The Role of ZBP-89 in Globin Regulation

Ali Aghajanirefah

ISBN: 978-94-6203-208-8

The research presented in this thesis was performed at the Department of Cell Biology within the Erasmus University Medical Center, Rotterdam, The Netherlands. The Department is member of Medisch Genetisch Centrum Zuid-West Nederland (MGC).

The research has been supported by NIH grant "Molecular basis of γ-globin gene activation", HL 073455, NGI, NWO, Landsteiner foundation for Blood Transfusion Research (NL), the Center for Biomedical Genetics (CBG, NL), the FP7 EuTRACC consortium of the European Commission and the Netherlands Genomics Initiative.

Cover Design: Sevda Jannatirad
Cover Background: www.docslide.com/erythrocytes-aka-rbc

Layout: Ali Aghajanirefah

Printed by Wöhrmann Print Service, Zutphen

The Role of ZBP-89 in Globin Regulation

De rol van ZBP-89 in de regulatie van globine

Thesis

To obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof.dr. H. G. Schmidt
and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Wednesday 14th November 2012 at 9:30 hours

by

Ali Aghajanirefah

born in Tabriz, Iran

Doctoral Committee:

Promotor:	Prof.dr. F.G. Grosveld
Other members:	Prof.dr. J.N.J. Philipsen Prof.dr. ir. D.N. Meijer Prof.dr. J.H. Gribnau

می‌کوش به هر ورق که خوانی
کان دانش را تمام دانی

All that you have browsed and read- Know the wisdom of paths you tread

(Nizami Ganjavi; Persian poet, 12th century)

To Sevda
&
My parents

Contents

Chapter 1	Introduction	9
Chapter 2	Targeted Chromatin Purification (TChP)	39
Chapter 3	The role of ZBP-89 in globin regulation	75
Chapter 4	Five Friends of Methylated Chtop, a complex linking arginie methylation to Desumoylation	109
Chapter 5	Discussion	143
	Summary	153
	Samenvatting	155
	Curriculum Vitae	157
	Phd Portfolio	159
	Abbreviations	160
	Acknowledgement	162

Scope of the thesis

The ameliorating effect of increased γ-globin gene expression in β-thalassemia and Sickle cell disease, favored γ-globin gene reactivation as the most attractive treatment approach. This has led to intense research efforts to elucidate the mechanism of γ-globin gene switching with the aim of reversing the switching process. Many DNA cis elements (e.g. LCR, -117 HPFH mutation, BGL3 region) as well as trans regulatory protein such as BCL11a, LDB1, and KLF1 has been identified that are important for developmental globin genes regulation. This and other evidences from literature suggest the involvement of other protein factors, as yet unknown, on γ-globin promoter or LCR that work remotely with previously known protein factors. We recently described an unbiased proteomics approach; Targeted Chromatin Purification (TChP), to identify other possible proteins involved in γ-globin silencing.

ZBP-89 was identified as one of the proteins in γ-globin chromatin purification. Here we report functional study on the role of ZBP-89 in globin regulation.

Chapter 1

1

Introduction

Hemoglobin and Disease

Blood is a circulating body fluid providing nutrients and oxygen to the cells and removing waste products from the same cells and it provides protection from infectious agents. Blood cells are suspended in plasma, which constitute 55% of the blood fluid. Blood cells comprise red blood cells, white blood cells and platelets. The most abundant cells in blood are red blood cells (erythrocytes), which contain mostly hemoglobin.

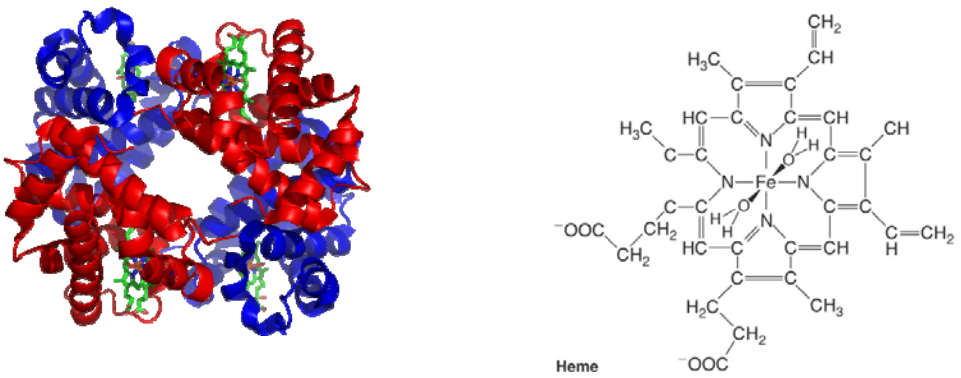
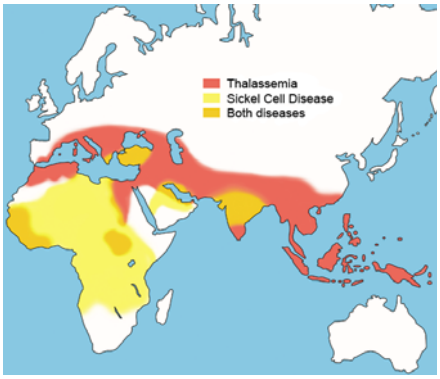


Figure 1. 3D structure of Hemoglobin molecule (left). Molecular structure of Heme group (right).

Hemoglobin is a protein complex containing four chains (two α-like and two β-like chains) with one heme group each whose iron atom binds oxygen temporarily to deliver it to all tissues where the oxygen is exchanged for carbon di-oxyde (Fig. 1). The α-and β-like chains change during the embryonic, fetal and adult stages of development (Steinberg MH, 2001). Early on the embryonic hemoglobins Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) are expressed. They are replaced by fetal hemoglobin (Hemoglobin F) ($\alpha_2\gamma_2$) and finally by Hemoglobin A ($\alpha_2\beta_2$) and Hemoglobin A_2 ($\alpha_2\delta_2$) in the adult. Many mutations have been characterized in the human globin loci, a number of which lead to the formation of pathogenic forms of hemoglobin due to the absence of α- or β-chains or to mutations in the protein chains. Hemoglobin H (β_4) and Hemoglobin Barts (γ_4) are found in α-thalassemias, Hemoglobin S ($\alpha_2\beta^S{}_2$) is found in people with sickle cell disease, while Hemoglobin C ($\alpha_2\beta^C{}_2$) and Hemoglobin E ($\alpha_2\beta^E{}_2$) are found in patients with chronic hemolytic anemia and

Hemoglobin C and Hemoglobin E are found in β-thalassmias.

Mutant hemoglobin was discovered more than the half century ago and many mutations in the α and β–globin gene loci have been determined. Many of the mutations are silent and don't show any clinical phenotype, but a number of mutations and deletions cause a severe phenotype and are common in areas traditionally infested with malaria. Sickle cell anemia and β-thalassemia are the most common β-hemoglobinopathies and one of the world's most frequent monogenic diseases, because heterozygote carriers are at an advantage with respect to *Plasmodium falciparum* malaria infection (Steinberg MH, 2001; Flint et al, 1993; Weatherall, 2005). Around 7% of the global population carry a globin mutation (Weatherall & Clegg, 2001). These diseases occur in South East Asia, Middle East, the Mediterranean area and Africa and to a lesser extent in all other areas of the world due to recent immigration of carriers from these areas to other parts of the world (Fig. 2).



Carrier frequency in the Netherlands

Dr. P.C. Giordano, Leiden

•Immigrant group origin	% carriers	HbP major births	HbP type.
•West Indies, Cape Verde Islands, •Central Africa, Suriname	10 - 24 %	40 - 60	SCD
•Turkey, Greeks, Balkan	2,5 - 8 %	2 - 11	SCD + β-thal
•Marocco, N-Africa	5,5 - 6,5 %	11 - 18	β-thal
•Southern Europe, Mediterranian	0,5 - 8 %	0,08 - 6,4	β-thal
•China, South Eastern Asia (excl. Indonesia)	3 - 26 %	0,6 - 40	β-thal + β-thal/HbE
•Middle- and West-Asia	3 - 9 %	3,4 - 34	SCD + β-thal
•TOTAL	4 - 13,6 % (140.000)	57 - 169	

Figure 2. Distribution of Thalassemia and sickle cell disease in worldwide (left) and carrier frequency in the Netherlands (right)(Dr. P.C.Giordano).

β-hemoglobinopathies are inherited disorders that result from mutations in one of the globin genes caused a decreased level of β-hemoglobin or a decreased ability of hemoglobin to bind oxygen and usually manifest themselves after the first months of life once γ-globin expression are silenced.

Thalassemia

Thalassemia is characterized by a reduction in the synthesis of globin chains. β-thalassemia is the most common autosomal recessive disorders worldwide; however dominant mutations have also been reported. β-thalassemia is caused by the reduced level (beta^{+}) or absence (beta^{0}) of β-globin

chains and unassembled α-chains precipitate causing oxidative damage to the cell membrane (Rund & Rachmilewitz, 2005; Weatherall, 2010). Such cells are recognized by macrophages quickly and removed from the circulation instead of having the normal 120-day life span. β-thalassemia is caused by a base substitution, small deletion or insertion and comes in many different forms (Table 1). Thalassemia major also called Cooley's anemia or Mediterranean anemia. The phenotypes of homozygous or genetic heterozygous compound β-thalassemias include thalassemia major and intermedia.

Types of β-thalassemia	
β-thalassemia	Thalassemia major Thalassemia intermedia Thalassemia minor
β-thalassemia associated with an Hb variant	HbC/β-thalassemia HbE/β-thalassemia HbS/β-thalassemia
Hereditary Persistence Fetal Hemoglobin and β-thalassemia	
β-thalassemia associated with other manifestations	β-thalassemia-tricothiodystrophy X-linked thrombocytopenia with thalassemia
Autosomal dominant forms	Hemoglobin Hakkari

Table 1. Different types of β-thalassemia (Renzo Galanello and Raffaella Origa; 2010).

The most common method of treatment of β-thalassemia major is blood transfusion plus chelation therapy (Figure 3). Blood transfusion compensates the hemoglobin shortage in the bloodstream. However frequent blood transfusion causes a buildup of excessive amounts of iron in the bloodstream, which needs to be regulated by chelation therapy.

Hemoglobin is a protein complex containing four chains (two α-like and two β-like chains, with one heme group each whose iron atom binds oxygen temporarily to deliver it to all tissues where the oxygen is exchanged for carbon di-oxyde (Fig. 1). The α-and β-like chains change during the embryonic, fetal and adult stages of development (Steinberg MH, 2001). Early on the embryonic hemoglobins Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) are expressed. They are replaced by fetal hemoglobin (Hemoglobin F) ($\alpha_2\gamma_2$) and finally by Hemoglobin A ($\alpha_2\beta_2$) and Hemoglobin A_2

($\alpha_2\delta_2$) in the adult. Many mutations have been characterized in the human globin loci, a number of which lead to the formation of pathogenic forms of hemoglobin due to the absence of α- or β-chains or to mutations in the protein chains like hemoglobin C and hemoglobin E which formed due to mutations in the β-globin gene. Hemoglobin H (β_4) and Hemoglobin Barts (γ_4) are found in α-thalassemias, Hemoglobin S ($\alpha_2\beta^S{}_2$) is found in people with sickle cell disease, while Hemoglobin C ($\alpha_2\beta^C{}_2$) and Hemoglobin E ($\alpha_2\beta^E{}_2$) are found in patients with chronic hemolytic anemia and Hemoglobin C and Hemoglobin E are found in β-thalassmias.

The most successful means of β-hemoglobinopathies treatment is hematopoietic stem cell transplantation (Krishnamurti et al, 2008). However it is very difficult to find an appropriate HLA matched donor and it is very expensive. Thalassemia was one of the first candidate diseases to gene therapy and the first such treatments have been carried out recently (Cavazzana-Calvo et al, 2010) while other centers are planning similar trials. However this treatment is still experimental and also very costly and dangerous. Due to the cost general application of these two treatments is out of reach of the vast majority of patients. As a result, treatment is often focused on raising the level of HbF as it can ameliorate the effects of the disease (see Fig. 3).

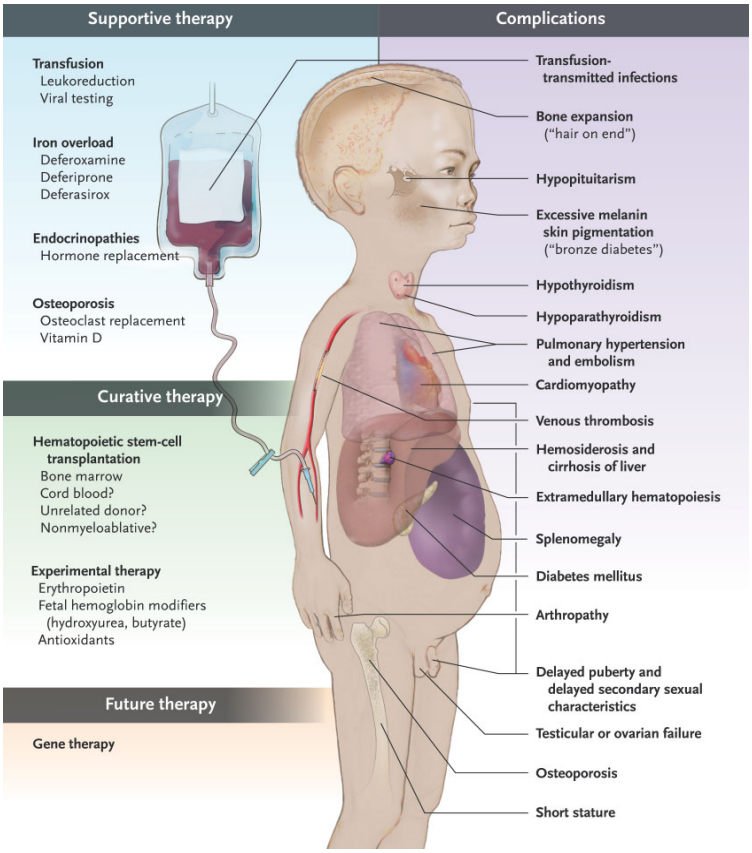


Figure 3. Symptoms and current treatments of Thalassemia (Rund et al, 2005).

Sickle cell disease (SCD)

Sickle cell disease emerges as the result of a mutation in the β-globin gene, which causes a substitution of the amino acid glutamic acid (a charged residue) to valine (a hydrophobic residue) at the sixth amino acid position (Ingram, 1956; Pauling et al, 1949). The presence of HbS can change red blood cells shape to a sickle-shape during deoxygenation, which can cause the blockage in one of the small blood vessels resulting in a local necrosis of the tissue (Fig. 4). SCD can lead to many complications, including anemia, strokes, haemolytic crisis, pain crisis, acute chest syndrome, organ damage, vision loss, infection and skin ulcerations. Individuals who are homozygous SS generally show the most severe sickle cell disease, and the primary hemoglobin present in their red blood cells is sickle hemoglobin. Individuals affected with other types of sickle cell disease are compound

heterozygotes and show a variable phenotype. They possess one copy of the HbS variant plus one copy of another variant such as HbC or Hb β-thalassemia. These individuals produce a mixture of variant hemoglobins (Okpala IE, 2004). Sickle cell trait individuals who possess one copy of the sickle variant and one copy of the normal β-globin gene, produce a mixture of sickle hemoglobin and normal hemoglobin and don't express symptoms of sickle cell disease (Eichner, 2007; Sears, 1978), although one study found that sickle cell trait may be a risk factor for sudden death during physical training (Kark et al, 1987). Importantly, individuals with sickle cell trait are more protected than normal individuals from malaria (Aluoch, 1997). The high frequency of HbS carriers is believed to be a result of this protective effect.

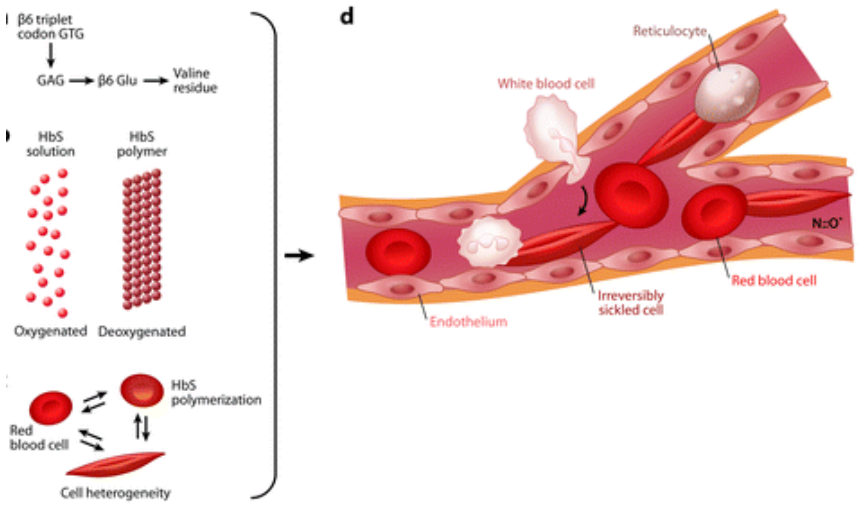


Figure 4. Effects of sickle cell anemia on RBC (Stuart et al, 2004). Under deoxygenated conditions, incorporation of HbS into the hemoglobin tetramer forms the fibrous polymers.

Structure and developmental expression of the β-globin locus

Mammalian enucleated red cells are unique in the animal kingdom, while they remain nucleated in fish, amphibians and birds (Gulliver, 1875). Two different populations of red cells are generated at different times of development in mammals: the first population has larger and nucleated red cells and are generated transiently in the yolk sac (primitive erythropoiesis) that these cells do eventually enucleated whereas the second population consists of smaller, enucleated red cells during fetal and

postnatal life continuously (definitive erythropoiesis). The cells differentiate as part of blood islands and obtain their unique concave shape late in the process by nuclear extrusion (McGrath et al, 2008; Palis, 2008; Tavassoli, 1991). The human α-like globin genes are located on chromosome 16, while the β-like globin genes are located on chromosome 11.

The human α-globin locus

The α-globin locus resides on chromosome 16 and contains 5'-ζ2-ζ1-ψα-ψα-α2-α1-θ-3' in the same order as they are expressed during development (Fig.6). There are four pseudogenes within the α-globin gene cluster (Forget et al, 2001). During the embryonic stage of development the ζ2-globin gene is expressed while the α2 and α1 genes are expressed during the fetal and adult stage. α2 and α1 encode identical α-globin chains (Liebhaber et al, 1986). A strong DNaseI hypersensitive site and transcription factor-binding site, known as HS-40 and located around 40 kb upstream of the α-globin gene cluster (Higgs et al, 1990) controls the expression of the genes(Fig. 5).

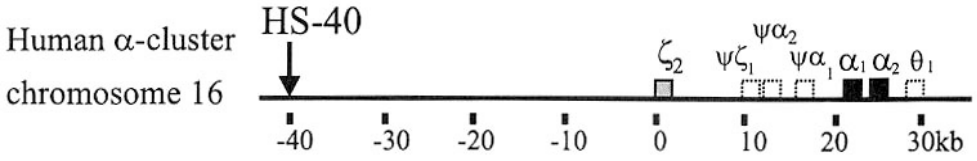


Figure 5. The location of α-globin locus elements on chromosome 16 (Wijgerde 1998)

The human β–globin locus

All of the β-globin genes present in the β-globin locus are regulated by a variety of cis-acting elements. Each of the globin genes has their own promoter. All contain three consensus motifs, namely CACCC, CCAAT and TATA boxes. The most important distant cis-acting element for the tissue and developmental stage specific expression is the Locus Control Region (LCR) (Grosveld et al 1987). In addition there are a number of other elements such as the enhancers 3' of the γ- and β-genes. The function of some of these is important for the appropriate developmental switching of

the genes together with the respective globin gene promoters. All these elements bind a variety of transcription factors (Bank, 2006). The human β-like globin locus contains five globin genes and DNA regulatory sequences located both proximal, distal to and inside the genes. The β-globin locus is surrounded by olfactory receptor (OR) genes which are not expressed in erythroid cells (Bulger et al, 2000; Bulger et al, 1999) as the result of a fairly recent evolutionary event, because the β-globin locus in other species such as birds are located in completely different surrounding sequences and genes. The globin coding region was first cloned as a cDNA (Rabbitts, 1976) and one of the first complete genes to be cloned and sequenced (Proudfoot, 1976; Rougeon et al., 1975). The cluster of five genes occurs in the order 5'- ε, Gγ, Aγ, δ and β which is the same order of their expression during development (Fig. 6).

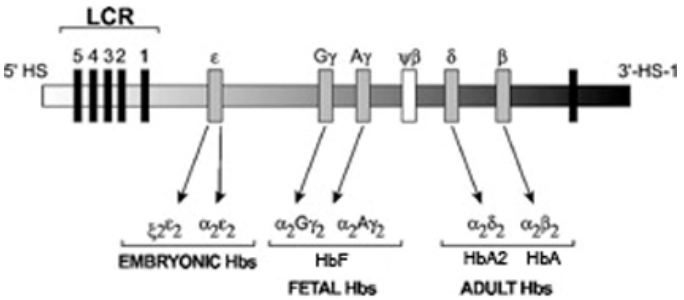


Figure 6. Location of human β-globin locus elements. The top line shows the genes and the major hypersensitive sites at the 5' and 3' side of the locus. The bottom line shows the different hemoglobin molecules formed with the α-like chains during development. (Testa , 2009).

The two γ-globin genes are the result of a 5 kb tandem duplication. The coding region of these two genes differs by one nucleotide in exon 3, where the Gγ gene codon is GGA and Aγ gene codon is GCA. Fetal hemoglobin is composed of 40% Gγ and 60% Aγ in adult cells, whereas throughout fetal development, it is approximately 70% Gγ and 30% Aγ (Schroeder et al, 1971). During development, different globin genes are expressed and the site of hematopoiesis also changes (Fig. 7). The embryonic ε-globin gene is expressed in the yolk sac in the first trimester. By the 13th week of gestation (first switch) the Gγ and Aγ globin genes are activated in the fetal liver and remain expressed during the remainder of fetal life. The β (97%) and δ (3%) globin genes are first activated in the fetal liver and take over after birth (the second switch) leading to the replacement of fetal

hemoglobin (HbF) with adult hemoglobin (Hb A & Hb A_2). HbF levels normally decrease to less than 1% of total globin expression by the end of first year of life. The β-globin and δ-globin genes are expressed in erythroid cells derived from bone marrow. The δ-globin gene is expressed at very low levels and hence the major hemoglobin is HbA while the minor hemoglobin (HbA2) is expressed between 1.5%-3.5% of the total hemoglobin in adult life (Rochette et al, 1994). The LCR is required for high level of β-globin expression in erythroid cells.

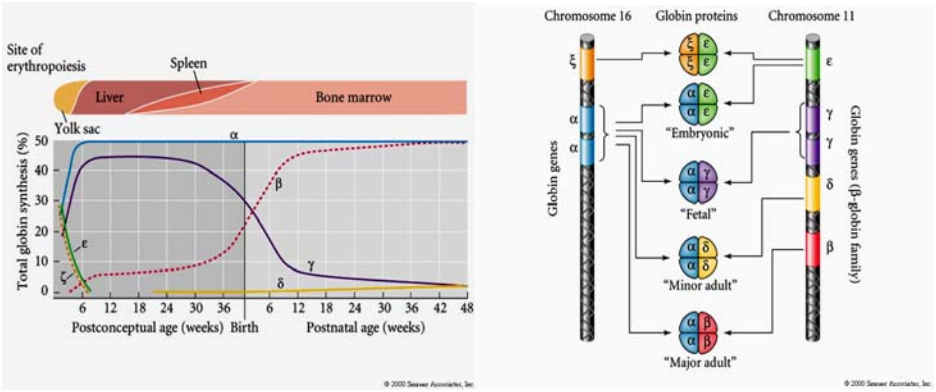


Figure 7. Globin expression. The left panel shows the site of hematopoiesis during development (top) and which globin genes are expressed (bottom). The panel on the right shows the composition of the different types of hemoglobin during development.

Hereditary persistence of fetal hemoglobin (HPFH)

The expression of γ-globin is maintained at a high level in individuals with HPFH and δβ thalassemia in adult life and HPFH individuals have a normal phenotype. Elevated levels of hemoglobin F ameliorate manifestation of sickle cell anemia and β-thalassemia (Steinberg, 2009). HPFH condition is subdivided into deletion (Fig. 8) and non-deletion HPFH. For non-deletional HPFH sites of changed DNA-protein interactions have been located between the -260 and +25 region of the γ-globin promoter (Gumucio et al, 1988; Ikuta & Kan, 1991; Jane et al, 1992; Mantovani et al, 1987; Mantovani et al, 1988; Martin et al, 1989; O'Neill et al, 1990). These point mutations in the promoter of the γ-globin genes have been found at positions -114, -175 and -200 in the 5' promoter regions HBG2 and HBG1 (Chakalova et al, 2005; Gibney et al, 2008; Grosso et al, 2008; Hoyer et al, 2002; Liu et al, 2005). Furthermore, the γ-globin gene remains active at the adult stage when the –117 G to A HPFH mutation is introduced in the promoter (Berry et al., 1992). In contrast many

different large deletions in the β-globin locus cause deletion HPFH, probably resulting from a rearrangement of regulatory regions either by deleting suppressor sequences or bringing activation sequences close to the γ-globin genes (Fig. 8). HPFH can result from mutations outside the locus. For example KLF1 haplo-insufficiency caused by point mutations such as K288X and S270X in the KLF1 gene (Borg et al, 2010; Satta et al, 2011) of this trans-acting factor leads to increased γ-globin expression. A correlation of high HbF levels with several quantitative trait loci (QTL) including the XmnI polymorphism in the promoter of Gγ-globin gene, picked up a single nucleotide polymorphism in the gene of BCL11a on chromosome 2, in the HBS1L-MYB intergenic region on 6q23, chromosome 8 and the X chromosome (Chang et al, 1995; Chang et al, 1997; Garner et al, 2002; Labie et al, 1985; Lettre et al, 2008; Solovieff et al, 2010; Thein et al, 1994; Uda et al, 2008; Wyszynski et al, 2004).

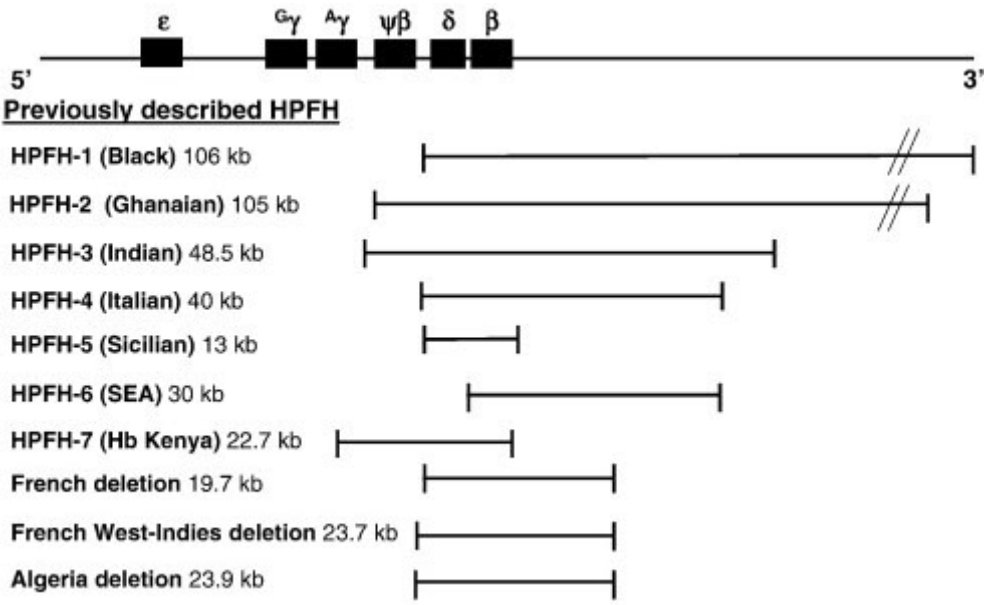


Figure 8. Location and size of HPFH deletions (Joly et al, 2009).

Pharmacological induction of γ-globin expression

The pharmacological treatment of the β-hemoglobinopathies is largely focused on increasing γ-globin expression or on decreasing α-globin expression. Sustained fetal hemoglobin expression of

more than 20% reduces the clinical events in SCD (Powars et al, 1984). Drug therapy of β-thalassemia has focused on stimulating γ-globin gene expression to reduce anemia and eliminate transfusion requirement. Agents which are able to reactivate γ–globin expression fall into two groups: cytotoxic/chemotherapeutic agents and epigenetic reagents, both of which result in some unfavorable side effects on cells. As a result, none of inducer agents is efficient enough (and safe enough in high doses) to "cure" patients. One of pharmacological chemicals is a ribonucleotide reductase inhibitor (Hydroxyurea), which stimulates expression of fetal hemoglobin and currently has been approved by FDA and it is widely used to treat patients with β-hemoglobinopathies (Charache et al, 1992; Platt et al, 1984; Veith et al, 1985). Increased expression of fetal hemoglobin is of great importance in sickle cell disease and β-thalassemia because this event can mainly neutralize the adverse excess of α–chains (Charache et al, 1995). So-called "HU-responder" patients can become transfusion independent and HU can convert more severe patients to a milder phenotype (Voskaridou et al, 2010). β-thalassemia and sickle cell anemia were the first disorders treated with epigenetic therapy (Charache et al, 1983a; Ley et al, 1983). Two of the main classes of therapeutic agents are histone deacetylase (HDAC) or DNA methyltransferase inhibitors. HDAC inhibitors such as short chain fatty acids (e.g. butyrate & valproate) lead to elevated levels of histone acetylation thereby inducing fetal hemoglobin expression and erythropoiesis (Bohacek et al, 2006; Castaneda et al, 2005; Pace et al, 2002; Torkelson et al, 1996). 5-azacytidine is incorporated into DNA and then bound to DNA methyltransferase (DNMT) leading to depletion this enzyme and selective degradation DNMT1, resulting in re-expression of γ-globin genes (Charache et al, 1983b; Creusot et al, 1982; DeSimone et al, 1983; Dover et al, 1983; Ghoshal et al, 2005; Ley et al, 1982).

The Locus Control Region

Locus control regions (LCRs) are defined by their ability to drive the expression of the linked genes independent of the position of integration in the genome resulting in copy number dependent expression. The first LCR was identified to regulate in the human β-globin locus (Fraser & Grosveld, 1998; Grosveld et al, 1987; Talbot et al, 1989) comprises 4 DNase I hypersensetive sites (HSs). Deletion of the LCR from the mouse genome resulted in a β-globin expression level at 1–4% of that of has found in wild type cells, which led to the suggestion that the LCR - at least in part - works though activation of transcriptional elongation (Bender et al, 2000; Schubeler et al, 2001).

The LCR is composed of several small elements that show high sensitivity to DNase I digestion (hypersensitive sites or HS). They are located between 6 to 22 kb upstream of the ε-globin gene and are required for high level expression at all developmental stages (Forrester et al, 1986; Fraser & Grosveld, 1998; Tuan et al, 1985). The β-globin locus remains DNase I resistant in cells in which globin genes are not expressed, while in erythroid cells they exhibit a high degree of sensitivity to DNase I (Bulger & Groudine, 1999) that changes during different developmental stages and erythropoiesis in erythroid cells (Gribnau et al, 2000). Transcription factors bind to the HSs of LCR for their activity. The core elements of HS are 200-400 bp in size that encompass binding sites for erythroid specific and ubiquitous transcription factors. The core elements are separated each other by more than 2 kb (Philipsen et al, 1990b). The enhancer activity of the LCR resides in 5'HS2, 3, 4, while 5'HS1 or 5 show very low activity (Fraser et al, 1990; Hardison et al, 1997; Tuan et al, 1989). Each HS contains one or several binding motif for GATA-1, TAL-1, NF-E2, KLF1, SP1, BRG-1 and a number of other DNA binding proteins (Gillemans et al, 1998; Goodwin et al, 2001; Im et al, 2005; Martin & Orkin, 1990; Ney et al, 1990a; pruzina et al, 1991; Talbot et al, 1991). Each HS has a different developmental specificity interact with other HSs (Ellis et al, 1996) and also interact with the globin genes. To determine the developmental influence of the HS, each HS region was introduced in combination with γ- and β-globin genes into transgenic mice. This showed that HS1 plays a small role during development. HS2 contributes equally to γ or β transgene expression throughout development. HS3 is the most active hypersensitive site during the embryonic/fetal stage and remains active in the adult. HS4 contributes to expression of the locus at low level during embryonic and fetal stage but is the most active site at β-globin expression during adult stage (Fraser et al, 1993b). Deletion of HS2 from the LCR resulted in a minor decrease of ε-, γ- and β-globin expression at all stages of development, while deletion of HS3 resulted in major decrease of ε-globin gene expression and an increase of γ-globin gene expression in embryonic stage and doesn't affect the timing of globin gene switching. Deletion of only the 200-300 bp of core regions of HS3 or HS4 has a big effect on expression of globin genes (Bungert et al, 1995; Ellis et al, 1993; Fraser et al, 1993b; Hug et al, 1996; Peterson et al, 1996; Philipsen et al, 1990a; Talbot et al, 1990). GATA consensus sites are present in all regulatory elements of the globin genes. Many transcription-binding sites were found within the core of the **HS2** that including GATA-1 and SP1 (which are important for erythroid specificity of HS2), KLF1 and TEF2/BKLF – the latter three binding to the same sequences in the LCR - which all the motifs are found in the promoter and enhancer of each of β-like globin gene. Moreover, a conserved motif (TGCTGA(C/G)TCA(T/C) in HS2 is essential for

powerful enhancer activity that is associated to NF-E2/Ap-1 binding YY1 and USF binding sites are essential for full HS2 activity (Caterina et al, 1991; Ney et al, 1990b; Talbot & Grosveld, 1991). The core region (225 bp) of **HS3** contains a triple repeat of GATA sites in combination with a CACC box and appears to be the only element that is able to activate the globin genes when present as a single copy locus in the absence of the other HS elements (Ellis et al, 1996; Philipsen et al, 1993; Philipsen et al, 1990b). KLF1 is an active factor in HS3, which induces DNase I sensitivity to this region. The core region of HS4 (280 bp) contains binding sites for erythroid specific transcription factors as GATA-1, NF-E2, SP1 and ubiquitous factors including TEF2/BKLF (Pruzina et al, 1991). HS4 is not able to activate the linked β-globin gene as a single copy transgene (Fig. 9).

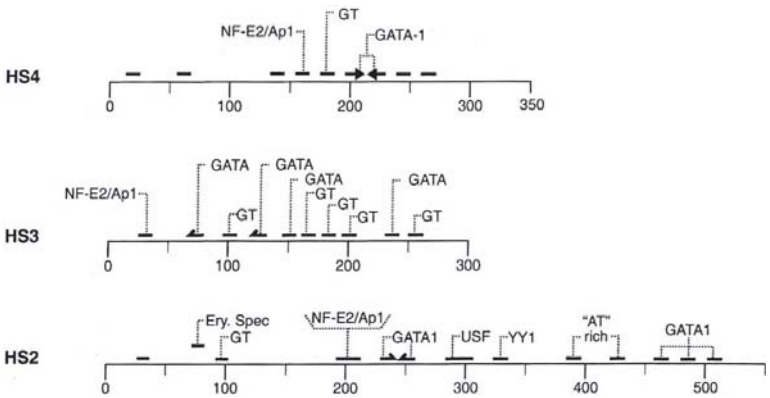


Figure 9. The position of various transcription factor-binding sites within HS2, HS3, HS4 (Stamatoyannopoulos and Grosveld 2001).

Transcriptional regulation of the β-globin locus

Trans-acting factors are proteins that function as DNA binding, transcriptional activator/repressor or protein interaction (Mitchell & Tjian, 1989). Most transcription factors act as part of a larger complex, which can function as an activator or repressor.

GATA-1

GATA-1 was identified as the first transcription factor in erythroid cells. GATA-1 is a zinc finger erythroid specific activator that recognizes the consensus sequence (A/T)GATA(A/G) found throughout β-globin locus (Martin & Orkin, 1990). GATA-1 contains two conserved zinc finger

domains. The C-terminal zinc finger is involved in DNA binding, while the N-terminal zinc finger domain is essential for protein-protein interaction (Blobel et al, 2001). It is expressed in megakaryocytic, mast, dendritic, erythroid, eosinophilic cells and at basal levels in hematopoietic progenitors (Evans & Felsenfeld, 1989; Sposi et al, 1992; Stachura et al, 2006). It mostly activates erythroid specific genes in erythroid cells but also represses genes associated with the immature and proliferative state including GATA-2 and c-kit cytokine receptor (Munugalavadla et al, 2005; Weiss et al, 1994; Welch et al, 2004). GATA-1 can force differentiation from hematopoietic cells to red blood cells (Evans & Felsenfeld, 1989; Fujiwara et al, 1996; Pevny et al, 1995; Pevny et al, 1991; Simon et al, 1992). GATA-1 overexpression induces the proliferation of pro-erythroid cells and inhibits the differentiation into mature erythroid cells (Whyatt & Grosveld, 2002). The consensus sequence of GATA-1 is often found in close proximity to that of KLF1 and SP1. GATA-1 binds to the LCR DNase I hypersensitive sites and gene promoters in β-globin locus (Johnson et al, 2002; Kim et al, 2007). It forms a number of different complexes most notably FOG-1 (Tsang et al, 1997), LMO-2 (Osada et al, 1995)/TAL-1, Klf1/SP1 (Merika & Orkin, 1995), CBP/P300 (Blobel et al, 1998). GATA-1 interacts with GATA-1 interacting factor FOG-1 (Crispino et al, 1999; Tsang et al, 1997), TAL-1 complex (Wadman et al, 1997), the repressive MeCP1 complex that FOG-1 serves as the bridging factor, the chromatin remodeling ACF/WCRF, the essential hematopoietic factor Gfi-b, distinctly (Rodriguez et al, 2005). GATA-1 forms several distinct complexes with the ACF/WCRF complex, FOG-1, FOG-1 and MeCP-1, TAL-1/Ldb-1 and Gfi-1b.

FOG-1

The first factor that was found to interact with GATA-1 was FOG-1. It is a multiple zinc finger protein that recruits other transcription factors to GATA-1 but does not bind to DNA directly itself (Chang et al, 2002). Both GATA-1 and FOG-1 are important factors in erythroid and megakaryocytic differentiation (Fujiwara et al, 1996; Pevny et al, 1991; Tsang et al, 1998). FOG-1 recruits components of NuRD complex like histone deacetylase, providing a mechanism for GATA-1-dependent gene repression (Hong et al, 2005; Rodriguez et al, 2005).

SCL/TAL1, LMO2, Ldb1/NLI1, E2A

A protein complex composed of SCL/TAL1, LMO2 (LIM protein domain), Ldb1/NLI1 (LIM domain interacting protein), E2A can associate to GATA-1 (Wadman et al, 1997). The E2A/TAL-1 heterodimer is bound to the E-box sequence, CANNTG, which are separated by 9 to 12 bp from GATA site (Soler

et al 2010). The complex composed of all five proteins binds to Gata/E-box motif through GATA-1 and E2A/TAL-1 heterodimer respectively. (Goardon et al, 2006; Wadman et al, 1997; Soler et al 2010). TAL-1 is a basic helix loop helix transcription factor that can function both as activator and repressor (Schuh et al, 2005). The repressor function of TAL-1 is primarily controlled through its interaction with ETO-2 (Soler et al, 2010; Tripic et al, 2009). The LDB-1 complex is essential to establish the interaction of the β-globin gene and the LCR to provide transcriptional activation during erythroid differentiation (Song et al, 2010; Xu et al, 2003).

EKLF/KLF1

Erythroid Krüppel-like factor (ELKF/KLF1) is an erythroid specific transcription factor that is expressed throughout all developmental stages and is critical for erythropoiesis (Miller & Bieker, 1993; Nuez et al, 1995; Perkins et al, 1995). It contains a N-terminal proline rich transactivation domain and three C-terminal zinc finger domains that recognize CCNCNCCCN motif on DNA. The β-globin gene promoter binds KLF1 and mutations in the CACCC box lead to β-thalassemia (Feng et al, 1994);Bieker & Southwood, 1995). KLF1 in combination with GATA-1 also binds at three consecutive sites in HS3 which is probably responsible for its unique property to provide copy number dependent expression in a single copy constuct (Gillemans et al, 1998). KLF1 is a bi-functional protein (transcriptional activator/repressor) that is characterized by post-translational modifications that promote protein-protein interaction. The C-terminus of KLF1 is acetylated on lysine 288 by CBP/P300 and phosphorylated on threonine 41 by casein kinase II resulting in an interaction of KLF1 with a SWI/SNF related complex (E-RC1) through BRG-1 (Armstrong et al, 1998; Bottardi et al, 2006; Brown et al, 2002; Ouyang et al, 1998; Sengupta et al, 2008; Zhang et al, 2001). Repression activity of KLF1 is associated with acetylation of lysine 302, enabling its interaction with the corepressors HDAC-1 and SIN3A and sumoylation of lysine 74, that enables its interaction with HDAC-1 and Mi2-β (Chen & Bieker, 2001; Chen & Bieker, 2004; Siatecka et al, 2007). KLF1 binds specifically to adult globin gene promoters and also to HS1, HS2, HS3 and HS4 throughout all development stages. The expression level of KLF1 increases during development (Zhou et al, 2006). Importantly KLF1 activates the BCL11a gene which has been identified as γ-globin repressor (Sankaran et al, 2008). Low levels of KLF1 during embryonic and fetal stages results in low expression level of β-globin and BCL11a and a high level of γ-globin expression (Borg et al 2011). A high expression level of KLF1 in adult stage promotes a high level of β-globin and BCL11a (Bieker, 2010).

NF-E2

NF-E2 is an erythroid specific transcription factor (Romeo et al, 1990). NF-E2 interacts with Maf recognition elements (MARE) within HS2 and HS3 sub-regions of β-globin locus LCR directly (Forsberg et al, 2000; Kang et al, 2002; Talbot & Grosveld, 1991). NF-E2 contains two distinct domains: the hematopoietic subunit P45 (bZip) and ubiquitous subunit P18 (MafK). The P45 subunit interacts with CBP and P300, ubiquitin ligase and TBP-associated factor TAFII130 (Amrolia et al, 1997; Cheng et al, 1997). Interaction of the P45 domain of NF-E2 with TAFII130 mediates looping between the LCR and the globin genes promoters (Amrolia et al, 1997). The CREB-binding protein (CBP) and P300, which possess histone acetytransferase activity, are recruited by NF-E2 and GATA-1 to the LCR and the distant β-globin gene (Armstrong & Emerson, 1996; Johnson et al, 2001; Letting et al, 2003). NF-E2 cooperates with GATA-1 to recruit RNA polymerase II to the β-globin promoters (Johnson et al, 2002).

BCL11a

BCL11a is a C2H2 type zinc-finger transcription factor, which is associated with γ-globin silencing and hemoglobin switching. Down-regulation of BCL11a induces γ-globin expression in human adult erythroid progenitor cells. BCL11a interacts with GATA-1, FOG-1 and subunits of NuRD complex including CHD4, HDAC1, 2 and MTA2 in erythroid cells. The full-length isoforms of BCL11a (XL/L) are expressed in adult bone marrow and at a lower level in fetal liver. They are absent in primitive erythroblasts, while short variants of BCL11a are expressed in primitive erythroblasts and fetal liver (Sankaran et al, 2008). BCL11a is not expressed in primitive erythroid cell of mice, and only full-length isoforms of BCL11a are expressed in definitive erythroid cells of mice (fetal liver and adult bone marrow). BCL11a binds strongly to HS3, the embryonic ε-globin gene, and 3 kb downstream of the Aγ–globin gene. The pattern of the binding site BCL11a is the distinct from GATA-1 and SOX-6 (Cheng et al, 2009; Yu et al, 2009).

PYR complex

The poly primidine (PYR) complex, a chromatin-remodeling complex, which is presence in adult life, contains SWI/SNF (BAF) activator complex subunits and nucleosome-remodeling deacetylase (NuRD) repressor complex subunits like CHD4. The PYR complex requires the presence of Ikaros to interact with DNA (Lopez et al, 2002). The 250 bp PYR-binding DNA sequences are located 1 kb upstream of the human δ-globin gene. Deletion of an intergenic DNA-binding site of PYR complex from a human

β-globin locus construct resulted in delayed human γ- to β-globin switching in transgenic mice involving the PYR complex good candidate for human globin switching (O'Neill et al, 1991; O'Neill et al, 1999).

ZBP-89/ZNF148

The Krüppel-type transcription factor ZBP-89 is located on human chromosome 3q21 encompassing 9 exons. The promoter contains high GC content and lacks canonical CAAT and TATA boxes. Acidic domain that interacts with coactivator P300, synergy control (SC) domains serve as sites for sumolyation, basic, zinc finger, two more basic, serine-rich, and PEST domains are found in ZBP-89 protein (Bai et al., 2000; Bai et al., 2003; Chupreta et al., 2007; Feo et al., 2001). ZBP-89 binds to a GC-rich region, and represses or activates known target genes. When acting as a repressor, it has been proposed that ZBP-89 and Sp1 compete for binding to the same or overlapping GC-rich sequences (Merchant et al., 1996). ZBP-89 has been reported to interact with GATA1 and Mafk and it is involved in erythroid development and differentiation (Brand et al., 2004; Woo et al., 2011; Woo et al., 2008a). Knock down of ZBP-89 in zebrafish results in the "bloodless" phenotype due to disruption of both primitive and definitive erythropoiesis (Li et al., 2006). It has been proposed that ZBP-89 represses gene transcription via specific recruitment of HDAC1 to promoters of its target genes (Wu et al., 2007).

Epigenetic regulation

Epigenetic regulation plays an important role in gene expression and regulation. Chromatin is the complex of DNA and histone proteins, is important for the packaging of chromosal DNA and regulation of expression of the genome. Epigentic regulation comprises DNA methylation and histone modifications. The DNA methyltransferse enzymes DNMT3B and DNMT1 mediate DNA methylation. The position of the methylation plays a role in controling the expression of gene in the transcriptional unit. For example, methylation in transcription start site blocks initiation whereas in the gene body can even stimulate transcription elongation. In the globin gene cluster, DNA methylation is associated with inactive DNA, while hypomethylation is associated with gene activity. For example the γ-globin gene promoter becomes methylated during development, while the β-globin gene promoter remains unmethylated (Goren et al, 2006; van der Ploeg & Flavell, 1980).

Histone modifications are a number of post-translational modifications of histone amino acids like acetylation, phosphorylation, methylation and uniquitination. Different amino acids on different histones are modified differently and are associated with functions such as gene activity, repression or replication (Fig. 10).

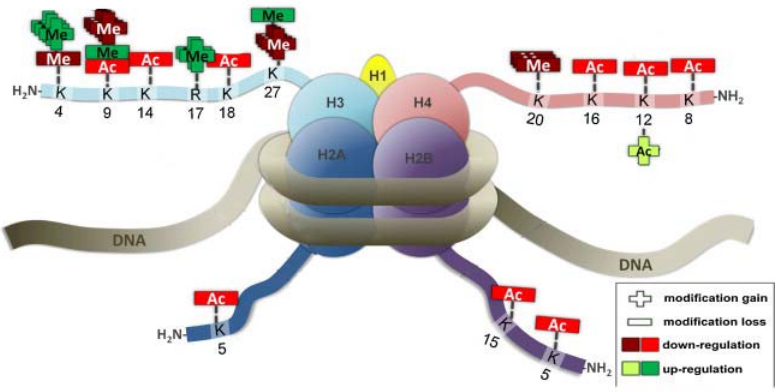


Figure 10. The histone modification marks (Gargalionis et al, 2012). The core nucleosomal structure with gain (cross) or loss (rectangular) of main acetylation (Ac) and methylation (Me) marks on specific histone residues. Up-regulating or down-regulating properties of these marks have been shown green and red respectively.

Chromatin remodellers such as the SWI/SNF complex modulate nucleosome positioning in an ATP dependant manner, which affects replication, transcription recombination and repair, while other enzymes such as histone acetyltransferase (HAT) and histone deacetyltransferse (HDAC) modify histones in an ATP independent manner. Some chromatin remodelling and deacetylase complexes like the NuRD complex possess both functions (Tong et al, 1998; Zhang et al, 1998). Chromatin remodelling modifies chromatin access to transcription factors or replication factors. Epigenetic regulation is also important in hemoglobin switching. Repressive complexes including the NuRD complex or arginine methyltransferase 5 that methylates histone H4 at position R3 can induce additional epigenetic marks at the promoter through recruitment of a multiprotein repressive complex (Rank et al, 2010; Zhao et al, 2009). Furthermore, PRMT1 has been linked to β-globin

activation. Its knockdown resulted in loss of H4R3me at the chicken β-globin gene, loss of acetylation on both H3 and H4 and also an increased in marks associated with inactive chromatin (H3K9 and H3K27) (Huang et al, 2005). Chromatin Target of Prmt1 (Chtop) expression rises during development and its knockdown resulted in depression of γ-globin (van Dijk et al, 2010).

Active chromatin hub (ACH) in globin regulation

The eukaryotic enhancers can be located far away from the gene they regulate. Long-range interaction between promoter and gene regulatory elements can activate gene expression via form a spatial organization of chromatin loop. Proteins bound to enhancers interact directly with proteins bound to promoters with intervening DNA looping out. The proximity between the promoter and the enhancer assembles the pre-initiation RNA polymerase complex on the promoter (Cook, 2003; Drissen et al, 2004a; Fraser, 2006; Jing et al, 2008; Kadauke & Blobel, 2009; Liu & Garrard, 2005; Tolhuis et al, 2002; Vilar & Saiz, 2005). The five HSs of the LCR interact with extra 5' and 3' HSs to form an ACH in the β-globin locus (Palstra et al, 2003) converting it to an active chromatin structure (Forrester et al, 1990; Kioussis et al, 1983). The intervening chromatin and inactive genes loop out. The β-globin gene remains in proximity to the ACH when the β-globin gene promoter is deleted and additional deletion of HS2 or HS3 shows that HS3 in combination of β-globin promoter is essential for maintenance of the ACH in definitive erythroid cells. These results show that the presence of HSs, but not globin promoters, is important for ACH formation (Patrinos et al, 2004)). Chromatin hub interacts with active embryonic and fetal globin genes and forms ACH in primitive stage. During development in definitive stage, it switches to adult globin genes (Fig. 11) (Palstra et al, 2003). Some of the proteins involved in the β-globin LCR interaction are erythroid specific transcriptional activators including KLF1, GATA-1, NF-E2, but also co-activators like Brg1, NLI/Ldb1 and insulator related proteins such as CTCF/cohesin (Drissen et al, 2004b; Song et al, 2007; Song et al, 2010; Vakoc et al, 2005). CTCF is involved in chromatin architecture and deletion of CTCF can destabilize long-range interactions and cause a change in active and repressive histone marks. The cohesin complex interacts with CTCF and co-occupies CTCF sites in the chromatin of the β–globin locus (Hou et al, 2010; Wendt & Peters, 2009). The cohesion complex consists of Smc1, Smc3, Rad21 and SA, which is loaded onto to chromatin through the loading factor (NIBL-1). A deficiency of cohesin or NIPBL-1 disrupts the LCR enhancer-promoter interactions and inhibits gene expression (Chien et al, 2011). The NF-E2 subunit (P18) is thought to interact with cohesion subunits (Rad21 and Smc1)(Brand et al, 2004).

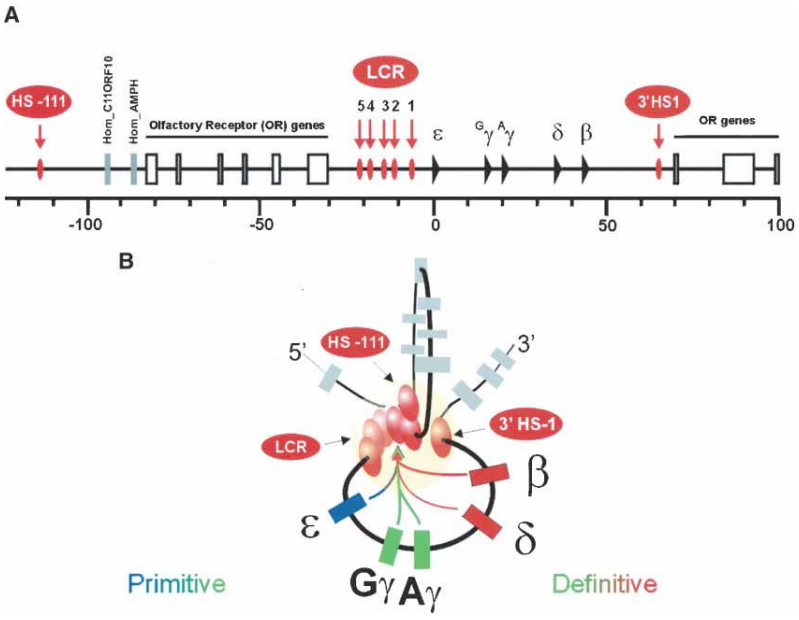


Figure 11. A. Location of human β-globin locus elements, B. The 3-D interactions between the HSs and the globin genes in erythroid cells (Patrinos et al 2004). The five HSs (The most 5' and 3' HSs) of the LCR form the chromatin hub, with which the globin genes interact to form the ACH resulting in high-levels of transcription.

Aim of the thesis

The aim of this thesis was to further investigate the proteins that are involved in hemoglobin switching with the eventual goal to be able to manipulate the repression of the γ-globin genes in adult β-thalassemia or sickle cell disease patients. Activation of the γ-globin genes in such patients would lead to the amelioration of the disease. The first part of this work was initiated by Dr Farzin Pourfarzad who developed a method to directly isolate the proteins that are bound to the β-globin chromatin hub, when the γ-globin genes are repressed during development. This resulted in the identification of a number of novel proteins that appear to be involved in the switching process including the transcription factor ZBP-89/ZNF148. The second part of the work describes the function of this protein in the switching process.

REFERENCES

Aluoch JR (1997) Higher resistance to Plasmodium falciparum infection in patients with homozygous sickle cell disease in western Kenya. *Trop Med Int Health* **2:** 568-571

Amrolia PJ, Ramamurthy L, Saluja D, Tanese N, Jane SM, Cunningham JM (1997) The activation domain of the enhancer binding protein p45NF-E2 interacts with TAFII130 and mediates long-range activation of the alpha- and beta-globin gene loci in an erythroid cell line. *Proc Natl Acad Sci U S A* **94:** 10051-10056

Armstrong JA, Bieker JJ, Emerson BM (1998) A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95:** 93-104

Armstrong JA, Emerson BM (1996) NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 in vitro. *Mol Cell Biol* **16:** 5634-5644

Bai, L., and J. L. Merchant (2000) Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells. *J Biol Chem* **275**:30725-33

Bai, L., and J. L. Merchant (2003) Transcription factor ZBP-89 is required for STAT1 constitutive expression. *Nucleic Acids Res* **31**:7264-70

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

Bender MA, Bulger M, Close J, Groudine M (2000) Œ≤-globin Gene switching and DNase I sensitivity of the endogenous Œ≤-globin locus in mice do not require the locus control region. *Molecular Cell* **5:** 387-393

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.

Bieker JJ (2010) Putting a finger on the switch. *Nat Genet* **42:** 733-734

Blobel GA, Nakajima T, Eckner R, Montminy M, Orkin SH (1998) CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* **95:** 2061-2066

Blobel GA, Weiss MJ (2001) Nuclear factors that regulate erythropoiesis. In Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management, Steinberg MH, Forget BG.

Bohacek R, Boosalis MS, McMartin C, Faller DV, Perrine SP (2006) Identification of novel small-molecule inducers of fetal hemoglobin using pharmacophore and 'PSEUDO' receptor models. *Chem Biol Drug Des* **67:** 318-328

Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, Cassar W, Galdies R, van Ijcken W, Ozgur Z, Gillemans N, Hou J, Bugeja M, Grosveld FG, von Lindern M, Felice AE, Patrinos GP, Philipsen S (2010) Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet* **42:** 801-805

Bottardi S, Ross J, Pierre-Charles N, Blank V, Milot E (2006) Lineage-specific activators affect beta-globin locus chromatin in multipotent hematopoietic progenitors. *EMBO J* **25:** 3586-3595

Brand M, Ranish JA, Kummer NT, Hamilton J, Igarashi K, Francastel C, Chi TH, Crabtree GR, Aebersold R, Groudine M (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11:** 73-80

Brown RC, Pattison S, van Ree J, Coghill E, Perkins A, Jane SM, Cunningham JM (2002) Distinct domains of erythroid Kruppel-like factor modulate chromatin remodeling and transactivation at the endogenous beta-globin gene promoter. *Mol Cell Biol* **22:** 161-170

Bulger M, Bender MA, Van Doorninck JH, Wertman B, Farrell CM, Felsenfeld G, Groudine M, Hardison R (2000) Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse Œ≤-globin gene clusters. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 14560-14565

Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13:** 2465-2477

Bulger M, Schubeler D, Bender MA, Hamilton J, Farrell CM, Hardison RC, Groudine M (2003) A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol Cell Biol* **23:** 5234-5244

Bulger M, Von Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M (1999) Conservation of sequence and structure flanking the mouse and human Œ≤- globin loci: The Œ≤-globin genes are embedded within an array of odorant receptor genes. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 5129-5134

Bungert J, Dave U, Lim KC, Lieuw KH, Shavit JA, Liu Q, Engel JD (1995) Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes & development* **9:** 3083-3096

Castaneda S, Boosalis MS, Emery D, Thies A, Faller DV, Perrine SP (2005) Enhancement of growth and survival and alterations in Bcl-family proteins in beta-thalassemic erythroid progenitors by novel short-chain fatty acid derivatives. *Blood Cells Mol Dis* **35:** 217-226

Caterina JJ, Ryan TM, Pawlik KM, Palmiter RD, Brinster RL, Behringer RR, Townes TM (1991) Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. *Proc Natl Acad Sci U S A* **88:** 1626-1630

Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K, Cavallesco R, Gillet-Legrand B, Caccavelli L, Sgarra R, Maouche-Chretien L, Bernaudin F, Girot R, Dorazio R, Mulder GJ, Polack A, Bank A, Soulier J, Larghero J, Kabbara N, Dalle B, Gourmel B, Socie G, Chretien S, Cartier N, Aubourg P, Fischer A, Cornetta K,

Galacteros F, Beuzard Y, Gluckman E, Bushman F, Hacein-Bey-Abina S, Leboulch P (2010) Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467:** 318-322

Chakalova L, Osborne CS, Dai YF, Goyenechea B, Metaxotou-Mavromati A, Kattamis A, Kattamis C, Fraser P (2005) The Corfu deltabeta thalassemia deletion disrupts gamma-globin gene silencing and reveals post-transcriptional regulation of HbF expression. *Blood* **105:** 2154-2160

Chang AN, Cantor AB, Fujiwara Y, Lodish MB, Droho S, Crispino JD, Orkin SH (2002) GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci U S A* **99:** 9237-9242

Chang YC, Smith KD, Moore RD, Serjeant GR, Dover GJ (1995) An analysis of fetal hemoglobin variation in sickle cell disease: the relative contributions of the X-linked factor, beta-globin haplotypes, alpha-globin gene number, gender, and age. *Blood* **85:** 1111-1117

Chang YP, Maier-Redelsperger M, Smith KD, Contu L, Ducroco R, de Montalembert M, Belloy M, Elion J, Dover GJ, Girot R (1997) The relative importance of the X-linked FCP locus and beta-globin haplotypes in determining haemoglobin F levels: a study of SS patients homozygous for beta S haplotypes. *Br J Haematol* **96:** 806-814

Charache S, Dover G, Smith K, Talbot CC, Jr., Moyer M, Boyer S (1983a) Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the gamma-delta-beta-globin gene complex. *Proc Natl Acad Sci U S A* **80:** 4842-4846

Charache S, Dover G, Smith K, Talbot CC, Jr., Moyer M, Boyer S (1983b) Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the gamma-delta-beta-globin gene complex. *Proceedings of the National Academy of Sciences of the United States of America* **80:** 4842-4846

Charache S, Dover GJ, Moore RD, Eckert S, Ballas SK, Koshy M, Milner PF, Orringer EP, Phillips G, Jr., Platt OS, et al. (1992) Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia. *Blood* **79:** 2555-2565

Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, McMahon RP, Bonds DR (1995) Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. *The New England journal of medicine* **332:** 1317-1322

Chen X, Bieker JJ (2001) Unanticipated repression function linked to erythroid Kruppel-like factor. *Mol Cell Biol* **21:** 3118-3125

Chen X, Bieker JJ (2004) Stage-specific repression by the EKLF transcriptional activator. *Mol Cell Biol* **24:** 10416-10424

Cheng X, Reginato MJ, Andrews NC, Lazar MA (1997) The transcriptional integrator CREB-binding protein mediates positive cross talk between nuclear hormone receptors and the hematopoietic bZip protein p45/NF-E2. *Mol Cell Biol* **17:** 1407-1416

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

Chien R, Zeng W, Kawauchi S, Bender MA, Santos R, Gregson HC, Schmiesing JA, Newkirk DA, Kong X, Ball AR, Jr., Calof AL, Lander AD, Groudine MT, Yokomori K (2011) Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J Biol Chem* **286:** 17870-17878

Chupreta, S., H. Brevig, L. Bai, J. L. Merchant, and J. A. Iniguez-Lluhi (2007) Sumoylation-dependent control of homotypic and heterotypic synergy by the Kruppel-type zinc finger protein ZBP-89. *J Biol Chem* **282**:36155-66

Cook PR (2003) Nongenic transcription, gene regulation and action at a distance. *Journal of Cell Science* **116:** 4483-4491

Creusot F, Acs G, Christman JK (1982) Inhibition of DNA methyltransferase and induction of Friend erythroleukemia cell differentiation by 5-azacytidine and 5-aza-2'-deoxycytidine. *The Journal of biological chemistry* **257:** 2041-2048

Crispino JD, Lodish MB, MacKay JP, Orkin SH (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

DeSimone J, Heller P, Schimenti JC, Duncan CH (1983) Fetal hemoglobin production in adult baboons by 5-azacytidine or by phenylhydrazine-induced hemolysis is associated with hypomethylation of globin gene DNA. *Prog Clin Biol Res* **134:** 489-500

Dover GJ, Charache SH, Boyer SH, Talbot CC, Jr., Smith KD (1983) 5-Azacytidine increases fetal hemoglobin production in a patient with sickle cell disease. *Prog Clin Biol Res* **134:** 475-488

Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W (2004a) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & development* **18:** 2485-2490

Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W (2004b) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* **18:** 2485-2490

Eichner ER (2007) Sickle cell trait. *J Sport Rehabil* **16:** 197-203

Ellis J, Talbot D, Dillon N, Grosveld F (1993) Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *The EMBO journal* **12:** 127-134

Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipsen S, Grosveld F (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *EMBO J* **15:** 562-568

Evans T, Felsenfeld G (1989) The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell* **58:** 877-885

Farrell CM, West AG, Felsenfeld G (2002) Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* **22:** 3820-3831

Feng WC, Southwood CM, Bieker JJ (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* **269:** 1493-1500

Feo, S., V. Antona, G. Cammarata, F. Cavaleri, R. Passantino, P. Rubino, and A. Giallongo (2001) Conserved structure and promoter sequence similarity in the mouse and human genes encoding the zinc finger factor BERF-1/BFCOL1/ZBP-89. *Biochem Biophys Res Commun* **283**:209-18

Flint J, Harding RM, Boyce AJ, Clegg JB (1993) The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol* **6:** 215-262

Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes & development* **4:** 1637-1649

Forrester WC, Thompson C, Elder JT, Groudine M (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* **83:** 1359-1363

Forsberg EC, Downs KM, Bresnick EH (2000) Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells. *Blood* **96:** 334-339

Fraser P (2006) Transcriptional control thrown for a loop. *Curr Opin Genet Dev* **16:** 490-495

Fraser P, Grosveld F (1998) Locus control regions, chromatin activation and transcription. *Current Opinion in Cell Biology* **10:** 361-365

Fraser P, Hurst J, Collis P, Grosveld F (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* **18:** 3503-3508

Fraser P, Pruzina S, Antoniou M, Grosveld F (1993a) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* **7:** 106-113

Fraser P, Pruzina S, Antoniou M, Grosveld F (1993b) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes & development* **7:** 106-113

Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93:** 12355-12358

Galanello R and Origa R (2010) beta-thalassemia. *Orphanet Journal of Rare Diseases* 2010, **5**:11

Gargalionis AN, Piperi C, Adamopoulos C, Papavassiliou AG (2012) Histone modifications as a pathogenic mechanism of colorectal tumorigenesis. *Int J Biochem Cell Biol*. **44(8)**: 1276-89.

Garner CP, Tatu T, Best S, Creary L, Thein SL (2002) Evidence of genetic interaction between the beta-globin complex and chromosome 8q in the expression of fetal hemoglobin. *Am J Hum Genet* **70:** 793-799

Ghoshal K, Datta J, Majumder S, Bai S, Kutay H, Motiwala T, Jacob ST (2005) 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Molecular and cellular biology* **25:** 4727-4741

Gibney GT, Panhuysen CI, So JC, Ma ES, Ha SY, Li CK, Lee AC, Yuen HL, Lau YL, Johnson DM, Farrell JJ, Bisbee AB, Farrer LA, Steinberg MH, Chan LC, Chui DH (2008) Variation and heritability of Hb F and F-cells among beta-thalassemia heterozygotes in Hong Kong. *Am J Hematol* **83:** 458-464

Gillemans N, Tewari R, Lindeboom F, Rottier R, de Wit T, Wijgerde M, Grosveld F, Philipsen S (1998) Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the *β-globin* locus control region in vivo. *Genes Dev*. **12(18)**: 2863–2873.

Goardon N, Lambert JA, Rodriguez P, Nissaire P, Herblot S, Thibault P, Dumenil D, Strouboulis J, Romeo PH, Hoang T (2006) ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J* **25:** 357-366

Goodwin AJ, McInerney JM, Glander MA, Pomerantz O, Lowrey CH (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

Goren A, Simchen G, Fibach E, Szabo PE, Tanimoto K, Chakalova L, Pfeifer GP, Fraser PJ, Engel JD, Cedar H (2006) Fine tuning of globin gene expression by DNA methylation. *PLoS One* **1:** e46

Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P (2000) Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* **5:** 377-386

Grosso M, Amendolara M, Rescigno G, Danise P, Todisco N, Izzo P, Amendola G (2008) Delayed decline of gamma-globin expression in infant age associated with the presence of Ggamma-158 (C-->T) polymorphism. *Int J Lab Hematol* **30:** 191-195

Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human ≤-globin gene in transgenic mice. *Cell* **51:** 975-985

Gumucio DL, Rood KL, Gray TA, Riordan MF, Sartor CI, Collins FS (1988) Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. *Mol Cell Biol* **8:** 5310-5322

Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205:** 73-94

Higgs DR, Wood WG, Jarman AP, Sharpe J, Lida J, Pretorius IM, Ayyub H (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev* **4:** 1588-1601

Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA (2005) FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J* **24:** 2367-2378

Hou C, Dale R, Dean A (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* **107:** 3651-3656

Hoyer JD, Penz CS, Fairbanks VF, Hanson CA, Katzmann JA (2002) Flow cytometric measurement of hemoglobin F in RBCs: diagnostic usefulness in the distinction of hereditary persistence of fetal hemoglobin (HPFH) and hemoglobin S-hPFH from other conditions with elevated levels of hemoglobin F. *Am J Clin Pathol* **117:** 857-863

Huang S, Litt M, Felsenfeld G (2005) Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes Dev* **19:** 1885-1893

Hug BA, Wesselschmidt RL, Fiering S, Bender MA, Epner E, Groudine M, Ley TJ (1996) Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Molecular and cellular biology* **16:** 2906-2912

Ikuta T, Kan YW (1991) In vivo protein-DNA interactions at the beta-globin gene locus. *Proc Natl Acad Sci U S A* **88:** 10188-10192

Im H, Grass JA, Johnson KD, Kim SI, Boyer ME, Imbalzano AN, Bieker JJ, Bresnick EH (2005) Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc Natl Acad Sci U S A* **102:** 17065-17070

Ingram VM (1956) A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* **178:** 792-794

Jane SM, Ney PA, Vanin EF, Gumucio DL, Nienhuis AW (1992) Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *EMBO J* **11:** 2961-2969

Jing H, Vakoc CR, Ying L, Mandat S, Wang H, Zheng X, Blobel GA (2008) Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular cell* **29:** 232-242

Johnson KD, Christensen HM, Zhao B, Bresnick EH (2001) Distinct Mechanisms Control RNA Polymerase II Recruitment to a Tissue-Specific Locus Control Region and a Downstream Promoter. *Molecular Cell* **8:** 465-471

Johnson KD, Grass JA, Boyer ME, Kiekhaefer CM, Blobel GA, Weiss MJ, Bresnick EH (2002) Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A* **99:** 11760-11765

Joly P, Lacan P, Garcia C, Couprie N, Francina A (2009) Identification and molecular characterization of four new large deletions in the beta-globin gene cluster. *Blood Cells Mol Dis.* 43(1):53-7

Kadauke S, Blobel GA (2009) Chromatin loops in gene regulation. *Biochim Biophys Acta* **1789:** 17-25

Kang SH, Vieira K, Bungert J (2002) Combining chromatin immunoprecipitation and DNA footprinting: a novel method to analyze protein-DNA interactions in vivo. *Nucleic Acids Res* **30:** e44

Kark JA, Posey DM, Schumacher HR, Ruehle CJ (1987) Sickle-cell trait as a risk factor for sudden death in physical training. *The New England journal of medicine* **317:** 781-787

Kim A, Song SH, Brand M, Dean A (2007) Nucleosome and transcription activator antagonism at human beta-globin locus control region DNase I hypersensitive sites. *Nucleic Acids Res* **35:** 5831-5838

Kioussis D, Vanin E, deLange T, Flavell RA, Grosveld FG (1983) Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306:** 662-666

Krishnamurti L, Bunn HF, Williams AM, Tolar J (2008) Hematopoietic Cell Transplantation for Hemoglobinopathies. *Current Problems in Pediatric and Adolescent Health Care* **38:** 6-18

Labie D, Pagnier J, Lapoumeroulie C, Rouabhi F, Dunda-Belkhodja O, Chardin P, Beldjord C, Wajcman H, Fabry ME, Nagel RL (1985) Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. *Proc Natl Acad Sci U S A* **82:** 2111-2114

Letting DL, Rakowski C, Weiss MJ, Blobel GA (2003) Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Molecular and Cellular Biology* **23:** 1334-1340

Lettre G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH (2008) DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A* **105:** 11869-11874

Ley TJ, DeSimone J, Anagnou NP, Keller GH, Humphries RK, Turner PH, Young NS, Keller P, Nienhuis AW (1982) 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. *N Engl J Med* **307:** 1469-1475

Ley TJ, DeSimone J, Noguchi CT, Turner PH, Schechter AN, Heller P, Nienhuis AW (1983) 5-Azacytidine increases gamma-globin synthesis and reduces the proportion of dense cells in patients with sickle cell anemia. *Blood* **62:** 370-380

Liebhaber SA, Cash FE, Ballas SK (1986) Human alpha-globin gene expression. The dominant role of the alpha 2-locus in mRNA and protein synthesis. *J Biol Chem* **261:** 15327-15333

Liu LR, Du ZW, Zhao HL, Liu XL, Huang XD, Shen J, Ju LM, Fang FD, Zhang JW (2005) T to C substitution at -175 or -173 of the gamma-globin promoter affects GATA-1 and Oct-1 binding in vitro differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. *J Biol Chem* **280:** 7452-7459

Liu Z, Garrard WT (2005) Long-range interactions between three transcriptional enhancers, active Vkappa gene promoters, and a 3' boundary sequence spanning 46 kilobases. *Molecular and cellular biology* **25:** 3220-3231

Lopez RA, Schoetz S, DeAngelis K, O'Neill D, Bank A (2002) Multiple hematopoietic defects and delayed globin switching in Ikaros null mice. *Proc Natl Acad Sci U S A* **99:** 602-607

Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* **118:** 1590-1605

Mantovani R, Malgaretti N, Giglioni B, Comi P, Cappellini N, Nicolis S, Ottolenghi S (1987) A protein factor binding to an octamer motif in the gamma-globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells. *Nucleic Acids Res* **15:** 9349-9364

Mantovani R, Malgaretti N, Nicolis S, Ronchi A, Giglioni B, Ottolenghi S (1988) The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors. *Nucleic Acids Res* **16:** 7783-7797

Martin DI, Orkin SH (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev* **4:** 1886-1898

Martin DI, Tsai SF, Orkin SH (1989) Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* **338:** 435-438

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9:** 356-369

McGrath KE, Kingsley PD, Koniski AD, Porter RL, Bushnell TP, Palis J (2008) Enucleation of primitive erythroid cells generates a transient population of "pyrenocytes" in the mammalian fetus. *Blood* **111:** 2409-2417

Merika M, Orkin SH (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* **15:** 2437-2447

Miller IJ, Bieker JJ (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* **13:** 2776-2786

Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371-378

Munugalavadla V, Dore LC, Tan BL, Hong L, Vishnu M, Weiss MJ, Kapur R (2005) Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation. *Mol Cell Biol* **25:** 6747-6759

Ney PA, Sorrentino BP, Lowrey CH, Nienhuis AW (1990a) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res* **18:** 6011-6017

Ney PA, Sorrentino BP, McDonagh KT, Nienhuis AW (1990b) Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* **4:** 993-1006

Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375:** 316-318

O'Neill D, Bornschlegel K, Flamm M, Castle M, Bank A (1991) A DNA-binding factor in adult hematopoietic cells interacts with a pyrimidine-rich domain upstream from the human delta-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 8953-8957

O'Neill D, Kaysen J, Donovan-Peluso M, Castle M, Bank A (1990) Protein-DNA interactions upstream from the human A gamma globin gene. *Nucleic Acids Res* **18:** 1977-1982

O'Neill D, Yang J, Erdjument-Bromage H, Bornschlegel K, Tempst P, Bank A (1999) Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. *Proc Natl Acad Sci U S A* **96:** 349-354

Osada H, Grutz G, Axelson H, Forster A, Rabbitts TH (1995) Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci U S A* **92:** 9585-9589

Ouyang L, Chen X, Bieker JJ (1998) Regulation of erythroid Kruppel-like factor (EKLF) transcriptional activity by phosphorylation of a protein kinase casein kinase II site within its interaction domain. *J Biol Chem* **273:** 23019-23025

Pace BS, White GL, Dover GJ, Boosalis MS, Faller DV, Perrine SP (2002) Short-chain fatty acid derivatives induce fetal globin expression and erythropoiesis in vivo. *Blood* **100:** 4640-4648

Palis J (2008) Ontogeny of erythropoiesis. *Curr Opin Hematol* **15:** 155-161

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

Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, Grosveld FG (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.***18(12**):1495-509.

Pauling L, Itano HA, et al. (1949) Sickle cell anemia a molecular disease. *Science* **110:** 543-548

Perkins AC, Sharpe AH, Orkin SH (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375:** 318-322

Peterson KR, Clegg CH, Navas PA, Norton EJ, Kimbrough TG, Stamatoyannopoulos G (1996) Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 6605-6609

Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, Costantini F (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121:** 163-172

Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, Orkin SH, Costantini F (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349:** 257-260

Philipsen S, Pruzina S, Grosveld F (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *EMBO J* **12:** 1077-1085

Philipsen S, Talbot D, Fraser P, Grosveld F (1990a) The beta-globin dominant control region: hypersensitive site 2. *The EMBO journal* **9:** 2159-2167

Philipsen S, Talbot D, Fraser P, Grosveld F (1990b) The beta-globin dominant control region: hypersensitive site 2. *EMBO J* **9:** 2159-2167

Platt OS, Orkin SH, Dover G, Beardsley GP, Miller B, Nathan DG (1984) Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. *The Journal of clinical investigation* **74:** 652-656

Powars DR, Weiss JN, Chan LS, Schroeder WA (1984) Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? *Blood* **63:** 921-926

Proudfoot NJ (1977) Nucleotide sequence from the coding region of rabbit beta-globin messenger RNA. *Nucleic Acids Res*. **3(7)**:1811-21.

Pruzina S, Hanscombe O, Whyatt D, Grosveld F, Philipsen S (1991) Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Res* **19:** 1413-1419

Rabbitts TH (1976) Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA. *Nature* **260:** 221-225

Rank G, Cerruti L, Simpson RJ, Moritz RL, Jane SM, Zhao Q (2010) Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. *Blood* **116:** 1585-1592

Rochette J, Craig JE, Thein SL (1994) Fetal hemoglobin levels in adults. *Blood Rev* **8:** 213-224

Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **24:** 2354-2366

Romeo PH, Prandini MH, Joulin V, Mignotte V, Prenant M, Vainchenker W, Marguerie G, Uzan G (1990) Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature* **344:** 447-449

Rougeon F, Kourilsky P, Mach B (1975) Insertion of a rabbit beta-globin gene sequence into an E. coli plasmid. *Nucleic Acids Res*. **2(12)**:2365-78.

Rund D, Rachmilewitz E (2005) Beta-thalassemia. *N Engl J Med* **353:** 1135-1146

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322:** 1839-1842

Satta S, Perseu L, Moi P, Asunis I, Cabriolu A, Maccioni L, Demartis FR, Manunza L, Cao A, Galanello R (2011) Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica* **96:** 767-770

Schroeder W.A, Huisman T.H.J, Brown A.K, Bouver N.G, Lerch P.O, Shelton J.R, Shelton J.B, Apell G (1971) Postnatal Changes in the Chemical Heterogeneity of Human Fetal Hemoglobin. *Pediatric Research* **5**: 493–493

SchVºbeler D, Groudine M, Bender MA (2001) The murine Œ≤-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 11432-11437

Schuh AH, Tipping AJ, Clark AJ, Hamlett I, Guyot B, Iborra FJ, Rodriguez P, Strouboulis J, Enver T, Vyas P, Porcher C (2005) ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol* **25:** 10235-10250

Sears DA (1978) The morbidity of sickle cell trait: a review of the literature. *Am J Med* **64:** 1021-1036

Sengupta T, Chen K, Milot E, Bieker JJ (2008) Acetylation of EKLF is essential for epigenetic modification and transcriptional activation of the beta-globin locus. *Mol Cell Biol* **28:** 6160-6170

Siatecka M, Xue L, Bieker JJ (2007) Sumoylation of EKLF promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Mol Cell Biol* **27:** 8547-8560

Simon MC, Pevny L, Wiles MV, Keller G, Costantini F, Orkin SH (1992) Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nat Genet* **1:** 92-98

Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van Ijcken W, Hou J, Steinhoff C, Rijkers E, Lenhard B, Grosveld F (2010) The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24:** 277-289

Solovieff N, Milton JN, Hartley SW, Sherva R, Sebastiani P, Dworkis DA, Klings ES, Farrer LA, Garrett ME, Ashley-Koch A, Telen MJ, Fucharoen S, Ha SY, Li CK, Chui DH, Baldwin CT, Steinberg MH (2010) Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood* **115:** 1815-1822

Song SH, Hou C, Dean A (2007) A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol Cell* **28:** 810-822

Song SH, Kim A, Ragoczy T, Bender MA, Groudine M, Dean A (2010) Multiple functions of Ldb1 required for beta-globin activation during erythroid differentiation. *Blood* **116:** 2356-2364

Sposi NM, Zon LI, Care A, Valtieri M, Testa U, Gabbianelli M, Mariani G, Bottero L, Mather C, Orkin SH, et al. (1992) Cell cycle-dependent initiation and lineage-dependent abrogation of GATA-1 expression in pure differentiating hematopoietic progenitors. *Proc Natl Acad Sci U S A* **89:** 6353-6357

Stachura DL, Chou ST, Weiss MJ (2006) Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. *Blood* **107:** 87-97

Steinberg MH (2009) Genetic etiologies for phenotypic diversity in sickle cell anemia. *ScientificWorldJournal* **9:** 46-67

Stuart MJ, Nagel RL (2004) Sickle-cell disease. *Lancet*.;364(9442):1343-60.

Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, Greaves DR (1989) A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* **338:** 352-355

Talbot D, 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

Talbot D, Philipsen S, Fraser P, Grosveld F (1990) Detailed analysis of the site 3 region of the human beta-globin dominant control region. *The EMBO journal* **9:** 2169-2177

Tanimoto K, Sugiura A, Omori A, Felsenfeld G, Engel JD, Fukamizu A (2003) Human beta-globin locus control region HS5 contains CTCF- and developmental stage-dependent enhancer-blocking activity in erythroid cells. *Mol Cell Biol* **23:** 8946-8952

Tavassoli M (1991) Embryonic and fetal hemopoiesis: an overview. *Blood Cells* **17:** 269-281; discussion 282-266

Testa U (2009) Fetal hemoglobin chemical inducers for treatment of hemoglobinopathies. *Ann Hematol* **88:** 505-528

Thein SL, Sampietro M, Rohde K, Rochette J, Weatherall DJ, Lathrop GM, Demenais F (1994) Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. *Am J Hum Genet* **54:** 214-228

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell* **10:** 1453-1465

Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395:** 917-921

Torkelson S, White B, Faller DV, Phipps K, Pantazis C, Perrine SP (1996) Erythroid progenitor proliferation is stimulated by phenoxyacetic and phenylalkyl acids. *Blood Cells Mol Dis* **22:** 150-158

Tripic T, Deng W, Cheng Y, Zhang Y, Vakoc CR, Gregory GD, Hardison RC, Blobel GA (2009) SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113:** 2191-2201

Tsang AP, Fujiwara Y, Hom DB, Orkin SH (1998) Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12:** 1176-1188

Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90:** 109-119

Tuan D, Solomon W, Li Q, London IM (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* **82:** 6384-6388

Tuan DY, Solomon WB, London IM, Lee DP (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* **86:** 2554-2558

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR, Schlessinger D, Cao A (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A* **105:** 1620-1625

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* **17:** 453-462

van der Ploeg LH, Flavell RA (1980) DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues. *Cell* **19:** 947-958

van Dijk TB, Gillemans N, Pourfarzad F, van Lom K, von Lindern M, Grosveld F, Philipsen S (2010) Fetal globin expression is regulated by Friend of Prmt1. *Blood* **116:** 4349-4352

Veith R, Galanello R, Papayannopoulou T, Stamatoyannopoulos G (1985) Stimulation of F-cell production in patients with sickle-cell anemia treated with cytarabine or hydroxyurea. *The New England journal of medicine* **313:** 1571-1575

Vilar JM, Saiz L (2005) DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* **15:** 136-144

Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J* **16:** 3145-3157

Weatherall DJ (2005) The global problem of genetic disease. *Ann Hum Biol* **32:** 117-122

Weatherall DJ (2010) The inherited diseases of hemoglobin are an emerging global health burden. *Blood* **115:** 4331-4336

Weatherall DJ, Clegg JB (2001) Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ* **79:** 704-712

Weiss MJ, Keller G, Orkin SH (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* **8:** 1184-1197

Welch JJ, Watts JA, Vakoc CR, Yao Y, Wang H, Hardison RC, Blobel GA, Chodosh LA, Weiss MJ (2004) Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104:** 3136-3147

Wendt KS, Peters JM (2009) How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Res* **17:** 201-214

Whyatt D, Grosveld F (2002) Cell-nonautonomous function of the retinoblastoma tumour suppressor protein: new interpretations of old phenotypes. *EMBO Rep* **3:** 130-135

Wyszynski DF, Baldwin CT, Cleves MA, Amirault Y, Nolan VG, Farrell JJ, Bisbee A, Kutlar A, Farrer LA, Steinberg MH (2004) Polymorphisms near a chromosome 6q QTL area are associated with modulation of fetal hemoglobin levels in sickle cell anemia. *Cell Mol Biol (Noisy-le-grand)* **50:** 23-33

Xu Z, Huang S, Chang LS, Agulnick AD, Brandt SJ (2003) Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. *Mol Cell Biol* **23:** 7585-7599

Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y, Yu D, Hardison R, Weiss MJ, Orkin SH, Bernstein BE, Fraenkel E, Cantor AB (2009) Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* **36:** 682-695

Zhang W, Kadam S, Emerson BM, Bieker JJ (2001) Site-specific acetylation by p300 or CREB binding protein regulates erythroid Kruppel-like factor transcriptional activity via its interaction with the SWI-SNF complex. *Mol Cell Biol* **21:** 2413-2422

Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95:** 279-289

Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, Cunningham JM, Jane SM (2009) PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **16:** 304-311

Zhou D, Pawlik KM, Ren J, Sun CW, Townes TM (2006) Differential binding of erythroid Krupple-like factor to embryonic/fetal globin gene promoters during development. *J Biol Chem* **281:** 16052-16057

Chapter 2

2

Targeted Chromatin Purification (TChP)

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

1. Department of Cell Biology, 3. Department of Biochemistry, 5. Department of Hematology, 6. Netherlands Proteomics Centre, 7. Biomics Department, 8. Center for Biomedical Genetics, Erasmus MC, Dr. Molewaterplein 50, 3015GE, Rotterdam, the Netherlands.

2. Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, Scotland, United Kingdom UK.

4. Computational Biology Unit-Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway.

ABSTRACT

Here we show for the first time 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 pulldown 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 knockdown of a number of these transcription factors, in human primary erythroid cells, induces γ-globin gene expression. Reactivation of γ-globin gene expression ameliorates the symptoms of β-thalassemia and sickle cell disease and hence these factors provide potential novel targets for the development of therapeutics to treat 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. The only procedure to identify such proteins directly has been published by Déjardin and Kingston (2009). Using an elegant hybridization technique called Proteomics of Isolated Chromatin (PICh), they identified factors bound to telomeres. Since 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). 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, constitute each less than 1 kilo base (kb) and are therefore almost two orders of magnitude less abundant than the combined telomeres. It is unlikely that such a small percentage of the genome can be purified sufficiently by PICh 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, since 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; 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 since they do not lead to a normal quality of life nor 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., 2011;

Xu et al., 2010). 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 pulldown of a protein hook that is crosslinked to the γ-globin gene promoter. Importantly, this will also purify interacting cis-regulatory regions as 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 novel 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

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. Firstly, the bait should not disturb the normal expression pattern of the γ-globin gene. Secondly, 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. Thirdly, 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::p53null mice) to facilitate the derivation of immortalized cell lines (Dolznig et al., 2001; Donehower et al., 1992; von Lindern, 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::p53null mice resulted in quadruple LCR-TetO-γ::TetR3T::BirA::p53null mice after several breeding and selection steps. Quadruple LCR-TetO-γ::TetR3T::BirA::p53null embryos were collected at 13.5dpc, 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, 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 (Figure 1C, D).

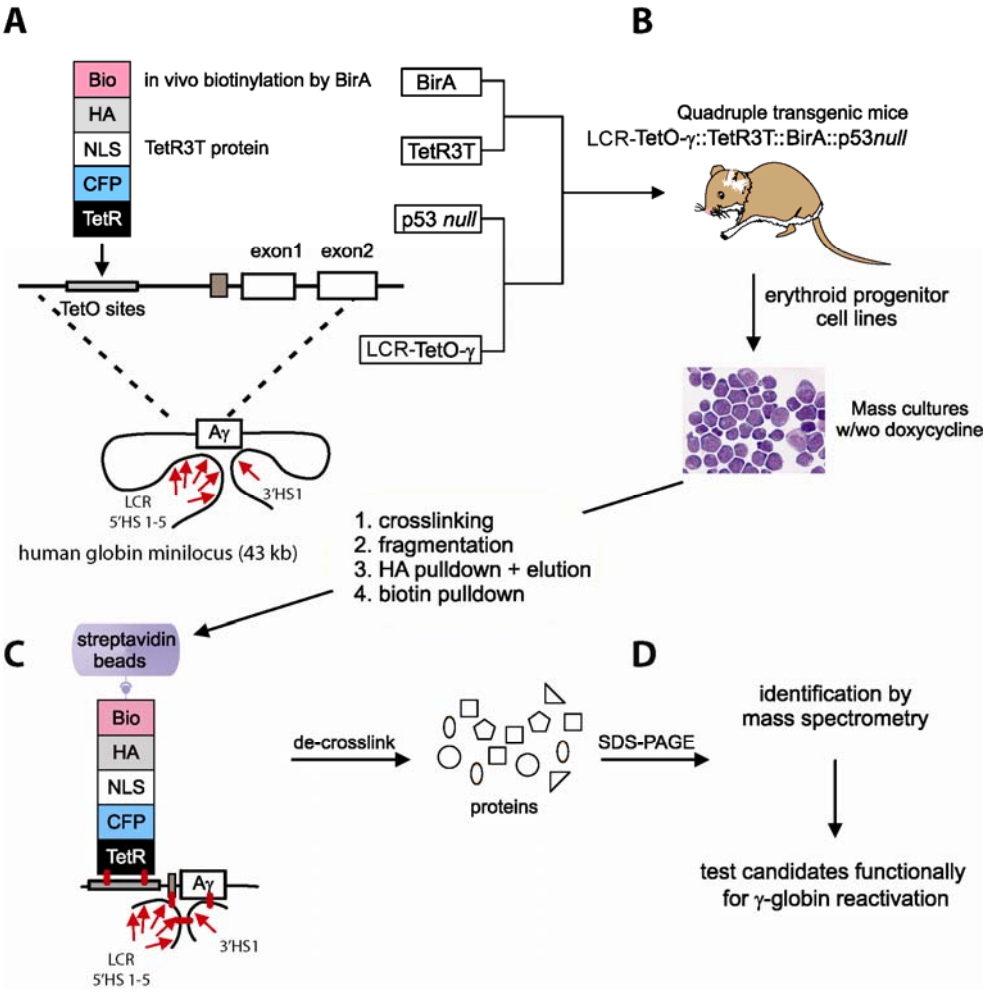


Figure 1| Scheme of the TChP approach. (**A**) Generation of the LCR-TetO-γ and TetR3T-G1HRD constructs resulting in the two transgenic mouse lines: TetR3T::BirA and LCR-TetO-γ::p53null mice (**B**) LCR-TetO-γ::TetR3T::BirA::p53null mice were used to derive the LCR-TetO-γ::TetR3T::BirA::p53null cell line. (**C**) Purification of the γ-globin chromatin hub after formaldehyde crosslinking, the LCR-TetO-γ::TetR3T::BirA::p53null cell line chromatin is fragmented and used for HA and biotin pulldown. = Hypothetical crosslink. (**D**) De-crosslinking of the isolated chromatin proteins, identification by mass spectrometry and functional testing for potential role in γ-globin regulation.

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 base pair (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 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 it remains inactive in the adult.

We next performed Chromosome Conformation Capture sequencing (3C-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 (Figure 1A and Figure 2C).

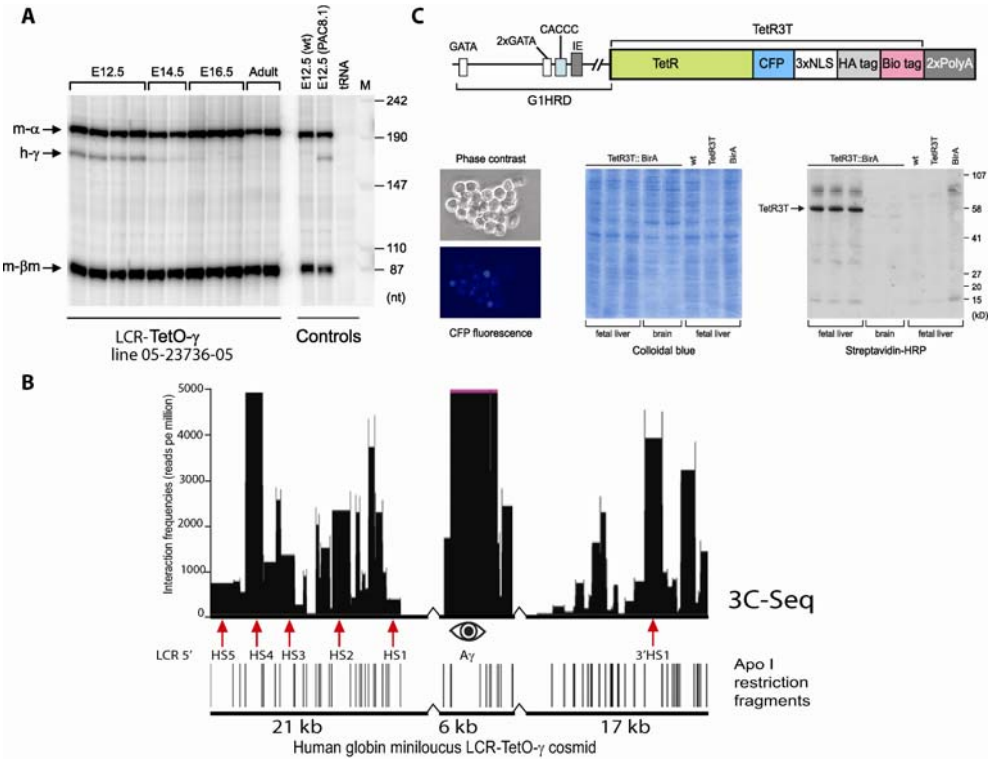


Figure 2| Transgenic mouse proof of principle experiments. (**A**) S1 nuclease protection analysis of globin expression in LCR-TetO-γ transgenic mice. RNA was prepared from fetal livers and adult blood of transgenic mice carrying two copies of the LCR-TetO-γ transgene (line 05-23736-05). Controls are 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 3C-seq using the γ-globin promoter as the viewpoint (indicated by the eye symbol). Bars indicate interactions of restriction fragments with the view point as measured by the number of normalized sequence reads. (**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).

Purification of the suppressed γ-globin chromatin hub

Erythroid progenitor cell lines (von Lindern, 2001) were derived from LCR-TetO-γ::TetR3T::BirA::p53null fetal livers at 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 streptavidin as well as anti-HA beads (Figure 3), in principle enabling consecutive pulldowns.

We next tested whether the chromatin captured with the HA beads could be eluted efficiently by the addition of HA peptide, allowing a subsequent pulldown step on the eluted chromatin by streptavidin. The results show that all of the bound material could be eluted efficiently by HA peptide (Figure 3C, D). Both pulldowns showed an enrichment of the sequences upstream and downstream of the TetO sites (Figure 3B, D).

Subsequently all parameters for TChP were optimized. This included fixation conditions, crosslinking, choice of beads for pulldown, buffer composition, blocking agents to prevent non-specific binding, large-scale culture and the elution conditions. The optimization details are shown in extended experimental procedures (supplementary document).

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

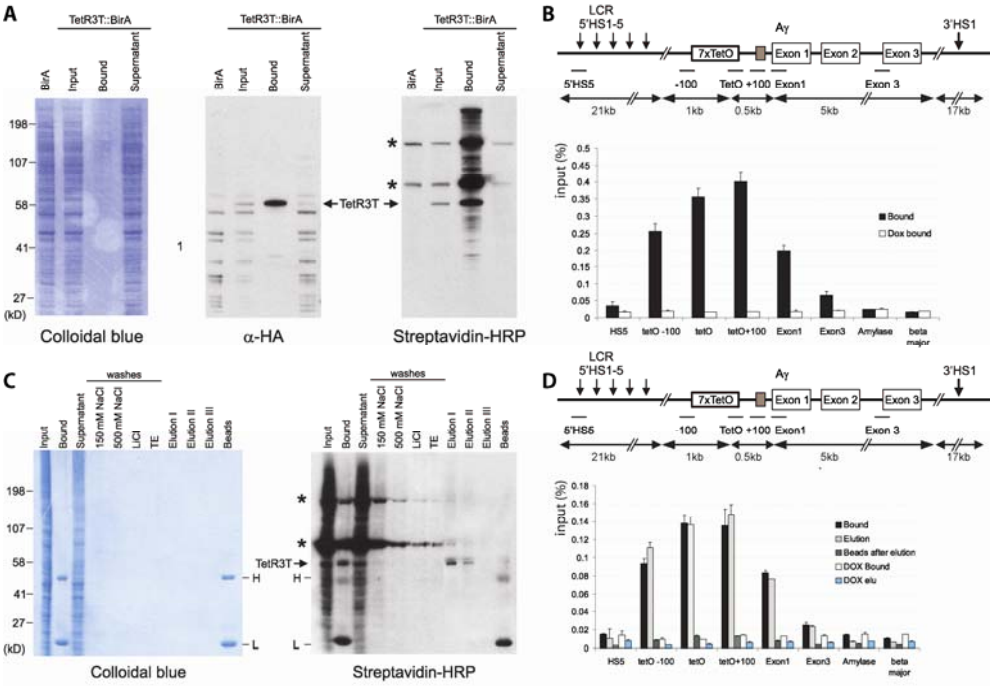


Figure 3| Affinity purification of the γ-globin chromatin hub. (**A**) Left panel; gel electrophoresis of the nuclear extracts after colloidal blue staining. The first lane contains a control extract from an identical cell line not containing the TetR3T construct. Middle panel; detection of the TetR3T 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 doxycyclin. 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 of the beads fractions. 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 doxycyclin. Bars in **B** and **D** represent averages of 3 independent measurements; error bars represent s.d..

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 $3x10^9$ 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 liter. We therefore repeated the purification several times to maximize identification of possible specific hits. To purify chromatin hub complexes 1.2 l of cells were grown to a density of $2.5x10^6$ cells per milliliter. For each experiment, we grew two cultures in parallel, one with and one without doxycycline. The cells were harvested, treated with 1% formaldehyde and the crosslinked chromatin was subjected to the isolation procedure. After the final step, the chromatin was de-crosslinked 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 729 proteins with a Mascot score of >40. Of these, 660 were also present in the doxycycline-treated purification experiments or were very unlikely to be involved in the suppressed γ-globin chromatin hub, such as abundant structural and cytoplasmic proteins, glutathione S transferase and ribosomal proteins. These were relegated to the bottom of the list, leaving 49 candidates (Table 1 and Table S1). Amongst these 49 candidates were a number of factors that have been reported previously to bind to the LCR or the γ-globin promoter: Apex1, Actl6a (Baf53A), Gata1, Dnmt1, Cdc5l, Zfp148 (ZBP-89) and Chd4 (Mi-2β) (Harju-Baker et al., 2008; Karmakar et al., 2010; Mahajan et al., 2005; Olave et al., 2007; Woo et al., 2011).

Identified protein	Description	Molecular weight (Da)	Number of unique detected peptides	Number of total identified peptide	Total coverage (%)
Apex1	apurinic/apyrimidinic endonuclease 1	35867	3	3	14.2
Actl6a	BAF53a	47913	4	4	16.8
Gata1	GATA binding protein 1	43274	2	2	9.0
Nap1l1	nucleosome assembly protein 1-like 1	45602	2	2	6.9
Zfp691	zinc finger protein 691	33075	2	3	6.4
Rad23a	RAD23a homolog	34813	2	4	6.2
Supt5h	suppressor of Ty 5 homolog	120988	3	3	4.5
Cdc5l	cell division cycle 5-like	95498	2	2	4.4
Zfp148	transcription factor BFCOL1	86469	2	11	3.8
Tcfeb	Transcription factor EB	52638	2	8	3.8
Sox30	SRY-box containing gene 30	68103	2	7	3.6
Chd4	chromodomain helicase DNA binding protein 4	94191	2	2	3.4
Ehmt1	euchromatic histone methyltransferase 1	139146	2	4	3.3
Ctnnbl1	nuclear protein NAP	65421	2	2	3.2
Fanci	Fanconi anemia, complementation group I	89597	2	2	3.1
Wdr12	nuclear protein Ytm1p	47770	1	1	2.8
Dnmt1	DNA methyltransferase 1	185624	3	3	2.2
Smarca2	SWI/SNF related subfamily	173244	2	3	1.9
Zfp532	zinc finger protein 532	151855	2	4	1.5
Ccdc88b	coiled-coil domain containing 88B	128640	2	4	0.8

Table 1|Summary of the selected proteins and their identified peptides. (See also Table S1)

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 LCR-TetO-γ-globin minilocus transgene in the mouse fetal liver-derived cell line. We tested for binding of Apex1, Actl6a (BAF53A), Gata1, Nap1l1, Supt5h, Cdc5l, Zfp148 (ZBP-89), ChD4 (Mi2-β), Ehmt1, Ctnnbl1, Fanci, and Son. Six out of these 12 proteins, Apex1, Gata1, Cdc5l, Zfp148, Chd4, and Fanci, showed binding to different elements within the minilocus. In particular, the 5'HS3 regulatory region showed binding of all six proteins, while the other elements showed binding of only one or two of these proteins. Particularly striking was the strong enrichment of Zfp148 (ZBP-89) at 5'HS2 (Table 2 and Figure 4). These results strongly suggest that the proof of principle experiments successfully identified a number of factors bound to the globin locus.

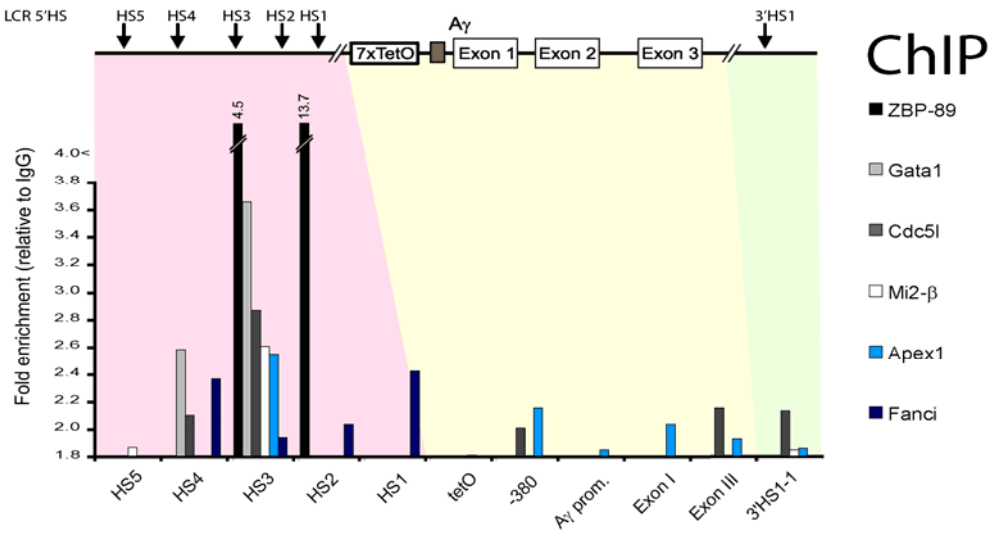


Figure 4| Chromatin occupancy along the LCR-TetO-γ minilocus in mouse erythroid fetal liver cell line for 12 selected proteins identified by TChP. Results are shown only for proteins with more than 1.8-fold enrichment. The enrichments are normalized relative to antibody IgG species and endogenous Amylase signals. Bars represent averages of two experiments. The different transcription factors tested are color-coded in the bar graph as shown on the right.

	Knock down efficiency	globin mRNA γ/(γ+β)	HbF % (HPLC)	HbF posetive cells % (IHC)	Binds to (ChIP)
Scramble	0%	0.1	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
CTNNBL1	85%	0.2	6.1	Not Determined	no binding detected
SUPT5	70%	0.2	5.3	Not Determined	no binding detected

Table 2| Summary of functional studies. 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.

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 homologues 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 five-fold rise in γ-globin mRNA expression and a rise in HbF (Figure 5A-C and Table 2). This was accompanied by an increase of the percentage of cells expressing high HbF levels as determined by immunohistochemistry (IHC) (Figure 5D and Table 2). These results were reproduced with HEP cells derived from 6 different healthy individuals (data not shown).

Identical sets of experiments were performed for CDC5L, APEX1, SUPT5H and CTNNBL1. In all cases at least two of the shRNA tested resulted in a knockdown of the 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 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 non-binding proteins in our ChIP experiments (Table 2 and Figure 4), SUPT5H and CTNNBL1, were also knocked down efficiently by different shRNAs but these KDs 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 and APEX1 is specific

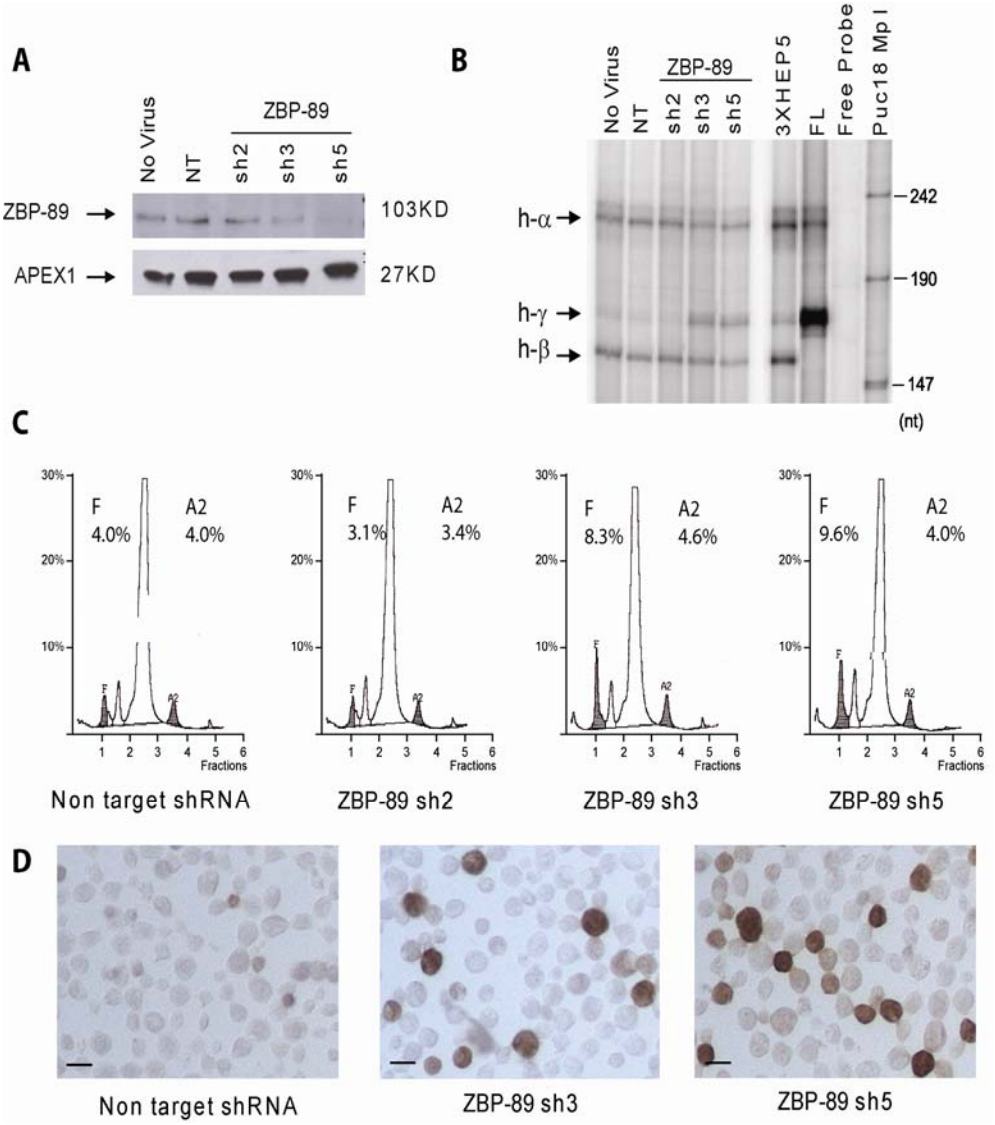


Figure 5| Lentiviral shRNA-mediated knockdown of ZBP-89. (**A**) Top panel, immunoblot for ZBP-89 shRNA-mediated knockdown in HEP cells. First lane: no virus control; NT: Non-target shRNA lentiviral transduced HEP cells. Bottom panel: APEX1 staining serving as loading control. (**B**) S1 nuclease protection analysis of globin expression in ZBP-89 depleted HEP cells. First lane: no virus control, NT; Non-target shRNA, ZBP-89 sh2, 3 and 5; short hairpins to human ZBP-89. Controls (to the right): 3xHEP; 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.

DISCUSSION

Here we show for the first time that chromatin factors bound to unique sequences can be identified directly through the purification of the target locus sequences (Targeted Chromatin Purification, TChP) in vivo. Although the results obtained are not directly comparable with those of Dejardin and Kingston (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 while 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 pulldown, probably due to mismatched hybridization and non-specific sticking of irrelevant chromatin. The latter is supported by the observation that 74 of the 200 PICh background proteins were also found among the TChP background proteins.

Here we have used a transgenic mouse approach and derived an erythroid cell line from it. This was necessary since the human γ-globin gene presents a specific developmentally regulated expression pattern which is not reproduced upon transfection. In many cases it would be possible to transfect the promoter or other sequence of interest into an appropriate cell line. An advantage of our method is that one can easily generate mice that are p53 null and express both the TetR protein and the BirA ligase ubiquitously, which would allow application of TChP to genes not expressed in erythroid cells (Driegen et al., 2005; Katsantoni et al., 2003).

Collectively, TChP identified 729 unique protein IDs in four independent purifications. Of these, 457 proteins hits were also identified in the doxycycline-treated control purifications and were therefore considered as background contaminants. We classified 223 proteins as unlikely candidates to interact with chromatin; these were abundant proteins such as ribosomal, structural and cytoplasmic proteins that were not seen in the control experiments for as yet unknown reasons. We considered the remaining 49 candidates for their possible role in γ-globin gene suppression. We selected a number of these candidates by applying the following criteria: are they expressed in human erythroid cells; are they likely to be bound to chromatin; are antibodies available for ChIP, and are shRNA clones available for functional analysis? That left approximately 20 candidates for testing, including Apex1, Cdc5l, Zfp148 (also known as ZBP-89), Supt5h, and Ctnnbl1. We investigated whether downregulation of these factors affects 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., 2011; Woo et al., 2008).

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 (Takashi and Kazuo, 1996), 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 with β-catenin like 1 protein (Ctnnbl1) (Conticello et al., 2008). Interestingly, CDC5L knockdown increased fetal hemoglobin, whilst 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 in 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; Talbort 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). SUPT5H and

CTNNBL1 knockdown did not result in γ-globin gene activation, indicating that these factors are not involved in γ-globin repression.

In the knockdown experiments we observed γ-globin upregulation for 3 out of 5 factors tested. Moreover, identifying specific proteins such as Gata1, Dnmt1, 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. For example, 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 an elution step of the streptavidin pulldown as part of the purification procedure, either by using biotin analogues (Hirsch et al., 2002) or 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. Finally incorporation of additional tags would allow an extra purification step to improve the signal over background ratio.

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 novel 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 basepairs upstream of the transcription initiation site and a loxP sequence inserted in the ClaI site (sequences are provided in extended experimental procedures online) (Figure 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 3x 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 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::p53null 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::p53null 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::p53null cells (von Lindern, 2001) was essentially as published with minor modifications provided in extended experimental procedures.

Chromatin Immunoprecipitation (ChIP) and γ-globin chromatin purification using Targeted Chromatin Purification (TChP)

Crosslinked chromatin was prepared by adding formaldehyde directly to culture medium to a final concentration of 1%, and incubated for 10 minutes 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 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 pulldown, HA elution, further purification by streptavidin pulldown, Chromatin Immunoprecipitation (ChIP),and the analysis of precipitated material by immunoprecipitation and quantitative PCR protocols are detailed in extended experimental procedures.

Mass spectrometry

1D 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) or an LTQ linear ion trap (Thermo), 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 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 40.

Chromosome Conformation Capture Sequencing (3C-Seq)

3C material was prepared essentially as described (Palstra et al., 2003; Soler et al., 2010). ApoI was used as the primary restriction endonuclease. Purified 3C DNA was used for 3C-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 adapter sequences. The resulting 3C 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 r3C-Seq pipeline (Thongjuea et al., unpublished). 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 hours 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

$200x10^5$ cells 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 minutes. 20 μl 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 hour in 1xTBS with 2% BSA and 0.05% Tween20 and incubated in primary (overnight) and secondary (1 hour) 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 (BioRad).

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 hrs at 4 °C, followed by peroxidase staining. Pictures were taken with an Olympus BX40 microscope (40x 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.

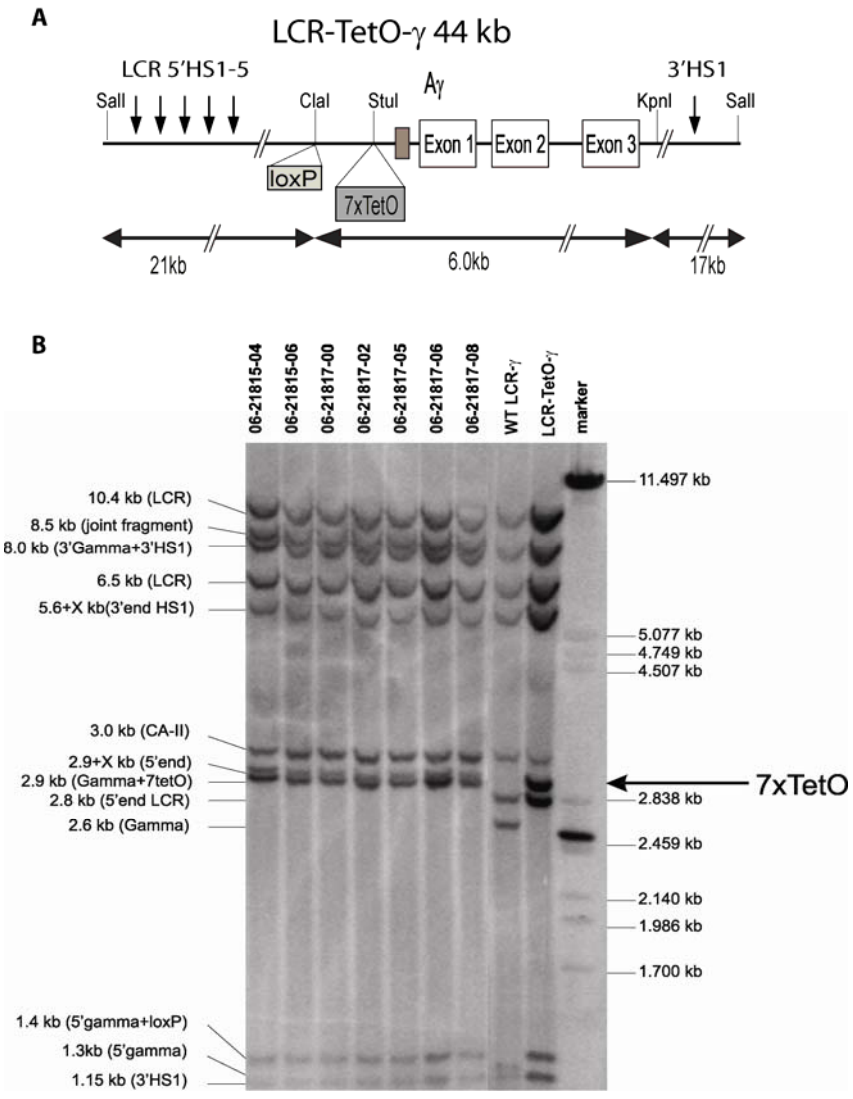


Figure S1| The LCR-TetO-γ minilocus (related to Figure 2). **A**) The drawing shows the γ-globin minilocus (Dillon and Grosveld, 1991) and the position of the insertion of the loxP site and 7 copies of the TetO sequences upstream of the γ-globin gene promoter. **B**) The EcoRI restriction digest of the normal γ-globin minilocus (WT LCR-Aγ) and the modified γ-globin locus (LCR-TetO-γ) cosmids. The fragment containing the TetO sequences is indicated. The lanes on the left show the Southern blot of several transgenic mice containing two copies of the locus and hybridized with the γ-globin minilocus as probe (Dillon and Grosveld, 1991) and a PstI fragment of the carbonic anhydrase gene as loading control. The transgene has two head to tail copies as determined with 5'HS5

and 3'HS1 probes (not shown). The identities of the bands are indicated on the left. The fragment sizes of the marker bands are indicated on the right.

	Identified protein	Score	Da	cover (%)	uniq pept.	total pept.	description
1	**Hmgb2**	505	24318	36.2	8	12	high mobility group box 2
2	**Hmgb3**	299	23167	26	4	6	high mobility group box 3
3	**Ybx1**	275	35822	22.4	3	3	Y box protein 1
4	**Actl6a**	264	47913	16.8	4	4	BAF53a
5	**Apex1**	258	35867	14.2	3	3	apurinic/apyrimidinic endonuclease 1
6	**Cdc5l**	232	92361	4.6	2	2	cell division cycle 5-like (S. pombe)
7	**Gtf2f2**	205	28421	20.1	4	4	general transcription factor IIF, polypeptide 2
8	**Dnmt1**	167	185624	2.2	3	3	DNA methyltransferase 1
9	**Mybbp1a**	156	152773	1.9	2	2	MYB binding protein (P160) 1a
10	**Gata1**	142	43274	9	2	2	GATA binding protein 1
11	**Nap1l1**	130	45602	6.9	2	2	nucleosome assembly protein 1-like 1
12	**Ddx3y**	112	28188	8.6	2	9	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3
13	**Son**	99	226000	2.2	2	2	truncated SON protein
14	**Ctnnbl1**	93	65421	3.2	2	2	catenin, beta like 1
15	**Snw1**	91	38806	3	1	1	SNW domain containing 1
16	**Zfp830**	82	40790	4.4	1	1	zinc finger protein 830
17	**Supt5h**	79	120988	4.5	3	3	suppressor of Ty 5 homolog (S. cerevisiae)
18	**Chd4**	77	94191	3.4	2	2	chromodomain helicase DNA binding protein 4
19	**Wdr12**	72	47770	2.8	1	1	WD repeat domain 12
20	**Ccdc88b**	68	128640	0.8	2	4	coiled-coil domain containing 88B
21	**Smc4**	64	147602	3.2	3	5	structural maintenance of chromosomes 4
22	**Ranbp2**	63	142089	3.2	3	13	RAN binding protein 2
23	**Sox30**	62	84285	2.9	2	7	SRY-box containing gene 30
24	**Csrp1**	58	21425	11.4	1	1	cysteine and glycine-rich protein 1
25	**Dbf4**	56	75042	4.2	2	3	DBF4 homolog (S. cerevisiae)
26	**Zfp148**	56	86469	3.8	2	11	zinc finger protein 148
27	**Tcfeb**	55	59616	3.4	2	8	transcription factor EB
28	**Nup50**	53	49921	5.4	1	1	nucleoporin 50
29	**Kif22**	52	73430	3.3	2	6	kinesin family member 22
30	**Uhrf1**	51	54565	2.1	2	3	ubiquitin-like, containing PHD and RING finger
31	**Cul3**	51	89519	4.2	2	11	cullin 3
32	**Zfp691**	50	33075	6.4	2	3	zinc finger protein 691
33	**Spopl**	49	45442	9.4	2	2	speckle-type POZ protein-like
34	**Smarca2**	48	173244	1.9	2	3	SWI/SNF relatedregulator of chromatin
35	**Lmcd1**	47	42167	6	2	3	LIM and cysteine-rich domains 1
36	**Irs1**	46	131783	2.2	2	3	insulin receptor substrate 1
37	**Ubr4**	46	149228	4.1	2	4	ubiquitin protein ligase E3 recognin 4
38	**Huwe1**	46	86058	6.2	2	2	HECT, UBA and WWE domain containing 1
39	**Ddb1**	46	128053	0.8	1	1	damage specific DNA binding protein 1
40	**Foxc1**	45	57261	5.8	2	2	forkhead box C1
41	**Zfhx4**	44	396922	2.2	2	2	zinc finger homeodomain 4
42	**Tep1**	44	294811	1	2	2	telomerase associated protein 1
43	**Nbn**	44	84703	6.8	2	6	nibrin
44	**Rad23a**	44	39802	5.5	2	4	RAD23a homolog (S. cerevisiae)
45	**Cand1**	44	25791	4.4	1	1	cullin associated and neddylation disassociated 1
46	**Zfp532**	43	151855	1.5	2	4	zinc finger protein 532
47	**Mbnl1**	43	37807	3.2	1	1	muscleblind-like 1 (Drosophila)
48	**Fanci**	43	89597	3.1	2	2	Fanconi anemia, complementation group I
49	**Ehmt1**	42	135382	3.4	2	4	euchromatic histone methyltransferase 1

Table S1 | Mass spectrometry results of TChP experiments (Related to Table 1).

Proteins found also in pulldowns of chromatin from LCR-TetO-γ::TetR3T::BirA::p53null cells treated with doxycyclin and unlikely to be associated with γ-globin suppression were considered as background and removed.

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., Gutierrez, 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.

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. Molecular 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. Nucl Acids Res *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., Boulme, F., Stangl, K., Deiner, E.M., Mikulits, W., Beug, H., and Mullner, 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.

Driegen, S., Ferreira, R., Van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res *14*, 477-482.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A *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.

Hirsch, J.D., Eslamizar, L., Filanoski, B.J., Malekzadeh, N., Haugland, R.P., Beechem, J.M., and Haugland, R.P. (2002). Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. Analytical Biochemistry *308*, 343-357.

Inusa, E.C.a.B. (2007). Sickle Cell Anemia: Targeting the Role of Fetal Hemoglobin in Therapy. Clinical Pediatrics *46*, 386 - 391.

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.

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 beta-globin locus. EMBO J *29*, 3260-3271.

Katsantoni, E.Z., Langeveld, A., Wai, A.W.K., Drabek, D., Grosveld, F., Anagnou, N.P., and Strouboulis, J. (2003). Persistent gamma-globin expression in adult transgenic mice is mediated by HPFH-2, HPFH-3, and HPFH-6 breakpoint sequences. Blood *102*, 3412-3419.

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. Biochemical and Biophysical Research Communications *344*, 1038-1047.

Leberbauer, C., Boulme, F., Unfried, G., Huber, J., Beug, H., and Mullner, 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 U S A *102*, 15012-15017.

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 â€“198 Mutation in the Î³-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.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K.A., 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.

Stamatoyannopoulos, G. (2005). Prospects for developing a molecular cure for thalassemia. Hematology *10*, 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.

Takashi, H., and Kazuo, S. (1996). A cdc5+ homolog of a higher plant,â€‰Arabidopsisâ€‰thaliana. Proc Natl Acad Sci U S A *93*, 13371-13376.

Talbort, 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.

von Lindern, M., Deiner, E.M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M.J., Mullner, E.W., 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.

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., 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.

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.

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*., et al.* (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993-996.

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.

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.

Supplemental Extended Experimental Procedures

Short random sequence containing SwaI and PmeI restriction enzyme digestion sites inserted into the StuI site of the γ-globin gene promoter 380 basepairs upstream of the transcription initiation site

30nt_Pme_Swa_Stu_EcoRV_AspF

5'>GTA CTA CGT AAT ACG ACT CAC TAG TGA GAT GTA TTT AAA TAG CTT TGT TTA AAC TGA GCG CCG GAG GCC TAG ATA TCG

30nt_Pme_Swa_Stu_EcoRV_AspR

5'>GTA CCG ATA TCT AGG CCT CCG GCG CTC AGT TTA AAC AAA GCT ATT TAA ATA CAT CTC ACT AGT GAG TCG TAT TAC GTA

LoxP sequence inserted in the ClaI site ClaI-KpnI fragment from the γ-globin minilocus (Dillon and Grosveld, 1991).

5'>CGA TAT AAC TTC GTA TAA TGT ATG CTA TAC GAA GTT ATG GCG CGC CTT

The sequence of overlapping oligonucleotides for HA-Bio tag synthesis:

HA-Bio-F1: 5'>GAT CCT ATC CTT ACG ATG TAC CCG ACT ATG CAT ATT CTG CTG GTC

HA-Bio-F2: 5'>ATC ACC ACC ACC ATC ACC ATC ACC ATC ACT TAG TTC CTC GTG GTT CTC CTG GTA TTT CTG

HA-Bio-F3: 5'>GCG GAG GGG GTG GCG CGT CAT CGT TGC GTC AGA TCC TTG ATA GTC AAA AAA TGG AGT

HA-Bio-F4: 5'>GGC GCA GCA ACG CTG GGG GTT CTT AAT AGC

His-Bio-R1: 5'>GGC CGC TAT TAA GAA CCC CCA GCG TTGC

His-Bio-R2: 5'>TGC GCC ACT CCA TTT TTT GAC TAT CAA GGA TCT GAC GCA ACG ATG ACG CGC CAC CCC

His-Bio-R3: 5'>CTC CGC CAG AAA TAC CAG GAG AAC CAC GAG GAA CTA AGT GAT GGT GAT GGT GAT GGT GGT

HA-Bio-R4: 5'>GGT GAT GAC CAG CAG AAT ATG CAT AGT CGG GTA CAT CGT AAG GAT AGG

Culturing of human erythroid proerythroblasts (HEPs)

Human buffy coat was obtained from the blood bank (Sanquin, The Netherlands), and blood peripheral mononuclear cells were collected from the interphase after Ficoll step gradient

purification, and washed. For initial expansion, $5x10^6$ cells ml^{-1} were cultivated in serum-free medium (StemSpan; Stem Cell Technologies, Vancouver, BC, Canada) supplemented with Epo (2 U ml^{-1} EPREX Ortho-Biotech, Tilburg, The Netherlands), the synthetic glucocorticoid Dexamethasone (Dex, 1 μM; Sigma, St Louis, MO), SCF (100 ng ml^{-1}; R&D Systems, Minneapolis, MN), and lipids (40 μg ml^{-1} cholesterol-rich lipid mix; Sigma) for about 5 days. The cells were fed every day by partial medium changes with the same medium. After 5 days, large proerythroblasts became visible and by day 7 they started to overgrow the culture. In order to purify proerythroblasts from lymphocytes and macrophages Percoll purification was carried out and the cells were isolated from the interphase. Homogeneous cultures of erythroid progenitors were kept in the same medium at 1.5-$2x10^6$ cells $ml^{-1}$by daily partial medium changes. Proliferation kinetics and size distribution of the cells were monitored daily using an electronic cell counter (CASY-1; Schärfe System, Reutlingen, Germany).

Culturing of LCR-TetO-γ::TetR3T::BirA::p53null cells

LCR-TetO-γ::TetR3T::BirA::p53null cells were grown in Stem-Pro 34 medium (Invitrogen) supplemented with 2mM L-Glutamine (Invitrogen), Epo (1 U ml^{-1} EPREX Ortho-Biotech, Tilburg, The Netherlands), the synthetic glucocorticoid Dexamethasone (1 μM; Sigma, St Louis, MO) and SCF (100 ng ml^{-1}; R&D Systems, Minneapolis, MN) (Dolznig et al., 2001). Cells were harvested at a density of $2-3x10^6$ cells ml^{-1}.

Chromatin Immunoprecipitation (ChIP) and γ-globin chromatin purification using Targeted Chromatin Purification (TChP*)*

Part A. Crosslinking and Chromatin preparation

Chromatin was crosslinked by adding formaldehyde directly to culture medium to a final concentration of 1%, and incubated for 10 minutes at room temperature. The reaction was stopped by the addition of 1 M glycine to a final concentration 0.125 M. The cells were spun at 450*g* for 3 min at 4 °C (Eppendorf centrifuge 5810R), the supernatant was decanted and the cells were washed 2 times with ice cold PBS. The cells were resuspended at a density of $2x10^8$ cells ml^{-1} in sonication buffer (10 mM Tris-HCl, 1 mM EDTA , 0.5 mM EGTA, pH 8.0) plus 1x complete EDTA-free protease inhibitor (Roche), and incubated for 10min on ice. The cells were disrupted and the DNA fragmented by sonication at 0 °C (Sanyo, SoniPrep150 large probe, 20 times, amplitude 9; 15" on; 45" off). The

sample was centrifuged for 10 minutes at 15,000*g* (Eppendorf centrifuge 5415D) at 4 °C, and the supernatant aliquoted. The pellet was discarded. The samples (chromatin) were snap-frozen in liquid nitrogen and stored at –80 °C. The efficiency of sonication was analyzed by adding 8 µl 5 M NaCl to 200 µl chromatin sample and the crosslinks were reversed at 65 °C for 4 hours. The DNA was recovered by phenol chloroform extraction and run on an agarose gel to analyze shearing efficiency.

Part B. Pulldown

Note: When proceeding to the PCR or immunoblot protocol a portion of the diluted chromatin (2%; ~20 µl) was kept to quantify the amount of DNA or protein. This sample was considered to be the starting material input.

Anti-HA-agarose beads pulldown

100µl of monoclonal Anti-HA-Agarose beads (Sigma Cat# A2095) were washed twice with 1 ml PBS. The agarose was pelleted by brief centrifugation for 1min at 400*g* (Eppendorf centrifuge 5415D) and the supernatant was discarded. The beads were equilibrated with 1 ml ChIP dilution buffer (0.01% SDS, 1.1% Triton X100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 150 mM NaCl) and the agarose pelleted by brief centrifugation for 1 min at 400*g*. The supernatant was discarded. The equilibration step was repeated two more times.

The beads were blocked for 1 hour with 1% fish skin gelatin (Sigma G7765), 0.2 mg ml^{-1} chicken egg albumin (Sigma A-5503) and 40 µl sonicated salmon sperm DNA (10 mg ml^{-1}). The agarose was pelleted by brief centrifugation (1 min) at 400*g* and the supernatant was discarded.

A chromatin sample equal to $20x10^6$ cells (i.e. 100 µl) was taken up in 1 ml ChIP dilution buffer containing 1x complete EDTA-free protease inhibitor (Roche).

The agarose was pelleted by brief centrifugation and the supernatant collected. The pre-cleared supernatant was added to the blocked beads and incubated at 4 °C overnight with rotation.

The agarose was pelleted by centrifugation (1 min at 400*g*) and the supernatant carefully removed. The protein HA-agarose::TetR-HA-biotag::chromatin complex was washed for 3-5 minutes on a rotating platform with 1 ml of each of the buffers in the following order, all at 4 °C; a Low Salt Immune Complex Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), one wash; a High Salt Immune Complex Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), one wash; LiCl Immune Complex Wash Buffer

(0.25 M LiCl, 1% NP-40, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.1), one wash; TE Buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), two washes.

At this stage the HA-agarose::TetR-HA-biotag::chromatin complex was ready for either an Immunoprecipitation and Immunoblot assay, PCR assay or elution by HA peptide followed by streptavidin purification.

Chromatin Immunoprecipitation (ChIP)

ChIP for specific protein factors was essentially as described in the previous section with minor adjustments. The HA-agarose beads were replaced with antibody against the protein of interest and incubated overnight at 4 °C with rotation. Depending on the antibody, Protein A or G agarose Salmon Sperm DNA beads, Millipore, were added the next day and incubation at 4 °C with rotation was carried out for another hour. After this step, washes were performed and the rest of the procedure was followed as described above.

Streptavidin paramagnetic beads pulldown

Streptavidin paramagnetic beads chromatin pulldown via TetR3T protein was performed essentially similar to Anti-HA-agarose beads pulldown with slight modifications. 50 µl beads were used per $20x10^6$ cells and the beads were pelleted with a magnet instead of centrifugation. In the PCR protocol beads were kept after elution and were discarded after first phenol chloroform extraction. This is because of strong binding affinity of biotin- streptavidin that can not be eluted in elution buffer only.

Optimization of parameters for TChP

The Optimized parameters for TChP were as follow:

- Titration of crosslinking agent → 1% formaldehyde, 10 min at 37 °C
- DNA shearing → Sonication
- Pre-clearing of the chromatin → NO
- Blocking the beads → 1% fish skin gelatin
- Sonication buffer vs. SDS lysis buffer → Sonication buffer
- Number of cells per 100 µl of HA beads → $2x10^7$
- Optimal streptavidin beads → Dynabeads® M-280

Immunoprecipitation and immunoblot protocol

Following washing of the beads the immunoprecipitated complex can be analyzed by immunoblot analysis. 25 µl of 1x Laemmli buffer was added per sample and boiled for 10 minutes. 20 µl was loaded per lane of NuPAGE 4-12% Bis-Tris gel (Life Technologies), and the immunoblot procedure carried out.

PCR protocol to amplify DNA bound to the beads

500 µl of freshly prepared elution buffer (1% SDS, 0.1 M NaHCO3) was added per sample of the pelleted agarose-antibody-protein or TetR-HA-biotag::DNA complex beads. The sample was vortexed briefly and then incubated at room temperature for 15 minutes with rotation. The agarose was pelleted by brief centrifugation (1 min) at 400*g* and the supernatant (eluted purified chromatin) was transferred to a fresh tube. 20 µl of 5 M NaCl was added to the eluate (500 µl) and the protein-DNA crosslinks were reversed by heating at 65 °C for at least 4 hours. 10 µl of 0.5 M EDTA and 20 µl 1 M Tris-HCl, pH 6.5 were added plus 2 µl of 10 mg ml^{-1} Proteinase K and incubated for one hour at 45°C. The DNA was cleaned by two phenol chloroform and one chloroform extraction. 1 µl of 20 µg ml^{-1} glycogen (Roche, 14267332) was added followed by 0.6 volumes of isopropanol. After 30 min at –20 °C the sample was centrifuged and the pellet washed with 70% ethanol and air dried. The pellet was resuspended in water, the input into 400 µl and pulldown sample into 200 µl; and 4 µl was used per PCR reaction.

Quantitative PCR conditions and primers

Quantitative real-time PCR (MyIQ, Bio Rad) was performed using 0.75 µl of SYBR Green I (Sigma S9430) 1/2,000 dilution in DMSO, Platinum Taq kit (Invitrogen), 10 pmol of each primer, 4 µl DNA sample under the following cycling conditions: 3 min at 95 °C followed by 40 cycles of 30 sec. at 95 °C, 20 sec. at 56 °C, 40 sec. at 60 °C, 15 sec. at 75 °C. Mouse β-major and Amylase was used as endogenous references. Enrichment of specific sequences was calculated using the comparative CT method (Livak and Schmittgen, 2001).

Primers used for Quantitative PCR:

Human genomic βLCR HS5-F	5'>CCCAAGCAAGGAAGTTGT
Human genomic βLCR HS5-R	5'>CAGATGTCCTGTCCCTGTA
Human genomic βLCR HS4-F	5'>TTTCTCTCTCCCACTCAGC
Human genomic βLCR HS4-R	5'>TGCTATCAAAGCCCTGACA

Human genomic βLCR HS3-F	5'>GCTCAGATAGGTGGTTAGGT
Human genomic βLCR HS3-R	5'>TGGTCTATCTCTCCTGGCT
Human genomic βLCR HS2-F	5'>CTCCATTAGTGACCTCCCA
Human genomic βLCR HS2-R	5'>TTACACAGAACCAGAAGGC
Human genomic βLCR HS1-F	5'>TGCGGTTGTGGAAGTTTAC
Human genomic βLCR HS1-R	5'>CACTAAGGGTGAGGATGCT
Human genomic tetO-100-F	5'>AAAAGTCACAAAGAGTATATTCAAAAAG
Human genomic tetO-100-R	5'>CAGGATTTTTGACGGGACAAA
tetO-F	5'>AAGTCGAGCTCGGTACTACG
tetO-R	5'>GACTTCTTTTGTCAGCCGTTTT
Human genomic tetO+100-F	5'>CGGCTGACAAAAGAAGTCCT
Human genomic tetO+100-R	5'>CCCAAGAGGATACTGCTGCT
Human genomic Aγpromoter-F	5'>GAAACGGTCCCTGGCTAAACTC
Human genomic Aγpromoter-R	5'>CCTCACTGGATACTCTAAGACTATTGG
Human genomic AγExonI-F	5'>ACCCTTCAGCAGTTCCACAC
Human genomic AγExonI-R	5'>CCCCACAGGCTTGTGATAGT
Human genomic AγExonIII-F	5'>GACCGTTTTGGCAATCCATTTC
Human genomic AγExonIII-R	5'>TTGTATTGCTTGCAGAATAAAGCC
Human genomic β3'HS1-R	5'>ATCTACTTTCATTCAGGCTTCTTCAG
Human genomic β3'HS1-F	5'>CTCTATGGTGGCAACATGGATAATAC
Mouse genomic Amylase-F	5'>CTCCTTGTACGGGTTGGT
Mouse genomic Amylase-R	5'>AATGATGTGCACAGCTGAA
Mouse genomic βmajor-F	5'>GGGAGAAATATGCTTGTCATC
Mouse genomic βmajor-R	5'>CAACTGATCCTACCTCACCTT

Supplemental References

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-[Delta][Delta]CT Method. Methods *25*, 402-408.

Chapter 3

3

The role of ZBP-89 in globin gene regulation

Ali Aghajanirefah[1*], Farzin Pourfarzad[1*], Ralph Stadhouders[1], Robert-Jan Palstra[1], Nynke Gillemans[1], Jan Christian Bryne[2], Kerstin Wendt[1], Petros Kolovos[1], Jeroen Demmers[3], Wilfred van IJcken[3], Eric Soler[1,4] and Frank Grosveld[1,4,5,6]

* These authors contributed equally to this work

1. Department of Cell Biology, 3. Department of Biomics, 4. Centre for Biomedical Genetics, 5. Cancer Genomics Centre, 6. NIRM Erasmus Medical Center, Dr Molewaterplein 50, 3015GE Rotterdam, The Netherlands.

2. Computational Biology Unit-Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway.

ABSTRACT

Reactivation of human fetal γ-globin expression in adult erythroid cells ameliorates the symptoms of sickle cell disease and thalassemia, which is why we are interested in the silencing mechanism of these genes. We recently described an unbiased proteomics approach, Targeted Chromatin Purification (TChP, Pourfarzad et al., submitted), to identify DNA binding proteins involved in the silencing process and identified ZBP-89 as one of the candidate proteins. Here we report on the functional role of ZBP-89 in globin gene regulation. We show that ZBP-89 interacts with subunits of suppressor as well as activating protein complexes in different cell lines (MEL versus K562). ZBP-89 binds to the β-globin LCR and is differentially recruited to the regulatory sequences adjacent to the γ- and β-globin genes in fetal liver and adult Human Primary Erythroid (HEP) progenitor cells. Reduced ZBP-89 expression in HEP cells correlates with reduced human adult globin and increased fetal/embryonic genes expression, including γ-globin. We show this is the consequence of a change in spatial conformation of the β-globin locus back partially to the fetal liver γ-globin expressing stage. These data show that ZBP-89 has an important regulatory role in globin gene regulation by changing long-range interactions during different stages of development.

INTRODUCTION

The human α and β-globin gene loci code for α- and β-globin, which together with heme are subunits of hemoglobin protein, responsible for the oxygen transport in the blood. The human β-globin locus consists of an upstream regulatory region, the Locus Control Region (LCR), and five different β-globin like genes: ε, Gγ, Aγ, δ and β. Expression of these genes is developmentally regulated with one switch in the embryonic period where expression of ε is replaced with γ and a second switch from γ to β around the time of birth (Stamatoyannopoulos and Grosveld, 2001). Conditions affecting the function of β-globin, known as β-thalassemia and Sickle Cell Disease (SCD), are among the most frequent inherited single gene disorders in the human population (Weatherall, 1996). Their severity is greatly ameliorated by increased γ-globin expression in response to drugs such as hydroxyurea (Steinberg, 2002; Steinberg et al., 1997) or in patients with hereditary persistence fetal hemoglobin (HPFH) (Stamatoyannopoulos and Grosveld, 2001). Thus reactivation of γ-globin expression in adults would provide a very elegant and attractive treatment of the disease. This has led to intense research efforts to elucidate the mechanism of γ-globin gene switching with the aim of reversing the normal switching process, resulting in the identification of DNA cis elements as well as transcription factors (TFs) that regulate the temporal expression of globin genes at each developmental stage (Akinsheye et al., 2011; Bank, 2006; Higgs et al., 2012; Stamatoyannopoulos, 2005).

Regulatory regions are present throughout the locus, the most prominent of which is the Locus Control Region (LCR) upstream of the locus. The LCR elements are bound by TF complexes like that formed by Ldb1 (Soler et al., 2010; Song et al., 2007) and promote the interaction of the LCR with the genes (Deng et al., 2012; Tolhuis et al., 2002). One of its components, ETO2 is present at HS2 and 3 of the LCR and a region 3.7 kb 3' to the Aγ-globin gene and its binding correlates with reduced γ-globin expression (Kiefer et al., 2011). These interactions change during development (Palstra et al., 2003a; Palstra et al., 2008; Patrinos et al., 2004; Schoenfelder et al., 2010; Tolhuis et al., 2002) resulting in temporal and developmental stage specific globin gene expression. BCL11a is one of the most important factors involved in the globin gene regulation by modulating such interactions. BCL11a binds inside the locus just downstream of the γ-globin genes and in cooperation with Sox6, GATA1, FOG and the NuRD complex suppresses the expression of the γ-globin genes (Sankaran et al.,

2008; Xu et al., 2011; Xu et al., 2010). Other factors in γ-globin gene expression such as KLF1 exert their action directly in the locus and indirectly by regulating BCL11a (Borg et al., 2010).

However a γ-globin gene linked to the LCR, but lacking the downstream BCL11a binding site is still suppressed in transgenic mice, suggesting that other factors binding to the promoter or LCR are also important (Dillon and Grosveld, 1991). For example the TR2/TR4 complex has been implied as one of the factors binding to the γ-promoter (Campbell et al., 2011), but their absence results only in a modest up-regulation of γ-globin suggesting that more factors are involved. We have recently described an unbiased proteomics approach, Targeted Chromatin Purification (TChP), to identify other possible proteins involved in γ-globin silencing (Pourfarzad et al., submitted) and identified several proteins whose reduced expression led to an increase in γ-globin gene expression in adult Human Erythroid Progenitor (HEP) cells. One of the most interesting proteins we identified by TChP was ZBP-89 (also known as ZNF148, Zfp148 in mouse, BERF1, and BFCOL1). It is a Krüppel-type, zinc-finger transcription factor that binds to a GC-rich region, and represses or activates known target genes. When acting as a repressor, it has been proposed that ZBP-89 and Sp1 compete for binding to the same or overlapping GC-rich sequences (Merchant et al., 1996). ZBP-89 has been reported to interact with GATA1 and Mafk and it is involved in erythroid development and differentiation (Brand et al., 2004; Woo et al., 2011; Woo et al., 2008a). Knock down of ZBP-89 in zebrafish results in the "bloodless" phenotype due to disruption of both primitive and definitive erythropoiesis (Li et al., 2006). It has been proposed that ZBP-89 represses gene transcription via specific recruitment of HDAC1 to promoters of its target genes (Wu et al., 2007). This would agree with the known role of HDACs in γ-globin silencing, as HDAC inhibitors can partially (re) activate γ-globin expression (Constantoulakis et al., 1989; Perrine et al., 1989; Perrine et al., 1987). We and others (Pourfarzad et al. submitted) (Woo et al., 2011; Woo et al., 2008b) showed strong binding of ZBP-89 to the human β-globin LCR DNA hypersensitive sites 2 and 3 (HS2 and 3) and to a lesser extent to sequences just upstream and downstream of γ-globin genes. ZBP-89 was also reported to bind to the GATA-1 enhancer, and a number of other genes (Ohneda et al., 2009; Vernimmen et al., 2007). Based on these observations, we decided to study the molecular function of ZBP-89 in globin regulation.

Here we describe the role of the zinc finger protein ZBP-89 in the regulation of β-globin gene cluster, using proteomic and genomic approaches. We show that ZBP-89 interacts with repressive chromatin modifying complexes and with other transcription factors important for globin gene regulation in the erythroid lineage. Interestingly, in the myeloid lineage the repressive interactions are lost or reduced and replaced with transcriptional activator interactions.

RESULTS

ZBP-89 is involved in repressive and activating complexes

We used an epitope-tagging affinity purification strategy to identify the ZBP-89 protein partners in erythroid and myeloid lineage. Mouse erythroid leukemia (MEL) cells expressing BirA biotin ligase (MEL-BirA) (de Boer et al., 2003) were stably transfected with a C-terminally tagged Bio-V5 ZBP-89 expression vector. Tagged ZBP-89 from undifferentiated and differentiated MEL cells was purified using streptavidin magnetic beads (Fig. 1A and B). Similarly, K562 a myeloid cell line that expresses globin genes was stably transfected with the same C-terminally tagged Bio-V5 ZBP-89 expression vector. Bio-V5-ZBP-89 was purified from K562 nuclear extracts using V5 antibody conjugated agarose beads. After pull down the proteins were digested and analyzed by mass spectrometry (LC-MS/MS). Table 1 shows the proteins that were found to interact with ZBP-89 in proliferating and differentiating MEL and in K562 cells. We identified several members of the NuRD complex (Chd4, Hdac 1 and 2, Mta1/2/3, Gatad2a/b, Rbbp4, and Mbd3) in both proliferating and differentiating MEL cells. Interestingly Chtop and several members of the 5FMC complex (Fanis et al; 2012) are detectable as ZBP-89 interacting partners before and after induction suggesting the involvement of desumoylation of ZBP-89 during differentiation. Furthermore, we also identified several other TFs (Zfp219, Zfp512 and Runx1). A number of these and chromatin remodeling protein interactions change upon differentiation e.g. Zfp512, Fog1 (Zfpm1), Chtop and Smarca4/5. Since we are interested in the mechanism of ZBP-89 and γ-globin activation and we confirmed the results obtained from mass spectrometry for the factors thought to be involved in γ-globin regulation as well as subunits of the NuRD complex and tested their interaction with ZBP-89 by western blotting (Fig 1D).

It has previously been reported that ZBP-89 is in complex with transcriptional activators in non-erythroid cell lines (Bai and Merchant, 2000; Woo et al., 2011). We mostly find ZBP-89 interacting with the NuRD suppressor complex in MEL cells suggesting that ZBP-89 forms a different complex in purely erythroid cells. We therefore also expressed the Bio-V5-ZBP-89 fusion protein in human chronic myelogenous leukemia cell-line K562 (Lozzio and Lozzio, 1975), which also has erythroid characteristics. We indeed found ZBP-89 in complex with activators of transcription namely Trrap/P400, and different mediator proteins (Table 1). Some of these were previously described by Woo et al (Woo et al., 2011). The interactions were confirmed by western blotting (Supplementary Fig. 1). In summary these data suggests that ZBP-89 acts as a positive regulator that interacts with

the basic transcription machinery at the transcription start site (TSS) in an erythroid/myeloid precursor cell (K562) as evidenced by its interactions with the Mediator proteins and NIBL/MAU1 which are known to bind at the TSS (Zuin and Wendt unpubl.). These complexes are lost in erythroid cells (MEL) with a switch to NuRD, a repressive complex (Fig. 1D and Supplementary Fig. 1). Interestingly ZBP-89 also interacts with the 5FMC complex (Fanis et al 2012), which switches to an interaction with a number of transcription factors and zinc finger proteins during terminal differentiation.

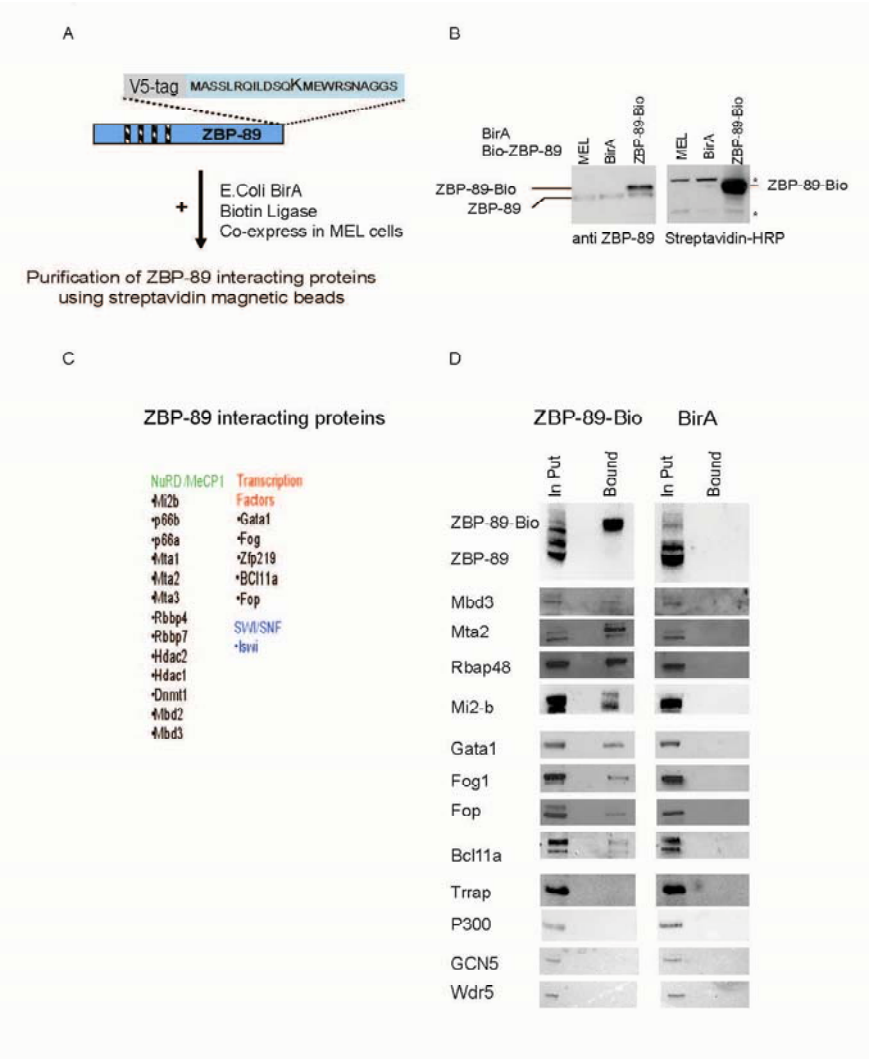


Figure 1) ZBP-89 interacts with subunits of NuRD repressor complex and transcription factors involved in γ-globin gene regulation. A) Schematic representation of C-terminal ZBP-89 epitope tagging with V5-Bio tag followed by its expression in MEL cells expressing Bir A biotin ligase (de Boer 2003). Speckled boxes indicate the positions of the four ZBP-89 Zinc-fingers. The biotinylation Lysine residue in the 23 amino acid biotin tag is indicated as large K letter. ZBP-89-biotinilated fusion protein precipitated using streptavidin beads. B) Taged ZBP-89 biotinylation in MEL cells. (Left) Western blot with ZBP-89 antibody for endogenous and tagged ZBP-89 protein detection in MEL cells nuclear extract. MEL; non-transfected MEL cells, Bir A; MEL cells expressing Bir A biotin ligase, ZBP-89-V5-Bio; MEL cells expressing ZBP-89-V5-Bio and Bir A biotin ligase (Right) Streptavidin–HRP conjugated staining of the same blot as in the left for detection of biotinylated ZBP-89. Asterisks indicate endogenously biotinylated cytochrome proteins (C) Mass spectrometry results of the proteins identified by streptavidin precipitation of biotinylated ZBP-89 in MEL Bir A cells. Subunits of the NuRD complex and transcription factors involved in erythroid differentiation/ globin regulation are among the most abundant and specific identified proteins. (D) Co-precipitation of ZBP-89 interacting proteins using streptavidin to confirm results obtained in Mass Spectrometry. Trrap, P300, GCN5 and Wdr5 were used as negative control.

	MEL				K562		
symbol	uniqe peptide (undiff.)	mascot score (undiff.)	Uniqe peptide (diff.)	mascot score (diff.)	uniqe peptide	mascot score	description
NuRD							
Chd4	24	1265	29	1579	4	166	Chromodomain-helicase-DNA-binding protein 4
Mta1	11	697	17	1009			metastasis associated 1
Gatad2a	11	645	13	924			Transcriptional repressor p66 alpha
Gatad2b	11	761	12	899			Transcriptional repressor p66-beta
Mta2	8	514	11	666			Metastasis-associated protein MTA2
Hdac1	8	415	9	473			Histone deacetylase 1
Hdac2	7	418	9	512			Histone deacetylase 2
Dnmt1	7	300	6	331			DNA (cytosine-5)-methyltransferase 1
Mta3	6	346	7	425			metastasis associated 3
Chd3	5	324					chromodomain helicase DNA binding protein 3 (Mi-2 alpha, Chd7)
Chd8	3	139	6	346			Chromodomain-helicase-DNA-binding protein 8
Mbd3	3	119	5	275	4	141	Methyl-CpG-binding domain protein 3
Rbbp4	8	549	11	723			Retinoblastoma binding protein 4
Gm10093			8	442			Histone deacetylase
SWI/SNF							
Smarca5	4	181	5	210	3	114	SWI/SNF-related regulator of chromatin
SMARCC1					5	137	SWI/SNF complex subunit SMARCC1
ACTL8					2	129	Actin-like protein 8
Mediator coplex							
MED1					5	195	Mediator of RNA polymerase II transcription subunit 1
MED17					5	209	Mediator of RNA polymerase II transcription subunit 17
MED16					5	229	Mediator of RNA polymerase II transcription subunit 16
MED7					3	163	Mediator of RNA polymerase II transcription subunit 7
MED6					3	120	Mediator of RNA polymerase II transcription subunit 6
MED12					3	117	Mediator complex subunit 12
MED14					3	92	Mediator of RNA polymerase II transcription subunit 14
TRRAP/P400 complex							
TRRAP					8	380	Transformation/transcription domain-associated protein
EP400					6	256	E1A-binding protein p400
Cohesin complex							
Nipbl			2	73	10	455	Nipped-B-like protein
MAU2					4	133	MAU2 chromatid cohesion factor homolog
Transcription factors							
Zfp219	7	574	8	610			zinc finger protein 219
Bclaf1	4	178					BCL2-associated transcription factor 1
Pelp1	2	96					proline, glutamic acid and leucine rich protein 1
Wdr18	2	82					WD repeat domain 18
Cdk1	2	87	2	91			Cyclin-dependent kinase 1
Runx1	2	134	2	166			runt related transcription factor 1
Senp3	1	78	1	63			SUMO/sentrin specific peptidase 3
Chtop			1	76			2500003M10Rik
Zfpm1			2	58			similar to FOG
Znf512b			4	160			zinc finger protein 512B
Znf512			4	195			Zinc finger protein 512
Znf622			2	85			Zinc finger protein 622
Trim28			2	97			Transcription intermediary factor 1-beta

Table 1) The proteins that were found to interact with ZBP-89 in proliferating and differentiating MEL and in K562 cells.

Differential recruitment of ZBP-89 to globin regulatory regions

We originally identified ZBP-89 via γ-globin TChP (Pourfarzad et al), hence we were interested to identify its genome wide binding sites and in particular those in the human globin locus. A genome wide ZBP-89 ChIP-Seq was carried out in Human Erythroid Progenitors (HEPs) grown out from adult peripheral blood mononuclear and fetal liver cells (AD-HEP and FL-HEP). We identified 23709 significant ChIP-Seq peaks in AD-HEPs. Of these 77% and 93% mapped within 10 kb and 50 kb from a gene, respectively. ZBP-89 binds primarily at promoters (Fig. 2A), within +/- 1 kb from transcription start site of the associated genes (Fig. 2B). The motif discovery search found enrichment for the C rich motif as CCCNNCCCCC (Fig. 2C). The result of ChIP-Seq experiment of FL-HEP appeared very similar to that of AD-HEPs

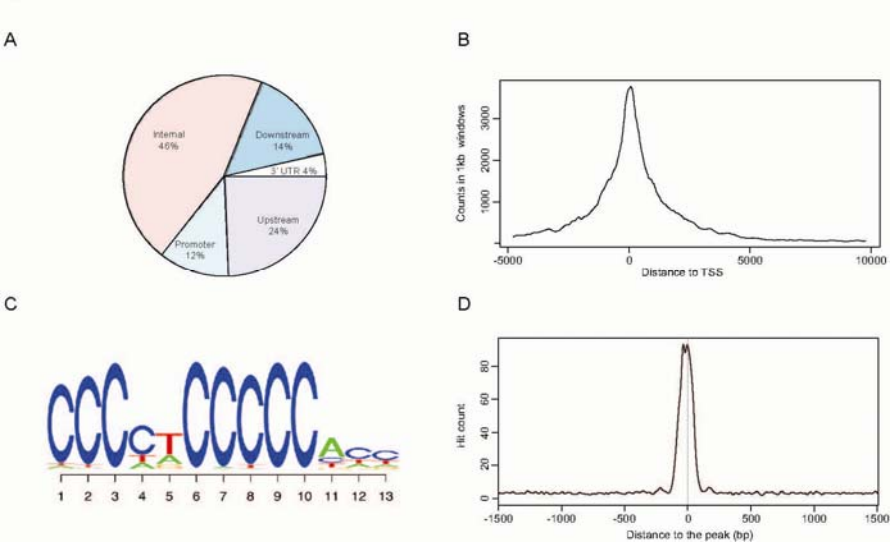


Figure 2) Endogenous ZBP-89 chromatin occupancy in Human Erythroid Progenitors (HEP). A) ZBP-89 recruitment relative to the genes. Different regions relative to the genes are defined as: Promoters; +/- 1 kb from transcription start site, 3' UTR; within 10kb of transcript end, Upstream; in the region between 1 to 10 kb upstream of transcription start site. Downstream; in the region between 1 to 10 kb downstream of the gene. D) Frequency of binding relative to transcription start site of the associated genes.

There is a considerable difference in ZBP-89 binding to the β-globin locus. There is much stronger binding of the factor to the locus in FL-HEP cells when compared to AD-HEP cells. The strongest

binding in both cell types is seen at HS2 and HS3 of the β-globin LCR. Much smaller peaks are found at HS1 of the LCR, in the middle of the β-globin locus 4kb upstream of Gγ, 3.7 kb down stream of Aγ, and 1.7 kb downstream of β-globin gene (Fig. 3). Interestingly the signal ratio at the weaker binding sites around the γ versus β-genes is reversed in the two cell types with relatively higher binding at the γ-genes in FL-HEP. The result indicates that the binding of ZBP-89 to these sites has a negative effect on γ-globin expression in AD-HEPs. Reducing ZBP-89 may therefore lead to an increase in γ-globin gene expression. Alternatively the weak sites may not be real binding sites but rather sites where interactions take place with regulatory regions that bind to ZBP-89. Thus these results suggest that ZBP-89 is likely to be important for globin gene regulation and possibly globin gene switching.

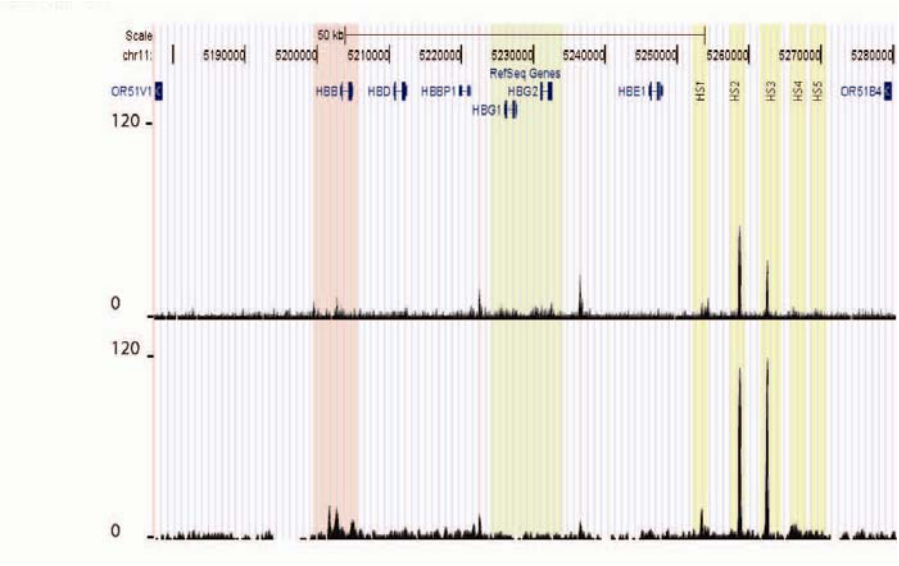


Figure 3) ZBP-89 binding to the human β-globin locus (top) Human β-globin locus on chromosome 11 between two olfactory receptors in UCSC genome browser (hg18). Coordinates are indicated above the locus. Position of LCR DNase I Hyper sensitive sites 1 to 5 (HS1-5) and β-globin like genes are indicated. HBE1; ε-globin, HBG2; Gγ-globin, HBG1; Aγ-globin, HBBP1; ψβ-globin, HBD; δ-globin, HBB; β-globin genes. The scale bar on the top represents 50 kb. (middle and bottom) ChIP sequencing results of endogenous ZBP-89 on human β-globin locus in fetal liver and adult Human Erythroid Progenitors (HEP), respectively. The y axis represent normalized reads per million.

Reduced ZBP-89 correlates with increased γ-globin expression

We therefore carried out a lentiviral shRNA-mediated knockdown of ZBP-89 to test the effect of this factor on globin gene expression in AD-HEP cells. We used several different ZBP-89 shRNAs and two of these resulted in a substantial depletion of ZBP-89 when compared to non-targeting shRNAs (Fig.4A and not shown). We carried out total mRNA sequencing to identify the target genes of ZBP-89 depletion.

The expression of many genes was induced upon ZBP-89 knock down in agreement with its binding of the NuRD complex. However there is also a much smaller set of down-regulated genes (Supplementary Table 1). This suggests that ZBP-89 has both a suppressor and activator function in HEP cells.

A functional annotation of genes bound by ZBP-89 and differentially expressed after its knock down (Supplementary Table 2) shows an overrepresentation of hemopoiesis, JNK and MAPKKK protein kinase cascades, WW protein domains and erythroid differentiation (Supplementary Table 3).

Interestingly the reduced level of ZBP-89 correlated with increased γ-globin mRNA and protein expression and an increase in the number of expressing cells in several donors (Fig. 4B and not shown). The embryonic ε-globin was also increased while the adult stage δ- and β-globins were decreased (Fig. 4C). It is very interesting to see that ZBP-89 knock down reduces BCL11a levels by 3.7 fold. This suggests that ZBP-89 has a suppressive role for embryonic and fetal stage globins and activates the adult globins. We therefore carried out a 3C-Seq analysis to study the interactions in the β-globin locus in normal HEP cells and HEP cells in which ZBP-89 has been knocked down.

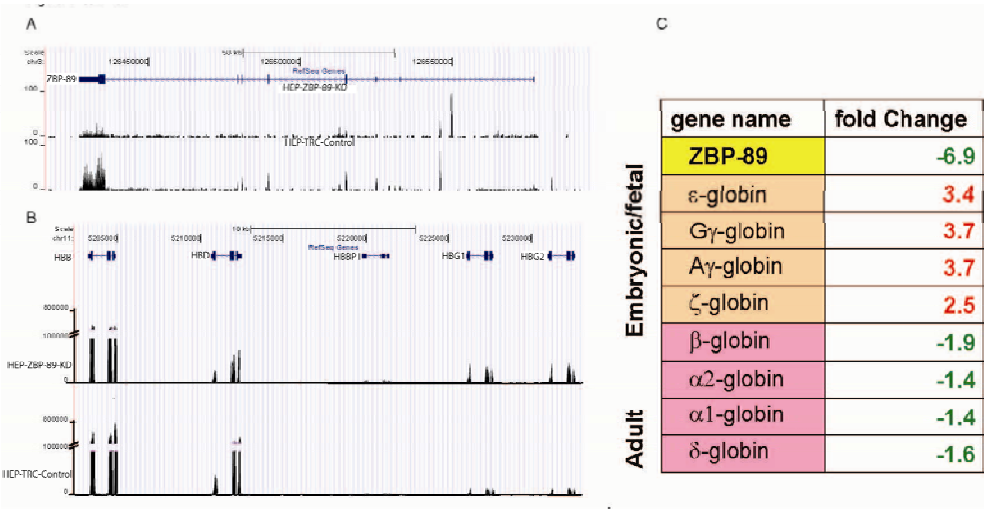


Figure 4) Reduced expression of ZBP-89 is associated with increase human embryonic and fetal and decreased adult globin genes. A) RNA sequencing of ZBP-89 transcripts in adult HEP cells after Lentiviral-mediated ZBP-89 shRNA knock down. (top) ZBP-89 locus on chromosome 3 in UCSC genome browser (hg18). Coordinates are indicated above the locus. The scale bare on the top represents 50 kb. (middle) ZBP-89 expression after shRNA knock down in adult HEP cells. (bottom) ZBP-89 expression after scrambled shRNA knock down in adult HEP cells. The y-axis represent normalized reads per million. B) RNA sequencing of β-globin like genes transcripts in adult HEP cells after ZBP-89 knock down. (top) β-globin locus on chromosome 11 in UCSC genome browser (hg18). Coordinates are indicated above the locus. The scale bare on the top represents 10 kb. (middle) β-globin like genes expression after ZBP-89 knock down in adult HEP cells. (bottom) β-globin like genes expression after scrambled shRNA knock down in adult HEP cells. The y axis represent normalized reads per million. The β-globin like gene symbols are the same as figure 3. C) Summary of expression fold change of ZBP-89, embryonic, fetal and adult globin genes upon ZBP-89 knock down in adult HEP cells compare to scrambled control. Down regulated genes are indicated in green, up regulated genes are indicated in red. Embryonic and fetal β-globin like genes are boxed light pink and adult β-globin like genes are boxed in dark pink.

Reduced ZBP-89 in adult primary human erythroid progenitors restores the β-globin locus back towards a fetal liver stage specific conformation.

ZBP-89 was knocked down by lentiviral-mediated shRNA (Supplementary Fig. 2), followed by Chromosome Conformation Capture sequencing (3C-Seq) (Fig. 5). The human β-globin locus contains a Locus Control Region, consisting of five Hyper Sensitive sites LCR (HS1-5), upstream of the globin

genes ε- (HBE), Gγ- (HBG2), Aγ- (HBG1), δ- (HBD) and β- (HBB) (Fig. 5A). The interaction of different anchor sequences with other restriction fragments throughout the locus, were measured by the number of normalised sequencing reads and represented by the height of restriction blocks (Figure 5B-G), in FL-HEPs (top), AD-HEPs after ZBP-89 knock down (middle) and AD-HEPs after a scrambled sequence shRNA knock down (bottom).

The decrease in ZBP-89 protein in AD-HEPs causes increased interaction of HS4 with HS3 and the HBG1 and 2 genes as well as their 3.7 kb downstream region similar to what is seen in FL-HEPs (Fig. 5B). The interaction of HS3 with HS1, 4 and 5 and the 3.7 kb region of HBG1 and 2 also increases, again similar to the FL-HEPs (Fig. 5C). The interaction of HS2 with the HBG 4 kb up-, 3.7 kb down-stream region and sequences close to the HBE promoter are increased, while the interactions with HS1 and HS3 is reduced (Fig. 5D). HS1 interactions increase with the HBG genes and their 3.7 kb downstream region while there is a decrease with HS2, 3, 5 and the HBB gene, very similar to the situation observed in FL-HEPs (Fig. 5E). HBG1 and 2 interactions increase with the region 3.7 kb downstream of the genes, and HS4 and HS2 (Fig. 5F). On the other hand, HBB interactions with the rest of the locus do not change upon ZBP-89 knock down when compared to scrambled shRNA lentiviral transduced cells (Fig. 5G). Collectively, we conclude from these data that reduced levels of ZBP-89 protein causes a change in spatial conformation towards that seen in fetal liver human primary erythroid progenitors.

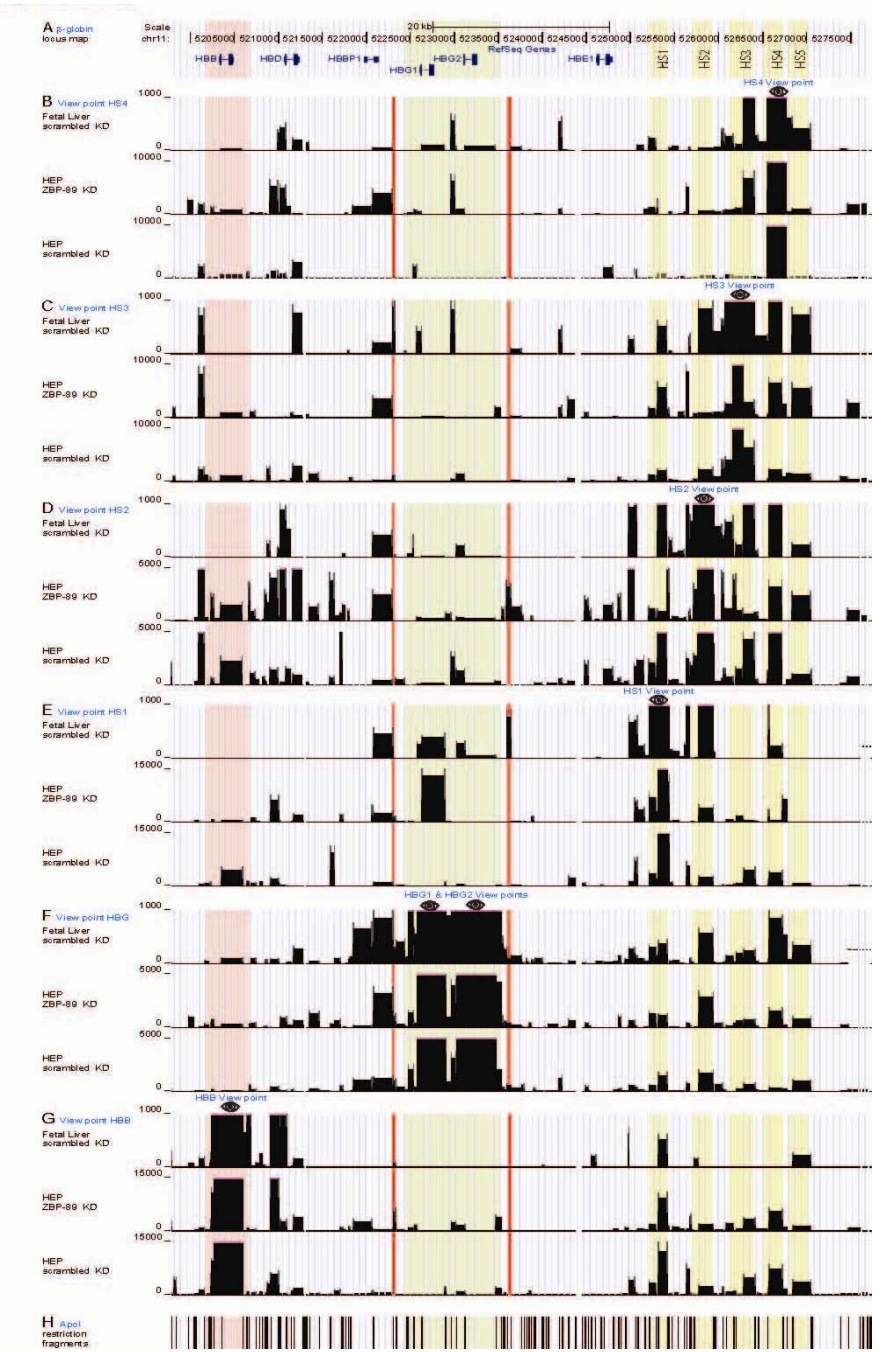

A β-globin locus map
Scale
chr11:
20 kb
5205000 5210000 5215000 5220000 5225000 5230000 5235000 5240000 5245000 5250000 5255000 5260000 5265000 5270000 5275000
RefSeq Genes
HBB
HBD
HBBP1
HBG1
HBG2
HBE1
HS1
HS2
HS3
HS4
HS5
B View point HS4
Fetal Liver scrambled KD
HEP ZBP-89 KD
HEP scrambled KD
HS4 View point
C View point HS3
HS3 View point
D View point HS2
HS2 View point
E View point HS1
HS1 View point
F View point HBG
HBG1 & HBG2 View points
G View point HBB
HBB View point
H ApoI restriction fragments

Figure 5) Reduced expression of ZBP-89 protein in adult HEP cells cause a spatial conformation change in β-globin locus that resembles to the fetal liver stage A) Human β-globin locus on chromosome 11 in UCSC genome browser (hg18). Coordinates are indicated above the locus. Position of the LCR DNase I Hyper sensitive sites 1 to 5 (HS1-5) (yellow shading), and β-globin like genes are indicated. HBE1; ε-globin, HBG1 and 2; Aγ- and Gγ-globins, respectively (pale green shading) HBBP1; ψβ-globin, HBD; δ-globin, HBB; β-globin gene (pink shading). The scale bare on the top represents 10 kb. The red lines indicate the positions of residual ZBP-89 signals in figure 3A, 4 kb upstream and 3.7 kb downstream of the γ-globin genes. B-G) Chromosome confirmation capture sequencing (3C-Seq) of β-globin locus in fetal liver HEP cells (top), ZBP-89 kd adult HEP cells (middle) and scrambled control shRNA kd HEP cells (bottom). Viewpoints are indicated as eye symbol on top of the fixed anchor fragment. B) HS4 viewpoint. C) HS3 viewpoint. D) HS2 viewpoint. E) HS1 viewpoint. F) HBG1 and 2 view points. G) HBB viewpoint. H) Human β-globin locus Apo I restriction digestion fragments.

DISCUSSION

ZBP-89 functions as both an activator and a suppressor TF via an interaction with activating and repressing complexes in different cell types. For instance, ZBP-89 represses vimentin gene expression, which has a complex tissue- and developmental-specific pattern, by recruitment of histone deacetylase-1 (Wu et al., 2007). However, ZBP-89 is reported to act as an activator by recruiting the co-activator P300 to $P21^{waf1/cip1}$ and the β-globin locus (Bai and Merchant, 2000; Petrovic et al., 2009; Woo et al., 2011). It has been suggested that the different modes of ZBP-89 activity is modulated by posttranslational modifications such as sumoylation resulting in different activator or suppressor interacting partners proteins (Chupreta et al., 2007). ZBP-89 desumoylation by Five Friends of Methylated Chtop (5FMC) complex for instance results in higher Pol II levels on ZBP-89 target genes (Fanis et al., 2012). These data suggest that the specific protein partners play an important role in determining the role of ZBP-89 in gene regulation and it would be interesting to determine what signals are responsible for the different states of post-transcriptional modification.

We identified ZBP-89 in Targeted Chromatin Purification (TChP) in embryonic day 13.5 mouse fetal liver erythroid progenitor cells carrying a transgenic human γ-globin minilocus as one of the proteins specifically binding to this locus (Pourfarzad et al. submitted). The regulation of the human γ-globin transgene in mouse embryonic day 13.5 mimics the human γ-globin gene switching process around the time of birth when it is in the process of being silenced. ZBP-89 which is known to be involved in

erythroid development and differentiation (Brand et al., 2004; Li et al., 2006; Woo et al., 2011; Woo et al., 2008a) seems to be also important in the switching process.

To investigate the role of ZBP-89 in human globin gene regulation we first identified ZBP-89 interacting partners in two intrinsically different erythroleukemia cell lines, Murine erythroleukemia (MEL) and human myelogenous leukemia (K562). MEL cells express adult globins and represent an erythropoietin-responsive stage of adult erythroid differentiation between the BFU-E and CFU-E (Friend et al., 1971; Marks and Rifkind, 1978). K562 cells, on the other hand, are derived from myelogenous leukemia line (Lozzio and Lozzio, 1975) and have fetal or embryonic erythroblast characteristics, since they express fetal and embryonic globin (Benz et al., 1980). Furthermore, whereas MEL cells can only be induced to erythroid differentiation, K562 cells can differentiate along a myeloid lineage or an erythroid pathway (Green et al., 1993).

We reasoned that if ZBP-89 is involved in differential regulation of β-globin like genes via recruitment of different protein partners to the β-globin locus regulatory region, it may interact with different subunits of activator and repressor complexes in K562 and MEL cells. Mass spectrometry of tagged ZBP-89 protein complexes in these two cell lines indeed resulted in two distinct set of proteins. While ZBP-89 interacts with subunits of the NuRD/MeCP1 repressor complexes as well as other transcription factors known to be involved in γ-globin gene suppression such as BCL11a in adult type cells (MEL), it interacts with the subunits of SWI/SNF, mediators and P400/Trrap in embryonic/fetal type cells (K562) (Table 1 and Fig. 1D). Interestingly we also detect NIPBL-1 and MAU1 in K562 with a high Mascot score. NIPBL generally co-occupies the enhancer and core promoter regions bound by mediator and key proteins involved in gene activation via enhancer promoter long-range interactions (Chien et al., 2011; Kagey et al., 2010).

Transcription factors control gene expression by binding to enhancer elements often located at considerable distance from promoter elements, where transcription initiation/elongation takes place (Panne, 2008; Stadhouders et al., 2011). Transcription factors bound to enhancers interact with mediator and P300 co-activator proteins, which in turn bind the transcription initiation complex (Conaway et al., 2005; Kornberg, 2005; Malik and Roeder, 2005; Roeder, 1998; Taatjes, 2010). ZBP-89 appears to be involved in differential β-globin gene regulation during development by recruiting different sets of proteins to the β-globin regulatory regions, and we were therefore interested to identify differential ZBP-89 recruitment to the β-globin locus using ChIP-Seq in AD- and FL-HEPs (Fig. 3). This shows that ZBP-89 binds to the β-globin locus LCR HS2 and 3 both in the embryonic and adult stage. There are also binding signals 4 kb upstream and 3.7 kb downstream of the γ-globin

genes and 1.2 kb downstream of the β-globin gene. Since these regions lack a ZBP-89 consensus-binding site, these signals suggest that the pull down of these sequences occurs via interaction partners that are cross-linked to ZBP-89. Whereas the upstream and downstream γ-globin peaks in FL-HEPs are much more prominent in FL-HEPs compare to AD-HEPs (Fig 3), these peaks change in favour of the peaks just downstream of the β-globin gene and enhancer region in AD-HEPs.

The region 3.7 kb downstream of the HBG1 gene is also co-occupied by Ldb1 and BCL11a and has been suggested to be involved in suppression of γ-globin (Kiefer et al., 2011; Xu et al., 2010). Among all other identified transcription factors binding sites in β-globin locus, ZBP-89 appears to be the only transcription factor that has preferential binding only to HS2 and HS3. Importantly lentiviral-mediated knock down of ZBP-89 leads to specific induction of the embryonic and fetal genes and a reduction of the adult globin genes (Fig. 4). 3C-Seq experiments show that increased γ-globin expression after ZBP-89 knockdown in the AD-HEPs is a consequence of a change in the spatial conformation of the human β-globin locus that is more like the β-globin locus conformation observed in FL-HEP cells. The HBG region and 3.7 kb HBG1 downstream sequence interaction increases with HS1-4, very similar to FL-HEPs (Fig. 5B-F), while the interaction of HS1 with HBB is somewhat decreased (Fig. 5E). This is in agreement with increased interaction of the same region in high HbF expressing cells (Kiefer et al., 2011). A small change in interaction is also seen when looking from HBB gene as the anchor sequence in HEP cells when compared to the scrambled control (Fig. 5G).

Thus this analysis of interactions implies that activation of γ-globin gene expression is accompanied by DNA loops re-configuration back to the fetal liver stage.

We took advantage of ENCODE data and zoomed in to the region between 4 kb upstream of HBG2 and 3.7 kb downstream of HBG1, where the ZBP-89 peaks are observed to determine which other transcription factors are bound at these regions and by which epigenetic changes they are accompanied in the K562 cell line (supplementary Fig 3). Interestingly the 4 kb upstream region is the starting point of H3K4me1. This epigenetic active mark indicative of enhancer sequences (Barski et al, 2007; Heintzman et al, 2007; Wang et al, 2008) continues up to the transcription start site of HBG2. From this point onwards, the histone mark changes to H3K4me3 and H3K27ac. The region 4 kb upstream of the HBG2 gene is also co-occupied by YY1, MAX, MYC, Pol II and NF-E2. The site is DNase hypersensitive and nucleosome depleted. The 3.7 kb downstream region of the HBG1 gene has another ZBP-89 peak and is also co-occupied by BCL11a, Ldb1, Myc and Max. The MYC-MAX heterodimer binds to an E-box sequence (CACGTG), which are present in the upstream and

downstream sites of HBG1 and 2 respectivly and recruits the co-activator TRRAP (transformation/transcription domain-associated protein) and histone acetyltransferases (HAT). Acetylation of H3K9 alters the chromatin structure into an active gene expression configuration. Conversely, the MAD transcription factor antagonizes c-MYC function by forming dimers with MAX and recruiting co-repressors such as YY1, SIN3 and HDACs to the target DNA, leading to histone deacetylation and subsequent repression (de Nigris et al., 2007; Pelengaris et al., 2002). Binding of MAX, YY1, MYC, BCL11a and Ldb1 and stage specific presence of ZBP-89 in the same motif, strongly suggest that ZBP-89 is involved in the preferential regulation expression of HBG by interaction with positive regulators at the embryonic/fetal stage of expression. This swithches to an interaction of with repressor complex subunits in adult stage cells (MEL). This suggests that the post transcriptional modification of ZBP-89 plays an important role in the dual function and consequent role in switching.

Material and methods

DNA constructs

ZBP-89 was C-terminally tagged by cloning the ZBP-89 cDNA (ATCC-10324191) into the pBud-V5-Bio expression vector as previously described (Kolodziej et al. 2009). Bio-tagged ZBP-89 was transfected into the BirA biotin ligase expressing MEL cells (de Boer 2003) and K562 cell line. Stable MEL cell clones expressing the tagged protein were isolated and used for subsequent experiments.

Cell culture

MEL and K562 cell lines were cultured in DMEM supplemented with 10% fetal calf serum and penicillin/streptomycin. MEL cells were differentiated in the log phase by adding 2% dimethylsulfoxide (DMSO) for 3 days as already described (de Boer 2003).

The culturing of human erythroid pro-erythroblasts (HEPs) was essentially as published with minor modifications (Leberbauer et al., 2005; von Lindern, 2001). Human buffy coat was obtained from the blood bank (Sanquin, The Netherlands), and blood peripheral mononuclear cells were collected from the interphase after Ficoll step gradient purification, and washed. For initial expansion, $5x10^6$ cells ml^{-1} were cultivated in serum-free medium (StemSpan; Stem Cell Technologies, Vancouver, BC, Canada) supplemented with Epo (2 U ml^{-1} EPREX Ortho-Biotech, Tilburg, The Netherlands), the synthetic glucocorticoid Dexamethasone (Dex, 1 μM; Sigma, St Louis, MO), SCF (100 ng ml^{-1}; R&D Systems, Minneapolis, MN), and lipids (40 μg ml^{-1} cholesterol-rich lipid mix; Sigma) for about 5 days. The cells were fed every day by partial medium changes with the same medium. After 5 days, large pro-erythroblasts became visible and by day 7 they started to overgrow the culture. In order to purify pro-erythroblasts from lymphocytes and macrophages a Percoll purification was carried out and the cells were isolated from the interphase. Homogeneous cultures of erythroid progenitors were kept in the same medium at 1.5-$2x10^6$ cells ml^{-1} by daily partial medium changes. Proliferation kinetics and size distribution of the cells were monitored daily using an electronic cell counter (CASY-1; Schärfe System, Reutlingen, Germany).

Nuclear protein extraction precipitation

Endogenous and tagged ZBP-89 interacting proteins were precipitated using either ZBP-89 antibody (ab69933), V5 agarose beads (Sigma) or Streptavidin paramagnetic beads (Invitrogen) as described earlier (Rodriguez et al., 2006; Kolodziej et al., 2009).

Western blotting

Samples were resolved by 4-12% gradient SDS-PAGE (Invitrogen), and transferred to 0.45-µm nitrocellulose membranes. The membranes were blocked for 1 hour in 1xTBS with 2% BSA and 0.05% Tween20 and incubated in primary (overnight) and secondary (1 hour) antibody. The primary antibodies were purchased from Abcam: Mta2 (ab66051), RbAp48 (ab488), ZNF148 (ab69933) and Santa Cruz: GATA1 (sc-265), MBD3 (sc-9402), P300 (sc-584), Trrap (sc-5405), GCN5 (sc-20698), Mi2beta (sc-8774), Bcl11a (sc-56013), Zfp219 (ab71279), Wdr5 (ab22512) or from Absea Biotechnology Ltd for Chtop.

Western blots were developed with the ECL Western Blotting Detection Reagents (GE Healthcare).

Mass spectrometry

1D 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) or an LTQ linear ion trap (Thermo), 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 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 40.

Chromatin preparation

Crosslinked chromatin was prepared by adding formaldehyde directly to culture medium to a final concentration of 1%, and incubated for 10 minutes at room temperature. The reaction was stopped by the addition of 1 M glycine to a final concentration 0.125 M. The cells were spun at 450*g* for 3 min at 4 °C (Eppendorf centrifuge 5810R), the supernatant was decanted and the cells were washed 2 times with ice cold PBS. The cells were resuspended at a density of $2x10^8$ cells ml^{-1} in sonication buffer (10 mM Tris-HCl, 1 mM EDTA, 0.5 mM EGTA, pH 8.0) plus 1x complete EDTA-free protease inhibitor (Roche), and incubated for 10 min on ice. The cells were disrupted and the DNA sheared to 200-800 bp fragments by sonication at 0 °C (Sanyo, SoniPrep150 large probe, 20 times, amplitude 9; 15" on; 45" off). The sample was centrifuged for 10 minutes at 15,000*g* (Eppendorf centrifuge 5415D) at 4 °C, and the supernatant aliquoted. The pellet was discarded. The samples (chromatin) were snap-frozen in liquid nitrogen and stored at –80 °C. The efficiency of sonication was analyzed by adding 8 μl 5 M NaCl to 200 μl chromatin sample and the crosslinks were reversed at 65 °C for 4 hours. The DNA was recovered by phenol chloroform extraction and run on an agarose gel to analyze shearing efficiency.

Chromatin Imunopercipitation (ChIP)

A chromatin sample equal to $10x10^6$ cells (i.e. 100 μl) was taken up in 1 ml ChIP dilution buffer (0.01% SDS, 1.1% Triton X100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 150 mM NaCl) containing 1x complete EDTA-free protease inhibitor (Roche). Chromatin samples were pre-cleared by incubation of the chromatin with 60μl (50% slurry) of Protein A agarose/salmon sperm DNA (Millipore) for 30 min at 4°C with rotation. The agarose was pelleted by brief centrifugation for 1min at 400*g* (Eppendorf centrifuge 5415D) and the supernatant was taken up in a fresh eppendorf tube. At this stage 20 μl (2%) of the pre-cleared chromatin was kept to serve as input and DNA was extracted along with eluted material in the end of procedure. The pre-cleared chromatin was incubated with 10μg of ZBP-89 antibody (ab69933) and incubated overnight at 4°C with rotation. Protein A agarose Salmon Sperm DNA beads (Millipore) were added the next day and incubated at 4 °C with rotation for another hour. The agarose was pelleted by centrifugation (1 min at 400*g*) and the supernatant carefully removed. The protein A agarose::antibody::protein::chromatin complex was washed for 3-5 minutes on a rotating platform with 1 ml of each of the buffers in the following order, all at 4 °C; a Low Salt Immune Complex Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), one wash; a High Salt Immune Complex Wash Buffer (0.1% SDS, 1% Triton X-

100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), one wash; LiCl Immune Complex Wash Buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.1), one wash; TE Buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), two washes. At this stage the protein::chromatin (ZBP-89::DNA) complex was eluted from protein A agarose::antibody by two addition of 250 μl of elution buffer (1% SDS, 0.1 M Na_2HCO_3), incubated at room temperature for 15 min with rotation. The elution was pooled (500 μl) and the 20 μl input sample that was saved in the beginning of procedure was taken up in 500 μl of elution buffer. The elute and input material were de-crosslinked by addition of 20 μl of 5 M NaCl incubated for 4 hours at 65°C on shaking platform. The de-crosslinked material was digested by addition of 10 μl of 0.5 M EDTA, 20 μl Tris-HCl, pH 6.5, and 2 μl of 10 mg/ml Proteinase K, incubated at 45°C for 1 hour. DNA was cleaned by two phenol/chloroform extraction and precipitated by addition of 20 μg glycogen (Roche) followed by ethanol precipitation. The DNA pellet was washed with 70% ethanol and resuspended in double distilled de-ionized water (200 μl for ZBP-89 precipitated sample) and (400 μl for in put control).

Quantitative PCR of ZBP-89 precipitated DNA

Quantitative real-time PCR (MyIQ, Bio Rad) was performed using 0.75 μl of SYBR Green I (Sigma S9430) 1/2,000 dilution in DMSO, Platinum Taq kit (Invitrogen), 10 pmol of each primer, 4 μl DNA sample under the following cycling conditions: 3 min at 95 °C followed by 40 cycles of 30 sec. at 95 °C, 20 sec. at 56 °C, 40 sec. at 60 °C, 15 sec. at 75 °C. Aγ promoter was used as endogenous references. Enrichment of specific sequences was calculated using the comparative CT method (Livak and Schmittgen, 2001).

Primers used for Quantitative PCR:

βLCR HS4-F	5'>TTTCTCTCTCCCACTCAGC
βLCR HS4-R	5'>TGCTATCAAAGCCCTGACA
βLCR HS3-F	5'>GCTCAGATAGGTGGTTAGGT
βLCR HS3-R	5'>TGGTCTATCTCTCCTGGCT
βLCR HS2-F	5'>CTCCATTAGTGACCTCCCA
βLCR HS2-R	5'>TTACACAGAACCAGAAGGC
βLCR HS1-F	5'>TGCGGTTGTGGAAGTTTAC
βLCR HS1-R	5'>CACTAAGGGTGAGGATGCT
Aγpromoter-F	5'>GAAACGGTCCCTGGCTAAACTC

Aγpromoter-R	5'>CCTCACTGGATACTCTAAGACTATTGG
Aγ 3.7 kb downstream-F	5'>AATGACCTAATGCCCAGCAC
Aγ 3.7 kb downstream –R	5'>AGTGTTGGGGGAGAAGTGTG
Aγ 4 kb upstream-F	5'>TCAGCAGAGGCAGTCAGG
Aγ 4 kb upstream –R	5'>TGATGGTCCTTCCTTACTCAC

ChIP Sequencing

Cross-linked chromatin from 1 × 10^8 cells were used per ChIP-Seq experiment. Chromatin preparations and precipitations were performed as described for the ChIP experiment above. The precipitated material was sequenced by Illumina Hi-Seq. 10 ng of ChIPped DNA was end-repaired, ligated to adapters, size selected on gel (200±25 bp range) and PCR amplified using Phusion polymerase as follow: 30sec at 98ºC, 18 cycles of (10sec at 98ºC, 30sec at 65ºC, 30sec at 72ºC), 5min at 72ºC final extension. Cluster generation was performed using the Illumina Cluster Reagents preparation, and the library was sequenced on the Illumina Genome Analyzer II platform to generate 36 bp reads. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP). The raw data from the Illumina Genome Analyzer were processed using the IPAR (Integrated Primary Analysis Reporting Software) and the GAP. The resultant sequences were mapped against NCBI build 37.1 of the human genome using the ELAND alignment software (Illumina).

Lentiviral-mediated shRNA Knock down

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 hours and the cells were kept in culture for another 5 days before harvesting and further analysis.

RNA sequencing

HEP cells were harvested after lentiviral-mediated shRNA knock down. Total RNA was extracted from cells using the TRI reagent (Sigma) and used directly for RNA sequencing analysis. RNA sequencing was performed on an Illumina Hi-Seq as described (Soleimani et al, 2012).

Chromosome Conformation Capture sequencing

3C material was prepared essentially as described (Palstra et al., 2003b; Soler et al., 2010). ApoI was used as the primary restriction endonuclease. Purified 3C DNA was used for 3C-Seq library preparation as described (Palstra et al., 2003b; Soler et al., 2010; Stadhouders et al., 2011), using DpnII as secondary restriction endonuclease. Viewpoint-interacting DNA fragments were PCR-amplified using viewpoint-specific divergent primers linked to standard Illumina adapter sequences. The sequence of the primers is follow:

HS1 viewpoint	
ApoI primer:	AGTCCTTAGCTTTATGTTGCTG
DpnII primer:	GCAAGGCTTAAATGGAAGAAG
HS2 viewpoint	
DpnII primer:	ATGTTTCTTTCCTCTCAGGATC
ApoI primer:	GTGACATGACTTAAGGAACTATAC
HS3 viewpoint	
ApoI primer:	CCCAAGGTTCTGAACATGAG
DpnII primer:	AGAGTGTGCATCTCCTTTGA
HS4 viewpoint	
DpnII primer:	TAGCCAGGCATGTGATGTA
ApoI primer:	TGTAGGGTTGAGAGGGAGAT
HS5 viewpoint	
ApoI primer:	ACACTAGAAGGCAGGTATTATCT
DpnII primer:	CAAAGAGTAAAAGGATTGAAGCC
HBG2 5'ZBP-89 binding site	
ApoI primer:	TGGAGTCTGCACACCAAA
DpnII primer:	TGGAGTCTGGTAGACCTGAT
HBG2 viewpoint	
ApoI primer:	TTGGCAATCCATTTCGGC
DpnII primer:	TCCTCTGCAGTTTCTTCACT
HBG1 viewpoint	
ApoI primer:	TTGGCAATCCATTTCGGC
DpnII primer:	CCTCTGCAGTTTCTTCACC

HBG1 3'ZBP-89 binding site view point

DpnII primer: GAAGAATGTCTTTGCCTGATTAG

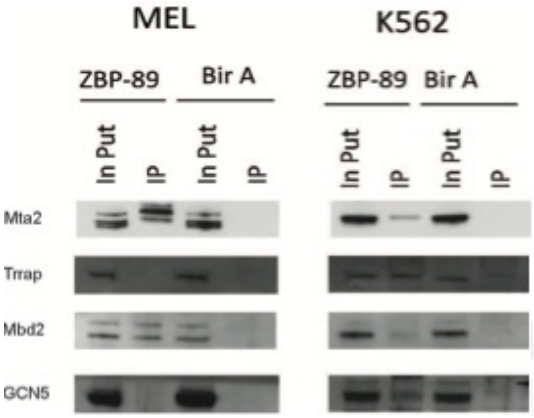
ApoI primer: CATGTGTATTTCAAGGAGACTCT

HBB-promoter viewpoint

ApoI primer: CCATCACTTTGGCAAAGAATTC

DpnII primer: GAAAACAATTGTTATGAACAGCAA

The resulting 3C 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 3C-Seq pipeline (Thongjuea et al., unpublished). 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.



Supplementary Figure 1) ZBP-89 forms distinct complexes in MEL cells and K562 cells. ZBP-89-V5-Bio precipitation with V5 agarose beads in MEL (left panel) and K562 (right panel) cells.

Supplementary Table 1
Differentially (up) regulated genes upon ZBP-89 knock down

Gene name	Fold change	padj	KD read count	WT read count	
CRYM	141.7	2E-23	4202	32	crystallin mu
ENC1	135.4	1E-06	251	2	ectodermal-neural cortex 1 (with BTB-like domain)
CTSE	119.3	3E-19	2321	21	cathepsin E
TMEM74	106.2	9E-13	975	10	transmembrane protein 74
KALRN	47.8	9E-04	133	3	kalirin RhoGEF kinase
RAB32	47.5	5E-02	44	1	RAB32 member RAS oncogene family
TNFRSF25	43.7	1E-06	405	10	tumor necrosis factor receptor superfamily member 25
ADRA1D	39.6	4E-03	110	3	adrenergic alpha-1D- receptor
CD52	38.0	1E-08	845	24	CD52 molecule
KCNK1	36.2	2E-07	637	19	potassium channel subfamily K member 1
ZFPM2	27.8	3E-04	232	9	zinc finger protein multitype 2
MRC2	27.8	7E-05	283	11	mannose receptor C type 2
NXF3	26.5	4E-07	761	31	nuclear RNA export factor 3
TIMP4	25.4	1E-04	283	12	TIMP metallopeptidase inhibitor 4
EPS8L3	23.6	7E-05	348	16	EPS8-like 3
ESPNP	22.4	3E-09	1824	88	espin pseudogene
NPY1R	21.8	2E-03	182	9	neuropeptide Y receptor Y1
ESPN	20.9	2E-11	20007	1032	espin
MT1F	20.8	5E-02	77	4	metallothionein 1F
G0S2	20.1	5E-08	1450	78	G0/G1switch 2
SNTA1	19.6	4E-05	489	27	syntrophin alpha 1 (dystrophin-associated protein A1 59kDa acidic component)
CCR7	19.2	6E-06	713	40	chemokine (C-C motif) receptor 7
SLC16A3	18.7	1E-04	382	22	solute carrier family 16 member 3 ([Source:HGNC Symbol Acc:10924]
SLC25A24	18.4	2E-07	1363	80	solute carrier family 25 member 24 [Source:HGNC Symbol Acc:20662]
MT1E	18.1	4E-05	519	31	metallothionein 1L (gene/pseudogene) [Source:HGNC Symbol Acc:7404]
PLXNB3	17.2	9E-06	766	48	plexin B3 [Source:HGNC Symbol Acc:9105]
CACHD1	16.6	5E-06	892	58	cache domain containing 1 [Source:HGNC Symbol Acc:29314]
MT1G	14.2	9E-06	999	76	metallothionein 1G [Source:HGNC Symbol Acc:7399]
FAM49A	13.3	1E-03	358	29	family with sequence similarity 49 member A [Source:HGNC Symbol Acc:25373]
PDGFA	13.0	6E-05	757	63	platelet-derived growth factor alpha polypeptide [Source:HGNC Symbol Acc:8799]
MT1A	12.8	1E-02	201	17	metallothionein 1A [Source:HGNC Symbol Acc:7393]
SNPH	12.1	3E-02	168	15	syntaphilin [Source:HGNC Symbol Acc:15931]
C11orf21	11.9	2E-07	3225	292	Uncharacterized protein C11orf21 [Source:UniProtKB/Swiss-Prot Acc:Q9P2W6]
TBC1D12	11.7	2E-04	653	60	TBC1 domain family member 12 [Source:HGNC Symbol Acc:29082]
RELN	11.6	1E-07	19436	1818	reelin [Source:HGNC Symbol Acc:9957]
MT1H	10.6	3E-02	196	20	metallothionein 1H [Source:HGNC Symbol Acc:7400]
EGFL7	10.0	3E-05	1368	147	EGF-like-domain multiple 7 [Source:HGNC Symbol Acc:20594]
TSPAN32	8.9	1E-04	1258	152	tetraspanin 32 [Source:HGNC Symbol Acc:13410]
FAM178B	7.7	2E-05	3716	520	family with sequence similarity 178 member B [Source:HGNC Symbol Acc:28036]
SLC27A4	7.6	5E-04	1256	179	solute carrier family 27 (fatty acid transporter) member 4 [Source:HGNC Symbol Acc:10998]
MT1X	7.5	2E-03	947	137	metallothionein 1X [Source:HGNC Symbol Acc:7405]
MT2A	7.3	4E-05	3102	461	metallothionein 2A [Source:HGNC Symbol Acc:7406]
PRG2	7.1	4E-05	20446	3102	proteoglycan 2 bone marrow [Source:HGNC Symbol Acc:9362]
MT1P1	7.1	5E-03	763	116	metallothionein 1 pseudogene 1 [Source:HGNC Symbol Acc:23681]
ORAI3	6.8	2E-03	1180	186	ORAI calcium release-activated calcium modulator 3 [Source:HGNC Symbol Acc:28185]
TPSB2	6.4	3E-03	1200	203	tryptase beta 2 (gene/pseudogene) [Source:HGNC Symbol Acc:14120]
LAMB3	6.3	1E-02	822	141	laminin beta 3 [Source:HGNC Symbol Acc:6490]
TPSAB1	6.3	2E-03	1426	246	tryptase alpha/beta 1 [Source:HGNC Symbol Acc:12019]
BEX1	6.2	2E-04	3435	594	brain expressed X-linked 1 [Source:HGNC Symbol Acc:1036]
SERINC2	6.2	3E-03	1204	209	serine incorporator 2 [Source:HGNC Symbol Acc:23231]
SFRP5	6.1	7E-04	2132	379	secreted frizzled-related protein 5 [Source:HGNC Symbol Acc:10779]
TNIK	6.1	2E-02	624	111	TRAF2 and NCK interacting kinase [Source:HGNC Symbol Acc:30765]
LRRC4	6.0	1E-02	806	144	leucine rich repeat containing 4 [Source:HGNC Symbol Acc:15586]
PFKFB4	5.8	7E-03	1148	215	6-phosphofructo-2-kinase/fructose-2 6-biphosphatase 4 [Source:HGNC Symbol Acc:8875]
TNFRSF19	5.5	3E-02	706	138	tumor necrosis factor receptor superfamily member 19 [Source:HGNC Symbol Acc:11915]
DHRS11	5.1	8E-03	1636	347	dehydrogenase/reductase (SDR family) member 11 [Source:HGNC Symbol Acc:28639]
SEMA4C	4.9	6E-03	1994	436	sema domain immunoglobulin domain (Ig) 4C [Source:HGNC Symbol Acc:10731]
STARD8	4.7	3E-02	1022	234	StAR-related lipid transfer (START) domain containing 8 [Source:HGNC Symbol Acc:19161]
STX3	4.6	3E-03	5428	1266	syntaxin 3 [Source:HGNC Symbol Acc:11438]
TJP1	4.6	5E-02	922	217	tight junction protein 1 (zona occludens 1) [Source:HGNC Symbol Acc:11827]
DUSP6	4.4	5E-02	1145	284	dual specificity phosphatase 6 [Source:HGNC Symbol Acc:3072]
C12orf23	4.2	4E-02	1382	354	UPF0444 transmembrane protein C12orf23 [Source:UniProtKB/Swiss-Prot Acc:Q8WUH6]
FASN	4.1	2E-02	30649	8078	fatty acid synthase [Source:HGNC Symbol Acc:3594]
MIF	4.0	2E-02	2830	762	macrophage migration inhibitory factor [Source:HGNC Symbol Acc:7097]
LDHA	3.9	2E-02	18086	4968	lactate dehydrogenase A [Source:HGNC Symbol Acc:6535]
AHNAK	3.9	2E-02	6312	1757	AHNAK nucleoprotein [Source:HGNC Symbol Acc:347]
RALB	3.8	5E-02	1886	530	v-ral simian leukemia viral oncogene homolog B [Source:HGNC Symbol Acc:9840]
PAQR9	3.8	2E-02	4295	1233	progestin and adipoQ receptor family member IX [Source:HGNC Symbol Acc:30131]
AP000350.4	3.7	4E-02	2350	681	Putative uncharacterized protein ENSP00000383961 [Source:UniProtKB/TrEMBL Acc:B5MCZ7]

Supplementary Table 1 (Continued)
Differentially (down) regulated genes upon ZBP-89 knock down

Gene name	Fold change	padj	KD read count	WT read count	
IMPA2	-3.4	5E-02	1440	5356	inositol(myo)-1(or 4)-monophosphatase 2 [Source:HGNC Symbol Acc:6051]
FOS	-3.6	4E-02	899	3506	FBJ murine osteosarcoma viral oncogene homolog [Source:HGNC Symbol Acc:3796]
RB1	-3.6	4E-02	3541	13833	retinoblastoma 1 [Source:HGNC Symbol Acc:9884]
TCP11L2	-3.7	2E-02	1528	6173	t-complex 11 (mouse)-like 2 [Source:HGNC Symbol Acc:28627]
PHGDH	-3.8	2E-02	1737	7031	phosphoglycerate dehydrogenase [Source:HGNC Symbol Acc:8923]
RGS16	-3.9	3E-02	662	2799	regulator of G-protein signaling 16 [Source:HGNC Symbol Acc:9997]
MAT2B	-4.0	1E-02	1847	7941	methionine adenosyltransferase II beta [Source:HGNC Symbol Acc:6905]
DAPK1	-4.0	2E-02	5987	25915	death-associated protein kinase 1 [Source:HGNC Symbol Acc:2674]
TMCC2	-4.2	1E-02	3600	16239	transmembrane and coiled-coil domain family 2 [Source:HGNC Symbol Acc:24239]
RNF11	-4.2	1E-02	2494	11305	ring finger protein 11 [Source:HGNC Symbol Acc:10056]
CA3	-4.4	1E-02	623	2944	carbonic anhydrase III muscle specific [Source:HGNC Symbol Acc:1374]
CCNA2	-4.6	5E-03	3700	16307	cyclin A2 [Source:HGNC Symbol Acc:1578]
CNN3	-4.8	3E-03	788	4049	calponin 3 acidic [Source:HGNC Symbol Acc:2157]
ARG1	-4.8	5E-02	177	914	arginase liver [Source:HGNC Symbol Acc:663]
FMNL3	-4.8	1E-02	330	1725	formin-like 3 [Source:HGNC Symbol Acc:23698]
PECR	-5.0	1E-02	302	1616	peroxisomal trans-2-enoyl-CoA reductase [Source:HGNC Symbol Acc:18281]
RASGRP3	-5.1	1E-02	247	1363	RAS guanyl releasing protein 3 (calcium and DAG-regulated) [Source:HGNC Symbol Acc:14545]
CLIC4	-5.2	2E-03	5120	28466	chloride intracellular channel 4 [Source:HGNC Symbol Acc:13518]
ID1	-5.9	5E-04	510	3239	inhibitor of DNA binding 1 dominant negative helix-loop-helix protein [Source:HGNC Symbol Acc:5360]
GPNMB	-6.0	3E-02	101	653	glycoprotein (transmembrane) nmb [Source:HGNC Symbol Acc:4462]
ZNF711	-6.0	1E-02	135	873	zinc finger protein 711 [Source:HGNC Symbol Acc:13128]
MTA3	-6.5	4E-02	63	442	metastasis associated 1 family member 3 [Source:HGNC Symbol Acc:23784]
PELI2	-6.6	2E-02	79	560	pellino homolog 2 (Drosophila) [Source:HGNC Symbol Acc:8828]
ZNF148	-6.9	8E-04	208	1556	zinc finger protein 148 [Source:HGNC Symbol Acc:12933]
C16orf45	-6.9	1E-02	87	652	Uncharacterized protein C16orf45 [Source:UniProtKB/Swiss-Prot Acc:Q96MC5]
RAP1GAP	-7.1	3E-05	1232	9393	RAP1 GTPase activating protein [Source:HGNC Symbol Acc:9858]
IL16	-7.3	3E-03	113	895	interleukin 16 (lymphocyte chemoattractant factor) [Source:HGNC Symbol Acc:5980]
ITGB1	-7.9	2E-05	3101	26317	integrin beta 1[Source:HGNC Symbol Acc:6153]
CHD3	-9.0	1E-05	280	2733	chromodomain helicase DNA binding protein 3 [Source:HGNC Symbol Acc:1918]
RGS9	-9.8	5E-02	18	190	regulator of G-protein signaling 9 [Source:HGNC Symbol Acc:10004]
IGFBP4	-16.8	2E-02	8	145	insulin-like growth factor binding protein 4 [Source:HGNC Symbol Acc:5473]
TSLP	-16.9	1E-04	27	491	thymic stromal lymphopoietin [Source:HGNC Symbol Acc:30743]

Supplementary Table 2 - ZBP-89 bound genes and up-regulated upon its knock down

Gene name	Fold change	padj	KD read count	TRC read count	description
KALRN	123,5	0,00	133	3	kalirin, RhoGEF kinase [Source:HGNC Symbol;Acc:4814]
TCTEX1D1	56,5	0,10	58	8	Tctex1 domain containing 1 [Source:HGNC Symbol;Acc:26882]
SOD3	23,2	0,36	30	1	superoxide dismutase 3, extracellular [Source:HGNC Symbol;Acc:11181]
ZMYND12	23,4	0,43	24	13	zinc finger, MYND-type containing 12 [Source:HGNC Symbol;Acc:21132]
PLXNB3	22,6	0,00	766	48	plexin B3 [Source:HGNC Symbol;Acc:9105]
LRRC6	15,1	0,43	31	11	leucine rich repeat containing 6 [Source:HGNC Symbol;Acc:16725]
PCSK9	14,6	0,00	827	201	proprotein convertase subtilisin/kexin type 9 [Source:HGNC Symbol;Acc:20001]
PTPN21	14,4	0,24	74	10	protein tyrosine phosphatase, non-receptor type 21 [Source:HGNC Symbol;Acc:9651]
ZFPM2	14,1	0,04	232	3	zinc finger protein, multitype 2 [Source:HGNC Symbol;Acc:16700]
CREB3L3	14,1	0,53	29	1	cAMP responsive element binding protein 3-like 3 [Source:HGNC Symbol;Acc:18855]
PDE1B	13,1	0,58	27	8	phosphodiesterase 1B, calmodulin-dependent [Source:HGNC Symbol;Acc:8775]
PRG2	11,2	0,00	20446	3102	proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein) [Source:HGNC Symbol;Acc:9362]
EGFL7	11,0	0,00	1368	147	EGF-like-domain, multiple 7 [Source:HGNC Symbol;Acc:20594]
TBC1D12	10,3	0,01	653	60	TBC1 domain family, member 12 [Source:HGNC Symbol;Acc:29082]
TNFRSF19	10,0	0,01	706	138	tumor necrosis factor receptor superfamily, member 19 [Source:HGNC Symbol;Acc:11915]
TPSD1	3,7	0,10	230	34	tryptase delta 1 [Source:HGNC Symbol;Acc:14118]
HHEX	8,1	0,21	184	30	hematopoietically expressed homeobox [Source:HGNC Symbol;Acc:4301]
TNIK	7,6	0,04	624	111	TRAF2 and NCK interacting kinase [Source:HGNC Symbol;Acc:30765]
SLC6A19	7,5	0,00	2301	715	solute carrier family 6 (neutral amino acid transporter), member 19 [Source:HGNC Symbol;Acc:27960]
MAP3K6	7,4	0,18	258	70	mitogen-activated protein kinase kinase kinase 6 [Source:HGNC Symbol;Acc:6858]
GPR98	7,2	0,37	126	29	G protein-coupled receptor 98 [Source:HGNC Symbol;Acc:17416]
AZGP1	6,6	0,37	143	13	alpha-2-glycoprotein 1, zinc-binding [Source:HGNC Symbol;Acc:310]
ANKRD55	6,6	0,59	74	16	ankyrin repeat domain 55 [Source:HGNC Symbol;Acc:25681]
NTN1	6,2	0,35	178	58	netrin 1 [Source:HGNC Symbol;Acc:8029]
PPM1H	6,1	0,01	2221	733	protein phosphatase, Mg2+/Mn2+ dependent, 1H [Source:HGNC Symbol;Acc:18583]
ROR2	6,1	0,59	87	21	receptor tyrosine kinase-like orphan receptor 2 [Source:HGNC Symbol;Acc:10257]
ORAI3	5,2	0,07	1180	186	ORAI calcium release-activated calcium modulator 3 [Source:HGNC Symbol;Acc:28185]
HSPG2	5,2	0,37	243	66	heparan sulfate proteoglycan 2 [Source:HGNC Symbol;Acc:5273]
TSEN2	4,6	0,25	598	253	tRNA splicing endonuclease 2 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:28422]
ENPP3	4,6	0,54	204	48	ectonucleotide pyrophosphatase/phosphodiesterase 3 [Source:HGNC Symbol;Acc:3358]
KIAA0802	4,6	0,25	583	214	KIAA0802 [Source:HGNC Symbol;Acc:23121]
AHNAK	4,5	0,03	6312	1757	AHNAK nucleoprotein [Source:HGNC Symbol;Acc:347]
SFRP5	4,4	0,07	2132	379	secreted frizzled-related protein 5 [Source:HGNC Symbol;Acc:10778]
EGR1	4,3	0,04	9402	14415	early growth response 1 [Source:HGNC Symbol;Acc:3238]
MOBKL2B	4,2	0,35	504	229	MOB1, Mps One Binder kinase activator-like 2B (yeast) [Source:HGNC Symbol;Acc:23825]
C17orf76	4,1	0,07	3509	1204	Leucine-rich repeat-containing protein C17orf76 [Source:UniProtKB/Swiss-Prot;Acc:Q8NAA5]
ACSL4	3,9	0,07	11394	5368	acyl-CoA synthetase long-chain family member 4 [Source:HGNC Symbol;Acc:3571]
LAMB3	3,9	0,30	822	141	laminin, beta 3 [Source:HGNC Symbol;Acc:6490]
RIMS3	3,9	0,37	580	330	regulating synaptic membrane exocytosis 3 [Source:HGNC Symbol;Acc:21292]
PLEC	3,7	0,16	2303	312	plectin [Source:HGNC Symbol;Acc:9063]
SFXN2	3,6	0,19	2005	1036	sideroflexin 2 [Source:HGNC Symbol;Acc:16086]
CHD7	3,6	0,12	5197	2639	chromodomain helicase DNA binding protein 7 [Source:HGNC Symbol;Acc:20626]
FBXW9	3,5	0,52	437	192	F-box and WD repeat domain containing 9 [Source:HGNC Symbol;Acc:28136]
BCORL1	3,5	0,41	788	303	BCL6 corepressor-like 1 [Source:HGNC Symbol;Acc:25657]
PIK3R2	3,4	0,15	3939	3853	phosphoinositide-3-kinase, regulatory subunit 2 (beta) [Source:HGNC Symbol;Acc:8980]
AC124798.1	3,3	0,45	885	338	Putative Ig-like domain-containing protein DKFZp686O24166/DKFZp686I21167 Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q68D85]
C9orf123	3,2	0,43	977	467	Transmembrane protein C9orf123 [Source:UniProtKB/Swiss-Prot;Acc:Q96GE9]
MLLT3	3,2	0,23	3281	1658	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3 [Source:HGNC Symbol;Acc:7136]
SLC30A10	3,2	0,35	1537	605	solute carrier family 30, member 10 [Source:HGNC Symbol;Acc:25355]
WWC2	3,1	0,46	1113	602	WW and C2 domain containing 2 [Source:HGNC Symbol;Acc:24148]
BOP1	2,9	0,31	5873	3664	block of proliferation 1 [Source:HGNC Symbol;Acc:15513]
ACOT9	2,9	0,45	1546	732	acyl-CoA thioesterase 9 [Source:HGNC Symbol;Acc:17152]
SLC25A17	2,9	0,35	2343	1158	solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17 [Source:HGNC Symbol;Acc:10987]
IQGAP2	2,9	0,34	3942	1846	IQ motif containing GTPase activating protein 2 [Source:HGNC Symbol;Acc:6111]
AC145291.1	2,9	0,34	4300	2629	KM-PA-2 protein [Source:UniProtKB/TrEMBL;Acc:Q96Q25]
TTLL4	2,9	0,36	3056	1108	tubulin tyrosine ligase-like family, member 4 [Source:HGNC Symbol;Acc:28376]
SACS	2,8	0,35	3602	2114	spastic ataxia of Charlevoix-Saguenay (sacsin) [Source:HGNC Symbol;Acc:10513]

Supplementary Table 2 (continued) - ZBP-89 bound genes and down-regulated upon its knock down

Gene name	Fold change	padj	KD read count	TRC read count	description
ULBP1	-2,8	0,53	450	1573	UL16 binding protein 1 [Source:HGNC Symbol;Acc:14833]
REEP4	-2,9	0,35	1204	3014	receptor accessory protein 4 [Source:HGNC Symbol;Acc:26176]
STIM1	-2,9	0,33	1537	3938	stromal interaction molecule 1 [Source:HGNC Symbol;Acc:11386]
DGUOK	-2,9	0,41	677	1783	deoxyguanosine kinase [Source:HGNC Symbol;Acc:2858]
PEAR1	-2,9	0,35	930	2386	platelet endothelial aggregation receptor 1 [Source:HGNC Symbol;Acc:33631]
FAM129A	-2,9	0,35	10383	29671	family with sequence similarity 129, member A [Source:HGNC Symbol;Acc:16784]
C6orf161	-2,9	0,38	717	1467	family with sequence similarity 165, member A [Source:HGNC Symbol;Acc:21400]
FMNL2	-2,9	0,54	342	865	formin-like 2 [Source:HGNC Symbol;Acc:18267]
LPIN2	-3,0	0,33	9282	32884	lipin 2 [Source:HGNC Symbol;Acc:14450]
TCEA1	-3,0	0,33	9879	30687	transcription elongation factor A (SII), 1 [Source:HGNC Symbol;Acc:11612]
ABHD4	-3,0	0,29	1453	3238	abhydrolase domain containing 4 [Source:HGNC Symbol;Acc:20154]
U1	-3,0	0,46	395	780	U1 spliceosomal RNA [Source:RFAM;Acc:RF00003]
YPEL5	-3,0	0,26	2244	5737	yippee-like 5 (Drosophila) [Source:HGNC Symbol;Acc:18329]
RTN3	-3,1	0,28	8272	23393	reticulon 3 [Source:HGNC Symbol;Acc:10469]
BCL11A	-3,1	0,55	227	918	B-cell CLL/lymphoma 11A (zinc finger protein) [Source:HGNC Symbol;Acc:13221]
H1F0	-3,2	0,51	61554	177342	H1 histone family, member 0 [Source:HGNC Symbol;Acc:4714]
FYN	-3,2	0,24	1178	3948	FYN oncogene related to SRC, FGR, YES [Source:HGNC Symbol;Acc:4037]
FOXO3	-3,2	0,34	21651	73827	forkhead box O3 [Source:HGNC Symbol;Acc:3821]
GATS	-3,2	0,56	194	507	GATS, stromal antigen 3 opposite strand [Source:HGNC Symbol;Acc:23954]
SMURF2	-3,3	0,46	262	692	SMAD specific E3 ubiquitin protein ligase 2 [Source:HGNC Symbol;Acc:16809]
RB1	-3,3	0,18	3541	13833	retinoblastoma 1 [Source:HGNC Symbol;Acc:9884]
PNRC1	-3,3	0,19	1178	3331	proline-rich nuclear receptor coactivator 1 [Source:HGNC Symbol;Acc:17278]
CBFB	-3,3	0,50	201	590	core-binding factor, beta subunit [Source:HGNC Symbol;Acc:1539]
TBC1D3C	-3,4	0,34	355	933	TBC1 domain family, member 3C [Source:HGNC Symbol;Acc:24883]
TBC1D3G	-3,4	0,35	322	878	TBC1 domain family, member 3G [Source:HGNC Symbol;Acc:29860]
CCDC77	-3,4	0,18	995	1976	coiled-coil domain containing 77 [Source:HGNC Symbol;Acc:28203]
TRPS1	-3,5	0,57	129	368	trichorhinophalangeal syndrome I [Source:HGNC Symbol;Acc:12340]
HIST1H4C	-3,5	0,46	176	554	histone cluster 1, H4c [Source:HGNC Symbol;Acc:4787]
DAPK1	-3,5	0,14	5987	25915	death-associated protein kinase 1 [Source:HGNC Symbol;Acc:2674]
CAB39L	-3,6	0,53	105	262	calcium binding protein 39-like [Source:HGNC Symbol;Acc:20290]
NFE2L1	-3,6	0,12	3612	11431	nuclear factor (erythroid-derived 2)-like 1 [Source:HGNC Symbol;Acc:7781]
MBNL2	-3,6	0,11	1733	6026	muscleblind-like 2 (Drosophila) [Source:HGNC Symbol;Acc:16746]
ITPR1	-3,6	0,35	222	634	inositol 1,4,5-triphosphate receptor, type 1 [Source:HGNC Symbol;Acc:6180]
SH2D3C	-3,7	0,57	90	241	SH2 domain containing 3C [Source:HGNC Symbol;Acc:16884]
RAB33B	-3,9	0,16	432	1580	RAB33B, member RAS oncogene family [Source:HGNC Symbol;Acc:16433]
SLC38A1	-3,9	0,28	228	814	solute carrier family 38, member 1 [Source:HGNC Symbol;Acc:13447]
OSM	-3,9	0,32	194	845	oncostatin M [Source:HGNC Symbol;Acc:8506]
ASS1	-4,0	0,20	294	648	argininosuccinate synthase 1 [Source:HGNC Symbol;Acc:758]
CNN3	-4,0	0,08	788	4043	calponin 3, acidic [Source:HGNC Symbol;Acc:2157]
RAB43	-4,1	0,16	301	812	RAB43, member RAS oncogene family [Source:HGNC Symbol;Acc:19383]
C6orf223	-4,2	0,05	1373	4030	Uncharacterized protein C6orf223 [Source:UniProtKB/Swiss-Prot;Acc:Q8N319]
CLIC4	-4,2	0,06	5120	28466	chloride intracellular channel 4 [Source:HGNC Symbol;Acc:13518]
PHGDH	-4,3	0,04	1737	7031	phosphoglycerate dehydrogenase [Source:HGNC Symbol;Acc:8923]
SH3D20	-4,3	0,33	120	397	Rho GTPase activating protein 27 [Source:HGNC Symbol;Acc:31813]
KIAA0427	-4,5	0,22	156	534	KIAA0427 [Source:HGNC Symbol;Acc:23925]
TNFRSF10B	-4,6	0,02	2503	7011	tumor necrosis factor receptor superfamily, member 10b [Source:HGNC Symbol;Acc:11905]
RNF11	-4,6	0,02	2434	11305	ring finger protein 11 [Source:HGNC Symbol;Acc:10056]
ASS1P1	-5,1	0,57	28	83	argininosuccinate synthetase 1 pseudogene 1 [Source:HGNC Symbol;Acc:753]
GAB2	-5,2	0,01	920	2562	GRB2-associated binding protein 2 [Source:HGNC Symbol;Acc:14458]
C16orf45	-5,2	0,22	87	652	Uncharacterized protein C16orf45 [Source:UniProtKB/Swiss-Prot;Acc:Q96MC5]
FBN1	-5,3	0,02	535	1857	fibrillin 1 [Source:HGNC Symbol;Acc:3603]
SLC37A1	-5,4	0,60	20	46	solute carrier family 37 (glycerol-3-phosphate transporter), member 1 [Source:HGNC Symbol;Acc:11024]
BIN2	-5,5	0,12	119	470	bridging integrator 2 [Source:HGNC Symbol;Acc:1053]
FMNL3	-5,6	0,02	330	1725	formin-like 3 [Source:HGNC Symbol;Acc:23696]
PLXNA4	-5,8	0,10	116	471	plexin A4 [Source:HGNC Symbol;Acc:9102]
GLIPR2	-5,9	0,05	172	563	GLI pathogenesis-related 2 [Source:HGNC Symbol;Acc:16007]
MTA3	-6,6	0,10	63	442	metastasis associated 1 family, member 3 [Source:HGNC Symbol;Acc:23784]
ITGB1	-7,3	0,00	3101	26317	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) [Source:HGNC Symbol;Acc:6153]
RAP1GAP	-7,4	0,00	1232	9393	RAP1 GTPase activating protein [Source:HGNC Symbol;Acc:9858]
CA3	-7,6	0,00	623	2944	carbonic anhydrase III, muscle specific [Source:HGNC Symbol;Acc:1374]
AC068580.3	-7,9	0,21	23	80	CD225 family protein FLJ76511 [Source:UniProtKB/Swiss-Prot;Acc:A6NMD0]
CELF6	-8,2	0,34	14	43	CUGBP, Elav-like family member 6 [Source:HGNC Symbol;Acc:14053]
SAMD12	-8,4	0,59	5	19	sterile alpha motif domain containing 12 [Source:HGNC Symbol;Acc:31750]
PLA2G3	-9,2	0,46	6	33	phospholipase A2, group III [Source:HGNC Symbol;Acc:17334]
NRSN2	-10,4	0,37	6	60	neurensin 2 [Source:HGNC Symbol;Acc:16223]
CYSLTR2	-11,0	0,35	6	31	cysteinyl leukotriene receptor 2 [Source:HGNC Symbol;Acc:18274]
CYFIP2	-11,3	0,18	11	84	cytoplasmic FMR1 interacting protein 2 [Source:HGNC Symbol;Acc:13760]
COL9A2	-11,6	0,21	3	85	collagen, type IX, alpha 2 [Source:HGNC Symbol;Acc:2218]
UNC5B	-14,4	0,35	3	28	unc-5 homolog B (C. elegans) [Source:HGNC Symbol;Acc:12568]
B4GALNT3	-22,6	0,46	1	4	beta-1,4-N-acetyl-galactosaminyl transferase 3 [Source:HGNC Symbol;Acc:24137]
LMOD1	-23,6	0,43	1	11	leiomodin 1 (smooth muscle) [Source:HGNC Symbol;Acc:6647]
LARP6	-40,1	0,20	1	21	La ribonucleoprotein domain family, member 6 [Source:HGNC Symbol;Acc:24012]

Enrichment Score: 2.3485744212350124

Term	Count	%	PValue	List T	Pop	Pop To	Fold Enric	Bonferr	Benjam	FDR
GO:0030097~hemopoiesis	8	6.2016	0.00105	92	236	13528	4.98452	0.613	0.2712	1.625
GO:0045321~leukocyte activation	8	6.2016	0.00121	92	242	13528	4.86094	0.666	0.24	1.878
h_ephA4Pathway:Eph Kinases and ephrins support pla	3	2.3256	0.00153	10	10	1437	43.11	0.077	0.0765	1.476
GO:0048534~hemopoietic or lymphoid organ developn	8	6.2016	0.00183	92	260	13528	4.52441	0.809	0.282	2.82
GO:0046649~lymphocyte activation	7	5.4264	0.00222	92	199	13528	5.17238	0.866	0.2845	3.409
GO:0002520~immune system development	8	6.2016	0.00257	92	276	13528	4.26213	0.902	0.2517	3.926
GO:0001775~cell activation	8	6.2016	0.00319	92	287	13528	4.09877	0.944	0.2306	4.857

Enrichment Score: 2.3218484921340874

Term	Count	%	PValue	List T	Pop	Pop To	Fold Enric	Bonferr	Benjam	FDR
GO:0007254~JNK cascade	5	3.876	6.57E-04	92	59	13528	12.4613	0.448	0.4478	1.02
GO:0031098~stress-activated protein kinase signaling	5	3.876	8.43E-04	92	63	13528	11.6701	0.533	0.3167	1.307
GO:0007243~protein kinase cascade	9	6.9767	0.00342	92	370	13528	3.57673	0.955	0.212	5.209
GO:0000165~MAPKKK cascade	5	3.876	0.03549	92	184	13528	3.99575	1	0.758	43.08
GO:0033554~cellular response to stress	9	6.9767	0.03651	92	566	13528	2.33815	1	0.7532	44.01

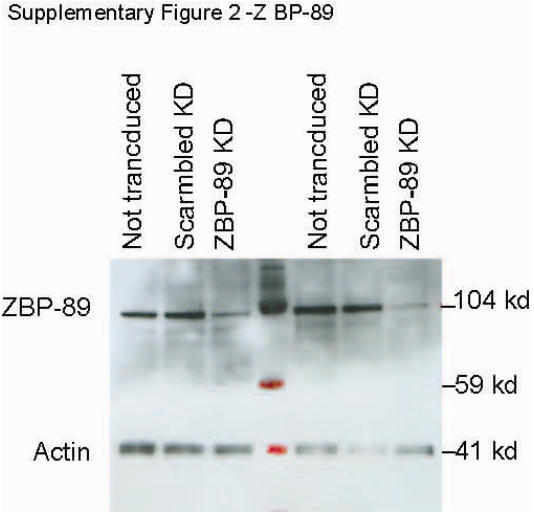
Enrichment Score: 1.9601275733470946

Term	Count	%	PValue	List T	Pop	Pop To	Fold Enric	Bonferr	Benjam	FDR
IPR001202:WW/Rsp5/WWP	4	3.1008	0.00543	113	53	16659	11.1264	0.827	0.5837	7.089
SM00456:WW	4	3.1008	0.01163	82	53	9079	8.35619	0.671	0.4264	12.07
domain:WW 2	3	2.3256	0.01512	125	29	19113	15.8177	1	0.9092	20.27
domain:WW 1	3	2.3256	0.01512	125	29	19113	15.8177	1	0.9092	20.27

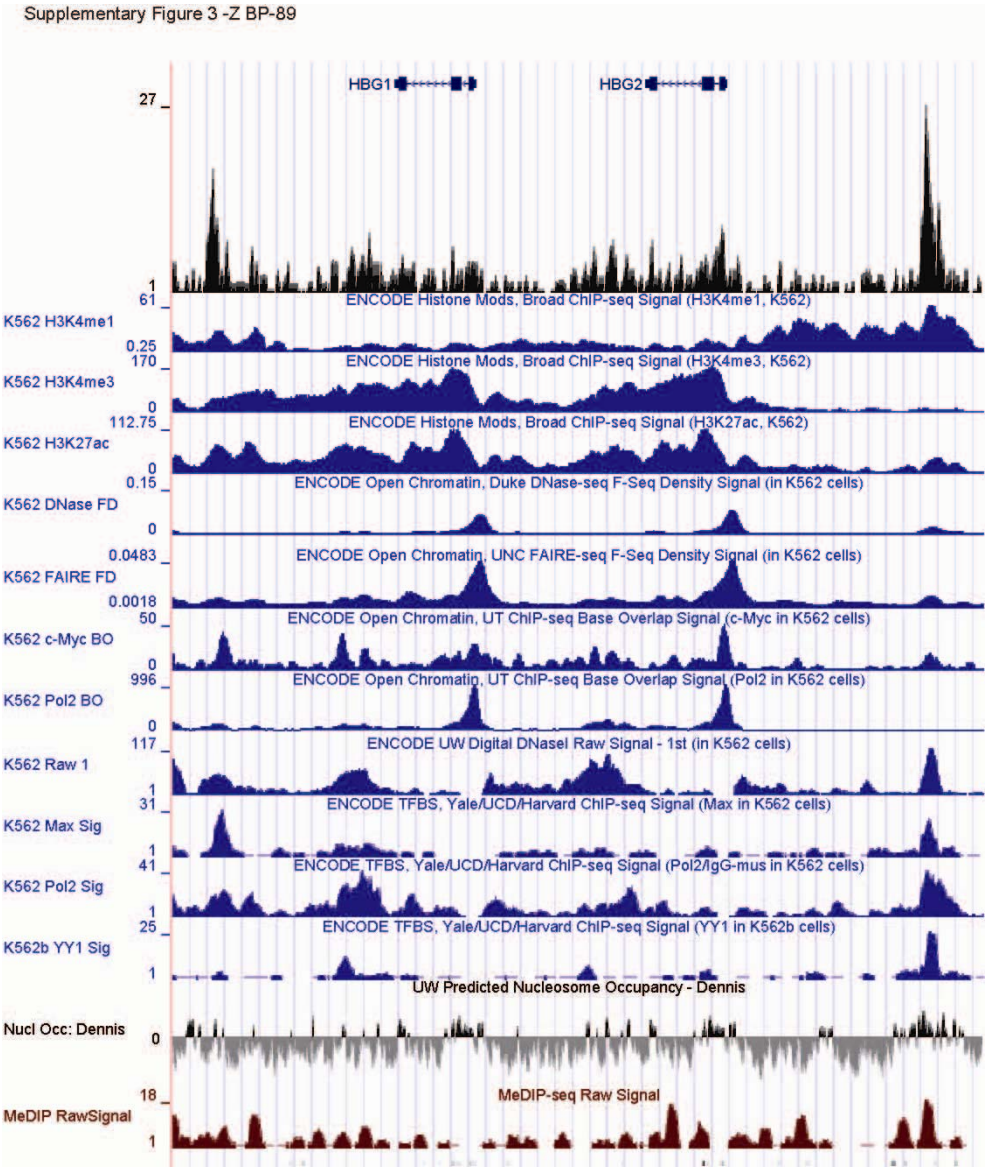
Enrichment Score: 1.7286581622760875

Term	Count	%	PValue	List T	Pop	Pop To	Fold Enric	Bonferr	Benjam	FDR
GO:0030097~hemopoiesis	8	6.2016	0.00105	92	236	13528	4.98452	0.613	0.2712	1.625
GO:0048534~hemopoietic or lymphoid organ developn	8	6.2016	0.00183	92	260	13528	4.52441	0.809	0.282	2.82
GO:0002520~immune system development	8	6.2016	0.00257	92	276	13528	4.26213	0.902	0.2517	3.926
GO:0030099~myeloid cell differentiation	4	3.1008	0.02474	92	93	13528	6.32445	1	0.6773	32.34
GO:0030218~erythrocyte differentiation	3	2.3256	0.03381	92	43	13528	10.2588	1	0.7563	41.51
GO:0034101~erythrocyte homeostasis	3	2.3256	0.04291	92	49	13528	9.00266	1	0.7448	49.54

Supplementary Table 3) Functional annotation terms of genes bound by ZBP-89 and differentially expressed upon its knock down



Supplementary Figure 2) Lentiviral-mediated shRNA knock down in HEP cells (related to figure 4 and 5).


Supplementary Figure 3 -Z BP-89
HBG1
HBG2
27
1
61
ENCODE Histone Mods, Broad ChIP-seq Signal (H3K4me1, K562)
K562 H3K4me1
0.25
170
ENCODE Histone Mods, Broad ChIP-seq Signal (H3K4me3, K562)
K562 H3K4me3
0
112.75
ENCODE Histone Mods, Broad ChIP-seq Signal (H3K27ac, K562)
K562 H3K27ac
0
0.15
ENCODE Open Chromatin, Duke DNase-seq F-Seq Density Signal (in K562 cells)
K562 DNase FD
0
0.0483
ENCODE Open Chromatin, UNC FAIRE-seq F-Seq Density Signal (in K562 cells)
K562 FAIRE FD
0.0018
50
ENCODE Open Chromatin, UT ChIP-seq Base Overlap Signal (c-Myc in K562 cells)
K562 c-Myc BO
0
996
ENCODE Open Chromatin, UT ChIP-seq Base Overlap Signal (Pol2 in K562 cells)
K562 Pol2 BO
0
117
ENCODE UW Digital DNaseI Raw Signal - 1st (in K562 cells)
K562 Raw 1
1
31
ENCODE TFBS, Yale/UCD/Harvard ChIP-seq Signal (Max in K562 cells)
K562 Max Sig
1
41
ENCODE TFBS, Yale/UCD/Harvard ChIP-seq Signal (Pol2/IgG-mus in K562 cells)
K562 Pol2 Sig
1
25
ENCODE TFBS, Yale/UCD/Harvard ChIP-seq Signal (YY1 in K562b cells)
K562b YY1 Sig
1
UW Predicted Nucleosome Occupancy - Dennis
Nucl Occ: Dennis
0
18
MeDIP-seq Raw Signal
MeDIP RawSignal
1

REFERENCES

Akinsheye, I., Alsultan, A., Solovieff, N., Ngo, D., Baldwin, C.T., Sebastiani, P., Chui, D.H., and Steinberg, M.H. (2011). Fetal hemoglobin in sickle cell anemia. Blood *118*, 19-27.

Bai, L., and Merchant, J.L. (2000). Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells. J Biol Chem *275*, 30725-30733.

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

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837.

Borg, J., Papadopoulos, P., Georgitsi, M., Gutierrez, 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.

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.

Campbell, A.D., Cui, S., Shi, L., Urbonya, R., Mathias, A., Bradley, K., Bonsu, K.O., Douglas, R.R., Halford, B., Schmidt, L.*, et al.* (2011). Forced TR2/TR4 expression in sickle cell disease mice confers enhanced fetal hemoglobin synthesis and alleviated disease phenotypes. Proc Natl Acad Sci U S A *108*, 18808-18813.

Chien, R., Zeng, W., Kawauchi, S., Bender, M.A., Santos, R., Gregson, H.C., Schmiesing, J.A., Newkirk, D.A., Kong, X., Ball, A.R., Jr.*, et al.* (2011). Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. J Biol Chem *286*, 17870-17878.

Chupreta, S., Brevig, H., Bai, L., Merchant, J.L., and Iniguez-Lluhi, J.A. (2007). Sumoylation-dependent control of homotypic and heterotypic synergy by the Kruppel-type zinc finger protein ZBP-89. J Biol Chem *282*, 36155-36166.

Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. Trends Biochem Sci *30*, 250-255.

Constantoulakis, P., Knitter, G., and Stamatoyannopoulos, G. (1989). On the induction of fetal hemoglobin by butyrates: in vivo and in vitro studies with sodium butyrate and comparison of combination treatments with 5-AzaC and AraC. Blood *74*, 1963-1971.

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. Nucl Acids Res *100*, 7480-7485.

de Nigris, F., Botti, C., Rossiello, R., Crimi, E., Sica, V., and Napoli, C. (2007). Cooperation between Myc and YY1 provides novel silencing transcriptional targets of alpha3beta1-integrin in tumour cells. Oncogene *26*, 382-394.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233-1244.

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

Friend, C., Scher, W., Holland, J.G., and Sato, T. (1971). Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. Proc Natl Acad Sci U S A *68*, 378-382.

Green, A.R., Rockman, S., DeLuca, E., and Begley, C.G. (1993). Induced myeloid differentiation of K562 cells with downregulation of erythroid and megakaryocytic transcription factors: a novel experimental model for hemopoietic lineage restriction. Exp Hematol *21*, 525-531.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311–318.

Higgs, D.R., Engel, J.D., and Stamatoyannopoulos, G. (2012). Thalassaemia. Lancet *379*, 373-383.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kiefer, C.M., Lee, J., Hou, C., Dale, R.K., Lee, Y.T., Meier, E.R., Miller, J.L., and Dean, A. (2011). Distinct Ldb1/NLI complexes orchestrate gamma-globin repression and reactivation through ETO2 in human adult erythroid cells. Blood *118*, 6200-6208.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci *30*, 235-239.

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

Li, X., Xiong, J.-W., Shelley, C.S., Park, H., and Arnaout, M.A. (2006). The transcription factor ZBP-89 controls generation of the hematopoietic lineage in zebrafish and mouse embryonic stem cells. Development *133*, 3641-3650.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-[Delta][Delta]CT Method. Methods *25*, 402-408.

Lozzio, C.B., and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood *45*, 321-334.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.

Marks, P.A., and Rifkind, R.A. (1978). Erythroleukemic differentiation. Annu Rev Biochem *47*, 419-448.

Merchant, J.L., Iyer, G.R., Taylor, B.R., Kitchen, J.R., Mortensen, E.R., Wang, Z., Flintoft, R.J., Michel, J.B., and Bassel-Duby, R. (1996). ZBP-89, a Kruppel-like zinc finger protein, inhibits epidermal growth factor induction of the gastrin promoter. Mol Cell Biol *16*, 6644-6653.

Ohneda, K., Ohmori, S.y., Ishijima, Y., Nakano, M., and Yamamoto, M. (2009). Characterization of a Functional ZBP-89 Binding Site That Mediates Gata1 Gene Expression during Hematopoietic Development. J Biol Chem *284*, 30187-30199.

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

Palstra, R.J., de Laat, W., and Grosveld, F. (2008). Beta-globin regulation and long-range interactions. Adv Genet *61*, 107-142.

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

Panne, D. (2008). The enhanceosome. Curr Opin Struct Biol *18*, 236-242.

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.

Pelengaris, S., Khan, M., and Evan, G. (2002). c-MYC: more than just a matter of life and death. Nat Rev Cancer *2*, 764-776.

Perrine, S.P., Miller, B.A., Faller, D.V., Cohen, R.A., Vichinsky, E.P., Hurst, D., Lubin, B.H., and Papayannopoulou, T. (1989). Sodium butyrate enhances fetal globin gene expression in erythroid progenitors of patients with Hb SS and beta thalassemia. Blood *74*, 454-459.

Perrine, S.P., Miller, B.A., Greene, M.F., Cohen, R.A., Cook, N., Shackleton, C., and Faller, D.V. (1987). Butryic acid analogues augment gamma globin gene expression in neonatal erythroid progenitors. Biochemical and Biophysical Research Communications *148*, 694-700.

Petrovic, I., Kovacevic-Grujicic, N., and Stevanovic, M. (2009). ZBP-89 and Sp3 down-regulate while NF-Y up-regulates SOX18 promoter activity in HeLa cells. Mol Biol Rep *36*, 993-1000.

Roeder, R.G. (1998). Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harb Symp Quant Biol *63*, 201-218.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K.A., 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.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., *et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet *42*, 53-61.

Soleimani VD, Yin H, Jahani-Asl A, Ming H, Kockx CE, van Ijcken WF, Grosveld F, Rudnicki MA (2012) Snail Regulates MyoD Binding-Site Occupancy to Direct Enhancer Switching and Differentiation-Specific Transcription in Myogenesis. *Mol Cell*.; **47(3)**: 457-68.

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.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell *28*, 810-822.

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.

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

Stamatoyannopoulos, G., and Grosveld, F. (2001). Hemoglobin switching. In The molecular basis of blood diseases (3rd ed), G. Stamatoyannopoulos, Majeurs P., Perlmutter, P., Varmus H., ed. (Philadelphia, W.B. Saunders), pp. 135-182.

Steinberg, M.H. (2002). Hydroxyurea treatment for sickle cell disease. ScientificWorldJournal *2*, 1706-1728.

Steinberg, M.H., Lu, Z.-H., Barton, F.B., Terrin, M.L., Charache, S., Dover, G.J., and the Multicenter Study of, H. (1997). Fetal Hemoglobin in Sickle Cell Anemia: Determinants of Response to Hydroxyurea. Blood *89*, 1078-1088.

Taatjes, D.J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem Sci *35*, 315-322.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and Interaction between Hypersensitive Sites in the Active [beta]-globin Locus. Molecular Cell *10*, 1453-1465.

Vernimmen, D., Gobbi, M.D., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J *26*, 2041-2051.

von Lindern, M., Deiner, E.M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M.J., Mullner, E.W., 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.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40: 897–903.

Weatherall, D.J., Ledingham J.G.G., Warrell DA. (1996). In Oxford textbook of medicine 3rd ed, L. Weatherall DJ, and W.D. JGG, eds. (Oxford, Oxford University Press).

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., 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.

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

Wu, Y., Zhang, X., Salmon, M., and Zehner, Z., E. (2007). The zinc finger repressor, ZBP-89, recruits histone deacetylase 1 to repress vimentin gene expression. Genes Cells *12*, 905-918.

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., *et al.* (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993-996.

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.

Chapter 4

4

Five Friends of Methylated Chtop, a complex linking arginine methylation to Desumoylation

Pavlos Fanis[1], Nynke Gillemans[1], Ali Aghajanirefah[1], Farzin Pourfarzad[1], Jeroen Demmers[2], Fatemehsadat Esteghamat[1], Ratna K. Vadlamudi[3], Frank Grosveld[1], Sjaak Philipsen[1], and Thamar B. van Dijk[1]

1. Department of Cell Biology, Erasmus MC, 3000 CA, Rotterdam, The Netherlands
2. Proteomics Center, Erasmus MC, 3000 CA, Rotterdam, The Netherlands
3. Department of Obstetrics and Gynecology, UTHSCSA, 7703 Floyd Curl Drive, San Antonio, TX 78229, Texas, United States of America

(Mol Cell Proteomics. 2012 Aug 7)

SUMMARY

Chromatin target of Prmt1 (Chtop) is a vertebrate-specific chromatin-bound protein that plays an important role in transcriptional regulation. As its mechanism of action remains unclear, we identified Chtop-interacting proteins using a biotinylation-proteomics approach. Here we describe the identification and initial characterization of Five Friends of Methylated Chtop (5FMC). 5FMC is a nuclear complex that can only be recruited by Chtop when the latter is arginine-methylated by Prmt1. It consists of the co-activator Pelp1, the Sumo-specific protease Senp3, Wdr18, Tex10, and Las1L. Pelp1 functions as the core of 5FMC, as the other components become unstable in the absence of Pelp1. We show that recruitment of 5FMC to Zbp-89, a zinc-finger transcription factor, affects its sumoylation status and transactivation potential. Collectively, our data provide a mechanistic link between arginine methylation and (de)sumoylation in the control of transcriptional activity.

INTRODUCTION

Transcription factor activity is often controlled by post-translational modifications such as acetylation, phosphorylation, methylation and sumoylation. Some modifications are associated with both gene activation and repression, while others appear to be more exclusive: asymmetrical dimethylation of arginine residues is restricted to transcriptional activation, while modification by sumoylation correlates with inhibition of transcription (1).

Arginine methylation occurs frequently within glycine-arginine-rich (GAR) regions and is catalyzed by members of the protein arginine methyltransferase (Prmt) family. These enzymes are subdivided in two major classes: type I enzymes catalyze the formation of asymmetrically dimethylated arginines (aDMA), while type II enzymes form symmetrically dimethylated arginines (sDMA) (2). Prmt1 and Prmt4/Carm1 (Coactivator-associated arginine methyltransferase 1) are the major type I enzymes and both are critical for mammalian development (3-4). Their substrates include RNA-binding proteins, nuclear matrix proteins, cytokines and transcriptional regulators (2). Prmt1 methylates transcription factors such as Runx1 and STAT1 thereby promoting their transcriptional activity (5-6). Furthermore, Prmt1 and Prmt4 are recruited by nuclear hormone receptors and other transcription factors including YY1, p53 and NF-κB (7-10), resulting in the methylation of additional coactivators and histones. Prmt4 methylates histone H3 at arginine 17 and 26, while Prmt1 targets histone H4 at arginine 3 for methylation promoting subsequent acetylation of histone H3 at lysine 9 and histone H3 at lysine 14 (11) and further activating events (12).

Small ubiquitin-like modifier (SUMO) has an important regulatory function in several cellular processes, including DNA repair, cell cycle progression, signal transduction, chromatin structure and transcriptional regulation (13). Mammalian cells express four SUMO paralogs (SUMO-1 to SUMO-4). SUMO-1 differs in sequence by about 50% from SUMO-2 and 3, whereas SUMO-2 and SUMO-3 are 97% identical to each other. Conjugation of SUMO to target proteins occurs by a series of reactions conducted by the E1 activating enzyme, E2 conjugating enzyme and an E3 SUMO ligase (14). The reverse desumoylation process is mediated by the isopeptidase activity of SUMO-specific proteases (Senps). In mammals, six members of Senps have been reported, known as Senp1-3 and Senp5-7. Sumoylation of multiple transcription factors, including Sp3, Sox6, Zeb1, and Zbp-89, has a negative effect on their transactivation potential, as it promotes the recruitment of repressive complexes (15-

17). Many components of the repressor complexes CoREST1, NuRD, PRC1, Setdb1, and MEC themselves are also sumoylated, or have SUMO interacting motifs (SIMs). This suggests that sumoylation plays an important role in the formation and/or stabilization of these complexes (18).

We previously identified Friend of Prmt1 (Fop), also known as Small protein rich in arginine and glycine (SRAG), encoded by the mouse 2500003M10Rik and human C1orf77 genes, respectively (19-20). Recently, the name Chromatin Target of Prmt1 (Chtop) has been assigned to this gene/protein by the HUGO Gene Nomenclature Committee (HGNC). For reasons of clarity, we will also use the name Chtop for the murine homolog. Chtop is a chromatin-associated protein that plays an important role in the ligand-dependent activation of estrogen target genes such as TFF1 (pS2) in breast cancer cells (19). In addition, it is a critical regulator of γ-globin gene expression (21). However, little is known about the molecular mechanism of transcriptional control mediated by Chtop.

Chtop contains a central glycine-arginine-rich (GAR) region that is recognized and methylated by Prmt1. Since arginine methylation controls protein-protein interactions, we used a biotinylation-proteomics approach to identify proteins that bind Chtop in the presence and absence of Prmt1. In this study we identified and characterized a protein complex that binds specifically to methylated Chtop. As this nuclear complex consists of five proteins – SUMO1/sentrin/SMT3 specific peptidase 3 (Senp3), proline-glutamate and leucine rich protein 1 (Pelp1), LAS1-like protein (Las1L), Testis expressed 10 protein (Tex10), and WD repeat domain 18 protein (Wdr18) – we call it Five Friends of Methylated Chtop (5FMC). We show that Pelp1 is critical for the integrity of 5FMC and that Chtop and 5FMC are recruited by Zinc finger binding protein-89 (Zbp-89), thereby regulating both (de)sumoylation of, and transactivation by, Zbp-89.

EXPERIMENTAL PROCEDURES

Plasmid Construction and Cells

The following plasmids have been described previously: GST-PELP1 deletions (22), T7-Pelp1 (23), pMT2_HA_Chtop, HA_mPRMT1 WT, and HA_mPRMT1 EQ (19). The cDNA of human SUMO-2 (hSMT3b) was kindly provided by Dr. Guntram Suske (Philipps-Universität Marburg, Germany). The cDNA of full-length LAS1L was obtained from Open Biosystems (Clone ID 3140243; Huntsville, AL, USA), full-length cDNA of Wdr18 was obtained from RZPD/imaGenes (clone IRAVp968G04150D6; Berlin, Germany) and full-length cDNA of Senp3 was obtained from RZPD/imaGenes (clone IRAVp968B0184D6; Berlin, Germany). The cDNA of PELP1 was subcloned from T7-PELP1. After the introduction of the 23-amino acid (aa) biotinylation tag into the pMT2_HA (24) and pMT2_HA_Chtop, LAS1L, Senp3 and PELP1 were cloned to pMT2_bio_HA using SalI and NotI, Wdr18 was cloned using SalI and SUMO-2 was cloned using SalI and EcoRI. Bio_HA_LAS1L, Bio_HA_Senp3, Bio_HA_PELP1, Bio_HA_Wdr18, Bio_HA_SUMO-2 and Bio_HA_Chtop were subcloned into the erythroid expression vector pEV-neo (25) and electroporated into mouse erythroleukemia (MEL) cells expressing the BirA biotin ligase (26). To make the Gateway pSG513_myc destination plasmid, the Attr1-Cm^R-ccdb-Attr2 fragment was subcloned from pDEST17 (Invitrogen) to a modified pSG5 (Stratagene) using HindIII, downstream the myc-tag sequence that was introduced to the pSG513 plasmid using EcoRI and BamHI. cDNAs for Senp3, Wdr18 and LAS1L were cloned into pDONR221 (Invitrogen), from which they were cloned by Gateway LR reaction to pSG513_myc. Internal deletion mutants of LAS1L and Wdr18 were generated using the QuikChange site-directed mutagenesis kit (Stratagene). MEL and 293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) supplemented with 10% fetal calf serum (FCS).

Transient transfection, Immunoprecipitation, GST pull-down assay and Western blot analysis

Transient transfections in 293T cells, immunoprecipitations and western blot analysis were performed as described previously (27). For immunoprecipitations combined with Benzonase (Novagen) incubation, 250 units of Benzonase were used followed by 3 hours incubation at 4^oC. Nitrocellulose membranes were blocked in 1% bovine serum albumin (BSA), probed with the appropriate primary antibodies and analyzed using the Odyssey Infrared Imaging System (Li-Cor Biosciences). GST pull-down assays were performed as described previously (22). Western blots were probed with the following primary antibodies: Prmt1 (07-404), Asym24 (07-414), and Cbx4 (09-

029) were from Upstate; Actin (clone I-19; sc-1616), Taf1β (clone H-120; sc-25564), Lamin B (sc-6216), HA (monoclonal F7; sc-7392), HA (polyclonal Y11; sc-805), Myc (monoclonal 9E10; sc-40), Myc (polyclonal A-14; sc-789) and Pol II (polyclonal N-20; sc-899) were from Santa Cruz Biotechnology; Tex10 (17372-1-AP), Las1L (16010-1-AP), Senp3 (17659-1-AP) and Wdr18 (15165-1-AP) were from Proteintech Group; Zbp-89 (ab69933) was from Abcam, SUMO 2/3 (clone 1E7; M114-3) was from MBL; Pelp1 (A300-876A) was from Bethyl Laboratories; T7 (69522-3) was from Novagen; Taf1α (B100-56353) was from Novus Biologicals and Chtop (KT64) was from Absea Biotechnology. Ring1B antibody was kindly provided from Dr. Miguel Vidal (Madrid, Spain).

Cell lysates and Mass Spectrometry (MS)

Preparation of nuclear and whole cell extracts from MEL and 293T cells were carried out as described previously (27). Purification of biotinylated proteins, digestion with trypsin (Promega, sequencing grade) on paramagnetic streptavidin beads and LC-MS/MS were performed as described previously (19, 26). Nanoflow LC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap mass spectrometer (Thermo), 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. 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 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 IPI_mouse database (version 3.83, containing 60,010 sequences and 27,475,843 residues). The peptide tolerance was set to 10 ppm and the fragment ion tolerance was set to 0.8 Da. A maximum number of 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. Search results were parsed into a home-built database system for further analysis. Entries were parsed if they had a minimum peptide Mascot score of 25, and a significance threshold of $p < 0.05$; the option "require red bold" was also selected. Using these parameters yields an estimated peptide false

discovery rate (FDR) of 3-5% against a target decoy database. The Mascot data have been uploaded to the PRIDE database (28) (www.ebi.ac.uk/pride) under accession numbers 21750-21767. The data was converted using PRIDE Converter (29) (http://pride-converter.googlecode.com).

Lentivirus Mediated Knockdown

The Chtop and Prmt1 shRNA lentiviral vectors were described previously (19, 21). For mouse Pelp1, Senp3, and human PELP1, SENP3, WDR18, LAS1L, clones from the TRC Mission shRNA library ((30); Sigma Aldrich, St. Louis, MO, USA) were used for knockdown experiments in MEL and 293T cells respectively, including a non-targeting shRNA control virus (SHC002). Lentivirus was produced by transient transfection of 293T cells as described before (31). The following clones were used from the TRC shRNA library: TRCN0000177043 (shPelp1 #1), TRCN0000178252 (shPelp1 #2), TRCN0000031014 (shSenp3 #1), TRCN0000031017 (shSenp3 #2), TRCN0000159617 (shPELP1 #2), TRCN0000159673 (shPELP1 #3), TRCN0000004106 (shSENP3 #1), TRCN0000004107 (shSENP3 #2), TRCN0000078088 (shWDR18 #1), TRCN0000078089 (shWDR18 #2), TRCN0000121835 (shLAS1L #2), TRCN0000142144 (shLAS1L #4), TRCN0000035931 (sh hPRMT1).

Size-Exclusion Chromatography and Subcellular Fractionation

Nuclear extracts from MEL cells expressing BirA biotin ligase enzyme, were chromatographed over a Superose 6 column (Amersham Biosciences) using an AKTA fast-performance liquid chromatography apparatus. Fractions were collected and precipitated with trichloroacetic acid and analyzed by western blotting. Subcellular fractionation was performed as described previously (19).

RT, QPCR, ChIP assay and statistical analysis

Reverse transcription (RT), RT-quantitative PCR (RT-QPCR) and ChIP were performed as described previously (32). Primers used for RT-QPCR and ChIP-QPCR are summarized in Supplemental Table 1. ANOVA statistical analysis was performed by GraphPad Prism 5.02.

RESULTS

Methylation dependent and independent interactions of Chtop

In order to identify interaction partners of Chtop, an N-terminal double-tagged version of Chtop protein (Bio_HA_Chtop) was expressed in mouse erythroleukemia (MEL) cells. These cells also stably expressed BirA, a bacterial biotin ligase which efficiently biotinylates the Bio-tag (26). Note that tagged Chtop was not overexpressed, as endogenous levels were reduced in Bio_HA-Chtop transfected cells (Fig. 1C). Protein complexes from nuclear lysates were recovered by streptavidin pull down followed by nanoflow liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) and were compared to samples from cells expressing BirA alone (Fig. 1A, Supplemental Table 2). This confirmed the association of Chtop with Prmt1 and Prmt5, factors that we have previously identified as Chtop binding proteins (16). MS analysis also revealed the binding of three members of the polycomb repressor complex 1 (PRC1; Pc2, Ring1B, and Phc2) (33-35), the SET nuclear oncoprotein (36), and the mRNA export protein 40-2-3 (37). Furthermore, high MASCOT scores were obtained for the proteins Pelp1, Las1L, Tex10, Senp3 and Wdr18, four of which were previously co-purified with the MLL complex and were recently described as regulators of ribosome biogenesis (38-40). Pelp1 is a coactivator involved in nuclear hormone signaling (41), while Senp3 is a SUMO-specific protease (42). The WD-repeat protein Wdr18, Tex10, and Las1L had not been characterized previously. To investigate whether arginine methylation of Chtop played a role in these interactions, we performed lentiviral-mediated knockdown of Prmt1 in Bio_HA_Chtop expressing MEL_BirA cells and used a similar purification approach. When compared to cells transduced with a control lentivirus, no major differences were observed in the binding of Chtop to Prmt5, Pc2, Ring1B, Phc2, 40-2-3 and Set (Fig. 1B, Supplemental Table 2). Interestingly, copurification of Las1L, Pelp1, Tex10, Senp3 and Wdr18 was strongly reduced or absent when Chtop was hypomethylated. Next, we confirmed the methylation (in)dependent interactions by streptavidin affinity purification, followed by Western blot analysis. Chtop is hypomethylated in the absence of Prmt1, as indicated by its faster migration pattern and by staining with an antibody that specifically recognizes aDMA (Asym24; Fig. 1C). Endogenous Chtop interactors were efficiently recovered in Bio_HA_Chtop pull downs, while no background staining was observed in MEL_BirA cells (Fig. 1D-E). In addition, no association was observed with Wdr18 in the absence of Prmt1, while

the interactions with Pelp1, Las1L, Tex10 and Senp3 were strongly reduced. These associations do not depend on the presence if Prmt1, as only wildtype but not enzymatic inactive Prmt1 could rescue the knockdown (Supplemental Fig. 1). Together, these results validate the interactions identified by MS and show that methylation of Chtop is required for the recruitment of Pelp1, Las1L, Tex10, Senp3 and Wdr18.

A

Strept. PD of Bio_HA_Chtop		
Protein	uniq. pept.	coverage (%)
Prmt1	13	39.9
Chtop	5	28.8
Set	3	20.5
Prmt5	3	8.6
40-2-3	1	5.4
Pc2	1	3.1
Ring1B	1	3.3
Phc2	1	2.5
Las1L	7	12.5
Pelp1	7	10
Senp3	6	14.8
Wdr18	6	20.9
Tex10	5	10.2

B

Strept. PD of Bio_HA_Chtop shControl			Strept. PD of Bio_HA_Chtop shPrmt1		
Protein	uniq. pept.	coverage (%)	Protein	uniq. pept.	coverage (%)
Prmt1	12	34	Prmt1	4	13.6
Chtop	3	24.3	Chtop	4	30.7
Prmt5	7	17.4	Prmt5	6	12.5
Set	4	18.3	Set	2	11.3
40-2-3	2	8.5	40-2-3	3	12.6
Pc2	2	5.6	Pc2	1	3.1
Ring1B	1	5.1	Ring1B	2	9.3
Phc2	1	3.4	Phc2	3	12.4
Las1L	11	18.4	not detected	x	x
Pelp1	10	14.2	not detected	x	x
Senp3	6	16.5	Senp3	1	2.5
Wdr18	7	23.2	Wdr18	1	4.2
Tex10	10	15.7	Tex10	1	2.3

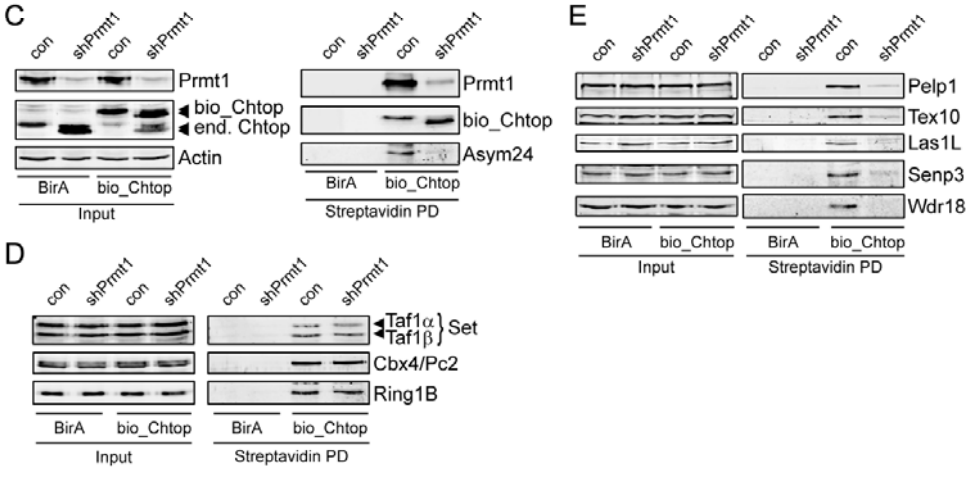


Figure 1. Identification of Chtop-interacting proteins

(A) List of Chtop interacting proteins indentified by mass spectrometry (MS). Protein complexes from nuclear lysates of MEL_BirA cells ectopically expressing Bio_HA_Chtop were recovered by MS. Proteins, unique peptides (uniq. pept.) and percent coverage (coverage %) are indicated.(B) List of Chtop interacting proteins in MEL_BirA cells treated with control lentivirus (shControl) and lentivirus expressing shRNA against Prmt1 (shPrmt1) identified by MS. Interactions with major differences between shControl and shPrmt1 treated cells are shown in

a white box.(C) Chtop is hypomethylated in the absence of Prmt1. Whole cell lysates (Input) and streptavidin pull downs (Streptavidin PD) were analyzed for Prmt1, Chtop and asymmetric dimethyl arginine (Asym24) residues. Actin staining serves as a loading control. Ectopically expressed (bio_Chtop) and endogenous Chtop (end. Chtop) are indicated. (D) Chtop methylation-independent interactions. Whole cell lysates (Input) and streptavidin pull-downs (Streptavidin PD) were analyzed with the antibodies indicated.(E) Chtop methylation-dependent interactions. Whole cell lysates (Input) and streptavidin pull-downs (Streptavidin PD) were analyzed with the antibodies indicated.

5FMC is a nuclear complex

We have previously shown that the majority of Chtop is bound to chromatin (19). To elucidate where the newly identified methylation-specific partners of Chtop localize in the cell, we performed biochemical fractionation of MEL cells. This revealed that all five proteins were mainly found in the nucleoplasm, while low levels were also detected in the cytoplasmic and chromatin fractions (Fig. 2A). To examine whether the five proteins form a complex, we first performed size-exclusion chromatography of MEL nuclear extracts. The elution patterns of Pelp1, Las1L, Tex10, Senp3 and Wdr18, as well as Chtop, overlapped substantially (Fig. 2B). The molecular mass of the positive fractions was >1MDa, indicating that the factors were presents in a high molecular weight protein complex or were bound to chromatin. Similar experiments in human 293T cells revealed comparable results, although larger proportions of LAS1L, TEX10, and WDR18 were detected in fractions corresponding to lower molecular mass (Supplemental Fig. 2). Next, doubled tagged (Bio_HA) Pelp1, Las1L, Senp3 and Wdr18 were stably expressed in MEL_BirA cells. Of note, we were not able to exogenously express the Tex10 protein, probably due to protein stability issues. Associated proteins were identified by streptavidin pull down in nuclear lysates followed by nanoLC-MS/MS and were compared to samples from MEL_BirA only cells. In all four experiments, the associated proteins with the highest MASCOT score were Pelp1, Las1L, Tex10, Senp3 and Wdr18 (Fig. 2C (Bio_HA_Senp3) and Supplemental Fig. 3 (Pelp1, Wdr18, and Las1L)). Moreover, in the MS analysis of Bio_HA_Senp3, Bio_HA_Wdr18 and Bio_HA_Pelp1 we observed an association with the Nol9 protein. The Nol9 ortholog in *S. pombe* (Grc3) was shown to associate with Las1 and the yeast IPI complex that consist of Rix1, Ipi1 and Crb3 (43). These proteins share homologous regions with Pelp1, Tex10 and Wdr18, respectively. Moreover, it was recently shown in human cells that NOL9 interacts with PELP1, TEX10, WDR18, LAS1L SENP3 (40). Chtop was only detected in the Bio_HA_Senp3 purification (Fig. 2C, left panel). The MS results were confirmed by immunoblot analysis of the streptavidin pull downs of

tagged proteins (Fig. 2C, right panel and Supplemental Fig. 3), and immunoprecipitations of endogenous proteins from both mouse and human cells (Fig. 2D and Supplemental Fig. 4, respectively). Taken together, these results show that Pelp1, Las1L, Tex10, Senp3 and Wdr18 form a nuclear multi-protein complex. As this complex binds selectively to methylated Chtop, we named it Five Friend of Methylated Chtop, or 5FMC.

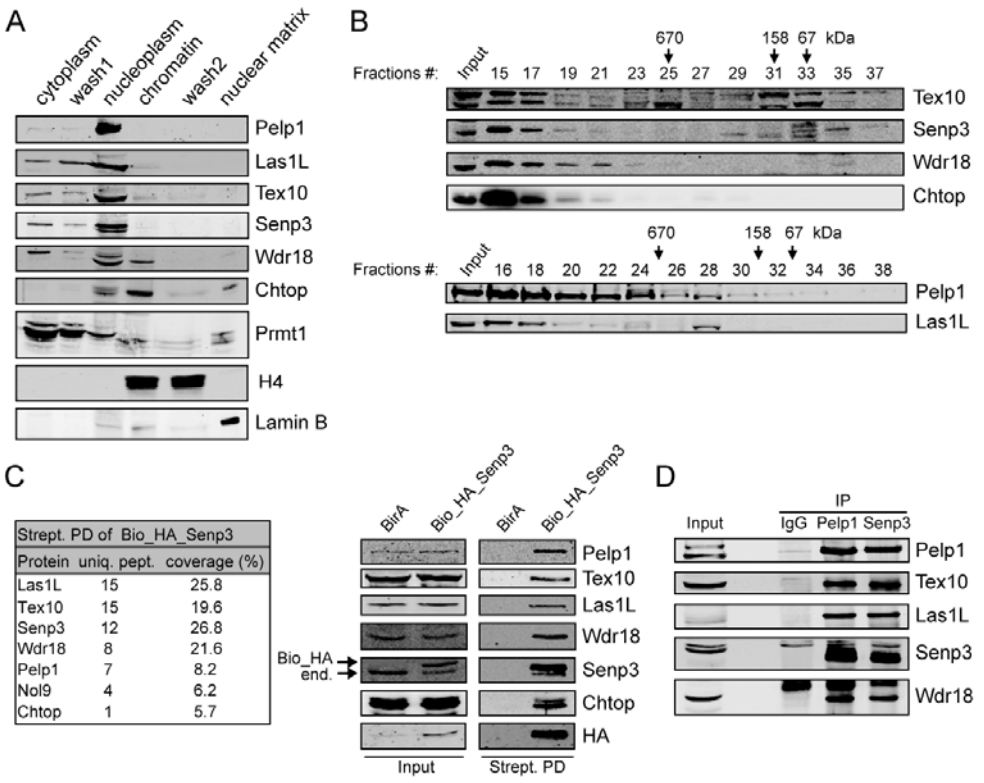


Figure 2. 5FMC is a nuclear complex

(A) Chtop methylation dependent interaction proteins are localized mainly in the nucleoplasm. MEL_BirA cells were biochemical fractionated as described in Experimental Procedures. Cytoplasmic, nucleoplasmic, chromatin and nuclear matrix were tested using Pelp1, Las1L, Tex10, Senp3, Wdr18 antibodies against endogenous proteins. Chtop, Prmt1, H4, and Lamin B served as controls for individual fractions. (B) MEL_BirA cell nuclear extracts were analyzed by sized-exclusion chromatography on a Superose 6 column. Proteins eluted from the indicated fractions were blotted with the indicated antibodies. Molecular mass markers are indicated at the

top.(C) Senp3 interactions in MEL cells. Whole cell lysates (Input) and streptavidin pull downs (Strept. PD) from MEL_BirA (BirA) and MEL_BirA cells expressing biotinylated Senp3 (Bio_HA_Senp3) were analyzed by MS (table) and western blotting. Immunoblot probed with the indicated antibodies. Arrows indicate endogenous (end.) and biotinylated (Bio_HA) Senp3.(D) 5FMC is a nuclear complex. Endogenous association between the 5FMC components. MEL_BirA cell nuclear lysates were analyzed by immunoprecipitation (IP) and western blotting with the antibodies indicated.

Pelp1 is the core subunit and critical for 5FMC stability

To further study the composition of the 5FMC complex, we transiently co-transfected tagged 5FMC components in 293T cells, followed by co-immunoprecipitation (co-IP). We found that T7_Pelp1 is efficiently recovered in HA_Wdr18 IPs, while Myc_Senp3 interacts with HA_Wdr18 (Fig. 3A). Moreover, Myc_Senp3 co-purifies with HA_Wdr18 only when T7_Pelp1 is co-transfected (Fig. 3A, lanes 3 and 4). In addition, we observed that cotransfection of T7_Pelp1 resulted in higher protein levels of HA_Wdr18 and Myc_Senp3 (Fig. 3A, lanes 2 and 4). Identical results were obtained when Myc_Senp3 was immunoprecipitated (data not shown). No conclusive results could be obtained with ectopically expressed Tex10 and Las1L, probably due to stability issues. These results show that Pelp1 is required for the interaction between Wdr18 and Senp3 and that Pelp1 might be required for the stability of these proteins. To further study the potential central role of Pelp1 within the 5FMC complex, we depleted endogenous mouse Pelp1 in MEL_BirA cells and endogenous human PELP1 in 293T cells by lentiviral-mediated knockdown using two different shRNAs (Fig. 3B and Supplemental Fig. 5A). Interestingly, the protein levels of Senp3 and Las1L were dramatically decreased, while Wdr18 and Tex10 could not be detected in the absence of endogenous Pelp1 (Fig. 3B, left panel). Quantitative RT-PCR showed that the reduced protein levels were due to protein stability rather than reduced mRNA levels (Fig. 3B, right panel). The individual knockdown of LAS1L, SENP3 or WDR18 had no significant effect on the protein levels of the other 5FMC subunits. Furthermore, partial complexes exist upon depletion of LAS1L, SENP3 and WDR18 (Supplemental Fig. 5). These results indicated that Pelp1 is the central component of the 5FMC complex and that the integrity of 5FMC is essential for the stability of its components. Analysis of the primary sequence of Pelp1 showed that Pelp1 contains a cysteine-rich region, two proline-rich regions and a C-terminal glutamine-rich region (41). To map the interactions between Pelp1 and other 5FMC components, we tested the ability of HA-tagged Las1L, Wdr18 and Senp3 proteins to bind to various domains of Pelp1 fused to GST. HA_Las1L and HA_Wdr18 interacted with the GST_Pelp1 fusion containing amino acids (aa) 401-600, while HA_Senp3 mainly interacted with domain 960-1130.

Interactions were further studied using a series of deletion mutants of Las1L. These experiments showed that the C-terminal part of Las1L (aa 552-734) mediated the interaction with Pelp1, while the central domain (aa 370-552) mediated the interaction with Wdr18. The binding to Senp3 could not be mapped in detail: any deletion between aa 188-734 disrupted the interaction (Supplemental Fig. 6A). Similar experiments with Wdr18 deletion constructs revealed that the region containing WD40 domains 4-6 were required for binding to Las1L, while deletion of any WD40 domain disrupted the binding to Pelp1 (Supplemental Fig. 6B). These initial domain-mapping experiments suggest complex multi-intermolecular interactions and are in line with the proposed model shown in Figure 3D.

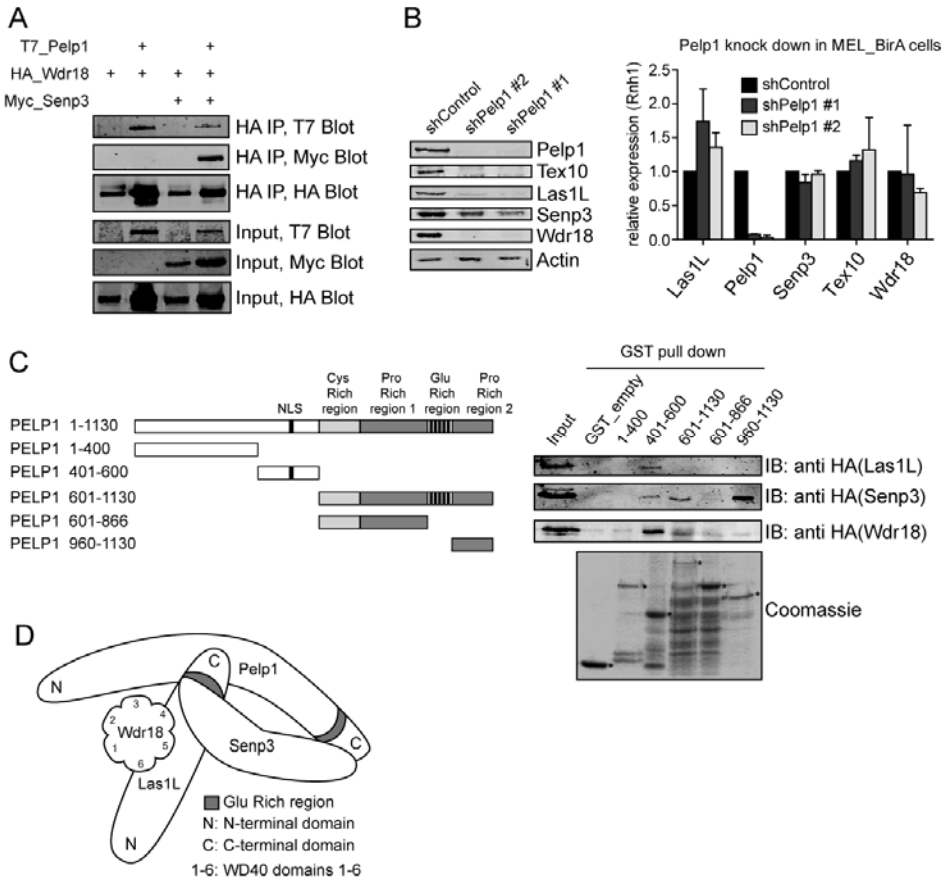


Figure 3. Pelp1 is the core component of the 5FMC complex

(A) Pelp1 is required for interaction between Senp3 and Wdr18. 293T cells were transiently transfected with expression vectors encoding T7_Pelp1, HA_Wdr18 and Myc_Senp3. Cell lysates were analyzed by IP and western blotting with the antibodies indicated. (B) Pelp1 is needed for the stability of the 5FMC complex. MEL_BirA cells were treated with the indicated shRNAs. Nuclear lysates were analyzed by western blotting with the indicated antibodies. Actin staining serves as a loading control. Total RNA was extracted from MEL_BirA cells transduced with the indicated shRNA and analyzed by RT-QPCR for Las1L, Pelp1, Senp3, Tex10 and Wdr18. Error bars: SD of triplicate experiment. (C) Mapping the interaction regions of Pelp1. Schematic representation of Pelp1 deletion constructs. GST fused Pelp1 domains and GST alone (GST_empty) were immobilized onto glutathione beads (lower panel stained with coomassie) and used to pull down nuclear cell lysates from 293T cells expressing HA_Las1L, HA_Senp3 and HA_Wdr18. Asterisks indicate GST fusion proteins. (D) Schematic representation of interaction between Pelp1, Las1L, Senp3 and Wdr18 protein domains.

Chtop recruits 5FMC to Zbp-89

MS analysis of the Zbp-89 interactome revealed Chtop and several Chtop-associated factors as potential interaction partners of Zbp-89 in MEL cells (manuscript in preparation). To further explore the possibility that 5FMC interacts with Zbp-89, we performed bio_Zbp-89 pull downs in MEL_BirA cells followed by immunoblotting. We observed that Chtop, Prmt1 and 5FMC components associated with Zbp-89 (Supplemental Fig. 7). These interactions were DNA independent, as degradation of DNA by Benzonase treatment did not affect the efficiency of co-purification. To investigate whether Chtop is required for the interaction between Zbp-89 and 5FMC, we examined the association of Zbp-89 and 5FMC complex components in Chtop knockdown cells. Bio_Zbp-89 was precipitated more efficiently when Chtop protein levels were reduced, suggesting that bio_Zbp-89 was more accessible in the absence of Chtop (Fig. 4A). Indeed, co-purification of 5FMC complex components was reduced to ~30% of control samples (Fig. 4A and 4B). These results indicate that Chtop is required for the association of Zbp-89 with the 5FMC complex.

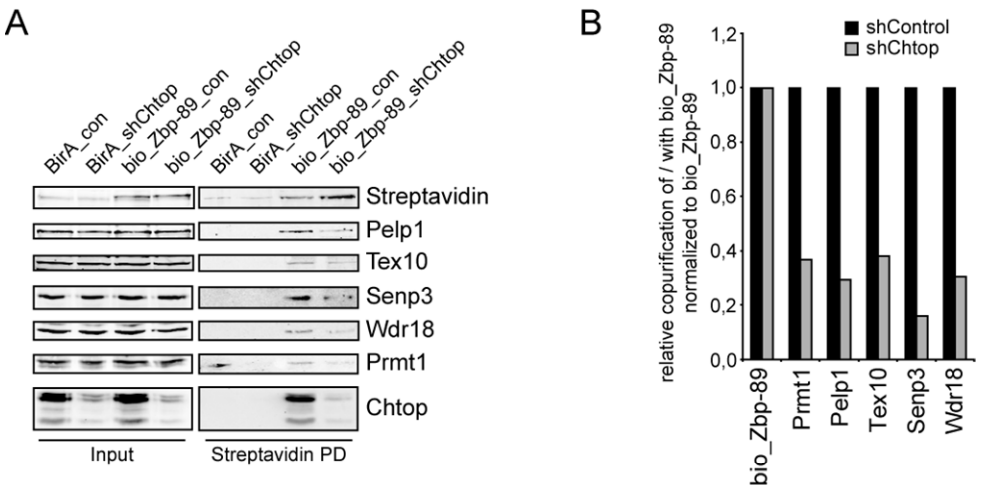


Figure 4. Chtop recruits 5FMC to Zbp-89

(A) Zbp-89 is associated with Chtop and 5FMC complex. Nuclear cell lysates (Input) of MEL_BirA cells (BirA) and MEL_BirA cells expressing biotinylated Zbp-89 (bio_Zbp-89) treated with lentivirus expressing shRNA against Chtop (shChtop) and control lentivirus (shControl) analyzed by streptavidin pull down (Streptavidin PD) and western blotting with the antibodies indicated. (B) Quantification of protein levels using the Odyssey Infrared Imaging System.

Senp3 regulates the sumoylation-status of Zbp-89

It has been reported that Zbp-89 can be post-translationally modified by SUMO in transient overexpression experiments (44). To investigate whether endogenous Zbp-89 could be sumoylated, we first performed streptavidin pull downs from bio_HA_SUMO-2 expressing MEL_BirA nuclear extracts. Staining with an antibody recognizing Zbp-89 detects multiple sumoylated Zbp-89 species (Fig. 5A, lower panel), in line with the observation that Zbp-89 contains at least 2 domains that can be sumoylated (44). Knockdown of Senp3, using two different shRNAs, in these cells led to a significant increase of SUMO-2 detection in whole cell lysates, as well as of sumoylated Zbp-89. It should be noted that the reduction of Senp3 expression affected cell growth and survival, thereby limiting the effect of the knockdown. Next, we performed similar experiments in MEL cells expressing bio_Zbp-89. Streptavidin pull downs probed with an anti-SUMO 2/3 antibody showed that Senp3 depletion resulted in an increase of the levels of SUMO modified bio_Zbp-89 and the appearance of a slower mobility form of bio_Zbp-89 when probed with anti-Zbp-89 antibody (Fig. 5B). The upper band is consistent with SUMO modification. Taken together, these results show that Zbp-89 is sumoylated in vivo, and that Senp3 plays a role in this process.

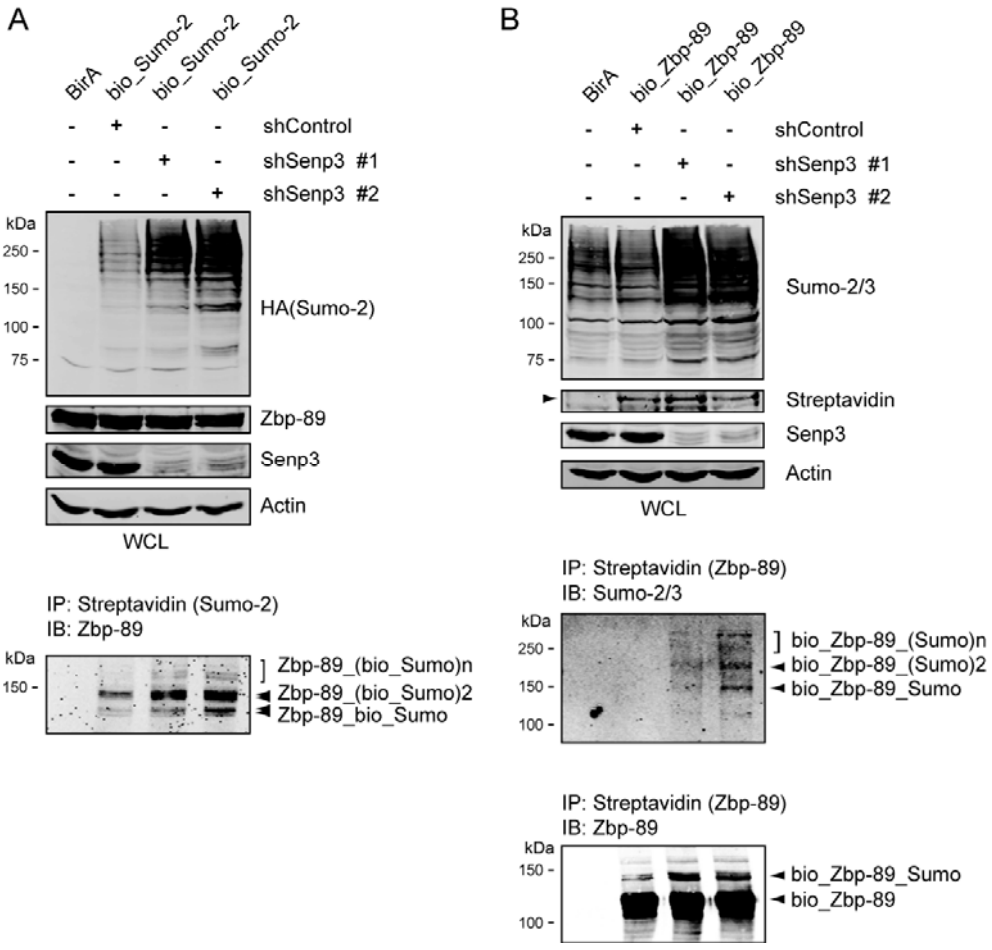


Figure 5. Senp3 regulates Zbp-89 sumoylation

(A-B) Senp3 plays a role in Zbp-89 sumoylation. (A) MEL_BirA (BirA) and MEL_BirA cells expressing bio_HA_Sumo2 (bio_Sumo2) were treated with lentiviruses expressing two different shRNAs against Senp3 (shSenp3 #1, shSenp3 #2) and control lentivirus (shControl). Whole cell lysates (WCL) analyzed by western blotting with the indicated antibodies. Actin staining serves as a loading control. Nuclear extracts were pull down using magnetic streptavidin beads and analyzed by western blotting with anti-Zbp-89 antibody. (B) MEL_BirA (BirA) and MEL_BirA cells expressing biotinylated Zbp-89 (bio_Zbp-89) were treated with lentiviruses expressing two different shRNAs against Senp3 (shSenp3 #1, shSenp3 #2) and control lentivirus (shControl). Whole cell lysates (WCL) analyzed by western blotting with the antibodies indicated. Actin staining serves as a loading

control. Nuclear extracts were pulled down using magnetic streptavidin beads and analyzed by western blotting with anti-Zbp-89 and anti Sumo-2/3 antibodies.

5FMC is critical for Zbp-89 dependent gene regulation

The observations that 5FMC is a desumoylating complex that is recruited by Zbp-89 suggest that it is involved in transcriptional regulation. To examine whether 5FMC is recruited to Zbp-89 target genes we performed chromatin immunoprecipitation (ChIP) experiments for Pelp1, the core component of the complex. We used MEL_BirA cells that ectopically expressed bio_Pelp1, as this increased sensitivity (not shown). After chromatin precipitation, the promoter regions of three (Dusp6, Zbp-89, Atf5) and the coding region of one (Tubb1) Zbp-89 target genes, that were identified by ChIP-sequencing as binding sites of Zbp-89 (manuscript in preparation) and confirmed by ChIP (Supplemental Fig. 8A), were analyzed with the corresponding primers. Occupancy by Pelp1 was indeed observed for these Zbp-89 target genes (Fig. 6A). To further investigate the potential role of 5FMC in transcription regulation, we depleted Pelp1, Senp3 and Chtop in MEL_BirA cells using two different shRNAs (Fig. 6B) and performed ChIP using an antibody against RNA polymerase II (Pol II). Reduced levels of Pelp1, Senp3 and Chtop resulted in reduced Pol II occupancy at the promoter regions of Dusp6, Zbp-89, Atf5 and the coding region of *Tubb1* when compared to cells treated with scrambled control shRNA (Fig. 6C). To validate the Pol II ChIP observations we performed quantitative RT-PCR analysis for the Zbp-89 target genes after depletion of Senp3 and Pelp1. As expected, Dusp6, Tubb1, Zbp-89 and Atf5 mRNA levels were decreased (Fig. 6D). To exclude that the observed differences in occupancy by Pol II were not due to altered occupancy by Zbp-89 we performed ChIP using an antibody against Zbp-89. Binding of Zbp-89 to the regions of its target genes was unaffected upon depletion of Senp3 and Pelp1, while binding of Zbp-89 was reduced when Chtop was depleted (Supplemental Fig. 8B). The reduction of Zbp-89 binding upon depletion of Chtop was due to reduced protein levels of Zbp-89 (Fig. 6B lower panel). The same regions were also tested for changes in the histone modifications H3K4 and H3K27, but no changes were detected (not shown). Collectively, our data indicate that 5FMC is recruited to Zbp-89 target genes and that it is involved in their transcriptional activation by Zbp-89. This most likely involves desumoylation of Zbp-89 and possibly of other factors.

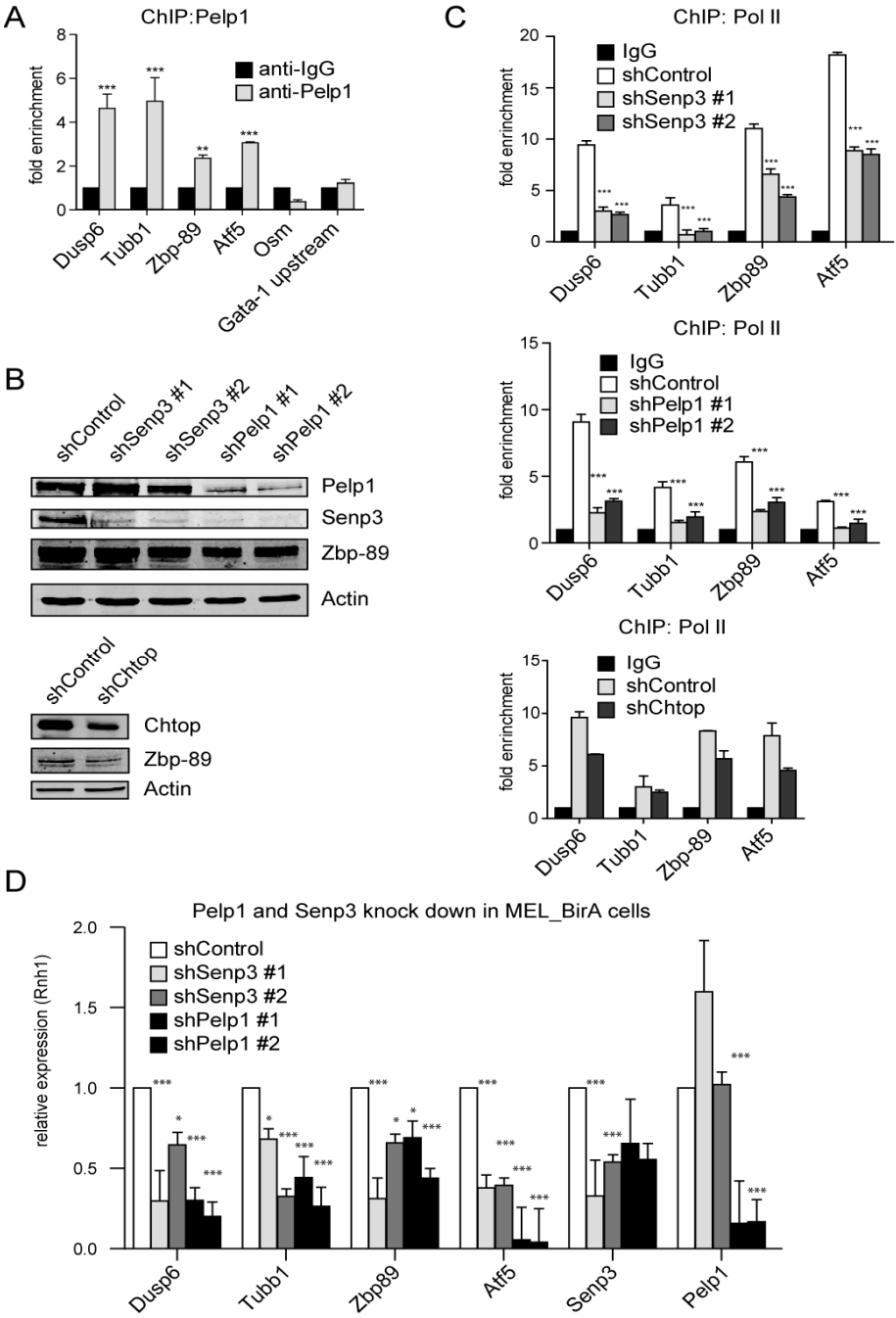

A
ChIP:Pelp1
fold enrinchment
anti-IgG
anti-Pelp1
Dusp6
Tubb1
Zbp-89
Atf5
Osm
Gata-1 upstream
B
shControl
shSenp3 #1
shSenp3 #2
shPelp1 #1
shPelp1 #2
Pelp1
Senp3
Zbp-89
Actin
shControl
shChtop
Chtop
Zbp-89
Actin
C
ChIP: Pol II
IgG
shControl
shSenp3 #1
shSenp3 #2
Dusp6
Tubb1
Zbp89
Atf5
ChIP: Pol II
IgG
shControl
shPelp1 #1
shPelp1 #2
ChIP: Pol II
IgG
shControl
shChtop
Zbp-89
D
Pelp1 and Senp3 knock down in MEL_BirA cells
relative expression (Rnh1)
shControl
shSenp3 #1
shSenp3 #2
shPelp1 #1
shPelp1 #2
Senp3
Pelp1

Figure 6. 5FMC is involved in the regulation of Zbp-89 target genes

(A) Pelp1 is recruited at the promoter regions of Dusp6, Zbp-89, Atf5 and the coding region of Tubb1. MEL_BirA cells that ectopically expressed bio_Pelp1 analysed by ChIP using Pelp1 antibody for the indicated gene promoter or coding regions. The promoter region of the *Osm* gene and the region upstream of Gata-1 promoter (Gata-1 upstream) were used as negative controls. *** Indicates P<0.001, ** indicates P<0.01. Error bars: SD of triplicate experiment. (B) Knockdown of Pelp1, Senp3 and Chtop in MEL_BirA cells. MEL_BirA cells were treated with the indicated shRNAs. Cell lysates were analyzed by western blotting with the indicated antibodies. Actin staining serves as a loading control. (C) Pelp1, Senp3 and Chtop knockdowns reduced RNA polymerase II (Pol II) occupancy at the promoter regions of *Dusp6, Zbp-89, Atf5* and the coding region of Tubb1. MEL_BirA cells were treated as in (B). ChIP analysis at the indicated regions was performed using Poll II antibody. *** indicates P<0.001. Error bars: SD of triplicate experiment. (D) Regulation of Zbp-89 target genes by Senp3 and Pelp1. Total RNA was extracted from MEL_BirA cells treated as in (B) and analyzed by RT-QPCR for Dusp6, Tubb1, Zbp-89, Atf5, Senp3 and Pelp1. *** indicates P<0.001, ** indicates P<0.01, * indicates P<0.05. Error bars: SD of triplicate experiment.

DISCUSSION

In the present study, we have identified 5FMC, a desumoylating protein complex that exclusively binds to arginine-methylated Chtop. It acts as a key regulator of Zbp-89 sumoylation and is required for full transcriptional activation of Zbp-89 dependent genes. To our knowledge, this is the first description of a mechanism that adds specificity to desumoylation processes.

The 5FMC complex is composed of five proteins: Pelp1, Las1L, Tex10, Senp3, and Wdr18. These factors are amongst the most abundant proteins present in Chtop purifications, but only when Chtop is methylated by Prmt1. little is known about most of the components of the 5FMC complex. Pelp1 is a coactivator involved in nuclear hormone signaling (41), Senp3 is a SUMO-specific protease (42), while Wdr18, Las1L, and Tex10 have not been characterized yet. In yeast, the proteins Rix1, Ipi1 and Ipi3 (S. cerevisiae) / Crb3 (S. pombe) share conserved regions with Pelp1, Tex10 and Wdr18, respectively. They form the IPI complex and have been shown to function in ribosomal RNA processing (45). It was recently shown that Las1, the yeast ortholog of Las1L, is also associated with the IPI complex (43). Furthermore, PELP1, TEX10, LAS1L, SENP3, and WDR18 were recently linked to ribosome biogenesis in human cells, indicating an evolutionary conserved complex with a role in ribosomal RNA processing. In human cells, the complex was localized mainly in the nucleoplasm with

a subfraction present in the nucleolus (39-40). In this study, we observed a similar distribution in mouse cells.

Additionally, several studies suggested a role for 5FMC components in transcriptional regulation and (de)sumoylation events, although this had not been explored further. Doseff and Arndt proposed in their initial identification of Las1 in S. cerevisiae that it functions as a transcription factor (46), while it was recently shown to localize to heterochromatic regions (43). In human cells, components of the 5FMC complex were first detected in the MS analysis of the MLL1-WDR5 complex, a complex that regulates transcription activation by H3K4 methylation (38). The 5FMC complex was also found in a recent study on human coregulator protein complex networks obtained from integrative mass spectrometry-based analysis of 3290 antibody-based affinity purifications (47). Thus, the data obtained with this approach independently support the composition of the 5FMC complex reported here. Furthermore, PELP1, LAS1L, TEX10 and SENP3 were also detected together with components of the CoREST1/HDAC1 co-repressor complex in a MS study for proteins that interact with SUMO-2 (48). Sumoylation is important for stability and recruitment of repressive complexes such as CoREST, NuRD, and SetDB1 (18), indicating that desumoylation of transcription factors and corepressors is required for derepression.

Our results indicate that Pelp1 is the core component of 5FMC. Pelp1 has the ability to interact with nuclear receptors (NRs) and enhances transcription of their target genes (23). Pelp1 has been shown to interact with the acetyltransferases CBP and p300 (22), KDM1, a member of the CoREST1 repressor complex (49), and deacetylases, including components of the NuRD repressor complex (50). Interestingly, Pelp1 was also identified in a blind screen for SUMO-2 interacting proteins (51). This opens the possibility that Pelp1 acts, in addition to its scaffold function within 5FMC, as a Sumo-2 sensor to detect Senp3 substrates.

As Chtop is strongly associated with chromatin (19), while 5FMC mainly resides in the nucleoplasm (this paper), the interactions between Chtop and 5FMC are most likely transient and highly dynamic. Possibly, Chtop recruits 5FMC complex, in order to desumoylate its substrates, in a “hit-and-run” manner. In addition, 5FMC may very well act as a desumoylation complex outside the chromatin environment and independent of Chtop, e.g. in ribosome biogenesis (39).

In line with our previous observation that Chtop colocalizes to H3K27me3 (19), we found that Chtop interacts with the Prc1 complex. In contrast to 5FMC, this association does not depend on the methylation status of Chtop. Intriguingly, the Pc2 component of Prc1 (also known as Cbx4) has been

identified as a SUMO-ligase for several transcriptional regulators (52-53). Although these studies mainly focused on SUMO-1 modification, Chtop may recruit 5FMC as an antagonist of Pc2/Cbx4.

We have shown that Chtop recruits 5FMC to Zbp-89 and that Zbp-89 is subsequently desumoylated by Senp3, resulting in higher Pol II levels on Zbp-89 target genes. We anticipated that depletion for Senp3 and Pelp1 would also affect H3K4 and/or H3K27 methylation of these regions, as 5FMC is might be connected with the MLL1-WDR5 (38) and Prc1 complexes (this paper). However, no changes were observed, which is in line with the observation that overexpression of catalytically inactive Senp3 does not affect diMeH3K4 of specific promoters (48).

Collectively, these observations suggest a model where methylated Chtop recruits the 5FMC complex to factors like Zbp-89. Subsequently, the Senp3 protease desumoylates Zbp-89 and possibly additional components of repressor complexes, resulting in the stimulation of transcription of target genes (Fig. 7).

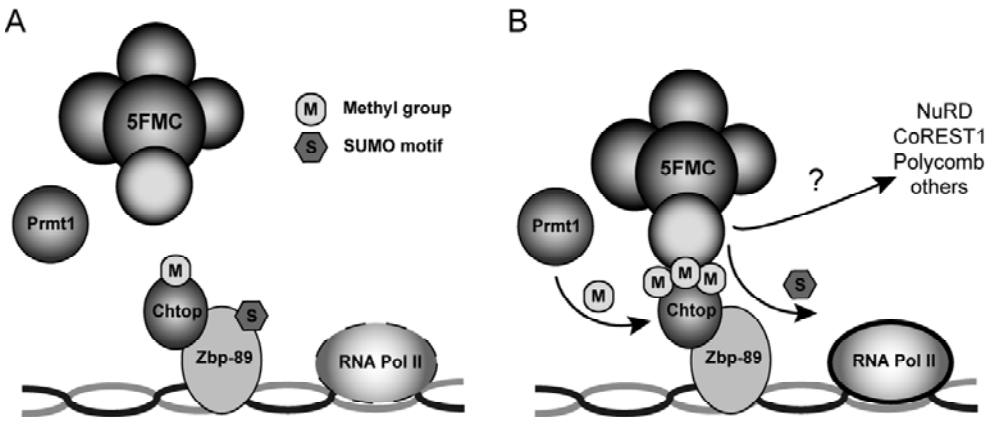


Figure 7. 5FMC complex recruitment to Zbp-89 stimulates transcription of target genes

(A-B) Model illustrating 5FMC complex function in Zbp-89 dependent gene expression. Methylated Chtop recruits the 5FMC complex to transcription factor Zbp-89. Upon 5FMC binding, Senp3 protease desumoylates Zbp-89 and possibly additional components of repressor complexes, resulting in the stimulation of transcription of target genes.

ACKNOWLEDGEMENTS

We thank Karel Bezstarosti and Erikjan Rijkers for technical assistance.

This work was supported by the Netherlands Genomics Initiative (93518009 and 93511036), the Landsteiner Foundation for Blood Transfusion Research (1040), the Netherlands Scientific Organization (NWO DN 82-301 and ZonMW 912-07-019 and 40-00812-98-08032) and the Centre for Biomedical Genetics.

AUTHOR CONTRIBUTIONS

F.G., S.P. and T.B. designed experiments. P.F., N.G., T.B., F.E., F.P. and A.A. performed experiments. P.F., T.B. analyzed results. J.D. provided expertise, analysis tools and infrastructure. J.D. analyzed data. P.F., S.P., and T.B. wrote the paper.

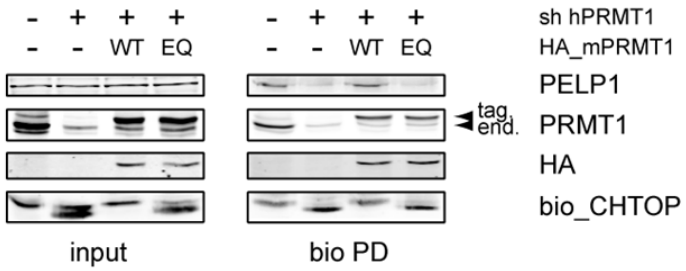
REFERENCES

1. Verger, A., Perdomo, J., and Crossley, M. (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO Rep* 4, 137-142.
2. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. *Mol Cell* 33, 1-13.
3. Pawlak, M. R., Scherer, C. A., Chen, J., Roshon, M. J., and Ruley, H. E. (2000) Arginine N-methyltransferase 1 is required for early postimplantation mouse development, but cells deficient in the enzyme are viable. *Mol Cell Biol* 20, 4859-4869.
4. O'Brien, K. B., Alberich-Jorda, M., Yadav, N., Kocher, O., Diruscio, A., Ebralidze, A., Levantini, E., Sng, N. J., Bhasin, M., Caron, T., Kim, D., Steidl, U., Huang, G., Halmos, B., Rodig, S. J., Bedford, M. T., Tenen, D. G., and Kobayashi, S. (2010) CARM1 is required for proper control of proliferation and differentiation of pulmonary epithelial cells. *Development* 137, 2147-2156.
5. Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Nimer, S. D. (2008) Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev* 22, 640-653.
6. Mowen, K. A., Tang, J., Zhu, W., Schurter, B. T., Shuai, K., Herschman, H. R., and David, M. (2001) Arginine methylation of STAT1 modulates IFNalpha/beta-induced transcription. *Cell* 104, 731-741.
7. Wang, H., Huang, Z. Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B. D., Briggs, S. D., Allis, C. D., Wong, J., Tempst, P., and Zhang, Y. (2001) Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* 293, 853-857.
8. An, W., Kim, J., and Roeder, R. G. (2004) Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell* 117, 735-748.
9. Rezai-Zadeh, N., Zhang, X., Namour, F., Fejer, G., Wen, Y. D., Yao, Y. L., Gyory, I., Wright, K., and Seto, E. (2003) Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. *Genes Dev* 17, 1019-1029.
10. Covic, M., Hassa, P. O., Saccani, S., Buerki, C., Meier, N. I., Lombardi, C., Imhof, R., Bedford, M. T., Natoli, G., and Hottiger, M. O. (2005) Arginine methyltransferase CARM1 is a promoter-specific regulator of NF-kappaB-dependent gene expression. *EMBO J* 24, 85-96.
11. Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., Ybarra, R., Qiu, Y., Felsenfeld, G., Bungert, J., and Huang, S. (2010) H4R3 methylation facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* 115, 2028-2037.
12. Cheng, D., Cote, J., Shaaban, S., and Bedford, M. T. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell* 25, 71-83.
13. Hay, R. T. (2005) SUMO: a history of modification. *Mol Cell* 18, 1-12.
14. Johnson, E. S. (2004) Protein modification by SUMO. *Annu Rev Biochem* 73, 355-382.
15. Stielow, B., Sapetschnig, A., Wink, C., Kruger, I., and Suske, G. (2008) SUMO-modified Sp3 represses transcription by provoking local heterochromatic gene silencing. *EMBO Rep* 9, 899-906.
16. Fernandez-Lloris, R., Osses, N., Jaffray, E., Shen, L. N., Vaughan, O. A., Girwood, D., Bartrons, R., Rosa, J. L., Hay, R. T., and Ventura, F. (2006) Repression of SOX6 transcriptional activity by SUMO modification. *FEBS Lett* 580, 1215-1221.
17. Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G. G., Krones, A., Ohgi, K. A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F. L., Korach, K. S., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446, 882-887.
18. Garcia-Dominguez, M., and Reyes, J. C. (2009) SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim Biophys Acta* 1789, 451-459.
19. van Dijk, T. B., Gillemans, N., Stein, C., Fanis, P., Demmers, J., van de Corput, M., Essers, J., Grosveld, F., Bauer, U. M., and Philipsen, S. (2010) Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases. *Mol Cell Biol* 30, 260-272.
20. Zullo, A. J., Michaud, M., Zhang, W., and Grusby, M. J. (2009) Identification of the small protein rich in arginine and glycine (SRAG): a newly identified nucleolar protein that can regulate cell proliferation. *J Biol Chem* 284, 12504-12511.
21. van Dijk, T. B., Gillemans, N., Pourfarzad, F., van Lom, K., von Lindern, M., Grosveld, F., and Philipsen, S. (2010) Fetal globin expression is regulated by Friend of Prmt1. *Blood* 116, 4349-4352.
22. Nair, S. S., Mishra, S. K., Yang, Z., Balasenthil, S., Kumar, R., and Vadlamudi, R. K. (2004) Potential role of a novel transcriptional coactivator PELP1 in histone H1 displacement in cancer cells. *Cancer Res* 64, 6416-6423.
23. Vadlamudi, R. K., Wang, R. A., Mazumdar, A., Kim, Y., Shin, J., Sahin, A., and Kumar, R. (2001) Molecular cloning and characterization of PELP1, a novel human coregulator of estrogen receptor alpha. *J Biol Chem* 276, 38272-38279.
24. Kaufman, R. J., Davies, M. V., Pathak, V. K., and Hershey, J. W. (1989) The phosphorylation state of eucaryotic initiation factor 2 alters translational efficiency of specific mRNAs. *Mol Cell Biol* 9, 946-958.
25. Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F., and Hollis, M. (1992) LCR/MEL: a versatile system for high-level expression of heterologous proteins in erythroid cells. *Nucleic Acids Res* 20, 997-1003.
26. 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 U S A* 100, 7480-7485.
27. van Dijk, T. B., van Den Akker, E., Amelsvoort, M. P., Mano, H., Lowenberg, B., and von Lindern, M. (2000) Stem cell factor induces phosphatidylinositol 3'-kinase-dependent Lyn/Tec/Dok-1 complex formation in hematopoietic cells. *Blood* 96, 3406-3413.

28. Vizcaino, J. A., Cote, R., Reisinger, F., Barsnes, H., Foster, J. M., Rameseder, J., Hermjakob, H., and Martens, L. (2010) The Proteomics Identifications database: 2010 update. *Nucleic Acids Res* 38, D736-742.
29. Barsnes, H., Vizcaino, J. A., Eidhammer, I., and Martens, L. (2009) PRIDE Converter: making proteomics data-sharing easy. *Nat Biotechnol* 27, 598-599.
30. Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepfer, A. M., Hinkle, G., Piqani, B., Eisenhaure, T. M., Luo, B., Grenier, J. K., Carpenter, A. E., Foo, S. Y., Stewart, S. A., Stockwell, B. R., Hacohen, N., Hahn, W. C., Lander, E. S., Sabatini, D. M., and Root, D. E. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124, 1283-1298.
31. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L., and Trono, D. (1997) Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* 15, 871-875.
32. Esteghamat, F., van Dijk, T. B., Braun, H., Dekker, S., van der Linden, R., Hou, J., Fanis, P., Demmers, J., van Ijcken, W., Ozgur, Z., Horos, R., Pourfarzad, F., von Lindern, M., and Philipsen, S. (2011) The DNA binding factor Hmg20b is a repressor of erythroid differentiation. *Haematologica*.
33. Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., and Kingston, R. E. (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98, 37-46.
34. Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van 't Veer, L. J., Berns, A., and van Lohuizen, M. (1997) Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev* 11, 226-240.
35. Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R. E. (2002) The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* 22, 6070-6078.
36. Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A., and Matsumoto, K. (1995) Replication factor encoded by a putative oncogene, set, associated with myeloid leukemogenesis. *Proc Natl Acad Sci U S A* 92, 4279-4283.
37. Hautbergue, G. M., Hung, M. L., Walsh, M. J., Snijders, A. P., Chang, C. T., Jones, R., Ponting, C. P., Dickman, M. J., and Wilson, S. A. (2009) UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. *Curr Biol* 19, 1918-1924.
38. Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 121, 873-885.
39. Finkbeiner, E., Haindl, M., and Muller, S. (2011) The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. *EMBO J* 30, 1067-1078.
40. Castle, C. D., Cassimere, E. K., and Denicourt, C. (2012) LAS1L interacts with the mammalian Rix1 complex to regulate ribosome biogenesis. *Mol Biol Cell* 23, 716-728.
41. Vadlamudi, R. K., and Kumar, R. (2007) Functional and biological properties of the nuclear receptor coregulator PELP1/MNAR. *Nucl Recept Signal* 5, e004.
42. Nishida, T., Tanaka, H., and Yasuda, H. (2000) A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized in the nucleolus at interphase. *Eur J Biochem* 267, 6423-6427.
43. Kitano, E., Hayashi, A., Kanai, D., Shinmyozu, K., and Nakayama, J. (2011) Roles of fission yeast grc3 protein in ribosomal RNA processing and heterochromatic gene silencing. *J Biol Chem* 286, 15391-15402.
44. Chupreta, S., Brevig, H., Bai, L., Merchant, J. L., and Iniguez-Lluhi, J. A. (2007) Sumoylation-dependent control of homotypic and heterotypic synergy by the Kruppel-type zinc finger protein ZBP-89. *J Biol Chem* 282, 36155-36166.
45. Krogan, N. J., Peng, W. T., Cagney, G., Robinson, M. D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D. P., Beattie, B. K., Lalev, A., Zhang, W., Davierwala, A. P., Mnaimneh, S., Starostine, A., Tikuisis, A. P., Grigull, J., Datta, N., Bray, J. E., Hughes, T. R., Emili, A., and Greenblatt, J. F. (2004) High-definition macromolecular composition of yeast RNA-processing complexes. *Mol Cell* 13, 225-239.
46. Doseff, A. I., and Arndt, K. T. (1995) LAS1 is an essential nuclear protein involved in cell morphogenesis and cell surface growth. *Genetics* 141, 857-871.
47. Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B. J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W., and Qin, J. (2011) Analysis of the human endogenous coregulator complexome. *Cell* 145, 787-799.
48. Ouyang, J., Shi, Y., Valin, A., Xuan, Y., and Gill, G. (2009) Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol Cell* 34, 145-154.
49. Nair, S. S., Nair, B. C., Cortez, V., Chakravarty, D., Metzger, E., Schule, R., Brann, D. W., Tekmal, R. R., and Vadlamudi, R. K. (2010) PELP1 is a reader of histone H3 methylation that facilitates oestrogen receptor-alpha target gene activation by regulating lysine demethylase 1 specificity. *EMBO Rep* 11, 438-444.
50. Mishra, S. K., Balasenthil, S., Nguyen, D., and Vadlamudi, R. K. (2004) Cloning and functional characterization of PELP1/MNAR promoter. *Gene* 330, 115-122.
51. Rosendorff, A., Sakakibara, S., Lu, S., Kieff, E., Xuan, Y., DiBacco, A., Shi, Y., and Gill, G. (2006) NXP-2 association with SUMO-2 depends on lysines required for transcriptional repression. *Proc Natl Acad Sci U S A* 103, 5308-5313.
52. Wotton, D., and Merrill, J. C. (2007) Pc2 and SUMOylation. *Biochem Soc Trans* 35, 1401-1404.
53. Kagey, M. H., Melhuish, T. A., and Wotton, D. (2003) The polycomb protein Pc2 is a SUMO E3. *Cell* 113, 127-137.

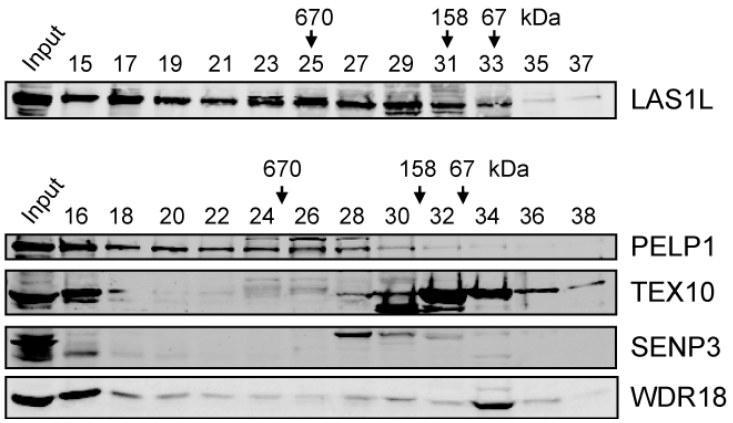
	Sequence (5´-3´)
QPCR	
Rnh1_Fwd	TCCAGTGTGAGCAGCTGAG
Rnh1_Rev	TGCAGGCACTGACTGAAGCACCA
Senp3_Fwd	GAGAGGGTCTCCACCAGTGCC
Senp3_Rev	CTGCGCTCCCCTTCAGGACCA
Wdr18_Fwd	CTGGCCTGGAGCCTCTGCAGT
Wdr18_Rev	CCAATGAAGCGGTGGCCACCC
Las1L_Fwd	CGAGTGCCAGGGGTATCGTCG
Las1L_Rev	GCCAGGGGCAGTTCGTTGCCT
Tex10_Fwd	GCTGCTTGGAGTCTCTTGGAGG
Tex10_Rev	GCCAGCCAACGGGAGAGCACT
Pelp1_Fwd	GCTGGGCCTCAGACCAGAGTG
Pelp1_Rev	CTCGCAGGCCAGCTGTTGGAG
Dusp6_Fwd	GGCGAGTTCAAATACAAGCAA
Dusp6_Rev	ACCAGGACACCACAGTTTTTG
Tubb1_Fwd	TAAGAAGTATGTGCCGCGAG
Tubb1_Rev	AGTTGTTACCAGCACCAGAGTTA
Zbp89_Fwd	CATGAGGAGACAGTGAAAAATG
Zbp89_Rev	CTGTAAGTCTACTGGCTCTCTG
Atf5_Fwd	CTGGCTCGTAGACTATGGGAA
Atf5_Rev	TCCAATCAGAGAAGCCGTCA
ChIP	
Amylase_Fwd	CTCCTTGTACGGGTTGGT
Amylase_Rev	AATGATGTGCACAGCTGAA
Dusp6_Fwd	ACACACGATCTAAAGGAGGAC
Dusp6_Rev	CAATTAGCAAGCACAAAAGC
Tubb1_Fwd	GGATGTACAAGTGTCTCTGAGC
Tubb1_Rev	TATCTTTCCGCTCATTTCC
Zbp-89_Fwd	CTGGGAGGAGGAAGAGAAG
Zbp-89_Rev	GAGAGAACTTTTGCTGTGGC
Atf5_Fwd	GGTTCCTCACTTCGTCTCC
Atf5_Rev	TTCACTCTCCGCTCACACC
Osm_Fwd	CTGGGGAGACTTTGGTTTT
Osm_Rev	ACTGGGTCCTGGTACTCTGG
Gata-1 Upstream_Fwd	TGATGGCTTCTACTAGGCACACG
Gata-1 Upstream_Rev	GGCTTCACTCCCAGGAATGTAGG

Supplementary Table 1. Oligonucleotides used in this study



Supplemental Figure 1. Methylation of Chtop is required for Pelp1 recruitment

293T cells were depleted for endogenous PRMT1 using an shRNA against PRMT1 (sh hPRMT1). PRMT1 depleted cells were transiently transfected with expression vectors encoding mouse wild type (WT) and enzymatic inactive (EQ) Prmt1 (HA_mPRMT1). Cell lysates were analyzed by streptavidin pull-down (bio PD) and Western blotting with the antibodies indicated. Arrows indicate endogenous (end.) and HA-tagged (tag.) Prmt1.

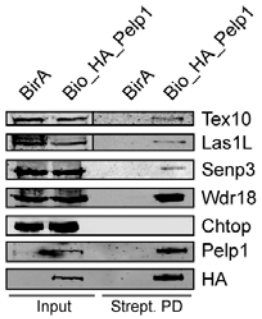


Supplemental Figure 2. 5FMC elution patterns in 293T cells

293T cell nuclear extracts were analyzed by sized-exclusion chromatography on a Superose 6 column. Proteins eluted from the indicated fractions were blotted with the indicated antibodies. Molecular mass markers are indicated at the top.

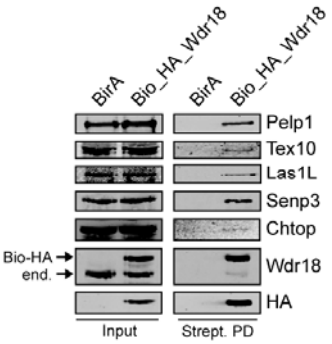
A

Strept. PD of Bio_HA_Pelp1		
Protein	uniq. pept.	coverage (%)
Las1L	19	26.3
Wdr18	14	39.7
Tex10	16	19.2
Pelp1	11	11.4
Senp3	8	16.9
Nol9	4	6.4



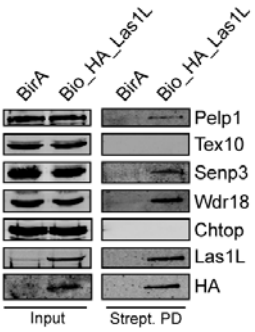
B

Strept. PD of Bio_HA_Wdr18		
Protein	uniq. pept.	coverage (%)
Las1L	20	31.3
Wdr18	15	41.3
Tex10	16	21
Pelp1	15	18.1
Senp3	15	30.1
Nol9	9	10.9



C

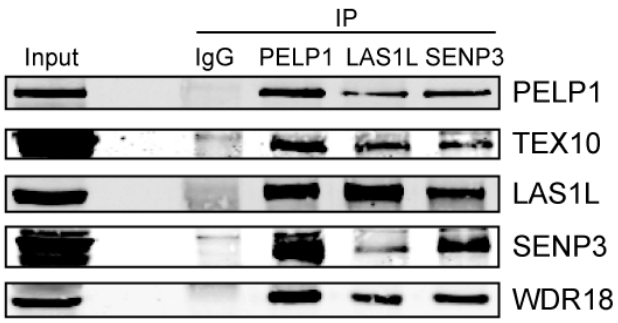
Strept. PD of Bio_HA_Las1L		
Protein	uniq. pept.	coverage (%)
Pelp1	9	9.7
Wdr18	9	23
Tex10	4	4.5
Las1L	4	5.9
Senp3	4	7.9



Supplemental Figure 3. 5FMC nuclear interactions

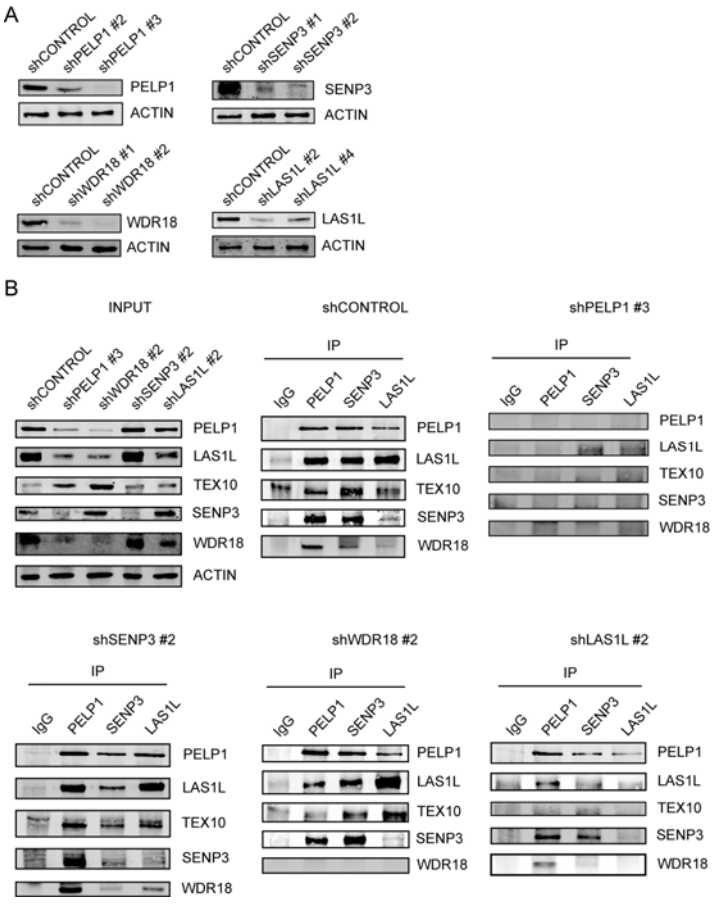
(A) Pelp1 interactions in MEL cells. Whole cell lysates (Input) and streptavidin pull downs (Strept. PD) from MEL_BirA (BirA) and MEL_BirA cells expressing biotinylated Pelp1 (Bio_HA_Pelp1) were analyzed by MS (table) and western blotting. Immunoblot probed with the antibodies indicated. (B) Wdr18 interactions in MEL cells. Whole cell lysates (Input) and streptavidin pull downs (Strept. PD) from MEL_BirA (BirA) and MEL_BirA cells expressing biotinylated Wdr18 (Bio_HA_Wdr18) were analyzed by MS (table) and western blotting. Immunoblot probed with the antibodies indicated. Arrows indicate endogenous (end.) and biotinylated (Bio_HA) Wdr18. (C) Las1L interactions in MEL cells. Whole cell lysates (Input) and streptavidin pull downs (Strept. PD) from MEL_BirA

(BirA) and MEL_BirA cells expressing biotinylated Las1L (Bio_HA_Las1L) were analyzed by MS (table) and western blotting. Immunoblot probed with the antibodies indicated.



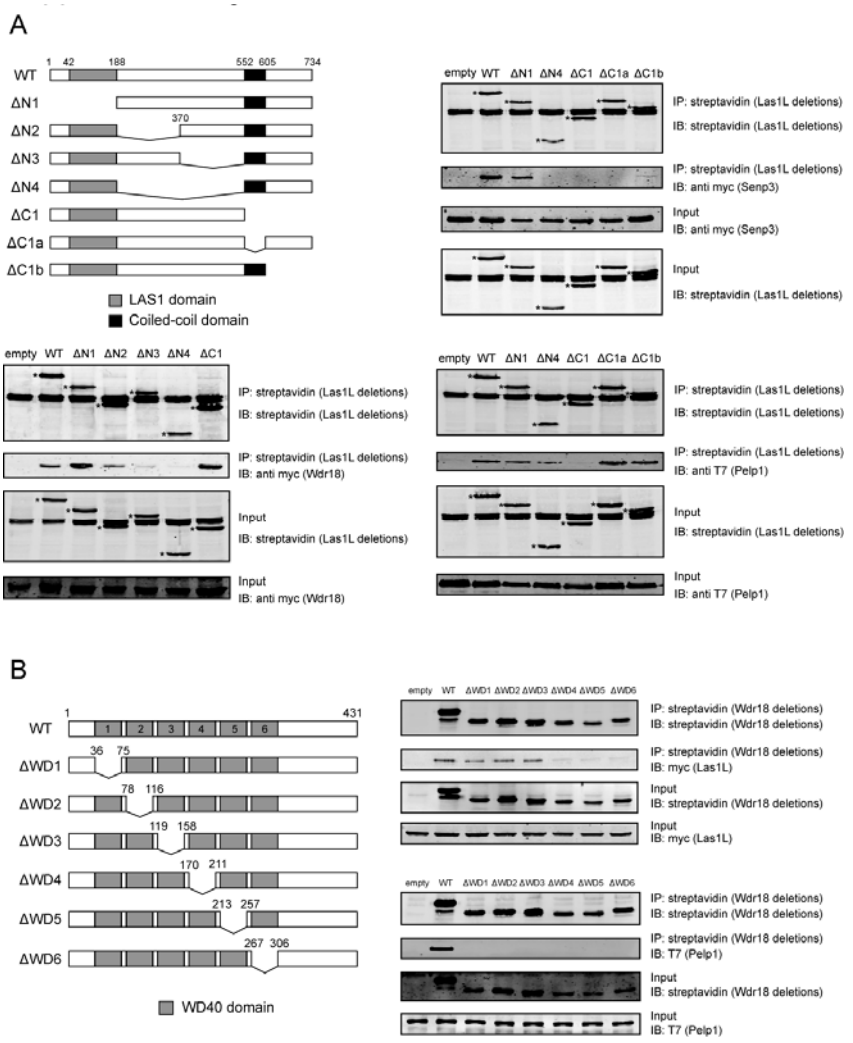
Supplemental Figure 4. 5FMC is a nuclear complex

Endogenous association between the 5FMC components. 293T cell nuclear lysates were analyzed by immunoprecipitation (IP) and western blotting with the antibodies indicated.



Supplemental Figure 5. PELP1 is the key stabilizing component of 5FMC complex

(A-B) 293T cells were transfected with the indicated shRNAs; whole cell lysates were analyzed by Western blotting using the indicated antibodies. Actin served as a loading control. (B) Whole cell lysates (Input) and immunoprecipitations (IP) from 293T cells were analyzed by Western blotting to study the stability and interactions of 5FMC components after depletion of individual 5FMC members. Immunoblots were probed with the indicated antibodies; actin serves as a loading control. Immunoprecipitation results are shown only for the indicated shRNA. Similar results were obtained using the second shRNA construct.

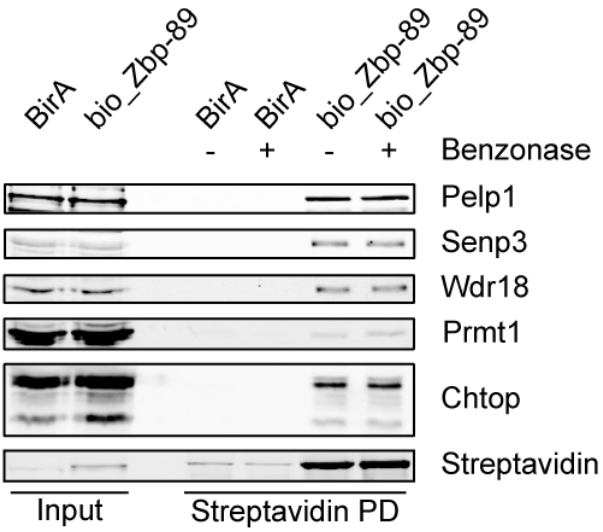


Supplemental Figure 6. Mapping the interaction regions of Las1L and Wdr18

(A) Bio_HA-tagged wild-type Las1L and its deletion mutants were expressed in 293T cells. Cell lysates were immunoprecipitated (IP) and blotted (IB) with the indicated antibodies. Whole cell lysates (Input) were blotted and ectopically expressed wild-type Wdr18 and its deletion mutants were visualized using a streptavidin antibody. Asterisks indicate wild-type Las1L and its deletion mutants.

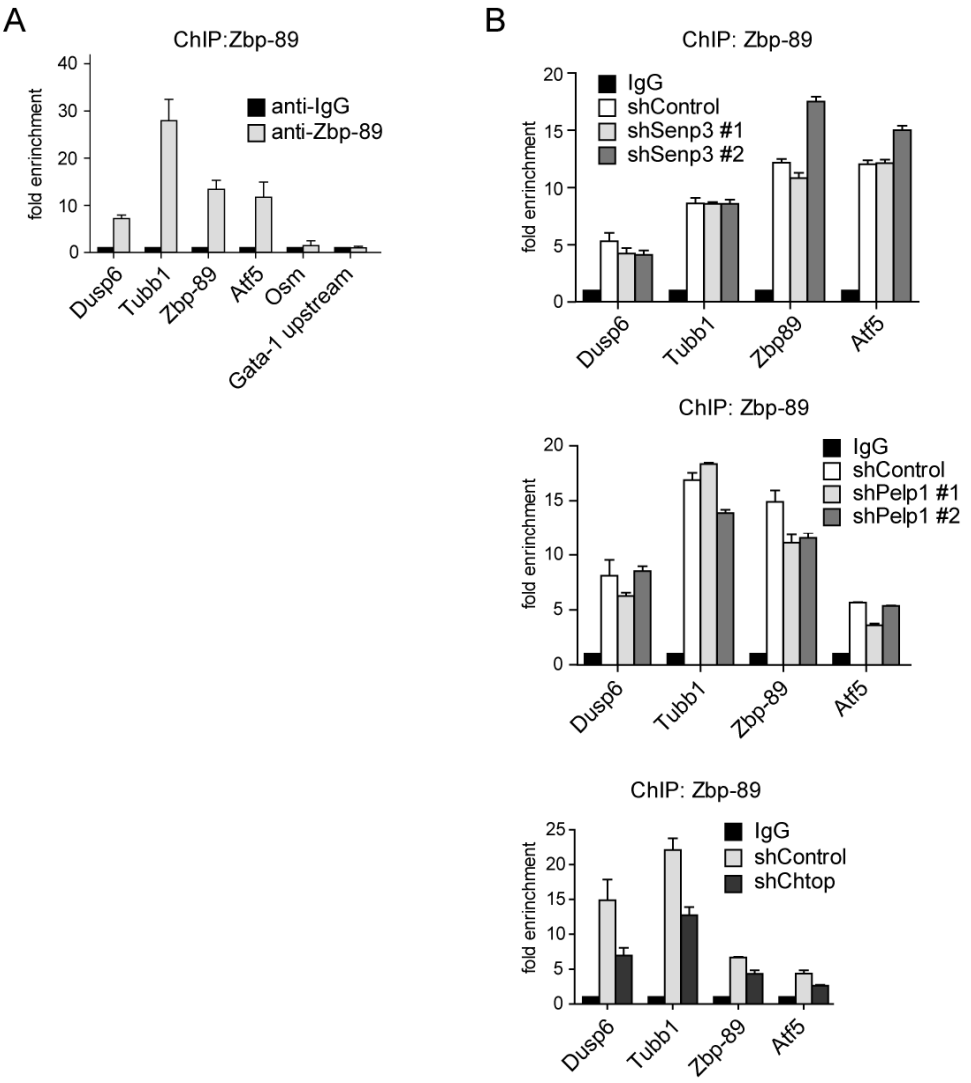
(B) Bio_HA-tagged wild-type Wdr18 and its deletion mutants were expressed in 293T cells. Cell lysates were immunoprecipitated (IP) and blotted (IB) with the antibodies indicated. Whole cell lysates (Input) were blotted

and ectopically expressed wild-type Wdr18 and its deletion mutants were visualized using a streptavidin antibody.



Supplemental Figure 7. Chtop, Prmt1 and 5FMC complex are associated with Zbp-89

Streptavidin pull downs (Streptavidin PD) from MEL_BirA (BirA) and MEL_BirA cells expressing biotinylated Zbp-89 (Bio_Zbp-89) were treated with Benzonase. Whole cell lysates (Input) and streptavidin pull downs were analyzed by western blotting. Immunoblot probed with the antibodies indicated.



Supplemental Figure 8. Recruitment of Zbp-89 to its target genes does not depend on 5FMC

(A) Zbp-89 is recruited to the promoter regions of Dusp6, Zbp-89, Atf5 and the coding region of Tubb1. MEL_BirA cells analysed by ChIP using Zbp89 antibody for the indicated gene promoter or coding regions. The promoter region of the *Osm* gene and the region upstream of Gata-1 promoter (Gata-1 upstream) were used as negative controls. (B) Pelp1, Senp3 and Chtop knockdowns do not affect Zbp-89 binding at the promoter or coding regions of Dusp6, Tubb1, Zbp-89 and Atf5. MEL_BirA cells were treated as in (Fig. 6B). ChIP analysis at the indicated regions was performed using Zbp-89 antibody. Error bars: SD of triplicate experiment.

5

Chapter 5

Discussion

DISCUSSION

β-thalassemia and sickle cell disease (SCD) are one of the most common single gene disorders in the worldwide. Around 7% of the population of the earth carries globin gene mutations. β-hemoglobinopathies are spread from south East Asia to the Middle East, Mediterranean area, South Europe, Africa and the USA. The available treatments for these diseases are risky, inconvenient and pleiotropic. The observation that elevated expression of fetal hemoglobin in patients ameliorates β-thalassemia and sickle cell disease phenotypes led to the idea of γ-globin reactivation for the treatment of these diseases. Expression of the γ-globin genes is restricted to human fetal life and hence understanding the molecular mechanism of γ-globin gene suppression is of great importance to develop new therapeutic strategies. In order to identify protein complexes responsible for γ-globin gene suppression, we developed an innovative and unbiased proteomics approach, Targeted Chromatin Purification (TChP), to purify and identify protein factors bound to γ-globin promoter and regulatory regions (Pourfarzad et al submitted). TChP can be used to isolate a unique sequence, which relies on the binding TetR protein to the TetO sequences. Dejardin and Kingston developed a fundamentally different approach in 2009 approach (proteomics of isolated chromatin segments = PICh), where a DNA probe is hybridized, to a specific locus to isolate its association proteins. There are advantages and disadvantages associated with each method. TChP is likely more efficient than PICh as a result of losses due to the hybridization process. However the introduction of bait sequence may change the behavior of the target sequence, which necessitates the addition of an extra control. On the other hand once the bait sequence has been introduced, using appropriate tags multiple purification steps can be carried out. To identify candidate-binding proteins, it is essential to have negative controls. As a result of mismatched hybridization and sticking of irrelevant chromatin/protein in the PICh method approximately 200 proteins were found to be background proteins and around 75 of these 200 background proteins were also found in TChP method.

We identified 729 unique proteins in four different pull down experiments in the TChP approach. 475 proteins were also detected in the doxycycline treated control pull down as background. Abundant proteins such as ribosomal and cytoplasmic proteins caused 223 proteins to be removed from the list of specific chromatin binding proteins. This leaves 49 proteins and we considered 20 of the remaining 49 proteins on basis of their specific expression in human erythroid cell, and the availability of specific antibodies or shRNA. These 20 included Supt5h, Cdc5l, Apex1, and Zfp148.

One of the most interesting proteins was ZBP-89 (also known as BFCOL1, BERF-1, ZNF148 in human and Zfp148 in mouse). Its 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 GATA-1 and Mafk, and it is involved in erythroid development and differentiation (Woo et al, 2011; Woo et al, 2008). ZBP-89 represses gene transcription via recruitment of HDAC1 to the promoter of its target genes (Wu et al, 2007). This would agree with the known role of HDACs in γ-globin silencing, as HDAC inhibitors can partially (re) activate γ-globin expression (Constantoulakis et al, 1989; Perrine et al, 1989; Perrine et al, 1987). Zbp-89 was also reported to bind to the GATA-1 enhancer and a number of other genes (Ohneda et al, 2009; Vernimmen et al, 2007). Moreover, ZBP-89 binds to the human β-globin LCR and to a lesser extend to sequences upstream and downstream of γ-globin genes. Mouse embryonic day 13.5 erythroid progenitor cells containing the γ-globin minilocus were used for identifying its association proteins. Mouse embryonic day 13.5 which mimics the switching process around the time of birth in humans mentioned above ZBP-89 can act as a suppressor e.g. it suppresses vimentin expression by recruiting HDAC1 (Wu et al, 2007). However, ZBP-89 can also act as an activator as it activates p21 $^{waf1/cip1}$ expression by recruitment of co-activator P300 to the P21 promoter and β-globin locus (Bai & Merchant, 2000; Petrovic et al, 2009; Woo et al, 2011). To distinguish between these different roles we identified the ZBP-89 partner proteins and their role in γ-globin regulation. We used tagged versions of ZBP-89 in two different cell lines, erythroid leukemia cell line MEL (erythroid leukemia cell line) and K562 (embryonic/fetal type cell lines). This analysis showed that ZBP-89 is involved in regulation of β-globin like genes differentially via recruitment of different complexes. ZBP-89 interacts with co-activator like proteins P400/Trrap, the mediator complex and the SWI/SNF complex in K562 cell line, while in MEL cells; it interacts with NuRD/MeCP1 complex, a co-repressor and known transcription factors which are involved in γ-globin repression like BCL11a. Importantly ZBP-89 appears to be involved in β-globin like gene regulation during development. We identified genome wide ZBP-89 binding sites during different stages of development (human fetal and adult stages) by ChIP-Seq. These results show that ZBP-89 binds to HS2 and HS3 both in fetal and adult stages. Furthermore, there are some weaker binding sites 4 kb upstream and 3.7 kb downstream of γ-globin genes and 1.2 kb downstream of β-globin gene. Unlike other transcription factors, which bind to all of the LCR hypersensitive sites, ZBP-89 binds exclusively to HS2 and HS3. The binding sites upstream and downstream of the γ-globin genes are more pronounced in the fetal stage when

compared to the adult stage. The site 3.7 kb downstream of the Aγ-globin gene is intergenic RNA transcript (sequence of BGL3) that also co-occupied by BCL11a and Ldb1 which suggests that it is involved in γ-globin silencing (Kiefer et al, 2011). BGL3 isn't transcribed in cells transcribing β-globin and is occupied by Ldb1 complex and corepressor ETO-2 and BCL11a. When γ-globin expression is reactivated, ETO-2 and BCL11a bound less to BGL3 and results in expression of γ-globin and BGL3 (Kiefer et al, 2011; Xu et al, 2010). We reduced the level of ZBP-89 protein using short hairpin RNA and lentiviral transduction. This experiment was carried out in human erythroid progenitor (HEP) cells taken from six different donors. ZBP-89 depletion induced γ-globin expression in HEP cells. HbF was elevated between 2 to 3 folds. Upon ZBP-89 depletion, the minority of cells showed high γ-globin expression, using immunohistochemistry (IHC). ZBP-89 knock down affects the expression of several other genes including BCL11a. It is very interesting to see that ZBP-89 knock down reduces BCL11a levels by 3.7 fold, while the expression of all embryonic and fetal globin genes is increased significantly upon ZBP-89 knock down. The expression of all the adult globin genes is reduced in the same cells. ZBP-89 knock down in adult HEP and the subsequent induction of γ-globin expression are accompanied by a partial restoration of the interaction between the LCR HSs and β-globin like genes to that seen in the fetal stage. For example upon ZBP-89 knock down in adult HEP cells increased the interaction of the γ-genes with HS1 and HS4 and decreased the HS1-β-globin gene interaction. Moreover, the 3.7 kb HBG1 downstream sequence interaction with HS1-4 increased again resembling the fetal β-globin locus conformation. Thus this analysis of interactions implies that activation of γ-globin gene expression is accompanied by DNA loops re-configuration back to the fetal liver stage.

ZBP-89 and co-factors

Transcription factors regulate gene expression by binding to enhancer elements to interact with proximal promoter elements. Transcription factors are often modulated by post-translational modification such as acetylation, methylation, sumoylation (Verger et al, 2003). SUMO conjugation catalyzes sumoylation by a series of enzymes including E1-activating enzyme, E2-conjugating enzyme and an E3-SUMO ligase (Johnson, 2004), while the desumoylation process is catalyzed by a SUMO- specific protease (Senp3). Sumoylation of ZBP-89 causes it to recruit repressor complexes such as NuRD or CoREST (Garcia-Dominguez & Reyes, 2009). Protein arginine N-methyltransferases (PRMTs) are a family of proteins that catalyze the methylation of arginine residues, a process of

post-translational modification found on both nuclear and cytoplasmic proteins. Depending on the methylated forms of arginine residues of target proteins, mammalian PRMTs are classified into two types. Asymmetrically dimethylated (aDMA) which termed the type I PRMTs, is including PRMT1, 3, 4, 6 and 8. The methylated arginine residues by Prmt1 are mostly located within glycine-arginine-rich (GAR) domain (Bedford & Clarke, 2009). The second category causes symmetrical dimethylation (sDMA), the type II PRMTs, which include PRMT5, 7 and 9 (Bedford & Richard, 2005). The Chromatin target of Prmt1 (Chtop), also known as Friend of Prmt1 (FOP), is a DNA binding protein that can regulate gene expression like the γ-globin gene (van Dijk et al, 2010). 5FMC, which is a desumoylating complex desumoylates ZBP-89 resulting in a higher level of RNA pol II on ZBP-89 target genes. 5FMC is composed of five proteins: Proline-glutamate and leucine rich protein (Pelp1) as the core component, the WD repeat domain 18 (Wdr18), SUMO1/sentrin/SMT3 specific peptidase 3 (Senp3), Testis expressed 10 protein (Tex10) and Las1 like protein (Las1L). Upon Prmt1 knock down, the 5FMC complex is not detected in a Chtop pull down. This reasons that Chtop recruits 5FMC in a methylation dependent manner by Prmt1. ZBP-89 purification in MEL cell revealed that ZBP-89 interacts with Chtop, Prmt1 and 5FMC components. Chtop depletion showed a better ZBP-89 purification relative to the control samples while the co-purification of 5FMC components was reduced to 30% of the control samples. These data suggest that Chtop can recruit 5FMC (Five Friends of Methylated Chtop) to ZBP-89 when Prmt1 methylates Chtop. A reduced level of Senp3 led to an increase of sumoylated ZBP-89 and demonstrates that Senp3 regulates sumoylation of ZBP-89. Different post-translational modification of ZBP-89 can probably change its function during development, a point that remains to be established. SILAC experiments are probably most suitable to confirm our hypothesis.

We postulate that different post-translational modifications of ZBP-89 result in alternative ZBP-89 complexes that mediate γ-globin activation or silencing during development by changing long-range interaction of the γ-globin genes with the LCR. In human adult cells, the most prominent interaction is between the LCR and β-globin promoter. KLF1 and Ldb1 complex occupy β-globin promoter and also involved in long-range interaction. It would be of great interest to test whether they act together or independently. For example one could imagine that ZBP-89 actively stimulates loop formation of the LCR to the γ- globin gene in the fetal liver (binding to activating factors), while it may function to inhibit long-range interactions in the adult stage but that the function of looping with the β-globin gene is taken over by factors such as KLF1 and Ldb1.

REFERENCES

Bai L, Merchant JL (2000) Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells. *J Biol Chem* **275:** 30725-30733

Bedford MT, Clarke SG (2009) Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33:** 1-13

Bedford MT, Richard S (2005) Arginine methylation an emerging regulator of protein function. *Mol Cell* **18:** 263-272

Constantoulakis P, Knitter G, Stamatoyannopoulos G (1989) On the induction of fetal hemoglobin by butyrates: in vivo and in vitro studies with sodium butyrate and comparison of combination treatments with 5-AzaC and AraC. *Blood* **74:** 1963-1971

Garcia-Dominguez M, Reyes JC (2009) SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim Biophys Acta* **1789:** 451-459

Johnson ES (2004) Protein modification by SUMO. *Annu Rev Biochem* **73:** 355-382

Kiefer CM, Lee J, Hou C, Dale RK, Lee YT, Meier ER, Miller JL, Dean A (2011) Distinct Ldb1/NLI complexes orchestrate gamma-globin repression and reactivation through ETO2 in human adult erythroid cells. *Blood* **118:** 6200-6208

Ohneda K, Ohmori S, Ishijima Y, Nakano M, Yamamoto M (2009) Characterization of a functional ZBP-89 binding site that mediates Gata1 gene expression during hematopoietic development. *J Biol Chem* **284:** 30187-30199

Perrine SP, Miller BA, Faller DV, Cohen RA, Vichinsky EP, Hurst D, Lubin BH, Papayannopoulou T (1989) Sodium butyrate enhances fetal globin gene expression in erythroid progenitors of patients with Hb SS and beta thalassemia. *Blood* **74:** 454-459

Perrine SP, Miller BA, Greene MF, Cohen RA, Cook N, Shackleton C, Faller DV (1987) Butryic acid analogues augment gamma globin gene expression in neonatal erythroid progenitors. *Biochem Biophys Res Commun* **148:** 694-700

Petrovic I, Kovacevic-Grujicic N, Stevanovic M (2009) ZBP-89 and Sp3 down-regulate while NF-Y up-regulates SOX18 promoter activity in HeLa cells. *Mol Biol Rep* **36:** 993-1000

van Dijk TB, Gillemans N, Pourfarzad F, van Lom K, von Lindern M, Grosveld F, Philipsen S (2010) Fetal globin expression is regulated by Friend of Prmt1. *Blood* **116:** 4349-4352

Verger A, Perdomo J, Crossley M (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO Rep* **4:** 137-142

Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR (2007) Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *Embo J* **26:** 2041-2051

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

Woo AJ, Moran TB, Schindler YL, Choe SK, Langer NB, Sullivan MR, Fujiwara Y, Paw BH, Cantor AB (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

Wu Y, Zhang X, Salmon M, Zehner ZE (2007) The zinc finger repressor, ZBP-89, recruits histone deacetylase 1 to repress vimentin gene expression. *Genes Cells* **12:** 905-918

Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH (2010) Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev* **24:** 783-798

Summary

Samenvatting

SUMMARY

γ-globin gene expression is restricted to human fetal life and is replaced by β-globin gene expression after birth. Re-activation of γ-globin gene expression in adult ameliorates the devastating effect of β-thalassemia and sickle cell anemia, which are the most important single gene disorders. Hence understanding the molecular mechanism of γ-globin gene suppression is of great importance to find new therapeutic strategies.

In chapter 2 we describe the development of an innovative and unbiased proteomics approach, Targeted Chromatin Purification (TChP), to identify protein complexes responsible for γ-globin gene. One of the protein factors we identified was ZBP-89.

Chapter 3 describes the functional role of ZBP-89 in γ-globin regulation. The protein partners of ZBP-89 were identified using proteomics and Mass Spectrometry in two different cell lines, the binding of ZBP-89 in the human genome was studied using ChIP sequencing and the effect of reduced ZBP-89 expression by shRNA Lentiviral mediated knock down (shRNA KD) was measured using RNA sequencing. The role of ZBP-89 in spatial organization of globin locus during development and differentiation was also studied using shRNA KD followed by Chromosome Conformation Capture Sequencing. These studies show that ZBP-89 interacts with NuRD repressor complexes as well as other proteins known to be involved in γ-globin regulation in MEL cells, whereas it interacts with co-activator like proteins P400/Trrap, the mediator complex and the SWI/SNF complex in K562 cells. The reduced expression of ZBP-89 leads to increased embryonic and fetal globin gene expression in human adult erythroid progenitors while the expression of the adult stage globins is decreased. Upon ZBP-89 knockdown in HEP cells the interaction of the γ-genes with HS1 and HS4 is increased while the interaction between the HS1 and the β-globin gene is decreased. Interestingly the interaction with a sequence 3.7 kb HBG1 downstream of the γ-globin genes with HS1-4 increased resembling the conformation of the β-globin locus during the fetal stage.

Chapter 4 describes Chtop as a DNA binding protein regulating transcription. Chtop-interacting proteins were identified by a biotinylation-proteomics approach. Chtop methylation by Prmt1 results in recruitment of Five Friends of Methylated Chtop (5FMC) to Chtop. 5FMC is a nuclear complex that consists of co-activator Pelp1, the Sumo-specific protease Senp3, Wdr18, Tex10 and

Las1L. Recruitment of 5FMC to ZBP-89 affects transactivation potential by increasing Polymerase II level on the promoters of ZBP-89 target genes.

Samenvatting

γ-Globine genexpressie is beperkt tot enkel het foetale deel van het menselijk leven en wordt na de geboorte vervangen door expressie van het β-globine gen. Re-activatie van γ-globine gen expressie na de geboorte vermindert de desastreuze gevolgen van thalassemie en sikkelcel anemie, enkele van de belangrijkste genetische ziektes van één enkel gen. Het is daarom belangrijk het moleculaire mechanisme achter de onderdrukking van γ-globine genexpressie te begrijpen, om zo to nieuwe therapeutische strategieën te komen.

Hoofdstuk 2 beschrijft de identificatie van de eiwitcomplexen verantwoordelijk voor de onderdrukking van γ-globine expressie, dmv een door ons ontwikkelde innovatieve en objectieve proteomics benadering (genaamd Targeted Chromatin Purification, TChP). Hierbij worden de eiwit factoren gebonden aan de γ-globine promoter en regulatoire regionen opgezuiverdn en geidentificeerd. Eén van deze eiwit factoren is ZBP-89.

Hoofdstuk 3 beschrijft de functionele rol van het ZBP-89 eiwit bij de onderdrukking van γ-globine expressie. De eiwit partners van ZBP-89 zijn bestudeerd in twee verschillende cellijnen met behulp van proteomics en massaspectrometrie, de bindingsplaatsen van ZBP-89 in het humane genome bepaald met ChIP sequencing en het effect bestudeerd van verminderde ZBP-89 expressie door middel van shRNA lentivirale knockdown (shRNA KD) gevolgd door RNA sequencing. Ook is de rol van ZPB-89 onderzocht bij de ruimtelijke organisatie van het globine locus tijdens de ontwikkeling door gebruik te maken van shRNA KD gevolgd door Chromosome Conformation Capture sequencing. Deze studies laten zien dat ZBP-89 een interactie aangaat met NuRD repressor complexen, alsmede met andere eiwitten betrokken bij de onderdrukking van γ-globine expressie. Anderzijds gaat het ook interacties aan met co-activator eiwtten as P400/Trrap, het Mediator complex and het SWI/SNF complex in de K562 cellijn. Verminderde expressie van ZBP-89 leidt to een verhoging van embryonale en foetale globine expressie in volwassen rode erythroïde voorlopercellen, terwijl de de expressie van globine genen die alleen na de geboorte tot expressie komen verlaagd wordt. ZPB-89 reductie in volwassen erythroïde voorlopercellen leidde tot een verhoogde interactie van de γ-globine genen met HS1 en HS4, en verlaagde de HS1-β-globine gen interactie. Tevens was er sprake van verhoogde interactie van de sequentie HBG1 die 3.7 kb naast de γ-globine genen ligt met HS1-4, wat wederom gelijkenis vertoont met de foetale β-globine locus conformatie.

Hoofdstuk 4 beschrijft Chtop als een DNA-bindend eiwit dat transcriptie reguleert. Eiwitten die een interactie aangaan met Chtop werden geïdentificeerd door een biotinylation-proteomics strategie.

De methylatie van Chtop door Prmt1 resulteert in het aantrekken van de Five Friends of Methylated Chtop (5FMC) door Chtop. 5FCM is een nucleair complex dat bestaat uit de co-activator Pelp1, de Sumo-specifieke protease Senp3, Wdr18, Tex10 en Las1L. Rekrutering van 5FMC door ZBP-89 beïnvloedt het transactivatie potentieel door het verhogen van Polymerase II niveaus op de promotoren van ZBP-89 doelwitgenen.

Curriculum Vitae

Personal Details

Name: Ali Aghajanirefah

Birth place: Tabriz, Iran

Birth date: 06-07-1980

Education and Research experience

1998- 2003	B.Sc. Biology, Major of Microbiology, Department of Basic Science, Karaj Azad University, Karaj, Iran.
2003- 2006	M.Sc. Molecular Genetics, Tehran Azad University, Science and Research Branch, Tehran, Iran.
2006- 2008	Research Assistant, Genetic Research Center, Social Welfare and Rehabilitation Sciences University, Tehran, Iran.
2008- 2012	PhD. Department of Cell Biology, Medical Genetic Cluster, Erasmus University Medical Center, Rotterdam, The Netherlands (Prof. dr. F.G. Grosveld).
.	
2012- Present	Postdoctoral fellow, Department of Molecular Biology of the Nijmegen Center for Molecular Life Science (NCMLS), Nijmegen, The Netherlands (Prof. dr. ir. H.G. Stunnenberg).

List of publications:

A Point Mutation AT the Calreticulin Gene Core Promoter conserved sequence in a Case of Schizophrenia.
Ali Aghajani, Alireza Rahimi, Farbod Fadai, Ahmad Ebrahimi, Hossein Najmabadi, Mina Ohadi. (Am J Med Genet B Neuropsychiatr Genet. 2006 Apr 5; 141B (3):294-5).

Attention-Deficit/Hyperactivity Disorder (ADHD) Association with the DAT1 Core Promoter -67 T-Allele.
Ohadi Mina, Shirazi Elham, Tehranidoost Mehdi, Keikhaee Mohammad R, Ehssani Sima, Ali Aghajani, Moghimi Narges, Najmabadi Hossein. (Brain Res. 2006 Jul 26; 1101(1):1-4. Epub 2006 Jun 19).

A mutation in the calreticulin gene promoter in a family case of schizoaffective disorder leads to its aberrant transcriptional activation.
Ana Nunes, Mina Ohadi, Alireza Rahimi, Ali Aghajani, Hossein Najmabadi, Salvador Soriano. (Brain Res. 2008 Nov 6; 1239:36-41. Epub 2008 Sep 4).

Carrier frequency of SMA by quantitative analysis of the SMN1 deletion in the Iranian population.
Hasanzad M, Azad M, Kahrizi K, Saffar BS, Nafisi S, Keyhanidoust Z, Azimian M, Refah AA, Also E, Urtizberea JA, Tizzano EF, Najmabadi H. (Eur J Neurol. 2010 Jan; 17(1):160-2. Epub 2009 Jun 15).

Diagnostic values of GHSR DNA methylation pattern in breast cancer.
S Botla, A Gholami, M Malekpour, E Moskalev, M Fallah, A Aghajani, I Bondar, R Omranipour, F Malekpour, A Mohajeri, A.J.Babadi, V.V. Bubnov, H Najmabadi, J.D. Hoheisel, Y Riazalhosseini. (Breast Cancer Res Treat. 2012 Aug 17. [Epub ahead of print]).

Five Friends of Methylated Chtop, a complex linking arginine methylation to desumoylation.
Pavlos F, Gillemans N, Aghajanirefah A ,Pourfarzad F, Demmers J, Esteghamat F, Vadlamudi RK, Grosveld F, Philipsen S, Van Dijk TB. (Mol Cell Proteomics. 2012 Aug 7. [Epub ahead of print).

In vivo isolation of suppressed γ-globin chromatin hub via an unbiased proteomic approach.
Pourfarzad F, Aghajanirefah A, de Boer E, Ten Have S, KheradmandKia S, Stadhouders R, Gillemans N, Linderen M, Demmers J, Philipsen S, Grosveld F. (Submitted).

The role of the ZBP89 protein complex in globin gene regulation.
Aghajanirefah A, Pourfarzad F, Stadhouders R, Gillemans N, Wendt K, Papadopoulos P, Demmers J, Van Ijcken W, Soler E, Grosveld F. (Manuscript in preparation).

PhD Portfolio

Summary of PhD training

Department of Cell Biology Promoter and supervisor: Prof.dr.F.G. Grosveld	October 2008- November 2012

PhD training

General course	Year
Radiation protection certificate	2008
Browsing genes and genome with ensemble	2009
Technology facilities course	2009
EuTRACC Proteomics course	2010
Epigenetics and chromatin	2010
Safety working in the laboratory	2010
From Development to Disease	2011
Literature course	2011
Biomedical English writing course	2011
Transgenesis, gene Targeting and Gene Therapy	2011

Seminars and workshops

Monday morning meetings	2008-2012
Erasmus lectures in cell biology and development	2008-2012
MGC PhD workshop (poster presentation)	2010
The 20th MGC-Symposium, Leiden, the Netherlands	2010
Workshop on InDesign CS5	2011
MGC PhD workshop (oral presentation)	2011

Presentations

Monday morning meetings	2008-2012
Work discussion	2008-2012

International Conferences

EUrythron annual meeting, Lisbon, Portugal	2009
17th Hemoglobin Switching (poster presentation), Oxford, UK	2010
18th Hemoglobin Switching (poster presentation), Asilomar, CA, USA	2012

Abbreviations

3C	Chromosome Conformation Capture
4C	Circular Chromosome Conformation Capture
5-Aza	5-Azacytidine
5-FMC	Five Friends of Methylated Chtop
ACH	Active Chromatin Hub
aDMA	asymmetric arginine dimethylation
Apex1	Apurinic/apyrimidinic endonuclease 1
Bio tag	Biotin tag
BirA	Biotin ligase
C/EBP	CCAAT/Enhancer-Binding Protein
CBF	CCAAT-Binding Factor
CBP	CREB Binding Protein
CDC5L	Cell Division Cycle 5-like
cDNA	complementary DNA
CDP	CCAAT Displacement Protein
CH	Chromatin Hub
ChIP	Chromatin Immunoprecipitation
Chtop	Chromatin Target of Prmt1
Ctnnbl1	Beta catenin like 1 protein
DNMT	DNA methyltransferase
DNMT3A	DNA Methyltransferase
DR-1	Direct Repeats
DRED	Direct Repeat Erythroid Definitive
EKLF	Erythroid Kruppel like Factor
EPO	Erythropoietin
FANC	Fanconi Anemia Complementation group
FL	Fetal Liver
FOP	Friend of Prmt1
H3ac	Histone 3 acetylated
H3K4me3	Histone 3 lysine 4 methylated
HA tag	Hemagglutinin tag
HAT	Histone Acetyltransferase
Hb	Hemoglobin
HbA	Adult Hemoglobin
HbF	Fetal Hemoglobin
HbS	Sickle Hemoglobin
HDAC	Histone deacetylases
HEP	Human Erythroid Progenitor cells
HIPK2	Homeodomain Interacting Protein Kinase 2
HPFH	Hereditary Persistence Fetal Hemoglobin
HPLC	High Performance Liquid Chromatography
HS DNase I	Hypersensitive Sites
HU	Hydroxyurea
IHC	Immunohistochemistry

Abbreviations

IP	Immunoprecipitation
LCR	Locus Control Region
MBD2	Methyl-CpG Binding Domain 2 protein
MeCP1	Methyl-CpG Binding Protein Repressor complex 1
MEL	Murine Erythroleukemia
MS	Mass Spectrometry
NAP	Nucleosome Assembly Protein
NuRD	Nucleosome Remodeling and Deacetylating
Pelp1	Proline, Glutamate and Leucine rich protein 1
PICh	Proteomics of Isolated Chromatin segments
PTM	Post-translational Modification
PYR	PYR complex
QTL	Quantitative Trait Loci
RB	Red Blood Cells
RT-QPCR	Reverse transcription quantitative PCR
SCD	Sickle Cell Disease
SCF	Stem Cell Factor
SENP	SUMO/sentrin specific peptidase
shRNA	short hairpin RNA
SUMO	Small Ubiquitin – like MOdifier
TChP	Target Chromatin Purification
TetO	Tetracycline Operator sequence
TetO-γ	TetO-modified γ-globin minilocus
TetR	Tetracycline Repressor protein
TetR3T	Triple Tag TetR protein
Tex10	Testis Expressed protein 10
TRRAP	Transformation/transcription domain-associated protein
TF	Transcription Factor
Wdr18	WD repeat domain 18
Zbp	Zinc finger binding protein
Yb-1	Y-box-binding protein 1

Acknowledgment

Completing a PhD is a marathon event, and it would not have been possible to finish this adventurous journey without the guidance, assistance and support of many people within the past four years.

Foremost, I would like to express my sincere appreciation to my supervisor, Prof.dr. Frank Grosveld. Dear Frank, I very much appreciate you giving me the opportunity to join your lab. I consider myself very fortunate for being able to work with a very considerate and encouraging professor like you. Your management, attention to scientific details and support have set an example I hope to match some day. I am very grateful for your patient guidance and all that I have learned from you.

I acknowledge my gratitude to my small committee Prof.dr. Sjaak Philipsen, Prof.dr. Dies Meijer, and Prof.dr. Joost Gribnau for reading my thesis and your helpful comments.

I would like to thank the other members of my committee Prof.dr. Ruud Delwel, Prof.dr. Marieke Von Lindern and Dr. Niels Galjart. I am thankful for them accepting to be the members of my committee.

A special thanks of mine goes to my friend and colleague, Farzin, who taught me different molecular biology techniques and helped me in finding the way at the beginning of this journey.

Dear Farzin and Parham, my paranifms, thank you very much for your help and guidance during this period. I am very happy to have you as one of my best friends. Parham, I wish you the best for the rest of your PhD.

Special thanks to Ali Imam (also for translating into English trochaic tetrameter), Eric, Charlotte, Thamar, Robert-Jan, Ralph (also for samenvatting), Tokameh, Nynke and Petros (Kolovos) for their help throughout my time in Cell Biology.

Many thanks go to the rest of members (past and present) of lab 702; Anita, Xiao, Mary, Andrea, Guillaume, Boet, Petros, Athina and Sanja. We spent happy times together during my stay in the lab.

I also want to thank Jeroen, Dick, Karel (Proteomics center), Rutger, Mirjam, Christel, Zeliha and Wilfred (Biomics center) for analyzing my samples.

I would also like to thank my dear colleagues of lab 710; Dubi, Rick, Alex, Rien, Rik, Ruud and Mariette. I had a wonderful time with you.

Also many thanks to Ton, Pim, Leo and Sjozef, Mario (from PC-Support), Marieke, Jasperine and Bep (Secretaries), Melle, Arthur, Leo, Marjoleine, Job (from P&O) and kitchen ladies for their help when I needed.

I am very much grateful to Mehrnaz, Saeid, Yaser, Alireza, Parisa, Raheleh, Kiavash, Farhad, Sima and Babak for your friendship and help. We have spent together a very memorable and pleasant time during the last four years.

Thanks to my former and present colleagues of labs 706 and 716; Ernie, Raymond, Erik, Umut, Maaike, Tiago, Debbie, Maureen, Miyata, Pavlos, Sahar, Sylvia, Jun, Rejane, Harald, Eva, Laura, Carla, Maria, Divine and Ana.

Thanks to all my colleagues in Erasmus MC for the good times we had together; Noorie, Siska, Ashok, Olaf, Reinier, Polynikis, Alvin, Dorota, Thomas, Kris, Nesrin, Fanny, Alex, Lalini, Joshua, Ilona, Dave, Zhe, Jeffry, Jessica, Peng, Francesca, Natasja, Kerstin, Joost, Nils, Arjan, Ozge, Karl, Gerben, Robbert, Gert, Catherine and many other people not mentioned here.

I express my sincere thanks to my teachers and colleagues in Iran, Prof. Najmabadi and Dr. Ohadi and all my colleagues in Social Welfare and Rehabilitation Sciences University.

I deeply appreciate the Iranian community; Elnaz, Omid, Ronak, Azadeh, Haleh, Amir, Farshid, Abbas, Mahban, Azadeh F.

Finally, I take this opportunity to express the profound gratitude from the bottom of my heart to my beloved family for their continuous support, countless efforts and endless contributions.

I really appreciate Sevda's family for their support and in particular, I must acknowledge my wife and best friend, Sevda, without whose love and encouragement, I would not have finished this thesis.

Ali