

Chromatin Architecture and Transcription Factor Binding Regulate Expression of Erythrocyte Membrane Protein Genes^{∇†}

Laurie A. Steiner,¹ Yelena Maksimova,¹ Vincent Schulz,¹ Clara Wong,¹ Debasish Raha,² Milind C. Mahajan,³ Sherman M. Weissman,³ and Patrick G. Gallagher^{1*}

Department of Pediatrics, Yale University School of Medicine, New Haven, Connecticut 06520-8021¹; Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520-8103²; and Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520-8005³

Received 15 June 2009/Returned for modification 27 July 2009/Accepted 9 August 2009

Erythrocyte membrane protein genes serve as excellent models of complex gene locus structure and function, but their study has been complicated by both their large size and their complexity. To begin to understand the intricate interplay of transcription, dynamic chromatin architecture, transcription factor binding, and genomic organization in regulation of erythrocyte membrane protein genes, we performed chromatin immunoprecipitation (ChIP) coupled with microarray analysis and ChIP coupled with massively parallel DNA sequencing in both erythroid and nonerythroid cells. Unexpectedly, most regions of GATA-1 and NF-E2 binding were remote from gene promoters and transcriptional start sites, located primarily in introns. Cooccupancy with FOG-1, SCL, and MTA-2 was found at all regions of GATA-1 binding, with cooccupancy of SCL and MTA-2 also found at regions of NF-E2 binding. Cooccupancy of GATA-1 and NF-E2 was found frequently. A common signature of histone H3 trimethylation at lysine 4, GATA-1, NF-E2, FOG-1, SCL, and MTA-2 binding and consensus GATA-1-E-box binding motifs located 34 to 90 bp away from NF-E2 binding motifs was found frequently in erythroid cell-expressed genes. These results provide insights into our understanding of membrane protein gene regulation in erythropoiesis and the regulation of complex genetic loci in erythroid and nonerythroid cells and identify numerous candidate regions for mutations associated with membrane-linked hemolytic anemia.

The erythrocyte membrane is a multifunctional, complex structure that provides the red cell the deformability and stability required to withstand its travels through macro- and microcirculation. It plays critical roles in erythropoiesis, including responding to erythropoietin, importing iron required for hemoglobin synthesis, and regulating cellular metabolism. Qualitative and quantitative disorders of erythrocyte membrane proteins have been associated with inherited abnormalities of red cell shape, including hereditary spherocytosis, hereditary elliptocytosis, and hereditary pyropoikilocytosis syndromes (65, 103). Despite biochemical and genetic linkage to specific erythrocyte membrane protein genes, e.g., ankyrin-1, α - or β -spectrin, and band 3, mutations are found in the coding exons and promoter regions of only ~75% of cases studied. This suggests that the disease-causing mutation is located in critical regulatory regions outside the promoters and exons in a quarter of cases.

Most erythrocyte membrane protein genes are large, comprised of >25 exons. They encode numerous diverse and complex isoforms, frequently generated by alternate splicing, alternate promoter usage, or alternate polyadenylation (18). In many cases, alternate promoters direct combinations of exons encoding diverse tissue-specific, cell type-specific, developmental-stage-specific, and/or differentiation stage-specific isoforms (6, 12, 13, 19, 21–24, 44, 52, 62, 78, 86, 108, 112–114). As

such, erythrocyte membrane protein genes serve as excellent models of complex gene locus structure and function. Study of the regulation of erythrocyte membrane protein genes has been hampered both by their large size and by their complexity. Limited information regarding their regulation in erythroid cells is available and consists primarily of *in vitro* studies of core erythroid cell promoters (6, 21, 22, 51, 53).

Advances in technology have permitted the rapid identification of critical regions of gene regulation on a genome-wide scale, identifying regions bound by transcription factors and other DNA-associated proteins, delineating regions of various histone architectures, and revealing the methylation status of regions of DNA. Techniques available for mapping protein-DNA interactions *in vivo* couple chromatin immunoprecipitation (ChIP) with microarrays that contain regions of genomic DNA (ChIP-chip) or with massively parallel DNA sequencing (ChIP-seq). These technologies are ideally suited for application to the study of the regulation of the large and complex membrane protein gene loci in erythroid cells.

To begin to understand the intricate interplay of transcription, dynamic chromatin architecture, transcription factor binding, and genomic organization in regulation of erythrocyte membrane protein genes, we performed ChIP-chip and ChIP-seq with erythroid and nonerythroid cells. Unexpectedly, most regions of GATA-1 and NF-E2 binding were remote from gene promoters and transcriptional start sites (TSS), located primarily in introns. Cooccupancy with FOG-1, SCL, and MTA-2 was found at all regions of GATA-1 binding. Interestingly, cooccupancy of SCL and MTA-2 was also found at regions of NF-E2 binding. Cooccupancy of GATA-1 and NF-E2 was found frequently. A common signature of histone H3 trimethyl-

* Corresponding author. Mailing address: Department of Pediatrics, Yale University School of Medicine, 333 Cedar Street, P.O. Box 208064, New Haven, CT 06520-8064. Phone: (203) 688-2896. Fax: (203) 785-6974. E-mail: patrick.gallagher@yale.edu.

† Supplemental material for this article may be found at <http://mc.manuscriptcentral.com/mcb>.

[∇] Published ahead of print on 17 August 2009.

ation at lysine 4 (H3Me3K4), GATA-1, NF-E2, FOG-1, SCL, and MTA-2 binding and consensus GATA-1-E-box binding motifs located 34 to 90 bp away from NF-E2 binding motifs was found frequently in erythroid cell-expressed genes.

These results provide insights into our understanding of membrane protein gene regulation in erythropoiesis and the regulation of complex genetic loci in erythroid and nonerythroid cells and identify numerous candidate regions for mutations associated with membrane-linked hemolytic anemia.

MATERIALS AND METHODS

Cell culture and selection. K562 cells (chronic myelogenous leukemia in blast crisis with erythroid characteristics, ATCC, CCL 243) were maintained in RPMI 1640 medium with 10% fetal calf serum. HeLa cells (epithelial-like carcinoma, cervix, CCL 2) were maintained in Eagle's minimal essential medium with 10% fetal calf serum. Human CD34-selected stem and progenitor cells were obtained from the Yale Center of Excellence in Molecular Hematology Cell Core and cultured in StemSpan SF expansion medium (StemSpan 09650) with estradiol (100 ng/ml), dexamethasone (10 ng/ml), human transferrin (200 ng/ml), insulin (10 ng/ml), Flt3 ligand (100 ng/ml), stem cell factor (100 ng/ml), interleukin-3 (50 ng/ml), interleukin-6 (20 ng/ml), insulin-like growth factor 1 (50 ng/ml), and erythropoietin (3 U/ml) for 9 to 14 days (63, 77). Fluorescence-activated cell sorter analysis was used to analyze the cellular expression of CD71 (transferrin receptor) and CD235a (glycophorin A). Magnetic bead selection for CD71 (MACS 130-046-201; Miltenyi Biotec) and CD235a (MACS 130-050-501; Miltenyi Biotec) was used to purify an R3/R4 population of erythroid cells (117).

mRNA expression. RNA prepared from K562 cells, HeLa cells, and primary, cultured human erythroid cells (RNeasy mini kit; Qiagen) was treated with amplification-grade DNase I and reverse transcribed with an oligo(dT) primer by use of a SuperScript first-strand synthesis system (Invitrogen). Primer pairs for each of the 15 target membrane protein genes were designed using Primer 3 software. Reverse transcription products were amplified by real-time PCR using an iCycler (Bio-Rad) with the primers listed in Table S1 in the supplemental material. PCR specificity was verified by assessing amplification product melting curves. Real-time PCR data were normalized to an ornithine decarboxylase antizyme 1 (OAZ1) mRNA control. The changes in specific mRNA levels were calculated using the $\Delta\Delta C_T$ method (where C_T is threshold cycle), with results presented as means \pm standard errors of the means. Results were normalized to the gene with the highest expression level in each group (100). Triplicate analyses were performed for each target gene.

ChIP-chip. ChIP assays were performed as previously described (20), with minor variation (ChIP kit 17-295; Upstate/Millipore). Briefly, 1% formaldehyde was added to $\sim 1 \times 10^8$ cells for 10 min at 37°C. Cells were washed twice in cold phosphate-buffered saline with protease inhibitors and then placed in hypotonic buffer for 20 min, followed by Dounce homogenization to isolate cross-linked nuclei. Nuclei were placed in sodium dodecyl sulfate lysis buffer for 30 min and then sonicated on ice with 21 10-s pulses, each followed by a 10-s rest period. Samples were diluted and then precleared at 4°C for 60 min with protein A or G agarose beads. Samples were immunoprecipitated for 12 to 18 h on a rotating platform at 4°C. Antibodies utilized for immunoprecipitation included H3Me3K4 (ab8580; Abcam), RNA polymerase II (RNAPII) (sc-899; Santa Cruz), GATA-1 (sc-265; Santa Cruz), and NF-E2 (sc-22827; Santa Cruz) antibodies, as well as nonspecific rabbit immunoglobulin G (IgG) (sc-2091; Santa Cruz). Antibody-bound DNA-protein complexes were collected using protein A or G agarose beads. Antibody-bound DNA-protein complexes were washed and eluted from the beads according to instructions. Cross-linking of DNA-protein adducts was reversed by incubation at 65°C for 4 h. The resulting DNA was cleaned with a QIAquick PCR purification kit (Qiagen) according to manufacturer instructions and amplified with a GenomePlex whole-genome amplification kit (Sigma) according to manufacturer instructions. Amplified DNA was cleaned using a QIAquick PCR purification kit (Qiagen) before amplification, labeling, and hybridization to arrays or before quantitative PCR (qPCR) analyses.

ChIP-seq. For ChIP-seq experiments, 1×10^8 K562 or HeLa cells were cross-linked with 1% formaldehyde for 10 min at room temperature, followed by Dounce homogenization. Cross-linked nuclei were isolated, followed by sonication to obtain chromatin-containing DNA fragments with an average size of ~ 500 bp. For each ChIP, 20 μ g of antibody or the appropriate control IgG species was used. The antigen-antibody complex was captured on protein G beads, washed four times with radioimmunoprecipitation assay buffer, and then

washed with phosphate-buffered saline. The DNA-protein complex was eluted from the protein G beads with 1% sodium dodecyl sulfate at 65°C, and cross-linking of DNA-protein adducts was reversed by overnight incubation at 65°C. After proteinase K and RNase digestion of the reverse-cross-linked sample, DNA was extracted with phenol-chloroform, precipitated with ethanol, suspended in 50 ml of 10 mM Tris-EDTA, pH 8.0, and used for Solexa deep sequencing and qPCR validation.

Quantitative ChIP analyses of factor cooccupancy. Additional ChIP experiments were performed with K562 cell chromatin using antibodies against FOG-1 (sc-9361; Santa Cruz), SCL (sc-12984; Santa Cruz), and MTA-2 (sc-9447; Santa Cruz) as described above for ChIP-chip, with the following modifications. Prior to the addition to formaldehyde, ethylene glycol-bis, dimethyl adipimidate, and di(*N*-succinimidyl) glutarate were added to the culture medium at a concentration of 0.15 mM, and the medium was incubated at room temperature for 30 min to maximize protein-protein cross-linking (116). The remainder of the ChIP assay was completed as described above.

ChIP-chip analyses. A custom, high-density genomic tiling array containing probes for the genomic regions of 100 genes expressed in erythroid cells, including the 15 target membrane protein genes, was designed with NimbleGen Systems software. The probes, designed to have an optimal melting temperature of $\geq 76^\circ\text{C}$, were typically ~ 50 bp in length, with some longer probes designed to increase the melting temperature to 76°C . Probes were tiled with 10- to 100-bp spacing, typically ~ 65 bp. Regions of repetitive DNA were excluded. Ten to 100 kb of flanking DNA was included on the chip for each locus, with 50 to 100 kb of flanking DNA included for the 15 target membrane protein genes. Labeling and hybridization of DNA samples for ChIP-chip analysis were performed by NimbleGen Systems.

Solexa sequencing. Approximately 500 ng of ChIP DNA was run on a 1.5% agarose gel to size select ChIP DNA in the 200- to 300-bp range. The size-selected ChIP DNA was purified using a gel extraction kit (Qiagen) and processed for Solexa sequencing as per manufacturer's protocol. Briefly, size-selected ChIP DNA was end repaired using an End-IT DNA end repair kit (ER0720; Epicenter), followed by addition of an adenine base at the 3' end by Klenow reaction and Solexa adaptor ligation. The modified DNA was then PCR amplified with one initial heating step of 98°C for 30 s, followed by 15 cycles of amplification with a melting temperature of 98°C for 30 s, an annealing temperature of 65°C for 30 s, and a product extension at 72°C for 30 s. At the end of the amplification, a final extension at 72°C for 5 min was performed. Amplified ChIP DNA was then size selected on a 1.5% agarose gel, purified by using a gel extraction kit (Qiagen), and subjected to deep sequencing. An average of 5 million reads was attained with each sequencing procedure.

Bioinformatics. Data obtained from ChIP-chip experiments were analyzed using the Tamalpais peak calling algorithm (4) to determine areas of DNA-protein interaction. Control and immunoprecipitation paired-data files were processed with R Smudgekit version 2.4 software (Jay Emerson, unpublished) to remove chip hybridization artifacts. Ratio GFF files of control and experimental data were generated, and the three replicate data files were processed together using the Tamalpais web server to generate candidate GATA-1 and NF-E2 binding regions. Peaks identified at all four levels of stringency were subjected to additional analyses because even though levels L1, L2, and L3 are more stringent and identify regions of binding with higher accuracy than L4, L4 peaks also often yield valid binding sites.

The conservation of each candidate region of GATA-1 and NF-E2 binding regions between corresponding genes of placental mammals was determined using the UCSC hg 18 genome browser database (43) with the 44-way placental-mammal PhastCons track (91). For each potential region of NF-E2 and GATA-1 binding, the maximum PhastCons score was determined using the galaxy aggregate function (25, 46, 48). The UCSC genome browser 7X regulatory potential table was used to determine the maximum regulatory potential (RP) scores for each region (16, 48). TESS, the Transcription Element Search System (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>), was used to identify potential GATA and NF-2E binding motifs in regions of GATA-1 and NF-E2 occupancy. Maximum PhastCons and RP scores were also determined for each GATA-1 and NF-E2 binding motif present in regions of GATA-1 and NF-E2 occupancy predicted by Tamalpais.

For ChIP-seq analyses of RNAPII binding, files for data visualization were created by the Yale CEGS Solexa processing pipeline. Each mapped sequence read was extended to a 200-bp window. The number of overlapping windows at each window start point was counted to generate the value displayed via an SGR file format.

Validation of ChIP-chip results. Primers were designed for each binding region in the target genes identified by the Tamalpais peak calling algorithm with a level of stringency of L1 to L4 (see Table S2 in the supplemental material).

TABLE 1. Major erythrocyte membrane genes

Gene	Designation	Chromosome	Start position (nt) ^a	End position (nt)	Amt of flanking DNA (kb)
Ankyrin-1	ANK1	8	41629900	41642241	100
α-Spectrin Band 3	SPTA1	1	156847119	156923130	100
β-Spectrin	SLC4A1	17	39682565	39700993	100
β-Adducin	SPTB	14	64282753	64359616	100
α-Adducin	ADD2	2	70742770	70793875	100
γ-Adducin	ADD1	4	2815381	2901587	100
ICAM-4	ADD3	10	111755715	111885313	100
ERMAP	ICAM-4	19	10258649	10260198	50
Protein 4.1R	ERMAP	1	43055362	43083247	50
Protein 4.2	EPB41	1	29086214	29319545	50
Dematin	EPB42	15	41276719	41289101	50
β-Actin	EPB49	8	21967101	21993917	50
Tropomodulin	ACTB	7	5533311	5535850	50
Tropomyosin	TMOD1	9	99303741	99403351	50
	TPM3	1	152394403	152411777	50

^a nt, nucleotide.

Immunoprecipitated DNA was analyzed by quantitative real-time PCR (iCycler; Bio-Rad) using the appropriate primers. SYBR green fluorescence in 25-μl PCRs was determined, and the amount of product was determined relative to a standard curve generated from a titration of input chromatin. Amplification of a single amplification product was confirmed by dissociation curve analysis. Enrichment of binding sites in target DNA over input was determined using ΔΔC_T analysis. Results are presented as means ± standard errors of the means, with triplicate analyses performed for each binding site. Student's *t* test was used to compare the level of enrichment attained with each specific antibody to the level of enrichment attained using nonspecific IgG. *P* values of >0.05 were considered to be significant.

Gel mobility shift analyses. Binding reactions were carried out as previously described (108). Competitor oligonucleotides were added at a 100-fold molar excess. Antibodies to GATA-1 were obtained from Santa Cruz Biotechnologies (M-20 and sc-1234; Santa Cruz, CA). Primers used in gel mobility shift analyses are listed in Table S3 in the supplemental material.

RESULTS

Expression of membrane protein genes in erythroid cells. In the erythrocyte membrane, there are 15 major proteins and hundreds of minor ones (66). The major membrane proteins are known to be critical for membrane assembly, structure, and function, and mutations in several of these genes result in inherited hemolytic anemia. Because there is limited information regarding the regulation of these genes in erythroid cells, to begin to identify and characterize the regulatory elements controlling their expression (Table 1), ChIP-chip or ChIP-seq was performed.

Quantitative reverse transcription-PCR was utilized to confirm that all 15 membrane genes were expressed both in K562 cells and in primary, cultured human erythroid cells. As expected, mRNA from all 15 genes was detectable both in K562 cells and in an R3/R4 population (117) of primary, cultured human erythroid cells. In HeLa cells, mRNA was detectable for 11 of 15 membrane genes, although generally at much lower levels than in K562 or primary, cultured erythroid cells (Fig. 1).

RNAPII binding and histone architecture of target genes. RNAPII preinitiation complex assembly is a feature of all protein-coding eukaryotic gene promoters. This process has been studied on a genome-wide scale via ChIP-chip and ChIP-seq with several cell types (8, 47, 83). RNAPII occupancy was examined across the 15 erythrocyte membrane genes in erythroid (K562) and nonerythroid (HeLa) cells by using ChIP-seq (Fig. 2; also see Fig. S1 in the supplemental material). In K562 cells, RNAPII binding was detected at a single promoter or putative promoter in 4 of 15 genes and at multiple promoters/putative promoters in 10 of 15 genes. The genes with multiple RNAPII peaks in K562 chromatin included ACTB, TPM3,

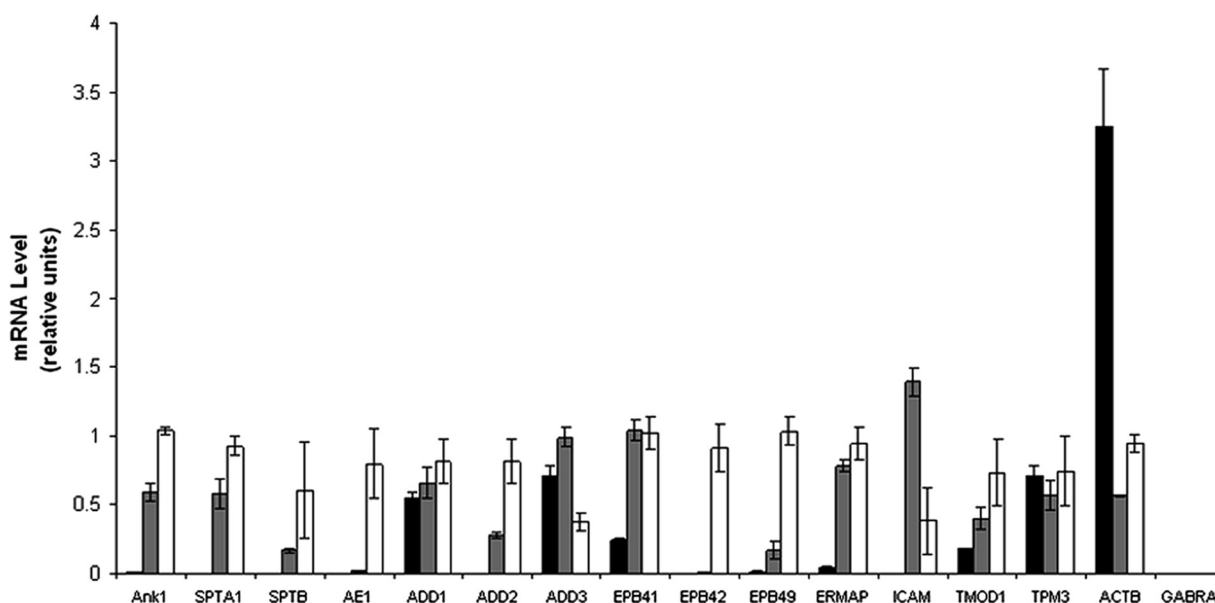


FIG. 1. Expression of major erythrocyte membrane genes. Reverse transcription-PCR was used to demonstrate mRNA expression levels of the 15 major erythrocyte membrane protein genes in erythroid- and nonerythroid-cell mRNA. Black bars represent mRNA expression in HeLa cells. Gray bars represent mRNA expression in K562 cells. White bars represent mRNA expression in primary, cultured erythroid cells. A region from the cerebellum-specific GABRA6 gene was included as a negative control.

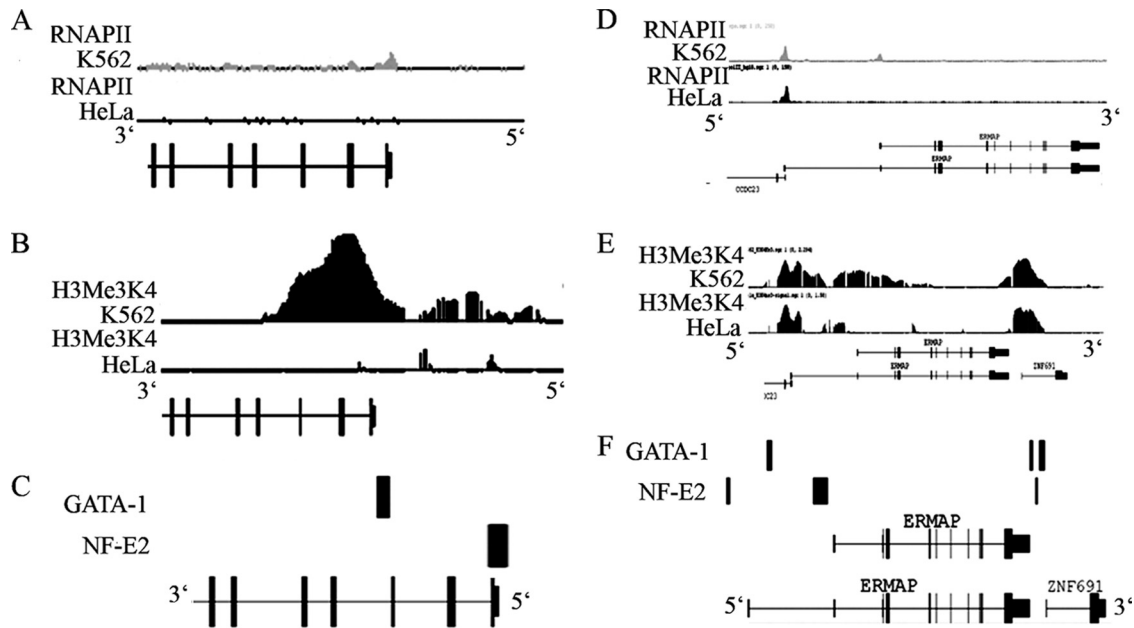


FIG. 2. Representative integrated genome browser (IGB) views of RNAPII, H3Me3K4, GATA-1, and NF-E2 binding at the 5' end of the α -spectrin gene (A to C) and the ERMAD gene locus (D to F). (A) IGB view of RNAPII binding at the 5' end of the α -spectrin promoter in erythroid (K562) and nonerythroid (HeLa) cells. There is a peak of RNAPII binding at the TSS of the α -spectrin gene in K562 cells but not in HeLa cells, demonstrating cell type-specific binding of RNAPII. (B) IGB view of H3Me3K4 binding at the 5' end of the α -spectrin gene in erythroid (K562) and nonerythroid (HeLa) cells. There is a peak of H3Me3K4 binding in K562 cells but not in HeLa cells, demonstrating cell type-specific enrichment for H3Me3K4. (C) IGB view of GATA-1 and NF-E2 binding at the 5' end of the α -spectrin gene, demonstrating NF-E2 binding at the promoter and GATA-1 binding in intron 3. (D) IGB view of RNAPII binding at the ERMAD locus in erythroid (K562) and nonerythroid (HeLa) cells. Two ERMAD isoforms are shown. There is a peak of RNAPII at the TSS in both K562 and HeLa cells for one isoform and a peak of RNAPII in K562 cells only for the alternate isoform, demonstrating cell type-specific binding of RNAPII. (E) IGB view of H3Me3K4 binding at the ERMAD locus in erythroid (K562) and nonerythroid (HeLa) cells. The ERMAD isoform with the erythroid cell-specific RNAPII binding has greater enrichment for H3Me3K4 in K562 cells than in HeLa cells, demonstrating cell type-specific enrichment of H3Me3K4. (F) IGB view of GATA-1 and NF-E2 binding at the ERMAD locus, demonstrating NF-E2 binding at the promoter, in intron 1, and at the 3' adjacent region and GATA-1 binding in intron 1 and at the 3' adjacent region of the ERMAD locus.

SPTB, ADD2, ANK1, EPB41, EPB49, ERMAD, ICAM4, and SLC4A1. RNAPII binding was not found in the region of the putative TSS (identified by the RefSeq Genes track in the UCSC genome browser) of the TMOD1 gene in K562 cells, but a peak of RNAPII enrichment was found ~ 20 kb 5' of the putative TSS. This corresponds to the TSS of "exon 0," an alternate transcript of TMOD1 (115).

Interestingly, several genes had intragenic peaks of RNAPII binding. For example, both the ANK1 and EPB41 loci had significant intragenic peaks of RNAPII binding that did not correlate with any known TSS (see Fig. S1 in the supplemental material). Both ANK1 and EPB41 are large complex loci which encode multiple isoforms with alternate 5' ends. There are two expressed sequence tags in the intragenic region of RNAPII binding in the ANK1 locus (GenBank accession no. BY800201 and DW421256) and one in the EPB41 locus (GenBank accession no. EH345598). The intragenic RNAPII peak in these genes likely represents either the first exon of an alternate 5' isoform or a separate gene embedded within the locus.

Previous studies have shown that over 10% of genes with detectable RNAPII/preinitiation complex binding at their promoter regions do not produce detectable transcripts, attributed to promoter stalling or pausing (54, 60, 71, 85). Transcripts of all 15 membrane protein genes were detected in K562 cell

mRNA; thus, this phenomenon was not observed with these genes.

In nonerythroid (HeLa) cells, RNAPII binding was detected at a single promoter/putative promoter/TSS in 7 of 15 genes and at multiple promoters/putative promoters/TSS of three genes, ANK1, ACTB, and EPB41. Five genes with undetectable or nearly undetectable mRNA levels in HeLa cells (Fig. 1) showed no RNAPII binding across their loci (SPTA, EPB42, EPB49, SPTB, and ADD2).

H3Me3K4 is both necessary for and a specific marker of transcriptionally active genes (72, 87, 88, 95). CHIP-chip technology using a custom-designed NimbleGen DNA microarray was employed to analyze H3Me3K4 binding in the 15 membrane protein genes in erythroid (K562) and nonerythroid (HeLa) cells. Microarray probes corresponding to the erythrocyte membrane protein genes spanned 3.35 Mb of the human genome and encompassed 11 different chromosomes (Table 1). Fourteen of 15 genes were enriched for H3Me3K4 at a putative promoter/promoter/TSS in K562 cells (Fig. 2; also see Fig. S1 in the supplemental material). Similarly to results for RNAPII occupancy, TMOD1 was not enriched at the putative TSS but did demonstrate a large peak of H3Me3K4 enrichment 20 kb 5' of the putative TSS, corresponding to the TSS of "exon 0." Two compact genes, β -actin and ICAM4, had H3K4me3 enrichment not only in the promoter region but also

TABLE 2. Regions of GATA-1 occupancy of major erythrocyte membrane protein genes in K562 chromatin, predicted by the Tamalpais peak calling algorithm

Gene	Region	Location ^a	Stringency level(s)
Ankryin-1	1	Intragenic (intron 1)	L2-L3
	2	Intragenic (intron 1)	L1
α -Spectrin	3	Intragenic (intron 3)	L4
α -Adducin	4	Intragenic (intron 1)	L4
γ -Adducin	5	Intragenic (intron 1)	L4
ERMAP	6	3' Adjacent	L2-L3
	7	3' Adjacent	L1
	8	Intragenic (intron 1)	L4
Protein 4.1R	9	5' Adjacent	L2-L3
	10	Intragenic (intron 1)	L2-L3
	11	Intragenic (intron 1)	L2-L3
	12	Intragenic (intron 1)	L2-L3
	13	Intragenic (intron 7)	L2-L3
	14	Intragenic (intron 3)	L4
Tropomodulin	15	Intragenic (intron 1-exon 2)	L4
	16	5' Intergenic	L1
	17	5' Intergenic	L4
Tropomyosin	18	Intragenic (intron 8)	L4
	19	Intragenic (intron 2)	L2-L3

^a Intragenic regions of GATA binding are within an exon or intron. Adjacent regions of GATA-1 binding are within 10 kb of the 5' or the 3' end of the gene. Intergenic regions of GATA-1 binding are greater than 10 kb from the 5' or 3' end of the gene.

throughout the body of the gene. In many cases, H3Me3K4 enrichment at the promoter/TSS was cell type specific, with some genes having dramatically lower or nondetectable binding of H3Me3K4 in HeLa cell chromatin (Fig. 2; also see Fig. S2 in the supplemental material).

Erythrocyte transcription factor binding in membrane protein genes. The genetic regulation of the erythrocyte membrane protein genes is complex, and the expression of these genes is influenced by many *cis*- and *trans*-acting factors. To gain insight into some of the *trans*-acting factors involved in the transcriptional regulation of these genes, regions of GATA-1 and NF-E2 binding were studied. GATA-1 is a zinc finger transcription factor that promotes erythroid cell, megakaryocyte, and mast cell development (70, 80, 81, 90, 93, 96, 106). Able to act as either an enhancer or a repressor, GATA-1 exerts its effects on target genes through a number of molecular mechanisms, including cofactor recruitment (5, 30, 33, 41) and higher-order chromatin organization (39, 45). NF-E2 is a heterodimer of an erythroid cell-specific 45-kDa subunit and a ubiquitously expressed 18-kDa subunit (1, 36). Like GATA-1, NF-E2 plays critical roles in erythropoiesis and megakaryocytogenesis (74) and influences higher-order chromatin structure (3, 27). Previous studies employing ChIP-chip technology to study GATA-1 binding have been limited to the region of human chromosome 11 containing the β -globin locus (34), the region of mouse chromosome 7 in and around the β -globin locus (10, 97), and the region of murine chromosome 6 in and around the GATA-2 locus (30). To date, p45 NF-E2 binding has not been studied using high throughput genomic strategies.

GATA-1 binding in major erythrocyte membrane genes. GATA-1 binding in the 15 erythrocyte membrane protein genes was studied. Tamalpais identified 19 regions of GATA-1 occupancy in K562 chromatin (Table 2 and Fig. 2; also see Fig.

S3 in the supplemental material), with predicted regions of GATA-1 occupancy averaging 539 bp in length (range, 300 to 1,200 bp). All 19 regions of GATA-1 occupancy in K562 chromatin were validated by use of ChIP-qPCR (Fig. 3A). K562 cells have been utilized as models of erythroid cells in numerous studies of gene structure and function. However, clonal variation in copy number produced by karyotypic abnormalities dictates that validation with primary erythroid cells be performed to confirm biologic relevance (49). The regions of GATA-1 occupancy identified for membrane protein genes in K562 cells were examined in chromatin from R3/R4 stage primary, cultured human erythroid cells by use of ChIP-qPCR (117). All 19 GATA-1 regions identified in K562 chromatin were also occupied in chromatin of R3/R4 primary erythroid cells (Fig. 3A).

GATA-1 binding was identified in 8 of 15 major erythrocyte membrane protein genes. There are several possible explanations why GATA-1 binding was not detected in the remaining seven genes. One possibility is that GATA-1 binding occurs at an earlier time point in erythrocyte differentiation (i.e., R1/R2) than our assay examined (R3/R4). Another possibility is that sites of low-level GATA-1 occupancy were not identified by the Tamalpais peak calling algorithm, which is very stringent in order to provide a low false-discovery rate. Due to the high stringency of the Tamalpais algorithm, it is likely, as reported by others utilizing similar genomic technologies (46), that not every region of GATA-1 binding in the major erythrocyte membrane genes was identified. Finally, it is possible that some of these erythrocyte genes exhibit GATA-1-independent expression.

H3Me3K4 has been associated with transcriptionally active chromatin, promoters, and enhancers. In the β -globin locus control region (LCR), disruption of GATA-1 binding to hypersensitive site 2 (HS2) leads to decreased levels of H3Me3K4, implying that GATA-1 is involved in the establishment of active histone methylation patterns (11, 84). The 19 regions of GATA-1 occupancy identified in the membrane genes were tested for the presence of H3me3K4 binding in erythroid (K562) and nonerythroid (HeLa) cells. In K562 chromatin, 16 of 19 regions of GATA-1 occupancy (Fig. 3B) had significant enrichment for H3Me3K4. In nonerythroid cell (HeLa) chromatin, only 6 of 19 regions had significant enrichment for H3Me3K4. These six regions of GATA-1 occupancy were present in four genes: ADD3, ERMAP, EPB41, and TMOD1. mRNA transcripts of all four genes could be detected in nonerythroid (HeLa) cells but at significantly lower levels than in erythroid (K562) cells. Correspondingly, in these regions, the degree of H3Me3K4 enrichment in nonerythroid cell (HeLa) chromatin (4.05 ± 0.45 relative units) was significantly lower than that in erythroid cell (K562) chromatin (9.55 ± 1.90 relative units).

Factor cooccupancy at sites of GATA-1 binding. GATA-1 interacts with numerous proteins and protein complexes during hematopoiesis, including but not limited to FOG-1, SP/X-KLF proteins, the NuRD complex, SCL complexes, RNAPII, RUNX1, CBP/P300, Cbf- β , Fli-1, PU.1, the MeCP1 complex, the ACF/WRF complex, and the Mediator complex. FOG-1, a cofactor usually found at sites of GATA-1 occupancy, is a transcriptional regulator that plays a critical role in erythroid cell and megakaryocyte differentiation. It is found at nearly all

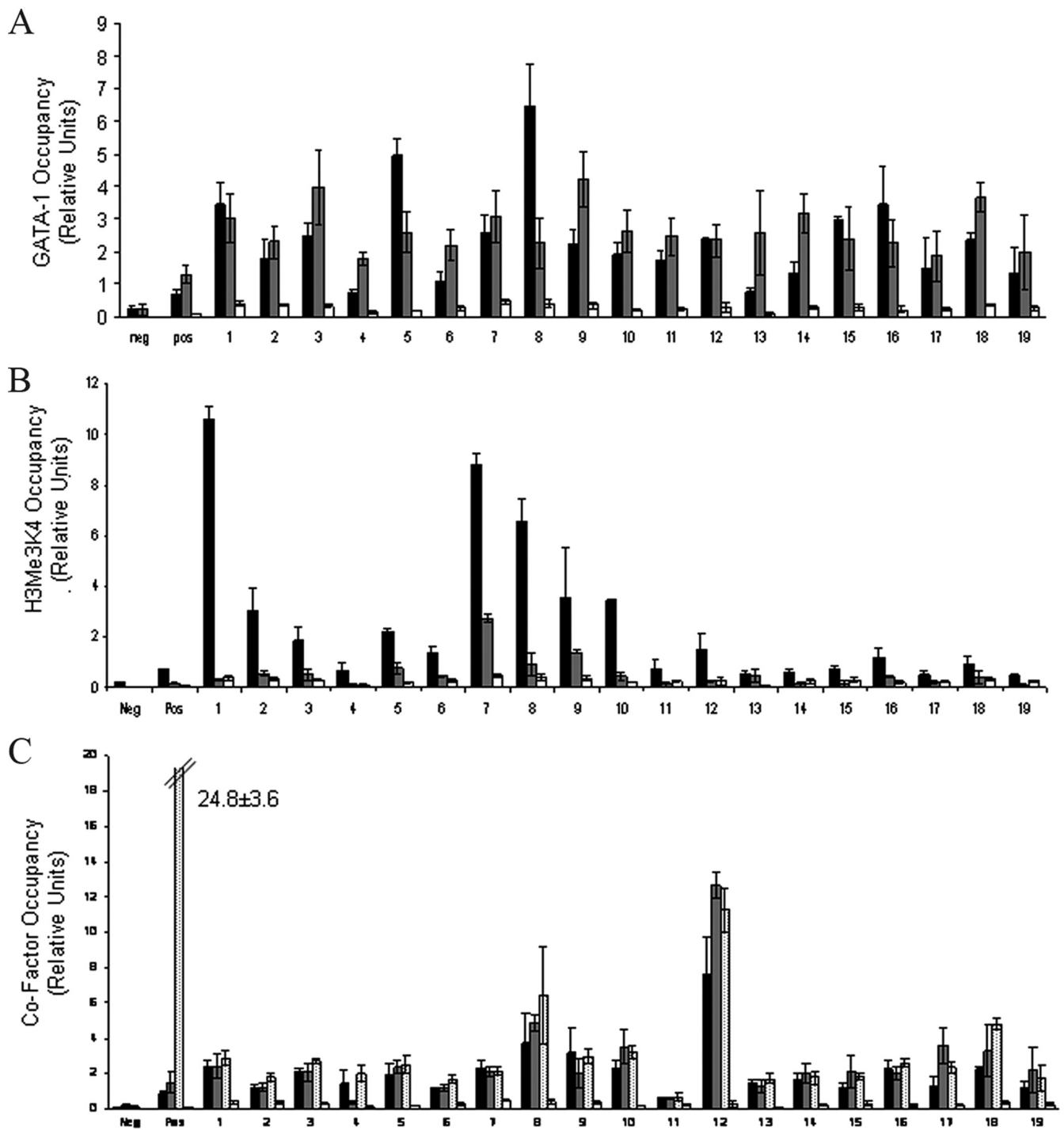


FIG. 3. Quantitative ChIP analyses of regions of GATA-1 occupancy in erythrocyte membrane protein genes. Data are presented as enrichment (*n*-fold) relative to total input. A region from HS2 of the β -globin LCR was included as a positive control (Pos), and a region from the cerebellum-specific GABRA6 gene was included as a negative control (Neg). (A) Validation of regions of GATA-1 occupancy in the 15 erythrocyte membrane protein genes. ChIP analyses of the 19 regions of GATA-1 occupancy in K562 and primary erythroid cell chromatin predicted by the Tamalpais peak calling algorithm were performed using antibodies against GATA-1. Black bars represent GATA1-precipitated K562 cell chromatin. Gray bars represent GATA1-precipitated primary erythroid cell chromatin. White bars represent IgG-precipitated K562 chromatin. (B) H3Me3K4 occupancy in regions of GATA-1 binding in K562 and HeLa chromatin. ChIP analyses of H3Me3K4 occupancy in regions of GATA-1 binding in K562 and HeLa cell chromatin were performed using an antibody against H3Me3K4. Black bars represent H3Me3K4-immunoprecipitated K562 chromatin. Gray bars represent H3Me3K4-immunoprecipitated HeLa chromatin. White bars represent IgG-precipitated K562 chromatin. (C) Factor cooccupancy at sites of GATA-1 binding. ChIP analyses of the 19 regions of GATA-1 occupancy in K562 cell chromatin were performed using antibodies against FOG-1, SCL, and MTA-2. FOG-1, SCL, and MTA-2 binding are present in all regions of GATA-1 occupancy. Black bars represent FOG1-precipitated K562 cell chromatin. Gray bars represent MTA-2-precipitated K562 chromatin. Speckled bars represent SCL-precipitated K562 chromatin. White bars represent IgG-precipitated K562 chromatin.

sites of GATA-1 binding in erythroid cells, where it acts via heterodimer formation with GATA-1, conferring either activating or repressive functions (14, 33, 40, 41, 57, 76). Not surprisingly, FOG-1 binding was demonstrated at all 19 regions of GATA-1 occupancy in erythrocyte membrane protein genes (Fig. 3C).

SCL is a basic helix-loop-helix transcription factor essential for hematopoiesis (26, 75, 89, 102). It is frequently found at activating GATA-1 sites and is depleted at repressive GATA-1 sites (97). In some but not all cases, a tandem E box, the consensus binding motif for SCL, and a GATA-1 site are found within 9 to 12 bp or even up to 27 bp apart at sites of SCL GATA-1 cooccupancy (17, 56). MTA-2 is a member of the NuRD (nucleosome remodeling and deacetylase) complex (118). It is recruited by FOG-1 to sites of GATA-1 binding and participates in GATA-1-mediated repression (33). Sites of GATA-1 occupancy in membrane protein genes were tested for the presence of SCL-1 and MTA-2 binding. Similar to results from a previous report, all regions of GATA-1 occupancy also demonstrated occupancy of SCL (Fig. 3C) (97). Interestingly, MTA-2, although a member of the NuRD complex described to be recruited by FOG-1 to GATA-1-repressed genes, was found at all sites of GATA-1 occupancy in these erythroid cell-expressed membrane protein genes. The function of the NuRD complex at these GATA-1 sites is unclear. A similar observation has recently been described (cited as unpublished data in reference 97). Of note, GATA-1 site number 11 had lower levels of enrichment for FOG-1, SCL, and MTA-2 than the other sites of GATA-1 occupancy. Although there was quantitatively less enrichment at site 11, there was statistically significant enrichment for all three factors ($P = 0.010$, $P = 0.001$, and $P = 0.033$, respectively).

Prediction of cis-regulatory modules (CRMs) in regions of GATA-1 binding. ChIP-chip analyses identified 19 regions of GATA-1 binding in the genes encoding the 15 major membrane protein genes. These genes and their flanking DNA span 3.35 Mb of the human genome. A large number of potential GATA-1 DNA binding motifs were present in this 3.35 Mb of DNA. The minimal GATA-1 binding site, GATA, was found 15,875 times, and the GATA-1 consensus site, WGATAR, was found 6,269 times.

Functional regions of GATA-1 binding are expected to be more conserved than nonfunctional sites (10, 31, 48). Recent studies have used PhastCons analyses and RP scores to predict whether or not a region of DNA contains a CRM (10, 46, 48). PhastCons uses a hidden Markov model method on aligned genomic sequences to estimate a probability that any nucleotide is conserved (92). The UCSC genome browser 44-way placental-mammal PhastCons scores were determined for each of the 19 regions of GATA-1 binding. Thirteen of 19 (68%) GATA-1 binding regions had PhastCons scores of >0.8 , suggesting that they contain a functional CRM (see Table S4 in the supplemental material).

An alternative way to predict the presence of CRMs is the RP score, which evaluates whether regions of DNA sequence have patterns more similar to those of regulatory elements or neutral DNA. Positive RP scores (RP scores of >0) (48) indicate conserved regions which contain a functional CRM. Fourteen of 19 regions (73%) of GATA-1 binding in erythrocyte

membrane genes had positive RP scores (see Table S4 in the supplemental material).

Regions of protein occupancy identified by ChIP assays do not always contain consensus DNA-protein binding motifs. Genome-wide ChIP-based studies of various transcription factors and other regulatory proteins have highlighted this observation. In some cases, corresponding DNA-protein binding motifs were identified infrequently for some proteins, e.g., E2F1 (4) and OCT4 (4, 38), and very frequently for others, e.g., CTCF (46) and SCL (107). In erythrocyte membrane protein genes, the minimal consensus GATA-1 binding motif, GATA, was found in 18 of 19 regions, with only a single region, located in the promoter of the protein 4.1R gene, lacking any GATA-1 binding motifs. The canonical GATA-1 binding motif, (A/T)GATA(A/G), was present in 15 regions of GATA-1 occupancy.

ChIP assays identify regions where DNA and proteins interact in vivo, but they lack the resolution to define the precise site of DNA-protein interaction. As the minimal GATA-1 binding motif, GATA, is common, most regions of GATA-1 occupancy predicted by Tamalpais had several potential sites of GATA-1 binding. The in vitro ability of each of these potential binding sites to bind GATA-1 was analyzed using electrophoretic mobility shift assays (EMSA). With GATA used as the minimal GATA-1 binding sequence, there were 67 potential GATA-1 binding motifs in the 18 regions of GATA-1 occupancy. Double-stranded, 20-bp oligonucleotide probes corresponding to each of the 67 potential GATA-1 binding motifs were created (see Table S3 in the supplemental material) and used for EMSA with K562 cell nuclear extracts. Nineteen GATA-1 motifs with in vitro GATA-1 binding ability were identified in 10 regions of GATA-1 occupancy (see Table S5 in the supplemental material). Four regions of GATA-1 occupancy contained multiple GATA-1 binding motifs which were able to bind GATA-1 by EMSA: intron 3 of the α -spectrin gene (four positive binding sites), the 3' flanking adjacent region of the ERMAP gene (two positive binding sites), intron 1 of the protein 4.1R gene (four positive binding sites), and intron 1/exon 2 of the protein 4.1R gene (two positive binding sites).

There were nine regions of GATA-1 occupancy that did not have any sites of GATA-1 binding in vitro. By a comparison using Fischer's exact test, no significant differences between regions of GATA-1 occupancy with positive GATA-1 binding in vitro and regions without GATA-1 binding in vitro were identified. This included no significant differences between the numbers of regions with PhastCons scores of <0.8 ($P = 0.6$), the locations of binding sites (e.g., intron, 5' flanking region, 3' flanking region, etc.) ($P = 0.11$), or the numbers of regions with a Tamalpais stringency level of L4 ($P = 0.99$) (see Table S5 in the supplemental material). In addition, PhastCons scores were calculated for each of the 67 potential GATA-1 binding motifs. These did not differ significantly between EMSA-positive and EMSA-negative motifs ($P = 0.46$, t test).

NF-E2 binding in erythrocyte membrane protein genes. Tamalpais identified 18 regions of NF-E2 occupancy in K562 chromatin of the 15 membrane protein genes (Table 3 and Fig. 2; also see Fig. S4 in the supplemental material), with predicted regions of NF-E2 binding averaging 672 bp (range, 250 to 1,200 bp). Seventeen of 18 regions of NF-E2 occupancy in

TABLE 3. Regions of NF-E2 occupancy of major erythrocyte membrane protein genes in K562 chromatin, predicted by the Tamalpais peak calling algorithm

Gene	Region	Location ^a	Stringency level(s)
Ankyrin-1	1	Intragenic (intron 1)	L4
α -Spectrin	2	Intragenic (promoter/intron 1)	L4
β -Adducin	3	5' Adjacent	L1
	4	Intragenic (intron 1)	L2–L3
	5	Intragenic (intron 6)	L2–L3
α -Adducin	6	Intragenic (intron 1)	L4
ERMAP	7	5' Adjacent	L4
	8	3' Adjacent	L4
	9	Intragenic (intron 1)	L1
Protein 4.1R	10	Intragenic (intron 1)	L4
	11	Intragenic (intron 1)	L4
	12	5' Adjacent	L1
	13	5' Adjacent	L2–L3
Protein 4.2	14	5' Adjacent	L1
Dematin	15	Intragenic (intron 1)	L4
Tropomodulin	16	5' Intergenic	L4
	17	5' Intergenic	L1
Tropomyosin	18	Intragenic (intron 1)	L4

^a Intragenic regions of NF-E2 binding are within an exon or intron. Adjacent regions of NF-E2 binding are within 10 kb of the 5' or the 3' end of the gene. Intergenic regions of NF-E2 binding are greater than 10 kb from the 5' or 3' end of the gene.

K562 chromatin were validated using ChIP-qPCR (Fig. 4A). All validated NF-E2 sites were also occupied in chromatin from R3/R4 primary, cultured erythroid cells (Fig. 4A).

Similarly to results for GATA-1, disruption of NF-E2 binding to HS2 in the β -globin LCR leads to decreased levels of H3Me3K4, implying that NF-E2 may also be involved in the establishment of active histone methylation patterns (11, 84). The 18 sites of NF-E2 occupancy identified in erythrocyte membrane genes were tested for the presence of H3Me3K4 in both erythroid (K562) and nonerythroid (HeLa) cells. As with regions of GATA-1 occupancy, regions of NF-E2 binding in erythroid cells were associated with enrichment for H3Me3K4. Sixteen of 19 NF-E2 sites (Fig. 4B) demonstrated significant enrichment for H3Me3K4 in erythroid cell (K562) chromatin, with only 9 of 18 regions demonstrating enrichment (and at lower levels) for H3Me3K4 in nonerythroid cell (HeLa) chromatin.

Prediction of CRMs in regions of NF-E2 binding. The UCSC genome browser 44-way placental-mammal PhastCons scores were calculated for each of the 18 regions of NF-E2 binding. Seventy-two percent (13/18) of the regions had PhastCons scores of >0.8, indicating that they are likely to contain a functional CRM (see Table S6 in the supplemental material). In addition, 83% of the NF-E2 regions (15/18) had positive RP scores.

GATA-1 and NF-E2 cooccupancy. GATA-1 and NF-E2 cooccupancy has been demonstrated to occur at several sites, including the HSs of the β -globin LCR and the flanking regions of the α -globin genes (29, 82, 94). Regions of GATA-1 occupancy identified in membrane protein genes were examined for NF-E2 cooccupancy. Five sites of GATA-1–NF-E2 cooccupancy were identified: one site in intron 1 of the α -adducin gene, one site in intron 2 of the tropomyosin gene, one site in the 5' adjacent region of the

protein 4.1R gene, and two sites in the 5' intergenic region of the tropomyosin gene (Table 4).

To determine if GATA-1–NF-E2 cooccupancy is a more widespread phenomenon in erythroid cell-expressed genes, GATA-1 and NF-E2 cooccupancy in other erythroid cell-expressed genes present on the high-density NimbleGen microarray was examined. The array contained a total of 100 erythroid cell-expressed genes, the 15 major membrane genes as well as 85 other erythroid cell-expressed genes. The probes corresponding to these erythroid cell-expressed genes span 12.80 Mb of the human genome, encompassing 20 different chromosomes. In the aggregate 100 erythroid cell-expressed genes, Tamalpais identified 74 regions of GATA-1 occupancy and 78 regions of NF-E2 occupancy. The majority of the GATA-1 (37/74) and the NF-E2 (37/78) binding regions were located in introns, with 21 of 37 GATA-1 sites and 25 of 37 NF-E2 sites located in intron 1.

Eighteen regions bound both GATA-1 and NF-E2 (Fig. 5A and Table 4), and all of these were validated using ChIP-qPCR with K562 chromatin (Fig. 5A). Ten of the 18 regions were located in introns. Fifteen of the 18 regions demonstrated significant occupancy of H3Me3K4 in K562 chromatin (Fig. 5B), while only 6 of the 18 regions demonstrated any detectable H3Me3K4 occupancy in HeLa chromatin (data not shown). Seventeen regions contained consensus GATA-1 binding sites, while 15 regions contained potential NF-E2 binding sites. A region of GATA-1 and NF-E2 cooccupancy in the 5' adjacent region of the protein 4.1R gene did not contain consensus binding sites for either GATA-1 or NF-E2. Half of the regions with GATA-1–NF-E2 cooccupancy had single consensus GATA-1 and NF-E2 binding sites separated by <50 bp in the corresponding genomic DNA. Thus, these sites differ slightly from sites of GATA-1–NF-E2 cooccupancy in the HS cores of the β -globin LCR, which contain consensus NF-E2 binding sequences ~50 bp from tandem inverted GATA-1 consensus sequences (28, 69, 94), possibly indicating that the mechanism(s) of cooperation at these sites differs from that in the β -globin LCR.

FOG-1, SCL, and MTA-2 occupancy in regions of GATA-1 and NF-E2 cooccupancy. In hematopoietic cells, in addition to GATA-1, SCL and MTA-2 associate with other regulatory proteins. The identification of FOG-1 or SCL binding at sites of GATA-1–NF-E2 cooccupancy would not be unexpected (97). Not surprisingly, in erythroid cell-expressed genes (Fig. 5C), all 18 regions of GATA-1–NF-E2 cooccupancy also bound FOG-1, MTA-2, and SCL. In half of these regions, consensus NF-E2 binding motifs were located 34 to 90 bp away from GATA-1–E-box motifs.

SCL and MTA-2 occupancy in regions of GATA-1-independent NF-E2 occupancy. Little is known about SCL binding at GATA-1-independent regions of NF-E2 occupancy. The 13 GATA-1-independent regions of NF-E2 binding in erythrocyte membrane protein genes were tested for the presence of SCL. All GATA-1-independent regions of NF-E2 occupancy also showed enrichment for SCL (Fig. 5C), demonstrating that SCL can associate with NF-E2 in the absence of GATA-1. Interestingly, the regions which bound both NF-E2 and SCL were also positive for MTA-2 (Fig. 5C), providing another, but different, example of cooccupancy of these proteins in erythroid cell-expressed genes and suggesting that cooperation between

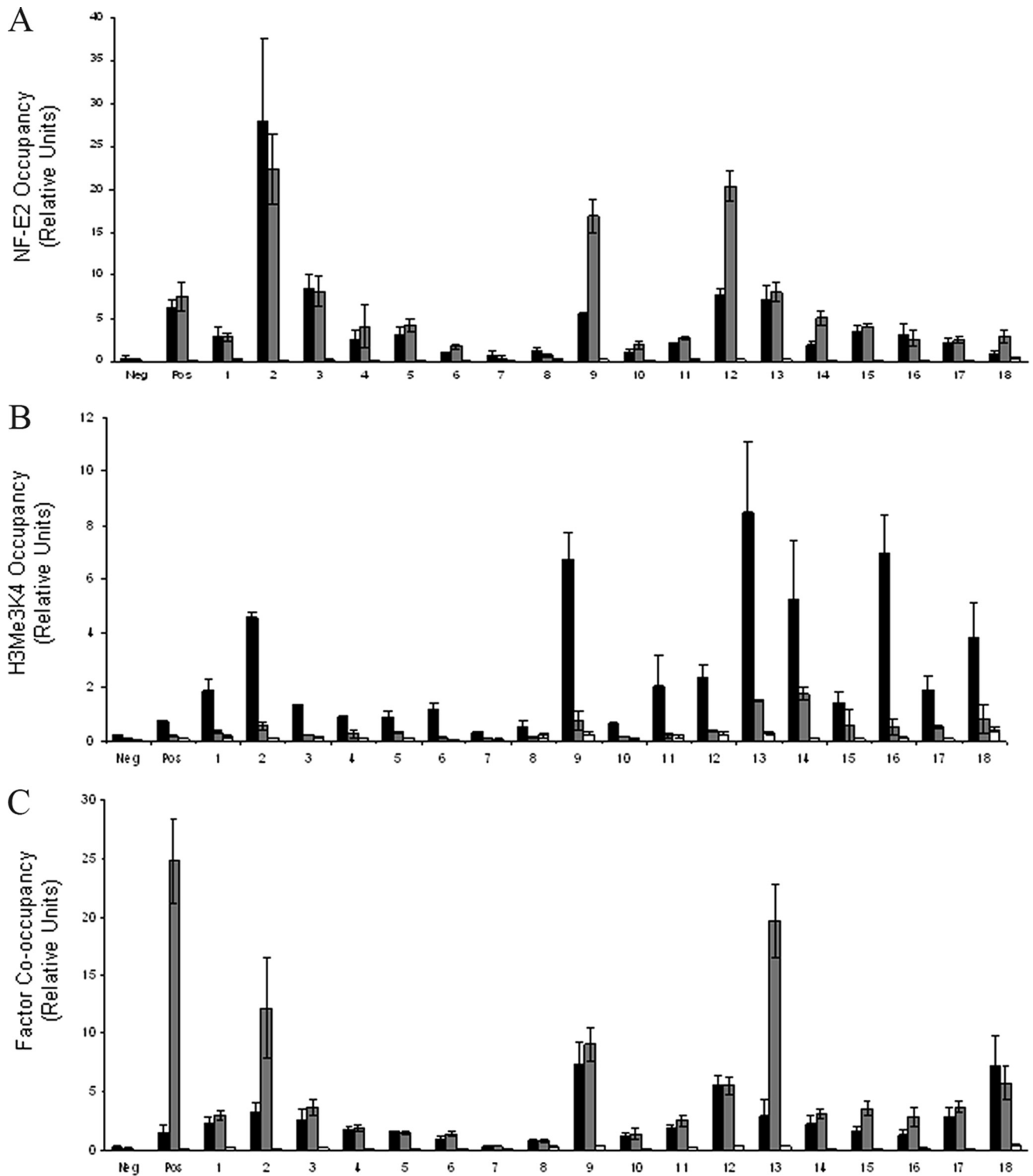


FIG. 4. Quantitative ChIP analyses of regions of NF-E2 occupancy in the erythrocyte membrane protein genes. Data are presented as enrichment (n -fold) relative to total input. A region from HS2 of the β -globin LCR was included as a positive control (Pos), and a region from the cerebellum-specific GABRA6 gene was included as a negative control (Neg). (A) Validation of regions of NF-E2 occupancy in the 15 erythrocyte membrane protein genes. ChIP analyses of the 18 regions of NF-E2 occupancy in K562 and primary erythroid cell chromatin predicted by the Tamalpais peak calling algorithm were performed using antibodies against NF-E2. Black bars represent NF-E2-precipitated K562 cell chromatin. Gray bars represent NF-E2-precipitated primary erythroid cell chromatin. White bars represent IgG-precipitated K562 chromatin. (B) H3Me3K4 occupancy in regions of NF-E2 binding in K562 and HeLa chromatin. ChIP analyses of H3Me3K4 occupancy in regions of NF-E2 binding in K562 and HeLa chromatin were performed using an antibody against H3Me3K4. Black bars represent H3Me3K4-immunoprecipitated K562 chromatin. Gray bars represent H3Me3K4-immunoprecipitated HeLa chromatin. White bars represent IgG-precipitated K562 chromatin. (C) Cofactor binding in regions of NF-E2 occupancy. ChIP analyses of the 19 regions of NF-E2 occupancy in K562 cell chromatin were performed using antibodies against SCL and MTA-2. SCL and MTA-2 binding were present in all regions of NF-E2 occupancy. Black bars represent MTA-2-precipitated K562 chromatin. Gray bars represent SCL-precipitated K562 chromatin. White bars represent IgG-precipitated K562 chromatin.

TABLE 4. Regions of GATA-1 and NF-E2 cooccupancy in 100 erythroid cell-expressed genes, predicted by the Tamalpais peak calling algorithm and validated with K562 cell chromatin

Gene	Region	Location ^a
KCC3	1	Intragenic (intron 12)
E2F3	2	Intragenic (intron 1)
	3	Intragenic (intron 1)
	4	Intragenic (intron 1)
	5	Intragenic (intron 1)
RPS24	6	Intragenic (intron 5–intron 7)
RBM15	7	3' Adjacent
C-FOS	8	5' Adjacent
C-JUN	9	3' Intergenic
p45 NF-E2	10	5' Intergenic
NRF2	11	Intragenic (intron 1)
RUNX1	12	Intragenic (intron 1)
Delta-like 3	13	5' Adjacent
α-Adducin	14	Intragenic (intron 1)
Protein 4.1R	15	5' Adjacent
Tropomodulin	16	5' Intergenic
	17	5' Intergenic
Tropomyosin	18	Intragenic (intron 2)

^a Intragenic regions of GATA binding are within an exon or intron. Adjacent regions of GATA-1 binding are within 10 kb of the 5' or 3' end of the gene. Intergenic regions of GATA-1 binding are greater than 10 kb from the 5' or 3' end of the gene.

them may be an important factor in erythrocyte gene regulation.

DISCUSSION

Previous studies attempting to identify critical regulatory elements of membrane protein genes have largely been limited to the study of core promoters, relying primarily on *in vitro* assays, such as reporter gene assays and EMSA (6, 21, 22, 51, 53). ChIP techniques have rarely been applied to the study of these genes (108, 111). When ChIP has been employed, the data have been of limited utility for a number of reasons, including the limited regions analyzed, the time and cost required to complete the studies, and investigator bias in choosing regions of study. High throughput ChIP-based technologies overcome many of these problems, allowing unbiased analyses of large contiguous genomic regions and using cost-effective commercially available platforms, as well as allowing comparisons and classifications of numerous genes in a single experiment.

This is the first report to utilize ChIP-chip to study in an unbiased manner both GATA-1 and NF-E2 binding in erythroid cell-expressed genes in numerous loci spread throughout the human genome. This approach has revealed that the majorities of regions of GATA-1 and NF-E2 occupancy are not at core erythrocyte promoters, as expected, but are located apart from the core promoter region, primarily in introns, frequently in intron 1. GATA-1 and NF-E2 *in vivo* binding outside core promoter regions has been described to occur in erythrocyte enhancers, most notably the β-globin LCR, but extensive non-promoter-related binding throughout numerous genomic loci was unexpected. High throughput, ChIP-based studies of other transcription factors have also demonstrated that the majority of binding sites of some, but not all, regula-

tory proteins may not necessarily be located at promoters or CpG islands. Together, these data provide additional evidence for the growing body of data supporting the critical role of long-range mechanisms, such as chromatin looping, in gene regulation (9, 39, 42, 50, 59, 73, 79, 98).

Genome-wide ChIP-based studies provide additional data supporting the role of long-range interactions in gene regulation. Several reports utilizing high throughput ChIP-based techniques have found multiple regions of factor occupancy *in vivo* that do not contain a consensus binding site in the corresponding DNA (4, 32, 38, 105, 110). Similarly, not all regions of GATA-1 occupancy in membrane protein genes contain a consensus GATA-1 binding site in the corresponding DNA. GATA-1 has been shown to play a critical role in long-range gene interactions at the *c-kit* and β-globin loci via formation of chromatin loops (39, 50, 101), a possible mechanism of action in regions of GATA-1 and NF-E2 binding that lack consensus GATA-1 or NF-E2 DNA binding sites.

Many different strategies have been employed to precisely identify and/or predict which GATA-1 consensus binding sites in genomic DNA interact with the GATA-1 protein *in vivo*. These have included various *in silico* and experimental techniques which have demonstrated that GATA-1 binding sites that regulate gene expression during erythropoiesis are under strong selection constraint (10, 104). As other reports have shown, regions of DNA in membrane protein genes demonstrating GATA-1 binding *in vivo* were more likely to demonstrate evolutionary conservation across species. However, even though these regions were identified as having conservation scores predictive of a *cis*-regulatory element, attempts to refine sites of GATA-1 binding by using *in vitro* binding with EMSA were unsuccessful, with no correlation between EMSA binding and PhastCons score. Numerous factors contribute to DNA-protein binding and subsequent erythrocyte gene expression, including *cis* sequences, concentration and stability of regulatory proteins, and chromatin architecture (7, 10, 35, 37, 41, 55, 58, 61, 64, 99). It is likely that a complex combination of factors regulates GATA-1 binding to its cognate DNA binding site.

As noted above, several recent studies have detailed binding partners at sites of GATA-1 and NF-E2 binding, as well as the composition of DNA-protein complexes. The GATA-1-associated proteins identified for membrane protein genes, including FOG-1, SCL, and MTA-2, have been described previously. In addition to heterodimerizing with small Maf proteins, p45 NF-E2 has been shown to interact with various proteins, such as MLL2, WWP1, and MCRS2 (2, 15, 67, 107, 109). With a few exceptions, proteins and multiprotein regulatory complexes interacting and associating with NF-E2 have not been characterized to the extent that GATA-1 has (68). Finding SCL and MTA-2 at all sites of NF-E2 binding implies that these transcription factors may play a broader role in the regulation of erythroid cell-expressed genes than may previously have been appreciated.

Importantly, these studies revealed a common erythroid cell type-specific chromatin signature located throughout the genomic loci of many erythroid cell-expressed genes. This mark includes H3Me3K4, with cooccupancy of GATA-1, NF-E2, FOG-1, SCL, and MTA-2 proteins, present at approximately a quarter of the GATA-1 binding sites. In the genomic DNA underlying half of these sites are single GATA-1-E-box consensus motifs

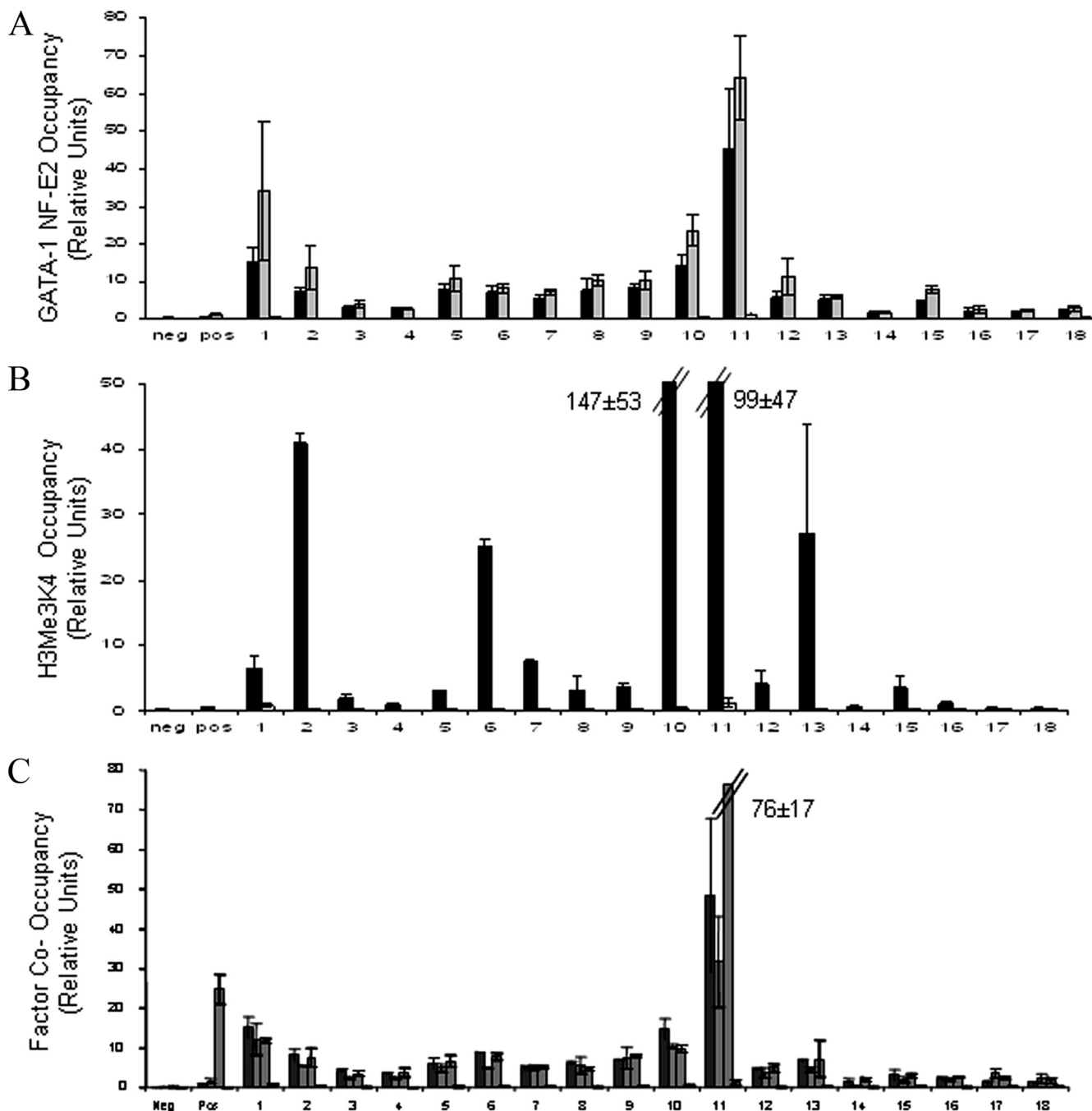


FIG. 5. Quantitative ChIP analyses of regions of GATA-1 NF-E2 cooccupancy. Data are presented as enrichment (*n*-fold) relative to total input. (A) Validation of regions of GATA-1 and NF-E2 cooccupancy. ChIP analyses of the 18 regions of GATA-1 and NF-E2 cooccupancy in K562 and primary erythroid cell chromatin predicted by the Tamalpais peak calling algorithm were performed using antibodies against GATA-1 and NF-E2. Black bars represent GATA1-precipitated chromatin. Gray bars represent NF-E2-precipitated chromatin. White bars represent IgG-precipitated chromatin. A region from HS2 of the β -globin LCR was included as a positive control (Pos), and a region from the cerebellum-specific GABRA6 gene was included as a negative control (Neg). (B) H3Me3K4 occupancy in regions of GATA-1 NF-E2 cooccupancy in K562 chromatin. ChIP analyses of H3Me3K4 occupancy in regions of GATA-1 binding in K562 chromatin were performed using an antibody against H3Me3K4. Black bars represent H3Me3K4-immunoprecipitated K562 chromatin. White bars represent IgG-precipitated K562 chromatin. (C) Factor co-occupancy at sites of GATA-1–NF-E2 binding. ChIP analyses of the 18 regions of GATA-1 occupancy in K562 cell chromatin were performed using antibodies against FOG-1, SCL, and MTA-2. FOG-1, SCL, and MTA-2 binding are present in all regions of GATA-1 occupancy. Black bars represent FOG1-precipitated K562 cell chromatin. Gray bars represent MTA-2-precipitated K562 chromatin. Dark gray bars represent SCL-precipitated K562 chromatin. White bars represent IgG-precipitated K562 chromatin.

separated from NF-E2 consensus motifs by 34 to 90 bp. An important goal in understanding erythrocyte gene regulation would be to have a cell type-specific topographic map of chromatin architecture with the interacting regulatory proteins, as well as the primary sequence and epigenetic configuration of the associated genomic DNA.

Disorders of erythrocyte shape comprise an important group of inherited hemolytic anemias. These disorders include the hereditary spherocytosis, elliptocytosis, and pyropoikilocytosis syndromes, which are often associated with qualitative and quantitative abnormalities of major erythrocyte membrane proteins. In many cases, causative mutations have been identified in the genes encoding these proteins. However, in as many as 25% of cases, despite specific protein deficiency and/or genetic linkage, the causative mutation is not identified, even after nucleotide sequence analysis of the coding exons, the immediate flanking intronic sequences, and the promoter regions (103). The high throughput genomic strategies employed in this study identify numerous excellent candidate regions for mutations associated with membrane-linked hemolytic anemia.

ACKNOWLEDGMENTS

This work was supported in part by grants K12HD000850, P30DK072442, HL65448, and DK62039 from the National Institutes of Health.

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.
- Anguita, E., J. Hughes, C. Heyworth, G. A. Blobel, W. G. Wood, and D. R. Higgs. 2004. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J.* **23**:2841–2852.
- Armstrong, J. A., and B. M. Emerson. 1996. NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 in vitro. *Mol. Cell. Biol.* **16**:5634–5644.
- Bieda, M., X. Xu, M. A. Singer, R. Green, and P. J. Farnham. 2006. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* **16**:595–605.
- Blobel, G. A., T. Nakajima, R. Eckner, M. Montminy, and S. H. Orkin. 1998. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc. Natl. Acad. Sci. USA* **95**:2061–2066.
- Boulanger, L., D. E. Sabatino, E. Y. Wong, A. P. Cline, L. J. Garrett, M. Garbarz, D. Dhermy, D. M. Bodine, and P. G. Gallagher. 2002. Erythroid expression of the human alpha-spectrin gene promoter is mediated by GATA-1- and NF-E2-binding proteins. *J. Biol. Chem.* **277**:41563–41570.
- Bresnick, E. H., K. D. Johnson, S. I. Kim, and H. Im. 2006. Establishment and regulation of chromatin domains: mechanistic insights from studies of hemoglobin synthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **81**:435–471.
- Brodsky, A. S., C. A. Meyer, I. A. Swinburne, G. Hall, B. J. Keenan, X. S. Liu, E. A. Fox, and P. A. Silver. 2005. Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6**:R64.
- Carter, D., L. Chakalova, C. S. Osborne, Y. F. Dai, and P. Fraser. 2002. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* **32**:623–626.
- Cheng, Y., D. C. King, L. C. Dore, X. Zhang, Y. Zhou, Y. Zhang, C. Dorman, D. Abebe, S. A. Kumar, F. Chiaromonte, W. Miller, R. D. Green, M. J. Weiss, and R. C. Hardison. 2008. Transcriptional enhancement by GATA1-occupied DNA segments is strongly associated with evolutionary constraint on the binding site motif. *Genome Res.* **18**:1896–1905.
- Cho, Y., S. H. Song, J. J. Lee, N. Choi, C. G. Kim, A. Dean, and A. Kim. 2008. The role of transcriptional activator GATA-1 at human beta-globin HS2. *Nucleic Acids Res.* **36**:4521–4528.
- Chu, X., D. Thompson, L. J. Yee, and L. A. Sung. 2000. Genomic organization of mouse and human erythrocyte tropomodulin genes encoding the pointed end capping protein for the actin filaments. *Gene* **256**:271–281.
- Costessi, L., G. Devescovi, F. E. Baralle, and A. F. Muro. 2006. Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. *Nucleic Acids Res.* **34**:243–253.
- Crispino, J. D., M. B. Lodish, J. P. MacKay, and S. H. Orkin. 1999. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol. Cell* **3**:219–228.
- Demers, C., C. P. Chaturvedi, J. A. Ranish, G. Juban, P. Lai, F. Morle, R. Aebersold, F. J. Dilworth, M. Groudine, and M. Brand. 2007. Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol. Cell* **27**:573–584.
- Elnitski, L., R. C. Hardison, J. Li, S. Yang, D. Kolbe, P. Eswara, M. J. O'Connor, S. Schwartz, W. Miller, and F. Chiaromonte. 2003. Distinguishing regulatory DNA from neutral sites. *Genome Res.* **13**:64–72.
- Elnitski, L., W. Miller, and R. Hardison. 1997. Conserved E boxes function as part of the enhancer in hypersensitive site 2 of the beta-globin locus control region. Role of basic helix-loop-helix proteins. *J. Biol. Chem.* **272**:369–378.
- Gallagher, P. G. 2003. Regulation of erythrocyte membrane protein gene expression. *Curr. Opin. Hematol.* **10**:115–122.
- Gallagher, P. G., and B. G. Forget. 1998. An alternate promoter directs expression of a truncated, muscle-specific isoform of the human ankyrin 1 gene. *J. Biol. Chem.* **273**:1339–1348.
- Gallagher, P. G., D. G. Nilson, L. A. Steiner, Y. D. Maksimova, J. Y. Lin, and D. M. Bodine. 2009. An insulator with barrier-element activity promotes alpha-spectrin gene expression in erythroid cells. *Blood* **113**:1547–1554.
- Gallagher, P. G., M. Romana, W. T. Tse, S. E. Lux, and B. G. Forget. 2000. The human ankyrin-1 gene is selectively transcribed in erythroid cell lines despite the presence of a housekeeping-like promoter. *Blood* **96**:1136–1143.
- Gallagher, P. G., D. E. Sabatino, M. Romana, A. P. Cline, L. J. Garrett, D. M. Bodine, and B. G. Forget. 1999. A human beta-spectrin gene promoter directs high level expression in erythroid but not muscle or neural cells. *J. Biol. Chem.* **274**:6062–6073.
- Gallagher, P. G., E. Wong, and C. Wong. 1998. A novel isoform of ankyrin 1 (ANK1) expressed in brain, heart and skeletal muscle is directed by an alternate promoter. *Blood* **92**:300a.
- Gascard, P., G. Lee, L. Coulombel, I. Auffray, M. Lum, M. Parra, J. G. Conboy, N. Mohandas, and J. A. Chasis. 1998. Characterization of multiple isoforms of protein 4.1R expressed during erythroid terminal differentiation. *Blood* **92**:4404–4414.
- Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**:1451–1455.
- Goardon, N., J. A. Lambert, P. Rodriguez, P. Nissaire, S. Herblot, P. Thibault, D. Dumenil, J. Strouboulis, P. H. Romeo, and T. Hoang. 2006. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J.* **25**:357–366.
- Gong, Q. H., J. C. McDowell, and A. Dean. 1996. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region. *Mol. Cell. Biol.* **16**:6055–6064.
- Goodwin, A. J., J. M. McInerney, M. A. Glander, O. Pomerantz, and C. H. Lowrey. 2001. In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1. *J. Biol. Chem.* **276**:26883–26892.
- Gourdon, G., F. Morle, J. Roche, N. Tourneur, V. Joulain, and J. Godet. 1992. Identification of GATA-1 and NF-E2 binding sites in the flanking regions of the human alpha-globin genes. *Acta Haematol.* **87**:136–144.
- Grass, J. A., H. Jing, S. I. Kim, M. L. Martowicz, S. Pal, G. A. Blobel, and E. H. Bresnick. 2006. Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. *Mol. Cell. Biol.* **26**:7056–7067.
- Hardison, R. C. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**:369–372.
- Hollenhorst, P. C., A. A. Shah, C. Hopkins, and B. J. Graves. 2007. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev.* **21**:1882–1894.
- Hong, W., M. Nakazawa, Y. Y. Chen, R. Kori, C. R. Vakoc, C. Rakowski, and G. A. Blobel. 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J.* **24**:2367–2378.
- Horak, C. E., M. C. Mahajan, N. M. Luscombe, M. Gerstein, S. M. Weissman, and M. Snyder. 2002. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIP-chip analysis. *Proc. Natl. Acad. Sci. USA* **99**:2924–2929.
- Hsu, M., C. A. Richardson, E. Olivier, C. Qiu, E. E. Bouhassira, C. H. Lowrey, and S. Fiering. 2009. Complex developmental patterns of histone modifications associated with the human beta-globin switch in primary cells. *Exp. Hematol.* **37**:799–806.
- Igarashi, K., K. Kataoka, K. Itoh, N. Hayashi, M. Nishizawa, and M. Yamamoto. 1994. Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367**:568–572.
- 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. USA* **102**:17065–17070.
38. Jin, V. X., H. O'Geen, S. Iyengar, R. Green, and P. J. Farnham. 2007. Identification of an OCT4 and SRY regulatory module using integrated computational and experimental genomics approaches. *Genome Res.* **17**: 807–817.
 39. Jing, H., C. R. Vakoc, L. Ying, S. Mandat, H. Wang, X. Zheng, and G. A. Blobel. 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell* **29**:232–242.
 40. Johnson, K. D., M. E. Boyer, J. A. Kang, A. Wickrema, A. B. Cantor, and E. H. Bresnick. 2007. Friend of GATA-1-independent transcriptional repression: a novel mode of GATA-1 function. *Blood* **109**:5230–5233.
 41. Johnson, K. D., S. I. Kim, and E. H. Bresnick. 2006. Differential sensitivities of transcription factor target genes underlie cell type-specific gene expression profiles. *Proc. Natl. Acad. Sci. USA* **103**:15939–15944.
 42. Kadauke, S., and G. A. Blobel. 2009. Chromatin loops in gene regulation. *Biochim. Biophys. Acta* **1789**:17–25.
 43. Karolchik, D., R. M. Kuhn, R. Baertsch, G. P. Barber, H. Clawson, M. Diekhans, B. Giardine, R. A. Harte, A. S. Hinrichs, F. Hsu, K. M. Kober, W. Miller, J. S. Pedersen, A. Pohl, B. J. Raney, B. Rhead, K. R. Rosenbloom, K. E. Smith, M. Stanke, A. Thakkapallayil, H. Trumbower, T. Wang, A. S. Zweig, D. Haussler, and W. J. Kent. 2008. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**:D773–D779.
 44. Kim, H. R., B. S. Kennedy, and J. D. Engel. 1989. Two chicken erythrocyte band 3 mRNAs are generated by alternative transcriptional initiation and differential RNA splicing. *Mol. Cell Biol.* **9**:5198–5206.
 45. Kim, S. I., S. J. Bultman, H. Jing, G. A. Blobel, and E. H. Bresnick. 2007. Dissecting molecular steps in chromatin domain activation during hematopoietic differentiation. *Mol. Cell Biol.* **27**:4551–4565.
 46. Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov, R. D. Green, M. Q. Zhang, V. V. Lobanov, and B. Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**:1231–1245.
 47. Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green, and B. Ren. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**:876–880.
 48. King, D. C., J. Taylor, L. Elnitski, F. Chiaromonte, W. Miller, and R. C. Hardison. 2005. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res.* **15**:1051–1060.
 49. Koefler, H. P., and D. W. Golde. 1980. Human myeloid leukemia cell lines: a review. *Blood* **56**:344–350.
 50. Kooren, J., R. J. Palstra, P. Klous, E. Splinter, M. von Lindern, F. Grosveld, and W. de Laat. 2007. Beta-globin active chromatin hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J. Biol. Chem.* **282**:16544–16552.
 51. Kudo, S., M. Onda, and M. Fukuda. 1994. Characterization of glycophorin A transcripts: control by the common erythroid-specific promoter and alternative usage of different polyadenylation signals. *J. Biochem.* **116**:183–192.
 52. Kudrycki, K. E., and G. E. Shull. 1993. Rat kidney band 3 Cl⁻/HCO₃⁻ exchanger mRNA is transcribed from an alternative promoter. *Am. J. Physiol.* **264**:F540–F547.
 53. Lahlii, R., E. Lecuyer, S. Herblot, and T. Hoang. 2004. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol. Cell Biol.* **24**:1439–1452.
 54. Law, A., K. Hirayoshi, T. O'Brien, and J. T. Lis. 1998. Direct cloning of DNA that interacts in vivo with a specific protein: application to RNA polymerase II and sites of pausing in *Drosophila*. *Nucleic Acids Res.* **26**: 919–924.
 55. Layon, M. E., C. J. Ackley, R. J. West, and C. H. Lowrey. 2007. Expression of GATA-1 in a non-hematopoietic cell line induces beta-globin locus control region chromatin structure remodeling and an erythroid pattern of gene expression. *J. Mol. Biol.* **366**:737–744.
 56. Lecuyer, E., and T. Hoang. 2004. SCL: from the origin of hematopoiesis to stem cells and leukemia. *Exp. Hematol.* **32**:11–24.
 57. Letting, D. L., Y. Y. Chen, C. Rakowski, S. Reedy, and G. A. Blobel. 2004. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc. Natl. Acad. Sci. USA* **101**:476–481.
 58. Letting, D. L., C. Rakowski, M. J. Weiss, and G. A. Blobel. 2003. Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol. Cell Biol.* **23**:1334–1340.
 59. Liang, S., B. Moghimi, T. P. Yang, J. Strouboulis, and J. Bungert. 2008. Locus control region mediated regulation of adult beta-globin gene expression. *J. Cell. Biochem.* **105**:9–16.
 60. Lis, J. 1998. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harbor Symp. Quant. Biol.* **63**:347–356.
 61. Lurie, L. J., M. E. Boyer, J. A. Grass, and E. H. Bresnick. 2008. Differential GATA factor stabilities: implications for chromatin occupancy by structurally similar transcription factors. *Biochemistry* **47**:859–869.
 62. McKeown, C. R., R. B. Nowak, J. Moyer, M. A. Sussman, and V. M. Fowler. 2008. Tropomodulin1 is required in the heart but not the yolk sac for mouse embryonic development. *Circ. Res.* **103**:1241–1248.
 63. Migliaccio, A. R., G. Migliaccio, A. Di Baldassarre, and K. Eddleman. 2002. Circulating hematopoietic progenitor cells in a fetus with alpha thalassemia: comparison with the cells circulating in normal and non-thalassemic anemia fetuses and implications for in utero transplantations. *Bone Marrow Transplant.* **30**:75–80.
 64. Miles, J., J. A. Mitchell, L. Chakalova, B. Goyenechea, C. S. Osborne, L. O'Neill, K. Tanimoto, J. D. Engel, and P. Fraser. 2007. Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus. *PLoS ONE* **2**:e630.
 65. Mohandas, N., and P. G. Gallagher. 2008. Red cell membrane: past, present, and future. *Blood* **112**:3939–3948.
 66. Morrow, J. S., D. L. Rimm, S. P. Kennedy, C. D. Cianci, J. H. Sinard, and S. A. Weed. 1997. Of membrane stability and mosaics: the spectrin cytoskeleton, p. 485–540. *In* J. Hoffman and J. Jamieson (ed.), *Handbook of physiology*. Oxford, London, United Kingdom.
 67. Mosser, E. A., J. D. Kasanov, E. C. Forsberg, B. K. Kay, P. A. Ney, and E. H. Bresnick. 1998. Physical and functional interactions between the transactivation domain of the hematopoietic transcription factor NF-E2 and WW domains. *Biochemistry* **37**:13686–13695.
 68. Motohashi, H., T. O'Connor, F. Katsuoaka, J. D. Engel, and M. Yamamoto. 2002. Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene* **294**:1–12.
 69. Motohashi, H., J. A. Shavit, K. Igarashi, M. Yamamoto, and J. D. Engel. 1997. The world according to Maf. *Nucleic Acids Res.* **25**:2953–2959.
 70. Muntean, A. G., and J. D. Crispino. 2005. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood* **106**:1223–1231.
 71. Muse, G. W., D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger, and K. Adelman. 2007. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**:1507–1511.
 72. Noma, K., C. D. Allis, and S. I. Grewal. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**:1150–1155.
 73. Noordermeer, D., and W. de Laat. 2008. Joining the loops: beta-globin gene regulation. *IUBMB Life* **60**:824–833.
 74. Orkin, S. H. 1990. Globin gene regulation and switching: circa 1990. *Cell* **63**:665–672.
 75. Osada, H., G. Grutz, H. Axelson, A. Forster, and T. H. Rabbitts. 1995. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc. Natl. Acad. Sci. USA* **92**:9585–9589.
 76. Pal, S., A. B. Cantor, K. D. Johnson, T. B. Moran, M. E. Boyer, S. H. Orkin, and E. H. Bresnick. 2004. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc. Natl. Acad. Sci. USA* **101**:980–985.
 77. Panzenbock, B., P. Bartunek, M. Y. Mapara, and M. Zenke. 1998. Growth and differentiation of human stem cell factor/erythropoietin-dependent erythroid progenitor cells in vitro. *Blood* **92**:3658–3668.
 78. Parra, M. K., S. L. Gee, M. J. Koury, N. Mohandas, and J. G. Conboy. 2003. Alternative 5' exons and differential splicing regulate expression of protein 4.1R isoforms with distinct N-termini. *Blood* **101**:4164–4171.
 79. Patrinos, G. P., M. de Krom, E. de Boer, A. Langeveld, A. M. Imam, J. Strouboulis, W. de Laat, and F. G. Grosveld. 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.* **18**:1495–1509.
 80. Pevny, L., C. S. Lin, V. D'Agati, M. C. Simon, S. H. Orkin, and F. Costantini. 1995. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**:163–172.
 81. 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.
 82. Pomerantz, O., A. J. Goodwin, T. Joyce, and C. H. Lowrey. 1998. Conserved elements containing NF-E2 and tandem GATA binding sites are required for erythroid-specific chromatin structure reorganization within the human beta-globin locus control region. *Nucleic Acids Res.* **26**:5684–5691.
 83. Radonjic, M., J. C. Andrau, P. Lijnzaad, P. Kemmeren, T. T. Kockelkorn, D. van Leenen, N. L. van Berkum, and F. C. Holstege. 2005. Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol. Cell* **18**:171–183.
 84. Rentoft, M., K. Kim, Y. Cho, C. H. Lee, and A. Kim. 2008. Enhancer requirement for histone methylation linked with gene activation. *FEBS J.* **275**:5994–6001.
 85. Rougvie, A. E., and J. T. Lis. 1990. Postinitiation transcriptional control in *Drosophila melanogaster*. *Mol. Cell Biol.* **10**:6041–6045.
 86. Sabatino, D. E., C. Wong, A. P. Cline, L. Pyle, L. J. Garrett, P. G. Gallagher, and D. M. Bodine. 2000. A minimal ankyrin promoter linked to a human gamma-globin gene demonstrates erythroid specific copy number depen-

- dent expression with minimal position or enhancer dependence in transgenic mice. *J. Biol. Chem.* **275**:28549–28554.
87. Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**:407–411.
 88. Schubeler, D., D. M. MacAlpine, D. Scalzo, C. Wirbelauer, C. Kooperberg, F. van Leeuwen, D. E. Gottschling, L. P. O'Neill, B. M. Turner, J. Delrow, S. P. Bell, and M. Groudine. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**:1263–1271.
 89. Schuh, A. H., A. J. Tipping, A. J. Clark, I. Hamlett, B. Guyot, F. J. Iborra, P. Rodriguez, J. Strouboulis, T. Enver, P. Vyas, and C. Porcher. 2005. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol. Cell. Biol.* **25**:10235–10250.
 90. Shivdasani, R. A., Y. Fujiwara, M. A. McDevitt, and S. H. Orkin. 1997. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J.* **16**:3965–3973.
 91. Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**:1034–1050.
 92. Siepel, A., and D. Haussler. 2004. Combining phylogenetic and hidden Markov models in biosequence analysis. *J. Comput. Biol.* **11**:413–428.
 93. Simon, M. C., L. Pevny, M. V. Wiles, G. Keller, F. Costantini, and S. H. Orkin. 1992. Rescue of erythroid development in gene targeted GATA-1 mouse embryonic stem cells. *Nat. Genet.* **1**:92–98.
 94. Stamatoyannopoulos, J. A., A. Goodwin, T. Joyce, and C. H. Lowrey. 1995. NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J.* **14**:106–116.
 95. Strahl, B. D., R. Ohba, R. G. Cook, and C. D. Allis. 1999. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc. Natl. Acad. Sci. USA* **96**:14967–14972.
 96. Takahashi, S., K. Onodera, H. Motohashi, N. Suwabe, N. Hayashi, N. Yanai, Y. Nabesima, and M. Yamamoto. 1997. Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J. Biol. Chem.* **272**:12611–12615.
 97. 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.
 98. Vakoc, C. R., D. L. Letting, N. Gheldof, T. Sawado, M. A. Bender, M. Groudine, M. J. Weiss, J. Dekker, and G. A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**:453–462.
 99. Vakoc, C. R., M. M. Sachdeva, H. Wang, and G. A. Blobel. 2006. Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* **26**:9185–9195.
 100. Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**:RESEARCH0034.
 101. 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.
 102. Wadman, I. A., H. Osada, G. G. Grutz, A. D. Agulnick, H. Westphal, A. Forster, and T. H. Rabbitts. 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.
 103. Walensky, L. D., M. Narla, and S. E. Lux. 2003. Disorders of the red blood cell membrane, p. 1709–1858. *In* R. I. Handin, S. E. Lux, and T. P. Stossel (ed.), *Blood: principles and practice of hematology*, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, PA.
 104. Wang, H., Y. Zhang, Y. Cheng, Y. Zhou, D. C. King, J. Taylor, F. Chiaromonte, J. Kasturi, H. Petrykowska, B. Gibb, C. Dorman, W. Miller, L. C. Dore, J. Welch, M. J. Weiss, and R. C. Hardison. 2006. Experimental validation of predicted mammalian erythroid cis-regulatory modules. *Genome Res.* **16**:1480–1492.
 105. Weinmann, A. S., P. S. Yan, M. J. Oberley, T. H. Huang, and P. J. Farnham. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.* **16**:235–244.
 106. Weiss, M. J., G. Keller, and S. H. Orkin. 1994. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8**:1184–1197.
 107. 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. Teichman, A. J. Bannister, and B. Gottgens. 2009. The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood* **113**:5456–5465.
 108. Wong, E. Y., J. Lin, B. G. Forget, D. M. Bodine, and P. G. Gallagher. 2004. Sequences downstream of the erythroid promoter are required for high level expression of the human alpha-spectrin gene. *J. Biol. Chem.* **279**:55024–55033.
 109. Wu, J. L., Y. S. Lin, C. C. Yang, Y. J. Lin, S. F. Wu, Y. T. Lin, and C. F. Huang. 2009. MCRS2 represses the transactivation activities of Nrf1. *BMC Cell Biol.* **10**:9.
 110. Xu, X., M. Bieda, V. X. Jin, A. Rabinovich, M. J. Oberley, R. Green, and P. J. Farnham. 2007. A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res.* **17**:1550–1561.
 111. Xu, Z., X. Meng, Y. Cai, H. Liang, L. Nagarajan, and S. J. Brandt. 2007. Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins. *Genes Dev.* **21**:942–955.
 112. Yamamoto, M. L., T. A. Clark, S. L. Gee, J. A. Kang, A. C. Schweitzer, A. Wickrema, and J. G. Conboy. 2009. Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. *Blood* **113**:3363–3370.
 113. Yao, W., J. Nathanson, I. Lian, F. H. Gage, and L. A. Sung. 2007. Mouse erythrocyte tropomodulin in the brain reported by lacZ knocked-in downstream from the E1 promoter. *Gene Expr. Patterns.* **8**:36–46.
 114. Yenerel, M. N., I. B. Sundell, J. Weese, M. Bulger, and D. M. Gilligan. 2005. Expression of adducin genes during erythropoiesis: a novel erythroid promoter for ADD2. *Exp. Hematol.* **33**:758–766.
 115. Yun, S., Y. Rim, and E. H. Jho. 2007. Induced expression of the transcription of tropomodulin 1 by Wnt5a and characterization of the tropomodulin 1 promoter. *Biochem. Biophys. Res. Commun.* **363**:727–732.
 116. Zeng, P. Y., C. R. Vakoc, Z. C. Chen, G. A. Blobel, and S. L. Berger. 2006. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques* **41**:694. **696**:698.
 117. 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.
 118. Zhang, Y., H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13**:1924–1935.