Cell Reports
Resource

Locus-Specific Proteomics by TChP: Targeted Chromatin Purification

Farzin Pourfarzad,[1,2,10] Ali Aghajanirefah,[1,2] Ernie de Boer,[1] Sara Ten Have,[8] Thamar Bryn van Dijk,[1] Sima Kheradmandkia,[3] Ralph Stadhouders,[1] Supat Thongjuea,[9] Eric Soler,[1,4] Nynke Gillemans,[1] Marieke von Lindern,[5,10] Jeroen Demmers,[3,6,7] Sjaak Philipsen,[1,4,6] and Frank Grosveld[1,2,4,5,6,*]

[1]Department of Cell Biology
[2]Center for Biomedical Genetics
[3]Department of Biochemistry
[4]Cancer Genomics Center
[5]Department of Hematology
[6]Netherlands Proteomics Centre
[7]Biomics Department
Erasmus MC, Dr. Molewaterplein 50, 3015GE Rotterdam, the Netherlands
[8]Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK
[9]Computational Biology Unit, Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway
[10]Present address: Department of Blood Cell Research and Hematopoiesis, Sanquin Research and Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, Plesmanlaan 125, 1066 CX Amsterdam, the Netherlands.
*Correspondence: f.grosveld@erasmusmc.nl

http://dx.doi.org/10.1016/j.celrep.2013.07.004

SUMMARY

Here, we show that transcription factors bound to regulatory sequences can be identified by purifying these unique sequences directly from mammalian cells in vivo. Using targeted chromatin purification (TChP), a double-pull-down strategy with a tetracycline-sensitive "hook" bound to a specific promoter, we identify transcription factors bound to the repressed γ-globin gene-associated regulatory regions. After validation of the binding, we show that, in human primary erythroid cells, knockdown of a number of these transcription factors induces γ-globin gene expression. Reactivation of γ-globin gene expression ameliorates the symptoms of β-thalassemia and sickle cell disease, and these factors provide potential targets for the development of therapeutics for treating these patients.

INTRODUCTION

The sensitivity of proteomics technology has been improved during recent years to a level where it becomes feasible to purify very small amounts of proteins and obtain their identity by mass spectrometry (Walther and Mann, 2010). As a result, it has become possible to attempt the direct identification of proteins bound to chromatin in vivo in mammalian cells. Attempts to identify such proteins directly have been published before (Déjardin and Kingston, 2009; Fujita and Fujii, 2011; Wang et al., 2013). Déjardin and Kingston used an elegant hybridization technique called proteomics of isolated chromatin (PICh) to identify factors bound to telomeres. Because there are two telomeres per chromosome and hence almost 100 telomeres per mouse or human cell, their total length approximates 0.01%–0.07% of the genome (Déjardin and Kingston, 2009). Fujita and Fujii applied insertional chromatin immunoprecipitation (iChIP) to directly identify components of β-globin chicken HS4 insulator complexes using an exogenous locus containing 24 copies of a chicken HS4-core plasmid possessing 16 copies of LexA binding sequences in a hematopoietic cell line (Fujita and Fujii, 2011). Wang et al. used chromatin-interacting protein MS (ChIP-MS) to identify MSL complex interactions and identified MSL-enriched histone modifications associated with the active genes (Wang et al., 2013). We are interested in identifying chromatin-bound proteins to promoter sequences and regulatory regions of a particular gene. Such DNA elements are normally present in two copies per genome, each constitute less than 1 kb, and, therefore, are almost two orders of magnitude less abundant than the combined telomeres, multicopy exogenous plasmid locus, and active gene bodies mentioned above. It is unlikely that such a small percentage of the genome can be purified sufficiently by PICh, iChIP, or ChIP-MS to allow the identification of specific proteins. We therefore devised a different method, using the human β-globin locus as a proof of principle in this report.

Our specific interest in this locus originates from the suppression of the human fetal γ-globin genes in adults, because a release of this suppression is of clinical interest for patients

suffering from β-thalassemia or sickle cell disease. Presently, the most common treatments for these diseases are blood transfusions combined with iron chelation therapy, or treatment with hydroxyurea or short chain fatty acids, the latter two leading to an increase of fetal γ-globin gene expression (Bank, 2006; Coleman and Inusa, 2007; Mankidy et al., 2006; Perrine, 2008; Stamatoyannopoulos, 2005). The γ-globin chains replace the absent or abnormal β-globin chains, thus ameliorating disease symptoms. However, these treatments are not satisfactory because they do not lead to a normal quality of life nor do they prevent a relatively early death. At present, bone marrow transplantation is the only effective cure, a risky procedure that is not available to the large majority of patients.

Despite intense research efforts by many laboratories, it is only partly understood how the human γ-globin genes are normally suppressed around the time of birth when expression switches to the adult β-globin gene. Several factors have been identified to be involved in the suppression of the γ-globin genes. For example, BCL11A and KLF1 were recently shown to lead to elevated γ-globin gene expression when their activity is suppressed (Borg et al., 2011; Sankaran et al., 2008; Xu et al., 2010, 2011). Interestingly, the promoters of the γ-globin genes were previously identified as the regions responsible for their suppression (Berry et al., 1992; Dillon and Grosveld, 1991; Li et al., 2001; Starck et al., 1994; Yu et al., 2006), whereas the BCL11A protein binds to a region downstream of the γ-globin genes. BCL11A interacts with SOX6, GATA1, FOG1, and the NuRD repressor complex, and these interactions are thought to act via the γ-globin gene promoter and regulatory regions (Sankaran et al., 2008; Xu et al., 2010). KLF1 is indirectly involved as it positively regulates the β-globin and *BCL11A* promoters. This suggests that the γ-globin promoter interacts with the surrounding regulatory sequences to achieve silencing.

The direct proteomics approach that we developed, targeted chromatin purification (TChP), entails the pull-down of a protein hook that is crosslinked to the γ-globin gene promoter. Importantly, this will also purify interacting *cis*-regulatory regions because these are spatially organized in a chromatin hub (Palstra et al., 2003; Patrinos et al., 2004). Here, we show that this approach successfully identified a number of protein factors bound to the globin chromatin hub whose reduced expression leads to the activation of the γ-globin genes in human erythroid cells. These factors shed light on the mechanism of γ-globin gene switching and may provide targets for the development of drugs that release the suppression of γ-globin genes in β-thalassemia and sickle cell disease patients.

RESULTS

Experimental Design

All experiments were carried out following the Dutch guidelines of the METC (Medical-Ethical Committee) and the DEC (Ethics Committee Animal Experiments). We opted for an unbiased proteomics approach by introducing Tet operator (TetO) sequences in the upstream γ-globin promoter. These sequences are bound by a tagged TetR protein that is used as a hook to pull down the γ-globin promoter together with its associated regulatory sequences by affinity purification. We considered that the experimental system should meet three important requirements. First, the bait should not disturb the normal expression pattern of the γ-globin gene. Second, the fragments of interest each constitute less than one-millionth of the genome; hence, many cells are needed to enable the purification of sufficient amounts of material for further analysis. Third, a large number of general (contaminating) chromatin-binding proteins will be found, because the relevant sequences constitute such a small part of the genome. The control sample should therefore be a close mimic of the experimental sample to control for these contaminants.

We used mouse transgenesis to generate a model system meeting these requirements. The basic design of the approach is schematically shown in Figure 1. Binding sites for the bacterial tetracycline repressor protein (TetR) were inserted into a human β-globin minilocus that contains the locus control region (LCR), the Aγ-globin gene, and the 3′ hypersensitive site 1 (Figure 1A). Previous experiments have shown that such a γ-globin minilocus is regulated properly in the mouse (Dillon and Grosveld, 1991). The γ-globin gene is expressed in the embryo and early mouse fetal liver; it is silenced around embryonic day 14 (E14) in the fetal liver and remains silenced in the adult (Peterson et al., 1998; Strouboulis et al., 1992).

The modified globin minilocus (LCR-TetO-γ) was introduced into transgenic mice that are p53 null (LCR-TetO-γ::p53-null mice) to facilitate the derivation of immortalized cell lines (Dolznig et al., 2001; Donehower et al., 1992; von Lindern et al., 2001). Next, we constructed a triple tag in the TetR protein (TetR3T, Figure 1A) that enables sequential purification on HA and streptavidin affinity beads. The TetR3T cDNA was cloned in the hematopoietic expression vector pIE3.9IntpolyAA (Ohneda et al., 2002), and this construct (TetR3T) was introduced in transgenic mice expressing the *E. coli* BirA biotin ligase from the pEV expression vector (de Boer et al., 2003); BirA efficiently biotinylates the Bio tag of the TetR3T protein as shown by probing western blots of crude nuclear extracts with HRP-streptavidin (see below). Crossing of the TetR3T::BirA mice with the LCR-TetO-γ::p53-null mice resulted in quadruple LCR-TetO-γ::TetR3T::BirA::p53-null mice after several breeding and selection steps. Quadruple LCR-TetO-γ::TetR3T::BirA::p53-null embryos were collected at 13.5 days postcoitum (dpc), a developmental time point coinciding with the suppression of the human γ-globin transgene, and used to generate fetal-liver derived erythroid progenitor cell lines (Dolznig et al., 2001; Peterson et al., 1998; Strouboulis et al., 1992; von Lindern et al., 2001) (Figure 1B). One of these cell lines was used for the suppressed Aγ-globin chromatin hub purification and identification of the bound proteins by mass spectrometry (Figures 1C and 1D).

Development of the Experimental System

First, the TetO site was tested in electrophoretic mobility shift assays with TetR3T::BirA fetal liver cell extracts. This showed strong binding of TetR3T to the TetO sequence (data not shown). We found no evidence for binding of other proteins. Next, seven copies of the TetO sequence were introduced into the StuI site 380 bp upstream of the transcription initiation site of the γ-globin gene in the γ-globin minilocus (Figure 1A). This LCR-TetO-γ locus was used for transgenesis, and several LCR-TetO-γ transgenic lines were selected for further analysis (Figure S1). The

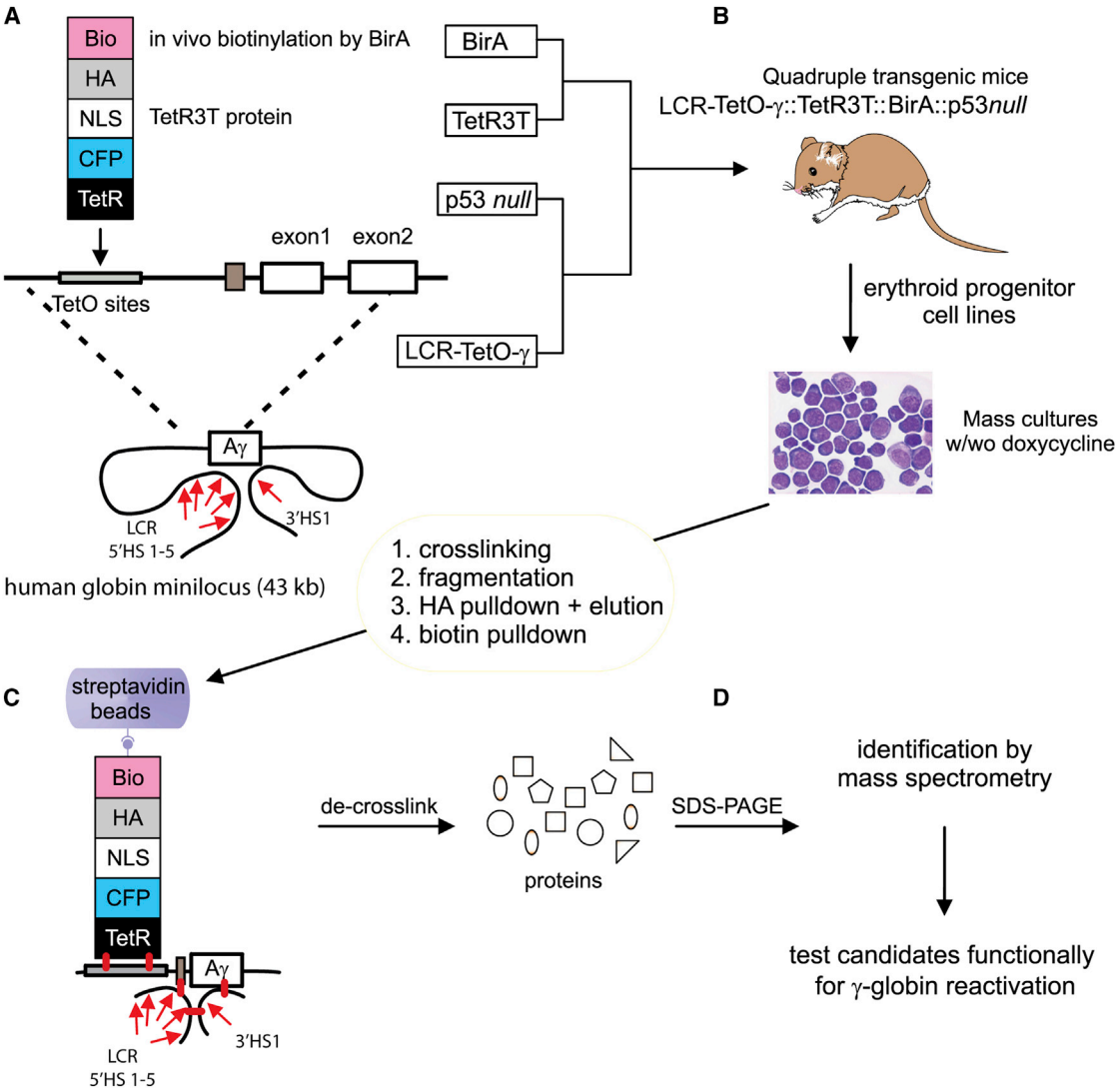

Figure 1. Scheme of the TChP Approach

(A) Generation of the LCR-TetO-γ and TetR3T-G1HRD constructs resulting in the two transgenic mouse lines: TetR3T::BirA and LCR-TetO-γ::p53-null mice.

(B) LCR-TetO-γ::TetR3T::BirA::p53-null mice were used to derive the LCR-TetO-γ::TetR3T::BirA::p53-null cell line.

(C) Purification of the γ-globin chromatin hub after formaldehyde crosslinking, the LCR-TetO-γ::TetR3T::BirA::p53-null cell-line chromatin is fragmented and used for HA and biotin pull-down. Small red dashes represent hypothetical crosslinks.

(D) Decrosslinking of the isolated chromatin proteins, identification by mass spectrometry, and functional testing for potential role in γ-globin regulation.

developmental mRNA expression pattern of γ-globin from the LCR-TetO-γ locus was analyzed, and the results obtained with line 05-23736-05, which contains two copies of the transgene, are shown in Figure 2A. From these data, we conclude that the LCR-TetO-γ locus is expressed properly during development. It is active during the embryonic period, suppressed during the late fetal liver stage, and remains inactive in the adult.

We next performed circularized chromosome conformation capture sequencing (4C-Seq) to visualize the spatial organization of the suppressed human LCR-TetO-γ-globin minilocus in the mouse fetal-liver-derived cell line. This showed that the suppressed human γ-globin gene promoter is indeed in close proximity to human LCR 5′HS2, 3, 4, and 3′HS1 (Figure 2B).

In parallel, the TetR3T cDNA was cloned in the pIE3.9IntpolyAA vector (Ohneda et al., 2002) resulting in the TetR3T construct. This vector recapitulates the hematopoietic expression pattern of the mouse Gata1 gene, driving expression at all developmental stages of erythroid cells, also when the γ-globin gene is suppressed. The TetR3T protein is composed of the TetR DNA binding domain coupled to CFP, three copies of a nuclear localization signal, an HA tag, and a biotinylation tag (Figures 1A and 2C).

Purification of the Suppressed γ-Globin Chromatin Hub

Erythroid progenitor cell lines (von Lindern et al., 2001) were derived from LCR-TetO-γ::TetR3T::BirA::p53null fetal livers at

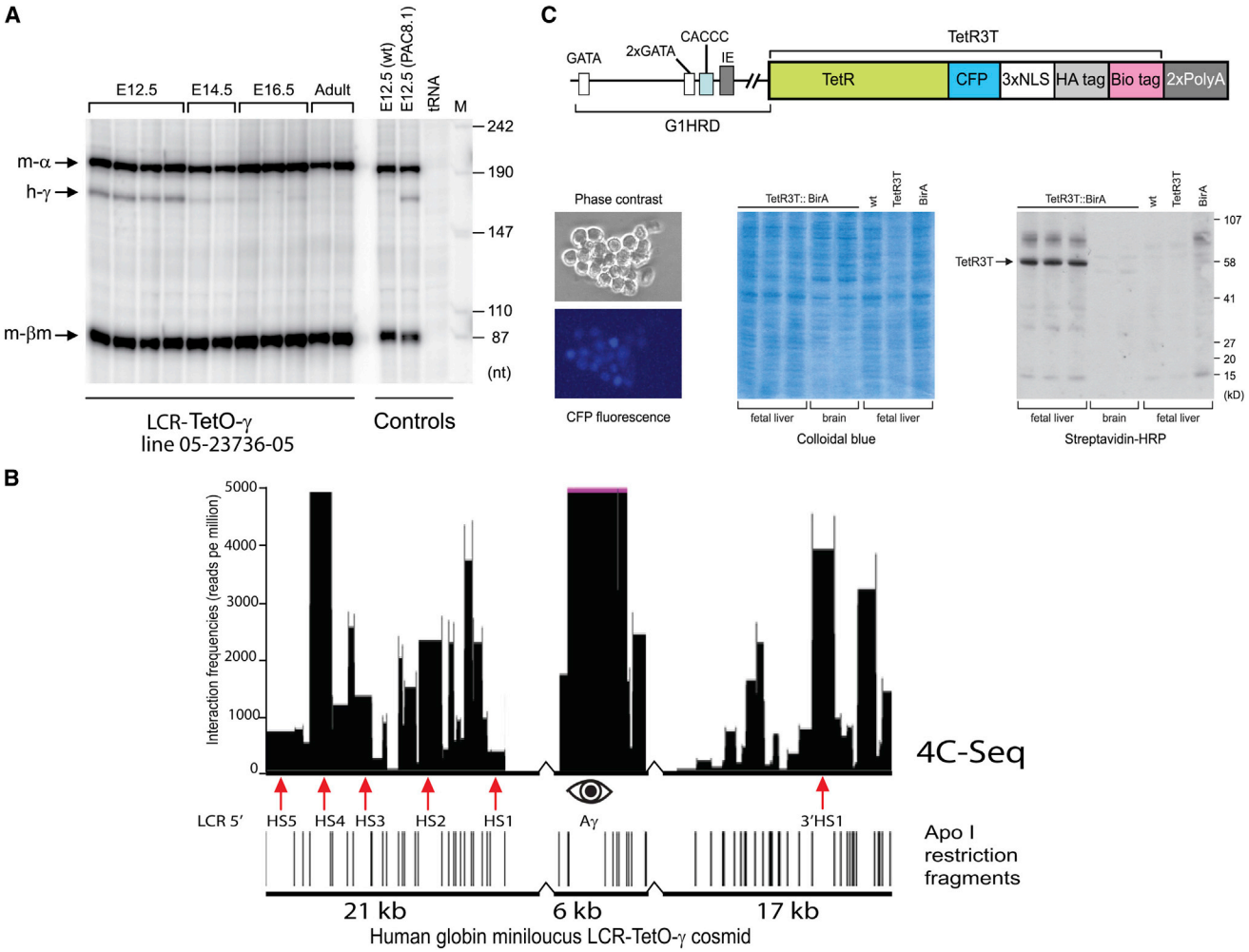

Figure 2. Transgenic Mouse Proof of Principle Experiments

(A) S1 nuclease protection analysis of globin expression in LCR-TetO-γ transgenic mice. RNA was prepared from fetal livers and adult blood of transgenic mice carrying two copies of the LCR-TetO-γ transgene (line 05-23736-05). Controls are shown on the right. E12.5 WT: RNA from liver of wild-type mouse E12.5 mouse embryo; E12.5 PAC8.1: RNA from liver of E12.5 mouse embryo carrying a complete human β-globin locus. The positions of the protected fragments for mouse α-globin (m-α), β-major globin (m-βm), and human γ-globin (h-γ) are indicated. Size marker: pUC18 cut with MspI.

(B) LCR-TetO-γ-globin chromatin hub. Analysis of the LCR-TetO-γ-globin mini locus spatial organization in mouse erythroid fetal liver cell line by 4C-seq using the γ-globin promoter as the viewpoint (indicated by the eye symbol). Bars indicate interactions of restriction fragments with the viewpoint as measured by the number of normalized sequence reads. The human β-globin minilocus LCR-TetO-γ cosmid fragments are indicated on the bottom; ^ indicates the junction between two fragments.

(C) Construction and expression of the TetR3T hook protein. Top panel: depiction of the TetR3T cDNA construction in the G1HRD erythroid-specific expression vector (Ohneda et al., 2002). Bottom panel left: nuclear localization of TetR3T in E13.5 TetR3T transgenic mouse fetal liver cells. Top picture shows phase contrast picture of a group of cells. Bottom picture shows nuclear CFP fluorescence in the same cells. Middle and left panel: TetR3T expression as shown by streptavidin-HRP-stained western blot of fetal liver and brain extracts of WT, TetR3T, BirA, and TetR3T::BirA compound transgenic E13.5 embryos. Molecular weights are shown on the right. See also Figure S1.

E13.5. The cell lines were cloned and further propagated in vitro. Out of these, a cell line was randomly selected that showed the expression pattern of mouse fetal liver at this stage of development; very low expression of the human γ-globin gene and high levels of mouse α- and β-globin expression (Figure 2A). This cell line was expanded for the subsequent optimization of each step of the biochemical purification of the suppressed γ-globin associated chromatin. First, the TetR3T precipitation by streptavidin beads was tested (Figure 3A). We also used chromatin immunoprecipitation (ChIP) to confirm that the TetO repeats were bound by the TetR3T protein, and released upon the addition of doxycycline (Figure 3B). The TetR3T protein and its bound TetO sequences could be efficiently pulled down by anti-HA beads as well as streptavidin (Figure 3), in principle, enabling consecutive pull-downs.

We next tested whether the chromatin captured with the HA beads could be eluted efficiently by the addition of HA peptide, allowing a subsequent pull-down step on the eluted chromatin

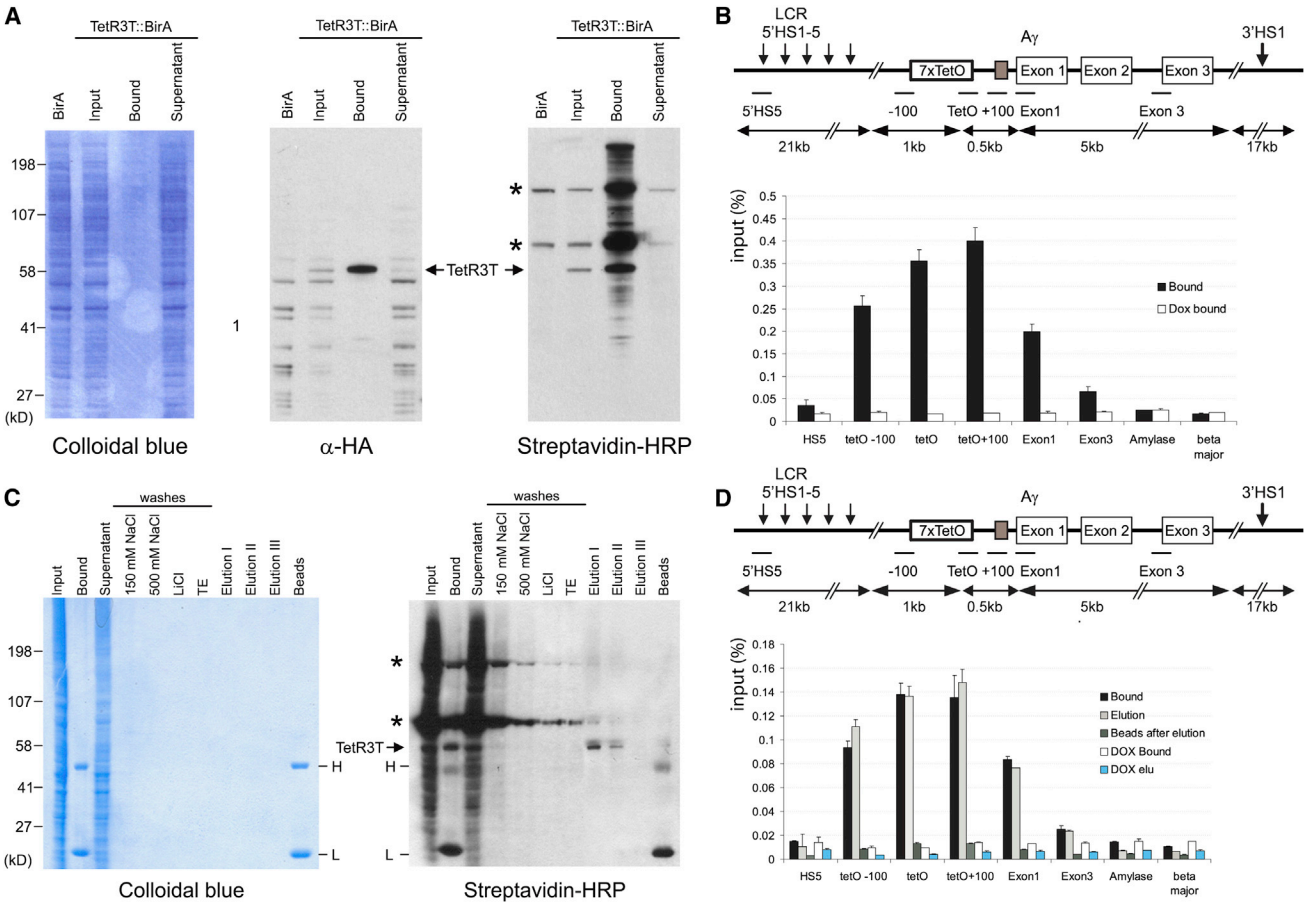

Figure 3. Affinity Purification of the γ-Globin Chromatin Hub

(A) Left panel; gel electrophoresis of the nuclear extracts after colloidal blue staining. The first lane contains a control extract from an identical cell line not containing the TetR3T construct. Middle panel; detection of the TetR3T protein with an HA-specific antibody on western blot. Right panel: detection of the TetR3T protein with streptavidin HRP on western blot. *Naturally occurring biotinylated proteins (de Boer et al., 2003).

(B) Top: map of the γ-globin locus (not to scale) and the distances of the various fragments from the TetO sites in the promoter. Bottom: chromatin precipitation of the different fragments on streptavidin beads in the presence or absence of doxycycline. The mouse β-major and Amylase genes were used as negative controls.

(C) Left panel: gel electrophoresis of the nuclear extracts after ChIP with an HA-specific antibody. The lanes show the bound, washed, and eluted fractions of the beads. H and L are the heavy and light chains of the HA-specific antibody. Right panel: western blot detection of the TetR3T protein with streptavidin HRP of the lanes shown in the left panel. *Naturally occurring biotinylated proteins (de Boer et al., 2003).

(D) Top: same as top in (B). Bottom: PCR detection of the various fragments of the γ-globin locus or the mouse β-major and amylase control genes after ChIP by binding to HA beads followed by HA peptide elution, in the presence or absence of doxycycline. Bars in (B) and (D) represent averages of three independent measurements; error bars represent SD.

See also Tables S1 and S2.

by streptavidin. The results show that all of the bound material could be eluted efficiently by HA peptide (Figures 3C and 3D). Both pull-downs showed an enrichment of the sequences upstream and downstream of the TetO sites (Figures 3B and 3D).

Subsequently all parameters for TChP were optimized (Table S1). This included fixation conditions, crosslinking, choice of beads for pull-down, buffer composition, blocking agents to prevent nonspecific binding, large-scale culture, and the elution conditions. The optimization details are shown in the Extended Experimental Procedures.

The optimization resulted in conditions that gave over 50-fold purification for both the HA- and the biotin pull-down (Figure 3). In the final optimized protocol the yield of the purification steps, measured as TetO DNA recovery after the HA and biotin pull-downs, was estimated at <2% of the starting material.

Proteomics Analysis of the Pulled-Down Chromatin

Currently, the sensitivity limit for peptide identification by LC-MS (Orbitrap) mass spectrometry is in the high attomole range. This suggests that a minimum number of 3×10^9 cells is required to recover sufficient amounts of suppressed γ-globin chromatin complexes with the optimized procedure. This would still be on the borderline of detection. The maximum culture volume we were able to handle, using the optimized purification protocol, was 1.2 l. We therefore repeated the purification several times to maximize identification of possible specific hits. To purify

Table 1. Summary of the Selected Proteins and Their Identified Peptides

	Identified Protein	Score	Da	Coverage (%)	Uniq Pept.	Total Pept.	Description	gil
1	Ybx1	275	35,822	22.4	3	3	Y box protein 1	gi|55451
2	Apex1	258	35,867	14.2	3	3	Apurinic/apyrimidinic endonuclease 1	gi|6753086
3	Cdc5l	232	92,361	4.6	2	2	Cell division cycle 5-like (*S. pombe*)	gi|22779899
4	Gtf2f2	205	28,421	20.1	4	4	General transcription factor IIF, polypeptide 2	gi|39930425
5	Mybbp1a	156	152,773	1.9	2	2	MYB binding protein (P160) 1a	gi|2645205
6	Gata1	142	43,274	9	2	2	GATA binding protein 1	gi|6679947
7	Nap1l1	130	45,602	6.9	2	2	Nucleosome assembly protein 1-like 1	gi|7657357
8	Ctnnbl1	93	65,421	3.2	2	2	Catenin, beta-like 1	gi|18460918
9	Supt5h	79	120,988	4.5	3	3	Suppressor of Ty 5 homolog	gi|22094123
10	Chd4	77	94,191	3.4	2	2	Chromodomain helicase DNA binding protein 4	gi|13543768
11	Sox30	62	84,285	2.9	2	7	SRY-box containing gene 30	gi|27734194
12	Dbf4	56	75,042	4.2	2	3	DBF4 homolog (*S. cerevisiae*)	gi|7304903
13	Zfp148	56	86,469	3.8	2	11	Zinc finger protein 148	gi|1724124
14	Tcfeb	55	59,616	3.4	2	8	Transcription factor EB	gi|4454797

See also Table S3.

chromatin hub, complexes 1.2 l of cells were grown to a density of 2.5×10^6 cells per ml. For each experiment, we grew two cultures in parallel, one with and one without doxycycline. The cells were harvested and treated with 1% formaldehyde, and the crosslinked chromatin was subjected to the isolation procedure. After the final step, the chromatin was decrosslinked, and the proteins were separated by SDS-PAGE. The lanes were sliced, and each slice was used for mass spectrometry analysis. We performed a total of four independent experiments, collectively identifying 441 proteins with a Mascot score of $\geq$55 and $\geq$2 unique identified peptides. Of these, 427 were present in the doxycycline-treated purification experiments including ubiquitous nuclear proteins such as histones and splicing factors (350), had low-quality peptide fragmentation spectra, or were very unlikely to be involved in the suppressed γ-globin chromatin hub (77), such as abundant structural and cytoplasmic proteins, glutathione S-transferase, and ribosomal proteins. These were relegated to the bottom of the list, leaving 14 candidates (Tables 1 and S3). Among these 14 candidates were a number of factors that have been reported previously to bind to the LCR or the γ-globin promoter: APEX1, GATA1, CDC5l, ZFP148 (ZBP-89), TCFEB, and CHD4 (Mi-2β) (Hirayama and Shinozaki, 1996; Harju-Baker et al., 2008; Karmakar et al., 2010; Mahajan et al., 2005; Olave et al., 2007; Reece-Hoyes et al., 2011; Woo et al., 2011).

Chromatin Hub Occupancy of Identified Proteins

To validate the chromatin occupancy of the identified proteins, we selected a set of proteins known to bind to the locus, and a number of proteins not known to bind to the locus for which we could obtain antibodies for ChIP experiments. ChIP experiments were carried out on the β-globin LCR and γ-globin locus in human primary erythroid progenitors (Figure 4) as well as the mouse fetal-liver-derived cell line (Figure S2). We tested a selected number of identified protein factors for binding (Table 1), namely, APEX1, GATA1, NAP1L1, SUPT5h, CDC5l, ZFP148 (ZBP-89), CHD4 (Mi2-β), CTNNBL1, TCFEB, and YBX1. Six out of these ten proteins, APEX1, GATA1, CDC5l, ZFP148 (ZBP-89), NAP1L1, and TCFEB, showed binding to different elements within the locus. In particular, the 5′HS3 and 5′HS4 regulatory regions showed binding of at least five proteins, whereas the other elements showed binding of only one or two of these proteins (Figure 4). Particularly striking was the strong enrichment of ZBP-89 (Zfp148 in mouse) at 5′HS2 (Figure 4). The ChIP experiments in mouse cells carrying the LCR-TetO-γ-globin minilocus transgene resulted in a very similar recruitment profile of these factors. We also tested some of the protein factors that were present in doxycycline-treated pull-downs as negative controls, these included, EHMT1, FANCI, SON, and HMGB3. However, none these showed recruitment to the locus (Figure S2). These results suggest that the experiments identified a number of factors that are bound to the globin locus, which might be involved in γ-globin transcription regulation.

Functional Assessment of Identified Proteins

We next carried out functional assays for a number of the proteins identified, to address whether they have any effect on the expression of the suppressed γ-globin genes. We tested this with the human homologs of these proteins in human erythroid progenitor (HEP) cells (Borg et al., 2010; Leberbauer et al., 2005).

ZBP-89 knockdown (65% at the protein level) resulted in a 2.5-fold rise in γ-globin mRNA expression and a rise in HbF (Figures 5A–5C; Table 2). This was accompanied by an increase of the percentage of cells expressing high HbF levels as determined by immunohistochemistry (IHC) (Figure 5D; Table 2). These results were reproduced with HEP cells derived from six different healthy individuals (data not shown).

Similar sets of experiments were performed for CDC5L, APEX1, TFEB, CTNNBL1, and SUPT5H. We also included the HMGB3 protein that was present in doxycycline-treated pull-down control experiments. In all cases, at least two of the small hairpin RNA (shRNA) tested resulted in a knockdown of the

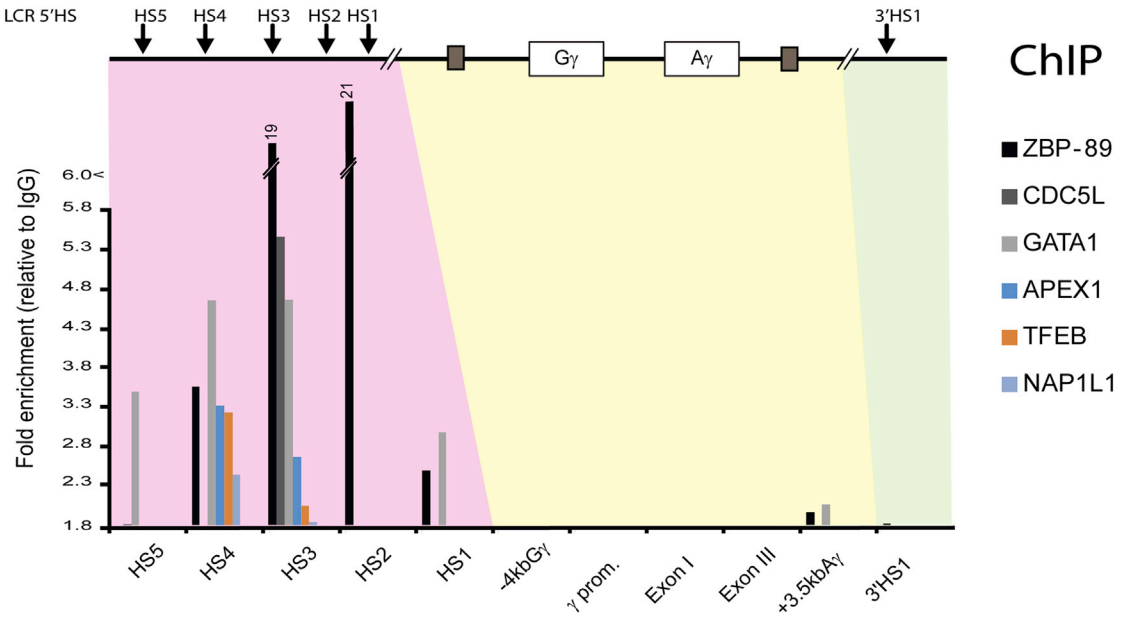

Figure 4. Chromatin Occupancy of Selected Set of Protein Factors Identified by TChP along the β-Globin Locus in Human Erythroid Progenitor Cells

Results are shown only for proteins with more than 1.8-fold enrichment. The enrichments are normalized relative to antibody immunoglobulin (Ig) G species and negative region 1 kb downstream of Aγ-globin gene. Bars represent averages of two experiments. The different transcription factors tested are color-coded in the bar graph as shown on the right. See also Figure S2 and Tables S2 and S3.

factor, and the experiments were reproduced in at least two independent donors. The results of these knockdown experiments are summarized in Table 2. Knockdown of the CDC5L and APEX1, TFEB proteins in HEP cells resulted in increased level of γ-globin mRNA and HbF, which was also observed at the cellular level by immunohistochemistry (Table 2). Two of the nonbinding proteins in our ChIP experiments (Table 2; Figure 4), SUPT5H and CTNNBL1 as well as the HMGB3 control protein, were also knocked down efficiently by different shRNAs, but these knockdowns showed no or very little change in the level of γ-globin at mRNA or protein level. We conclude from these knockdown experiments that a number of the factors we identified by TChP indeed appear to be involved in suppression of γ-globin gene expression. Importantly, knockdowns of some of the factors had no effect strongly suggesting that the observed induction of γ-globin gene expression upon knockdown of ZBP-98, CDC5L, APEX1, and TFEB is specific.

DISCUSSION

Here, we show that chromatin factors bound to unique sequences can be identified directly through the purification of the target locus sequences (TChP) in vivo. Although the results obtained are not directly comparable with those of Déjardin and Kingston (2009), the TChP approach presented here may be more generally applicable and suitable to the isolation of unique sequences. The two approaches are fundamentally different, and each method has its advantages and disadvantages. The PICh method involves a hybridization step to capture the target chromatin, whereas TChP relies on the binding of the TetR protein to the TetO sequences. The latter is likely to be more efficient than hybridization but has the disadvantage that it requires the introduction of sequences that serve as a "bait" into the naturally occurring sequences. Therefore, a control should be carried out to ensure that the introduction of the bait has not changed the behavior of the target sequence. The advantage is that, once the sequences have been introduced, multiple purification steps can be carried out with the same bait, using the appropriate tags and elution conditions. In each approach, it is essential to run a comparable negative control to allow the identification of candidate binding proteins. Dejardin and Kingston list approximately 200 proteins in the control PICh pull-down, probably due to mismatched hybridization and nonspecific sticking of irrelevant chromatin. The latter is supported by the observation that more than one-third of the PICh background proteins were also found among the TChP background proteins.

Using TChP, we identified 14 unique protein IDs and selected a number of these candidates for further functional analysis by analysis of their recruitment to the human β-globin locus and the effect of their reduced protein level on γ-globin expression in adult human HEP cells.

ZBP-89 (also known as BFCOL1, BERF-1, ZNF148 in human and Zfp148 in mouse) knockdown resulted in γ-globin gene activation. ZBP-89 is a Krüppel-type zinc finger transcription factor that binds to a GC-rich region, and subsequently represses or activates its target genes. ZBP-89 interacts with GATA1 and MAFK, and it is involved in erythroid development and differentiation (Brand et al., 2004; Woo et al., 2008, 2011). We recently showed that the 5FMC complex is recruited to ZBP-89 target

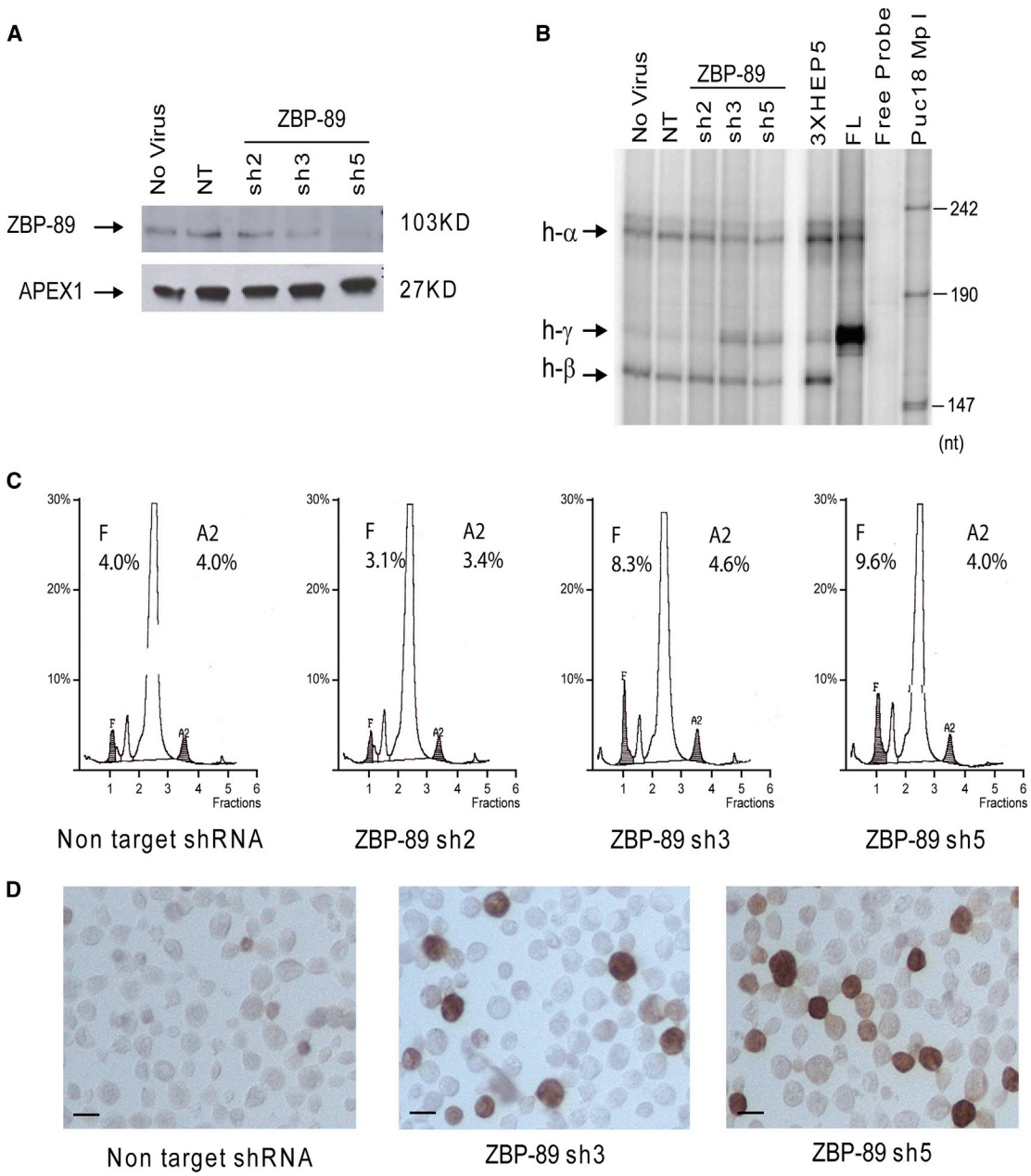

Figure 5. Lentiviral shRNA-Mediated Knockdown of ZBP-89

(A) Top panel: immunoblot for ZBP-89 shRNA-mediated knockdown in HEP cells. First lane: no virus control. NT, nontarget shRNA lentiviral transduced HEP cells. Bottom panel: APEX1 staining serving as loading control.

(B) S1 nuclease protection analysis of globin expression in ZBP-89 depleted HEP cells. First lane: no virus control, NT; nontarget shRNA, ZBP-89 sh2, 3, and 5; short hairpins to human ZBP-89. Controls (to the right): 3 × HEP; triple amount of RNA from non-virus-treated HEP and FL (human fetal liver) as the positive control; size marker: pUC18xMspI. The positions of human α-, γ-, and β-globin protected fragments are indicated.

(C) HPLC hemoglobin analysis of the ZBP-89 shRNA knockdown in HEP cells.

(D) HbF Immunohistochemistry of ZBP-89 shRNA knockdown HEP cells. Scale bars, 10 μm.

See also Table S3.

genes, and together they are involved in transcriptional regulation (Fanis et al., 2012).

CDC5l (cell division cycle 5-like) is a cell-cycle regulator important for G2 to M transition and has sequence-specific DNA binding activity (CTCAGCG). Its DNA binding domain has similarities to MYB, which is known to be involved in γ-globin repression (Jiang et al., 2006; Kuroyanagi et al., 2006). Cdc5l is able to bind to the γ-globin promoter (Olave et al., 2007), and it interacts

Table 2. Summary of Functional Studies

	Knockdown Efficiency	Globin mRNA γ/(γ+β)	HbF % (HPLC)	HbF-Positive Cells % (IHC)	Binds to (ChIP)
Scramble	0%	0.2	4.0	5	
ZBP-89	65%	0.5	9.0	18	HS3 and HS2
CDC5L	90%	0.7	12.8	25	HS4, HS3, γ-globin gene, and 3HS1
APEX1	55%	0.8	19.0	19	HS3 and gamma gene
TCFEB	65%	0.6	Not determined	Not determined	HS4 and HS3
HMGB3	60%	0.1	Not determined	Not determined	No binding detected
CTNNBL1	85%	0.2	6.1	Not determined	No binding detected
SUPT5	70%	0.2	5.3	Not determined	No binding detected

The experiments were performed in HEP cells grown from at least two independent donors. The knockdown experiments are expressed as average of at least two different shRNA viruses. See also Table S3.

with β-catenin-like 1 protein (CTNNBL1) (Conticello et al., 2008). Interestingly, CDC5L knockdown increased fetal hemoglobin, whereas a reduced level of CTNNBL1 had no substantial effect on γ-globin expression (Table 2).

Moderate knockdown of APEX1 (Apurinic-apyrimidinic endonuclease 1) also increased γ-globin expression. Apex1 is a multifunctional protein involved in DNA repair activity, proofreading exonuclease activity, and modulating DNA binding activity of several transcription factors including AP-1, CREB, and NRF2. Interestingly, AP-1 and CREB bind to the human β-globin LCR 5′HS2 (Johnson et al., 2002; Talbot and Grosveld, 1991). APEX1 belongs to a multiprotein complex necessary for both transcription and DNA replication at the β-globin locus (Karmakar et al., 2010).

TFEB was recently identified as a transcription factor that binds to human β-globin regulatory elements in a high-throughput regulatory network mapping yeast one-hybrid assay (Reece-Hoyes et al., 2011). Reduced level of TFEB also resulted in more than 2-fold increase in γ-globin expression.

HMGB3, CTNNBL1, and SUPT5H protein knockdown did not result in γ-globin gene activation. In addition, we failed to show any binding of these proteins to the human β-globin regulatory regions, suggesting that these factors are not involved in γ-globin repression.

In the knockdown experiments, we observed γ-globin upregulation for four out of six identified factors tested. Moreover, identifying specific proteins such as GATA1, CHD4, CDC5l that are already known to interact with the γ-globin promoter or β-globin locus LCR provides confidence regarding the validity of the TChP procedure developed in this paper and suggests that this method could be used more generally. Using the *p53*-null mouse background, it could be applied to any other (immortalized) cell type that could either be grown in large amounts in culture or be isolated directly, e.g., from liver or brain.

The TChP method may be improved further in order to increase the yield and decrease the background contamination by incorporating some extra steps, other crosslinking agents, and/or tags. The next generation of the TChP can be more efficient and generally applicable by designing zinc finger proteins. This makes TChP much easier as no artificial sequences (tetO) would have to be introduced and endogenous loci could be targeted directly. Moreover, this could be carried out in the appropriate cell lines rather than using mice. In addition, the multiple tag used for sequential pull-downs could be further improved, for instance, by including other affinity tags such as V5, MYC, or FLAG that are not dependent on BirA and known to work well upon formaldehyde crosslinking will further reduce the complexity of TChP (Kolodziej et al., 2009). In addition, a density fractionation step to isolate the chromatin after crosslinking could be used to remove contaminating proteins. Probably the most important loss of efficiency is the crosslinking of the TetR protein to the TetO sequence. TetR binds as a dimer to an inverted repeat sequence (Orth et al., 2000), and its crosslinking may be improved substantially by the incorporation of additional lysine or cysteine amino acids next to the binding sequence, and the incorporation of dG next to the TetO binding site. It is known that lysine-dG and cysteine-dG are the most efficient combinations for formaldehyde crosslinking (Lu et al., 2010; Orth et al., 2000). Further improvements may be achieved by adding a restriction site to the bait that would cleave the target DNA sequence from the TetO sequences. In addition, a sequence-specific protease site could be introduced between the tag and DNA binding moiety of the TetR3T protein such as TEV, which recognizes a linear protein sequence (Kapust et al., 2002), or SUMOstar protease (Lifesensors), which recognizes a 3D SUMO structure (Malakhov et al., 2004). These adjustments significantly reduce the complexity of TChP and make it adoptable for other laboratories that would be likely to identify in vivo chromatin-bound proteins.

In summary, we have shown that proteins bound to unique sequences in the genome can be purified sufficiently for identification by mass spectrometry. This resulted in the identification of a number of known but also proteins that are potentially involved in the suppression of the human γ-globin genes.

EXPERIMENTAL PROCEDURES

DNA Constructs

A human γ-globin gene ClaI-KpnI fragment from the γ-globin minilocus (Dillon and Grosveld, 1991) was modified as a subclone by insertion of the TetO heptamer (Gossen and Bujard, 1992) ligated to a short random sequence containing SwaI and PmeI restriction enzyme digestion sites into the StuI site of the γ-globin gene promoter 380 bp upstream of the transcription initiation site and a loxP sequence inserted in the ClaI site (sequences are provided in

the Extended Experimental Procedures) (Figures 1 and S1). The modified ClaI-KpnI fragment was cloned back into the minilocus by standard λ phage packaging (Stratagene) and transduction into *E. coli* DH10B. DNA was isolated and the integrity of the modified minilocus established by cleavage with EcoRI and gel electrophoresis (Figure S1).

The 642 bp TetR binding domain (Gossen and Bujard, 1992) was cloned in-frame with CFP cDNA (pECFP-N1, Clontech). The insert was recloned into pEYFP-Nuc replacing eYFP (Clontech) thereby gaining the 3× NLS and SV40 polyA sequence. A biotinylation tag (de Boer et al., 2003) and HA tag (YPYDVPDYA), separated by a few glycine residues as linker and restriction sites, were cloned in-frame from ligated overlapping oligonucleotides between the NLS and poly A resulting in TetR3T cDNA. The sequences of overlapping oligonucleotides for HA-Bio tag synthesis are provided in the Extended Experimental Procedures.

The TetR3T fragment was excised by EcoRI-NotI digestion and inserted into the NotI site of the G1HRD expression vector (Ohneda et al., 2002). The resulting TetR3T-RG1HRD vector was tested for integrity and standard transfection and expression in MEL cells, which demonstrated its localization to the nucleus by CFP fluorescence.

Transgenesis and Cell-Line Derivation

The minilocus DNA was cut with SalI, the insert fragment isolated by gel electrophoresis and injected into fertilized oocytes (Dillon and Grosveld, 1991). A number of γ minilocus transgenic mice was obtained, one of which contained two copies of the minilocus (line 05-23736-05). Its integrity was established by EcoRI digestion and Southern blots of genomic DNA (Figure S1). Similarly The Asp718-PvuI TetR3T restriction fragment was isolated from the vector by gel electrophoresis and injected into fertilized eggs (de Boer et al., 2003). The resulting mice expressed the TetR3T protein in the erythroid lineage. Finally, the mouse strains were crossed to yield LCR-TetO-γ::TetR3T::BirA::p53-null mice (Donehower et al., 1992). These mice were intercrossed and E13.5 fetal liver cells were cultured to obtain the LCR-TetO-γ::TetR3T::BirA::p53-null cell line (Dolznig et al., 2001).

Cell Cultures

The culturing of human erythroid proerythroblasts (HEPs) (Leberbauer et al., 2005) and LCR-TetO-γ::TetR3T::BirA::p53-null cells (von Lindern et al., 2001) was essentially as published with minor modifications provided in the Extended Experimental Procedures.

Chromatin Immunoprecipitation and γ-Globin Chromatin Purification with Targeted Chromatin Purification

Crosslinked chromatin was prepared by adding formaldehyde directly to culture medium to a final concentration of 1% and incubated for 10 min at room temperature. The reaction was stopped by the addition of 1 M glycine to a final concentration 0.125 M. The cells were disrupted, and the DNA was fragmented by sonication at 0°C. γ-Globin locus, DNA sequences, and their bound proteins were precipitated using HA-agarose beads. The precipitated material was eluted by HA peptides followed by a sequential Dynabeads M-280 streptavidin (Invitrogen) pull-down for further purification. The efficiency of purification was analyzed by immunoprecipitation and quantitative PCR. Purified material was analyzed by mass spectrometry for identification of proteins bound to the γ-globin locus and its associated DNA sequences. The anti-HA-agarose beads pull-down, HA elution, further purification by streptavidin pull-down, ChIP, and the analysis of precipitated material by immunoprecipitation and quantitative PCR protocols are detailed in the Extended Experimental Procedures.

Mass Spectrometry

One-dimensional SDS-PAGE gel lanes were cut into 2 mm slices using an automatic gel slicer and subjected to in-gel reduction with dithiothreitol, alkylation with iodoacetamide, and digestion with trypsin (Promega, sequencing grade), essentially as described (Wilm et al., 1996) Nanoflow LC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to either an LTQ-Orbitrap mass spectrometer (Thermo Scientific) or an LTQ linear ion trap (Thermo Scientific), both operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed phase column (Dr. Maisch GmbH; column dimensions 1.5 cm × 100 μm, packed in-house) at a flow rate of 8 μl min^{-1}. Peptide separation was performed on ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 μm, packed in-house) using a linear gradient from 0% to 80% B (A = 0.1% formic acid; B = 80% [v/v] acetonitrile, 0.1% formic acid) in 70 min and at a constant flow rate of 200 nl min^{-1} using a splitter. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode. Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.3; MatrixScience). The Mascot search algorithm (version 2.2, MatrixScience) was used for searching against the NCBI database (taxonomy: *Mus musculus*; release NCBInr_20081130.fasta). The peptide tolerance was set to 10 ppm (Orbitrap) or to 2 Da (ion trap) and the fragment ion tolerance was set to 0.8 Da. A maximum number of two missed cleavages by trypsin were allowed, and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. As would be expected, the protein hits in TChP are quite weak. Using the statistical procedures that are commonly used (typically a false discovery rate [FDR] of 1% for the protein database search algorithm), the hits would not be indicated as valid hits. The FDRs applied to this data set were therefore loosened up. The Mascot score cutoff value for a positive protein hit was set to 55 with more than two unique peptides. Selection of hits to follow up was based on gene ontology analysis (gene ontology terms used "transcription factor", "chromatin remodeling factors").

Circularized Chromosome Conformation Capture Sequencing

4C material was prepared essentially as described (Palstra et al., 2003; Soler et al., 2010). ApoI was used as the primary restriction endonuclease. Purified 4C DNA was used for 4C-Seq library preparation as described (Palstra et al., 2003; Soler et al., 2010; Stadhouders et al., 2011), using NlaIII as secondary restriction endonuclease. Viewpoint-interacting DNA fragments were PCR-amplified using viewpoint-specific divergent primers (forward-specific for Aγ 5′-CAG GTA GTT GTT CCC CTT CA; reverse 5′- AAT CCA TTT CGG CAA AGA ATT C) linked to standard Illumina adaptor sequences. The resulting 4C library was single-read sequenced on the Illumina Genome Analyzer II platform generating 76 bp reads. Images were recorded and analyzed by the GAP pipeline. The resulting reads were trimmed to remove viewpoint-specific primer sequences and mapped against NCBI build hg18 of the human genome using ELAND alignment software. To calculate the coverage, aligned reads were extended to 56 bp in the 3′ direction using the 4C-Seq pipeline (Thongjuea et al., 2013). The interaction-enriched regions per ApoI fragment were measured by calculating the number of reads per million (RPM) per restriction fragment. Data were visualized using a local mirror of the UCSC genome browser.

Lentiviral shRNA-Mediated Knockdown

shRNA vectors were obtained from the TRC Mission human and mouse library from Sigma. Packaging was done as described by the Sigma Mission library protocol. HEP cells were transduced with lentiviruses containing shRNAs, puromycin was added at a concentration of 1 μg ml^{-1} for 24 hr, and the cells were kept in culture for another 5 days before harvesting and further analysis.

RNA Purification

Total RNA was extracted from cells using the TRI reagent (Sigma) and used directly for S1 nuclease protection analysis of globin expression.

S1 Nuclease Protection Assays

S1 nuclease protection assays were performed as described (Hanscombe et al., 1991). The dried gels were exposed to a Typhoon TRIO phosphorimager screen (GE Healthcare), and signal intensities of individual bands were quantified with Image Quant 5.2.

Western Blotting

Cells (200 × 10^5) were lysed in 100 μl of RIPA buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1% Nonidet P-40, 1% NaDOC, 0.1% SDS), the same volume of Laemmli buffer was added and the samples were boiled for 5 min. Twenty microliters of each sample was resolved by 4%–12% gradient SDS-PAGE (Invitrogen) and transferred to 0.45 μm nitrocellulose membrane. The membranes were blocked for 1 hr in 1 × TBS with 2% BSA and 0.05% Tween 20 and incubated in primary (overnight) and secondary (1 hr) antibody. Primary antibodies: Abcam FANCI (KIA1794) ab15344, ZBP-89 ab69933, CTNNBL1 ab76243, CDC5L ab51320, NAP1l1 ab21630, EHMT1 ab41969, SUPT5 sc-101158, BAF53 ab3882, Santa Cruz APEX1 (Ref-1 C4)X sc-17774 X, CHD4 (Mi2β) sc-8774, GATA1 sc-265, SPT5 (Supt5) sc-101158, and Bethyl MCM5 A300-195A. Western blots were developed with the ECL Western Blotting Detection Reagents (GE Healthcare) and quantified with a Kodak Image Station 440CF.

HPLC Analysis

Hemoglobin subtypes were measured by HPLC (Bio-Rad).

Immunohistochemistry

Cells were spotted on poly-prep slides (Sigma), fixed with 4% paraformaldehyde, permeabilized in 10 mM citric acid (pH 6.0), and blocked with 5% BSA. Primary antibody incubation was performed in blocking solution for 16 hr at 4°C, followed by peroxidase staining. Pictures were taken with an Olympus BX40 microscope (40× objective, NA 0.65) equipped with an Olympus DP50 CCD camera and Viewfinder Lite 1.0 acquisition software. HbF-positive cells were determined using Photoshop at a tolerance level of 80.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, two figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.07.004.

ACKNOWLEDGMENTS

We are grateful to Angus Lamond for proteomics support. This work was supported by grants to F.G. and S.P. from the NIH (NIH 5 RO1 HL 73455), the Netherlands Genomics Initiative, the Landsteiner Foundation for Blood Transfusion Research (project grant 1040), the NWO Zon-MW program (grants "Human γ globin gene activation to treat thalassemia and sickle cell anemia," TOP 40-00812-98-12128 and DN 82-301), the EU fp6 Integrated Research Project EuTRACC (LSHG-CT-2007-03744)5, and the EU fp7 Specific Cooperation Research Project THALAMOSS (306201).

Received: May 30, 2013
Revised: July 1, 2013
Accepted: July 7, 2013
Published: August 1, 2013

REFERENCES

Bank, A. (2006). Regulation of human fetal hemoglobin: new players, new complexities. Blood *107*, 435–443.

Berry, M., Grosveld, F., and Dillon, N. (1992). A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. Nature *358*, 499–502.

Borg, J., Papadopoulos, P., Georgitsi, M., Gutiérrez, L., Grech, G., Fanis, P., Phylactides, M., Verkerk, A.J., van der Spek, P.J., Scerri, C.A., et al. (2010). Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat. Genet. *42*, 801–805.

Borg, J., Patrinos, G.P., Felice, A.E., and Philipsen, S. (2011). Erythroid phenotypes associated with KLF1 mutations. Haematologica *96*, 635–638.

Brand, M., Ranish, J.A., Kummer, N.T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T.H., Crabtree, G.R., Aebersold, R., and Groudine, M. (2004). Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. Nat. Struct. Mol. Biol. *11*, 73–80.

Coleman, E., and Inusa, B. (2007). Sickle cell anemia: targeting the role of fetal hemoglobin in therapy. Clin. Pediatr. *46*, 386–391.

Conticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M.S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. Mol. Cell *31*, 474–484.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. USA *100*, 7480–7485.

Déjardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic Loci. Cell *136*, 175–186.

Dillon, N., and Grosveld, F. (1991). Human gamma-globin genes silenced independently of other genes in the beta-globin locus. Nature *350*, 252–254.

Dolznig, H., Boulmé, F., Stangl, K., Deiner, E.M., Mikulits, W., Beug, H., and Müllner, E.W. (2001). Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays. FASEB J. *15*, 1442–1444.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215–221.

Fanis, P., Gillemans, N., Aghajanirefah, A., Pourfarzad, F., Demmers, J., Esteghamat, F., Vadlamudi, R.K., Grosveld, F., Philipsen, S., and van Dijk, T.B. (2012). Five friends of methylated chromatin target of protein-arginine-methyltransferase[prmt]-1 (chtop), a complex linking arginine methylation to desumoylation. Mol. Cell. Proteomics *11*, 1263–1273.

Fujita, T., and Fujii, H. (2011). Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS ONE *6*, e26109.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA *89*, 5547–5551.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., and Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. Genes Dev. *5*, 1387–1394.

Harju-Baker, S., Costa, F.C., Fedosyuk, H., Neades, R., and Peterson, K.R. (2008). Silencing of Agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. Mol. Cell. Biol. *28*, 3101–3113.

Hirayama, T., and Shinozaki, K. (1996). A cdc5+ homolog of a higher plant, Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA *93*, 13371–13376.

Jiang, J., Best, S., Menzel, S., Silver, N., Lai, M.I., Surdulescu, G.L., Spector, T.D., and Thein, S.L. (2006). cMYB is involved in the regulation of fetal hemoglobin production in adults. Blood *108*, 1077–1083.

Johnson, K.D., Norton, J.E., and Bresnick, E.H. (2002). Requirements for utilization of CREB binding protein by hypersensitive site two of the beta-globin locus control region. Nucleic Acids Res. *30*, 1522–1530.

Kapust, R.B., Tözsér, J., Copeland, T.D., and Waugh, D.S. (2002). The P1′ specificity of tobacco etch virus protease. Biochem. Biophys. Res. Commun. *294*, 949–955.

Karmakar, S., Mahajan, M.C., Schulz, V., Boyapaty, G., and Weissman, S.M. (2010). A multiprotein complex necessary for both transcription and DNA replication at the β-globin locus. EMBO J. *29*, 3260–3271.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. BMC Mol. Biol. *10*, 6.

Kuroyanagi, Y., Kaneko, Y., Muta, K., Park, B.-S., Moi, P., Ausenda, S., Cappellini, M.D., and Ikuta, T. (2006). cAMP differentially regulates gamma-globin gene expression in erythroleukemic cells and primary erythroblasts through c-Myb expression. Biochem. Biophys. Res. Commun. *344*, 1038–1047.

Leberbauer, C., Boulmé, F., Unfried, G., Huber, J., Beug, H., and Müllner, E.W. (2005). Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. Blood *105*, 85–94.

Li, Q., Duan, Z.J., and Stamatoyannopoulos, G. (2001). Analysis of the mechanism of action of non-deletion hereditary persistence of fetal hemoglobin mutants in transgenic mice. EMBO J. *20*, 157–164.

Lu, K., Ye, W., Zhou, L., Collins, L.B., Chen, X., Gold, A., Ball, L.M., and Swenberg, J.A. (2010). Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. J. Am. Chem. Soc. *132*, 3388–3399.

Mahajan, M.C., Narlikar, G.J., Boyapaty, G., Kingston, R.E., and Weissman, S.M. (2005). Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. Proc. Natl. Acad. Sci. USA *102*, 15012–15017.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., and Butt, T.R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J. Struct. Funct. Genomics *5*, 75–86.

Mankidy, R., Faller, D.V., Mabaera, R., Lowrey, C.H., Boosalis, M.S., White, G.L., Castaneda, S.A., and Perrine, S.P. (2006). Short-chain fatty acids induce gamma-globin gene expression by displacement of a HDAC3-NCoR repressor complex. Blood *108*, 3179–3186.

Ohneda, K., Shimizu, R., Nishimura, S., Muraosa, Y., Takahashi, S., Engel, J.D., and Yamamoto, M. (2002). A minigene containing four discrete cis elements recapitulates GATA-1 gene expression in vivo. Genes Cells *7*, 1243–1254.

Olave, I.A., Doneanu, C., Fang, X., Stamatoyannopoulos, G., and Li, Q. (2007). Purification and identification of proteins that bind to the hereditary persistence of fetal hemoglobin -198 mutation in the gamma-globin gene promoter. J. Biol. Chem. *282*, 853–862.

Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nat. Struct. Biol. *7*, 215–219.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. *35*, 190–194.

Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M.A., Strouboulis, J., de Laat, W., and Grosveld, F.G. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. Genes Dev. *18*, 1495–1509.

Perrine, S.P. (2008). Fetal globin stimulant therapies in the beta-hemoglobinopathies: principles and current potential. Pediatr. Ann. *37*, 339–346.

Peterson, K.R., Navas, P.A., and Stamatoyannopoulos, G. (1998). beta-YAC transgenic mice for studying LCR function. Ann. N Y Acad. Sci. *850*, 28–37.

Reece-Hoyes, J.S., Barutcu, A.R., McCord, R.P., Jeong, J.S., Jiang, L., MacWilliams, A., Yang, X., Salehi-Ashtiani, K., Hill, D.E., Blackshaw, S., et al. (2011). Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. Nat. Methods *8*, 1050–1052.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839–1842.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. *24*, 277–289.

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A., et al. (2011). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J. *31*, 986–999.

Stamatoyannopoulos, G. (2005). Prospects for developing a molecular cure for thalassemia. Hematology *10*(*Suppl 1*), 255–257.

Starck, J., Sarkar, R., Romana, M., Bhargava, A., Scarpa, A.L., Tanaka, M., Chamberlain, J.W., Weissman, S.M., and Forget, B.G. (1994). Developmental regulation of human gamma- and beta-globin genes in the absence of the locus control region. Blood *84*, 1656–1665.

Strouboulis, J., Dillon, N., and Grosveld, F. (1992). Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. Genes Dev. *6*, 1857–1864.

Talbot, D., and Grosveld, F. (1991). The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J. *10*, 1391–1398.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. *41*, e132. Published online May 13, 2013.

von Lindern, M., Deiner, E.M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M.J., Mullner, E.W., and Beug, H. (2001). Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. Oncogene *20*, 3651–3664.

Walther, T.C., and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. J. Cell Biol. *190*, 491–500.

Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E., Gorchakov, A.A., Britton, L.M., Elledge, S.J., Kharchenko, P.V., Garcia, B.A., and Kuroda, M.I. (2013). Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. Nat. Struct. Mol. Biol. *20*, 202–209.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature *379*, 466–469.

Woo, A.J., Moran, T.B., Schindler, Y.L., Choe, S.-K., Langer, N.B., Sullivan, M.R., Fujiwara, Y., Paw, B.H., and Cantor, A.B. (2008). Identification of ZBP-89 as a novel GATA-1-associated transcription factor involved in megakaryocytic and erythroid development. Mol. Cell. Biol. *28*, 2675–2689.

Woo, A.J., Kim, J., Xu, J., Huang, H., and Cantor, A.B. (2011). Role of ZBP-89 in human globin gene regulation and erythroid differentiation. Blood *118*, 3684–3693.

Xu, J., Sankaran, V.G., Ni, M., Menne, T.F., Puram, R.V., Kim, W., and Orkin, S.H. (2010). Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. *24*, 783–798.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., and Orkin, S.H. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993–996.

Yu, M., Han, H., Xiang, P., Li, Q., and Stamatoyannopoulos, G. (2006). Autonomous silencing as well as competition controls gamma-globin gene expression during development. Mol. Cell. Biol. *26*, 4775–4781.