2006; Metzzen 2007
<i>Prdx2</i>	158	0.11 ^a	peroxiredoxin 2	Oxidation reduction, apoptosis	Lee et al. 2003; Low et al. 2007
Heme pathway, heme complex <i>Abcb10</i>	739, 1367	1.69 ^a	ATP-binding cassette, sub-family B (MDR/TAP), member 10	Transport (heme)	Shirihai et al. 2000
<i>Hbb-b2</i>	327, 403, 977 (Mare)	20 ^a	hemoglobin, beta 2	Transport	Nakajima et al. 1999; Trimborn et al. 1999
<i>Alas2</i>	164		aminolevulinic acid, delta-, synthase 2	Heme metabolism	Ponka 1997, 1999
<i>Alad</i>	141		aminolevulinic acid, delta-, dehydratase	Heme metabolism	Ponka 1997, 1999
<i>Hmbs</i>	109, 112	0.38 ^a	hydroxymethylbilan synthase (PBGD)	Heme metabolism	Ponka 1997, 1999
<i>Uros</i>	364, 488, 2426	0.67 ^a	uroporphyrinogen III synthase	Heme metabolism	Ponka 1997, 1999
<i>Urod</i>	427		uroporphyrinogen decarboxylase	Heme metabolism	Ponka 1997, 1999
<i>Cpox</i>	21		coproporphyrinogen oxidase	Heme metabolism	Ponka 1997, 1999
<i>Ppox</i>	30, 4232	0.6 ^a	protoporphyrinogen oxidase	Heme metabolism	Taketani et al. 1995; Ponka 1997, 1999
<i>Fech</i>	276		ferrochelatase (protoporphyrin)	Heme metabolism	Taketani et al. 1995; Ponka 1997, 1999; Lyoumi et al. 2007
Transcription, cell cycle, proliferation <i>Tal1</i>	69		stem cell leukemia	Transcription (regulation)	Porcher et al. 1999; Hall et al. 2003; Mikkola et al. 2003; Kassouf et al. 2008
<i>E2A</i>	2166		transcription factor 3 (Tcf3)	Transcription (regulation)	Wadman et al. 1997; Porcher et al. 1999; Massari and Murre 2000
<i>Lmo2</i>	35		LIM domain only 2	Transcription (regulation)	Warren et al. 1994; Wadman et al. 1997
<i>Ldb1</i>	2369		LIM domain binding 1	Transcription (regulation)	Wadman et al. 1997; Xu et al. 2003
<i>Gata1</i>	61		GATA binding protein 1	Transcription (regulation)	Weiss and Orkin 1995; Fujiwara et al. 1996; Wadman et al. 1997
<i>Lyl1</i>	3		lymphoblastic leukemia derived sequence 1	Transcription (regulation)	Capron et al. 2006; Chan et al. 2007; Souroullas et al. 2009
<i>Nfe2</i>	27		nuclear factor (erythroid-derived 2), 45kDa	Transcription (globin expression)	Shivadasani et al. 1997; Sawado et al. 2001; Loyd et al. 2003;
<i>Nfe2l2</i>	122		nuclear factor (erythroid-derived 2)-like 2	Transcription (redox homeostasis)	Lee et al. 2004
<i>Klf1</i>	206		Kruppel-like factor 1 (erythroid)	Transcription (regulation)	Drissen et al. 2005; Hodge et al. 2006
<i>Bach1</i>	1823		BTB/CNC homology 1, basic leu/zipper transcription fact. 1	Transcription (regulation)	Brand et al. 2004; Igarashi and Sun 2006
<i>Trim10</i>	108	0.26 ^b	tripartite motif-containing 10	Transcription (terminal differentiation)	Harada et al. 1999; Blaybel et al. 2008
<i>Sfp1</i>	423		SFFV proviral integration 1 (Sfp1)	Transcription (antagonist of erythropoiesis)	Stopka et al. 2005; Liew et al. 2006; Rimmele et al. 2007
<i>Fli1</i>	1269		Friend leukemia virus integration 1	Transcription (antagonist of erythropoiesis)	Pereira et al. 1999, Kawada et al. 2001, Starck et al. 2003
<i>Zfpm1</i>	89, 272, 1093, 1616, 2144, 2575, 2689		Friend of Gata-1	Transcription (partner of GATA1)	Tsang et al. 1997
<i>Cbfa2t3</i>	33		core-binding factor, runt domain, alpha subunit 2	Transcription (repressor of erythropoiesis)	Schuh et al. 2005; Goardon et al. 2006; Meier et al. 2006
<i>Runx1</i>	41, 140, 181, 301, 798, 1309, 1855, 2091		Runt related transcription factor 1	Transcription (regulation)	Wang et al. 1996a; Yokomizo et al. 2008
<i>Izkf1</i>	324	0.57 ^a	IKAROS family zinc finger 1 (Ikaros)	Transcription (regulation)	Wang et al. 1996b; Bottardi et al. 2009
<i>E2f2</i>	179		E2F transcription factor 2	Transcription (cell cycle progression)	Pilon et al. 2008
<i>E2f4</i>	1933		E2F transcription factor 4, p107/p130-binding	Transcription (proliferation)	Humbert et al. 2000

(continued)

Table 1. Continued

Gene	Peak nos.	Fold change	Official full name	Processes	References
<i>Max</i>	1241		MYC associated factor X	Transcription (regulation)	Canelles et al. 1997
<i>Ddit3</i>	1997		DNA-damage-inducible transcript 3	Transcription (regulation)	Cui et al. 2000
<i>Dyrk3</i>	467		dual-specificity tyrosine-phosphorylation regulated kinase 3	Transcription (EpoR pathway)	Geiger et al. 2001; Wojchowski et al. 2006; Bogacheva et al. 2008
<i>Foxo3</i>	1184, 1331		forkhead box O3	Transcription (pro-apoptotic)	Bakker et al. 2004; Marinkovic et al. 2007; Nguyen et al. 2009
<i>Sox6</i>	701		SRY (sex determining region Y)-box 6	Transcription (regulation)	Dumitriu et al. 2006; Yi et al. 2006; Cohen-Barak et al. 2007
<i>Hmgn</i>	172, 325	0.37 ^a	hemogen	Transcription (proliferation)	Li et al. 2004, 2007
B. Genes bound by TAL1 with functions potentially relevant in erythropoiesis					
Adhesion					
<i>Micall2</i>	54, 258		Mical-like 2	Cytoskeleton, adhesion	Nishimura and Sasaki 2008
Signaling					
<i>Dapp1</i>	696, 1012, 1956	3.57 ^a	dual adaptor of phosphotyrosine and 3-phosphoinositides	Signaling	Han et al. 2003; Richards et al. 2008
<i>Emilin2</i>	1049, 1907, 2543	2.22 ^a	elastin microfibril interfacer 2	Signaling, adhesion	Mongiati et al. 2007
Redox processes, hypoxia					
<i>P4ha2</i>	2204	0.57 ^a	prolyl 4-hydroxylase, alpha polypeptide II	Oxidation, reduction (hypoxia)	Hofbauer et al. 2003
<i>Sphk1</i>	392, 593, 903, 2591	0.34 ^a	sphingosine kinase 1	Oxidation reduction, (hypoxia)	Ader et al. 2008
Lipid metabolism					
<i>Lpl</i>	1885	0.04	lipoprotein lipase	Metabolism	Ven Murthy et al. 1996
<i>Prdss2</i>	556, 1230, 1916, 2911		phosphatidylserine synthase 2	Metabolism (phospholipid)	Banerjee et al. 2008
<i>Ddh1</i>	1554		DDHD domain containing 1	Metabolism (phospholipid)	Aoki et al. 2002
<i>Lpin2</i>	330, 2058		lipin2	Metabolism (phospholipid)	Grimsey et al. 2008

^aFold change in expression (RER/WT) as detected by microarray experiments.^bFold change in expression (RER/WT) as detected by qPCR.

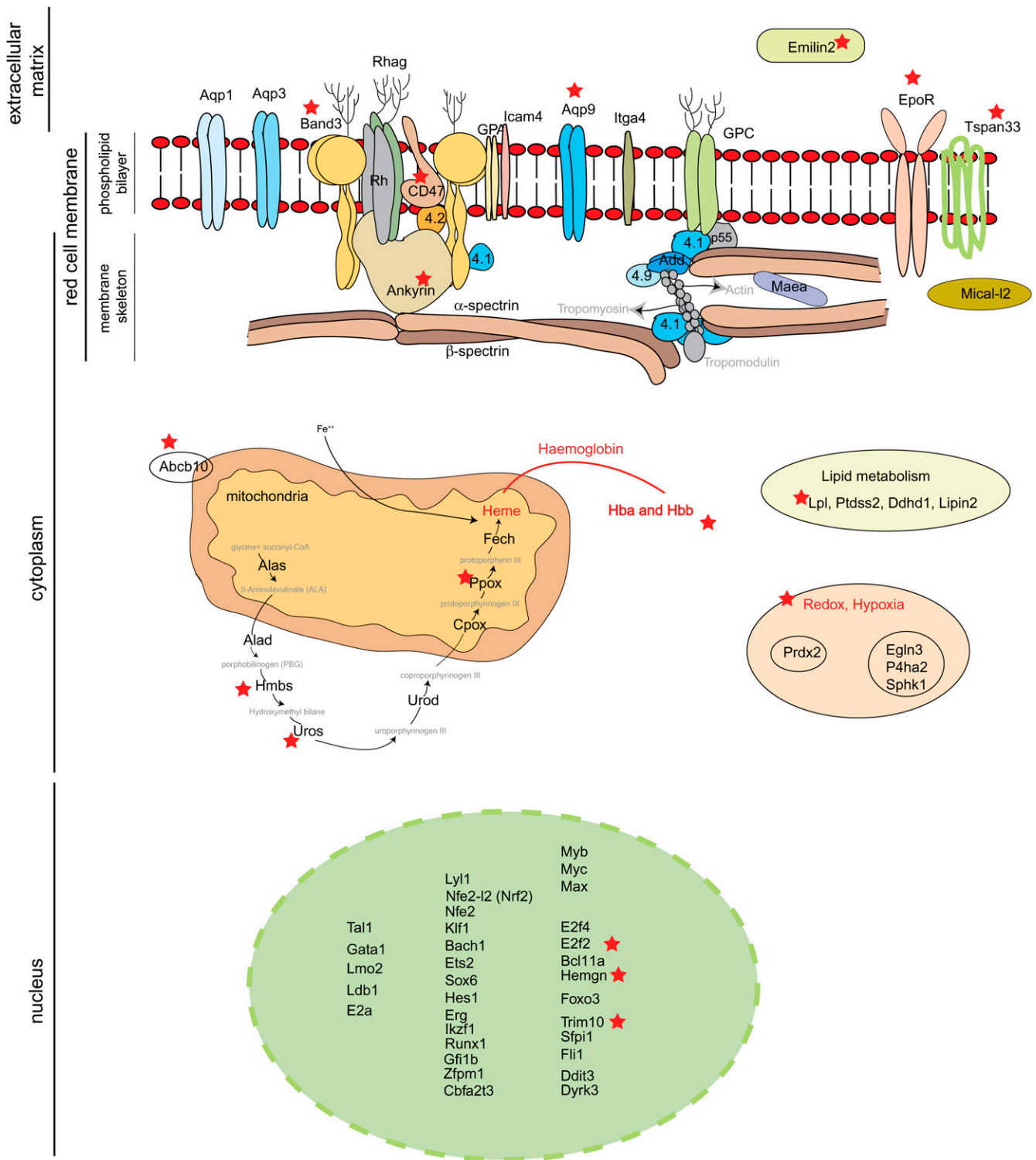


Figure 4. Schematic representation of selected pathways and molecules identified in this study as functional, direct targets of TAL1 in red cells. The red star indicates genes whose expression is perturbed in *Tal1^{RE/RE} Ter119⁺* fetal liver cells, when compared with wild-type controls.

this observation, TAL1 may also regulate expression of *Lipin2*, *Ptdss2*, and *Ddhd1*, all involved in phospholipid metabolism. Additional examples are *P4ha2* and *Sphk1*, which encode proteins involved in response to hypoxia; *Dapp1* (or *Bam32*),

which codes for a hematopoietic adaptor protein thought to control proliferation; finally, *Emilin2*, encoding an extracellular matrix protein that activates the extrinsic apoptosis pathway, is one of a number of apoptotic genes that are putative targets of TAL1 (see

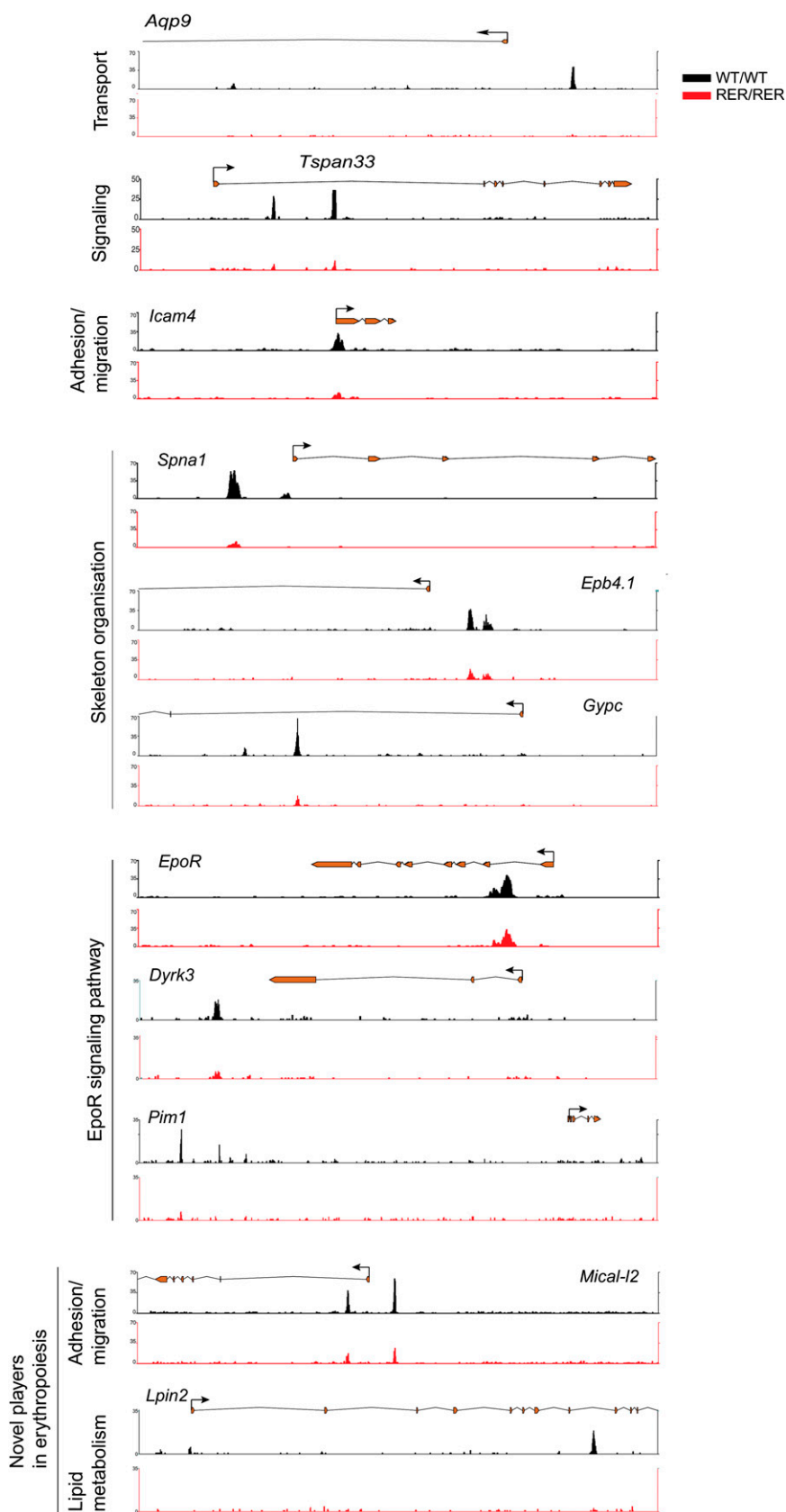


Figure 5. (Legend on next page)

Supplemental Tables 5, 6). Our data would suggest that *Emilin2* is normally repressed by TAL1 (Table 1), in agreement with TAL1's proposed role in forestalling apoptosis (Zeuner et al. 2003; Martin et al. 2004; Souroullas et al. 2009).

In conclusion, these few examples emphasize the strength of ChIP-seq in characterizing the whole repertoire of *cis*-elements bound by TAL1 and, therefore, candidate target genes. The comprehensive nature of this study provides a platform for assembling and testing networks of processes coordinated by TAL1 in red cells. It has also unveiled putative novel players in normal erythropoiesis and opened up new investigative routes for understanding inherited and acquired anemias.

De novo search for motifs underlying TAL1 peaks reveal E-boxes, GATA, and CACC sequences, as well as two novel motifs

To better understand the molecular mechanisms underlying TAL1's recruitment to its genomic targets in red cells, de novo motif finding was performed on the genomic sequences bound by TAL1. We first used the algorithm Weeder to search sequences underlying the 2994 peaks bound by wild-type TAL1 (Fig. 6A). GATA motifs were the most prevalent consensus sequences. Our data highlighted A_1 GATAA (and its extension to ${}^C_1G^A_1$ TGATAAG) as the *in vivo* preferred GATA sequence associated with TAL1 peaks, a subset of the WGATAR consensus motif initially identified *in vitro* (Orkin 1992). The same consensus site (WGATAA) was recently identified in two independent reports as the preferred sequence for occupancy by GATA1 (Yu et al. 2009; Zhang et al. 2009). WGATAA is therefore the dominant motif for GATA1.

After masking the GATA sites, E-box motifs were the next most prominent motif detected. CAGCTG was the TAL1 preferred consensus sequence, rather than the *in vitro* described variations of CAGATG (Hsu et al. 1994) or CAGGTG (Wadman et al. 1997).

After masking both GATA and E-box sequences, CAC motifs (consensus CAC CC) were identified, as well as a novel motif $CTGCC^A_1$ GNNG. This motif presents similarities with the first five positions of the recently described motif GCCAGC that is significantly enriched in GATA1-occupied DNA segments in erythroid cells (Zhang et al. 2009). It also

overlaps in a similar way with a consensus sequence present in occupied composite E-box/GATA sequences in red cells (${}^T/C/C/T/TG^G/T/C^C/G^A/TG^T/G$, the sequence common to all three motifs, GCCAG, is underlined) (Wozniak et al. 2008). Altogether, this further supports a strong correspondence between TAL1 and GATA1 occupancy in erythroid cells, as recently observed by Cheng et al. (2009).

No other motifs were identified, even when the three categories (WT only and RER/WT ratios 0.1–0.8 and 0.8–1.8) were interrogated individually (data not shown).

In a parallel analysis using the MEME algorithm (Fig. 6A), the enrichment of GATA and E-box motifs was confirmed. In addition, two longer motifs were characterized: the well-described E-box-GATA element [TATC(n_9)CAGCTG], and a new, related, composite motif consisting of a GATA sequence and the trinucleotide CTG, separated by 9 bp [CTG(n_9)GATA]. To confirm the existence of this composite motif, the 2994 TAL1-bound peaks were searched for WGATAA and sequences 5' and 3' of this motif examined. A significant proportion of the sequences were found to be associated to a CTG motif (Fig. 6A, bottom). The trinucleotide is half an E-box, inviting the hypothesis that such a sequence may allow less stringent binding of a TAL1 dimer with bound GATA factors stabilizing TAL1 recruitment. Supporting a biological meaning to this finding, a similar motif [CTG(n_{7-8})WGATA] was recently described in peaks co-occupied by GATA1 and TAL1 (Soler et al. 2010).

To measure the frequency of the motifs generated by Weeder and MEME, an in-house Perl script calculated the mean frequency of a motif per sequence over a region encompassing the peaks and flanking 100 bp and against a background distribution (see Methods; Table 2). Motif occurrence (number of peaks with the motifs) was also calculated (Table 2).

Although CAGCTG only appeared at a frequency of 0.18 motif/sequence, when relaxed to CAGNTG, E-boxes were the most frequent motifs (0.68 motif/sequence, $P = 0$; present in 39% of all peaks); the degenerated motif CANNTG was not found to be significantly enriched over the background distribution (frequency 0.71, $P = 0.97$). We therefore propose that the relaxed CAGNTG motif is TAL1-preferred E-box sequence in erythroid cells. The ${}^A/T$ -GATAA motif was found at a frequency of 0.53 motif/sequence (46% of all peaks) and confirmed to be dominant over ${}^A/T$ -GATAG (frequency 0.19, occurrence 17%). CAC motifs were also frequent (0.42 motif/sequence, occurrence 34%). The new motif, simplified to CTGCC ${}^A/T$ G, was present at a frequency of 0.16 (occurrence 14%). As for composite motifs, the novel sequence [CTG(n_9)GATA] dominated with a frequency of 0.19 motif/sequence (occurrence 13%) whereas, surprisingly, the E-box-GATA motif was associated with a frequency of only 0.03 motif/sequence (occurrence 4.5%).

Enrichment in novel motifs in a subset of peaks

We repeated the motif analysis on subsets of peaks that had been grouped by various parameters: peak localization, whether the peaks were associated with activated or repressed genes and according to the mechanisms of recruitment of TAL1 to DNA (Table 2).

The frequency or nature of DNA motifs did not vary according to peak localization (as defined in Fig. 1E). The enrichment in re-

cruitment of the DNA-binding mutant protein on distal *cis*-elements as opposed to proximal promoter regions (Fig. 1F) is not, therefore, due to specific arrangements of the motifs identified de novo.

We then looked at peak composition according to the transcriptional state of the gene. There was no substantial change in the overall frequency of E-boxes or ${}^A/T$ -GATAA motifs between active and repressed genes. However, our search of motifs in sequences underlying TAL1 peaks pointed to a fourfold enrichment of composite E-box/GATA sequences in the peaks associated with activated genes when compared with repressed genes (12% versus 3% of peaks, respectively). This is in agreement with recent data suggesting that TAL1 cooperates with GATA1 in red cells mainly to activate target loci (Cheng et al. 2009; Tripic et al. 2009; Yu et al. 2009). In addition, the novel motif CTGCC ${}^A/T$ G, shows a 2.3-fold increase in occurrence in TAL1 peaks associated with activated genes when compared with repressed genes (21% versus 9%), highlighting its potential role in regulatory regions of activated genes.

Finally, to better define the mechanisms of recruitment of TAL1 to its genomic targets, the sequences underlying the peaks in the three categories reflecting the ratios of direct versus indirect DNA binding were investigated (Table 2). As expected, the percentage of peaks containing the E-box motif CAGNTG was observed at the highest occurrence (42%) in the sequences underlying the "WT only" peak category. This decreased by 1.5-fold to 30% and 28% in the sequences where the RER/WT ratios were 0.1–0.8 and 0.8–1.8, respectively. In contrast, the ${}^A/T$ -GATAA motif occurrence increased by ~1.5-fold in the "WT only" sequences when compared with sequences with RER/WT ratios 0.1–0.8 and 0.8–1.8 (42%, 62%, and 55%, respectively). In agreement with this, the occurrence of the composite motif CTG(n_9)GATA was increased by two- to threefold in the sequences bound by TAL1^{RER} (10%, 27%, and 22%, in sequences associated to "WT only" and RER/WT ratios 0.1–0.8 and 0.8–1.8, respectively). Although of moderate amplitude, these changes reflect the two mechanisms of recruitment of TAL1 to DNA: The frequency of E-boxes is the highest when TAL1 direct DNA-binding activity is exclusively employed ("WT only" category) and the presence of GATA motifs alone or in the new composite element becomes more frequent in the categories where TAL1 recruitment requires other DNA-binding proteins to be tethered to DNA. Investigating what binds the sequence CTG(n_9)GATA in addition to GATA proteins would give insight into what makes this composite motif more tolerant of a defective DNA-binding form of TAL1. Finally, motif overrepresentation analysis was performed on all peak categories using consensus sites associated to blood development-specific factors including ETS factors (SFPI1, SPI1, FLI1), NFE2, RUN1, GFI1B, and SP1. No specific enrichment was detected for any of the motifs in any of the categories studied (data not shown).

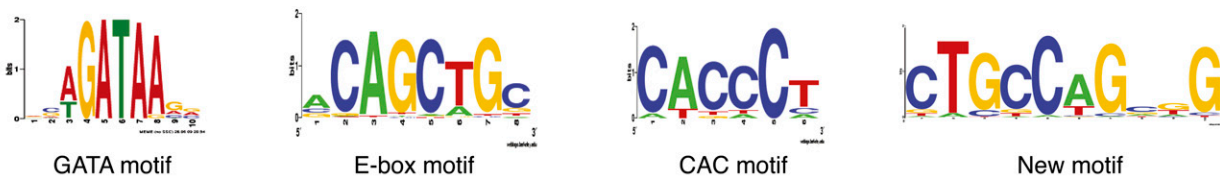
TAL1's binding is required prior or simultaneously to that of GATA1 at sites of co-occupancy

As the frequency of GATA1 binding site motifs increases when TAL1 recruitment requires other DNA-binding proteins to be tethered to DNA, we hypothesized that GATA1 might be the main protein to recruit TAL1 to its target sequences in *Tal1*^{RER/RER} cells. To test this, we performed GATA1 ChIP on selected loci, from material isolated from *Tal1*^{WT/WT}

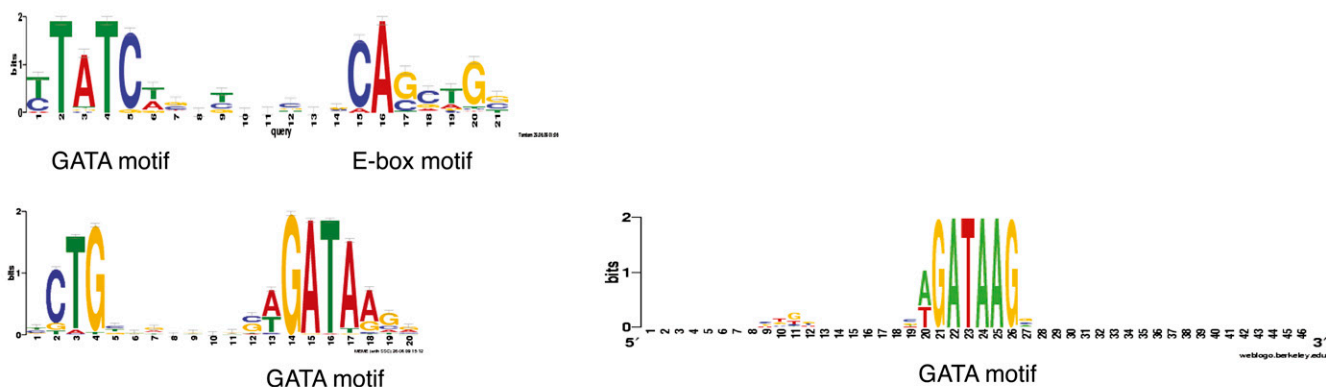
Figure 5. Profile of TAL1 binding on gene loci involved in red cell-specific processes or with functions potentially relevant in erythropoiesis. For each locus are shown (from top to bottom): the RefSeq annotation of the gene or part of the gene (orange, exons; thin lines, introns; arrow, position of the TSS); the ChIP-seq profiles in *Tal1*^{WT/WT} (black tracks) and *Tal1*^{RER/RER} (red tracks) cells. GO biological processes are indicated at the left of the figure.

A

Weeder search



Meme search



B

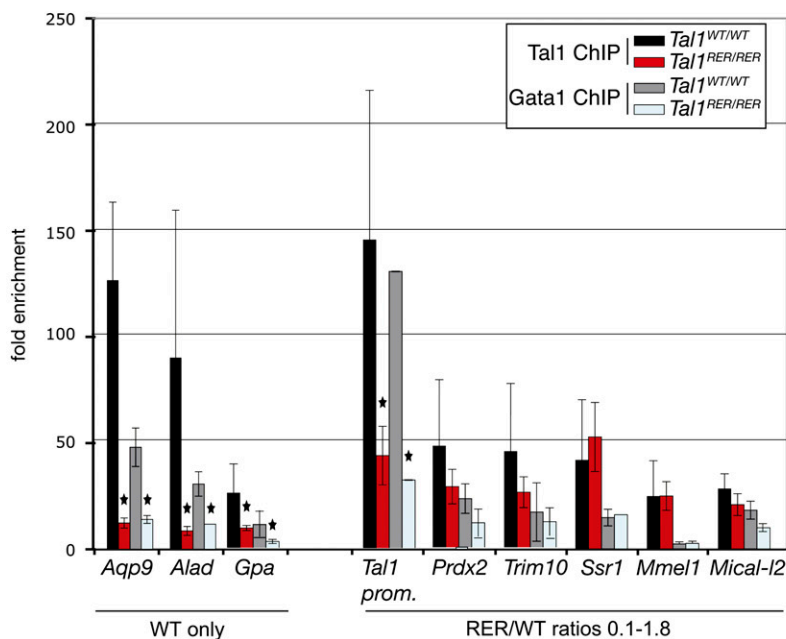


Figure 6. DNA motifs underlying the TAL1 peaks. (A) Logos representing the motifs identified in the sequences underlying the TAL1 peaks using de novo Weeder and Meme searches. (B) Real-time PCR analysis of anti-TAL1 and anti-GATA1 ChIP on selected loci. Chromatin derived from Ter119⁺ populations from *Tal1*^{WT/WT} and *Tal1*^{RER/RER} fetal liver culture cells was immunoprecipitated using anti-TAL1 and anti-GATA1 antibodies and the loci indicated on the graph analyzed by real-time PCR. The y-axis represents the enrichment over input DNA, normalized to a control sequence in the *Gapdh* gene. Error bars, ± 1 SD from at least three independent experiments (* $P < 0.01$). Below the graph are shown the categories the genes belong to, as detected by ChIP-seq.

Table 2. Motif frequency and motif occurrence

Motifs	Peak localization							
	All peaks 2994 peaks		Proximal promoter and first intron 1003 peaks		Body of gene 746 peaks		>5 kb upstream of and downstream from gene 1245 peaks	
	Frequency	Occurrence	Frequency	Occurrence	Frequency	Occurrence	Frequency	Occurrence
CAGCTG	0.18 (0)	16%	0.21 (0)	19%	0.22 (0)	20%	0.26 (0)	23%
CAGNTG	0.68 (0)	39%	0.86 (0)	46%	0.88 (0)	49%	1.03 (0)	54%
CANNTG	0.71 (0.968)	51%	0.96 (0)	60%	0.96 (0)	61%	1.12 (0)	67%
GATAG	0.16 (0.9)	21%	0.26 (0)	29%	0.23 (0.001)	25%	0.24 (0)	29%
GATAA	0.6 (0)	50%	0.77 (0)	60%	0.79 (0)	59%	0.89 (0)	65%
^A _T GATAA	0.53 (0)	46%	0.67 (0)	55%	0.71 (0)	55%	0.78 (0)	61%
^A _T GATAG	0.19 (0)	17%	0.27 (0)	23%	0.23 (0)	20%	0.3 (0)	26%
CACCC	0.42 (0)	34%	0.68 (0)	48%	0.59 (0)	44%	0.6 (0)	45%
CTGCC ^A _T G	0.16 (0)	14%	0.22 (0)	19%	0.21 (0)	18%	0.23 (0)	20%
CANNTG(9)GATA	0.03 (0)	4.50%	0.03 (0)	6%	0.03 (0)	4.90%	0.04 (0)	5.90%
CTG(9)GATA	0.19 (0)	13%	0.26 (0)	16%	0.25 (0)	17%	0.25 (0)	16%

Motifs	Gene expression				RER/WT ratios					
	Activated 104 peaks		Repressed 34 peaks		WT only 2400 peaks		0.1–0.8 457 peaks		0.8–1.8 137 peaks	
	Frequency	Occurrence	Frequency	Occurrence	Frequency	Occurrence	Frequency	Occurrence	Frequency	Occurrence
CAGCTG	0.29 (0)	28%	0.17 (0)	18%	0.18 (0)	17%	0.15 (0)	13%	0.17 (0)	17%
CAGNTG	1.15 (0)	57%	0.79 (0)	47%	0.72 (0)	42%	0.51 (0)	30%	0.5 (0)	28%
CANNTG	1.29 (0)	70%	1.17 (0)	70%	0.76 (0.6)	53%	0.53 (1)	41%	0.49 (1)	36%
GATAG	0.26 (0)	28%	0.29 (0)	26%	0.16 (0.98)	20%	0.16 (0.98)	24%	0.21 (0.05)	20%
GATAA	0.96 (0)	68%	1.06 (0)	82%	0.55 (0)	46%	0.83 (0)	67%	0.77 (0)	60%
^A _T GATAA	0.84 (0)	64%	0.97 (0)	79%	0.48 (0)	42%	0.73 (0)	62%	0.69 (0)	55%
^A _T GATAG	0.25 (0)	20%	0.26 (0)	26%	0.18 (0)	16%	0.23 (0)	21%	0.2 (0)	18%
CACCC	0.68 (0)	44%	0.79 (0)	41%	0.42 (0)	33%	0.46 (0)	36%	0.37 (0)	29%
CTGCC ^A _T G	0.27 (0)	21%	0.09 (0.001)	9%	0.15 (0)	14%	0.21 (0)	18%	0.12 (0)	10%
CANNTG(9)GATA	0.07 (0)	12%	0 (0.94)	3%	0.03 (0)	4.80%	0.02 (0)	3.50%	0.007 (0)	2.20%
CTG(9)GATA	0.22 (0)	17%	0.29 (0)	15%	0.16 (0)	10%	0.36 (0)	27%	0.32 (0)	22%

Motif frequency, number of motifs/number of peaks; *P*-values are in parentheses. Motif occurrence, percentage of peaks with the motif. The values in boldface are discussed in the text.

and *Tal1*^{RER/RER} samples (Fig. 6B). In agreement with our hypothesis, we observed binding of GATA1 on all the sequences able to recruit the TAL1^{RER} protein (peaks belonging to the RER/WT ratios 0.1–1.8) both in wild-type and *Tal1*^{RER/RER} cells (with the exception of one locus, *Mme11*). However, the levels of enrichment of GATA1 on these loci in wild-type and *Tal1*^{RER/RER} cells paralleled that of TAL1; for example, decreased levels in TAL1 enrichment at the *Tal1* promoter (*Tal1* prom.) in *Tal1*^{RER/RER} cells correlated with a proportional decrease in enrichment of GATA1 binding; little or no variation in TAL1 binding on sequences such as *Prdx2*, *Trim10*, *Ssr1*, and *Mical-12* in *Tal1*^{RER/RER} cells correlated with similar changes in GATA1 enrichments. As *Gata1* expression is not affected in *Tal1*^{RER/RER} cells (Kassouf et al. 2008), these data suggested that recruitment or stabilization of GATA1 on DNA relies on the presence of TAL1. Supporting these observations, GATA1 binding was severely affected at sequences not able to recruit TAL1 in *Tal1*^{RER/RER} cells (WT only, *Aqp9*, *Alad*, *Gpa* loci).

Therefore, our analysis of transcription factor recruitment in *Tal1*^{RER/RER} cells shows that, when co-localizing on target sequences, TAL1 and GATA1 might cooperate and stabilize each other's binding. This does not exclude the possibility that TAL1 is recruited to DNA by other, yet unidentified mechanisms. Analyses of the recruitment of these factors on "WT only" sequences suggests that TAL1's binding is necessary prior or simultaneous to that of GATA1. This agrees with earlier studies performed on the *Hba* locus, showing that TAL1 is present on the locus in early hema-

topoietic progenitors, before binding of GATA1 is firmly established in erythroid precursors (Anguita et al. 2004). Whether TAL1 can open compacted chromatin to allow subsequent recruitment of additional transcriptional regulators, thereby acting as a "pioneer" transcription factor (Lupien et al. 2008; Sekiya et al. 2009), remains to be investigated.

Conclusion

In conclusion, our attempt to elucidate the mechanisms of recruitment of TAL1 to DNA emphasized the robust and well-characterized observation of an E-box/GATA interplay at the genome-wide scale. The functional relationship between TAL1 and GATA1 is, however, likely to be quite complex and will need to be further dissected mechanistically. Novel motifs [CTGCC^A_TG and CTG(*n*₉)GATA] were described that remain to be studied functionally. This will help distinguish activated from repressed sequences and define which elements, in addition to GATA and CACC sites, help recruit the TAL1 DNA-binding mutant. However, as previously reported (Wozniak et al. 2008; Steiner et al. 2009), these studies have limitations; chromosomal environment and chromatin structure, dynamic recruitment of additional trans-acting factors, DNA looping, and influences of long-range regulation are elements to take into account when analyzing DNA/protein interactions that cannot be appreciated solely through ChIP-seq and motif prediction analyses. The biological intricacies

of tissue development and underlying molecular dynamics are a challenge to capture at the genome-wide scale.

Methods

Cell culture

Fetal liver cells from day E12.5 *Tal1*^{WT/WT} and *Tal1*^{REK/REK} embryos were expanded for 3 d, and Ter119⁻ erythroid progenitors were purified as previously described (von Lindern et al. 2001; Schuh et al. 2005).

ChIP assay

Anti-TAL1 and anti-GATA1 ChIP assays were performed on chromatin prepared from *Tal1*^{WT/WT} and *Tal1*^{REK/REK} Ter119⁻ erythroid purified progenitor populations as described (Schuh et al. 2005). Anti-TAL1 antibody has been described previously (Porcher et al. 1999). Anti-GATA1 antibody was from Abcam.

ChIP-seq sample preparation, library construction, and data processing

Tal1^{WT/WT} and *Tal1*^{REK/REK} anti-TAL1 and "no antibody" (input controls from both *Tal1*^{WT/WT} and *Tal1*^{REK/REK} cells) ChIP DNA were processed for Illumina high-throughput sequencing at the Center for Biomix, Erasmus MC, Rotterdam. Linker annealing, amplification, and gel purification were performed according to Illumina protocol. Data analysis was undertaken by the Computational Biology Research Group CBRG (Oxford University). ChIP-seq data sets have been submitted to the Gene Expression Omnibus (GEO) database under accession number GSE18720.

Approximately 12 million reads of 35 bp each were produced from *Tal1*^{WT/WT} and *Tal1*^{REK/REK} anti-TAL1 and "no antibody" ChIP DNA. Sequences were mapped to the repeat-masked reference mouse (build m37) genome using MAQ (Li and Durbin 2009). Repeat masking included simple, complex, and ribosomal repeats using data from the UCSC Table Browser "rmsk," and was done to avoid complications of full or partial multiple mapping of sequences when using the MAQ program which usually randomly maps such sequences to a single position in the genome. Approximately 4 million uniquely mapped reads were counted for every nucleotide position. The data were displayed on the Generic Genome Browser (Gbrowse, <http://tinyurl.com/SCLtargets>) (Stein et al. 2002). Peak height (as seen in GBrowse) reflected the number of sequences that map to the genomic region. Cisgenome (Ji et al. 2008), using the two-sample method (which allows comparison of antibody versus no antibody data) was used to call and quantify the peaks with an FDR cutoff of 0.1, positive/negative background ratio of 0.5, and minimum read number per window of 3 over a window size of 100. Quality filtering was performed by visual inspection of the peaks in GBrowse. An arbitrary cutoff was set up to reduce the noise level. To confirm that no potentially relevant peaks had been removed, a motif search was run on these peaks (as described below); no significant enrichments were observed. The identified peaks were associated to the nearest RefSeq genes to aid investigation of the data.

The genomic sequences of the Cisgenome-determined peaks were extracted using a custom in-house PERL scripts and analyzed for overrepresented sequence motifs using Weeder version 1.3 (Pavesi et al. 2004) and MEME (Bailey and Elkan 1994). Sequence logos of the motifs were generated using WebLogo (Crooks et al. 2004).

The frequency of motif occurrence (mf) was tested for significance against a background distribution which was generated

by drawing 1000 regions *S* times from a repeat masked version of the mm9 genome. In each case, the random peaks were twice the size of the real TAL1 peaks (200 bp versus an average of 107 bp) and matched for repeat content. For each round of sampling (*s*), the motif frequency $mf^{(s)}$ was calculated, and *P* was calculated by:

$$P^* = \sum_{s=1}^S \frac{I(|mf^{(s)}| > mf)}{S},$$

where *I* is 1, if the condition inside the parentheses is true. In this case *S* was set to 1000.

To test the applicability of the background model, sequences of 200 bp were isolated 500 bp and 1 kb upstream of and downstream from the TAL1 peaks, again controlling for useable sequence content. The frequency of the motifs within these data sets is now not significantly different from the background model.

All scripts for the analysis and Makefiles that were used to run the peak calling pipeline are available on request.

Quantitative real-time PCR analysis

For expression analysis, RNA was extracted from *Tal1*^{WT/WT} and *Tal1*^{REK/REK} Ter119⁻-purified erythroid progenitors using the RNeasy Micro RNA isolation kit (Qiagen), DNase-treated (Qiagen), and cDNA was synthesized using the Sensiscript kit (Qiagen). Primers for *Ank1*, *Tspan33*, and *Trim10* were designed using MacVector (MacVector, Inc.). SYBR Green-based quantitative qPCR (ABI SYBR Green PCR master mix, Applied Biosystems Inc.) was performed on three independent populations. Samples were analyzed in duplicates using an ABI Prism 7000 sequence detection system (Applied Biosystems Inc.). Data were normalized relative to *Gapdh*.

For ChIP experiments, primers and 5'-6-carboxyfluorescein-3'-6-carboxy tetramethylrhodamine-labeled probes were selected from unique sequences in the *Epb4.2*, *Gypa*, *Hba*, and *Hbb* loci using Primer Express (Kassouf et al. 2008). Primers for *Aqp9*, *Alad*, *Tal1*, *Prdx2*, *Trim10*, *Ssr1*, and *Mme11*, designed using MacVector, were used with SYBR Green PCR master mix (Applied Biosystems Inc.) for ChIP quantitation. Input and immunoprecipitated material were analyzed in duplicates relative to a sequence in the *Gapdh* locus as previously described (Anguita et al. 2004) on three independent *Tal1*^{WT/WT} and *Tal1*^{REK/REK} samples. Primers for negative points were selected on all the loci, and "no antibody" ChIP reactions were used as controls. Primer sequences are available upon request.

Expression microarray

Expression profiling was performed using Sentrix Mouse-6 Expression BeadChip arrays from Illumina (Illumina Inc.) on three independent *Tal1*^{WT/WT} and *Tal1*^{REK/REK} Ter119⁻-purified erythroid progenitor populations. RNA was extracted using RNeasy (Qiagen) and assessed for integrity using the Agilent Bioanalyzer 2100 (Agilent Technologies). All samples presented RNA integrity (RIN) scores above 9.5. Samples were then processed for array hybridization and data accumulation at the Wellcome Trust Center for Human Genetics, Genomics Group. In brief, amplification was performed using the Illumina TotalPrep RNA Amplification kit (Ambion) according to the manufacturer's instructions. Amplified cRNA was hybridized to the BeadChip arrays according to the manufacturer's guidelines and detected with Fluorolink Streptavidin-Cy3 (Amersham Biosciences). The raw intensity values obtained for the scanned array images were compiled using Illumina BeadStudio. The data were filtered so that any probe with a detection score of <0.95 across all samples was removed from the analysis prior to log transformation (base2) and quantile

normalization. Differentially expressed genes were identified using *limma* (Smyth 2004) for R. Expression array data sets have been submitted to the Gene Expression Omnibus (GEO) database under accession number GSE21877.

Acknowledgments

We thank Wilfred van Ijcken from Erasmus MC Centre for Biomics for the Illumina data. We thank the Genomic Service (The Wellcome Trust Centre of Human Genetics, Oxford) for the Illumina gene expression analyses. We thank D. Higgs for critical reading of the manuscript. M.T.K. was funded by the Leukaemia Research Fund. P.V. acknowledges funding from the MRC Disease Team Award, the MRC Molecular Haematology Unit, and the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health's NIHR Biomedical Research Centres. This work was supported by the Medical Research Council.

References

- Ader I, Brizuela L, Bouquerel P, Malavaud B, Cuvillier O. 2008. Sphingosine kinase 1: A new modulator of hypoxia inducible factor 1 α during hypoxia in human cancer cells. *Cancer Res* **68**: 8635–8642.
- Anderson KP, Crable SC, Lingrel JB. 1998. Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J Biol Chem* **273**: 14347–14354.
- Anderson KP, Crable SC, Lingrel JB. 2000. The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. *Blood* **95**: 1652–1655.
- Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WG, 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.
- Aoki J, Nagai Y, Hosono H, Inoue K, Arai H. 2002. Structure and function of phosphatidylserine-specific phospholipase A1. *Biochim Biophys Acta* **1582**: 26–32.
- Appelhoff RJ, Tian YM, Raval RR, Turley H, Harris AL, Pugh CW, Ratcliffe PJ, Gleadle JM. 2004. Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. *J Biol Chem* **279**: 38458–38465.
- Arimitsu N, Akimitsu N, Kotani N, Takasaki S, Kina T, Hamamoto H, Kamura K, Sekimizu K. 2003. Glycophorin A requirement for expression of O-linked antigens on the erythrocyte membrane. *Genes Cells* **8**: 769–777.
- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36. AAAI Press, Menlo Park, CA.
- Bakker WJ, Blazquez-Domingo M, Kolbus A, Besooyen J, Steinlein P, Beug H, Coffier PJ, Lowenberg B, von Lindern M, van Dijk TB. 2004. FoxO3a regulates erythroid differentiation and induces BTG1, an activator of protein arginine methyl transferase 1. *J Cell Biol* **164**: 175–184.
- Banerjee D, Saha S, Basu S, Chakrabarti A. 2008. Porous red cell ultrastructure and loss of membrane asymmetry in a novel case of hemolytic anemia. *Eur J Haematol* **81**: 399–402.
- Blanc L, Liu J, Vidal M, Chasis JA, An X, Mohandas N. 2009. The water channel aquaporin-1 partitions into exosomes during reticulocyte maturation: Implication for the regulation of cell volume. *Blood* **114**: 3928–3934.
- Blaybel R, Theoleyre O, Douablin A, Baklouti F. 2008. Downregulation of the Spi-1/PU.1 oncogene induces the expression of TRIM10/HERF1, a key factor required for terminal erythroid cell differentiation and survival. *Cell Res* **18**: 834–845.
- Bodine DMIV, Birkenmeier CS, Barker JE. 1984. Spectrin deficient inherited hemolytic anemias in the mouse: Characterization by spectrin synthesis and mRNA activity in reticulocytes. *Cell* **37**: 721–729.
- Bogacheva O, Bogachev O, Menon M, Dev A, Houde E, Valoret EI, Prosser HM, Creasy CL, Pickering SJ, Grau E, et al. 2008. DYRK3 dual-specificity kinase attenuates erythropoiesis during anemia. *J Biol Chem* **283**: 36665–36675.
- Bottardi S, Ross J, Bourgoin V, Fotouhi-Ardakani N, Affar el B, Trudel M, Milot E. 2009. Ikaros and GATA-1 combinatorial effect is required for silencing of human gamma-globin genes. *Mol Cell Biol* **29**: 1526–1537.
- 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.
- Canelles M, Delgado MD, Hyland KM, Lerga A, Richard C, Dang CV, Leon J. 1997. Max and inhibitory c-Myc mutants induce erythroid differentiation and resistance to apoptosis in human myeloid leukemia cells. *Oncogene* **14**: 1315–1327.
- Cantor AB, Orkin SH. 2002. Transcriptional regulation of erythropoiesis: An affair involving multiple partners. *Oncogene* **21**: 3368–3376.
- Capron C, Lecluse Y, Kaushik AL, Foudi A, Lacout C, Sekkai D, Godin I, Albagli O, Poullion I, Svinartchouk F, et al. 2006. The SCL relative LYL-1 is required for fetal and adult hematopoietic stem cell function and B-cell differentiation. *Blood* **107**: 4678–4686.
- Chan WY, Follows GA, Lacaud G, Pimanda JE, Landry JR, Kinston S, Knezevic K, Piltz S, Donaldson IJ, Gambardella L, et al. 2007. The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl^{-/-} phenotype. *Blood* **109**: 1908–1916.
- Chang SH, Low PS. 2001. Regulation of the glycophorin C-protein 4.1 membrane-to-skeleton bridge and evaluation of its contribution to erythrocyte membrane stability. *J Biol Chem* **276**: 22223–22230.
- Charrin S, le Naour F, Silvie O, Milhiet PE, Boucheix C, Rubinstein E. 2009. Lateral organization of membrane proteins: Tetraspanins spin their web. *Biochem J* **420**: 133–154.
- Chen H, Khan AA, Liu F, Gilligan DM, Peters LL, Messick J, Haschek-Hock WM, Li X, Ostafin AE, Chishti AH. 2007. Combined deletion of mouse dematin-headpiece and β -adducin exerts a novel effect on the spectrin-actin junctions leading to erythrocyte fragility and hemolytic anemia. *J Biol Chem* **282**: 4124–4135.
- Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, King DC, Chen KB, Zhang Y, Drautz D, et al. 2009. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19**: 2172–2184.
- Cohen-Barak O, Erickson DT, Badowski MS, Fuchs DA, Klassen CL, Harris DT, Brilliant MH. 2007. Stem cell transplantation demonstrates that Sox6 represses $\epsilon\gamma$ globin expression in definitive erythropoiesis of adult mice. *Exp Hematol* **35**: 358–367.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14**: 1188–1190.
- Cui K, Coutts M, Stahl J, Sytkowski AJ. 2000. Novel interaction between the transcription factor CHOP (GADD153) and the ribosomal protein FTE/S3a modulates erythropoiesis. *J Biol Chem* **275**: 7591–7596.
- Daleke DL. 2008. Regulation of phospholipid asymmetry in the erythrocyte membrane. *Curr Opin Hematol* **15**: 191–195.
- Drissen R, von Lindern M, Kolbus A, Driegen S, Steinlein P, Beug H, Grosveld F, Philipsen J. 2005. The erythroid phenotype of EKLF-null mice: Defects in hemoglobin metabolism and membrane stability. *Mol Biol Cell* **25**: 5205–5214.
- D'Souza SL, Elefanty AG, Keller G. 2005. SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast, but not for its development. *Blood* **105**: 3862–3870.
- Dumitriu B, Patrick MR, Petschek JP, Cherukuri S, Klingmuller U, Fox PL, Lefebvre V. 2006. Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood* **108**: 1198–1207.
- Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci* **93**: 12355–12358.
- Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, Choi K, Farnham PJ, Bresnick EH. 2009. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* **36**: 667–681.
- Gauthier E, Rahuel C, Wautier MP, El Nemer W, Gane P, Wautier JL, Cartron JP, Colin Y, Le Van Kim C. 2005. Protein kinase A-dependent phosphorylation of Lutheran/basal cell adhesion molecule glycoprotein regulates cell adhesion to laminin α 5. *J Biol Chem* **280**: 30055–30062.
- Geiger JN, Knudsen GT, Panek L, Pandit AK, Yoder MD, Lord KA, Creasy CL, Burns BM, Gaines P, Dillon SB, et al. 2001. mDYRK3 kinase is expressed selectively in late erythroid progenitor cells and attenuates colony-forming unit-erythroid development. *Blood* **97**: 901–910.
- Giroux S, Kaushik AL, Capron C, Jalil A, Kelaidi C, Sablitzky F, Dumenil D, Albagli O, Godin I. 2007. *lyl-1* and *tal-1/scl*, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. *Gene Expr Patterns* **7**: 215–226.
- Goardon N, Lambert JA, Rodriguez P, Nissaire P, Herblot S, Thibault P, Dumenil D, Strouboulis J, Romeo P-H, Hoang T. 2006. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J* **25**: 357–366.
- Goossens D, Trinh-Trang-Tan MM, Debbia M, Ripoche P, Vilela-Lamego C, Louache F, Vainchenker W, Colin Y, Cartron JP. 2010. Generation and characterisation of *Rhd* and *Rhag* null mice. *Br J Haematol* **148**: 161–172.

- Grimsey N, Han GS, O'Hara L, Rochford JJ, Carman GM, Siniosoglou S. 2008. Temporal and spatial regulation of the phosphatidate phosphatases lipin 1 and 2. *J Biol Chem* **283**: 29166–29174.
- Hall MA, Curtis DJ, Metcalf D, Elefanty AG, Sourris K, Robb L, Gothert JR, Jane SM, Begley CG. 2003. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci* **100**: 992–997.
- Hall MA, Slater NJ, Begley CG, Salmon JM, Van Stekelenburg LJ, McCormack MP, Jane SM, Curtis DJ. 2005. Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. *Mol Biol Cell* **25**: 6355–6362.
- Hamon Y, Broccardo C, Chambenoit O, Luciani MF, Toti F, Chaslin S, Freyssinet JM, Devaux PF, McNeish J, Marguet D, et al. 2000. ABC1 promotes engulfment of apoptotic cells and transbilayer redistribution of phosphatidylserine. *Nat Cell Biol* **2**: 399–406.
- Han A, Saijo K, Mecklenbrauer I, Tarakhovskiy A, Nussenzweig MC. 2003. Bam32 links the B cell receptor to ERK and JNK and mediates B cell proliferation but not survival. *Immunity* **19**: 621–632.
- Hanspal M, Smockova Y, Uong Q. 1998. Molecular identification and functional characterization of a novel protein that mediates the attachment of erythroblasts to macrophages. *Blood* **92**: 2940–2950.
- Harada H, Harada Y, O'Brien DP, Rice DS, Naevae CW, Downing JR. 1999. HRF1, a novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells. *Mol Cell Biol* **19**: 3808–3815.
- Heikens MJ, Cao TM, Morita C, Dehert SL, Tsai S. 2007. Penumbrin encodes a novel tetraspanin that is highly expressed in erythroid progenitors and promotes effective erythropoiesis. *Blood* **109**: 3244–3252.
- Hodge D, Coghill E, Keys J, Maguire T, Hartmann B, McDowall A, Weiss M, Grimmond S, Perkins A. 2006. A global role for EKLf in definitive and primitive erythropoiesis. *Blood* **107**: 3359–3370.
- Hoffbauer KH, Gess B, Lohaus C, Meyer HE, Katschinski D, Kurtz A. 2003. Oxygen tension regulates the expression of a group of procollagen hydroxylases. *Eur J Biochem* **270**: 4515–4522.
- Hsu H-L, Huang L, Tsan JT, Funk W, Wright WE, Hu J-S, Kingston RE, Baer R. 1994. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol Cell Biol* **14**: 1256–1265.
- Hu X, Li X, Valverde K, Fu X, Noguchi C, Qiu Y, Huang S. 2009. LSD1-mediated epigenetic modification is required for TAL1 function and hematopoiesis. *Proc Natl Acad Sci* **106**: 10141–10146.
- Huang S, Qiu Y, Stein RW, Brandt SJ. 1999. p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. *Oncogene* **18**: 4958–4967.
- Huang S, Qiu Y, Shi Y, Xu Z, Brandt SJ. 2000. P/CAF-mediated acetylation regulates the function of the basic helix-loop-helix transcription factor TAL1/SCL. *EMBO J* **19**: 6792–6803.
- Humbert PO, Rogers C, Ganiatsas S, Landsberg RL, Trimarchi JM, Dandapani S, Brugnara C, Erdman S, Schrenzel M, Bronson RT, et al. 2000. E2F4 is essential for normal erythrocyte maturation and neonatal viability. *Mol Cell* **6**: 281–291.
- Igarashi K, Sun J. 2006. The heme-Bach1 pathway in the regulation of oxidative stress response and erythroid differentiation. *Antioxid Redox Signal* **8**: 107–118.
- Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**: 1293–1300.
- Johnstone RM, Mathew A, Setchenska MS, Grdisa M, White MK. 1998. Loss of glucose transport in developing avian red cells. *Eur J Cell Biol* **75**: 66–77.
- Kamp D, Haest CW. 1998. Evidence for a role of the multidrug resistance protein (MRP) in the outward translocation of NBD-phospholipids in the erythrocyte membrane. *Biochim Biophys Acta* **1372**: 91–101.
- Kassouf MT, Chagraoui H, Vyas P, Porcher C. 2008. Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. *Blood* **112**: 1056–1067.
- Kawada H, Ito T, Pharr PN, Spyropoulos DD, Watson DK, Ogawa M. 2001. Defective megakaryopoiesis and abnormal erythroid development in Fli-1 gene-targeted mice. *Int J Hematol* **73**: 463–468.
- Khanna R, Chang SH, Andrabi S, Azam M, Kim A, Rivera A, Brugnara C, Low PS, Liu SC, Chishti AH. 2002. Headpiece domain of dematin is required for the stability of the erythrocyte membrane. *Proc Natl Acad Sci* **99**: 6637–6642.
- Klinken SP. 2002. Red blood cells. *Int J Biochem Cell Biol* **34**: 1513–1518.
- Koritschoner NP, Bartunek P, Knespel S, Blendinger G, Zenke M. 1999. The fibroblast growth factor receptor FGFR-4 acts as a ligand dependent modulator of erythroid cell proliferation. *Oncogene* **18**: 5904–5914.
- Lahlil R, Lecuyer E, Herblot S, Hoang T. 2004. SCL assembles a multifactorial complex that determines glycoporphin A expression. *Blood* **24**: 1439–1452.
- Lecuyer E, Hoang T. 2004. SCL: From the origin of hematopoiesis to stem cells and leukemia. *Exp Hematol* **32**: 11–24.
- Lecuyer E, Herblot S, Saint-Denis M, Martin R, Begley CG, Porcher C, Orkin SH, Hoang T. 2002. The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **100**: 2430–2440.
- Lee TH, Kim SU, Yu SL, Kim SH, Park DS, Moon HB, Dho SH, Kwon KS, Kwon HJ, Han YH, et al. 2003. Peroxiredoxin II is essential for sustaining life span of erythrocytes in mice. *Blood* **101**: 5033–5038.
- Lee JM, Chan K, Kan YW, Johnson JA. 2004. Targeted disruption of Nrf2 causes regenerative immune-mediated hemolytic anemia. *Proc Natl Acad Sci* **101**: 9751–9756.
- Lee G, Lo A, Short SA, Mankelov TJ, Spring F, Parsons SF, Yazdanbakhsh K, Mohandas N, Anstee DJ, Chasis JA. 2006. Targeted gene deletion demonstrates that the cell adhesion molecule ICAM-4 is critical for erythroblastic island formation. *Blood* **108**: 2064–2071.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li CY, Zhan YQ, Xu CW, Xu WX, Wang SY, Lv J, Zhou Y, Yue PB, Chen B, Yang XM. 2004. EDAG regulates the proliferation and differentiation of hematopoietic cells and resists cell apoptosis through the activation of nuclear factor- κ B. *Cell Death Differ* **11**: 1299–1308.
- Li CY, Zhan YQ, Li W, Xu CW, Xu WX, Yu DH, Peng RY, Cui YF, Yang X, Hou N, et al. 2007. Overexpression of a hematopoietic transcriptional regulator EDAG induces myelopoiesis and suppresses lymphopoiesis in transgenic mice. *Leukemia* **21**: 2277–2286.
- Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, Mackay JP. 2006. Molecular analysis of the interaction between the hematopoietic master transcription factors GATA-1 and PU.1. *J Biol Chem* **281**: 28296–28306.
- Lin C-S, Lim SK, D'Agati V, Costantini F. 1995. Differential effects of an erythropoietin receptor gene disruption on primitive and definitive erythropoiesis. *Genes Dev* **10**: 154–164.
- Liu C, Shi Y, Han Z, Pan Y, Liu N, Han S, Chen Y, Lan M, Qiao T, Fan D. 2003. Suppression of the dual-specificity phosphatase MKP-1 enhances HIF-1 trans-activation and increases expression of EPO. *Biochem Biophys Res Commun* **312**: 780–786.
- Liu C, Shi Y, Du Y, Ning X, Liu N, Huang D, Liang J, Xue Y, Fan D. 2005. Dual-specificity phosphatase DUSP1 protects overactivation of hypoxia-inducible factor 1 through inactivating ERK MAPK. *Exp Cell Res* **309**: 410–418.
- Liu Y, Promeneur D, Rojek A, Kumar N, Frokiaer J, Nielsen S, King LS, Agre P, Carberry JM. 2007. Aquaporin 9 is the major pathway for glycerol uptake by mouse erythrocytes, with implications for malarial virulence. *Proc Natl Acad Sci* **104**: 12560–12564.
- Low FM, Hampton MB, Peskin AV, Winterbourn CC. 2007. Peroxiredoxin 2 functions as a noncatalytic scavenger of low-level hydrogen peroxide in the erythrocyte. *Blood* **109**: 2611–2617.
- Loyd MR, Okamoto Y, Randall MS, Ney PA. 2003. Role of AP1/NFE2 binding sites in endogenous alpha-globin gene transcription. *Blood* **102**: 4223–4228.
- Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**: 958–970.
- Lyouni S, Abitbol M, Andrieu V, Henin D, Robert E, Schmitt C, Gouya L, de Verneuil H, Deybach JC, Montagutelli X, et al. 2007. Increased plasma transferrin, altered body iron distribution, and microcytic hypochromic anemia in ferrochelatase-deficient mice. *Blood* **109**: 811–818.
- Manwani D, Bieker JJ. 2008. The erythroblastic island. *Curr Top Dev Biol* **82**: 23–53.
- Marinkovic D, Zhang X, Yalcin S, Luciano JP, Brugnara C, Huber T, Ghaffari S. 2007. Foxo3 is required for the regulation of oxidative stress in erythropoiesis. *J Clin Invest* **117**: 2133–2144.
- Martin R, Lahlil R, Damert A, Miquerol L, Nagy A, Keller G, Hoang T. 2004. SCL interacts with VEGF to suppress apoptosis at the onset of hematopoiesis. *Development* **131**: 693–702.
- Massari ME, Murre C. 2000. Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**: 429–440.
- McCormack MP, Hall MA, Schoenwaelder SM, Zhao Q, Ellis S, Prentice JA, Clarke AJ, Slater NJ, Salmon JM, Jackson SP, et al. 2006. A critical role of the transcription factor Scl in platelet production during stress thrombopoiesis. *Blood* **108**: 2248–2256.
- Meier N, Krpic S, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosfeld F. 2006. Novel binding partners of Ldb1 are required for haematopoietic development. *Development* **133**: 4913–4923.
- Menon MP, Fang J, Wojchowski DM. 2006. Core erythropoietin receptor signals for late erythroblast development. *Blood* **107**: 2662–2672.
- Metzen E. 2007. Enzyme substrate recognition in oxygen sensing: How the HIF trap snaps. *Biochem J* **408**: 231–240.
- Mikkola HK, Klintman J, Yang H, Hock H, Schlaeger TM, Fujiwara Y, Orkin SH. 2003. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia *SCL/tal-1* gene. *Nature* **421**: 547–551.

- Mohandas N, Gallagher PG. 2008. Red cell membrane: Past, present, and future. *Blood* **112**: 3939–3948.
- Mongiati M, Ligresti G, Marastoni S, Lorenzon E, Doliana R, Colombatti A. 2007. Regulation of the extrinsic apoptotic pathway by the extracellular matrix glycoprotein EMILIN2. *Mol Cell Biol* **27**: 7176–7187.
- Nakajima O, Takahashi S, Harigae H, Furuyama K, Hayashi N, Sassa S, Yamamoto M. 1999. Heme deficiency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. *EMBO J* **18**: 6282–6289.
- Nguyen T, Nioi P, Pickett CB. 2009. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J Biol Chem* **284**: 13291–13295.
- Nishimura N, Sasaki T. 2008. Regulation of epithelial cell adhesion and repulsion: Role of endocytic recycling. *J Med Invest* **55**: 9–16.
- Ogilvy S, Ferreira R, Piltz SG, Bowen JM, Gottgens B, Green AR. 2007. The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. *Mol Cell Biol* **27**: 7206–7219.
- Oldenburg PA, Gresham HD, Chen Y, Izui S, Lindberg FP. 2002. Lethal autoimmune hemolytic anemia in CD47-deficient nonobese diabetic (NOD) mice. *Blood* **99**: 3500–3504.
- Orkin SH. 1992. GATA-binding transcription factors in hematopoietic cells. *Blood* **80**: 575–581.
- Patterson LJ, Gering M, Patient R. 2005. Scl is required for dorsal aorta as well as blood formation in zebrafish embryos. *Blood* **105**: 3502–3511.
- Pavesi G, Merghetti P, Mauri G, Pesole G. 2004. Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res* **32**: W199–W203.
- Paw BH, Davidson AJ, Zhou Y, Li R, Pratt SJ, Lee C, Trede NS, Brownlie A, Donovan A, Liao EC, et al. 2003. Cell-specific mitotic defect and dyserythropoiesis associated with erythroid band 3 deficiency. *Nat Genet* **34**: 59–64.
- Pereira R, Quang CT, Lesault I, Dolznig H, Beug H, Ghysdael J. 1999. FLI-1 inhibits differentiation and induces proliferation of primary erythroblasts. *Oncogene* **18**: 1597–1608.
- Peters LL, Shivdasani RA, Liu SC, Hanspal M, John KM, Gonzalez JM, Brugnara C, Gwynn B, Mohandas N, Alper SL, et al. 1996. Anion exchanger 1 (band 3) is required to prevent erythrocyte membrane surface loss but not to form the membrane skeleton. *Cell* **86**: 917–927.
- Peters LL, Jindel HK, Gwynn B, Korsgren C, John KM, Lux SE, Mohandas N, Cohen CM, Cho MR, Golan DE, et al. 1999. Mild spherocytosis and altered red cell ion transport in protein 4.2-null mice. *J Clin Invest* **103**: 1527–1537.
- Pilon AM, Arcasoy MO, Dressman HK, Vayda SE, Maksimova YD, Sangerman JI, Gallagher PG, Bodine DM. 2008. Failure of terminal erythroid differentiation in EKLf-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol Cell Biol* **28**: 7394–7401.
- Ponka P. 1997. Tissue-specific regulation of iron metabolism and heme synthesis: Distinct control mechanisms in erythroid cells. *Blood* **89**: 1–25.
- Ponka P. 1999. Cell biology of heme. *Am J Med Sci* **318**: 241–256.
- Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. 1996. The T-cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**: 47–57.
- Porcher C, Liao EC, Fujiwara Y, Zon LI, Orkin SH. 1999. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **126**: 4603–4615.
- Qin X, Krumrei N, Grubissich L, Dobarro M, Aktas H, Perez G, Halperin JA. 2003. Deficiency of the mouse complement regulatory protein mCd59b results in spontaneous hemolytic anemia with platelet activation and progressive male infertility. *Immunity* **18**: 217–227.
- Qin X, Hu W, Song W, Grubissich L, Hu X, Wu G, Ferris S, Dobarro M, Halperin JA. 2009. Generation and phenotyping of mCd59a and mCd59b double-knockout mice. *Am J Hematol* **84**: 65–70.
- Rank G, Sutton R, Marshall V, Lundie RJ, Caddy J, Romeo T, Fernandez K, McCormack MP, Cooke BM, Foote SJ, et al. 2009. Novel roles for erythroid Ankyrin-1 revealed through an ENU-induced null mouse mutant. *Blood* **113**: 3352–3362.
- Richards S, Watanabe C, Santos L, Craxton A, Clark EA. 2008. Regulation of B-cell entry into the cell cycle. *Immunol Rev* **224**: 183–200.
- Rimmele P, Kosmider O, Mayeux F, Moreau-Gachelin F, Guillouf C. 2007. Spi-1/PU.1 participates in erythroleukemogenesis by inhibiting apoptosis in cooperation with Epo signaling and by blocking erythroid differentiation. *Blood* **109**: 3007–3014.
- Robledo RF, Ciciotte SL, Gwynn B, Sahr KE, Gilligan DM, Mohandas N, Peters LL. 2008. Targeted deletion of α -adducin results in absent β - and γ -adducin, compensated hemolytic anemia, and lethal hydrocephalus in mice. *Blood* **112**: 4298–4307.
- Roudier N, Ripoche P, Gane P, Le Pennec PY, Daniels G, Cartron JP, Bailly P. 2002. AQP3 deficiency in humans and the molecular basis of a novel blood group system, GIL. *J Biol Chem* **277**: 45854–45859.
- Sahr KE, Lambert AJ, Ciciotte SL, Mohandas N, Peters LL. 2009. Targeted deletion of the γ -adducin gene (*Add3*) in mice reveals differences in α -adducin interactions in erythroid and nonerythroid cells. *Am J Hematol* **84**: 354–361.
- Salomao M, Zhang X, Yang Y, Lee S, Hartwig JH, Chasis JA, Mohandas N, An X. 2008. Protein 4.1R-dependent multiprotein complex: New insights into the structural organization of the red blood cell membrane. *Proc Natl Acad Sci* **105**: 8026–8031.
- Sathyanarayana P, Menon MP, Bogacheva O, Bogachev O, Niss K, Kapelle WS, Houde E, Fang J, Wojchowski DM. 2007. Erythropoietin modulation of podocalyxin and a proposed erythroblast niche. *Blood* **110**: 509–518.
- Sathyanarayana P, Dev A, Fang J, Houde E, Bogacheva O, Bogachev O, Menon M, Browne S, Pradeep A, Emerson C, et al. 2008. EPO receptor circuits for primary erythroblast survival. *Blood* **111**: 5390–5399.
- Sawado T, Igarashi K, Groudine M. 2001. Activation of beta-major globin gene transcription is associated with recruitment of NF-E2 to the beta-globin LCR and gene promoter. *Proc Natl Acad Sci* **98**: 10226–10231.
- Schlaeger TM, Mikkola HK, Gekas C, Helgadottir HB, Orkin SH. 2005. Tie2Cre-mediated gene ablation defines the stem-cell leukemia gene (*SCL/tal1*)-dependent window during hematopoietic stem-cell development. *Blood* **105**: 3871–3874.
- 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.
- Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev* **23**: 804–809.
- Shi ZT, Afzal V, Coller B, Patel D, Chasis JA, Parra M, Lee G, Paszty C, Stevens M, Walensky L, et al. 1999. Protein 4.1R-deficient mice are viable but have erythroid membrane skeleton abnormalities. *J Clin Invest* **103**: 331–340.
- Shirihai OS, Gregory T, Yu C, Orkin SH, Weiss MJ. 2000. ABC-me: A novel mitochondrial transporter induced by GATA-1 during erythroid differentiation. *EMBO J* **19**: 2492–2502.
- Shivadasani RA, Fujiwara Y, McDevitt MA, Orkin SH. 1997. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J* **16**: 3965–3973.
- Shivdasani R, Mayer E, Orkin SH. 1995. Absence of blood formation in mice lacking the T-cell leukemia oncoprotein tal-1/SCL. *Nature* **373**: 432–434.
- Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: doi: 10.2202/1544-6115.1027.
- Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF. 2001. Ineffective erythropoiesis in Stat5a^{-/-}5b^{-/-} mice due to decreased survival of early erythroblasts. *Blood* **98**: 3261–3273.
- Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van Ijcken W, et al. 2010. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24**: 277–289.
- Song SH, Hou C, Dean A. 2007. A positive role for NLI/Ldb1 in long-range β -globin locus control region function. *Mol Cell* **28**: 810–822.
- Souroullas GP, Salmon JM, Sablitzky F, Curtis DJ, Goodell MA. 2009. Adult hematopoietic stem and progenitor cells require either Lyl1 or Scl for survival. *Cell Stem Cell* **4**: 180–186.
- Starck J, Cohet N, Gonnet C, Sarrazin S, Doubeikovskaia Z, Doubeikovski A, Verger A, Duterque-Coquillaud M, Morle F. 2003. Functional cross-antagonism between transcription factors FLI-1 and EKLf. *Mol Cell Biol* **23**: 1390–1402.
- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al. 2002. The Generic Genome Browser: A building block for a model organism system database. *Genome Res* **12**: 1599–1610.
- Steiner LA, Maksimova Y, Schulz V, Wong C, Raha D, Mahajan MC, Weissman SM, Gallagher PG. 2009. Chromatin architecture and transcription factor binding regulate expression of erythrocyte membrane protein genes. *Mol Cell Biol* **29**: 5399–5412.
- Stockmann C, Fandrey J. 2006. Hypoxia-induced erythropoietin production: A paradigm for oxygen-regulated gene expression. *Clin Exp Pharmacol Physiol* **33**: 968–979.
- Stopka T, Amanatullah DE, Papetti M, Skultchi AI. 2005. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *EMBO J* **24**: 3712–3723.
- Swiers G, Patient R, Loose M. 2006. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev Biol* **294**: 525–540.
- Taketani S, Yoshinaga T, Furukawa T, Kohno H, Tokunaga R, Nishimura K, Inokuchi H. 1995. Induction of terminal enzymes for heme biosynthesis

- during differentiation of mouse erythroleukemia cells. *Eur J Biochem* **230**: 760–765.
- Trimborn T, Gribnau J, Grosveld F, Fraser P. 1999. Mechanisms of developmental control of transcription in the murine α - and β -globin loci. *Genes Dev* **13**: 112–124.
- 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.
- Tsang AC, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH. 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**: 109–119.
- Van Kim CL, Colin Y, Cartron JP. 2006. Rh proteins: Key structural and functional components of the red cell membrane. *Blood Rev* **20**: 93–110.
- Ven Murthy MR, Julien P, Singh P, Levy E. 1996. Human lipoprotein lipase deficiency: Does chronic dyslipidemia lead to increased oxidative stress and mitochondrial DNA damage in blood cells? *Acta Biochim Pol* **43**: 227–240.
- Visvader J, Begley CG, Adams JM. 1991. Differential expression of the Lyl, SCL, and E2a helix-loop-helix genes within the hemopoietic system. *Oncogene* **6**: 187–194.
- von Lindern M, Deiner EM, Dolznig H, Parren-Van Amelsvoort M, Hayman MJ, Mullner EW, Beug H. 2001. Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene* **20**: 3651–3664.
- Vyas P, McDevitt MA, Cantor AB, Katz SG, Fujiwara Y, Orkin SH. 1999. Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development* **126**: 2799–2811.
- 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, GATA-1 and Ldb1/NLI proteins. *EMBO J* **16**: 3145–3157.
- Wandersee NJ, Birkenmeier CS, Gifford EJ, Mohandas N, Barker JE. 2000. Murine recessive hereditary spherocytosis, sph/sph, is caused by a mutation in the erythroid alpha-spectrin gene. *Hematol J* **1**: 235–242.
- Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA. 1996a. Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci* **93**: 3444–3449.
- Wang JH, Nichogiannopoulou A, Wu L, Sun L, Sharpe AH, Bigby M, Georgopoulos K. 1996b. Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity* **5**: 537–549.
- Warren AJ, Colledge WH, Carlton MBL, Evans MJ, Smith AJH, Rabbitts TH. 1994. The oncogenic cysteine-rich LIM domain protein Rbta2 is essential for erythroid development. *Cell* **78**: 45–57.
- Wautier MP, El Nemer W, Gane P, Rain JD, Cartron JP, Colin Y, Le Van Kim C, Wautier JL. 2007. Increased adhesion to endothelial cells of erythrocytes from patients with polycythemia vera is mediated by laminin α 5 chain and Lu/BCAM. *Blood* **110**: 894–901.
- Weiss MJ, Orkin SH. 1995. GATA transcription factors: Key regulators of hematopoiesis. *Exp Hematol* **23**: 99–107.
- Wilson NK, Miranda-Saavedra D, Kinston S, Bonadies N, Foster SD, Calero-Nieto F, Dawson MA, Donaldson IJ, 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.
- Wojchowski DM, Menon MP, Sathyanarayana P, Fang J, Karur V, Houde E, Kapelle W, Bogachev O. 2006. Erythropoietin-dependent erythropoiesis: New insights and questions. *Blood Cells Mol Dis* **36**: 232–238.
- Wozniak RJ, Keles S, Lugus JJ, Young KH, Boyer ME, Tran TM, Choi K, Bresnick EH. 2008. Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. *Mol Cell Biol* **28**: 6681–6694.
- Wu H, Liu X, Jaenisch R, Lodish HF. 1995. Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83**: 59–67.
- Xu Z, Huang S, Chang L-H, Agulnick AD, Brandt SJ. 2003. Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. *Mol Cell Biol* **23**: 7585–7599.
- Yang B, Fukuda N, van Hoek A, Matthay MA, Ma T, Verkman AS. 2000. Carbon dioxide permeability of aquaporin-1 measured in erythrocytes and lung of aquaporin-1 null mice and in reconstituted proteoliposomes. *J Biol Chem* **275**: 2686–2692.
- Ye TZ, Gordon CT, Lai YH, Fujiwara Y, Peters LL, Perkins AC, Chui DH. 2000. *Ermap*, a gene coding for a novel erythroid specific adhesion/receptor membrane protein. *Gene* **242**: 337–345.
- Yi Z, Cohen-Barak O, Hagiwara N, Kingsley PD, Fuchs DA, Erickson DT, Epner EM, Palis J, Brilliant MH. 2006. Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet* **2**: doi: 10.1371/journal.pgen.0020014.
- Yokomizo T, Hasegawa K, Ishitobi H, Osato M, Ema M, Ito Y, Yamamoto M, Takahashi S. 2008. Runx1 is involved in primitive erythropoiesis in the mouse. *Blood* **111**: 4075–4080.
- Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y, Yu D, Hardison R, Weiss MJ, Orkin SH, et al. 2009. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* **36**: 682–695.
- Zeuner A, Eramo A, Testa U, Felli N, Pelosi E, Mariani G, Srinivasula SM, Alnemri ES, Condorelli G, Peschle C, et al. 2003. Control of erythroid cell production via caspase-mediated cleavage of transcription factor SCL/Tal-1. *Cell Death Differ* **10**: 905–913.
- Zhang Y, Wu W, Cheng Y, King DC, Harris RS, Taylor J, Chiaromonte F, Hardison RC. 2009. Primary sequence and epigenetic determinants of in vivo occupancy of genomic DNA by GATA1. *Nucleic Acids Res* **37**: 7024–7038.
- Zhu Y, Paszty C, Turetsky T, Tsai S, Kuypers FA, Lee G, Cooper P, Gallagher PG, Stevens ME, Rubin E, et al. 1999. Stomatocytosis is absent in “stomatin”-deficient murine red blood cells. *Blood* **93**: 2404–2410.

Received January 6, 2010; accepted in revised form May 19, 2010.