The EMBO Journal (2010), 1–16 | © 2010 European Molecular Biology Organization | Some Rights Reserved 0261-4189/10
www.embojournal.org

THE EMBO JOURNAL

EMBO open

Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages

Carmen G Palii[1,4], Carolina Perez-Iratxeta[1,4], Zizhen Yao[2], Yi Cao[2], Fengtao Dai[1,3], Jerry Davidson[2], Harold Atkins[1], David Allan[1], F Jeffrey Dilworth[1,3], Robert Gentleman[2], Stephen J Tapscott[2] and Marjorie Brand[1,3,*]

[1]The Sprott Center for Stem Cell Research, Department of Regenerative Medicine, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, [2]Department of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA and [3]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada

TAL1/SCL is a master regulator of haematopoiesis whose expression promotes opposite outcomes depending on the cell type: differentiation in the erythroid lineage or oncogenesis in the T-cell lineage. Here, we used a combination of ChIP sequencing and gene expression profiling to compare the function of TAL1 in normal erythroid and leukaemic T cells. Analysis of the genome-wide binding properties of TAL1 in these two haematopoietic lineages revealed new insight into the mechanism by which transcription factors select their binding sites in alternate lineages. Our study shows limited overlap in the TAL1-binding profile between the two cell types with an unexpected preference for ETS and RUNX motifs adjacent to E-boxes in the T-cell lineage. Furthermore, we show that TAL1 interacts with RUNX1 and ETS1, and that these transcription factors are critically required for TAL1 binding to genes that modulate T-cell differentiation. Thus, our findings highlight a critical role of the cellular environment in modulating transcription factor binding, and provide insight into the mechanism by which TAL1 inhibits differentiation leading to oncogenesis in the T-cell lineage.

The EMBO Journal advance online publication, 21 December 2010; doi:10.1038/emboj.2010.342

Subject Categories: chromatin & transcription; molecular biology of disease

Keywords: erythroid cells; ETS1; RUNX1; SCL/TAL1; T-cell acute lymphoblastic leukaemia (T-ALL)

*Corresponding author. The Sprott Center for Stem Cell Research, Department of Regenerative Medicine, Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, Ontario, Canada K1H 8L6. Tel.: +1 613 737 7700/Ext 70336; Fax: +1 613 739 6294; E-mail: mbrand@ohri.ca

[4]These authors contributed equally to this work

Received: 19 August 2010; accepted: 30 November 2010

Introduction

Cell differentiation is regulated by finely tuned mechanisms directed by cell-specific and ubiquitous transcription factors. Mutations (e.g. deletions, fusions) that affect the integrity of transcription factors by altering their DNA-binding specificity and/or capacity to interact with cofactors can transform these proteins into potent oncogenes. At the same time, wild-type (WT) (non-mutated) transcription factors can also become oncogenic when aberrantly expressed in an inappropriate cell type (Tenen, 2003; O'Neil and Look, 2007). While this argues for an important role of the cellular context in modifying transcription factors' ability to control cell fate, the extent to which the cellular environment affects the function of transcription factors is unclear (Pan *et al*, 2009).

The basic helix-loop-helix (bHLH) protein TAL1 (also called SCL) displays distinct, sometimes opposite, functions in different cell types (Begley and Green, 1999; Lecuyer and Hoang, 2004). Indeed, TAL1 expression is necessary for the specification, survival and competence of haematopoietic stem cells and for the differentiation of megakaryocytes and erythrocytes (Lecuyer and Hoang, 2004; Reynaud *et al*, 2005; Brunet de la Grange *et al*, 2006; Souroullas *et al*, 2009; Lacombe *et al*, 2010). Yet TAL1, which is normally turned off early in the lymphoid lineage, exhibits oncogenic properties when aberrantly expressed in lymphoid tissue (Condorelli *et al*, 1996; Kelliher *et al*, 1996). Importantly, wild-type TAL1 is aberrantly expressed in over 60% of T-cell acute lymphoblastic leukaemia (T-ALL) patients and is considered a major factor in initiating leukaemic transformation via perturbation of the transcriptional regulatory network (Aifantis *et al*, 2008). TAL1-mediated leukaemogenesis has been linked to both an early arrest in the T-cell differentiation program and elevated levels of anti-apoptotic genes (Ferrando *et al*, 2002). While the mechanism of TAL1-mediated leukaemogenesis is unclear, it has been proposed that TAL1 interferes with the function of bHLH E-proteins (i.e. E2A, HEB or E2-2), which are important regulators of T-cell differentiation and whose inactivation leads to T-cell tumours in mice (Quong *et al*, 2002). Indeed, TAL1 binding to E-box DNA motifs (CANNTG) requires heterodimerization with an E-protein and *in vitro* binding selection experiments have identified a TAL1/E-protein heterodimer's preferred E-box (CAGATG), which differs from the E-protein homodimers' preferred E-box (CAGGTG) (Hsu *et al*, 1994). Interestingly, E-box recognition is not always an important determinant of TAL1 binding as it has been proposed to be tethered to genes via other DNA-binding transcription factors, including GATA3 in leukaemic T cells (Ono *et al*, 1998), and SP1 (Lecuyer *et al*, 2002) or GATA1 (Wadman *et al*, 1997) in erythroid cells. Recent ChIP-seq experiments in erythroid

cells have revealed a strong correlation between GATA and TAL1 recognition motifs, with genomic sites bound by TAL1 being frequently associated to GATA motifs while GATA1-bound sites are enriched in E-boxes (Cheng *et al*, 2009; Fujiwara *et al*, 2009; Kassouf *et al*, 2010; Soler *et al*, 2010). In addition, GATA1 and TAL1 cooccupancy appears to correlate with active genes in erythroid cells, although these two transcription factors can be cobound to genes that are repressed (Cheng *et al*, 2009; Tripic *et al*, 2009; Soler *et al*, 2010). Interestingly, degenerate selection experiments for TAL1 binding *in vitro* have identified a composite E-box/Gata motif where the two DNA-binding sites are separated by 8–10 bp (Wadman *et al*, 1997). This particular distance is thought to be important for binding of a pentameric protein complex in which a TAL1/E2A heterodimer and a GATA factor are bridged by LMO2 and LDB1 proteins (Wadman *et al*, 1997). While this composite E-box/Gata motif was recently shown to be enriched under TAL1 peaks identified in erythroid cells (Kassouf *et al*, 2010; Soler *et al*, 2010), it has not been identified in ChIP-microarray studies performed in T-ALL cells (Palomero *et al*, 2006). As such, our lack of knowledge regarding the mechanism of how TAL1 recognizes binding sites *in vivo* represents one of the major limitations to our understanding of the role of this bHLH protein in promoting different cell fates depending on the lineage.

Results

TAL1 promotes erythroid differentiation while it blocks T-cell differentiation

To identify features that distinguish the role of TAL1 in different cell types, we employed a comparative strategy whereby the transcriptional network of TAL1 is contrasted between an erythroid environment in which TAL1 promotes cellular differentiation and a T-cell context in which TAL1 promotes oncogenic transformation. Our strategy combines phenotypic analysis and gene expression profiling after TAL1 knockdown (KD) with chromatin immunoprecipitation and deep sequencing (ChIP-seq).

To study TAL1 in the erythroid lineage we used primary erythroid cells differentiated *ex vivo* from human haematopoietic multipotential progenitors, a system that mimics the differentiation of erythroid cells *in vivo* (Giarratana *et al*, 2005) (Supplementary Figure S1 and data not shown). TAL1 KD was induced in pro-erythroblasts using lentivirus-delivered shRNA (Figure 1A). Following TAL1 KD (Figure 1B and C), we observed a strong diminution in cell growth (Figure 1D), which is due to both a decrease in cell proliferation (Figure 1E), and an increase in apoptosis (Figure 1F). Cell cycle analysis demonstrates accumulation of cells in the G0/G1 phases, suggesting a block at the G1/S transition (Figure 1G). To determine whether TAL1 KD also affects erythroid differentiation, we analysed accumulation of haemoglobin (Figure 1H; Supplementary Figure S2B), CD36, CD71 and GPA cell surface markers (Supplementary Figure S2C) as well as *Gpa* (Figure 1I) and β-*globin* (Figure 4C) transcripts. We found that these erythroid markers are all decreased in TAL1 KD cells confirming the importance of TAL1 for terminal erythroid differentiation.

To study TAL1 in a T-cell environment, we first used the TAL1-expressing Jurkat cell line, which was originally derived from a T-ALL patient and represents a prototypical immature transformed T cell (Schneider *et al*, 1977). To KD TAL1 in Jurkat cells, clonal lines expressing a Dox-inducible shRNA against *Tal1* were generated (Figure 2A and B). Similarly to erythroid cells, we observed a dramatic decrease in the growth of Jurkat cells upon TAL1 KD (Figure 2C). This is mostly due to apoptosis as shown by a 10-fold increase in Annexin V positive cells (Figure 2E), as well as a decrease in mitochondrial transmembrane potential and an increase in caspases 3 and 8 activities (data not shown). While TAL1 KD also led to a limited decrease in BrdU incorporation (Figure 2D), progression through the cell cycle is not affected (Figure 2F). Comparable phenotypic effects were observed in a second TAL1 KD clone stably expressing a distinct anti-Tal1 shRNA sequence (data not shown).

Gene expression analysis upon TAL1 KD

To gain further insight into the genes affected by TAL1 KD in erythroid and Jurkat cells, expression profiling by microarray was performed on WT and KD cells (Supplementary Figure S3). These experiments used Jurkat cells (WT) and their counterparts treated with Dox for 72 h (KD), as well as pro-erythroblasts at day 12 of differentiation that were either infected with lentiviruses expressing scrambled shRNA (WT) or infected with lentiviruses expressing anti-Tal1 shRNA (KD) as indicated on Figure 1A. We found that in erythroid cells, the majority of differentially expressed transcripts are downregulated upon TAL1 KD—442 transcripts downregulated versus 148 transcripts upregulated (Supplementary Figure S3). In contrast, the majority of differentially expressed transcripts in Jurkat cells are upregulated upon TAL1 KD—370 transcripts upregulated versus 249 transcripts downregulated. Microarray results were confirmed by RT–qPCR for all five tested TAL1-dependent genes in erythroid cells (Supplementary Figure S4A) and for 44 of the 45 genes tested in Jurkat cells (Supplementary Figure S4B and data not shown). In agreement with the observed phenotypic effects, Gene Ontology (GO) analysis of genes that are downregulated upon TAL1 KD in erythroid cells identified biological process categories related to cell cycle control, DNA replication and erythroid-related functions (Supplementary Figure S3C; Supplementary Table I). The same analysis of genes that are upregulated upon TAL1 KD in Jurkat cells identified categories related to apoptosis, negative regulation of growth and T-cell differentiation (Supplementary Figure S3B; Supplementary Table II). This last GO category suggested to us that upon TAL1 KD, Jurkat cells might have partially re-entered the T-cell differentiation transcriptional program. For example, four transcription factors that act as master regulators of T-cell differentiation (i.e. *Gata3* (Ting *et al*, 1996), *Sox4* (Schilham *et al*, 1997), *Ikzf3* (coding for the Ikaros homolog Aiolos (Morgan *et al*, 1997)) and the thymocyte selection-associated gene *Tox* (Aliahmad and Kaye, 2008)), are among the genes identified by microarray (and confirmed by RT–qPCR (Supplementary Figure S4)) as being upregulated upon TAL1 KD. Furthermore, the increased expression of GATA3 and Aiolos upon TAL1 KD was confirmed at the protein level (Figure 3D).

Genes that are differentially expressed at particular stages during human CD4+ T-cell differentiation were previously identified from purified human cells, and sorted into signature sets (Lee *et al*, 2004). To further examine the possibility that a decrease of TAL1 in CD4+ Jurkat cells leads to partial

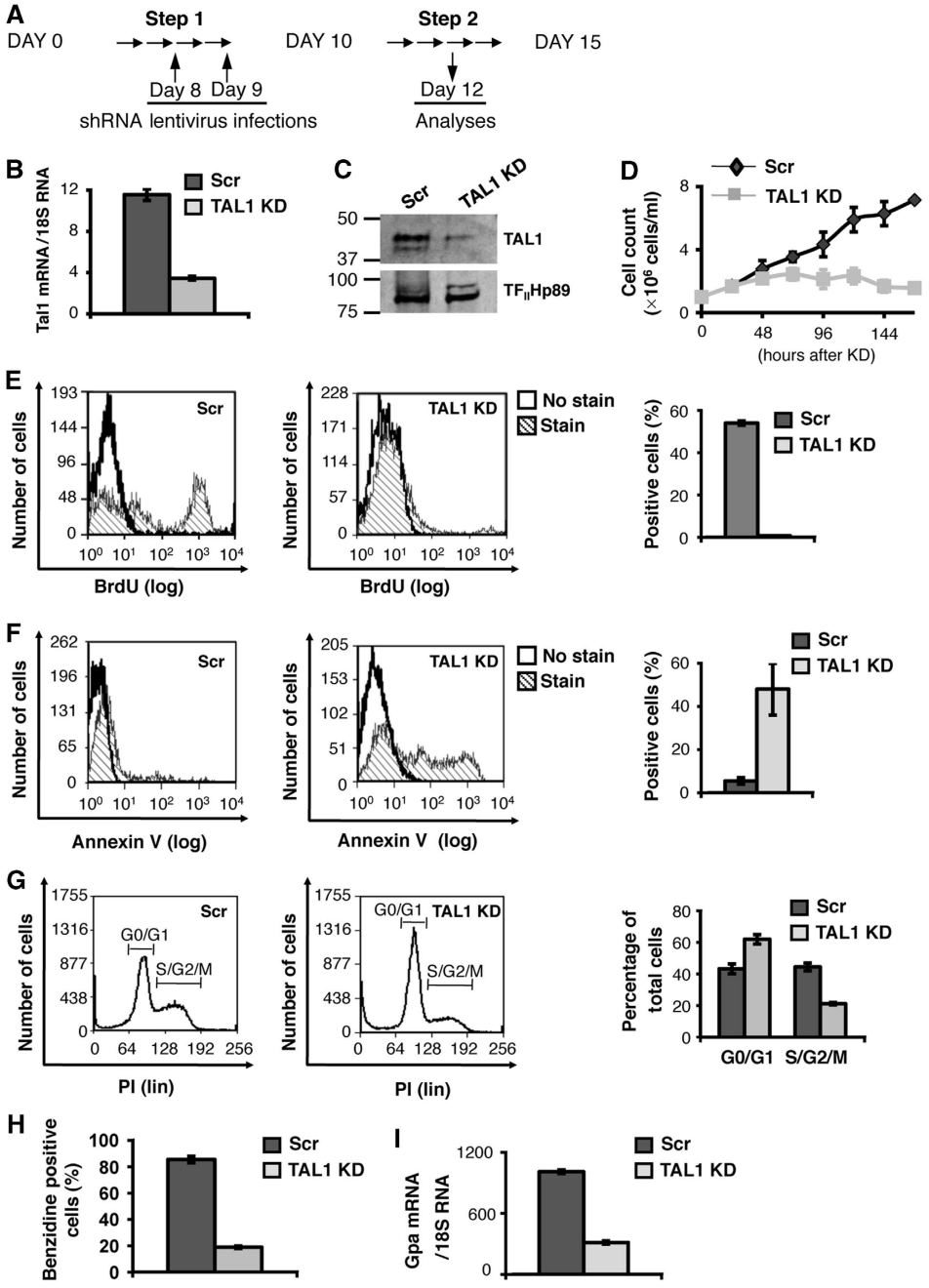


Figure 1 TAL1 knockdown in pro-erythroblasts leads to decreased proliferation and differentiation. (**A**) Timeline for induction of TAL1 KD in primary pro-erythroblasts using a lentiviral shRNA delivery system. (**B**, **C**) TAL1 level is decreased after infection with lentivirus expressing anti-Tal1 (TAL1 KD) but not scramble (Scr) shRNA, as verified by RT–qPCR (**B**) and western blot (**C**). Molecular weight markers (in kDa) are indicated on the left. (**D**) TAL1 KD leads to a profound decrease in cell growth. (**E**–**G**) Characterization of TAL1 KD. (Left panels) FACS histograms of representative experiments. (Right panels) Mean percentage of positive cells in three independent experiments. (**E**) TAL1 KD leads to a decrease in cell proliferation, as measured by BrdU incorporation. (**F**) TAL1 KD leads to an increase in apoptosis, as measured by Annexin V staining. Necrotic cells that are positive for 7AAD are excluded. (**G**) TAL1 KD leads to cell cycle arrest at the G1/S transition. DNA content is measured by propidium iodide (PI) staining. (**H**) TAL1 KD leads to a decrease in haemoglobin synthesis, as measured by benzidine staining. (**I**) TAL1 KD leads to a decrease in Glycophorin A (*Gpa*) transcription as measured by RT–qPCR. Error bars represent standard deviations of biological triplicates. (**B**, **C**, **E**–**I**) correspond to analyses performed on day 12 of (**A**).

re-entry into the T-cell transcriptional program, we used gene set enrichment analysis (GSEA) to compare genes that are differentially expressed upon TAL1 KD in Jurkat cells to these signature sets (Figure 3). We found that gene sets corresponding to early CD4+ intrathymic T-cell progenitors (sets 1 and 2) are enriched in Jurkat cells, while gene sets upregulated in further differentiated T-cell states are enriched in the TAL1 KD condition (sets 3–6) (Figure 3). Notably, an increase in *Tox*, *Sox4* and *Gata3* transcript level upon TAL1 KD was again identified in this GSEA analysis (Figure 3A). Combined results from all six gene sets enrichments indicate that Jurkat cells resemble T cells blocked at the CD4+ immature

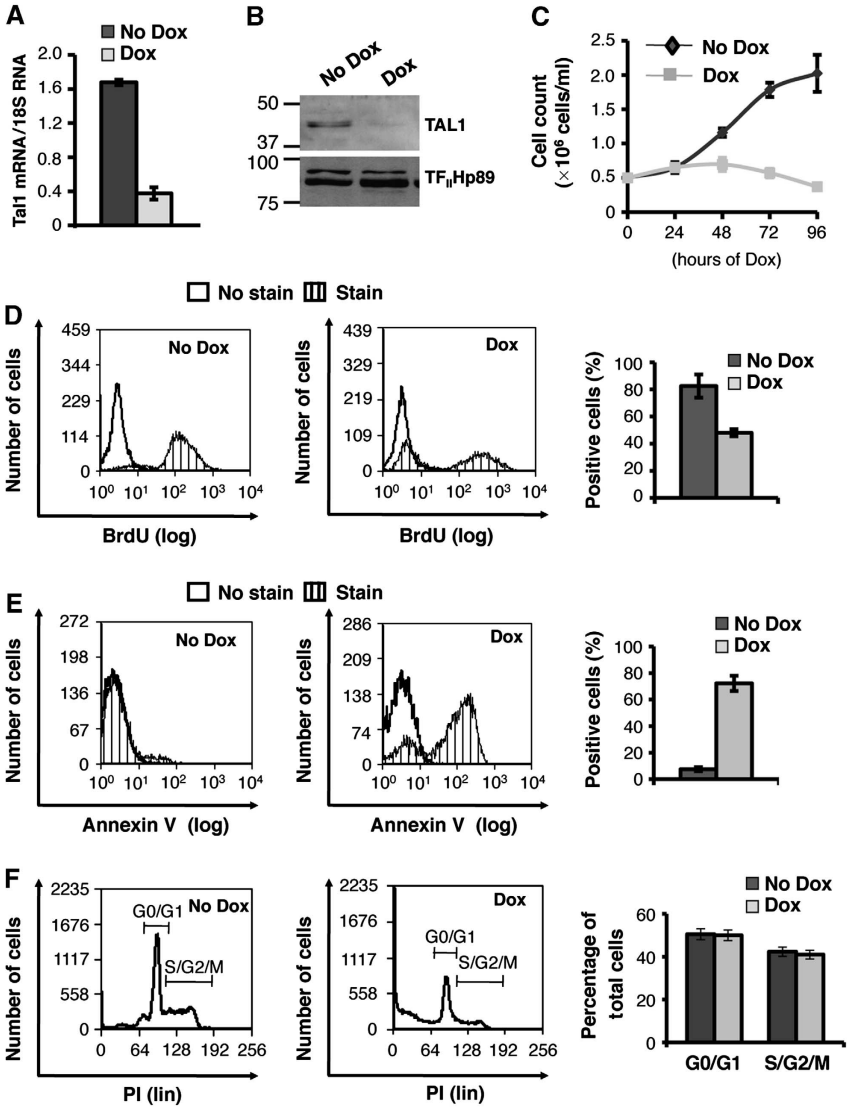


Figure 2 TAL1 knockdown in Jurkat cells leads to apoptosis. (**A**, **B**) TAL1 level is decreased upon 72 h Dox treatment in a Jurkat clone stably expressing a Dox-dependent TAL1 shRNA, as verified by RT–qPCR (**A**) and western blot (**B**). Molecular weight markers (in kDa) are indicated on the left. (**C**) TAL1 KD leads to a profound decrease in cell growth. (**D**–**F**) Characterization of TAL1 KD after 72 h of Dox. (Left panels) FACS histograms of representative experiments. (Right panels) Mean percentage of positive cells in three independent experiments. (**D**) TAL1 KD leads to a decrease in cell proliferation, as measured by BrdU incorporation. (**E**) TAL1 KD leads to an increase in apoptosis, as measured by Annexin V staining. Necrotic cells that are positive for 7AAD are excluded. (**F**) TAL1 KD has no effect on cell cycle progression. DNA content is measured by PI staining. Error bars represent standard deviations of biological triplicates.

single-positive stage (CD4ISP), and that upon TAL1 KD these cells partially resume the transcriptional program of T-cell differentiation before dying by apoptosis.

Identification of TAL1 genomic binding sites by ChIP sequencing

To understand how TAL1 might regulate the expression of genes identified by microarray, we performed genome-wide binding analysis. For each cell type (i.e. Jurkat cells and pro-erythroblasts at day 12 of differentiation, see Figure 1A), two biologically independent TAL1 ChIPs were performed, and the resulting DNA was amplified, subjected to high-throughput sequencing and aligned to the human reference genome (Supplementary Figure S5A). Examination of chromosome 2 indicates that TAL1 peaks are more abundant in erythroid compared with Jurkat cells (Supplementary Figure S5C). Genome wide, we counted 6315 TAL1 peaks in erythroid cells while in Jurkat cells the number of peaks decreases down to 2547 (Figure 4A). Considering the high frequency of E-boxes in the human genome, these numbers of TAL1-binding sites reflect a restricted genomic binding. An analysis based on the rate at which background signal is converted to foreground signal shows that we have attained sufficient sequencing depth for genome coverage (data not shown). Confirming the quality of our data set, ChIP-seq analysis identified a number of previously characterized TAL1 targets, including *Epb42* (Xu *et al*, 2003), *Gypa* (Lahlil *et al*, 2004), DNase I hypersensitive site (HS) 2 of the β-globin locus (Song *et al*, 2007), *c-kit* (Lecuyer *et al*, 2002), *Runx1* (Wilson *et al*, 2009) and *Lmo2* (Landry *et al*, 2009) in erythroid cells, and *Aldh1a2* (Ono *et al*, 1998), *Tcra* (Bernard *et al*, 1998), *Chrna5* (Palomero *et al*, 2006) and *Nfkb1* (Chang *et al*, 2006) in Jurkat cells (Figure 4; Supplementary Figures S6 and S7 and data not shown). In addition, we identified TAL1 peaks in

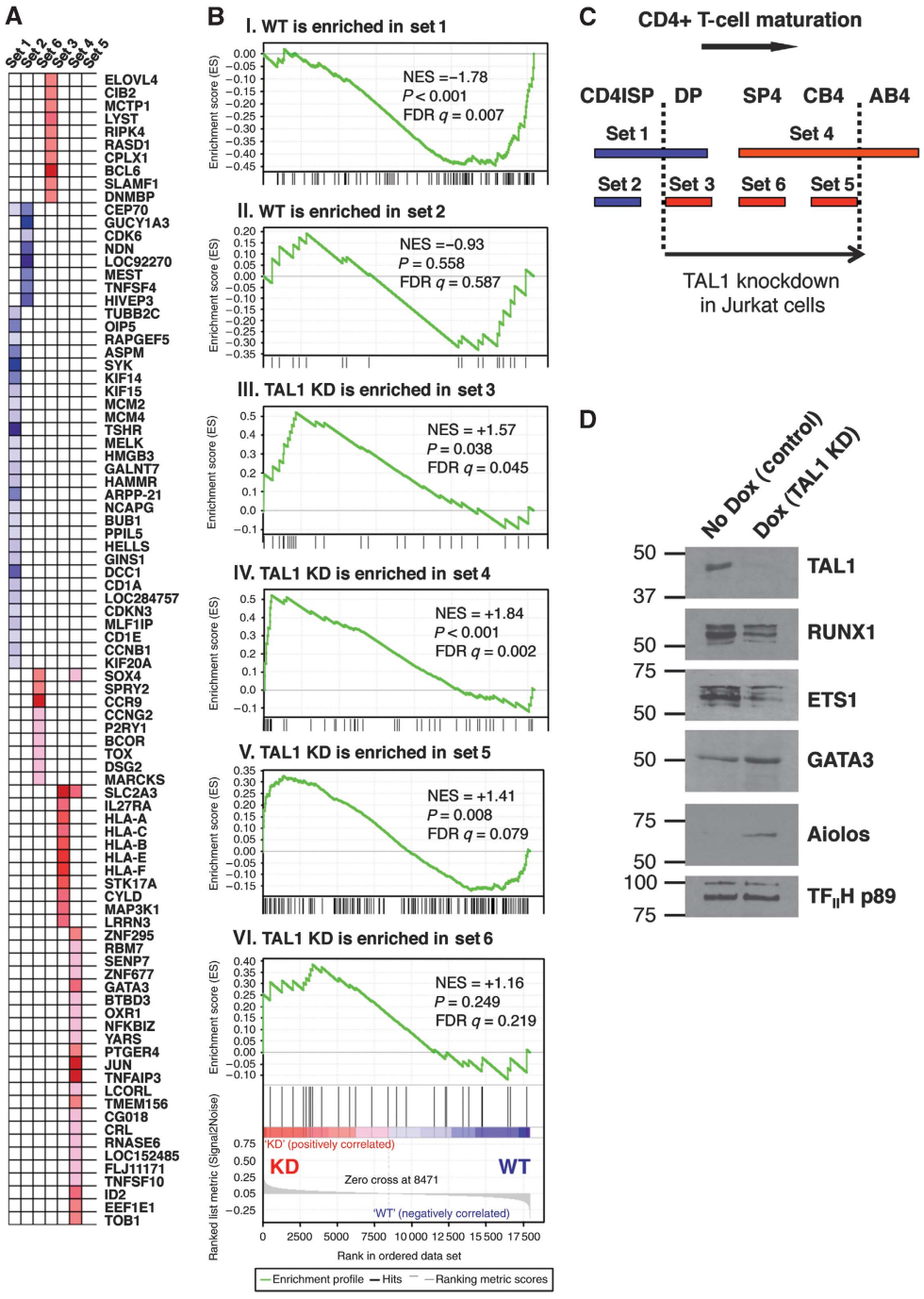


Figure 3 Jurkat cells are blocked at an immature stage resembling CD4+ intrathymic progenitors but partially re-enter the transcriptional program of T-cell differentiation upon TAL1 knockdown. (**A–C**) Gene set enrichment analysis (GSEA) was performed comparing microarray expression profiling of Jurkat cells upon TAL1 KD with signature gene sets established from expression profiling of purified normal primary human CD4+ T cells at different stages of differentiation (Lee *et al*, 2004). This analysis indicates that signature sets corresponding to early CD4+ intrathymic T-cell progenitors (sets 1 and 2) are enriched in the WT phenotype (in blue) while sets corresponding to more differentiated states (sets 3–6) are enriched in TAL1 KD phenotype (in red). For a detailed description of gene sets, see Supplementary data. (**A**) GSEA leading edge analysis heat map and list of enriched genes. (**B**) Enrichment plots for the six T-cell signature sets. NES, normalized enrichment score; FDR, false discovery rate. (**C**) Diagram summarizing the correlation between changes in gene expression profiles upon TAL1 KD in Jurkat cells and during CD4+ T-cell maturation. CD4ISP, CD4 immature single positive; DP, CD4 CD8 double positive; SP4, CD4 single positive; CB4, CD4 single positive from cord blood; AB4, CD4 single positive from adult blood. (**D**) TAL1 KD leads to profound changes in the protein levels of T-cell transcription factors. TAL1 KD was induced in Jurkat cells by the addition of Dox for 72 h. Nuclear extracts were prepared and analysed by western blot using antibodies indicated on the right. Molecular weight markers (in kDa) are indicated on the left.

pericentromeric regions (Supplementary Figure S5C and data not shown), which is consistent with reported TAL1 binding to satellite DNA (Wen *et al*, 2005). Validating the quality of our ChIP-seq results, all 23 known and novel TAL1 genomic targets that we tested (with peak heights ranging from 8 to 178 reads) were confirmed by independent ChIP-qPCR experiments (Figure 4C; Supplementary Figures S6 and S7 and data not shown).

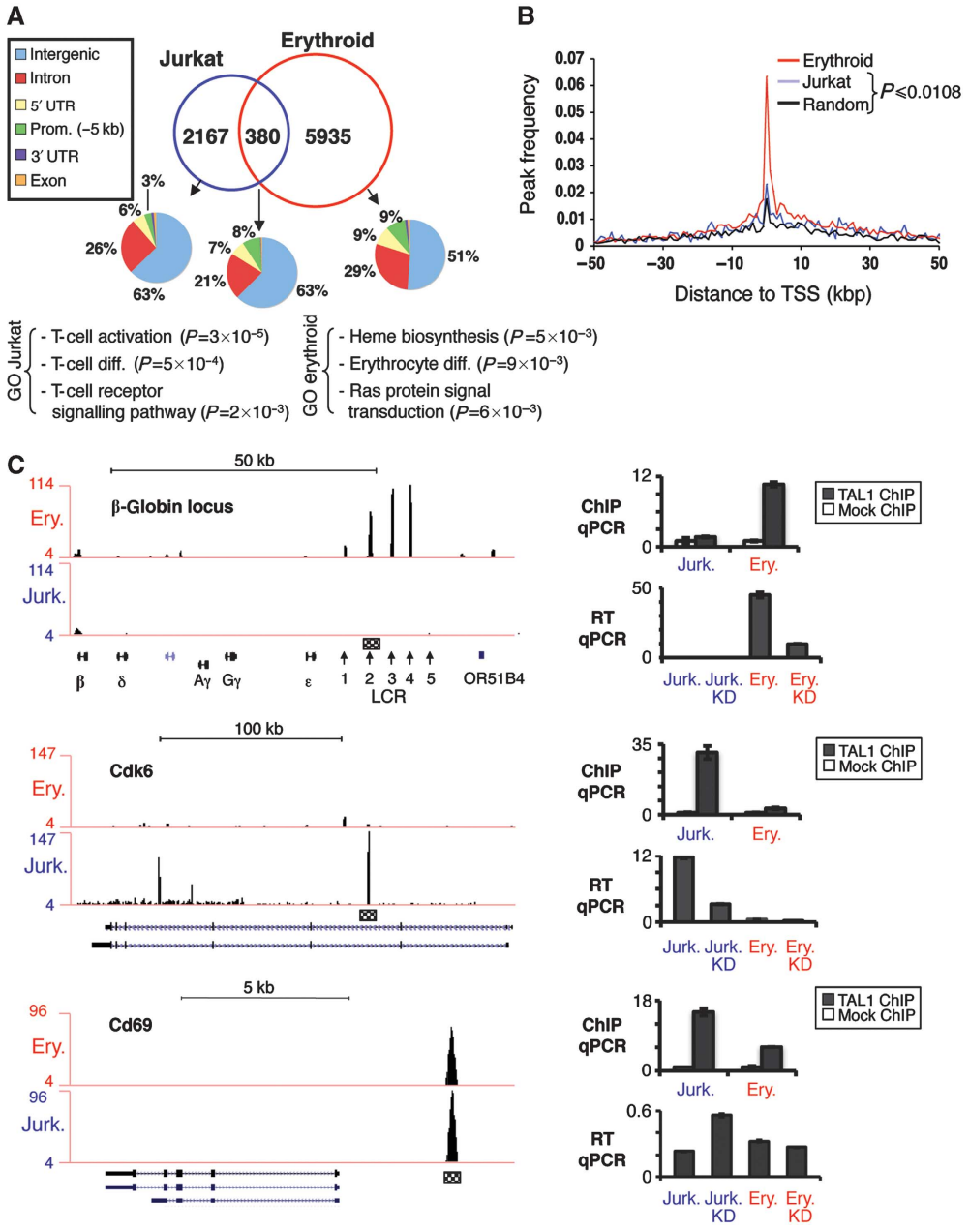


Figure 4 ChIP-seq analysis of TAL1 genomic binding in erythroid versus Jurkat cells. (**A**) Less than 10% of TAL1 peaks are located within promoters (defined as the region covering 5 kb upstream of TSS). GO analysis of genes closest to TAL1 peaks is represented below the Venn diagrams. *P*-values (*P*) are indicated. For a full list of GO terms, see Supplementary Tables III and IV. (**B**) Frequency of TAL1 peaks location relative to the TSS of their closest gene shows a bias around TSS. The *P*-value for Jurkat cells was calculated using Wilcoxon matched-pairs signed-ranks test. (**C**) Representative examples of ChIP-seq results showing TAL1 binding to the β-*globin* locus (LCR, locus control region), to the *cdk6* second intron and to the *cd69* promoter. Binding of TAL1 to locations indicated by a hatched square was validated by independent ChIP–qPCR. ChIP–qPCR values are expressed as a fraction of the input with error bars corresponding to standard deviations. Transcription was analysed by RT–qPCR with or without KD of TAL1. RT–qPCR values are normalized using 18S RNA with error bars corresponding to standard deviations.

In agreement with the different functions of TAL1 in erythroid versus T-ALL cells, the proportion of overlapping TAL1 peaks between the two cell types is relatively small representing 6% of total peaks in erythroid cells and 15% of total peaks in Jurkat cells (Figure 4A). Functional annotation analyses of genes associated to the nearest TAL1 peak in erythroid and Jurkat cells identified overrepresented GO terms related to erythroid and T-cell differentiation, respectively (Figure 4A; Supplementary Tables III and IV). We also observed that while there is a higher local density of TAL1 peaks near TSSs in both erythroid and Jurkat cells (Figure 4B), the majority of TAL1 peaks are located away from promoter regions of known genes, mostly within introns and intergenic regions (Figure 4A). Binding at intergenic regions could represent in some cases binding to distal regulatory elements such as enhancers. For example, in erythroid cells TAL1 is bound to the four erythroid-specific DNase I HSs that comprise the distal β-globin locus control region LCR (Figure 4C). TAL1 is also frequently bound to introns as shown for the *Cdk6* regulator of T-cell differentiation (Grossel and Hinds, 2006), whose expression is upregulated by TAL1 in T-ALL cells (Figure 4C). An example of TAL1

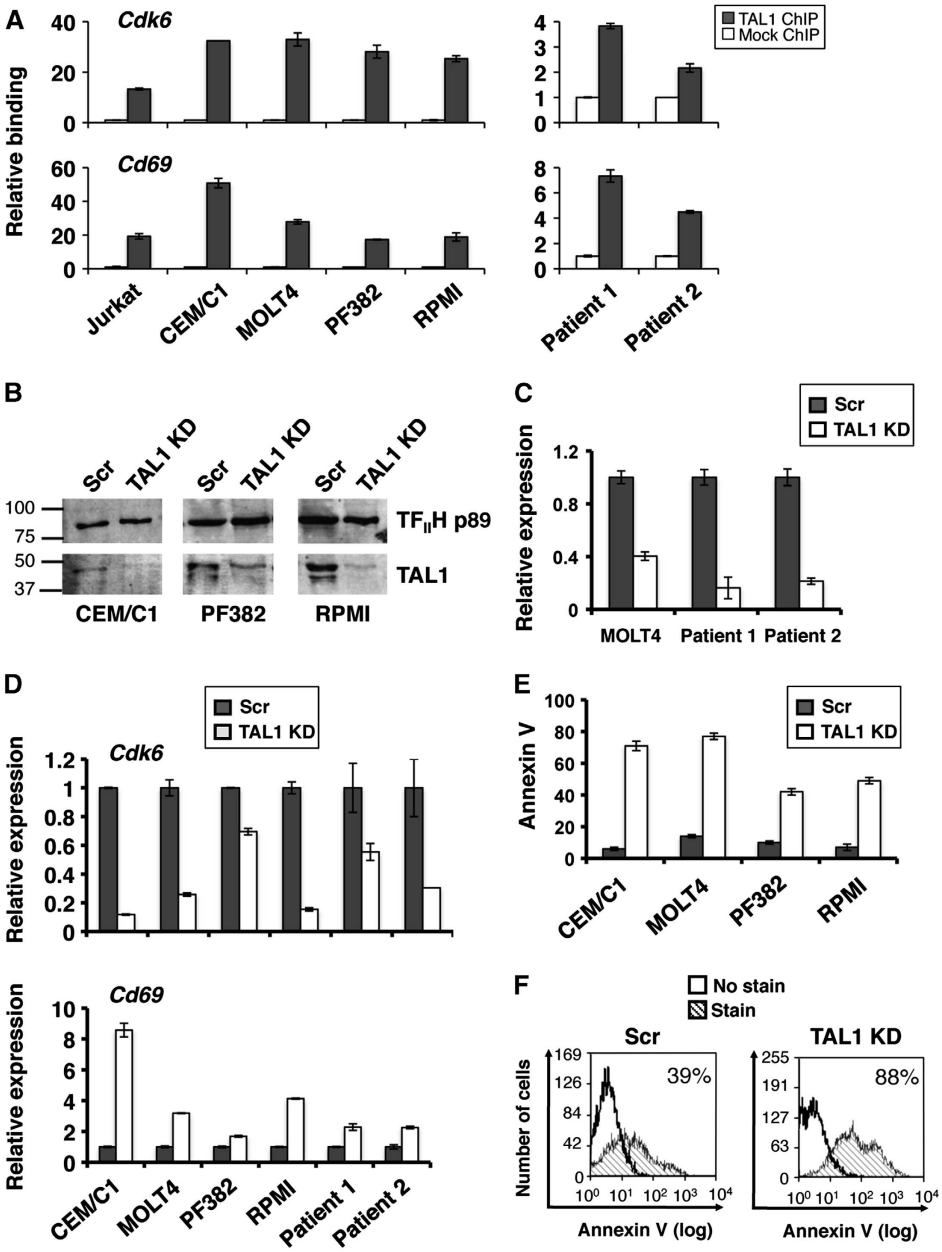


Figure 5 Validation of TAL1 properties in TAL1-expressing T-ALL cell lines and primary blasts from T-ALL patients. (**A**) TAL1 binding to representative sites on *cdk6* and *cd69* genes (Figure 4C) in Jurkat cells is validated by ChIP-qPCR in four additional TAL1-expressing cell lines (i.e. CEM/C1, MOLT4, PF382 and RPMI) and in TAL1-expressing blasts from two T-ALL patients. TAL1 binding is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG. See Supplementary Figure S7 for validation of TAL1 binding at additional sites. (**B**, **C**) TAL1 level is decreased after infection with lentivirus expressing anti-Tal1 (TAL1 KD) but not scramble (Scr) shRNA, as verified by western blot (**B**) or RT–qPCR (**C**). RT–qPCR values are normalized using 18S RNA and are expressed relative to the Scr control. Molecular weight markers (in kDa) are indicated on the left of the western blot. (**D**) TAL1 KD leads to a decrease in *cdk6* and an increase in *cd69* transcripts, as shown by RT–qPCR. RT–qPCR values are normalized using 18S RNA and are expressed relative to the Scr control. (**E**) TAL1 KD leads to an increase in apoptosis in T-ALL cell lines, as measured by Annexin V staining. Necrotic cells that are positive for 7AAD are excluded. Numbers represent the percentage of positive cells. (**F**) TAL1 KD leads to an increase in apoptosis in primary T-ALL blasts. Patient 2 is shown as a representative example. Apoptosis is measured by Annexin V staining. Necrotic cells that are positive for 7AAD are excluded.

binding to a promoter region is shown on the *Cd69* gene (Figure 4C), which is expressed transiently in immature thymocytes undergoing positive selection (Bendelac *et al*, 1992) and is also one of the earliest inducible cell surface glycoprotein acquired during lymphoid activation (Sancho *et al*, 2005). In contrast to *Cdk6*, the *Cd69* gene is down-regulated by TAL1 in Jurkat cells. Interestingly, the KD of TAL1 in an erythroid environment does not affect *Cd69* expression despite TAL1 binding to this gene's promoter. Together, these ChIP-seq data provide us with a number of novel TAL1 genomic targets in both erythroid and T-ALL cellular environments.

Identification of TAL1-target genes functionally regulated by TAL1

A major question arising in genome-wide studies of transcription factor binding is that of the association between binding events and regulated genes. As a first approximation,

we associated TAL1 peaks to their closest genes. Functional annotation of these genes led to the identification of biological categories that are consistent with TAL1 function (Figure 4A), providing confidence that TAL1 indeed regulates some genes by binding to their promoter (e.g. *Cd69*; Figure 4C). However, in many cases, TAL1 is bound away from promoters. Therefore, a simple association of TAL1 peaks to their closest genes may miss target genes regulated by TAL1 via binding to a distal regulatory element.

To identify TAL1 targets that are also functionally regulated by this factor, we took advantage of our identification of TAL1-dependent genes by microarray and restricted our analysis to genes that are differentially expressed upon TAL1 KD. To remain permissive to distal regulatory elements, differentially expressed genes were associated to TAL1 peaks located within 50 kb upstream or downstream of their TSS. Using this approach, 289 genes were identified in erythroid cells, including 246 that are downregulated upon TAL1 KD (thereafter called 'TAL1-activated genes') and 43 that are upregulated (thereafter called 'TAL1-repressed genes') (Supplementary Table V). This result is consistent with previous studies, which have predominantly associated TAL1 to active genes in erythroid cells (Cheng *et al*, 2009; Tripic *et al*, 2009; Soler *et al*, 2010). In agreement with our phenotypic analyses (Figure 1), many TAL1-activated genes in erythroid cells are involved in the regulation of DNA replication, cell cycle and erythroid differentiation. In Jurkat cells, we identified 73 TAL1-repressed genes and 44 TAL1-activated genes (Supplementary Table VI). Among them, we note the presence of genes that are upregulated upon TAL1 KD and code for important T-cell-specific transcription factors such as TOX (Aliahmad and Kaye, 2008) and Aiolos/ikzf3 (Morgan *et al*, 1997). Conversely, the TAL1-target gene *Cdk6*, which is normally downregulated during T-cell differentiation (Grossel and Hinds, 2006), is decreased upon TAL1 KD (Figure 4C). In addition, five genes with previously described functions in apoptosis were identified as direct targets of TAL1, including four pro-apoptotic genes (i.e. *Pmaip1/Noxa* (Oda *et al*, 2000); *Stk17a/Drp1* (Inbal *et al*, 2000); *Isg20l1/Aen* (Kawase *et al*, 2008) and *Plk3* (Xie *et al*, 2001)) that are repressed by TAL1. Interestingly, TAL1 may also decrease apoptosis via upregulation of *Cd226*, which is involved in the resistance of thymocytes to apoptosis (Fang *et al*, 2009). Finally, we identified three tumour suppressor genes: *Tnfaip3* (Malynn and Ma, 2009), *Pten* (Li *et al*, 1997; Gutierrez *et al*, 2009) and *Lrp12/st7* (Zenklusen *et al*, 2001)) that are repressed by TAL1. Finally, we observed a significant enrichment (2.8-fold; *P*-value = 5.488e−13, proportions test with continuity correction) of TAL1 peaks near the signature genes associated with TAL1 expression in T-ALL patients (Ferrando *et al*, 2002), suggesting that many of these signature genes are direct targets of TAL1.

Overall, our results, which were confirmed by ChIP–qPCR (Figure 4; Supplementary Figures S6 and S7 and data not shown), RT–qPCR (Figure 4; Supplementary Figure S4 and data not shown) and western blot (Figure 3D), identified several possible mechanisms through which TAL1 may contribute to leukaemogenesis: repression of pro-apoptotic genes, tumour suppressor genes and T-cell differentiation genes, as well as activation of anti-apoptotic and anti-T-cell differentiation genes.

Validation of the role of TAL1 in additional T-ALL cell lines and in primary blasts from T-ALL patients

We next sought to confirm our findings in four additional TAL1-expressing T-ALL cell lines (i.e. CEM/C1, MOLT4, PF382 and RPMI8402) and in primary blasts from two TAL1-expressing T-ALL patients. We first used ChIP to verify the binding of TAL1 to 17 previously identified sites associated to TAL1-regulated genes, including T-cell marker genes (e.g. *Aiolos/Ikzf3*, *Tox*, *Ccr9*), pro- and anti-apoptotic genes (e.g. *Pmaip1/Noxa*, *Plk3* and *Cd226*) and other genes (Figure 5A; Supplementary Figure S7 and data not shown). TAL1 binding was confirmed at all the sites we have tested, not only in the four T-ALL cell lines, but also in the primary blasts from both T-ALL patients. Next, the KD of TAL1 was induced in these cells by infection with lentivirus expressing anti-TAL1 or scrambled (scr) shRNA (Figure 5B and C). Similarly to the phenotype observed in Jurkat cells, the KD of TAL1 leads to a decrease of *Cdk6* transcription and a reactivation of *Cd69* in all T-ALL cell lines and in the primary blasts from both T-ALL patients (Figure 5D). Finally, the apoptotic phenotype induced upon TAL1 KD is also conserved in the T-ALL cell lines and primary blasts (Figure 5E and F and data not shown). Therefore, the results we have obtained in four additional T-ALL cell lines that express TAL1 and in primary blasts from two TAL1-expressing T-ALL patients confirm our findings from Jurkat cells.

Figure 6 Differential selectivity of TAL1 genomic binding in erythroid and Jurkat cells. (**A**) DNA motifs that are overrepresented within TAL1 peaks in Jurkat (left panel) or erythroid (middle panel) cells versus control, and in Jurkat versus erythroid cells (right panel) were identified using a *de novo* motif discovery method. The motifs scores are defined as *z*-values of logistic regression (see Supplementary data). *Reverse complement of the consensus. (**B**) TAL1 displays the same preference for E-box variants in erythroid and Jurkat cells except for the E-protein homodimer's preferred variant (red squared), which is bound by TAL1 preferentially in Jurkat cells. TAL1 peak-centred E-box variants were ranked in both cell types. E-box variants are represented in the first column by their two variable nucleotides corresponding to N in the sequence: CANNTG. The genomic distribution of E-box variants is shown in the last column as percentage of total E-box sequences. (**C**) Runx and Ets motifs are more prevalent under TAL1 peaks in Jurkat cells. Cumulative distribution function (CDF) of the indicated motifs is represented relative to the absolute distance (Abs. dist.) from TAL1 peak summit in base pairs. The black line represents motifs common to Jurkat and erythroid cells. (**D**) A preferred distance of 8–10 bp between an E-box and a Gata motif is present in 10% of TAL1 peaks in erythroid cells and in 4% of TAL1 peaks in Jurkat cells (top panels). A composite motif with juxtaposed E-box and Ets motifs is present in 1% of TAL1 peaks in Jurkat cells only (bottom panels). E-box/Gata and E-box/Ets composite motifs are represented as a percentage of E-box-containing TAL1 peaks. The *x* axis corresponds to the distance from the end of the E-box to the end of the other motif. The numbers above the blue lines correspond to the distance between the two motifs. For a logo of these composite motifs, see Supplementary Figure S5D. (**E**) TAL1 interacts with RUNX1 and ETS1 but not GABPA in Jurkat nuclear extracts (NE), as shown by reciprocal immunoprecipitations (IP). Abs used in western blot are indicated on the right. Molecular weight markers (in kDa) are indicated on the left. Mock IP was done using normal IgG from corresponding organism used to generate antibodies. SN, supernatant.

TAL1 displays distinct genomic binding profiles in erythroid and Jurkat cells

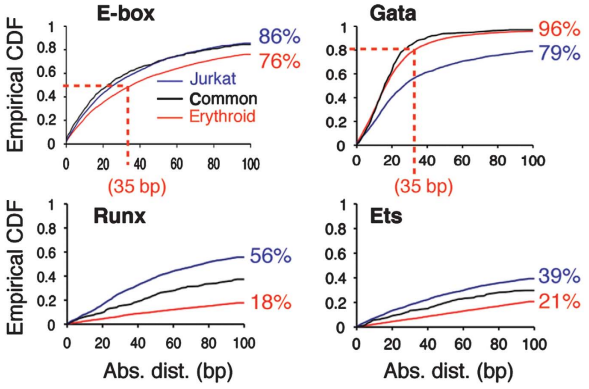
To determine what restricts TAL1 binding *in vivo*, motif analyses were performed on DNA sequences underlying TAL1 peaks (Figure 6). As TAL1 is a bHLH transcription factor known to bind to E-boxes, we first measured the proportion of E-box variants centred at TAL1 peak summits (Figure 6B). We found that the *in vitro* preferred TAL1/E-protein E-box (CAGATG) (Hsu *et al*, 1994) ranks first among TAL1 peak-centred E-boxes in both erythroid and Jurkat cells (Figure 6B), confirming that it is also the TAL1 preferred E-box *in vivo*. The frequency ranking of E-box variants is identical between Jurkat and erythroid cells, except for the E-protein homodimer's preferred E-box (CAGGTG), whose proportion doubles in Jurkat cells (Figure 6B, red square). Consistent with this, a *de novo* search for motifs overrepresented in Jurkat compared with erythroid cells also identified this same E-protein homodi-

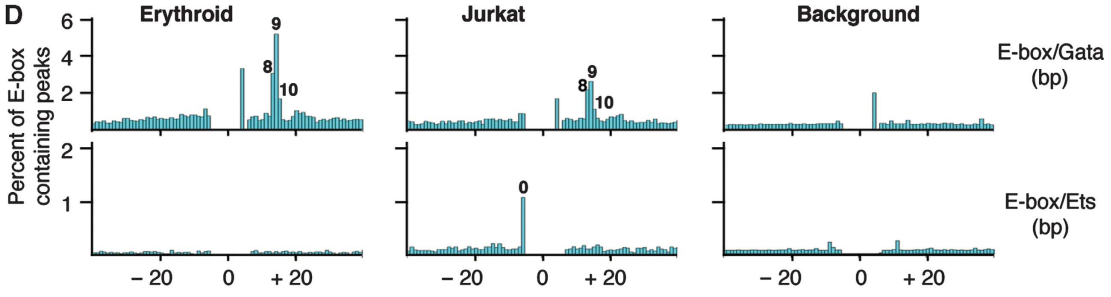
A

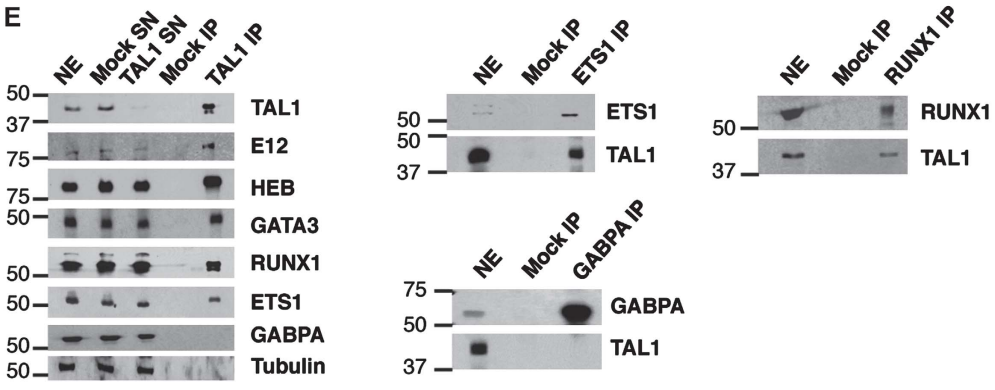
	Jurkat versus control			Erythroid versus control			Jurkat sp. versus Erythroid sp.		
	Class	Motif	Score	Class	Motif	Score	Class	Motif	Score
1	Runx	VCCACA*	49	Gata	(C)TTATCW	68	Runx	ACCACA*	33
2	Gata	(C)TTATCT*	45	E-box(GC)	CAGCTG	42	Ets	CAGGAARY	24
3	E-box(GC)	CAGCTG	40	Gata	AGATAA	38	E-box(GG)	CAGGTG	23
4	Ets	CAGGAAR	28	?	CAGGGC	30	?	CCACAD	18
5	Gata	AGATAA	19	GC-box	MCACCC*	30			
6	Runx	AACCACA	17	?	CTGGCA	25			
7	?	GCAGVC	17	Gata	(C)TAATCW	15			

B

E-box variant	Jurkat Rank	Jurkat Peak#	Jurkat %	Erythroid Rank	Erythroid Peak#	Erythroid %	Genome %
GA	1	717	24.5	1	1390	20.8	11
GC	3	573	19.6	2	1368	20.4	12
CA	4	274	9.4	3	892	13.3	12
GG	2	620	21.2	4	733	10.9	14
AC	5	224	7.7	5	663	9.9	7
AA	6	202	6.9	6	626	9.4	15
AG	7	192	6.5	7	579	8.6	10
TA	8	54	1.8	8	211	3.2	10
CG	9	45	1.5	9	153	2.3	3
AT	10	26	0.9	10	80	1.2	6







mer's preferred E-box (Figure 6A, right panel). The finding that TAL1 binds to the E-protein homodimer's preferred E-box more frequently in Jurkat than erythroid cells is consistent with the long-standing model in which TAL1 deregulates E-protein homodimers' function in T-ALL (Begley and Green, 1999; O'Neil and Look, 2007).

On the other hand, E-boxes are absent from an unexpectedly high proportion of TAL1 peaks (14% in Jurkat and 24% in erythroid cells) (Figure 6C). Consistent with this, the *de novo* motif search did not identify E-boxes as the top overrepresented motif in erythroid or Jurkat cells (Figure 6A, left and middle panels). Instead, in erythroid cells, a Gata motif ranks first within overrepresented sites, and two variants of this motif are also among the top 7 overrepresented motifs (Figure 6A, middle panel). Furthermore, virtually all TAL1 peaks (96%) contain a Gata motif while only 76% contain an E-box within a 100-bp radius of the peak summit (Figure 6C). In erythroid cells, Gata motifs are also on average closer to TAL1 peak summits than E-boxes, with 80% of TAL1 peaks containing a Gata site within 35 bp of the peak summit versus only 50% containing an E-box within this distance (Figure 6C). This is consistent with previous studies showing cooccupation of TAL1 and GATA1 at many genomic sites in erythroid cells (Cheng *et al*, 2009; Tripic *et al*, 2009; Soler *et al*, 2010). In Jurkat cells, Gata sites are also highly prevalent since they are present in 79% of TAL1 peaks (Figure 6A and C). These results suggest that GATA factors are important for targeting TAL1 to specific regions of the genome in both erythroid and Jurkat cells.

Strikingly, a Runx-binding site was identified as the top-ranking overrepresented motif in Jurkat cells when comparing to either control regions that share the same GC content distribution (Figure 6A, left panel), or regions specifically bound in Jurkat cells (Jurkat sp) versus those specifically bound in erythroid cells (erythroid sp) (Figure 6A, right panel). Furthermore, Runx motifs occur in 56% of TAL1 peaks in Jurkat cells (versus 18% only in erythroid cells) (Figure 6C). An Ets-binding site is the second most overrepresented motif when comparing Jurkat versus erythroid cells (Figure 6A, right panel) and is also overrepresented in Jurkat peaks versus control sequences (Figure 6A, left panel). In agreement with this, Ets motifs are present in 39% of TAL1 peaks in Jurkat cells (versus 21% in erythroid cells) (Figure 6C). Interestingly, the Ets motif identified under TAL1 peaks in Jurkat cells by our *de novo* search (CAGGAA(A/G); Figure 6A) resembles the recently identified ETS1-specific motif variant (CAGGA(A/T)GT) that is predominantly associated to T-cell-specific enhancers as opposed to the 'redundant' variant (CCGGAAGT) that can also be bound by other ETS-family members (e.g. GABPA) and is mostly located in promoter regions of housekeeping genes (Hollenhorst *et al*, 2009).

To get further insight into a potential binding pattern of TAL1 relative to the other transcription factors for which we detected overrepresented binding sites (i.e. GATA, RUNX and ETS), we searched within TAL1 peaks for preferred distances between E-boxes and the DNA motifs underlying binding of these factors (Figure 6D; Supplementary Figure S5D). About 10% of TAL1 peaks in erythroid cells and 4% in Jurkat cells display a preferred distance of 8–10 bp between E-box and Gata sites (Figure 6D; Supplementary Figure S5D). Remarkably, this corresponds to the composite E-box/Gata motif previously identified *in vitro* (Wadman *et al*, 1997) and *in vivo* in erythroid cells (Fujiwara *et al*, 2009; Kassouf *et al*, 2010; Soler *et al*, 2010). Precisely in erythroid cells, we counted 218/397/116 occurrences of the E-box-Gata motif with spacing of 8/9/10 bp, respectively. In Jurkat cells, we counted 94/99/50 occurrences of the E-box-Gata motif with spacing of 8/9/10 bp, respectively. We also found a significant preference (occurring in about 1% of E-box-containing TAL1 peaks; 99 occurrences) for a novel composite juxtaposed Ets/E-box motif specific to Jurkat cells (Figure 6D; Supplementary Figure S5D). Finally, no preferred distance was detected for Runx motifs with respect to E-boxes.

The identification of Runx- and Ets-binding sites within TAL1 peaks suggest that similarly to GATA factors, RUNX and ETS proteins might be involved in guiding TAL1 in close proximity to their binding sites in T-ALL cells. In agreement with this possibility, reciprocal co-IP experiments show that TAL1 interacts with RUNX1 and ETS1 factors in Jurkat nuclear extracts (Figure 6E). In contrast, no interaction was detected between TAL1 and GABPA, another ETS-family member that is highly expressed in Jurkat cells (Hollenhorst *et al*, 2004) and predominantly targets promoters of housekeeping genes, as opposed to ETS1 which targets T-cell-specific enhancers (Hollenhorst *et al*, 2009). The fact that TAL1 interacts specifically with ETS1 is highly consistent with our identification of the ETS1-specific motif (but not the redundant Ets motif that also binds GABPA) as being enriched under TAL1 peaks (Figure 6A). Interestingly, the *Runx1* and *Ets1* genes are direct targets of TAL1 (Supplementary Figure S7) and both genes are downregulated upon TAL1 KD at the transcript (Supplementary Figure S4B) and protein (Figure 3D) levels.

RUNX1/3 and ETS1 are required for TAL1 binding to genomic loci in Jurkat cells

Since genome-wide localization of RUNX1/3 (using an antibody that recognizes both RUNX1 and RUNX3) and ETS1 factors has recently been determined in Jurkat cells (Hollenhorst *et al*, 2009), we looked for genomic overlap of these factors with TAL1. Notably, we found colocalization of TAL1, RUNX1/3 and ETS1 on the representative TAL1 peaks of the *Cdk6* and *Cd69* genes (Figure 7A). Furthermore, genome-wide comparisons indicate that 50% of TAL1 peaks containing a Runx motif are bound by RUNX1/3 in Jurkat cells while 65% of TAL1 peaks containing an Ets motif are bound by ETS1. While we did find the previously described ETS1-RUNX-cobound composite motif (Hollenhorst *et al*, 2009) in a subset of our TAL1/RUNX/ETS1 shared peaks, this motif was also present in our TAL1/ETS1 (no RUNX) cobound sites.

To assess the importance of the interplay between TAL1, RUNX1/3 and ETS1, we used lentiviral-mediated KD of RUNX1/3 or ETS1 in Jurkat cells. We found that decreased levels of either RUNX1/3 or ETS1 (Figure 7B and C) lead to a decrease of TAL1 genomic binding at all sites tested (Figure 7D; Supplementary Figure S8), which confirms an important role for these transcription factors in targeting TAL1 to specific loci. Notably, the binding of TAL1 to the important T-cell-specific genes (i.e. *Tox*, *Ikzf3/Aiolos*, *Ccr9*, *Cd69*, *Tcrbv*), apoptotic genes (i.e. *Cd226*, *Pmaip1/Noxa*, *Plk3*) and other genes requires RUNX1/3 and/or ETS1 (Figure 7; Supplementary Figure S8), suggesting that these

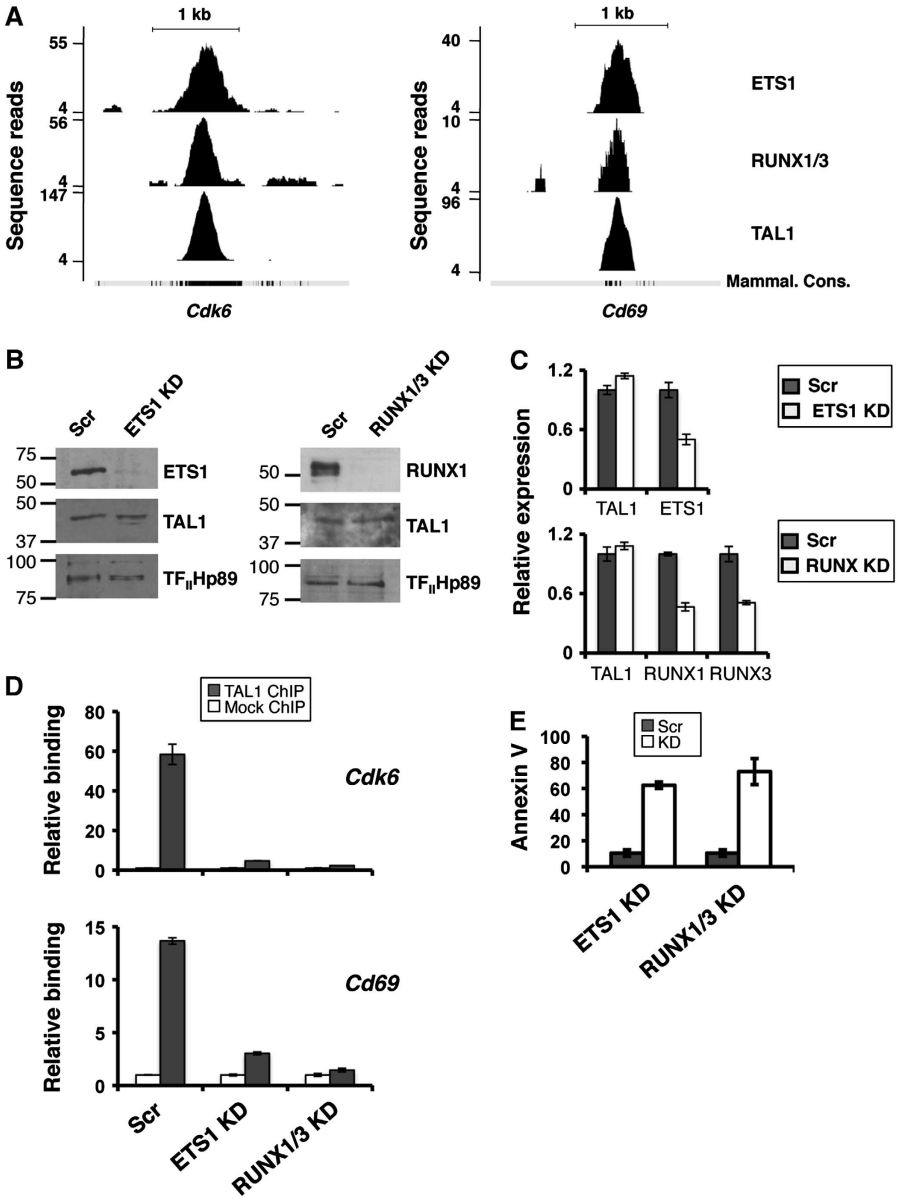


Figure 7 TAL1 genomic binding in Jurkat cells requires ETS1 and RUNX transcription factors. (**A**) ChIP-seq results showing ETS1, RUNX1/3 and TAL1 colocalization on representative sites in *cdk6* second intron and *cd69* promoter (see Figure 4C). ETS1 and RUNX1/3 binding data were extracted from ChIP-seq published by Hollenhorst *et al* (2009). The *y* axis indicates the number of overlapping sequence reads. Mammalian conservation is indicated at the bottom. (**B**, **C**) Single knockdown of ETS1 (ETS1 KD) and double knockdown of RUNX1 and RUNX3 (RUNX1/3 KD) is induced by infection with lentivirus expressing anti-Ets1 and anti-Runx1/Runx3 respectively, but not scramble (Scr) shRNA as verified by western blot (**B**) and RT–qPCR (**C**). TAL1 level is not affected by the KD of ETS1 and RUNX1/3. (**D**) ETS1 and RUNX1/3 KD lead to a decrease of TAL1 binding to representative sites in *cdk6* second intron and *cd69* promoter (Figure 4C) as measured by ChIP-qPCR. TAL1 binding is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG. See Supplementary Figure S8 for additional TAL1-binding sites. (**E**) ETS1 and RUNX1/3 KD leads to an increase in apoptosis as measured by Annexin V staining. Necrotic cells that are positive for 7AAD are excluded. Numbers represent the percentage of positive cells. Error bars represent standard deviations of biological triplicates.

additional transcription factors may be important contributors to TAL1-mediated leukaemogenesis. Consistent with this possibility, an apoptotic phenotype is observed in Jurkat cells upon KD of RUNX1/3 and ETS1 (Figure 7E).

Discussion

Composite DNA motifs have an important function in selecting TAL1-binding sites in vivo

The question of how a transcription factor selects its binding site *in vivo* has been the centre of recent focus (Pan *et al*, 2009). To address this question, we have examined the genomic binding of TAL1 in two cellular contexts. Importantly, in our study, the cellular context reflects both the lineage (erythroid versus T cell) and the state (normal versus malignant) of the cell. While these contributing factors are not distinguished, both are physiologically relevant since TAL1 is oncogenic in the T-cell lineage (Condorelli *et al*, 1996; Kelliher *et al*, 1996). Strikingly, we found that the binding profile of TAL1 and its preference for particular composite DNA motifs vary depending on the cellular context. Common properties of TAL1-binding sites in

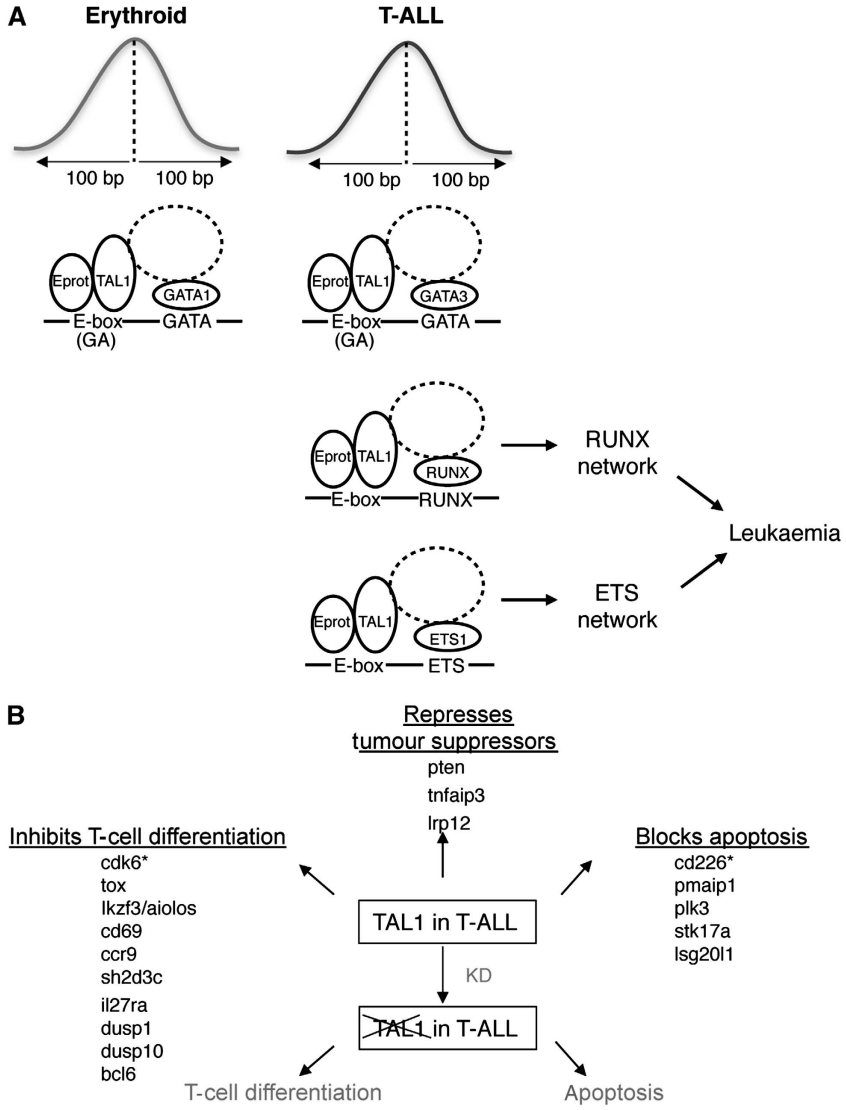


Figure 8 Model of TAL1-mediated leukaemogenesis. (**A**) TAL1 binds to different composite DNA motifs depending on the cellular environment. This alternate binding site selection provides TAL1 with access to RUNX and ETS transcriptional regulatory networks preferentially in T-ALL. Deregulation of transcriptional networks controlled by these important activators of T-cell differentiation could have a crucial role in the arrest of T-cell differentiation, potentially leading to T-cell leukaemia. (**B**) In the T-cell lineage, TAL1 inhibits tumour suppressors and mediators of differentiation and apoptosis. Upon TAL1 KD, T-ALL cells resume part of the T-cell transcriptional program and enter apoptosis. TAL1-regulated genes identified in our study, which could mediate these functions are indicated. All genes are repressed by TAL1, except those with an asterisk whose expression is increased by TAL1 in T-ALL.

erythroid and T-ALL cells include a high prevalence of E-boxes and Gata motifs, and a preference for the same E-box variant CAGATG. DNA motifs that are recognized by TAL1 preferentially in a T-ALL environment include Runx- and Ets-binding sites as well as another E-box variant (CAGGTG), which *in vitro* is preferentially bound by E-protein homodimers (Hsu *et al*, 1994). Our finding that the same transcription factor displays distinct DNA-binding profiles in two cell types was unexpected and reveals an additional level of complexity in the mechanism by which transcription factors select their binding sites *in vivo*. Specifically, we found that TAL1 recognizes a combination of DNA motifs comprised of TAL1's direct binding site (E-box) and that of other transcription factors with which it interacts, including GATA factors as well as RUNX and ETS proteins (Figure 8A).

An important role for GATA factors in targeting TAL1 has been suggested previously. For example, Gata motifs have an important function in mediating TAL1 binding to specific sites (Ono *et al*, 1998; Vyas *et al*, 1999; Song *et al*, 2007). In addition, recent genome-wide studies in erythroid cells have shown that a large proportion of GATA1-bound genomic sites are also occupied by TAL1 (Cheng *et al*, 2009; Soler *et al*, 2010). Furthermore, the E-box/Gata composite motif, which underlies the binding of a GATA1–TAL1 pentameric complex (Wadman *et al*, 1997), is enriched under TAL1 peaks (Kassouf *et al*, 2010) and under GATA1 peaks (Fujiwara *et al*, 2009) in erythroid cells. Finally, a recent ChIP-seq study in mouse primary fetal erythroid cells showed that a TAL1 mutant that has lost its E-box-mediated DNA-binding activity can still be recruited to one fifth of TAL1 targets, and TAL1 and GATA1 cooperate to stabilize each other at these sites (Kassouf *et al*,

2010). Our results are consistent with these findings and provide evidence that a similar mechanism likely occurs in T-ALL cells. Indeed, Gata motifs are largely overrepresented within the 200-bp region underlying TAL1 binding in both cell types studied. Furthermore, between 50 and 80% of Gata sites (depending on cell type) are centred within 35 bp of TAL1 peak summit. Finally, our ChIP-seq study detected an overrepresentation of the E-box/Gata composite motif in both erythroid and T-ALL. Therefore, TAL1-interacting GATA factors appear to have an important function in mediating TAL1 DNA binding *in vivo* in both erythroid and T-ALL cells.

Specificity of TAL1 binding in T-ALL

In addition to Gata motif, we found an overrepresentation of Ets and Runx motifs within TAL1-binding sites in T-ALL cells, and comparative bioinformatic analysis shows extensive binding of these TAL1-associated sites by RUNX1/3 and ETS1 factors in Jurkat cells. Furthermore, our co-IP experiments show that TAL1 interacts with RUNX1 and ETS1 proteins and knocking down RUNX1/3 or ETS1 factors leads to a disruption of TAL1 genomic binding at specific genes. Spatial analysis of TAL1 sites containing an Ets motif identified a novel T-ALL-specific composite E-box/Ets motif where the two DNA elements are juxtaposed. These findings suggest that RUNX1 and ETS1 may be involved in guiding TAL1 to specific genomic sites in the T-cell lineage.

Analysis of the Ets motifs found under TAL1 peaks in Jurkat cells revealed an enrichment for the recently identified ETS1-specific motif but not for the 'redundant' Ets motif, which can be bound by either ETS1 or the other ETS-family member GABPA (Hollenhorst *et al*, 2009). This is particularly interesting since the 'redundant' Ets motif is found within housekeeping gene promoters while the ETS1-specific motif is specifically enriched within enhancers of T-cell genes (Hollenhorst *et al*, 2009). Furthermore, our co-IP studies show that TAL1 does not interact with GABPA. These key findings are consistent with the idea that TAL1 acts on T-cell-specific genes in T-ALL. Furthermore, this specificity might at least partially be mediated through TAL1 interaction with ETS1. Taken together, our results reveal that RUNX1/3 and ETS1 are important mediators of TAL1 genomic binding in leukaemic T cells, notably at genes involved in T-cell differentiation (e.g. *Tox*, *Ikzf3/Aiolos*) and apoptosis (e.g. *Cd226*, *Pmaip1*, *Plk3*). Combined with the apoptotic phenotype observed upon their KD in T-ALL, these findings provide strong support for a model in which RUNX1/3 and ETS1 transcription factors are important contributors to the mechanism of TAL1-mediated leukaemogenesis (Figure 8A).

Interplay between TAL1, ETS1 and RUNX factors in T-ALL leukaemogenesis

RUNX and ETS proteins have essential roles in regulating T-cell differentiation (Anderson, 2006). In addition, both families of transcription factors often become oncogenic when mutated or overexpressed (Blyth *et al*, 2005; Gallant and Gilkeson, 2006). In light of our finding that RUNX1/3 and ETS1 are important for mediating TAL1 genomic binding, it is interesting to note that TAL1 also appears to directly regulate the expression and function of RUNX and ETS proteins, which suggests a multipronged mechanism for TAL1-mediated leukaemogenesis. Indeed, TAL1 directly targets the *Runx1* gene whose expression is upregulated by TAL1 in both Jurkat cells and blasts from T-ALL patients. In addition, TAL1 represses the *Pim1* kinase gene (Supplementary Table VI), whose product enhances RUNX1 activity via phosphorylation (Aho *et al*, 2006). Therefore, TAL1 acts at multiple levels to disrupt the transcriptional network regulated by RUNX1. This is of particular interest since RUNX1 is part of the 'gene expression signature' of TAL1-positive T-ALL (Ferrando *et al*, 2002).

Similarly, TAL1 appears to target the ETS1 transcriptional network. Indeed, the *Ets1* gene is also a direct target of TAL1, its expression is upregulated by TAL1 in both Jurkat cells and primary T-ALL blasts, and the ETS1 protein interacts with TAL1. Since ETS1 promotes the survival of T cells (Muthusamy *et al*, 1995) and is overexpressed in T-ALL (Sacchi *et al*, 1988), deregulation of its transcriptional network by TAL1 might also have a major role in TAL1-mediated leukaemogenesis. Interestingly, Runx- and Ets-bound motifs have also been identified as TAL1 targets in a murine haematopoietic progenitor cell line (Wilson *et al*, 2010), suggesting that aberrant expression of TAL1 in a T-cell precursor environment could cause these cells to acquire/retain some properties of multipotent progenitors.

Target genes as potential mediators of TAL1-mediated leukaemogenesis

A number of our newly identified TAL1-regulated target genes code for proteins that are involved in T-cell differentiation and apoptosis. In addition, tumour suppressors that are targeted by TAL1 are upregulated upon TAL1 KD, suggesting that TAL1 represses these genes in T-ALL (Figure 8B). While all TAL1-regulated direct targets (Supplementary Table VI) are candidate effectors of TAL1-mediated leukaemogenesis, *Cdk6*, which is activated by TAL1, presents a particular interest. Indeed, this gene must be downregulated for T-cell differentiation to proceed (Grossel and Hinds, 2006) and forced expression of CDK6 contributes to the development of lymphoid malignancies in mice (Schwartz *et al*, 2006). Importantly, CDK6 is overexpressed in human T-ALL (Chilosi *et al*, 1998) and mice lacking CDK6 are completely resistant to T-cell malignancies induced by constitutively active AKT1 (Hu *et al*, 2009). Since constitutive activation of Akt has been proposed to be involved in resistance to the NOTCH1 inhibition therapy by small molecule γ-secretase inhibitors (Palomero *et al*, 2008), CDK6 (which acts downstream of the AKT pathway) may represent a valuable alternative therapeutic target for treating T-ALL.

Here, we have performed the first genome-wide comparison of TAL1 function in an erythroid versus T-cell environment. This study revealed an unexpected contribution of the cellular environment in selecting particular composite DNA motifs for TAL1 binding *in vivo*. T-ALL-specific genomic targeting provides the TAL1 protein with access to transcriptional regulatory networks of key regulators of T cell, and leads to inhibition of differentiation in this lineage. Validation in T-ALL patients of key functional targets of TAL1 identified by this approach demonstrates the utility of such a comparative strategy in identifying mediators/effectors of oncogenesis. Taken together, our findings underscore the cellular environment as an important factor in the genomic binding selectivity of transcription factors and suggest how changing the cellular context can render a transcription factor oncogenic.

Materials and methods

Cell culture

Light-density mononuclear cells were isolated by ficoll density gradient centrifugation (GE Healthcare) from G-CSF-mobilized adult blood from donors without haematological malignancies (Ottawa Hospital Research Ethics Board #2007804-01H). CD34 + cells were enriched through positive immunomagnetic selection using the CD34 MicroBead Kit (Miltenyi Biotec Inc.) (purity $>95 \pm 3\%$) and differentiated *ex vivo* as described in Giarratana *et al* (2005) except that we used 20% BIT (StemCell Technologies). Jurkat cells stably expressing the Tet repressor (Invitrogen) were cultured in RPMI 1640 supplemented with 10% (v/v) Tet-free FBS, 100 U/ml penicillin, 100 μg/ml streptomycin and 20 μg/ml blasticidin. CEM/C1, MOLT4 (obtained from ATCC) and PF382, RPMI8402 (obtained from DSMZ) were cultured in RPMI 1640 supplemented with 10% (v/v) FBS, 100 U/ml penicillin, 100 μg/ml streptomycin.

Expression profiling on Affymetrix microarray

Total RNA was purified using the RNeasy Mini Kit (Qiagen). Labelling and hybridization to the Affymetrix Human Gene 1.0 ST gene expression microarray were performed following standard Affymetrix procedures. See Supplementary data. The microarray raw data are deposited into GEO (accession number GSE20546).

Gene set enrichment analysis

GSEA (Subramanian *et al*, 2005) with gene sets from Lee *et al* (2004) was performed on the Jurkat array set with the default parameters. For a description of the genes sets used, see Supplementary data.

Chromatin immunoprecipitation

ChIP experiments using anti-TAL1 Ab or normal IgG as a negative control were carried out as described in Chaturvedi *et al* (2009) except that chromatin was fragmented to a size range of 100–300 bp. For Solexa sequencing, ChIPed DNA was prepared according to the Illumina protocol with two modifications: (1) DNA fragments ranging from 150 to 300 bp were selected at the gel selection step and (2) 21 instead of 18 cycles of PCR were done at the amplification step, as previously described (Cao *et al*, 2010). The samples were sequenced using the Illumina Genome Analyzer II. For qPCR, DNA was quantified using SYBROGreen. Immunoprecipitated DNA was quantified using a standard curve generated with genomic DNA and were normalized by dividing by the amount of the corresponding target in the input fraction. The enrichment at each DNA target is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG.

High-throughput DNA sequencing data analysis

Sequences were extracted using the Firecrest and Bustard programs from package GApipeline-0.3.0. Reads were aligned using MAQ (version 0.6.6) to the human reference genome (hg18) as described in Cao *et al* (2010). For a complete description of data analysis including sequencing depth coverage, see Supplementary data. ChIP-seq data are deposited in GEO (accession number GSE25000).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank L Douay and MC Giarratana (Université Paris VI, France) and R Pasha (OHRI) for advice on the *ex vivo* erythroid differentiation, D Trono (Ecole Polytechnique Federale de Lausanne) for lentiviral vectors, L Filion (University of Ottawa) for advice on FACS, J Wu (OHRI) for technical help, S Huang (University of Florida), B Gottgens (University of Cambridge), L Megeney, Y Kawabe and M Rudnicki (OHRI) for discussion, shared data and reagents. This project was funded by a CIHR grant MOP-82813 to MB and an NIH NIAMS R01AR045113 grant to SJT. CGP is a recipient of a Ontario Research Fund Computational Regulomics Training Postdoctoral Fellowship. MB holds the Canada Research Chair in the Regulation of Gene Expression. FJD holds the Canada Research Chair in the Epigenetic Regulation of Transcription.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aho TL, Sandholm J, Peltola KJ, Ito Y, Koskinen PJ (2006) Pim-1 kinase phosphorylates RUNX family transcription factors and enhances their activity. *BMC Cell Biol* **7:** 21

Aifantis I, Raetz E, Buonamici S (2008) Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat Rev Immunol* **8:** 380–390

Aliahmad P, Kaye J (2008) Development of all CD4 T lineages requires nuclear factor TOX. *J Exp Med* **205:** 245–256

Anderson MK (2006) At the crossroads: diverse roles of early thymocyte transcriptional regulators. *Immunol Rev* **209:** 191–211

Begley CG, Green AR (1999) The SCL gene: from case report to critical hematopoietic regulator. *Blood* **93:** 2760–2770

Bendelac A, Matzinger P, Seder RA, Paul WE, Schwartz RH (1992) Activation events during thymic selection. *J Exp Med* **175:** 731–742

Bernard M, Delabesse E, Smit L, Millien C, Kirsch IR, Strominger JL, Macintyre EA (1998) Helix-loop-helix (E2-5, HEB, TAL1 and Id1) protein interaction with the TCRalphadelta enhancers. *Int Immunol* **10:** 1539–1549

Blyth K, Cameron ER, Neil JC (2005) The RUNX genes: gain or loss of function in cancer. *Nat Rev Cancer* **5:** 376–387

Brunet de la Grange P, Armstrong F, Duval V, Rouyez MC, Goardon N, Romeo PH, Pflumio F (2006) Low SCL/TAL1 expression reveals its major role in adult hematopoietic myeloid progenitors and stem cells. *Blood* **108:** 2998–3004

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, Tapscott SJ (2010) Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18:** 662–674

Chang PY, Draheim K, Kelliher MA, Miyamoto S (2006) NFKB1 is a direct target of the TAL1 oncoprotein in human T leukemia cells. *Cancer Res* **66:** 6008–6013

Chaturvedi CP, Hosey AM, Palii C, Perez-Iratxeta C, Nakatani Y, Ranish JA, Dilworth FJ, Brand M (2009) Dual role for the methyltransferase G9a in the maintenance of {beta}-globin gene transcription in adult erythroid cells. *Proc Natl Acad Sci USA* **106:** 18303–18308

Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, King DC, Chen KB, Zhang Y, Drautz D, Giardine B, Schuster SC, Miller W, Chiaromonte F, Blobel GA, Weiss MJ, Hardison RC (2009) Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19:** 2172–2184

Chilosi M, Doglioni C, Yan Z, Lestani M, Menestrina F, Sorio C, Benedetti A, Vinante F, Pizzolo G, Inghirami G (1998) Differential expression of cyclin-dependent kinase 6 in cortical thymocytes and T-cell lymphoblastic lymphoma/leukemia. *Am J Pathol* **152:** 209–217

Condorelli GL, Facchiano F, Valtieri M, Proietti E, Vitelli L, Lulli V, Huebner K, Peschle C, Croce CM (1996) T-cell-directed TAL-1 expression induces T-cell malignancies in transgenic mice. *Cancer Res* **56:** 5113–5119

Fang L, Zhang X, Miao J, Zhao F, Yang K, Zhuang R, Bujard H, Wei Y, Yang A, Chen L, Jin B (2009) Expression of CD226 antagonizes apoptotic cell death in murine thymocytes. *J Immunol* **182:** 5453–5460

Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, Lander ES, Golub TR, Look AT (2002) Gene expression signatures define

novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1:** 75–87

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

Gallant S, Gilkeson G (2006) ETS transcription factors and regulation of immunity. *Arch Immunol Ther Exp (Warsz)* **54:** 149–163

Giarratana MC, Kobari L, Lapillonne H, Chalmers D, Kiger L, Cynober T, Marden MC, Wajcman H, Douay L (2005) *Ex vivo* generation of fully mature human red blood cells from hematopoietic stem cells. *Nat Biotechnol* **23:** 69–74

Grossel MJ, Hinds PW (2006) Beyond the cell cycle: a new role for Cdk6 in differentiation. *J Cell Biochem* **97:** 485–493

Gutierrez A, Sanda T, Grebliunaite R, Carracedo A, Salmena L, Ahn Y, Dahlberg S, Neuberg D, Moreau LA, Winter SS, Larson R, Zhang J, Protopopov A, Chin L, Pandolfi PP, Silverman LB, Hunger SP, Sallan SE, Look AT (2009) High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. *Blood* **114:** 647–650

Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, Graves BJ (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5:** e1000778

Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **32:** 5693–5702

Hsu HL, Huang L, Tsan JT, Funk W, Wright WE, Hu JS, Kingston RE, Baer R (1994) Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol Cell Biol* **14:** 1256–1265

Hu MG, Deshpande A, Enos M, Mao D, Hinds EA, Hu GF, Chang R, Guo Z, Dose M, Mao C, Tsichlis PN, Gounari F, Hinds PW (2009) A requirement for cyclin-dependent kinase 6 in thymocyte development and tumorigenesis. *Cancer Res* **69:** 810–818

Inbal B, Shani G, Cohen O, Kissil JL, Kimchi A (2000) Death-associated protein kinase-related protein 1, a novel serine/threonine kinase involved in apoptosis. *Mol Cell Biol* **20:** 1044–1054

Kassouf MT, Hughes JR, Taylor S, McGowan SJ, Soneji S, Green AL, Vyas P, Porcher C (2010) Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome Res* **20:** 1064–1083

Kawase T, Ichikawa H, Ohta T, Nozaki N, Tashiro F, Ohki R, Taya Y (2008) p53 target gene AEN is a nuclear exonuclease required for p53-dependent apoptosis. *Oncogene* **27:** 3797–3810

Kelliher MA, Seldin DC, Leder P (1996) Tal-1 induces T cell acute lymphoblastic leukemia accelerated by casein kinase IIalpha. *EMBO J* **15:** 5160–5166

Lacombe J, Herblot S, Rojas-Sutterlin S, Haman A, Barakat S, Iscove NN, Sauvageau G, Hoang T (2010) Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. *Blood* **115:** 792–803

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

Landry JR, Bonadies N, Kinston S, Knezevic K, Wilson NK, Oram SH, Janes M, Piltz S, Hammett M, Carter J, Hamilton T, Donaldson IJ, Lacaud G, Frampton J, Follows G, Kouskoff V, Gottgens B (2009) Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood* **113:** 5783–5792

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

Lecuyer E, Hoang T (2004) SCL: from the origin of hematopoiesis to stem cells and leukemia. *Exp Hematol* **32:** 11–24

Lee MS, Hanspers K, Barker CS, Korn AP, McCune JM (2004) Gene expression profiles during human CD4+ T cell differentiation. *Int Immunol* **16:** 1109–1124

Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275:** 1943–1947

Malynn BA, Ma A (2009) A20 takes on tumors: tumor suppression by an ubiquitin-editing enzyme. *J Exp Med* **206:** 977–980

Morgan B, Sun L, Avitahl N, Andrikopoulos K, Ikeda T, Gonzales E, Wu P, Neben S, Georgopoulos K (1997) Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J* **16:** 2004–2013

Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. *Nature* **377:** 639–642

O'Neil J, Look AT (2007) Mechanisms of transcription factor deregulation in lymphoid cell transformation. *Oncogene* **26:** 6838–6849

Oda E, Ohki R, Murasawa H, Nemoto J, Shibue T, Yamashita T, Tokino T, Taniguchi T, Tanaka N (2000) Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis. *Science* **288:** 1053–1058

Ono Y, Fukuhara N, Yoshie O (1998) TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. *Mol Cell Biol* **18:** 6939–6950

Palomero T, Dominguez M, Ferrando AA (2008) The role of the PTEN/AKT Pathway in NOTCH1-induced leukemia. *Cell Cycle* **7:** 965–970

Palomero T, Odom DT, O'Neil J, Ferrando AA, Margolin A, Neuberg DS, Winter SS, Larson RS, Li W, Liu XS, Young RA, Look AT (2006) Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. *Blood* **108:** 986–992

Pan Y, Tsai CJ, Ma B, Nussinov R (2009) How do transcription factors select specific binding sites in the genome? *Nat Struct Mol Biol* **16:** 1118–1120

Quong MW, Romanow WJ, Murre C (2002) E protein function in lymphocyte development. *Annu Rev Immunol* **20:** 301–322

Reynaud D, Ravet E, Titeux M, Mazurier F, Renia L, Dubart-Kupperschmitt A, Romeo PH, Pflumio F (2005) SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. *Blood* **106:** 2318–2328

Sacchi N, de Klein A, Showalter SD, Bigi G, Papas TS (1988) High expression of ets-1 gene in human thymocytes and immature T leukemic cells. *Leukemia* **2:** 12–18

Sancho D, Gomez M, Sanchez-Madrid F (2005) CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol* **26:** 136–140

Schilham MW, Moerer P, Cumano A, Clevers HC (1997) Sox-4 facilitates thymocyte differentiation. *Eur J Immunol* **27:** 1292–1295

Schneider U, Schwenk HU, Bornkamm G (1977) Characterization of EBV-genome negative 'null' and 'T' cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer* **19:** 621–626

Schwartz R, Engel I, Fallahi-Sichani M, Petrie HT, Murre C (2006) Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. *Proc Natl Acad Sci USA* **103:** 9976–9981

Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van Ijcken W, Hou J, Steinhoff C, Rijkers E, Lenhard B, Grosveld F (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 beta-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

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545–15550

Tenen DG (2003) Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3:** 89–101

Ting CN, Olson MC, Barton KP, Leiden JM (1996) Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* **384:** 474–478

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

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

Wen J, Huang S, Pack SD, Yu X, Brandt SJ, Noguchi CT (2005) Tal1/SCL binding to pericentromeric DNA represses transcription. *J Biol Chem* **280:** 12956–12966

Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, Chilarska PM, Kinston S, Ouwehand WH, Dzierzak E, Pimanda JE, de Bruijn MF, Gottgens B (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7:** 532–544

Wilson NK, Miranda-Saavedra D, Kinston S, Bonadies N, Foster SD, Calero-Nieto F, Dawson MA, Donaldson IJ, Dumon S, Frampton J, Janky R, Sun XH, Teichmann SA, Bannister AJ, Gottgens B (2009) The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood* **113:** 5456–5465

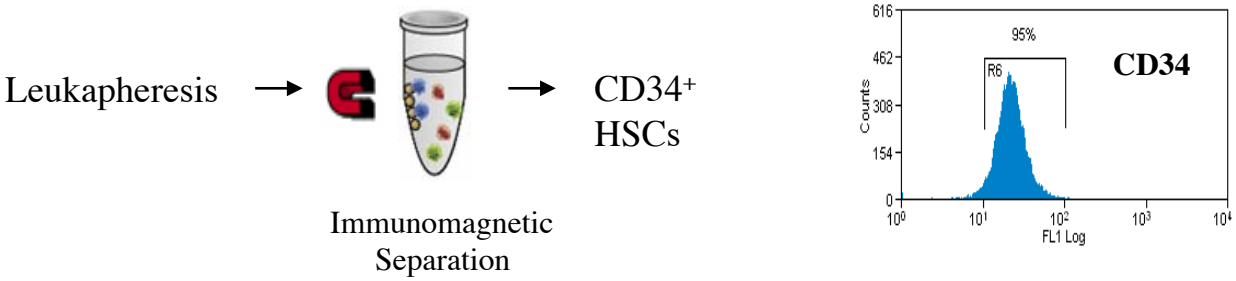
Xie S, Wu H, Wang Q, Cogswell JP, Husain I, Conn C, Stambrook P, Jhanwar-Uniyal M, Dai W (2001) Plk3 functionally links DNA damage to cell cycle arrest and apoptosis at least in part via the p53 pathway. *J Biol Chem* **276:** 43305–43312

Xu Z, Huang S, Chang LS, 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

Zenklusen JC, Conti CJ, Green ED (2001) Mutational and functional analyses reveal that ST7 is a highly conserved tumor-suppressor gene on human chromosome 7q31. *Nat Genet* **27:** 392–398

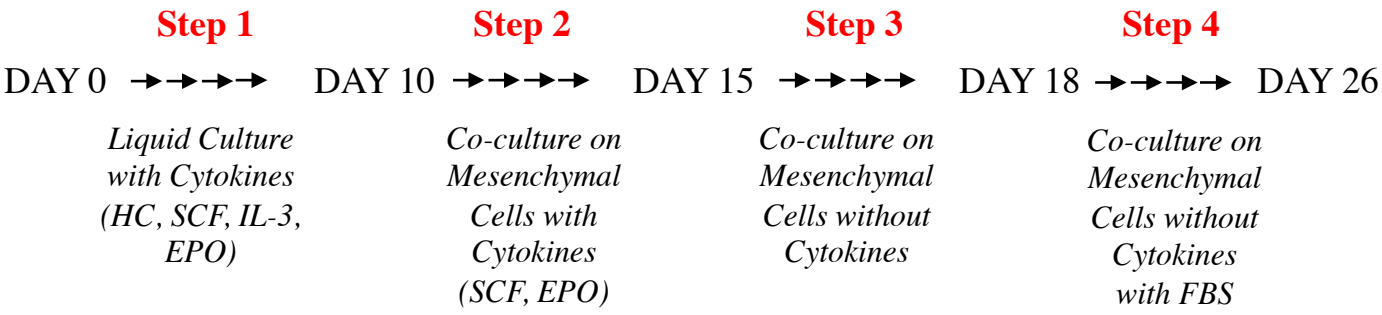
A

STEP 1 ***Positive Isolation of CD34+ human hematopoietic stem (HSCs) cells***



B

STEP 2 ***Serum-free Liquid Culture and Erythroid Differentiation***



C

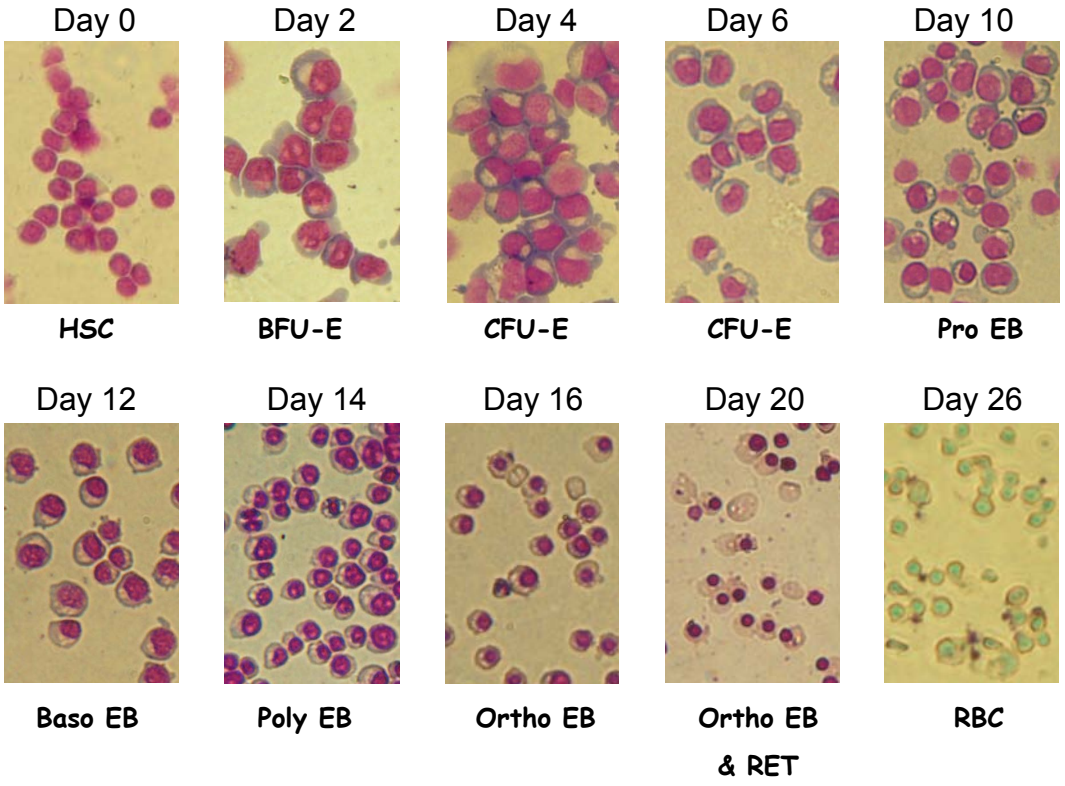


Figure S1

Figure S1: Protocol outline for the isolation and *ex vivo* erythroid differentiation of human primary hematopoietic progenitors

(A) CD34+ hematopoietic progenitors (purity of 95 ± 3%) are isolated from human peripheral blood mobilized with G-CSF (leukapheresis) using positive immunomagnetic selection.

(B) CD34+ cells are grown in liquid culture according to the indicated four-step protocol adapted from (Giarratana et al., 2005). In the first step, cells are grown in a serum-free medium supplemented with hydrocortisone (HC), stem cell factor (SCF), interleukin 3 (IL-3) and erythropoietin (EPO). In the second step, cells are co-cultured on mesenchymal MS-5 cells in the presence of SCF and EPO. In the third step, cells are co-cultured on MS-5 without growth factors. In the last step, erythroid cells are maintained on MS-5 in a medium supplemented with 10% fetal bovine serum (FBS).

(C) Cell morphological analysis during *ex vivo* erythroid differentiation. Cells are stained with May-Grünwald-Giemsa reagent at the indicated time points during differentiation (magnification x 40). HSC: hematopoietic stem cells; BFU-E: burst forming units –erythroid; CFU-E: colony forming units –erythroid; Pro EB: pro-erythroblasts; Baso EB: basophilic erythroblasts; Poly EB: polychromatophilic erythroblasts; Ortho EB: orthochromatic erythroblasts; RET: reticulocytes; RBC: red blood cells.

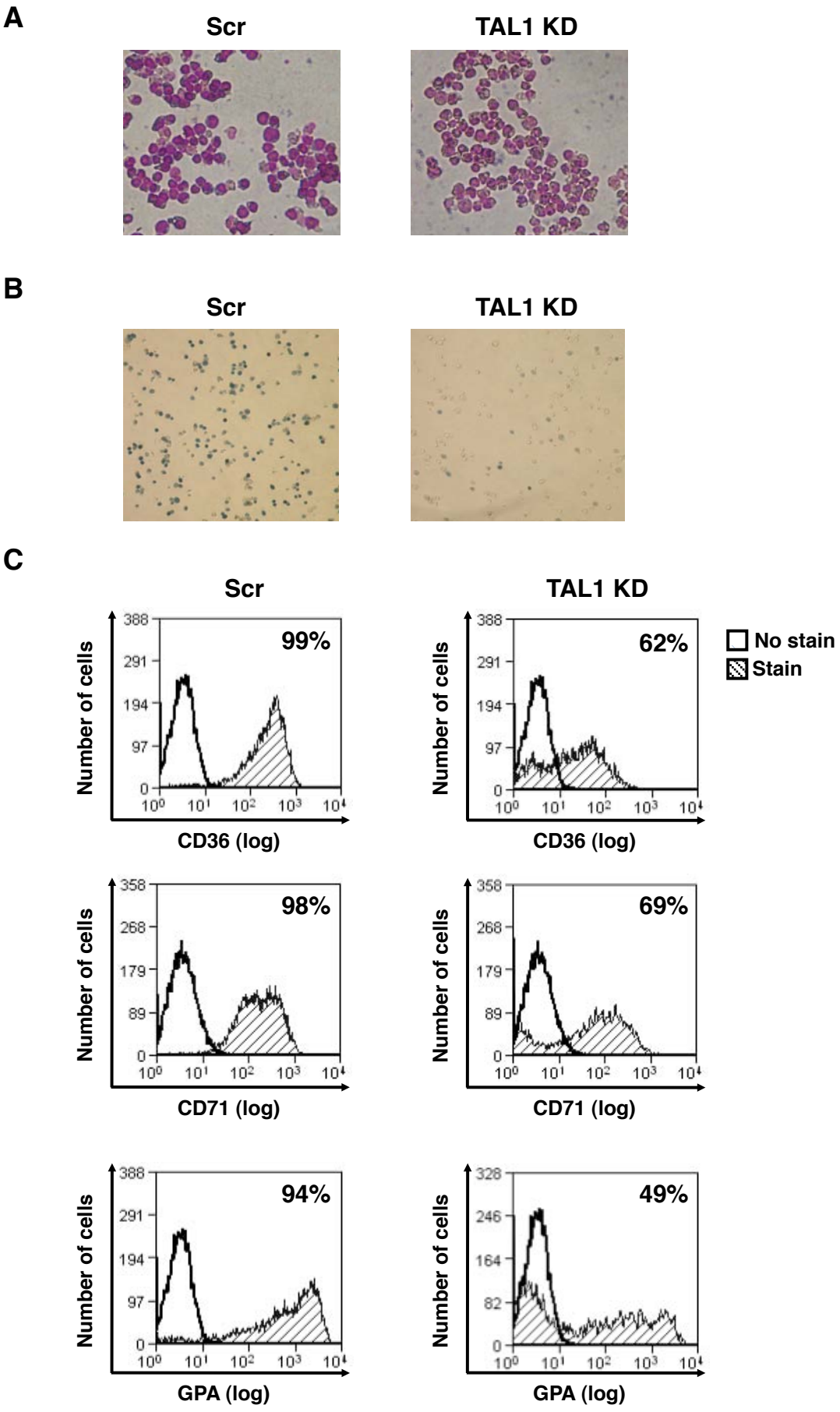

A
Scr
TAL1 KD
B
Scr
TAL1 KD
C
Scr
TAL1 KD
No stain
Stain
Number of cells
99%
62%
CD36 (log)
98%
69%
CD71 (log)
94%
49%
GPA (log)

Figure S2

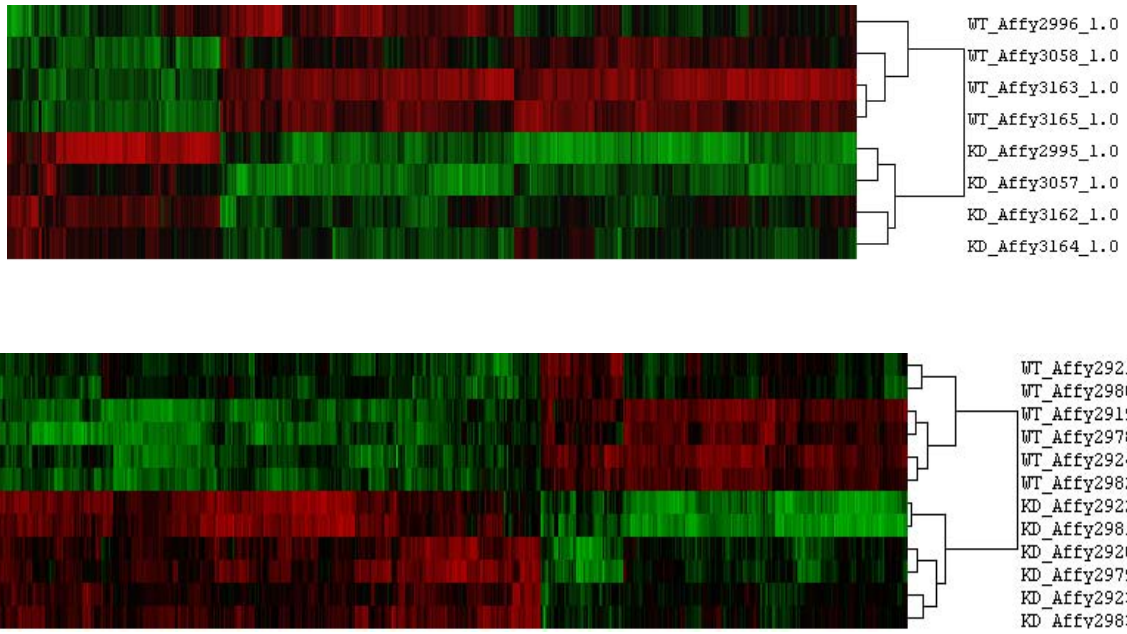
Figure S2: TAL1 knockdown in pro-erythroblasts alters erythroid differentiation

(A) Morphological analysis of human pro-erythroblasts infected with a lentivirus expressing anti-Tal1 (TAL1 KD) or scramble (Scr) shRNA. Cells are stained with May-Grünwald-Giemsa reagent (magnification x 40).

(B) TAL1 KD leads to a decrease in hemoglobin content, as measured by benzidine staining (magnification x 20).

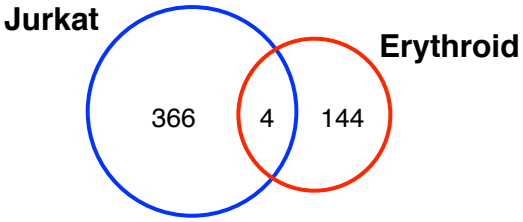
(C) FACS analysis showing that TAL1 KD leads to decrease in the indicated cell surface markers (i.e. CD36, CD71, GPA). Percentages of positive cells are indicated.

A



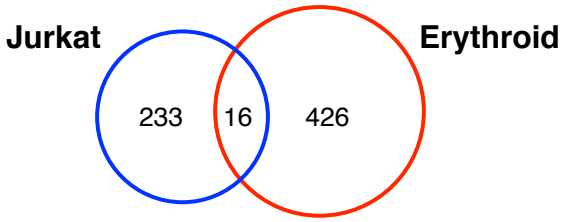
B

Upregulated transcripts (upon TAL1 KD)



C

Downregulated transcripts (upon TAL1 KD)



Jurkat (partial list)

GO Term	Number of Genes	p value
Antigen processing	9	2.74E-12
Apoptosis	36	1.12E-05
mRNA catabolic process	6	5.02E-05
Cell differentiation	59	1.08E-03
Negative regulation of growth	6	2.45E-03
Regulation of cell motility	6	3.53E-03
T cell differentiation during immune response	2	2.56E-02

Erythroid (partial list)

GO Term	Number of Genes	p value
Actin cytoskeleton organization	8	4.13E-04
Peptide metabolic process	3	7.61E-04

Jurkat (partial list)

GO Term	Number of Genes	p value
Amine metabolic process	16	5.95E-05
Organic acid metab. process	18	2.07E-04
DNA replication checkpoint	2	3.52E-03

Erythroid (partial list)

Mitosis	28	6.94E-20
DNA replication	26	1.94E-19
Cell cycle process	44	9.49E-14
Nucleosome assembly	12	4.34E-07
Regulation of cell cycle	24	8.57E-07
Chromosome organization	23	6.92E-06
Traversing start control point of mitotic cell cycle	3	2.34E-04
Heme biosynthetic process	4	2.89E-04

Figure S3

Figure S3: TAL1 regulates different subsets of genes in erythroid and Jurkat cells

Gene expression profiling upon TAL1 KD indicates that in erythroid cells, TAL1 is mostly involved in the activation of genes that are important for cell proliferation and differentiation whereas in Jurkat cells, TAL1 is mostly involved in repressing genes that are important for apoptosis and differentiation. Four (erythroid) and six (Jurkat) biological replicates of wild-type (WT) and TAL1- knocked-down (KD) cells were analyzed by Affymetrix Human Gene 1.0 ST Array, which probes 28,869 genes.

(A) Unsupervised hierarchical clustering of the genes up- and down- regulated upon TAL1 KD in erythroid and Jurkat cells.

(B) Venn diagram displaying the overlap between transcripts upregulated upon TAL1 KD in Jurkat and erythroid cells. The most relevant GO categories of these TAL1-downregulated genes were selected and are presented. For a full list of GO terms, see Supplementary Tables I and II.

(C) Venn diagram displaying the overlap between transcripts downregulated upon TAL1 KD in Jurkat and erythroid cells. The most relevant GO categories of these TAL1-upregulated genes were selected and are presented. For a full list of GO terms, see Supplementary Tables I and II.

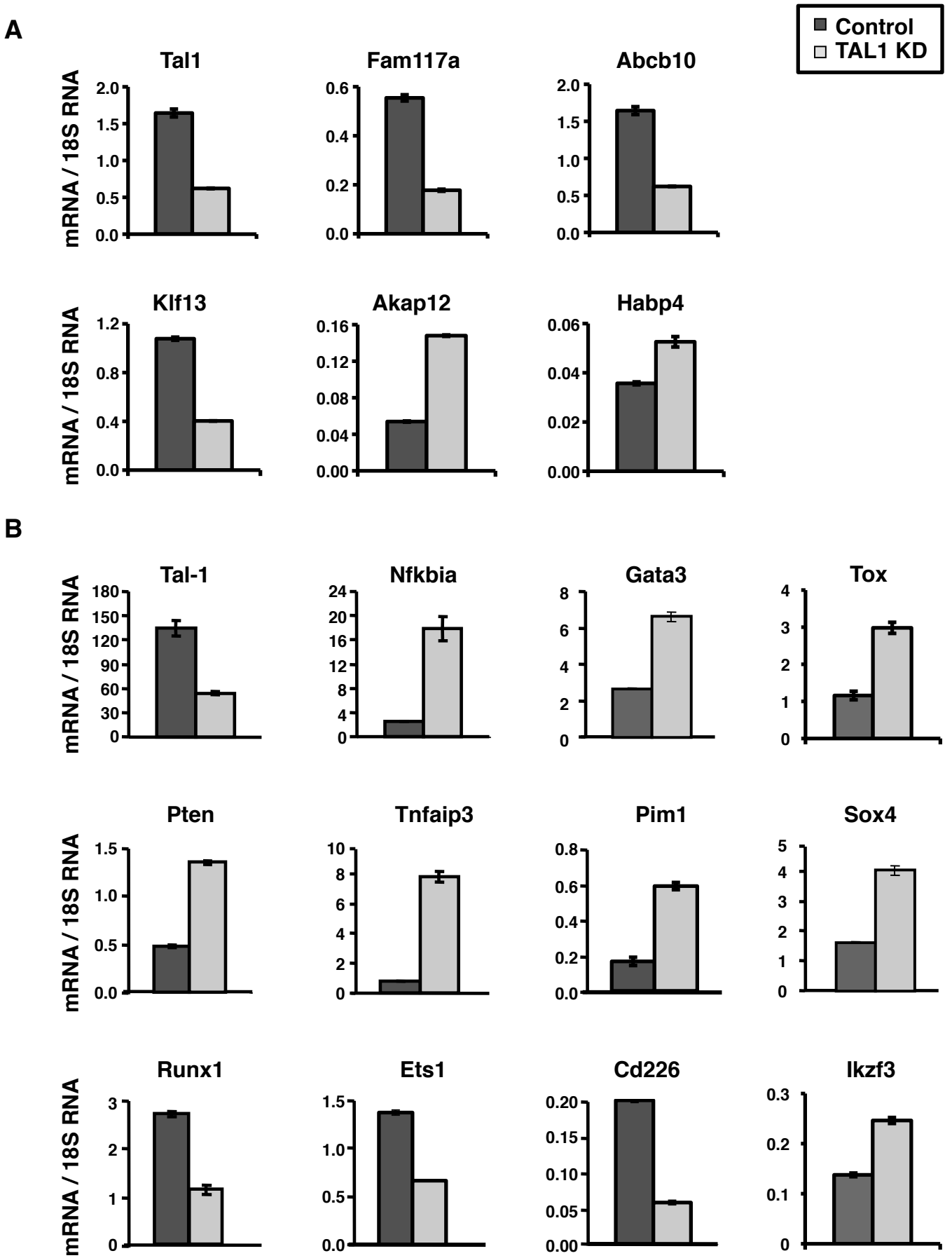

A
Control
TAL1 KD
Tal1
Fam117a
Abcb10
Klf13
Akap12
Habp4
mRNA / 18S RNA
B
Tal-1
Nfkbia
Gata3
Tox
Pten
Tnfaip3
Pim1
Sox4
Runx1
Ets1
Cd226
Ikzf3

Figure S4

Figure S4: RT-qPCR analysis of TAL1-dependent genes in erythroid and Jurkat cells

(A-B) Differentially expressed genes are analyzed by RT-qPCR with error bars corresponding to standard deviations.

(A) TAL1 KD is induced in primary pro-erythroblasts using lentivirus delivered anti-Tal1 shRNA (Tal-1 KD). As a control, cells are infected with a scrambled shRNA (Scr).

(B) TAL1 KD is induced by doxycyclin (Dox) treatment in a Jurkat clone stably expressing a Dox-dependent TAL1 shRNA.

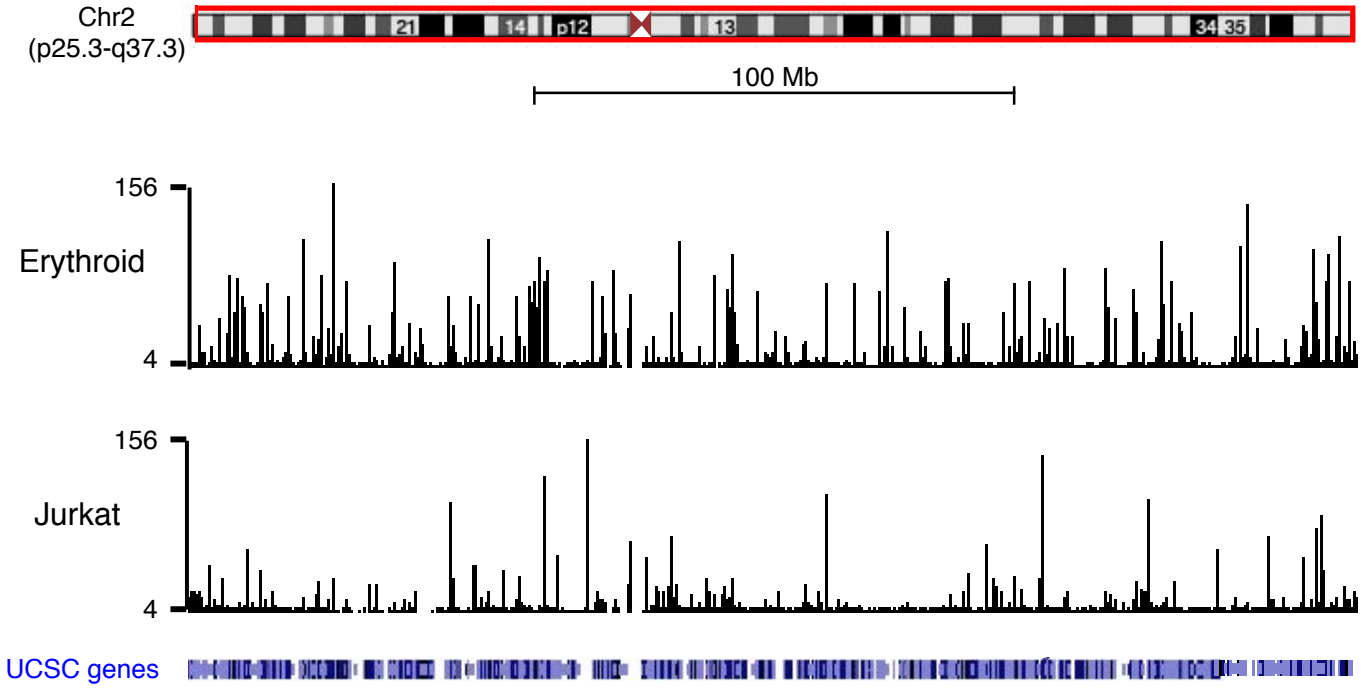
A

Lanes	nRead (M)	nAligned (M)
Jurkat Tal1 ChIP 1	5.21	4.27
Jurkat Tal1 ChIP 2	6.53	3.42
Erythroid Tal1 ChIP 1	6.96	4.01
Erythroid Tal1 ChIP 2	6.61	4.59
Jurkat IgG ChIP	6.54	5.94

B

Mean value of peak height	Median height	Mean value of peak width	Median width
20	28	388	374
14	18	396	372

C



D

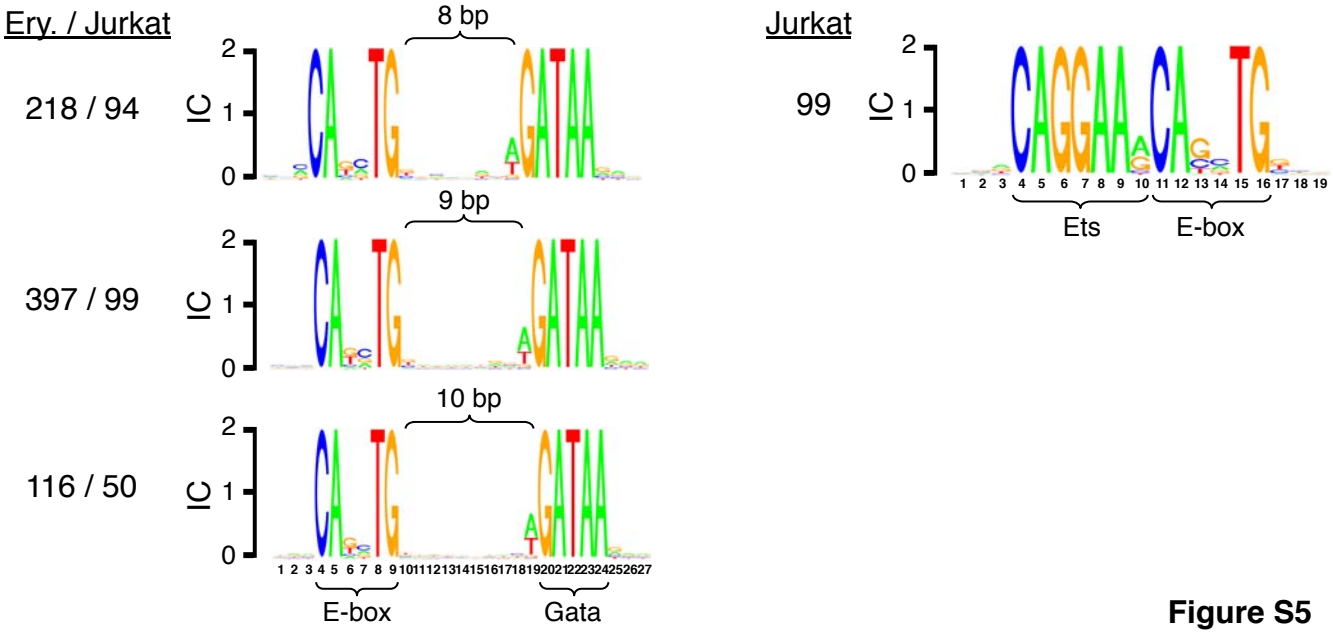


Figure S5

Figure S5: Identification of TAL1 genomic binding sites in erythroid and Jurkat cells by ChIP-seq

Two biologically independent TAL1 ChIPs were performed in erythroid and Jurkat cells, sequenced and aligned to the human genome.

(A) Number of sequence reads (nReads) and reads aligned to the human genome (nAligned) (in millions) for each lane of Solexa sequencing. A threshold of 8 reads was chosen for an approximate false discovery rate (FDR) of 0.005 in each cell type. We also required that the peaks for each cell type have at least a four-fold enrichment relative to the control.

(B) Characteristics of TAL1 peaks in erythroid (red) and Jurkat (blue) cells.

(C) TAL1 binding profile on chromosome 2 in erythroid and Jurkat cells. Numbers of sequence reads are on the y-axis. The UCSC genes are represented below.

(D) Logo representation of E-box/Gata and E-box/Ets composite motifs. The x-axis corresponds to the position of each nucleotide, and the y-axis corresponds to the information content (IC) expressed in bits. The height of each nucleotide indicates conservation at that position. The number of occurrences of each motif in erythroid (Ery.) and Jurkat cells is indicated on the left of the panels.

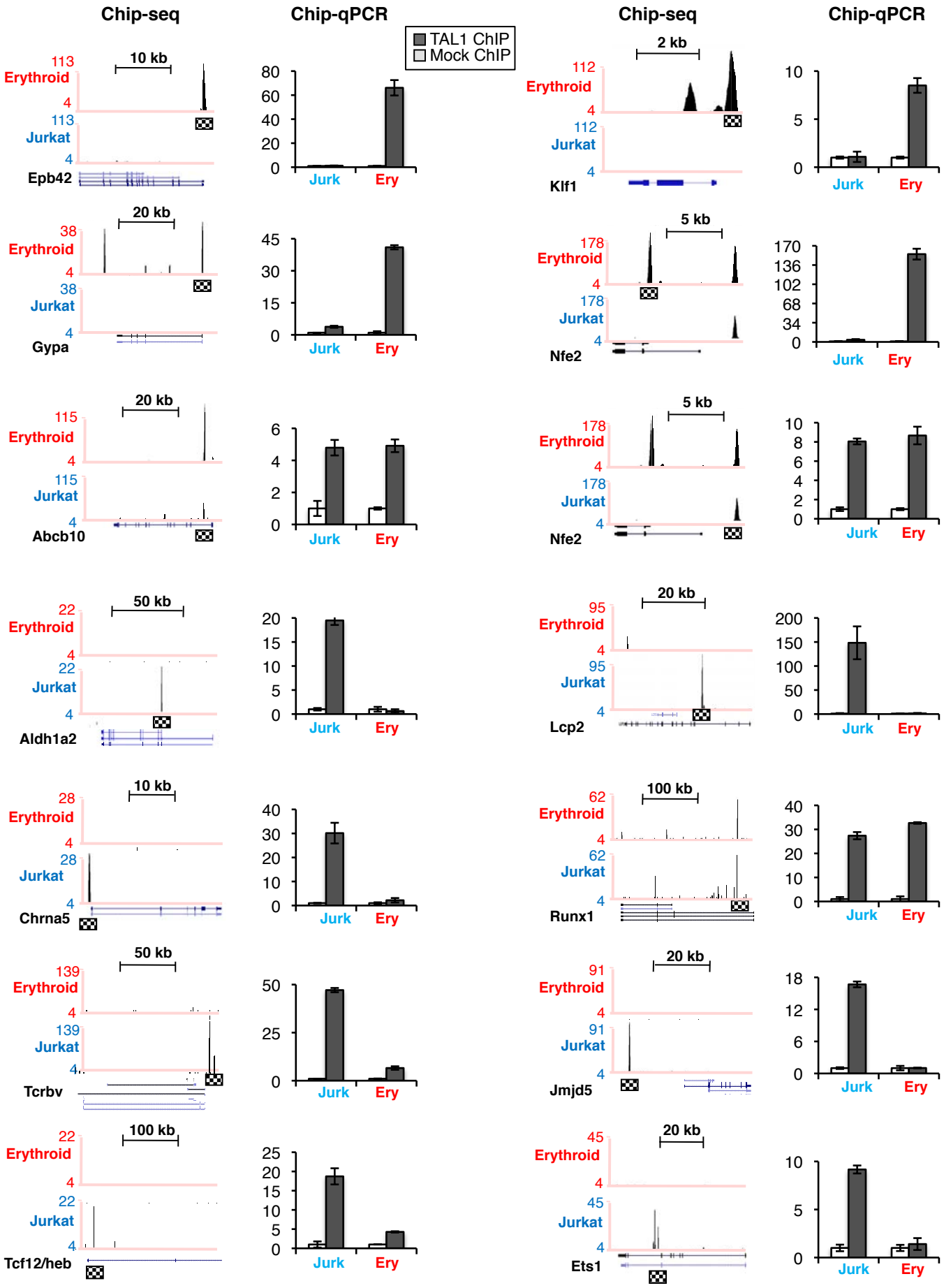

Chip-seq
Chip-qPCR
TAL1 ChIP
Mock ChIP
Chip-seq
Chip-qPCR
10 kb
Erythroid
Jurkat
Epb42
Jurk
Ery
2 kb
Klf1
20 kb
Gypa
5 kb
Nfe2
20 kb
Abcb10
5 kb
Nfe2
50 kb
Aldh1a2
20 kb
Lcp2
10 kb
Chrna5
100 kb
Runx1
50 kb
Tcrbv
20 kb
Jmjd5
100 kb
Tcf12/heb
20 kb
Ets1

Figure S6

Figure S6: Validation of ChIP-seq enrichment profiles by ChIP-qPCR in Jurkat and erythroid cells

Representative ChIP-seq results were verified by independent ChIP-qPCR experiments in erythroid and Jurkat cells using primer pairs indicated by a hatched square. TAL1 binding is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG with error bars corresponding to standard deviations.

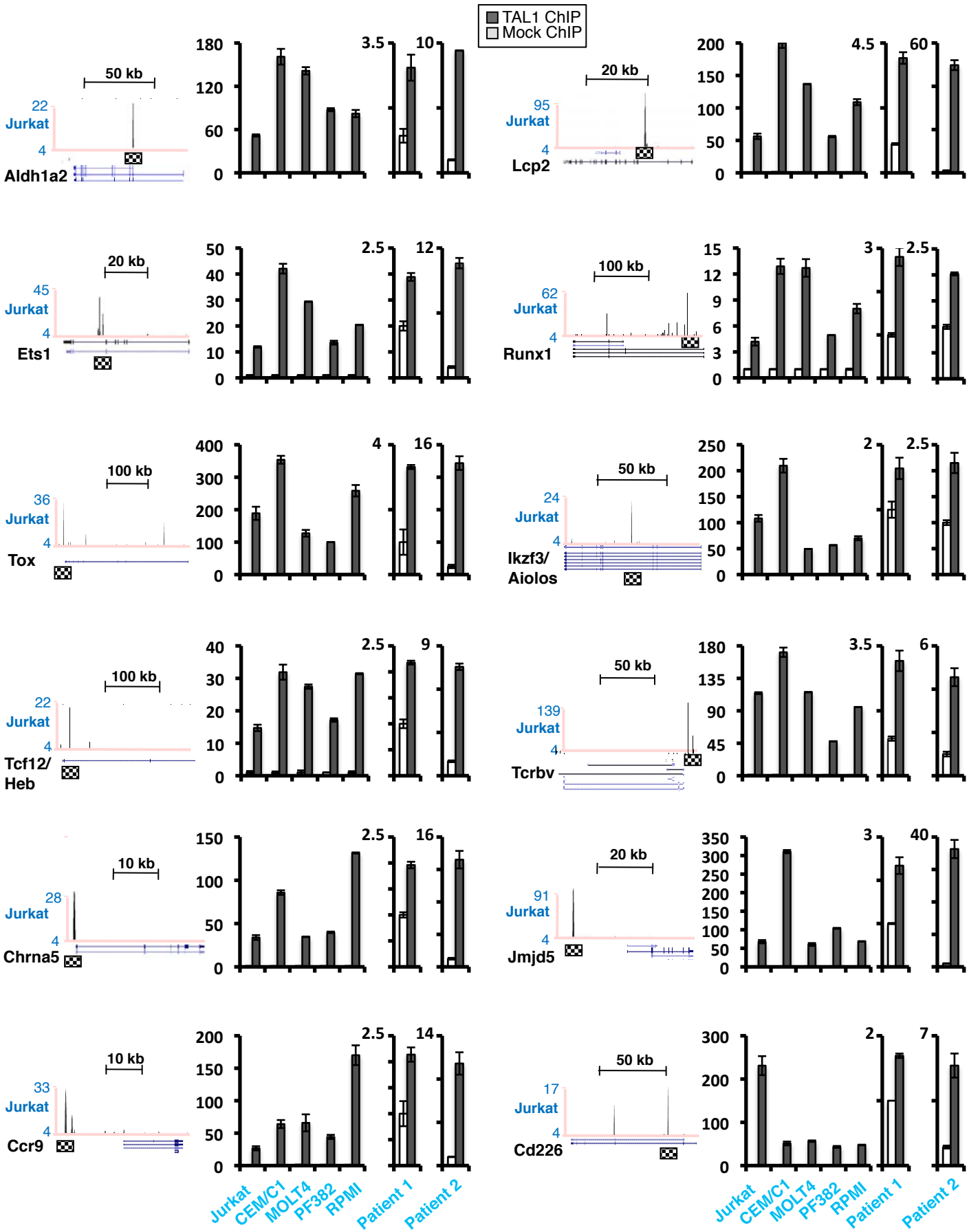

TAL1 ChIP
Mock ChIP
50 kb
22
Jurkat
4
Aldh1a2
180
120
60
0
3.5
10
20 kb
95
Jurkat
4
Lcp2
200
150
100
50
0
4.5
60
20 kb
45
Jurkat
4
Ets1
50
40
30
20
10
0
2.5
12
100 kb
62
Jurkat
4
Runx1
15
12
9
6
3
0
3
2.5
100 kb
36
Jurkat
4
Tox
400
300
200
100
0
4
16
50 kb
24
Jurkat
4
Ikzf3/
Aiolos
250
200
150
100
50
0
2
2.5
100 kb
22
Jurkat
4
Tcf12/
Heb
40
30
20
10
0
2.5
9
50 kb
139
Jurkat
4
Tcrbv
180
135
90
45
0
3.5
6
10 kb
28
Jurkat
4
Chrna5
150
100
50
0
2.5
16
20 kb
91
Jurkat
4
Jmjd5
350
300
250
200
150
100
50
0
3
40
10 kb
33
Jurkat
4
Ccr9
200
150
100
50
0
2.5
14
50 kb
17
Jurkat
4
Cd226
300
200
100
0
2
7
Jurkat
CEM/C1
MOLT4
PF382
RPMI
Patient 1
Patient 2
Jurkat
CEM/C1
MOLT4
PF382
RPMI
Patient 1
Patient 2

Figure S7

Figure S7: Validation of ChIP-seq enrichment profiles by ChIP-qPCR in T-ALL cell lines and primary blasts from T-ALL patients

Representative ChIP-seq results obtained in Jurkat cells were verified by independent ChIP-qPCR experiments in five TAL1-expressing cell lines (i.e. Jurkat, CEM/C1, MOLT4, PF382 and RPMI) and in TAL1-expressing blasts from two T-ALL patients. Primer pairs are indicated by a hatched square. TAL1 binding is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG with error bars corresponding to standard deviations.

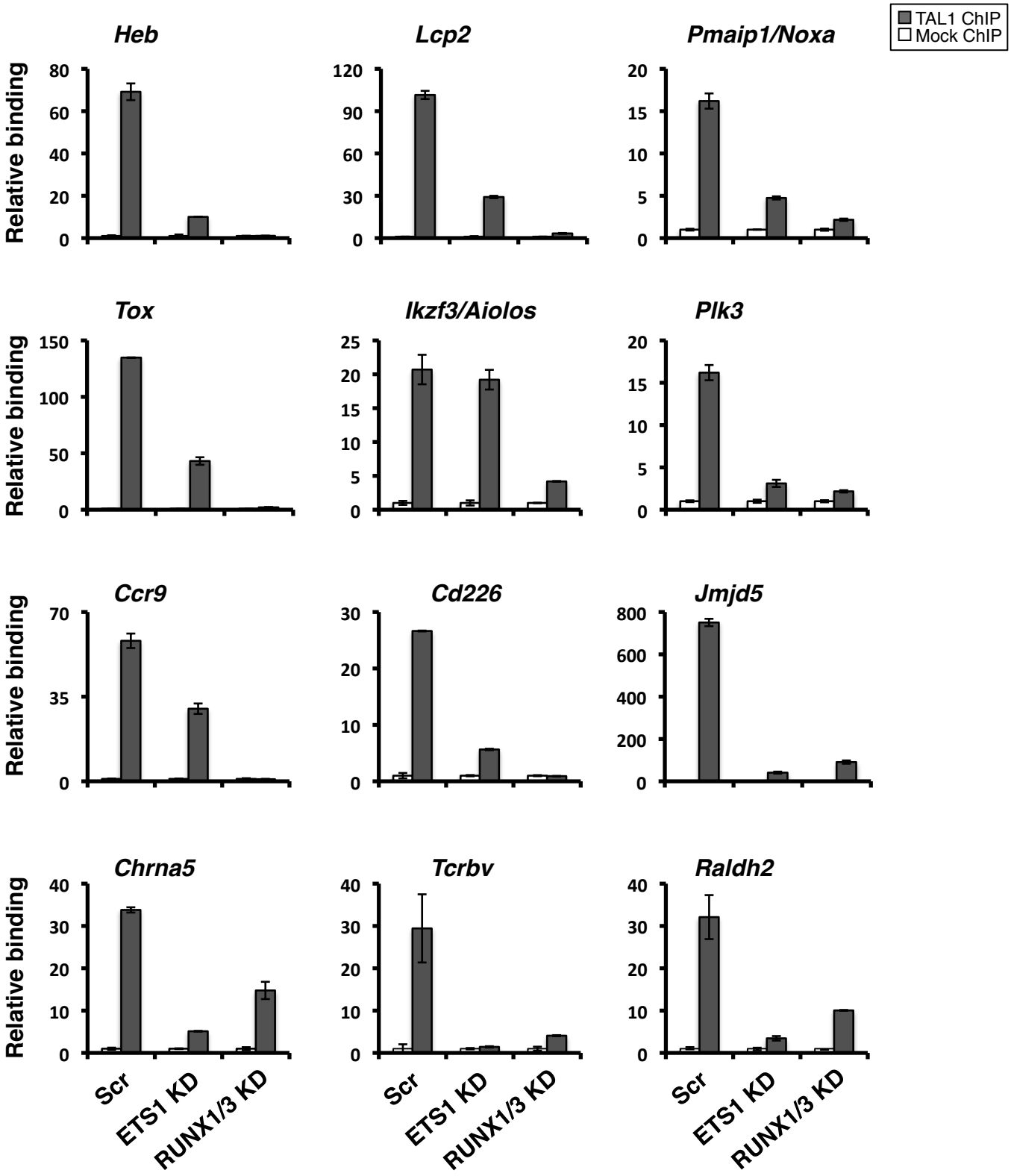

TAL1 ChIP
Mock ChIP
Heb
Lcp2
Pmaip1/Noxa
Tox
Ikzf3/Aiolos
Plk3
Ccr9
Cd226
Jmjd5
Chrna5
Tcrbv
Raldh2
Relative binding
Scr
ETS1 KD
RUNX1/3 KD

Figure S8

Figure S8: TAL1 genomic binding in Jurkat cells requires ETS1 and RUNX1/3 transcription factors

ETS1 and RUNX1/3 KD were performed using lentiviral-mediated delivery of shRNA. ETS1 and RUNX1/3 KD lead to a decrease of TAL1 binding to representative genes in Jurkat cells, as measured by ChIP-qPCR. TAL1 binding is expressed as the fold enrichment relative to a mock ChIP performed with normal IgG.

Supplemental Experimental Procedures

Inducible knockdown of TAL1 in Jurkat cells

Jurkat cell lines stably expressing a Dox-dependent shRNA sequence targeting TAL1 (Brunet de la Grange et al, 2006) were established as described previously (Demers et al, 2007). A second Jurkat cell line expressing a different shRNA sequence (5'- CGGACAAGAAGCTCAGCAA-3') and giving rise to a 50% KD of TAL1 was also established. The knockdown of TAL1 was induced by incubation with Doxycyclin at 5μg/ml.

Knockdown of TAL1, ETS1 and RUNX1/3 by lentiviral infection

Small hairpin (sh)RNA against TAL1 (Brunet de la Grange et al, 2006) or a scrambled sequence were cloned into the BLOCK-iT Lentiviral RNAi Expression Vector (Invitrogen). Mission shRNA lentiviral plasmids anti-Ets1 (TRCN0000005591 and TRCN0000005592), anti-Runx1 (TRCN0000013660 and TRCN0000013659), and anti-Runx3 (TRCN0000235674, TRCN0000235673 and TRCN0000235672) were obtained from SIGMA. Lentiviral particles were produced by cotransfection of 293T cells with the shRNA expressing vector, psPAX2 (Addgene 12260) and pMD2.G (Addgene 12259) plasmids (both from D. Trono laboratory). Viral culture supernatants were concentrated by ultracentrifugation at 100,000g for 90 min. Pro-erythroblasts (Day 8 of differentiation), primary blasts from T-ALL patients obtained from the Banque

Leucémique du Québec (Ottawa Hospital Research Ethics Board #2009009-01H) and T-ALL cell lines (Jurkat, CEM/C1, MOLT4, PF382 and RPMI8402) were infected twice at a 24h interval in the presence of 8 μg/ml polybrene (Sigma). After 72h (36h for primary blasts), cells were harvested for RNA extraction and phenotypic analyses.

May-Grünwald Giemsa staining

Cytospun cells at different stages of erythroid differentiation were fixed in methanol for 2 min. at RT prior to May-Grünwald staining for 5 min. and Giemsa staining for 10 min.

Hemoglobin staining with benzidine

Cells were stained in culture using a solution containing 0.5M acetic acid, 0.2% benzidine dihydrochloride (Sigma) and 0.3% hydrogen peroxide. After 2 min. incubation on ice, blue cells (containing hemoglobin) were counted under microscope.

Flow cytometry

5-Bromodeoxy-uridine (BrdU) incorporation was measured by flow cytometry using a monoclonal anti-BrdU antibody (BD Biosciences) according to manufacturer's instructions. Externalization of phosphatidylserine was measured with Annexin V using the Apoptosis Detection Kit I (BD-Pharmingen). 7-Amino-

actinomycin D (7-AAD) was used to exclude non-viable cells. For cell cycle analysis, cells were fixed in cold 70% ethanol for 1h at 4^0C, washed and resuspended in a staining buffer containing 250 μg/ml RNase A and 50 μg/ml propidium iodide (PI) for 30 minutes at 37°C. The DNA content was determined by flow cytometry. Antibodies to CD36 and GPA (BD Pharmingen, San Diego, CA) as well as CD71 (Immunotech, Marseille, France) were used for phenotyping. Flow cytometry controls were unstained cells, except for the BrdU test where an isotype-matched Alexa Fluor 488-conjugated control antibody (Invitrogen) was used. For all experiments, 10,000 cells per sample were recorded on a BD LSR flow cytometer (Becton Dickinson) and analyzed using CELL Quest and Summit v4.3. For cell cycle analysis 50,000 cells were recorded.

Affymetrix microarray data analysis

Differentially expressed transcripts were determined with two class paired analyses using SAM (Tusher et al, 2001) on intensity values previously normalized and summarized with RMA (Irizarry et al, 2003). Cut-offs of FDR of 0.037 and 0.13 where chosen for Jurkat and erythroid experiments, respectively, to obtain comparable numbers of transcripts changing. Gene Ontology analyses restricted to the “biological process” category were performed using GOstat (Beissbarth & Speed, 2004) with no p-value correction.

Description of genes sets used for GSEA analysis

All gene sets from Lee et al. (Lee et al, 2004) correspond to transcripts with a greater than 3-fold enrichment compared to genes expressed in control thymic stromal cells. Genes sets 1 to 4 also correspond to transcripts with a greater than 3-fold enrichment compared to other sets:

Set 1: CD4ISP+DP > SP4+CB4+AB4 (see Sup Table 3-1 in (Lee et al, 2004)).

Set 2: CD4ISP > DP+SP4+CB4+AB4 (see Sup Table 3-3 in (Lee et al, 2004)).

Set 3: DP > CD4ISP+SP4 +CB4+AB4 (see Sup Table 3-4 in (Lee et al, 2004)).

Set 4: SP4+CB4+AB4 > CD4ISP+DP (see Sup Table 3-2 in (Lee et al, 2004)).

Gene set 5 also corresponds to transcripts showing a greater than 2-fold higher expression in CB4 than in AB4:

Set 5: CB4 > AB4 (see Sup Table 4-2 in (Lee et al, 2004)).

Gene set 6 also corresponds to transcripts showing the following pattern of expression:

Set6: SP4 > CB4 > AB4 (see Sup Table 4-1 in (Lee et al, 2004)).

Nuclear extraction and Immunoprecipitations

Nuclear extraction and immunoprecipitations (IPs) were performed as previously described (Ranish et al, 2007) with the following modifications. Antibodies (Abs) were crosslinked to Dynabeads (Dynal) coupled to protein-A (in case of rabbit Abs) or -G (in case of murine or goat Abs) with DMP following manufacturer's instructions (Dynal). Nuclear extract was supplemented with 100 μg/ml final

ethidium bromide, incubated for 1h with rotation at 4^0C, and added to Abs-bound beads. After incubation overnight at 4^0C with rotation and washing with IP buffer (25 mM Tris, pH 7.9, 5mM $MgCl_2$, 10% (v/v) glycerol, 0.1 % (v/v) NP40, 150mM KCl, 0.3 mM DTT and protease inhibitors), bound proteins were eluted by heating at 95^0C for 5 min. in SDS-PAGE loading buffer.

Antibodies

For western blot we used the following antibodies: anti- GATA3 (sc-268), E2A (sc-416), HEB (sc-357), ETS1 (sc-56674), TAL1 (sc-12984), TFIIHp89 (sc-293), GABPA (sc-28312), Aiolos (sc-101982) from Santa-Cruz Biotechnology; anti-TAL1 (Cat# 39066) from Active Motif; anti-RUNX1 (ab35962) from Abcam; anti-tubulin monoclonal Ab from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa (gift from L. Megeney, OHRI, Ottawa, ON, Canada). For IP, we used: anti-TAL1 (sc-12984), ETS1 (sc-350), GABPA (sc-28312), normal IgG (sc-2028) from Santa-Cruz Biotechnology; anti-RUNX1 (39000) from Active Motif.

High-throughput DNA sequencing data analysis

Sequences were extracted using the Firecrest and Bustard programs from package GApipeline-0.3.0. Reads were aligned using MAQ (version 0.6.6) to the human reference genome (hg18). We only kept one of the duplicated sequences to minimize the artifacts of PCR amplification. Each read was extended in the

sequencing orientation to a total of 200 bases. To identify a reasonable coverage cutoff, we considered a null negative binomial distribution. It can be viewed as a continuous mixture of Poisson distribution where the mixing distribution of the Poisson rate is modeled as a Gamma prior. In the context of chip-seq, the variation of background reads density on the genome is modeled as a Gamma distribution. We estimate the parameters of the negative binomial distribution by fitting the truncated distribution on the number of nucleotides with coverage one to three, to avoid the problem of inferring effective genome size excluding the non-mappable regions, and to eliminate contamination of any foreground signals in the high coverage regions. The estimated distribution fit numerous control datasets very well in our studies. According to the null distribution, the false distribution rate (FDR) is calculated as the ratio of the expected and observed number of nucleotides at the given cutoff. We selected peak coverage cutoff 8 for both cell types, which corresponds to an approximate FDR of 0.005. We also required that the peaks for each cell type have at least a four-fold enrichment relative to the control.

For comparative analysis with TAL1 binding, RUNX1/3 and ETS1 ChIP-seq raw data from Jurkat cells obtained by Hollenhorst et al. (Hollenhorst et al, 2009) were processed using the same pipeline. The identified 5,267 RUNX peaks and 23,189 ETS1 peaks were compared with TAL1 peaks and the overlap between TAL1 and RUNX or ETS binding was determined based on the interval +/- 500 nt from the peak of the TAL1 site.

Estimation of read coverage

To address whether we had sequenced deeply enough, we developed an approach that is based on the rate at which background signal is being converted to foreground. Briefly, if we had not sequenced deeply enough, we would have anticipated that adding more sequencing reads will convert regions that are currently background into foreground. We randomly divided our existing data into 10 disjoint subsets, and sequentially added one subset at a time. At each step, we computed the proportion of the previous background reads (defined as the reads that do not overlap with any region of coverage 6 or more) that get classified as foreground reads (the reads that overlap with a region of coverage 6 or more) when the current subset is added. We expect this rate to decrease as more data are added, ideally to an asymptote of zero when all foreground signals have been identified. In practice, once the rate of conversion has flattened out, the background noise begins to look like foreground signal, and the benefit of collecting more data will be marginal.

***De novo* Motif Discovery**

High throughput chip-seq experiments suggest that transcription factor binding sites (TFBS) are prevalent. Due to the computational cost, most alignment-based motif finding algorithm, such as MEME and AlignAce, are too slow to run on hundreds of thousands of potential binding sites inferred by the chip-seq

experiment. As a result, many chip-seq studies use only top bindings sites for motif inference. To avoid the computation cost of the alignment-based motif finding algorithms, we use the following strategy: enumerate all n-mer patterns, choose good candidates, and refine each candidate iteratively. Different from many traditional motif finding algorithms which search for a motif model that maximizes the log-likelihood ratio of the data given the model vs. a null background, we aim to find discriminate motifs that distinguish the input foreground and background dataset. We use this optimization criterion to target applications in which the background datasets cannot be simply modeled by a null distribution. For example, there are cases in which a background sequence dataset contains a similar set of enriched motifs as the foreground, while the user is only interested in identifying the different motifs.

To solve this problem, we formulate this problem in a regression paradigm. The training dataset includes a positive set that contains putative TFBS, and a negative set that provides appropriate background. The goal is to find a motif whose occurrences best discriminate the positive and negative dataset. Unlike other motif search methods that model a motif by position weight matrix (PWM) or consensus pattern using IUPAC alphabets, we define motif simply as a collection of oligo-nucleotides. This flexible representation is capable to describe complex dependencies among positions of motifs. As more training sequences become available, our approach enables us to capture delicate but statistically

significant signals beyond the simple PWM model. The regression framework also enables us adjust effects of biases such as GC content. This method is capable to predict both enriched motifs and depleted motifs.

The problem is formulated in the following logistic regression model:

$$y \sim x + z,$$

where y is the binary response, 1 if the sequence is in the positive set and 0 otherwise, x the motif count in each sequence, and z for the optional term that biases the positive or negative datasets. For a good motif, the coefficient of corresponding x vector should be highly statistically significant, measured either by p-values or z-values. We use z-values, as the signs indicate whether the motifs are enriched or depleted, and our optimization criterion is to find a motif representation with maximum absolute z-value.

Our method can be divided into two components: first, find an IUPAC motif pattern with maximum z-value; and second, find positional dependency among the positions.

The details are described below:

Find an optimum IUPAC motif pattern.

1. Enumerate all n-mers with a given width, fitting the above regression model, and sort the n-mers by the absolute values of the corresponding z-values. The most significant n-mer is chosen as the seed for the top motif.

2. Extend the motifs. Append *f* “N"s (where *f* is usually 1 to 3) at both sides of the motifs. Enumerate all replacement of one “N" letter to a more specific letter in the IUPAC alphabets, and choose the one with maximum absolute z-value. If the best extended pattern has better z-value, then use it to replace the current “seed", and repeat this process. If no further improvement can be made at current motif length, add “N"s to both ends, so that each side still has *f* “N"s. If no more “N"s can be added, stop and trim the flanking “N"s, otherwise, repeat the current step.

3. At each position of the pattern, replace it with a different IUPAC latter.

Compute the z-values of all such mutations, and choose the best one, if it is better than the current “seed", let it be the new “seed". Repeat this process until no updates can be made. If any update has been made, go to the previous extension step.

Find positional dependencies.

For distance d from 1 to m, where m is pattern length minus 1, enumerate all pairs of positions at distance d with degenerate letters at the motif pattern. At each such pair of positions (p1; p2), enumerate all pairs of di-nucleotides (n1; n2) allowed by the degenerate letters of the motif pattern. Remove all occurrences of (n1; n2) at position (p1; p2), and re-compute the z-value. Sort all pairs of di-nucleotides at the distance d based on z-values, and test iteratively whether eliminating the given di-nucleotide pair resulting improved z-value, if it does,

update the z-value, and put the pair to the filter list, a blacklist of forbidding pairs. The final motif set defined by the IUPAC pattern filtered by the di-nucleotide black list. It can be easily decomposed into a set of IUPAC patterns.

After we finish refining the top motif, we mask all its occurrences in the sequence dataset, and repeat the whole process to find the next motif. The ranks for patterns that overlap with previous motifs tend to drop significantly after masking, which prevents reporting redundant motifs.

Annotate motifs

To annotate motifs found de novo discovery, we searched them against Jasper and/or Transfac database to identify matches to the motifs of known transcription factor.

Composite DNA motifs identification

To study whether Gata, Runx or Ets motifs interact with E-box at a preferred orientation and/or distance, we collected the PWMs of these transcription factors from the Jaspar database (http://jaspar.genereg.net/cgi-bin/jaspar_db.pl), and scanned them on TAL1 peak sequences and the human genome. For each E-box match, we computed the closest match of Gata/Runx/Ets motifs and calculated the distance distribution both within the peak sequences and on the whole genome, the latter of which serves as background for comparison.

References

Beissbarth T, Speed TP (2004) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20:** 1464-1465

Brunet de la Grange P, Armstrong F, Duval V, Rouyez MC, Goardon N, Romeo PH, Pflumio F (2006) Low SCL/TAL1 expression reveals its major role in adult hematopoietic myeloid progenitors and stem cells. *Blood* **108:** 2998-3004

Demers C, Chaturvedi CP, Ranish JA, Juban G, Lai P, Morle F, Aebersold R, Dilworth FJ, Groudine M, Brand M (2007) Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell* **27:** 573-584

Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, Graves BJ (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5:** e1000778

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-264

Lee MS, Hanspers K, Barker CS, Korn AP, McCune JM (2004) Gene expression profiles during human CD4+ T cell differentiation. *Int Immunol* **16:** 1109-1124

Ranish JA, Brand M, Aebersold R (2007) *Quantitative proteomics by mass spectrometry*, Vol. 359, Clifton, N.J.: Totowa, N.J. : Humana Press.

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98:** 5116-5121