



## A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells

Michael R Tallack, Tom Whittington, Wai Shan Yuen, et al.

*Genome Res.* published online May 27, 2010

Access the most recent version at doi:[10.1101/gr.106575.110](https://doi.org/10.1101/gr.106575.110)

---

**P<P** Published online May 27, 2010 in advance of the print journal.

**Accepted Manuscript** Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2010, Cold Spring Harbor Laboratory Press

## KLF1 regulatory networks in primary erythroid cells

OR

### A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells

Michael R. Tallack<sup>1,4</sup>, Tom Whittington<sup>1,4</sup>, Wai Shan Yuen<sup>1</sup>, Elanor N. Wainwright<sup>1</sup>, Janelle R. Keys<sup>1</sup>, Brooke B. Gardiner<sup>1,2</sup>, Ehsan Nourbakhsh<sup>1,2</sup>, Nicole Cloonan<sup>1,2</sup>, Sean M. Grimmond<sup>1,2</sup>, Timothy L. Bailey<sup>1</sup>, and Andrew C. Perkins<sup>1,3</sup>

<sup>1</sup>Institute for Molecular Bioscience; The University of Queensland; Brisbane, Queensland, 4072; Australia

<sup>2</sup>Queensland Centre for Medical Genomics, Institute for Molecular Bioscience; The University of Queensland; Brisbane, Queensland, 4072; Australia

<sup>3</sup>The Princess Alexandra Hospital; Brisbane, Queensland, 4102; Australia

<sup>4</sup>These authors contributed equally to this work.

Correspondence: Andrew C. Perkins  
Institute for Molecular Bioscience  
The University of Queensland  
St Lucia 4072  
QLD  
Australia  
E: [a.perkins@imb.uq.edu.au](mailto:a.perkins@imb.uq.edu.au)  
P: +617 3346 2077  
F: +617 3346 2101

Timothy L. Bailey  
Institute for Molecular Bioscience  
The University of Queensland  
St Lucia 4072  
QLD  
Australia  
E: [t.bailey@imb.uq.edu.au](mailto:t.bailey@imb.uq.edu.au)  
P: +617 3346 2614  
F: +617 3346 2101

Sean M. Grimmond  
Queensland Centre for Medical Genomics  
Institute for Molecular Bioscience  
The University of Queensland  
St Lucia 4072  
QLD  
Australia  
E: [s.grimmond@imb.uq.edu.au](mailto:s.grimmond@imb.uq.edu.au)  
P: +617 3346 2057  
F: +617 3346 2101

**Running Title: KLF1 Erythroid Regulatory Networks**

**Keywords: erythropoiesis, transcription, KLF1, EKLF, GATA1, ChIP-seq**

## Abstract

KLF1 regulates a diverse suite of genes to direct erythroid cell differentiation from bi-potent progenitors. To determine the local *cis*-regulatory contexts and transcription factor networks in which KLF1 operates, we performed KLF1 ChIP-seq in the mouse. We found at least 945 sites in the genome of E14.5 fetal liver erythroid cells which are occupied by endogenous KLF1. Many of these recovered sites reside in erythroid gene promoters such as *beta-globin*, but the majority are distant to any known gene. Our data suggests KLF1 directly regulates most aspects of terminal erythroid differentiation including production of alpha and beta-globin protein chains, heme biosynthesis, co-ordination of proliferation and anti-apoptotic pathways, and construction of the red cell membrane and cytoskeleton by functioning primarily as a transcriptional activator. Additionally, we suggest new mechanisms for KLF1 co-operation with other transcription factors, in particular the erythroid transcription factor GATA1, to maintain homeostasis in the erythroid compartment.

ChIP-seq data reported in this study has been deposited into the NCBI Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE20478. The accompanying Supplemental Material includes Supplemental Methods, Supplemental Data comprising five figures and six tables, and Supplemental References. This can be found with this article online.

## Introduction

Erythropoiesis, the production of red blood cells (erythrocytes), is characterized by an intricate co-ordination between proliferation and differentiation of hematopoietic stem cells into mature enucleated cells (Koury et al. 2002). As differentiation proceeds, there are well described changes in cell surface protein expression (e.g. CD71 and Ter119), reduction in cell size, progressive hemoglobinization and nuclear condensation which finally results in extrusion of the nucleus, RNA and mitochondria (Cantor and Orkin 2002; Fraser et al. 2007; Testa 2004; Zhang et al. 2003). During erythropoiesis there is also an activation of red cell specific biochemical pathways to harness energy in the absence of mitochondria (i.e. anaerobic glycolysis) and to generate anti-oxidant small molecules to protect the red cell against oxidant stressors. The co-ordination of this complex biology and biochemistry remains poorly understood.

A relatively small number of transcription factors are critical for erythropoiesis (Cantor and Orkin 2002); two of the best understood are GATA1 and KLF1 (formerly known as EKLF), which bind to GATA (5'-WGATAR-3') and extended CACC (5'-CCN-CNC-CCN-3') DNA motifs in the promoters, enhancers and locus control regions (LCRs) of many erythroid genes (Perkins 1999; Weiss and Orkin 1995). Our current understanding of GATA1- and KLF1-regulated genes has in the most part arisen from gene knockout studies (Fujiwara et al. 1996; Nuez et al. 1995; Perkins et al. 1995; Pevny et al. 1991), expression profiling experiments (Drissen et al. 2005; Hodge et al. 2006; Pilon et al. 2008; Welch et al. 2004), and gene-by-gene promoter-reporter studies. These approaches are likely to have elucidated only a small subset of the true direct target genes for these factors. This has become particularly evident through recent global studies of *in vivo* GATA1 binding that show it underpins extensive erythroid gene regulatory networks (Cheng et al. 2009; Fujiwara et al. 2009; Yu et al. 2009).

KLF1 is the founding member of a family of 17 transcription factors in mammals, which is defined by the presence of three highly similar C<sub>2</sub>H<sub>2</sub> type zinc fingers at the C-terminus (Miller and Bieker 1993; van Vliet et al. 2006). Expression of *Klf1* is remarkably restricted to erythroid cells and their precursors (Miller and Bieker 1993) and is critical in erythroid versus megakaryocyte lineage choice (Siatecka et al. 2007; Starck et al. 2003; Tallack and Perkins

2009). *Klf1* *-/-* mice die from anemia by E15.5, with severe defects in differentiation, hemoglobinisation, enucleation and membrane-cytoskeleton organization of red blood cells (Nuez et al. 1995; Perkins et al. 1995). Direct transcriptional target genes of KLF1 are few to date, but loss of some are likely to explain these phenotypes, at least in part. These KLF1 target genes include *beta-globin*, *AHSP*, *Dematin/Epb4.9*, *Klf3/Bklf*, *E2f2* and *p18<sup>INK4c</sup>* (Drissen et al. 2005; Funnell et al. 2007; Hodge et al. 2006; Keys et al. 2007; Miller and Bieker 1993; Nuez et al. 1995; Perkins et al. 1995; Pilon et al. 2008; Pilon et al. 2006; Tallack et al. 2009; Tallack et al. 2007).

At present, the best prediction of KLF1 binding site preference has arisen from the crystal structure of the three zinc fingers of a related protein, Zif268/Krox24, bound to its cognate DNA motif (Feng et al. 1994). However, the extent to which this reflects the true *in vivo* binding of KLF1 is not known. Descriptions of transcription factor binding sites determined *in vitro* have been shown in some cases to be information poor compared to those defined *in vivo*, as is the case for the tumor suppressor protein p53 (Wei et al. 2006).

The aim of this study was to gain a better understanding of how KLF1 regulates erythropoiesis through determination of the sites bound by KLF1 *in vivo*. By employing Chromatin Immunoprecipitation (ChIP) in primary fetal liver erythroid cells, coupled with high-throughput DNA sequencing, we have described the transcriptional networks that are underpinned by KLF1 in erythropoiesis. Additionally, by *de novo* motif discovery and motif enrichment analysis we have provided insights into the co-operation of KLF1 with other transcription factors *in vivo*. In particular, we describe that KLF1 often works with GATA1 under specific rules of engagement. Together, these results shed new light on how erythropoiesis is programmed and how KLF1 functions *in vivo*.

## Results

### **KLF1 Occupies CACC Motifs Located Proximally and Distally to Transcription Start Sites**

In order to broaden our understanding of how KLF1 regulates erythropoiesis, we set out to determine all of the binding sites occupied by endogenous KLF1 *in vivo*. We performed ChIP-seq on E14.5 fetal liver, a rich source of definitive erythroid cells at various stages of maturation, to uncover a comprehensive list of direct KLF1 target genes (Supplemental Table S1). DNA recovered from KLF1 ChIP was used to construct fragment libraries and sequenced with the SOLiD next-generation sequencing platform (Applied Biosystems). We generated 17,438,921 and 47,356,494 uniquely mapping ChIP-seq reads to the mm9 genome for KLF1 ChIP-seq and Input DNA libraries respectively and performed various peak calling approaches to estimate between 945 and 1380 sites are bound by KLF1 *in vivo* (Supplemental Table S1).

Many (~16%) of these *in vivo* KLF1 binding sites occur in the vicinity (within 1kb) of transcription start sites (TSSs), but the majority (~52%) are located at distances of >10kb from any known TSS (Figure 1A). A list of the most highly occupied sites (based on greatest peak height) reveals well known direct target genes which are critical for erythroid function such as *Mare* (*alpha-globin* LCR), *Alad*, *Gypc* and *Ermap* (Table 1). However, many of these highly occupied sites are located within or close to genes with no previously described erythroid function or are located at vast genomic distances from any known gene (Table 1, Figure 1A).

In order to evaluate the quality of our peak calls we measured the frequency of CACC motif occurrences in the peak sequences. KLF1 is known to bind DNA via a CACC motif, so the peak sequences should contain a non-randomly high number of CACC motif occurrences, if the peak calls are valid. In order to obtain an unbiased estimate of CACC motif frequency, we used the independently derived CACC motif for KLF7 (Newburger and Bulyk 2009), whose DNA-binding domain is highly similar to that of KLF1. Using an independently derived CACC motif avoids the problem of circularity that could arise if we instead employed a CACC motif

Genome Research  
KLF1 Erythroid Regulatory Networks

derived using *de novo* motif discovery on our peak sequences. We measured average motif affinity, using the program AMA (Buske et al. 2010), rather than simply counting the number of sequences that had at least one CACC motif occurrence at a specified score threshold. By calculating average motif affinity, we more accurately model the thermodynamic interaction of the transcription factor with the DNA. Thus, we account for sequences that have a high ChIP-seq signal due to multiple weak CACC motif occurrences.

By performing a *q*-value analysis (Storey and Tibshirani 2003) applied to the AMA scores of binding regions, we estimated that 75% of sequences are bound directly via the CACC motif. Applying the same analysis to our 330 strongest binding regions, we estimate 95% of such sequences are bound via the CACC motif. The fact that 95% of strongest peaks are estimated to be directly bound via the CACC motif indicates the high quality of the ChIP-seq data. The 75% statistic does not imply 25% of the sequences are noise, as CACC motif presence is only one factor contributing to binding. In particular, cooperative transcription factor binding (Whittington et al. 2010) and chromatin accessibility (Whittington et al. 2009) are critical to *in vivo* occupancy.

By interrogating the same 330 strongest occupied sites using the *de novo* pattern recognition algorithm MEME, we have refined the KLF1 *in vivo* binding site preference (Figure 1B) to 5' R CCM CRC CCN 3' (where R represents A or G and M represents A or C), thus providing additional specificity to the "N" positions previously predicted from *in vitro* studies (Feng et al. 1994). This site is very similar to the KLF7 binding site deposited in UniPROBE (Newburger and Bulyk 2009) and the KLF4 binding site as determined by ChIP-seq in ES cells (Chen et al. 2008) as we would expect based upon the similarity of the amino acid sequences at the DNA binding domains of most KLFs. This refined KLF1 binding motif (Figure 1B) appears in the majority of our 945 KLF1 ChIP-seq peak sequences (572/945 with a *p*-value threshold of 0.0001).

*De novo* motif discovery using MEME also uncovered a GATA-type motif in KLF1 ChIP-seq peak regions (Figure 1C). This motif is almost identical to the TAL1(SCL)-GATA1 motif from the JASPAR CORE database (TOMTOM *E*-value of  $6.6 \times 10^{-11}$ ). This GATA-type motif

(Figure 1C) was frequently observed in KLF1 ChIP-seq peak sequences (279/945 with a p-value threshold of 0.0001). This result suggested an interaction between KLF1 and the SCL-GATA1 complex at erythroid gene promoters and other regulatory elements might be common.

### **KLF1 directly regulates the production of the alpha-globin and beta-globin subunits of hemoglobin**

We expected KLF1 ChIP-seq peaks would be found at the proximal promoters of genes within the *beta-globin* gene cluster (eg. *Hbb-beta<sub>maj</sub>*) (Im et al. 2005; Keys et al. 2007) and within the *beta-globin* LCR (Im et al. 2005; Shyu et al. 2006). Indeed, we observed KLF1 binding at the *Hbb-beta<sub>maj</sub>* promoter and at HS1, HS2, HS4 and HS3 (weakly) of the *beta-globin* LCR (as defined in (Hardison et al. 1997)) (Figure 2B). In each of these cases there is concomitant occupancy of GATA1, as determined by ChIP-seq in G1-ER cells (Cheng et al. 2009), suggesting the two factors co-operate at each of these *cis*-regulatory elements (Figure 2B). We also determined KLF1 occupancy within the *alpha-globin* gene cluster at sites corresponding to HS-31, HS-26, HS-21, HS-8 and an intergenic site between the *Hba-theta<sub>2</sub>* and *Hba-alpha<sub>1</sub>* genes (Figure 2A). KLF1 occupancy at HS-26 has been observed previously (Vernimmen et al. 2007), however this is the first report of occupancy at additional DNaseI hypersensitive sites in the *alpha-globin* gene cluster. Again, we observed concomitant occupancy of GATA1 at all of these sites consistent with co-operation between these two factors in regulation of the *alpha-globin* gene cluster (Figure 2A). In particular, the peak of KLF1 binding at HS-26 (Corresponding to HS-40 in humans) has a high tag count suggesting strong affinity (Table 1). These observations suggest KLF1 is a transcriptional regulator of both the *alpha* and *beta-globin* loci and may co-ordinate the balanced production of globin chains. It also provides further validation of our ChIP-seq dataset and peak calling specificity.

### **Heme Synthesis and Iron Procurement are directly co-ordinated by KLF1**

Our study demonstrated KLF1 occupancy at many (~2/3) of the genes that are required to produce enzymes of the heme synthesis pathway (Ponka 1997), or those required for iron procurement (Figure 3A) (Dunn et al. 2007). Remarkably the majority of the genes in the heme synthesis pathway have KLF1 ChIP-seq peaks at their promoters, within introns (often the first intron) or in neighboring intergenic regions (Figure 3A, Supplemental Figure S2 and

Supplemental Table S2). In order to confirm that occupancy coincided with expression of each of these genes, we performed qRT-PCR of some of these heme synthesis related genes in WT versus *Klf1*<sup>-/-</sup> fetal liver (Figure 3B). All of the genes that we tested with the exception of *UROS* and *FECH*, which do not contain KLF1 ChIP-seq peaks, have decreased expression in *Klf1*<sup>-/-</sup> fetal liver compared to WT. Some of these genes appear to produce erythroid specific transcripts (eg. *ALADe*, *PBGDe/HMBS*, *BZRPe*) which are also specifically reduced at the mRNA level in *Klf1*<sup>-/-</sup> erythroid tissue (Figure 3B).

KLF1 also appears to regulate importation of Coproporphyrinogen III (COPRO) into the mitochondria where ferrous iron can be safely incorporated into heme (Figure 3A). This process is regulated by a recently described multiprotein complex known as the mitochondrial permeability transition pore (MPTP) complex (Papadopoulos et al. 2006) (Figure 3A, inset). KLF1 binds with GATA1 to the promoters of two genes, *BZRAP1* and *VDAC1*, which encode proteins that contribute to the MPTP (Figure 3A and Supplemental Figure S2).

Many of the genes required for iron transport have been discovered only recently, and more are likely to await discovery (Dunn et al. 2007). We have known for many years that surface expression of the transferrin receptor (*TFR1*, *CD71*) on erythroid cells is reduced in the absence of *Klf1* in a gene dosage-dependent manner, suggesting it might be a direct target of KLF1 (Hodge et al. 2006). We therefore asked whether KLF1 directly regulates either of the two transferrin receptor genes (*TFR1* and *TFR2*) and other genes involved in trafficking of iron from the receptor to the heme molecule inside mitochondria (Figure 3A). Indeed we found KLF1 ChIP-seq peaks in the vicinity of *TFR1*, the mitochondrial iron transporter mitoferrin (*MFRN1*), and the *DMT1* and *STEAP3* genes which encode proteins involved in iron transport and its reduction in the endosome (Figure 3A). Additionally, *MFRN1* is markedly down-regulated in the *Klf1*<sup>-/-</sup> fetal liver (Figure 3B).

The final step in the assembly of the hemoglobin molecule requires a fully constructed iron containing heme group to be transported outside of the mitochondria and co-ordinated with the appropriate globin protein chains (Figure 3A). KLF1 additionally appears to control this process by direct regulation of the *ABC-me* and *ABCG2* genes that produce heme exporter

proteins (Figure 3A). Previous studies have also described KLF1 as a critical regulator of the gene encoding the alpha-globin chaperone protein AHSP (Keys et al. 2007; Pilon et al. 2006). Unfortunately the *AHSP* gene is missing from the mm8 and mm9 genome assemblies, preventing its inclusion in the comprehensive list of KLF1 ChIP-seq target genes reported in this study.

We chose to validate that the heme synthesis related genes with strong KLF1 ChIP-seq peaks (red boxes, Figure 3A) were directly induced by KLF1 in a *Klf1* *-/-* erythroid cell line with tamoxifen inducible KLF1 activity (Coghill et al. 2001) (Figure 3C). This system facilitates testing of direct chromatin-dependent induction of gene expression and is more physiological than reporter assays with plasmids (Brown et al. 2002; Coghill et al. 2001). We compared the induction of these genes (*ALAD*, *MFRN1* and *ABC-me*) to that of *beta-globin* (*Hbb-beta<sub>maj</sub>*). We found *beta-globin* to be strongly induced (~50-fold by 14h) by tamoxifen in these cells as expected (Keys et al. 2007). We also found *ALADe* (erythroid specific transcript of *ALAD*) and *MFRN1* to be rapidly induced by tamoxifen, albeit at levels much lower than for *beta-globin* (~6-fold and ~3-fold respectively). Expression of *ABC-me* was only mildly induced by tamoxifen in these cells (~1.5-fold), although these cells do not fully differentiate as erythroid cells do *in vivo*. These observations confirm that KLF1 is a critical direct regulator of many genes acting throughout the production of hemoglobin, from heme synthesis to iron procurement to the assembly of hemoglobin itself.

### **KLF1 regulates erythroid cell survival, proliferation and integrity by direct regulation of specific target genes**

The site of additional KLF1 ChIP-seq peaks provided us with an interesting insight into other processes that were likely to be controlled by KLF1 activity. We found peaks within the genes *Bcl-X*, *Xiap* and *Pim1*, which are known to be anti-apoptotic, pro-survival components of erythroid cells (Supplemental Figure S3). We also confirmed our previous identification of *E2f2* as a direct KLF1 target gene via an enhancer located in the first intron (Tallack et al. 2009), however we were unable to identify KLF1 binding at the gene promoter (Supplemental Figure S4). Additionally we determined a region of strong KLF1 occupancy in intron 5 of the *E2f4* gene (Supplemental Figure S4). We previously could not identify KLF1 binding to this region in K1-ER cells that have been retrovirally immortalized (Tallack et al. 2009). However,

the use of primary tissue for ChIP-seq in this case provides evidence that *E2f4* is indeed a direct KLF1 target gene. Further KLF1 target genes that play roles in the maintenance and integrity of the erythroid cell membrane and cytoskeleton such as the *Ermap/Scianna* (Ye et al. 2000), *Glycophorin C (Gypc)*, and *Ankyrin 1 (Ank1)* (Nilson et al. 2006) genes were uncovered (Figure 4B and Supplemental Figure S4). Also, KLF1 occupies the *Fn3k*, *Glut4*, and *Pigq* genes (Table 1 and Figure 4B) suggesting a direct role in sugar metabolism and protein modification such as glycosylation and GPI biosynthesis. In short, our study shows KLF1 directly regulates virtually all processes necessary for the production and survival of erythroid cells.

### **KLF1 acts primarily as a transcriptional activator *in vivo* and not exclusively via promoter regions**

Our qRT-PCR experiments only permitted testing of a small fraction of the set of KLF1-bound genes for KLF1-dependent changes in expression. Thus, we compared our KLF1 ChIP-seq data with the KLF1 expression profiling data of Hodge *et al.* (Hodge et al. 2006), to elucidate the relationship between KLF1 binding and gene activation or repression, on a global scale. Microarray-probed genes were divided into three groups; “Down-regulated” in the *Klf1* *-/-* fetal liver (inferred to be KLF1 activated, 730 genes), “Up-regulated” in the *Klf1* *-/-* fetal liver (inferred to be KLF1 repressed, 369 genes), or “Non-responsive” to loss of *Klf1* (12349 genes). We recorded the presence of KLF1 ChIP-seq peaks in the promoter (within 2kb of TSS), first intron, intragenic region, and proximal downstream region (within 2kb of transcript end) of each gene (shown by the cartoon in Figure 4A). Figure 4A displays a barplot indicating the fraction of regions that are estimated as bound, for each (responsiveness, region) combination. A much larger fraction of “Down-regulated” genes are bound by KLF1, compared with the “Up-regulated” or “Non-responsive” genes. This supports the view that KLF1 primarily activates gene expression. In contrast, there is no significant difference between the fraction of “Up-regulated” genes bound and the fraction of “Non-responsive” genes bound, for any gene interval. Thus, we conclude that KLF1 rarely functions as a direct transcriptional repressor *in vivo*. The genes that are bound by KLF1 and “Down-regulated” (64 genes) in Figure 4A (maroon bars) provide a very high confidence subset of direct KLF1 target genes for future studies. These are provided in Supplemental Table S3.

We were not surprised to observe that a high proportion of the “Down-regulated” genes (presumably activated by KLF1) were not bound within their promoter region but were in fact bound at sites within the first intron or at other intragenic regions (Figure 4A). For example, *ALAS2*, *ALAD*, *HMBS*, *UROD*, *ABCG2*, *DMT1*, *Bcl-X* and *Ankyrin 1* all have KLF1 binding sites in the first intron (Supplemental Figures S2, S3 and S4 and Supplemental Table S4). This result supports the notion that KLF1 might be activating transcription of target genes at alternative erythroid specific promoters (either known or supported by mRNA, EST or CAGE data, Supplemental Table S4) in many cases, or via intronic enhancer elements, as for *E2f2* and *E2f4* (Tallack et al. 2009).

We also investigated the possibility that the strength or frequency of KLF1 binding (peak height) to the regulatory element of a gene might relate to the degree to which that gene is activated by KLF1. We hypothesized that the genes which are most “Down-regulated” in the absence of KLF1 might contain the strongest peaks. However, we found no relationship between strength of binding and degree of responsiveness to KLF1 (Figure 4B). Indeed some of the most “Down-regulated” genes have only modest peak heights (eg. *Dematin/Epb4.9*, Figure 4B). Likewise, some of the strongest peaks occur in genes that are only modestly “Down-regulated” in the Hodge *et al.* dataset (eg. *Gypc*, Figure 4B). However, in many cases it is difficult to assign a peak to a TSS, since for distal peaks the nearest TSS is often not the true target gene. Nevertheless, our study clearly shows that KLF1 binding leads to activation of target gene expression, but the degree to which gene activation occurs depends on many additional factors besides the binding of this single transcription factor.

### **KLF1 and GATA1 co-occupy many sites in the erythroid genome in a distinct non-random configuration**

Previous work has suggested co-operation between KLF1 and GATA1 in erythroid gene regulation (Gregory et al. 1996; Merika and Orkin 1995). However, KLF1 has not been found with GATA1 containing protein complexes *in vivo* (Rodriguez et al. 2005) and we have never been able to co-immunoprecipitate KLF1 and GATA1 in the absence of DNA. Nevertheless, to investigate a functional DNA-dependent interaction between KLF1 and GATA1 at a global level we took advantage of a recent ChIP-seq study performed by Cheng *et al.* which defined the full complement of GATA1 binding sites in G1-ER cells (Cheng et al. 2009). We

Genome Research  
KLF1 Erythroid Regulatory Networks

determined the distance to the nearest GATA1 ChIP-seq peak for all of our KLF1 ChIP-seq peaks by performing a liftOver of GATA1 ChIP-seq peaks onto the mm9 genome (see Methods). A histogram of these distances shows that approximately 48% of our KLF1 ChIP-seq peaks have a corresponding GATA1 ChIP-seq peak within 1kb ( $\text{Log}_{10}3$ ) (Shared, Figure 5A), suggesting *in vivo* co-operation between the two factors is likely. GATA1 is known to function together with the transcription factor SCL/TAL1 (hereafter SCL) to form an activating complex (Tripic et al. 2009). Our *de novo* motif discovery (Figure 1C) suggested that a complex containing KLF1, GATA1 and SCL might be present in erythroid cells. However, it was interesting to find that by incorporating *in vivo* erythroid ChIP-seq data for SCL (Wilson et al. 2009) into our analysis we find that KLF1-GATA1 and SCL-GATA1 co-bound regions show very little overlap *in vivo* (Venn diagram, Figure 6). The SCL ChIP-seq dataset may not be comprehensive; nevertheless most SCL bound sites are not bound by KLF1.

Although KLF1 and GATA1 can act synergistically at artificial promoters in reporter assays (Merika and Orkin 1995), co-occupancy of KLF1 and GATA1 *in vivo* does not necessarily indicate co-operation. So we next sought to determine if there were particular “rules of engagement” that might provide evidence for such a functional interaction. We used the algorithm SpaMo (Whittington et al. 2010) to determine whether CACC and GATA motifs exhibited any preferred spatial configuration in the KLF1 ChIP-seq peak sequences. SpaMo produces two histograms indicating the frequency of each possible displacement between the two input motifs. These histograms (one for same strand motif hits, one for opposite strand motif hits) clearly demonstrate an enrichment for CACC-GATA motif displacement at 15-45bp apart on opposite strands (Figure 5B). SpaMo did not detect enriched spacing for any non-GATA motifs in our set of 259 input motifs. This supports the veracity of the observed CACC-GATA enrichment in KLF1 ChIP-seq peak sequences.

The most highly represented displacement between CACC and GATA motifs is at a distance of 18bp apart on opposite strands. We used this information to construct a 3D model of cooperative binding, by placing the crystal structures for the zinc finger domains of GATA1 (pink) and Zif268 (a surrogate for KLF1, blue) onto a short segment of DNA (Figure 5C). We have no knowledge of the structure of the non-DNA-binding domains of either KLF1 or

GATA1, but we speculate that the non-DNA-binding domains of KLF1 and GATA1 as well as co-factors are likely to exist in the centre of the space between the CACC and GATA sites (Figure 5C). Based on our *de novo* motif discovery analysis with MEME, which recovered an SCL-GATA1 site in KLF1 ChIP-seq peak regions (Figure 1C) and the preferential orientations of CACC and GATA motifs, we suggest that SCL is likely to be one factor at the centre of the KLF1-GATA1 erythroid complex in a small subset of erythroid *cis*-regulatory modules (Figure 5C).

### **Transcription factor partners for KLF1**

Our discovery of an interaction between KLF1 and GATA1 at a large number of genomic sites prompted us to ask what other factors might also participate at these sites, and more importantly, what interactions might be at play for KLF1 exclusive sites (ie. those with no GATA1 binding). We utilized the motif enrichment analysis tool Clover to interrogate all KLF1 bound genomic locations as well as KLF1-GATA1 co-bound genomic locations for over-representation of specific known DNA motifs (see Methods). We found an over-representation of GATA and KLF motifs in both the total KLF1 bound regions and co-bound regions as expected (Figure 6). More interestingly we found enrichment of an AP-1 type motif (represented by TCF11-MafG bZIP, (Johnsen et al. 1998)) and Smad (represented by Smad3) motifs in both types of regions (Figure 6). Over-representation of a number of other motifs was also observed, including a FORKHEAD motif (FOXC1 FORKHEAD) and a homeobox half site motif, TAAT (Ubx HOMEODOMAIN, Figure 6). Taken together, these results suggest that KLF1 may interact in transcriptional complexes that also include proteins capable of binding to AP-1, Smad, and homeobox motifs (see Discussion).

## Discussion

We have described in this study a broad repertoire of direct KLF1 target genes that regulate virtually all processes associated with erythropoiesis. Depending on the specificity of our peak calling we declare between 945 and 1380 *in vivo* KLF1 binding sites in erythroid cells. We have deliberately favored high specificity over sensitivity in our peak calling approach to reduce the risk of declaring false positive KLF1 binding sites. Thus, much greater sequencing depth is likely to find a small number of additional lower affinity sites, but will mostly re-sequence the same high affinity sites. These declared binding sites describe the majority of the genome wide transcriptional regulatory activities of KLF1 during erythropoiesis.

In the vicinity of our declared KLF1 ChIP-seq peaks we identified by *de novo* motif discovery a SCL-GATA1 binding motif suggesting that these two factors might interact with KLF1 to drive the appropriate expression of erythroid genes (Figure 1C). In addition, by using the newly developed tool SpaMo we identified a set of specific constraints governing the positioning between CACC and GATA motifs in KLF1 ChIP-seq peaks (Figure 5B and 5C). SCL is likely to bind between KLF1 and GATA1 in some instances, yet when we consider the overlap between KLF1-GATA1 shared peaks and SCL-GATA1 co-bound regions it is only minimal (Figure 6). We feel that the high specificity and relatively low sensitivity of the SCL ChIP-seq dataset might partly explain for this apparent contradiction, whereby a less stringent peak calling approach in the SCL ChIP-seq dataset would likely lead to a greater number of peaks and more significant overlap with KLF1 ChIP-seq peaks. Nevertheless, many SCL-GATA1 bound regions in the genome function independently of KLF1 and vice versa.

Additionally, our study suggests there is a significant advantage in using primary tissue rather than engineered cell lines for identifying all sites of binding *in vivo*. For example, we previously conducted ChIP experiments in an immortalized *Klf1* *-/-* erythroid cell line with inducible KLF1 activity (K1-ER cells) and were unable to find any binding of KLF1 at the *E2f4* gene locus (Tallack et al. 2009). In direct contrast we describe here that the *E2f4* gene locus is robustly bound by KLF1 in primary fetal liver cells at a site located within intron 5 (Supplemental Figure S4). This is likely due to KLF1 occupying different regions of the genome as an erythroid precursor differentiates into a mature late normoblast. Both E2F2 and

E2F4 have roles in erythroid differentiation, perhaps balancing proliferation with differentiation (Dirlam et al. 2007; Kinross et al. 2006). Such dynamic changes in occupancy are sometimes difficult to observe in most cell lines. This is an example of a binding event we missed in K1-ER, and would probably miss in MEL cell lines, that we feel highlights the advantage of performing ChIP-seq experiments in primary tissue when it is possible.

We also describe a set of *cis*-elements we have associated with KLF1 target genes (Table 1) that significantly overlaps with that previously postulated from expression profiling experiments (Figure 4) (Drissen et al. 2005; Hodge et al. 2006; Pilon et al. 2008). This includes not only genes we might have expected to find such as those responsible for the production of hemoglobin (Figures 2 and 3) and the integrity and deformability of the red cell (Supplemental Figure S4), but also genes required for cell survival and nuclear condensation like *Bcl-X*, *Xiap* and *Pim1* (Supplemental Figure S3). Thus it appears that KLF1 (together with GATA1 50% of the time) regulates virtually all of the processes at work to produce and maintain erythroid cells.

In addition to the discovery of a new suite of KLF1 target genes, we have also produced evidence to suggest unique modes of gene regulation by KLF1. We noticed that the vast majority of KLF1 binding sites (peaks) are located at large genomic distances from known transcription start sites (Figure 1A). Many of these sites have high tag counts (strongly bound) and many are co-occupied by GATA1. KLF1 has recently been described as a critical looping factor particularly in the context of the *beta-globin* locus (Drissen et al. 2004). It has also been shown very recently that KLF1 is critical in the formation of nuclear transcription factories consisting of RNA pol II and highly transcribed erythroid genes (Schoenfelder et al. 2010). These studies strongly suggest that KLF1 is not only capable of, but necessary for, long distance gene regulation within erythroid cells. We suggest many of the *in vivo* binding sites for KLF1 that are found at vast distances from any known gene might function in chromatin looping, or be involved in recruiting genomic regions into transcription factories, and may not function as canonical enhancers. However, we cannot be sure that these binding sites aren't functional as very distant enhancers. In such cases it is very difficult to know which gene(s) might be responsive to such a putative enhancer, making functional testing extremely difficult.

We also determined that for direct KLF1 target genes, regulation of transcription was frequently occurring via binding at intragenic regions, the first intron in particular (Figure 4A, Supplemental Figures S2, S3 and S4). Some of these instances are previously reported alternative erythroid specific promoters, others have supporting EST evidence for alternative promoters (Supplemental Table S4), but many have not yet been discovered. Erythroid cells have a unique metabolic demand and appear to have co-evolved a unique set of promoters together with KLF1 so that the coding outputs of many genes in general biochemical pathways can be specifically tailored to the particular metabolic needs of erythroid cells. Further studies describing the comprehensive transcriptional landscape of erythroid cells will need to be performed to validate this hypothesis.

In addition to the KLF1-GATA1 interaction, we have uncovered enrichment of several transcription factor DNA binding motifs in the vicinity of KLF1 occupied sites that suggest the involvement of transcription factors capable of binding to AP-1 like (TCF11-MafG bZIP), Smad (Smad3) and Homeobox (Ubx HOMEO) motifs (Figure 6). The AP-1 like motif (TCF11-MafG bZIP, (Johnsen et al. 1998)) is most likely bound by bZIP heterodimers like p45-NFE2/p18-Maf which are important in erythropoiesis (Andrews et al. 1993; Johnsen et al. 1998; Kotkow and Orkin 1995). Interestingly a member of the Smad family of transcription factors, Smad5, is partly responsible for the induction of *Klf1* expression in erythroid cells (Lohmann and Bieker 2008). *Smad5* continues to be expressed in erythroid cells and is critical for the proper expansion of erythroid progenitors in the fetal liver and in stress erythropoiesis (Porayette and Paulson 2008). In addition, our data suggest KLF1 co-operates directly with “ventral” Smads such as with Smad1/5/8 to mediate erythroid expansion in response to BMP4 signaling. This hypothesis is consistent with experiments performed in *Xenopus* which show the KLF1-like protein, *neptune*, is essential for BMP4 induction of red blood cell formation (Huber et al. 2001). In short, we suggest Smad5 activates KLF1 gene expression and then works with KLF1 in the nucleus to regulate the erythroid response to stress.

Genome Research  
KLF1 Erythroid Regulatory Networks

In summary, our study has described the *in vivo* DNA-binding activities of KLF1 by performing primary erythroid tissue ChIP-seq. We have revealed a complex interaction between KLF1 and GATA1 that is responsible for regulating the expression of erythroid genes, and we provide future challenges and considerations for the interpretation of ChIP-seq studies particularly with regard to erythropoiesis.

## Methods

### Primary Fetal Liver ChIP

ChIP assays were performed as previously described (Hodge et al. 2006) using a rabbit polyclonal antibody raised against the N-terminus of KLF1 (Keys et al. 2007; Perkins et al. 1995) in *Klf3*<sup>-/-</sup> or wild type (WT) fetal livers collected from E14.5 mouse embryos. At this stage of development the fetal liver is entirely erythroid and composed of cells at various stages of maturation from pro-erythroblasts to late enucleating normoblasts. *Klf3*<sup>-/-</sup> fetal livers were utilized in addition to WT due to anticipated stronger KLF1 ChIP occupancy based on a recently proposed model of competition between KLF1 and KLF3 (BKLF) for CACC box occupancy *in vivo* (Eaton et al. 2008). In brief, ChIP assays were performed on pools of four homogenized E14.5 fetal livers ( $\sim 8 \times 10^7$  cells) and validated by quantitative real-time PCR as previously described to determine occupancy at an *E2f2* enhancer (*E2f2-i1en*) and the *AHSP* promoter (Keys et al. 2007; Tallack et al. 2009).

### ChIP-seq Library Construction

The SOLiD System 2.0 workflow for Lower Input/Lower Complexity DNA fragment library preparation (Applied Biosystems) was followed to prepare libraries containing 80-130bp of ChIP DNA (or Input DNA) flanked by the appropriate adaptors. A detailed description of the method is provided in the Supplemental Methods.

### SOLiD Chemistry Sequencing

SOLiD ChIP-seq DNA fragment libraries were sequenced by SOLiD system 2.0 and SOLiD system 3.0 chemistries to produce DNA sequence reads of 35 and 50 nucleotides respectively. Sequencing reads were mapped to the mm9 mouse genome using the mapreads algorithm (Applied Biosystems). In order to maximize the number of mappable reads we removed the final 5 nucleotides for unmapped reads in a recursive strategy until the reads reached 25 nucleotides in length, similar to an approach previously described (Cloonan et al. 2009). We recovered 17,438,921 and 47,356,494 unique reads (defined by a unique sequencing read start site) that mapped exactly once to the mm9 genome for the KLF1 ChIP-seq and Input DNA libraries respectively.

### **KLF1 ChIP-seq Peak Detection**

We employed the same method used by Chen *et al.* (Chen et al. 2008) in order to declare a set of KLF1-bound regions for downstream analysis. When generating the tag intensity profile used in peak declaration, we added a count of 1 to each position up to 100bp 3' from the start of each tag. To be declared a peak, a genomic position must pass both a tag intensity threshold and a fold-change threshold. We applied a false discovery rate (FDR) threshold of 1%, which resulted in a tag intensity profile threshold of 9. We applied negative control tag library fold-change threshold of 5.0 at the centre of each declared peak, yielding the final set of 945 declared peaks. We additionally declared a less stringent set of 1380 peaks by applying a fold-change threshold of 3.0 (Supplemental Table S1).

### **Motif Analyses**

We used several motif analysis tools to study the relationship between KLF1 binding and transcription factor DNA-binding sequence specificity.

AMA (average motif affinity) (Buske et al. 2010) was used to estimate the fraction of regions around peaks that have a non-randomly strong KLF1 affinity as predicted by a CACC box PWM.

The *de novo* motif discovery tool MEME (Bailey 2002) was used to determine KLF1 DNA-binding sequence specificity without prior information.

We used the tool Clover (Frith et al. 2004) to measure enrichment of *known* motifs in sets of KLF1 ChIP-seq peak regions of interest, in order to identify motifs of possible co-binding transcription factors. This was necessary in order to detect enrichments that are invisible to *de novo* discovery, due to the inherently large search space of *de novo* discovery.

We used SpaMo (Whittington et al. 2010) to detect significant spacing relationships between known motifs.

For the methodology and parameters employed when running these programs, refer to the Supplemental Methods.

### **Inference of KLF1-DNA-GATA1 Ternary Complex Structure**

The enriched motif spacing information inferred using SpaMo allowed us to predict the likely structure of the corresponding KLF1-DNA-GATA1 ternary complex, using existing transcription factor-DNA X-ray crystallography structures. See Supplemental Methods for an explanation of this approach.

### **Comparing KLF1 Binding and Transcriptional Response to Perturbation of KLF1**

To investigate the relationship between *in vivo* KLF1 binding and transcriptional response to perturbation of *Klf1*, we compared our ChIP-seq data with *Klf1* *-/-* expression profiling data from primary fetal liver (Hodge et al. 2006). Refer to Supplemental Methods for an explanation of this analysis.

### **K1-ER Cell Line Induction**

K1-ER cells (Coghill et al. 2001) were maintained and treated with 4-hydroxytamoxifen (Tam, Sigma) or vehicle control (EtOH) as previously described (Keys et al. 2007). The assessment of induced specific changes in gene expression was performed as described below.

### **Gene Expression Profiling by qRT-PCR**

Fetal liver or K1-ER cell line RNA was isolated and used to produce cDNA as described in the Supplemental Methods. Gene expression profiling was performed by qRT-PCR using specific primer sets as described in the Supplemental Methods and Supplemental Table S6.

## Acknowledgements

This work was supported by an Australian Research Council Discovery Grant (DP0770471 / ACP) and a grant from the Cancer Council Queensland (519718 / ACP). MRT and TW were the recipients of an Australian Postgraduate Award. NC is the recipient of an Australian Research Council Postdoctoral Fellowship. SMG is the recipient of an Australian NH&MRC Senior Research Fellowship.

## Figure Legends

### Figure 1. KLF1 ChIP-seq reveals a role in both proximal and distal gene regulation, and a refined *in vivo* binding site motif

(A) Characteristics of KLF1 binding with respect to transcriptional start sites (TSSs). Distances between all KLF1 ChIP-seq peak sites and the nearest mm9 UCSC Genome Browser annotated TSS were determined and sorted into upstream and downstream sets as shown. Distances are plotted as a histogram on a  $\text{Log}_{10}$  scale (eg. 3 represents 1000bp).

(B) KLF1 *in vivo* consensus binding site predicted by the motif discovery algorithm MEME 4.3 and presented as a sequence LOGO.

(C) Analysis of KLF1 ChIP-seq peaks with MEME 4.3 also uncovered the motif shown here which strikingly resembles the motif for murine TAL1(SCL)-GATA1 from the JASPAR CORE database (TOMTOM  $E$ -value =  $6.6 \times 10^{-11}$ ).

### Figure 2. KLF1 binds the mouse alpha and beta-globin loci at proximal promoters and locus control regions

(A) An image of the murine *alpha-globin* gene cluster and the upstream locus control region regulatory element (HS-31 to HS-8) from the UCSC Genome Browser. Genes and DNaseI Hypersensitivity (HS) sites are represented in the top track. KLF1 ChIP-seq signal (maroon) and Input ChIP-seq signal (blue) tracks are also shown together with KLF1 peak calls (KLF1 Peaks). An additional track describing GATA1 ChIP-seq peak calls from Cheng *et al.* is shown (GATA1 Peaks) together with a mammalian conservation track (Mammal Cons).

(B) An image of the murine *beta-globin* gene cluster and locus control region flanked by olfactory receptor genes from the UCSC Genome Browser. Genes and DNaseI HS sites are represented in the top track. Other tracks are presented as for (A).

**Figure 3. KLF1 directly regulates synthesis of heme and procurement of iron**

(A) Schematic diagram of an erythroid cell simplistically depicting the pathways of heme synthesis, iron procurement, heme export, and hemoglobin assembly. Genes are shown in *italics* at their point of involvement in the pathways. The components of the Mitochondrial Permeability Transition Pore (MPTP) complex are shown inset for clarity. Genes in the figure are color coded as shown in the Peak Height Key according to peak height from KLF1 ChIP-seq.

(B) Relative gene expression in *Klf1* <sup>-/-</sup> fetal liver (compared to WT) for many of the genes depicted in (A). Gene expression as determined by qRT-PCR is normalized to the housekeeping gene *HPRT* and shown for *Klf1* <sup>-/-</sup> fetal liver relative to WT (set to 1 as shown by the red dotted line). Data is presented as Mean + SEM. n ≥ 4 in each instance.

(C) Relative gene expression in 4-hydroxytamoxifen (Tam) induced K1-ER cells at 0, 3, 7, and 14h following induction. Gene expression has been determined by qRT-PCR and normalized to the housekeeping gene *HPRT*. Induced gene expression (Tam, maroon) is shown together with the corresponding vehicle control (EtOH, blue) on a Log<sub>2</sub> scale. Data are presented as Mean ± SEM, n = 4.

**Figure 4. KLF1 acts primarily as a direct activator of target genes and not exclusively via promoter regions**

(A) Each of the genes from the Hodge *et al.* expression profiling study were classified as “Down-regulated” in *Klf1* <sup>-/-</sup>, “Up-regulated” in *Klf1* <sup>-/-</sup>, or Non-responsive, when compared to WT expression. The plot shown represents the fraction of regions in the genes from each of these groups containing a KLF1 ChIP-seq peak binned into categories as shown in the cartoon underneath where the size of the Promoter and 3' Region are each 2kb. Each error bar represents the 95% confidence interval for the given fraction, assuming the number of regions with a peak follows a binomial distribution with N equal to the number of regions of that class. \* Note: genes consisting of a single exon do not contribute to the intragenic region calculation.

(B) A plot describing the relationship between fold-change in expression in *Klf1* <sup>-/-</sup> fetal liver compared to WT for genes that are “bound” by KLF1 within the gene or 2kb either side. The down-regulated in *Klf1* <sup>-/-</sup> fetal liver gene set is shown (ie. genes represented by maroon bars in (A)). The genes represented by some of the points are shown for clarity.

**Figure 5. Correlation with GATA1 ChIP-seq reveals extensive co-occupancy by KLF1 and GATA1 throughout the erythroid genome**

(A) A histogram of the distance between all KLF1 ChIP-seq peaks and the nearest GATA1 ChIP-seq peak as determined by Cheng *et al.* plotted on a  $\text{Log}_{10}$  scale as for Figure 1A. The bimodal distribution of distances indicates “Shared” and “Unique” binding sites for KLF1 as shown.

(B) Histogram plots of the distance between the KLF1 motif and GATA1 motifs in the same (Same Strand) and opposing (Opposite Strand) directions for the KLF1 ChIP-seq peaks.

(C) A model for KLF1-GATA1 co-operation based on the data presented in (B). Crystal structures of the GATA1 DNA binding zinc fingers and Zif268 (representative of KLF1) imposed onto a DNA backbone are shown in pink and blue respectively. The preferred directions and spacing of the motifs for co-operation is shown. The hypothesized existence of SCL between the two factors binding to a “CTG” motif is indicated in grey.

**Figure 6. Analysis of KLF1 and GATA1 bound genomic regions reveals potential interacting transcription factor partners**

A schematic representation of the overlap between KLF1, GATA1 (Cheng *et al.*) and SCL (Wilson *et al.*) erythroid ChIP-seq datasets defines two major groups of KLF1 binding sites, total binding sites (945) and shared KLF1-GATA1 binding sites (397). Interrogation of binding site sequences using the Clover algorithm identifies the motifs shown as being over-represented in the groups shown at 0.05 and 0.1 *p*-value significance levels.

Genome Research  
KLF1 Erythroid Regulatory Networks

Table 1. Highest Ranking KLF1 ChIP-seq Peaks

Peak Height	Genomic Location (mm9)	Nearest Gene Symbol	Nearest Gene Name	Previously Identified Target	Peak Position
57	Chr17:35878969-35878969	<b>Ddr1</b>	<i>Discoidin domain receptor family, member 1</i>		Intergenic
47	Chr6:72229669-72229671	<b>Sftpb</b>	<i>Surfactant associated protein B</i>		Intergenic
47	Chr8:19984277-19984280	<b>6820431F20Rik<sup>#</sup></b>	<i>RIKEN non-coding RNA</i>		Other Intronic
44	Chr17:26651596-26651596	<b>Dusp1</b>	<i>Dual specificity phosphatase 1</i>		Intergenic
42	Chr14:70866215-70866217	<b>Phyhip</b>	<i>Phytanoyl-CoA hydroxylase interacting protein</i>		Other Intronic
37	Chr16:8655920-8655920	<b>Pmm2</b>	<i>Phosphomannomutase 2</i>		Other Intronic
36	Chr15:85935903-85935904	<b>Gramd4</b>	<i>GRAM domain containing 4</i>		Other Intronic
35	Chr11:32151066-32151067	<b>Mare</b>	<i>Alpha globin regulatory element containing</i>	(Vernimmen et al. 2007)	Other Intronic
35	Chr3:103708832-103708836	<b>Ptpn22</b>	<i>Protein tyrosine phosphatase, non-receptor type 22</i>		Other Intronic
35	Chr4:62177796-62177808	<b>Alad</b>	<i>Aminolevulinate, delta-, dehydratase</i>		Intron 1
34	Chr1:36148310-36148311	<b>Hs6st1</b>	<i>Heparan sulphate 6-O-sulfotransferase 1</i>		Intron 1
34	Chr15:80591622-80591622	<b>Tnrc6b</b>	<i>Trinucleotide repeat containing 6b isoform 1</i>		Other Intronic
34	Chr18:32702272-32702272	<b>Gypc</b>	<i>Glycophorin C</i>	(Hodge et al. 2006)	Intron 1
34	Chr4:151100744-151100749	<b>Camta1</b>	<i>Calmodulin binding transcription activator 1</i>		Other Intronic
34	Chr4:151334799-151334802	<b>Dnajc11</b>	<i>DnaJ homolog, subfamily C, member 11</i>		Other Intronic
33	Chr17:29067432-29067432	<b>Pxt1</b>	<i>Peroxisomal testis-specific protein 1</i>		Intergenic
32	Chr15:56455310-56455316	<b>Has2</b>	<i>Hyaluronan synthase 2</i>		Intergenic
31	Chr17:26079773-26079775	<b>Pigq</b>	<i>Phosphatidylinositol N-acetylglucosaminyltransferase subunit Q</i>		Promoter
31	Chr4:34497718-34497723	<b>Akirin2</b>	<i>Akirin2</i>		Promoter
31	Chr8:126512748-126512749	<b>Abc-me</b>	<i>ABC mitochondrial erythroid</i>		Intergenic
30	Chr4:118862630-118862630	<b>Ermap</b>	<i>Erythroid membrane-associated protein</i>	(Hodge et al. 2006)	Promoter
30	Chr5:65256112-65256116	<b>Klf3</b>	<i>Kruppel-like factor 3</i>	(Funnell et al. 2007)	Intergenic
30	Chr6:86028020-86028021	<b>Add2</b>	<i>Adducin 2</i>		Promoter
29	Chr1:137838412-137838414	<b>Igfn1</b>	<i>Immunoglobulin-like and fibronectin type III domain-containing protein 1</i>		Intergenic
29	Chr10:40726672-40726673	<b>Gpr6</b>	<i>G protein-coupled receptor 6</i>		Intergenic
29	Chr4:124283662-124283662	<b>Pou3f1</b>	<i>POU domain, class 3, transcription factor 1</i>		Intergenic
29	Chr4:135734089-135734108	<b>E2f2</b>	<i>E2f transcription factor 2</i>	(Pilon et al. 2008; Tallack et al. 2009)	Intron 1
29	Chr5:139673057-139673057	<b>Unc84a</b>	<i>Unc-84 homolog A</i>		Promoter
29	Chr8:107824107-107824107	<b>E2f4</b>	<i>E2f transcription factor 4</i>	(Tallack et al. 2009)	Intronic
29	Chr8:24164688-24164690	<b>Ank1</b>	<i>Ankyrin 1</i>	(Nilson et al. 2006)	Intron 1/Promoter
28	Chr10:41632340-41632340	<b>Armc2</b>	<i>Armado repeat containing 2</i>		Intergenic
28	Chr12:112797042-112797043	<b>Eif5</b>	<i>Eukaryotic translation initiation factor 5</i>		Intergenic
28	Chr2:103962300-103962302	<b>A930018P22Rik</b>	<i>RIKEN cDNA for hypothetical protein LOC68243</i>		Promoter
28	Chr2:125973640-125973654	<b>4930525F21Rik</b>	<i>RIKEN cDNA for hypothetical protein LOC75823</i>		Promoter
27	Chr10:111262060-111262060	<b>Krr1</b>	<i>HIV-1 Rev binding protein 2</i>		Intergenic
27	Chr13:112833977-112833981	<b>Map3k1</b>	<i>Mitogen-activated protein kinase kinase kinase 1</i>		Intergenic
27	Chr17:12397892-12397892	<b>Agpat4</b>	<i>1-acylglycerol-3-phosphate O-acyltransferase 4</i>		Other Intronic
27	Chr17:25039865-25039870	<b>Mapk8ip3</b>	<i>Mitogen-activated protein kinase 8 interacting</i>		Other Intronic
26	Chr11:104890558-104890569	<b>Efcab3</b>	<i>EF-hand calcium binding domain 3</i>		Intergenic
26	Chr14:69923035-69923039	<b>D930020E02Rik</b>	<i>Envelope glycoprotein syncytin-B</i>		Intergenic
26	Chr17:29575527-29575538	<b>Pim1</b>	<i>Proviral integration site 1</i>		Intergenic
26	Chr2:164599599-164599600	<b>Ube2c</b>	<i>Ubiquitin-conjugating enzyme E2C</i>	(Hodge et al. 2006)	Intergenic
26	Chr2:172116029-172116043	<b>Mc3r</b>	<i>Melanocortin 3 receptor</i>		Intergenic
26	Chr4:143002576-143002578	<b>Pramef8</b>	<i>PRAME family member 8</i>		5' UTR
26	Chr9:108243328-108243330	<b>Gpx1</b>	<i>Glutathione peroxidase 1</i>		Intergenic
26	Chr9:66351957-66351962	<b>Herc1</b>	<i>Hect domain and RCC1-like domain 1</i>		Other Intronic
25	Chr13:114606283-114606286	<b>Arl15</b>	<i>ADP-ribosylation factor-like 15</i>		Intron 1
25	Chr13:35769443-35769443	<b>Cdyl</b>	<i>Chromodomain protein, Y chromosome-like</i>		Other Intronic
25	Chr15:78243638-78243639	<b>Mpst</b>	<i>Mercaptopyruvate sulfurtransferase</i>		Other Intronic
25	Chr6:88586555-88586556	<b>Kbtbd12</b>	<i>Kelch repeat and BTB domain containing 12</i>		Intergenic
25	Chr7:128851130-128851134	<b>Usp31</b>	<i>Ubiquitin specific peptidase 31</i>		Promoter
25	ChrX:39372687-39372688	<b>Xiap</b>	<i>X-linked inhibitor of apoptosis</i>		Intergenic

<sup>#</sup>The **6820431F20Rik** peak appears to be an artefact of chromatin conformation. This is further discussed in Supplemental Figure S1.

## References

- Andrews, N.C., H. Erdjument-Bromage, M.B. Davidson, P. Tempst, and S.H. Orkin. 1993. Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**: 722-728.
- Bailey, T.L. 2002. Discovering novel sequence motifs with MEME. *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2 4.
- Brown, R.C., S. Pattison, J. van Ree, E. Coghill, A. Perkins, S.M. Jane, and J.M. Cunningham. 2002. Distinct domains of erythroid Kruppel-like factor modulate chromatin remodeling and transactivation at the endogenous beta-globin gene promoter. *Mol Cell Biol* **22**: 161-170.
- Buske, F.A., M. Boden, D.C. Bauer, and T.L. Bailey. 2010. Assigning roles to DNA regulatory motifs using comparative genomics. *Submitted for publication*.
- Cantor, A.B. and S.H. Orkin. 2002. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**: 3368-3376.
- Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, and H.H. Ng. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.
- Cheng, Y., W. Wu, S.A. Kumar, D. Yu, W. Deng, T. Tripic, D.C. King, K.B. Chen, Y. Zhang, D. Drautz, B. Giardine, S.C. Schuster, W. Miller, F. Chiaromonte, Y. Zhang, G.A. Blobel, M.J. Weiss, and R.C. Hardison. 2009. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19**: 2172-2184.
- Cloonan, N., Q. Xu, G.J. Faulkner, D.F. Taylor, D.T. Tang, G. Kolle, and S.M. Grimmond. 2009. RNA-MATE: a recursive mapping strategy for high-throughput RNA-sequencing data. *Bioinformatics* **25**: 2615-2616.
- Coghill, E., S. Eccleston, V. Fox, L. Cerruti, C. Brown, J. Cunningham, S. Jane, and A. Perkins. 2001. Erythroid Kruppel-like factor (EKLF) coordinates erythroid cell proliferation and hemoglobinization in cell lines derived from EKLF null mice. *Blood* **97**: 1861-1868.
- Dirlam, A., B.T. Spike, and K.F. Macleod. 2007. Deregulated E2f-2 underlies cell cycle and maturation defects in retinoblastoma null erythroblasts. *Mol Cell Biol* **27**: 8713-8728.
- Drissen, R., R.J. Palstra, N. Gillemans, E. Splinter, F. Grosveld, S. Philipsen, and W. de Laat. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* **18**: 2485-2490.
- Drissen, R., M. von Lindern, A. Kolbus, S. Driegen, P. Steinlein, H. Beug, F. Grosveld, and S. Philipsen. 2005. The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol* **25**: 5205-5214.
- Dunn, L.L., Y.S. Rahmanto, and D.R. Richardson. 2007. Iron uptake and metabolism in the new millennium. *Trends Cell Biol* **17**: 93-100.
- Eaton, S.A., A.P. Funnell, N. Sue, H. Nicholas, R.C. Pearson, and M. Crossley. 2008. A network of Kruppel-like Factors (Klfs). Klf8 is repressed by Klf3 and activated by Klf1 in vivo. *J Biol Chem* **283**: 26937-26947.

Genome Research  
KLF1 Erythroid Regulatory Networks

- Feng, W.C., C.M. Southwood, and J.J. Bieker. 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* **269**: 1493-1500.
- Fraser, S.T., J. Isern, and M.H. Baron. 2007. Maturation and enucleation of primitive erythroblasts during mouse embryogenesis is accompanied by changes in cell-surface antigen expression. *Blood* **109**: 343-352.
- Frith, M.C., Y. Fu, L. Yu, J.F. Chen, U. Hansen, and Z. Weng. 2004. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* **32**: 1372-1381.
- Fujiwara, T., H. O'Geen, S. Keles, K. Blahnik, A.K. Linnemann, Y.A. Kang, K. Choi, P.J. Farnham, and E.H. Bresnick. 2009. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* **36**: 667-681.
- Fujiwara, Y., C.P. Browne, K. Cunniff, S.C. Goff, and S.H. Orkin. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**: 12355-12358.
- Funnell, A.P., C.A. Maloney, L.J. Thompson, J. Keys, M. Tallack, A.C. Perkins, and M. Crossley. 2007. Erythroid Kruppel-like factor directly activates the basic Kruppel-like factor gene in erythroid cells. *Mol Cell Biol* **27**: 2777-2790.
- Gregory, R.C., D.J. Taxman, D. Seshasayee, M.H. Kensinger, J.J. Bieker, and D.M. Wojchowski. 1996. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood* **87**: 1793-1801.
- Hardison, R., J.L. Slightom, D.L. Gumucio, M. Goodman, N. Stojanovic, and W. Miller. 1997. Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205**: 73-94.
- Hodge, D., E. Coghill, J. Keys, T. Maguire, B. Hartmann, A. McDowall, M. Weiss, S. Grimmond, and A. Perkins. 2006. A global role for EKLF in definitive and primitive erythropoiesis. *Blood* **107**: 3359-3370.
- Huber, T.L., A.C. Perkins, A.E. Deconinck, F.Y. Chan, P.E. Mead, and L.I. Zon. 2001. neptune, a Kruppel-like transcription factor that participates in primitive erythropoiesis in *Xenopus*. *Curr Biol* **11**: 1456-1461.
- Im, H., J.A. Grass, K.D. Johnson, S.I. Kim, M.E. Boyer, A.N. Imbalzano, J.J. Bieker, and E.H. Bresnick. 2005. Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc Natl Acad Sci U S A* **102**: 17065-17070.
- Johnsen, O., P. Murphy, H. Prydz, and A.B. Kolsto. 1998. Interaction of the CNC-bZIP factor TCF11/LCR-F1/Nrf1 with MafG: binding-site selection and regulation of transcription. *Nucleic Acids Res* **26**: 512-520.
- Keys, J.R., M.R. Tallack, D.J. Hodge, S.O. Cridland, R. David, and A.C. Perkins. 2007. Genomic organisation and regulation of murine alpha haemoglobin stabilising protein by erythroid Kruppel-like factor. *Br J Haematol* **136**: 150-157.
- Kinross, K.M., A.J. Clark, R.M. Iazzolino, and P.O. Humbert. 2006. E2f4 regulates fetal erythropoiesis through the promotion of cellular proliferation. *Blood* **108**: 886-895.
- Kotkow, K.J. and S.H. Orkin. 1995. Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol Cell Biol* **15**: 4640-4647.
- Koury, M.J., S.T. Sawyer, and S.J. Brandt. 2002. New insights into erythropoiesis. *Curr Opin Hematol* **9**: 93-100.

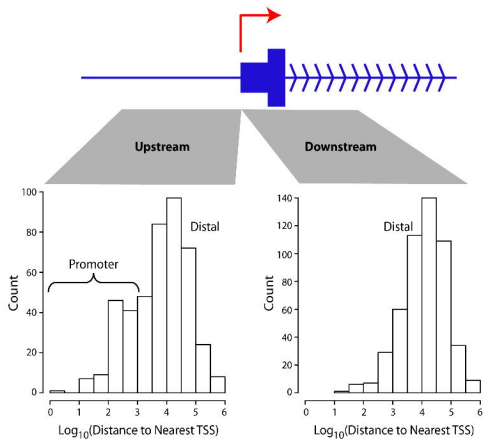
- Lohmann, F. and J.J. Bieker. 2008. Activation of Eklf expression during hematopoiesis by Gata2 and Smad5 prior to erythroid commitment. *Development*.
- Merika, M. and S.H. Orkin. 1995. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* **15**: 2437-2447.
- Miller, I.J. and J.J. Bieker. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* **13**: 2776-2786.
- Newburger, D.E. and M.L. Bulyk. 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **37**: D77-82.
- Nilson, D.G., D.E. Sabatino, D.M. Bodine, and P.G. Gallagher. 2006. Major erythrocyte membrane protein genes in EKLF-deficient mice. *Exp Hematol* **34**: 705-712.
- Nuez, B., D. Michalovich, A. Bygrave, R. Ploemacher, and F. Grosveld. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**: 316-318.
- Papadopoulos, V., M. Baraldi, T.R. Guilarte, T.B. Knudsen, J.J. Lacapere, P. Lindemann, M.D. Norenberg, D. Nutt, A. Weizman, M.R. Zhang, and M. Gavish. 2006. Translocator protein (18kDa): new nomenclature for the peripheral-type benzodiazepine receptor based on its structure and molecular function. *Trends Pharmacol Sci* **27**: 402-409.
- Perkins, A. 1999. Erythroid Kruppel like factor: from fishing expedition to gourmet meal. *International Journal of Biochemistry & Cell Biology* **31**: 1175-1192.
- Perkins, A.C., A.H. Sharpe, and S.H. Orkin. 1995. Lethal Beta-Thalassemia in Mice Lacking the Erythroid Cacc-Transcription Factor Eklf. *Nature* **375**: 318-322.
- Pevny, L., M.C. Simon, E. Robertson, W.H. Klein, S.F. Tsai, V. D'Agati, S.H. Orkin, and F. Costantini. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**: 257-260.
- Pilon, A.M., M.O. Arcasoy, H.K. Dressman, S.E. Vayda, Y.D. Maksimova, J.I. Sangerman, P.G. Gallagher, and D.M. Bodine. 2008. Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol Cell Biol* **28**: 7394-7401.
- Pilon, A.M., D.G. Nilson, D. Zhou, J. Sangerman, T.M. Townes, D.M. Bodine, and P.G. Gallagher. 2006. Alterations in Expression and Chromatin Configuration of the Alpha Hemoglobin-Stabilizing Protein Gene in Erythroid Kruppel-Like Factor-Deficient Mice. *Mol. Cell. Biol.* **26**: 4368-4377.
- Ponka, P. 1997. Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood* **89**: 1-25.
- Porayette, P. and R.F. Paulson. 2008. BMP4/Smad5 dependent stress erythropoiesis is required for the expansion of erythroid progenitors during fetal development. *Dev Biol* **317**: 24-35.
- Rodriguez, P., E. Bonte, J. Krijgsveld, K.E. Kolodziej, B. Guyot, A.J. Heck, P. Vyas, E. de Boer, F. Grosveld, and J. Strouboulis. 2005. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J* **24**: 2354-2366.
- Schoenfelder, S., T. Sexton, L. Chakalova, N.F. Cope, A. Horton, S. Andrews, S. Kurukuti, J.A. Mitchell, D. Umlauf, D.S. Dimitrova, C.H. Eskiw, Y. Luo, C.L. Wei, Y. Ruan, J.J. Bieker, and P. Fraser. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**: 53-61.

- Shyu, Y.C., S.C. Wen, T.L. Lee, X. Chen, C.T. Hsu, H. Chen, R.L. Chen, J.L. Hwang, and C.K. Shen. 2006. Chromatin-binding in vivo of the erythroid kruppel-like factor, EKLF, in the murine globin loci. *Cell Res* **16**: 347-355.
- Siatecka, M., L. Xue, and J.J. Bieker. 2007. Sumoylation of EKLF promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Mol Cell Biol* **27**: 8547-8560.
- Starck, J., N. Cohet, C. Gonnet, S. Sarrazin, Z. Doubeikovskaia, A. Doubeikovski, A. Verger, M. Duterque-Coquillaud, and F. Morle. 2003. Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Mol Cell Biol* **23**: 1390-1402.
- Storey, J.D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**: 9440-9445.
- Tallack, M.R., J.R. Keys, P.O. Humbert, and A.C. Perkins. 2009. EKLF/KLF1 controls cell cycle entry via direct regulation of E2f2. *J Biol Chem* **284**: 20966-20974.
- Tallack, M.R., J.R. Keys, and A.C. Perkins. 2007. Erythroid Kruppel-like factor regulates the G1 cyclin dependent kinase inhibitor p18INK4c. *J Mol Biol* **369**: 313-321.
- Tallack, M.R. and A.C. Perkins. 2009. Megakaryocyte-erythroid lineage promiscuity in EKLF null mouse blood. *Haematologica*.
- Testa, U. 2004. Apoptotic mechanisms in the control of erythropoiesis. *Leukemia* **18**: 1176-1199.
- Tripic, T., W. Deng, Y. Cheng, Y. Zhang, C.R. Vakoc, G.D. Gregory, R.C. Hardison, and G.A. Blobel. 2009. SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113**: 2191-2201.
- van Vliet, J., L.A. Crofts, K.G. Quinlan, R. Czolij, A.C. Perkins, and M. Crossley. 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* **87**: 474-482.
- Vernimmen, D., M. De Gobbi, J.A. Sloane-Stanley, W.G. Wood, and D.R. Higgs. 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *Embo J* **26**: 2041-2051.
- Wei, C.L., Q. Wu, V.B. Vega, K.P. Chiu, P. Ng, T. Zhang, A. Shahab, H.C. Yong, Y. Fu, Z. Weng, J. Liu, X.D. Zhao, J.L. Chew, Y.L. Lee, V.A. Kuznetsov, W.K. Sung, L.D. Miller, B. Lim, E.T. Liu, Q. Yu, H.H. Ng, and Y. Ruan. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**: 207-219.
- Weiss, M.J. and S.H. Orkin. 1995. GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol* **23**: 99-107.
- Welch, J.J., J.A. Watts, C.R. Vakoc, Y. Yao, H. Wang, R.C. Hardison, G.A. Blobel, L.A. Chodosh, and M.J. Weiss. 2004. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**: 3136-3147.
- Whittington, T., M.C. Frith, and T.L. Bailey. 2010. Inferring transcription factor complexes from ChIP-seq data. *Manuscript in preparation*.
- Whittington, T., A.C. Perkins, and T.L. Bailey. 2009. High-throughput chromatin information enables accurate tissue-specific prediction of transcription factor binding sites. *Nucleic Acids Res* **37**: 14-25.
- Wilson, N.K., D. Miranda-Saavedra, S. Kinston, N. Bonadies, S.D. Foster, F. Calero-Nieto, M.A. Dawson, I.J. Donaldson, S. Dumon, J. Frampton, R. Janky, X.H. Sun, S.A. Teichmann, A.J. Bannister, and B. Gottgens. 2009. The transcriptional program

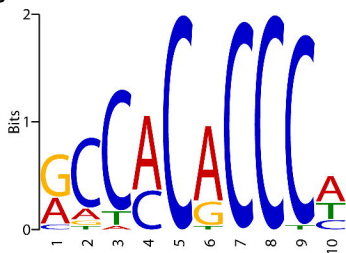
Genome Research  
KLF1 Erythroid Regulatory Networks

- controlled by the stem cell leukemia gene *Scl/Tal1* during early embryonic hematopoietic development. *Blood* **113**: 5456-5465.
- Ye, T.Z., C.T. Gordon, Y.H. Lai, Y. Fujiwara, L.L. Peters, A.C. Perkins, and D.H. Chui. 2000. Ermap, a gene coding for a novel erythroid specific adhesion/receptor membrane protein. *Gene* **242**: 337-345.
- Yu, M., L. Riva, H. Xie, Y. Schindler, T.B. Moran, Y. Cheng, D. Yu, R. Hardison, M.J. Weiss, S.H. Orkin, B.E. Bernstein, E. Fraenkel, and A.B. Cantor. 2009. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* **36**: 682-695.
- Zhang, J., M. Socolovsky, A.W. Gross, and H.F. Lodish. 2003. Role of Ras signaling in erythroid differentiation of mouse fetal liver cells: functional analysis by a flow cytometry-based novel culture system. *Blood* **102**: 3938-3946.

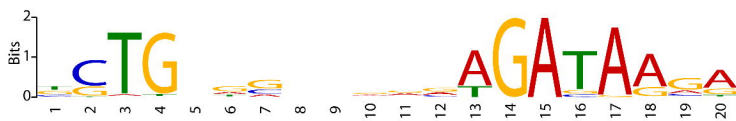
A



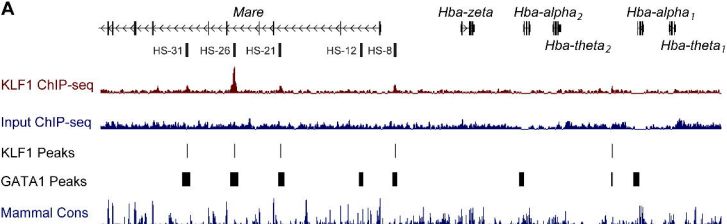
B



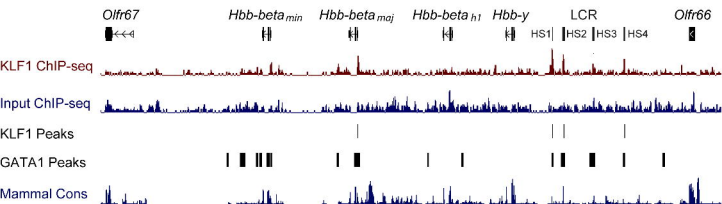
C

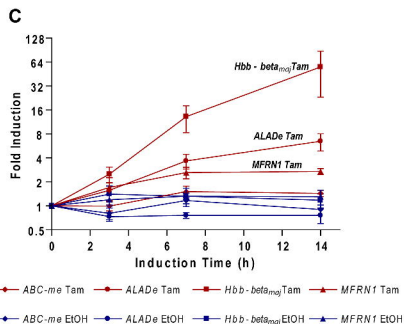
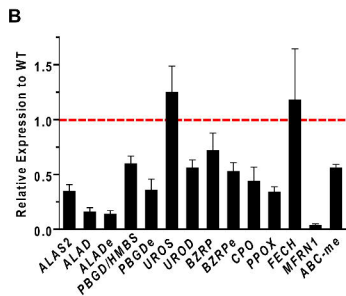
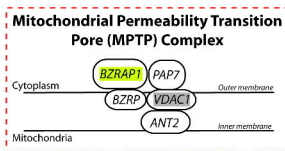
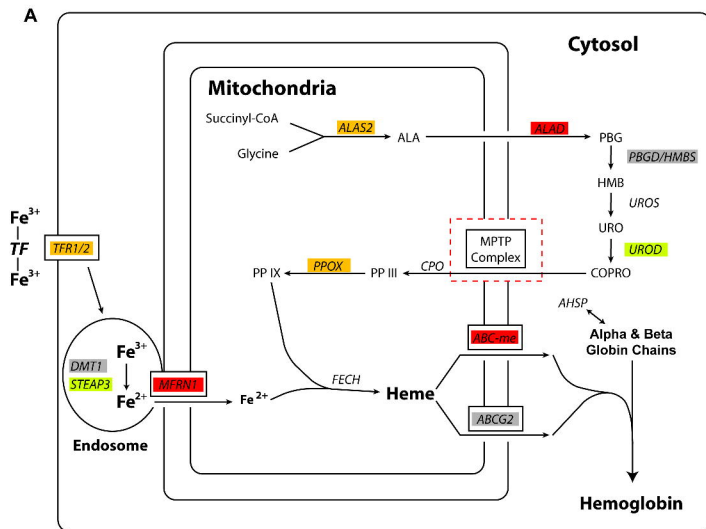


A

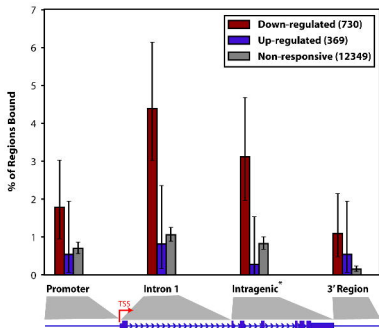


B

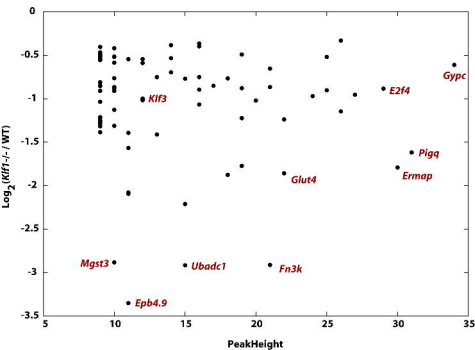


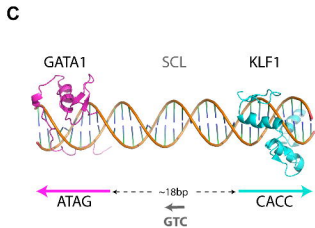
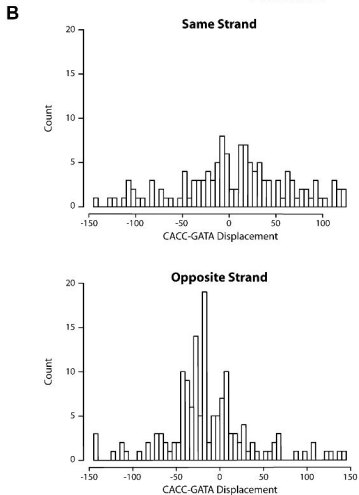
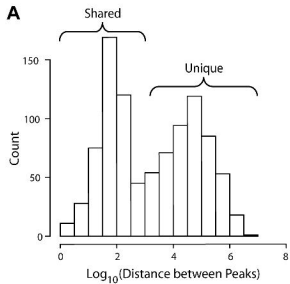


A



B





**FIGURE 6**

■ GATA1 (15339)  
■ KLF1 (945)  
■ SCL (222)

