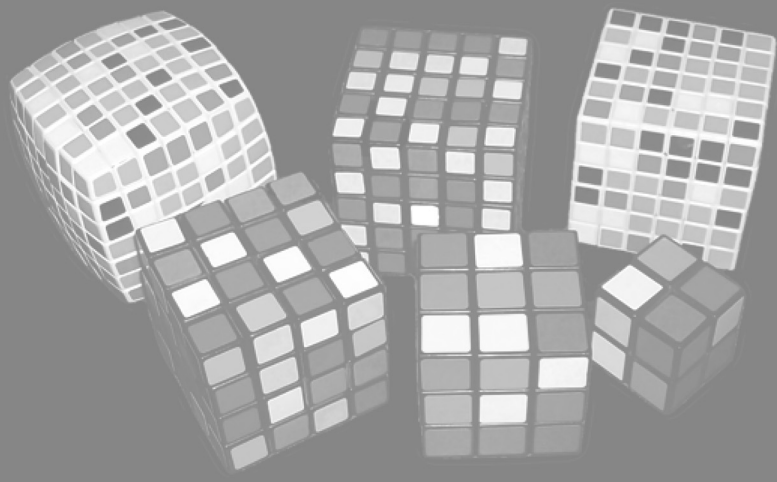
Studies on Human γ-globin Gene Regulation

Transcription Factors in a Rubik's Cube



Petros Papadopoulos

Printed in CPI-Wöhrmann Print Service, Zutphen.

ISBN: 978-90-8570-433-1

The studies presented in this thesis manuscript were performed in the Department of Cell Biology of the Erasmus MC in Rotterdam, The Netherlands. The Department is member of the Medisch Genetisch Centrum Zuid-West Nederland (MGC). The financiation was partially covered by the NWO (“Activation of the globin locus”) and NIH (“Molecular basis of human gamma globin gene activation”, HL 073455).

Studies on Human γ-globin Gene Regulation
Transcription Factors in a Rubik's Cube

Studies van menselijke γ-globine genregulatie
Transcriptiefactoren in een Rubik's kubus

Proefschrift

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 17 December 2009 om 11.30 uur

door

Petros Papadopoulos

geboren te Athene, Griekenland

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Prof.dr. J.N.J. Philipsen Prof.dr. D.N. Meijer Dr.ir. N.J. Galjart

....μετὰ μεγάλων δὲ σημείων καὶ οὐ δή τοι ἁμάρτυρόν γε τὴν δύναμιν παρασχόμενοι τοῖς τε νῦν καὶ τοῖς ἔπειτα θαυμασθησόμεθα, καὶ οὐδὲν προσδεόμενοι οὔτε Ὁμήρου ἐπαινέτου οὔτε ὅστις ἔπεσι μὲν τὸ αὐτίκα τέρψει, τῶν δ᾽ ἔργων τὴν ὑπόνοιαν ἡ ἀλήθεια βλάψει, ἀλλὰ πᾶσαν μὲν θάλασσαν καὶ γῆν ἐσβατὸν τῇ ἡμετέρα τόλμη καταναγκάσαντες γενέσθαι, πανταχοῦ δὲ μνημεῖα κακῶν τε κἀγαθῶν ἀίδεια ξυγκατοικίσαντες....

Θουκυδίδης (Περικλέους Επιτάφιος, 431 π.Χ.).

Στους γονείς μου.

CONTENTS

INTRODUCTION

Blood and hemoglobin

Blood consists of erythrocytes, leukocytes, platelets and nutrients that travel throughout the whole body. Blood cells are essential for life functions, such as transport of gas (O_2, CO_2) and they form part of the immune defense system. Examination of the blood is obviously crucial for the diagnosis and management of hematological diseases, but a simple blood smear for microscopic analysis or quantitative measures of cells in a blood cell counter can also very rapidly provide information about the more general physiological condition of the body. One of the main parameters is hemoglobin content of the erythrocytes, which is measured by a combination of different values and provides the major diagnostic tool for anemia. The interest to study hemoglobin has increased throughout the years and still is an important part of current research providing us with fundamental information about protein structure biology and regulation of gene expression.

The molecule of hemoglobin, the oxygen transporter of erythrocytes, is the best understood allosteric protein. It is a nearly spherical structure with a diameter of 55Å that consists of four polypeptide chains, two α-like chains and two β-like chains, packed together in a tetrahedral array (Fig. 1A). Each polypeptide contains a heme group with a single oxygen binding site. The heme groups are located in crevices near the exterior of the molecule providing the oxygen binding sites far apart from each other (Fig. 1A) (Perutz 1978; Fermi, Perutz et al. 1984).

The capacity of hemoglobin to bind oxygen depends on the presence of the non-polypeptide unit of the heme group (Fig. 1B). It consists of an organic part and an iron atom. The organic part, protoporphyrin, is made up of four pyrrole rings linked together by methine bridges to form a tetrapyrrole ring. The iron atom binds to the four nitrogens in the center of the ring and can additionally form two extra bonds on either side of the heme plane. When oxygen is bound, the iron atom of the corresponding form of hemoglobin, ferrihemoglobin, has the ferrous (+3) oxidative state (Stryer 1995).

Hemoglobin transports H^+ and CO_2 in addition to O_2. These molecules bind to spatially distinct sites that communicate with each other by means of complex conformational changes within the protein. Thus, O_2 binding at one heme group facilitates the binding of oxygen at the other heme groups on the same tetramer, and *vice versa*: the unloading of oxygen at one heme group facilitates the unloading of oxygen at the others. The cooperative binding of oxygen by hemoglobin enables

it to deliver 1,83 times as much oxygen under typical physiological conditions as it would, if the sites would be independent binding sites (Stryer 1995).

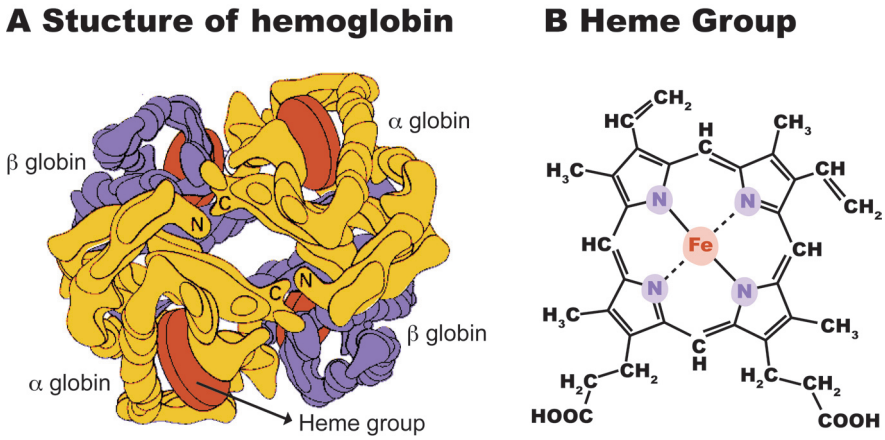


Figure 1
A. Structure of hemoglobin. **B.** Heme group.

Furthermore, the hemoglobin affinity for oxygen is reduced by increasing concentration of CO_2 and pH reduction. This competition occurs in tissues with high metabolic activities, such as contracting muscle, where considerable amounts of CO_2 and acid are produced. The presence of higher levels of CO_2 and H^+ in the capillaries of muscles promotes the release of oxygen from oxyhemoglobin. The reciprocal effect occurs in the alveolar capillaries of the lungs. The high concentration of oxygen unloads H^+ and CO_2 from hemoglobin. It is remarkable that the oxygen affinity of hemoglobin within red cells is lower than that of hemoglobin in free solution. This is due to a highly anionic organic phosphate, 2,3- bis-phosphoglycerate (BPG or DPG), that is present in the red cells at about the same molar concentration as hemoglobin. BPG, which binds only to deoxy-hemoglobin, lowers the affinity of hemoglobin for oxygen by a factor of 26, which is essential in enabling hemoglobin to unload oxygen in tissue capillaries (Benesch and Benesch 1969).

Another interesting feature of BPG is that it binds less strongly to fetal hemoglobin (HbF). Consequently, although the adult hemoglobin (HbA) affinity for oxygen is greater than that of HbF (Tyuma and Shimizu 1970), BPG decreases it. This is how HbF is oxygenated at the expense of HbA in the placenta (Fig. 2).

The binding of different molecules (O_2, CO_2) as well as their concentration,

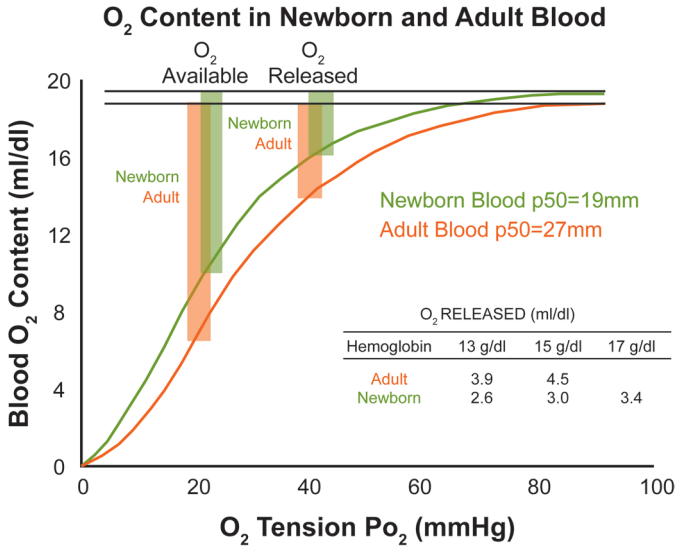


Figure 2

Oxygen equilibrium curves of adult and newborn blood based on figure 6-3 of Chapter 6 (Williams Hematology). The shift to the left of the oxygen equilibrium curve explains the less oxygen released in the newborn when pO_2 drops from the arterial to the venous blood due to the higher oxygen affinity of cord blood as compared to maternal blood.

cause structural changes that alter the light absorbance of hemoglobin (Fig. 3). Studies on expression of fluorescent proteins in red blood cells have reported variegated expression and autofluorescence, depending on the levels of expression of the fluorescent protein, especially for the ones with emission wavelength within the range of 500-600nm, as the Green Fluorescence Protein (GFP), which falls in the middle of the absorbance spectrum of hemoglobin (Fig. 3) (Spangrude, Cho et al. 2006; Swenson, Price et al. 2007).

Hemoglobin, as an allosteric protein, does not have fixed properties. Rather, its functional characteristics are regulated by specific molecules in its environment. That is why an imbalance in the production of hemoglobin or the factors regulating its functions has a tremendous impact on the red cells.

The course of erythropoiesis

Erythropoiesis, the process in which red cells are produced, occurs in two waves during embryogenesis, *i.e.* primitive and definitive. In human, primitive

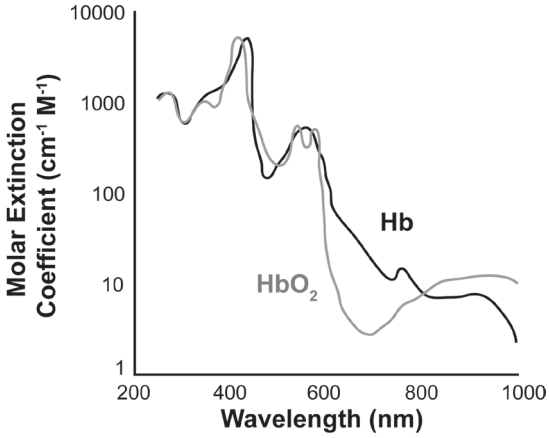


Figure 3
Absorbance spectra of hemoglobin (Hb) and oxyhemoglobin (HbO_2) (with permission of Scott Prahl, Oregon Medical Laser Center).

erythropoiesis starts in the yolk sac by day 18 of gestation, giving rise to primitive erythroid cells that enter the embryo proper at day 21 and circulate until approximately 12 weeks of gestation. The primitive erythroid cells are nucleated, they contain embryonic hemoglobins and they proliferate and terminally differentiate within the vascular network (Peschle, Mavilio et al. 1985).

As early as 4 weeks of gestation, definitive erythroid progenitors, *i.e.* burst forming units-erythroid (BFU-E), are found in the yolk sac and by 5 weeks in the fetal liver. After the 7th week of gestation there are no hematopoietic progenitors in the yolk sac. The fetal liver takes over as the main definitive hematopoietic tissue from the 7th to the 25th week of gestation. The bone marrow then takes the final lead in hematopoiesis (Fig. 4B). Definitive erythroid cells differentiate in proximity to macrophages and they extrude their nuclei before entering the bloodstream. Erythropoiesis has been thoroughly studied at the molecular level, and despite many questions remaining unanswered, the essential role of a number of transcription factors, such as Myb (Mucenski, McLain et al. 1991), Eklf (Nuez, Michalovich et al. 1995), Tal1 (Shivdasani, Mayer et al. 1995), Lmo2 (Warren, Colledge et al. 1994) and Gata1 (Fujiwara, Browne et al. 1996), has been unraveled through genetic studies.

In the mouse, primitive erythropoiesis starts in the yolk sac by 7.5*dpc*, and primitive erythroid cells are seen in circulation until 15*dpc*. Recently it was shown that in the mouse primitive circulating erythroid cells enucleate in the fetal liver at the end of their maturation process, and continue circulating until the end of their cycle (Isern, Fraser et al. 2008). Definitive erythropoiesis starts at 10.5*dpc* in the

fetal liver and later on during gestation it moves to the spleen and bone marrow (Fig. 4A). Definitive cells appear in the blood at 12.5*dpc* (Wong, Chung et al. 1985). They are generated from committed progenitors that go through a fixed number of divisions, cell size reduction, protein production, protein degradation, cell cycle arrest and enucleation (Fig. 5).

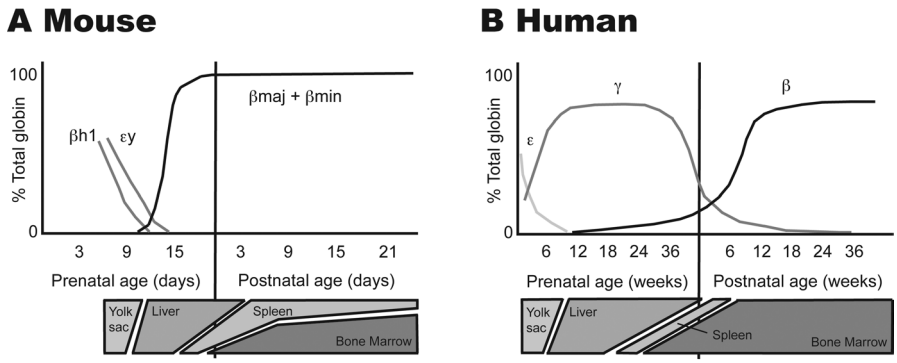


Figure 4

A. Mouse and **B.** human developmental expression of globin genes.

Hemoglobin synthesis and development

As mentioned above, hemoglobin consists of two α-like chains and two β-like chains packed together in a tetrahedral structure. The human α-globin cluster is located at chromosome 16 and contains three globin genes in 5'-3' direction, ζ-α2-α1. In the mouse, the conserved α-globin locus 5'-ζ-α1-α2-3' resides on chromosome 11. The β-globin locus in humans lies on chromosome 11 whereas in the mouse on chromosome 7. It consists of five functional globin genes in humans, placed in the same order as their developmental expression, 5'-ε-Gγ-Aγ-δ-β-3'. In the mouse, the β-globin locus genes are four and also positioned in the same order

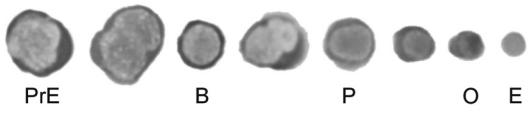


Figure 5

Erythroid cell differentiation of bone marrow cells from proerythroblast stage (PrE) to basophilic (B), polychromatic (P), orthochromatic (O) and finally erythrocyte (E) devoid of nucleus. Cells start accumulating hemoglobin at the polychromatic stage (taken from Gutiérrez *et al*, Exp. Hem 2005).

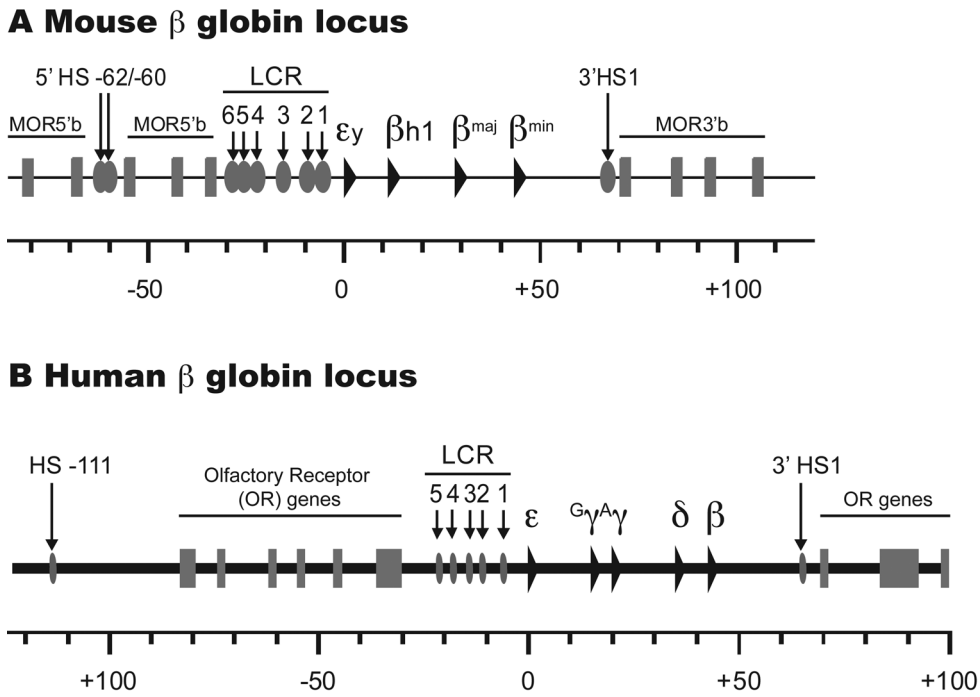


Figure 6

A. Map of the mouse and **B.** human beta (β) globin locus.

as they are expressed during mouse ontogeny, 5'-εy-βh1-βmaj-βmin-3' (Fig. 6).

Expression of globins in humans and mice corresponds with changes in sites of erythropoiesis. A major difference between them is that in human, two major switches occur, one from embryonic to fetal and the second from fetal to adult hemoglobin, while in mice only one switch occurs from embryonic to adult hemoglobin.

In human, during the first switch ζ and ε globins are replaced by α and γ globins as the liver replaces the yolk sac as the main hematopoietic site. HbF ($\alpha_2\gamma_2$) is the major hemoglobin in fetal life. The second switch occurs when γ globin is replaced by β globin. HbA ($\alpha_2\beta_2$) is detected in fetuses as young as 9 weeks and rises up to 21 weeks to reach 13% of the total hemoglobin. After 34 to 36 weeks HbA levels rise and HbF levels decrease, a process that progresses rapidly from the moment of birth (Fig. 4B).

In the mouse, yolk sac derived erythroid cells coexpress high levels of ζ and α globins with εy-βh1 and small amounts of βmaj and βmin (Brotherton, Chui et al. 1979; Wawrzyniak and Popp 1987; Whitelaw, Tsai et al. 1990). At 11.5*dpc,* when

the site of hematopoiesis changes to the fetal liver, expression of α1 and α2 globin genes as well as βmaj and βmin occurs (Fig. 4A).

Hemoglobinopathies

Hemoglobinopathies are a broad spectrum of diseases which can be divided in those resulting from an inherited structural alteration in one of the globin chains, like ***sickle cell anemia***, and those resulting from inherited defects in the rate of synthesis of one or more globin chains, the ***thalassemias***. The latter results in imbalanced globin chain production, ineffective erythropoiesis, hemolysis and a variable degree of anemia.

The sickle hemoglobin is a mutant hemoglobin in which valine, the sixth amino acid of the β-globin chain, has been substituted by glutamic acid. As a result, the mutant hemoglobin polymerizes under low oxygen conditions and becomes poorly soluble. Red cells in turn become rigid and distorted. This leads to capillary vessel occlusion and local necrosis of tissue. Sickle cell disease occurs in homozygosity for the sickle cell mutation or as a compound heterozygote for sickle hemoglobin and β-thalassemia (hemoglobin C) and some less common β-globin mutations. The disease is characterized by hemolytic anemia and different types of crisis.

There are many different types of thalassemias depending on the globin chain(s) affected (α-, β-, γ-, δ-, δβ-, εγδβ-) that lead to defective hemoglobin production. In fact, in these cases the damage to the red cells or their precursors comes from the globin chains that are produced in relative excess, like in the case of α-thalassemia in which a reduced rate of α-chain synthesis results in excess of γ chains in fetal life and γ_4 tetramers, or hemoglobin Bart's. In adult life, deficiency of α-chain synthesis results in excess of β chains and β_4 tetramers, or hemoglobin H. The molecular basis of thalassemias falls into deletional and non-deletional mutations and is extremely diverse. An open access database, the HbVar database, has been created with details of the mutations and deletions worldwide, leading to hemoglobin variants, all types of thalassemia and other hemoglobinopathies (***http://globin.bx.psu.edu/hbvar/menu.html***).

A heterogeneous group of conditions is the Hereditary Persistence of Fetal Hemoglobin (HPFH), which is characterized by persistent fetal hemoglobin (Pauling, Itano et al. 1949; Conley 1980). They are divided in deletion and non-deletion forms with the latter ones further classified into mutations that map within

the β-globin gene cluster (Collins, Stoeckert et al. 1984; Giglioni, Casini et al. 1984; Gelinas, Endlich et al. 1985) and those that segregate independently. They are also classified according to the population in which they occur, for example, the Greek HPFH, the British, Italian, Chinese and others. The distribution of the fetal hemoglobin in HPFH cases can be pancellular (*e.g*. deletion type Indian HPFH) or heterocellular (*e.g.* non-deletional type Atlanta HPFH).

There are also cases of HPFH characterized by low levels of heterocellular fetal hemoglobin. A small proportion of individuals with this type of HPFH was found to have increased amounts of fetal hemoglobin and F cells, which are the cells expressing fetal hemoglobin. Although in the beginning it was called Swiss HPFH, later studies found it in every racial group and some evidence suggested X-linked genetic determination of the number of F cells, located at Xp22.2 (Hebbel 1992; Briehl and Nikolopoulou 1993), even though this was not the case for all the low fetal hemoglobin HPFH forms (Thevenin, Crandall et al. 1997; Parise and Telen 2003).

Strong evidence of genetic links have been reported recently in support to the notion that specific DNA polymorphism (SNPs) are related to the observed variation in fetal hemoglobin levels among individuals (Menzel, Garner et al. 2007; Thein, Menzel et al. 2007; Lettre, Sankaran et al. 2008; Uda, Galanello et al. 2008). The genomic sequences of these SNPs were mapped to the gene BCL11A on chromosome 2, the intergenic region of HBS1L-MYB on chromosome 6 and upstream of the Gγ-globin in the β-globin locus on chromosome 11. MYB has been studied widely for its role in hematopoiesis, HBS1L is a poorly characterized gene with unknown biological functions and BCL11A that is expressed in erythrocyte precursors and implicated in lymphoid malignancies (Satterwhite, Sonoki et al. 2001; Liu, Keller et al. 2003) was recently shown to be a regulator of fetal hemoglobin in loss of function studies in human (Sankaran, Menne et al. 2008) and mouse (Sankaran, Xu et al. 2009). These results further support the importance of identifying mutated forms of transcription factors present in groups of individuals with specific phenotypes and hemoglobin indices in their erythrocytes (Singleton, Burton et al. 2008).

Therapeutic approaches

Currently, blood transfusions and iron chelation (deferoxamine) are the main therapeutic procedures for thalassemias. Bone marrow transplantation is also a solution but the outcome depends on the age and clinical condition of the individual and there are also potential post-transplant complications such as infections or graft-versus-host disease. There has been a considerable number of thalassemic and sickle cell patients (Lucarelli, Giardini et al. 1995; Di Bartolomeo P 1997) who underwent bone marrow transplantation and overcame the disease but one out of three developed acute or chronic graft-versus-host disease of mild to severe grades (Gaziev, Polchi et al. 1997; Bernaudin, Socie et al. 2007).

The observation that patients recovering from cytotoxic drug therapy or other periods of erythroid expansion may reactivate fetal hemoglobin synthesis has led to the use of many agents that increase the production of HbF in order to decrease the probability of *in vivo* sickling and reduce ineffective erythropoiesis. Such agents include erythropoietin, cytotoxic drugs (5-Azacytidine, Hydroxyurea) and butyrate analogs. All of them have been used in clinical trials (Olivieri and Weatherall 1998; Swank and Stamatoyannopoulos 1998; Weatherall 2003) as potential inducers of HbF as well as in combinations in order to test for optimal doses and additive or synergistic effects of the agents. Even though there has been a considerable advance in understanding how these agents work biochemically and new treatments have been used with some success in patients, overall progress has been slow due to the fact that the wide range of mutations or deletions of the β-globin gene cluster resulting in thalassemia are not equally susceptible to this type of approach.

Hydroxyurea (HU) has been used most successfully in patients with sickle cell anemia since it results in amelioration of crises. Administration of HU increases HbF hemoglobin, but there are probably other reasons for the success of the treatment (Charache, Terrin et al. 1995), possibly the reduction of the white blood cell count, changes in red cell rheology and indirect vascular effects.

Many pharmacological agents apart from hydroxyurea and azacytidine have been tested for their capability to reactivate fetal hemoglobin (HbF) in reporter assays in human cell lines and human primary progenitor cells (Haley, Smith et al. 2003) but the maximal γ-globin induction was always compromised by cytotoxicity. Histone deacetylase (HDAC) inhibitors have been of particular interest

in respect to γ-globin induction and especially a class of short-chain fatty acids (SCFAs) derivatives that have been implicated in the displacement of repressor complexes from the promoter of γ-globin and induce transcriptional activation of the gene (Mankidy, Faller et al. 2006). These are now entering clinical trials. Still, the molecular mechanisms underlying the γ-globin transcriptional activation and silencing are not yet uncovered.

Bone marrow transplantation appears to be a promising therapeutic approach, especially through the major advances of ES-cell-based therapy and the establishment of genetically identical ES-like cells by somatic nuclear transfer (SCNT) using donor cells from the patient (Jaenisch 2004). Human and mouse fibroblasts were reprogrammed *in vitro* into pluripotent stem cell-like cells (iPS) through retroviral transduction of combinations of transcription factors (OCT4, KLF4, SOX2, c-MYC) (Yu, Vodyanik et al. 2007; Hanna, Saha et al. 2009). Using the sickle cell anemia mouse model, iPS cells were corrected for the β^{sickle} allele by homologous recombination and then transplanted to irradiated sickle mice. This resulted in the correction of the disease phenotype (Hanna, Wernig et al. 2007). The drawback of applying ES or iPS cell based therapy to humans is mainly the possibility of developing malignancies that can not be excluded unless safer methods of gene delivery, retroviral vector-free, guarantee the efficient expression of the reprogramming factors in a controlable system (Hanna, Markoulaki et al. 2008), as well as developing alternatives for the oncogenes currently required for reprogramming. Gene therapy using the donor's bone marrow remains as another possibility particularly now the first patient has been treated successfully. This patient is still blood transfusion independent two years after therapy and has not developed any type of leukemia even though cells of one particular integration site of the lentiviral vector constitute the majority of the cells in the blood (Leboulch personal communication). More patients will be treated in the near future.

Regulation of γ-globin gene expression

Developmental analysis of the expression pattern of the human globin genes has been extensively studied in the mouse as a mammalian model organism. Experiments using transgenic mice showed that flanking sequences 5' or 3' of the coding region of the γ- and β-globin genes are sufficient to drive correct

developmental expression similar to that of the mouse homologues βh1- and βmaj-globin genes (Magram, Chada et al. 1985; Townes, Lingrel et al. 1985; Chada, Magram et al. 1986). However, the expression of the transgenes was highly variable, not correlated with the copy number of the transgene in the mouse genome and in the case of the β-globin gene, the level of expression per copy was always lower than that of the endogenous mouse homologue.

Studies in patients with thalassemia (Kioussis, Vanin et al. 1983; Taramelli, Kioussis et al. 1986) initiated the discovery of a region of regulatory sequences (Locus Control Region or LCR), 6-18Kb upstream of the ε-globin gene and another region 20Kb downstream (3' Hypersensitivity Site 1 or 3'HS1) of the β-globin gene (Forrester, Takegawa et al. 1987; Grosveld, van Assendelft et al. 1987). The LCR contains a number of sites hypersensitive to DNAseI digestion, even more than hypersensitive sites surrounding the individual genes when they are actively expressed (Groudine, Kohwi-Shigematsu et al. 1983; Forrester, Thompson et al. 1986). The LCR confers high-level expression and position-independence to human globin transgenes in transgenic mice and transfected cultured erythroid cells (Blom van Assendelft, Hanscombe et al. 1989), which is essential for studies in globin gene regulation during development and upon different treatments.

Independent studies from different groups, using mouse models containing single globin genes cloned under the control of the regulatory elements of the LCR or short versions of it, put forward a model of transcriptional competition between the fetal γ- and the adult β-globin genes (Hanscombe, Vidal et al. 1989; Enver, Raich et al. 1990). Developmental switching of the expression patterns of the two globin genes occurred only when the two genes where jointly cloned under the control of the LCR. As was shown shortly after (Dillon and Grosveld 1991), the inclusion of the genomic sequences flanking the γ- and β-globin genes in transgenic mice produced a strikingly different result, pointing out that copy numbers of the transgene and complete genomic sequences of the HSs of the LCR are extremely important. Silencing of the γ-globin gene was independent of the presence of β-globin, revealing the contribution of local sequences surrounding the γ-globin gene to this process. The relative distance of the globin genes to the LCR provided clues to the complex expression patterns found in mice harbouring constructs with altered globin gene order within the transgene or harbouring partial globin locus constructs. Position in the β-globin locus is an important parameter (Hanscombe, Whyatt et al. 1991), which defines the frequency of interactions

between the LCR and the promoters of the genes that are transcribed during a particular developmental stage.

The picture of understanding interactions in the whole β-globin locus got clearer when mice were generated with a transgene covering 70Kb of the β-globin gene cluster (Strouboulis, Dillon et al. 1992). All five functional human globin genes linked to the full LCR revealed differences between human and mouse globin expression patterns reflecting further differences in erythropoiesis between species (Iwasaki, Mizuno et al. 2006; Sankaran, Xu et al. 2009). However, there is appropriate expression of ε and γ-globins in the embryonic stage with the latter extending to the early fetal liver period of erythropoiesis and β-globin behaving as a fetal-adult gene. Interestingly, γ-globins were silenced, but much later (by 14*dpc*) than the mouse embryonic genes βh1 and εy which are silenced by 11.5*dpc* and 12.5*dpc* respectively, underscoring the importance of surrounding genomic sequences responsible for this effect rather than competition with the β-globin gene as was indicated before (Dillon and Grosveld 1991). Many laboratories have aimed to identify protein complexes that participate in γ-globin gene silencing but until recently these efforts have been inconclusive and inadequate to explain the complete molecular mechanism. Instead, many known regulators have been investigated for their role in the silencing process of the γ-globin, including MYB (Jiang, Best et al. 2006), the orphan receptors TR2-TR4 (Tanabe, Katsuoka et al. 2002), NF-E2 (Shivdasani and Orkin 1995; Zhao, Zhou et al. 2006), COUP-TF (Aerbajinai, Zhu et al. 2009) and SOX6 (Sripichai, Kiefer et al. 2009).

There are other genes that undergo postnatal silencing, including α-fetoprotein (AFP) (Vacher and Tilghman 1990), myogenin (Berghella, De Angelis et al. 2008) and others that can be re-activated/de-repressed at certain occasions. This begs the question whether there are common mechanisms or steps during these de-repressing events that in combination with specific proteins and regulatory elements orchestrate or re-program cell fate in a tissue specific manner? The answer probably resides in the promoter of the gene (Fig. 7) and the stage-specific interactions with upstream and downstream regulatory sequences of the β-globin locus, where binding of the transcriptional regulators is important to initiate the cascade of subsequent chromatin modifications, resulting in transcription initiation.

Gata1 transcription factor is one of the major hematopoietic players (Pevny, Simon et al. 1991; Fujiwara, Browne et al. 1996) and has been reported to be involved in activation and repression events (Welch, Watts et al. 2004) through its

numerous coregulator partners like Fog-1, Brg-1, CBP/p300, Med1, NuRD complex and interactions with other regulators, including Sp1, Eklf, and PU.1. All erythroid genes, including Gata1 itself, contain *GATA* sites in their promoter and so does the LCR (Fig. 7 and 8).

Interestingly, Gata1 was implicated in the Greek -117 HPFH (Berry, Grosveld et al. 1992) as part of the potential repressor complex responsible for the silencing of the γ-globins in the adult stage, in combination with C/EBPγ. However, C/EBPγ overexpression in transgenic mice showed no change in the levels of γ-globin gene expression in the adult mouse. C/EBPγ is a member of the *CCAAT*/enhancer binding protein (C/EBP) family of transcription factors and overexpression in transgenic mice resulted in increased γ-globin expression in the fetal liver but not

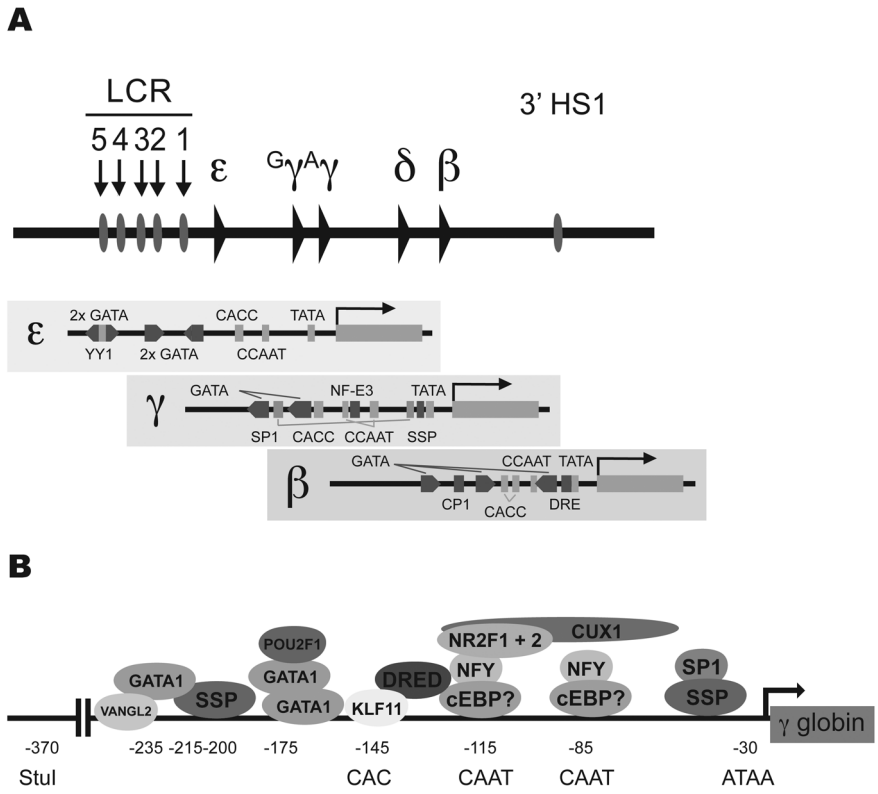


Figure 7

A. Comparative scheme of the binding sites of known transcription factors in the promoters of ε–, γ- and β-globin genes. **B.** Transcription factors and binding sites of the proximal and distal γ-globin gene promoter.

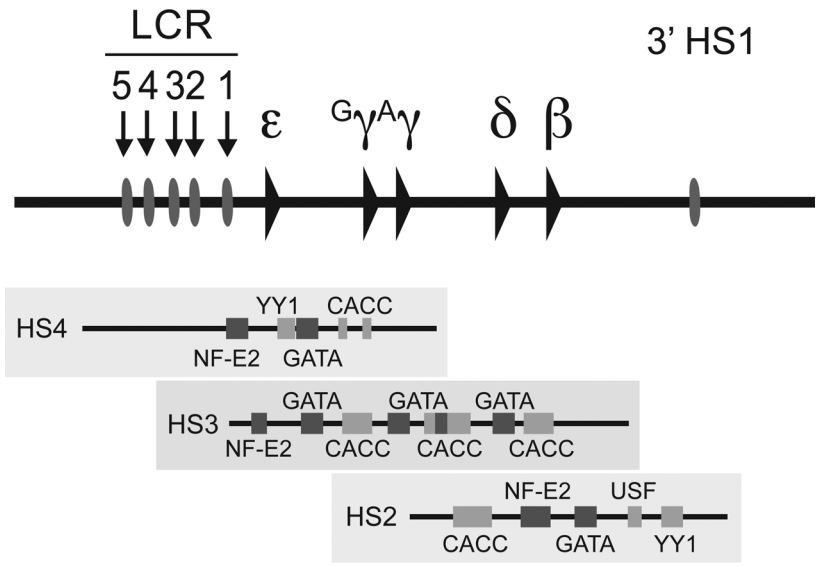


Figure 8
Depiction of selected binding sites of different transcription factors in the LCR hypersensitive sites HS2, HS3, and HS4.

in the adult stage (Zafarana, Rottier et al. 2000). Additionally, in mice carrying the Greek -117 HPFH mutation and the C/EBPγ transgene there was no change in the levels of γ- or β-globin expression, opposed to the idea that C/EBPγ is an activator of β-globin (Wall, Destroismaisons et al. 1996).

During studies of clustered mutations in the ε-globin promoter of the human β-globin locus transgenic mouse, incorporating high affinity binding sites for the erythroid specific transcription factor Eklf (Tanimoto, Liu et al. 2000), shown to be essential for the expression of β-globin (Perkins, Sharpe et al. 1995; Wijgerde, Gribnau et al. 1996), two nuclear orphan receptors (TR2 and TR4) were purified and held responsible for the suppression of the ε-globin gene orthologue, the murine εy gene *in vivo* (Tanabe, Katsuoka et al. 2002). These two receptors can homo- or heterodimerize and bind to a direct repeat (DR) *AGGTCA* sequence separated by 0- to 6- nucleotides (Lee, Chinpaisal et al. 1998). TR2/TR4 are the core DNA binding subunits of the large repressor complex DRED (Direct Repeat Erythroid-Definitive). DRED binds to the Greek -117 HPFH point mutation with 3-fold lower affinity than it does to the WT Aγ gene promoter DR1 site (Fig. 7B). Surprisingly, Tr2 and Tr4 receptors were shown to repress Gata1 transcription

through a conserved DR element in the Gata1 hematopoietic enhancer (G1HE) (Tanabe, Shen et al. 2007), that binds to the γ-globin promoter and could be involved in the HPFH phenotype.

Developmental expression of TR2 and TR4 orphan receptors is ubiquitous. In order to explain a possible repression activity on the γ-globin gene we have to envision a stage-specific effect that could be exerted by changes in the relative abundance of regulators of the γ-globin promoter, thus re-establishing new interactions with coregulators and initiating the proper chromatin environment for transcriptional repression. Recent studies show that DRED appears to bind to the β globin promoter rather than the γ globin gene promoter in the adult stage (Kolodziej unpublished).

A similar situation is encountered with the EKLF transcription factor. It binds *CACC*-box sequences (*CCACACCCT*) (Donze, Townes et al. 1995) frequently found in erythroid-specific gene promoters. Two such sites are present in the human and mouse β-globin promoters. Eklf's promoter has a Gata1 functional binding site (Crossley, Tsang et al. 1994). The fact that Eklf is active in primitive and definitive hematopoiesis and not required for yolk sac erythropoiesis and erythroid commitment (Nuez, Michalovich et al. 1995; Perkins, Sharpe et al. 1995), suggested that it is important for the transition from fetal to adult globin expression in human. Additionally, base substitutions in the EKLF binding sites in the β-globin promoter cause β-thalassemia (Orkin, Kazazian et al. 1982). Studies on transgenic mice bearing the human β-globin locus and inactivated Eklf explained the induction of γ-globin expression, in the absence of transcription of β-globin, as a consequence of productive LCR-promoter interactions of the γ-globins (Wijgerde, Gribnau et al. 1996).

The aforementioned transcriptional regulators have been shown to bind globin proximal promoters (Fig. 7B) and so their effect is attributable directly to transcriptional levels of the globin genes. Recently, a gene that has not been studied for its role in the red blood cell lineage, BCL11A, was identified as a stage-specific regulator of HbF expression (Sankaran, Menne et al. 2008) in human, and embryonic genes (Sankaran, Xu et al. 2009) εy and βh1 in the mouse. Expression profile data in primary human erythroid cells after knockdown and subsequent differentiation showed no differences in expression of well-characterized transcriptional regulators of globin genes, GATA1, FOG-1, NF-E2 and EKLF, suggesting that the effect of BCL11A on the γ-globin gene regulation is unlikely to

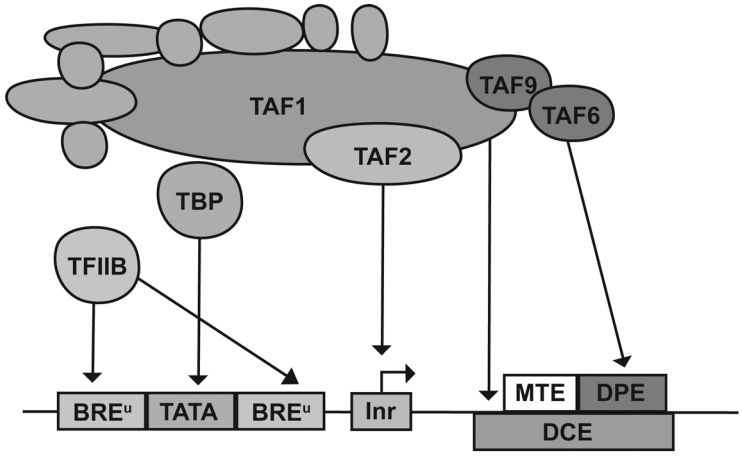
be mediated by these transcriptional regulators. Still, in immunoprecipitation assays (IP), GATA1, FOG-1 and components of the NuRD complex were confirmed to interact with and explain partly the repressor activity of BCL11A in erythroid cells. Surprisingly, BCL11A was not found in the proximal promoter of γ-globin, neither showed robust binding to other regions of the β-globin cluster. However, it was found to bind to three regions: the HS3 of the LCR, a region of high HbF-associated Corfu deletion upstream of the δ-globin gene and another region downstream the Aγ-globin gene that is commonly deleted in certain forms of HPFH (Bank 2006) in primary human erythroid progenitors.

The current experimental approaches to decipher the molecular mechanism of the silencing of the γ-globin genes focus on transcription factors that are found mutated in individuals with higher than normal levels of HbF, as it appears to be the case for the BCL11A. Still, the promoter of the γ-globin genes is of major importance for resolving such a complex gene regulation puzzle, even though so far single protein loss of function experiments have not assigned a critical role for any of the known factors found to be present in their promoters. The basal transcription machinery and the numerous general transcription factors are also thought to play a significant role in promoter-specific gene activation and should be investigated regarding their role in globin switching.

TFIID complex, promoter specificity and cell differentiation programmes

The DNA sequence that has been defined to accurately direct initiation of transcription by the RNAPolII machinery is the 'core promoter' (Struhl 1987). It contains the site of transcription initiation and extends either upstream or downstream for an additional ~35nt. There are several motifs that are commonly found in core promoters, like the TATA box, the initiator (Inr), the TFIIB recognition element (BRE^u and BRE^d), the downstream core promoter element (DPE), the motif ten element (MTE) and the downstream core element (DCE). It is important to note that each one of these core promoter elements are found in some but not all core promoters (Fig. 9).

DNA in chromatin is organized in nucleosomes (Kornberg 1974) that consist of 2 copies of each histone protein H2A, H2B, H3 and H4 assembled into an octamer



Core Promoter Element	Position	Consensus Sequence (5' to 3')	Bound Protein
BRE^u	-38 to -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 to -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE^u	-23 to -17	(G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 to +5	PyPyAN(T/A)PyPy	TAF1/TAF2
MTE	+18 to +29	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	n.a.
DPE	+28 to +34	(A/G)G(A/T)CGTG	TAF6/TAF9
DCE	3 subelements +6 to +11 +16 to +21 +30 to +34	core sequence: CTTC CTGT AGC	TAF1

Figure 9
Recognition of core promoter elements by TFIID and TFIIB. Consensus sequences and positions relative to the start site of transcription (+1) for each element are indicated in the table (based on Figure 2 on Review of M.C Thomas and Cheng-Ming Chiang, Critical Reviews in Biochemistry and Molecular Biology, 41:105-178, 2006).

with 145-147 base pairs (bp) wrapped around it. This nucleoprotein complex is repeated every 200+/-40 bp in eukaryotic genomes. Higher ordered structures of the nucleosomes are stabilized by the histone H1 (Graziano, Gerchman et al. 1994) and the overall compaction of the DNA can reach a factor of 30-40. The packaging of DNA in nucleosomes inhibits in general the binding of non-histone proteins, such as transcription factors, by distorting the DNA in their binding sites. Accessibility of the DNA by transcription factors requires dissociation of the DNA from the nucleosome in the region of the binding site to allow access (Luger, Mader et al. 1997). Thus, while the promoter region may be incorporated into a nucleosome,

the TATA box can be inaccessible for TBP (TATA box binding protein).

In addition to the core promoters there are other cis-acting elements that regulate RNAPolII transcription, including the proximal promoter, enhancers, silencers and boundary/insulator elements (Grosveld, van Assendelft et al. 1987; Blackwood and Kadonaga 1998; Bulger and Groudine 1999; West, Gaszner et al. 2002). The proximal promoter is the region in the immediate vicinity of the transcription start site (roughly -250 to +250nt). Accurate and efficient transcription from the core promoter requires the RNAPolII together with auxiliary factors termed as 'basal' or 'general' transcription factors (GTFs), which include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH to finally assemble the preinitiation complex (PIC). TFIID is a multi-subunit protein that consists of TBP (TATA box-binding protein) and approximately 14 TBP-associated factors (TAFs) (Burley and Roeder 1996; Albright and Tjian 2000; Berk 2000; Verrijzer 2001; Tora 2002). During the transcription preinitiation complex (PIC) formation, TFIIB and TFIID are the first two factors that interact with the core promoter.

TAFs can recognize core promoter elements, such as Inr and DPE and, in combination with TBP binding to TATA box, nucleate the assembly of other GTFs into a functional PIC.

The identification of these promoter elements indicates the broad diversity of eukaryotic promoters. In addition, promoter analysis of EPD (Eukaryotic promoter database) and DBTSS (database of human transcriptional start sites) revealed that less than 22% of the human genes contain TATA-containing promoters, 62% have an Inr, 24% include a DPE, and 12% hold a BRE^{u} (Gershenzon and Ioshikhes 2005). The other 78% of the human genes that are TATA-less, 45% contain an Inr, 25% have a DPE and 28% harbor a BRE^{u}. From functional studies, it seems that the promoters of human housekeeping genes, oncogenes, growth factors, and transcription factors often lack a TATA-box (Zhang 1998). Taken together these data, and the fact that the assembly/recruitment of the eukaryotic RNAPolII to the promoter region occurs either as a holoenzyme complex or in a stepwise manner as a separate entity along with the individual GTFs, support the notion that differential utilization of core promoter elements plays a critical role in regulating gene expression in a spatial, temporal or lineage specific manner (Orphanides, Lagrange et al. 1996; Lemon and Tjian 2000).

Genes are regulated by mixing and matching different types of activators and repressors in a coordinated fashion. It is evident that a given regulator can partner

and function with multiple types of coactivators and corepressors and *vice versa*. It has been shown that different classes of activators target distinct TAFs to initiate transcription (Verrijzer and Tjian 1996). However, all these interactions have to be permitted by a chromatin state, which will allow primarily TBP to access the TATA box. Chromatin modifiers can reconfigure chromatin structure by covalently modifying histones to reduce protein-DNA interactions or by using the energy of ATP hydrolysis to alter histone-DNA contacts. To date there are many histone-modifying enzymes identified, involved in different cellular functions (transcription, repair, replication, chromatin condensation) through a diverse group of chromatin modifications that they confer (acetylation, methylation of lysines (K) or arginines (R), phosphorylation, ubiquitylation, sumoylation, ADP-ribosylation, deimination (Arginine of H3 or H4 converted to citrulline), proline isomerization) (Kouzarides 2007). Apart from chromatin 'unravelling', these modifications recruit further non-histone proteins that can modify chromatin in a higher order and mediate activation or repression of promoters. The "crosstalk" between modifications takes place in the same or different histone tails. There are many examples demonstrating this communication, like the GCN5 acetyltransferase, which may recognize H3 more effectively when it is phosphorylated at serine 10 of histone 3 (H3S10) (Clements, Poux et al. 2003) or ubiquitylation of H2B being required for tri-methylation of lysine 4 of histone 3 (H3K4me3) (Mulder, Brenkman et al. 2007). Specific domains of proteins are responsible for their recruitment to modifications, as methylation is recognized by chromo-like domains and non-related plant homeodomain (PHD) and acetylation by bromodomains. Interestingly, TAF1 possesses histone acetyltransferase (HAT) and ubiquitin ligase activities, and was shown that TFIID targets acetylated or lysine 4 trimethylated H3 (H3K4me3) marks through the PHD domain of TAF3, involving nucleosome modifications in active transcription (Vermeulen, Mulder et al. 2007).

The picture of complex regulation by TFIID became more complicated by the identification of TBP-free TAF-containing complex TFTC (Wieczorek, Brand et al. 1998) and tissue specific TAF-containing TFIID complexes (Freiman, Albright et al. 2001). Additionally, the discovery of TBP-related factors (TBP2/TRF3, TLF/TRF2) that can replace TFIID function and support RNAPolII-mediated transcription, established the idea that GTFs are not only involved in executing transcriptional activation determined by gene-specific activators, but also they are themselves gene-specific factors contributing to the generation of cell and pathway specific

expression patterns (Bell and Tora 1999; Albright and Tjian 2000; Muller, Demeny et al. 2007). The replacement of TFIID by TAF3-contaning TBP2/TRF3 complex on the promoters of muscle specific genes during terminal differentiation of myoblasts to myotubes (Deato and Tjian 2007) is an example of the broad diversity of gene regulation by non-canonical (non-TBP based) TFIID complexes.

Recently it was shown that TAF10 inactivation in the mouse liver in the adult stage de-repressed several postnatally silenced hepatic genes, while less than 5% of the active genes were affected. The result of losing TAF10 was proven to be the dissociation of the TFIID complex into its components with a major effect on genes that were not actively transcribed at the moment of inactivation. Thus, it was proposed that repressors bound to TAFs will be displaced as a result of the TFIID disassembly and genes that were repressed can be reactivated (Tatarakis, Margaritis et al. 2008). This model is particularly interesting for the γ-globin gene silencing mechanism, since so far different regulators have been reported to affect the expression of γ-globin but not in such a way to indicate master control of transcriptional regulation.

In the mouse β-globin locus there have been reports of interactions of erythroid specific transcription factors with TFIID components. The p45 NF-E2 subunit has been reported to interact with TAF4 and Eklf with TAF9 (Amrolia, Ramamurthy et al. 1997; Sengupta, Cohet et al. 2009). Even though these interactions provide evidence of promoter specificity defined by DNA-binding proteins and GTFs, there is an advanced level of complexity introduced by interactions with chromatin remodelers, like Brg1, and histone acetylation or deacetylation complexes (HATs, HDACs), that is defined by the order in which they act, in order to produce the correct conformational changes in the promoter region to finally initiate transcription (Agalioti, Lomvardas et al. 2000; Hatzis and Talianidis 2002; Bottardi, Ross et al. 2006; Kim, Bultman et al. 2009).

Brg1, part of the SWI/SNF chromatin remodeling complex, has been reported to interact with different key hematopoietic regulators, like Gata1 (Kim, Bultman et al. 2007), Eklf (Zhang, Kadam et al. 2001), p45/NF-E2 (Brand, Ranish et al. 2004), and consequently it is found in the LCR HSs and the β-globin promoter in chromatin immunoprecipitation assays (ChIPs). Kinetic studies revealed that the recruitment of Brg1 is mediated by Gata1. Mice with a mutant Brg1 protein that appears to abolish its nucleosome remodeling properties are embryonic lethal between 11.5*dpc* and 14.5*dpc* exhibiting a hematopoietic block in the definitive

stage affecting the expression mainly of β-maj and β-min globin genes. Acetylation of histone H3 (AcH3) in the LCR is reduced as well as the hypersensitivity of the LCR (Bultman, Gebuhr et al. 2005).

How do all these regulators and coregulators work together to deploy promoter specific chromatin rearrangements in order to form the preinitiation complex (PIC) and support transcription initiation and elongation? How is PolII recruited to the promoter ready to be transcribed? Does the LCR play a fundamental role in the initiation of the proper chromatin conformation or it facilitates further stability of interactions between transcriptional regulators and coregulators? What is the timing of all these extremely complex events? All these challenging questions are still under extensive ongoing research.

Hypersensitive sites (HS), chromatin architecture and transcriptional activation of the β-globin locus

Deletion of the Locus Control Region (LCR) has been shown to affect negatively efficient transcription of the β-like globin genes (Grosveld, van Assendelft et al. 1987; Bender, Bulger et al. 2000; Sawado, Halow et al. 2003). When single hypersensitive sites (HS) were linked to the β-globin gene to generate transgenic mice or cell lines (Collis, Antoniou et al. 1990; Fraser, Hurst et al. 1990; Philipsen, Talbot et al. 1990), HS2 and HS3 appeared to account for 40-50% of the full LCR hypersensitivity whereas HS5 functions as a developmental stage specific border (Farrell, West et al. 2002; Wai, Gillemans et al. 2003) rather than a transcription enhancer (Reik, Telling et al. 1998; Li, Zhang et al. 2002).

Interactions between the LCR and the globin genes downstream have been the center of research trying to define the way that these interactions occur and attribute changes in the chromatin architecture of the locus relative to active transcription of the globin genes. The description of the Active Chromatin Hub (ACH) gave a 3D image of the human and mouse β-globin locus (Tolhuis, Palstra et al. 2002; Palstra, Tolhuis et al. 2003; Patrinos, de Krom et al. 2004) and revealed a dynamic structure that communicates enhancers, promoters and specific regulators and coregulators to execute gene transcription. In this model, the intervening sequences and non-transcribed genes are looping out of the active site of transcription (Fig. 10), supporting a looping model for transcriptional activation

of the globin genes. The formation of a basal Chromatin Hub (CH) consists of the LCR hypersensitive sites (HS), the 3'HS1 and the 5' HS-60/-62 in the mouse. Information about the HS -110, the equivalent of the mouse HS-60/-62, of the human β-globin locus (Bulger, Bender et al. 2000) regarding the formation of CH has not been revealed yet since transgenic mice do not carry this hypersensitive site in the globin locus. Loss of the β-globin gene promoter has a very mild effect on the formation of the ACH, but results in upregulation of γ-globin at 12.5*dpc* since there is absence of gene competition between γ- and β-globin genes. Instead, loss of HS3 combined with the deletion of the β-globin promoter resulted in the absence of γ-globin gene expression at 12.5*dpc*, as well as hypersensitivity of the LCR sites and the promoters of γ- and β-globin genes, and significantly reduced acetylation levels of globin promoters and HS5, suggesting a severely affected chromatin structure (Patrinos, de Krom et al. 2004). These results implicated HS3 directly in the stability of the ACH (Fang, Xiang et al. 2007) and in further interactions with sequences other than the promoter of the β-globin important for the transcription of the gene and for the maintenance of the ACH.

It is known that 3'HS1 and the 5'HS (-60/-62) for the mouse and HS-110 for the human β-globin locus are not required for high level expression of the globins (Strouboulis, Dillon et al. 1992; Bulger, Schubeler et al. 2003). Still, they form part of the CH. Interestingly, Ctcf binds to these sites, apart from other transcription factors and chromatin remodeling proteins that could be involved in the ACH formation. CTCF is ubiquitously expressed and can act as an enhancer- blocking protein, prevent the spreading of heterochromatin when anchored to boundary elements or function as both a transcriptional activator and repressor (Wallace and Felsenfeld 2007). Most importantly, it has been suggested that CTCF mediates at least some of its gene regulatory functions through the three–dimensional organization of the genome (Majumder, Gomez et al. 2008) but still it is not clear whether the long-range interactions are a consequence rather than the cause of the transcription process. Since CTCF sites are numerous all over the genome (Kim, Abdullaev et al. 2007), redundancy may explain why lack of 3'HS1 does not have an effect in β-globin gene expression nor on the establishment of LCR-promoter contact during erythroid differentiation. However, Ctcf is important for the looped conformation of the β-locus (Fig. 10) in erythroid progenitor cells and the maintenance of local euchromatic epigenetic marks (Splinter, Heath et al. 2006). A plethora of transcription factors and chromatin remodeling proteins have been reported to bind

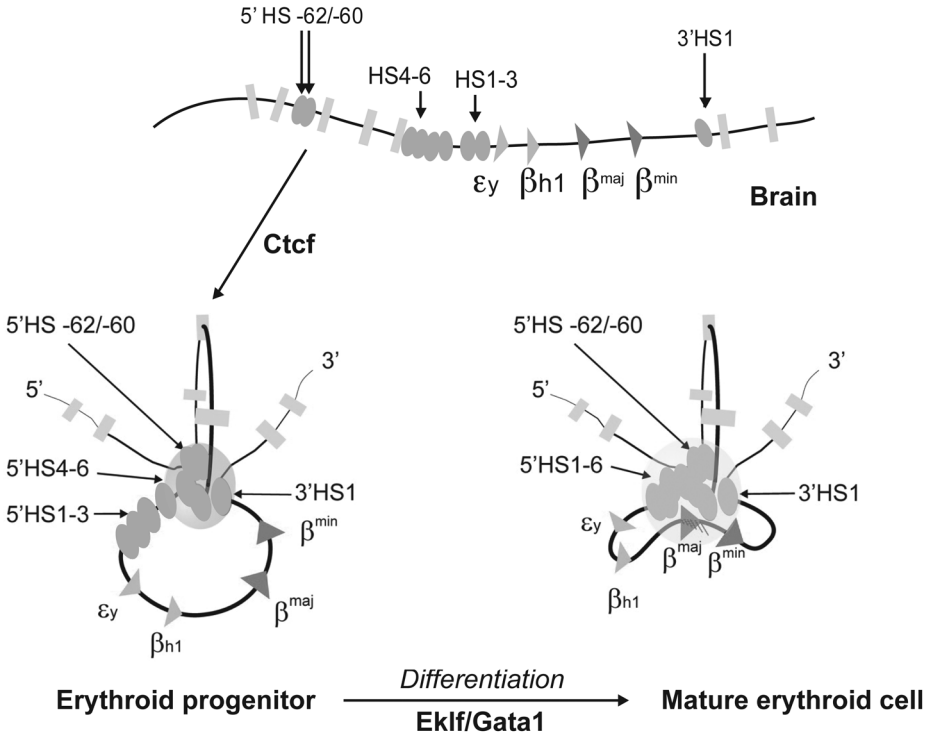


Figure 10
In the fetal brain cells the β-globin locus adopts a "linear" conformation and globin genes are not transcribed. In the contrary, in erythroid progenitors a structure called the chromatin hub (CH) is present which is sustained by interactions between the Ctcf-bound regulatory elements HS-60.5/-62.7, the 3'HS1 and the LCR HS(4-6). Ctcf itself is essential for these interactions to take place. During differentiation of the erythroid progenitors, β-globin gene expression is the result of the interactions of the remaining HS(1-3), the CH and transcription factors with the β-globin promoter to form the Active Chromatin Hub (ACH). Stability of the functional ACH is dependent on the presence of transcription factors, including Eklf and Gata1 (Taken from Palstra *et al*, 2008).

hypersensitive sites (HS) and proximal or distal regulatory elements of the globin genes. Well studied GATA1, EKLF, NF-E2, and Sp1 transcription regulators are common examples that are found in different stoichiometries from the HSs of the LCR to the promoters of the globin genes (Fig. 7 and 8). Even more, mutations in the promoter of γ-globin (Berry, Grosveld et al. 1992), the downstream promoter element (DPE) (Oner, Agarwal et al. 1991; Cai, Eng et al. 1992; Lewis, Kim et al. 2000) of β-globin in thalassemic patients or deletions of the HS of the LCR (Driscoll, Dobkin et al. 1989) and mutated transcription factors affect their concentration in

the β-globin locus. Consequently, chromatin conformational changes are introduced influencing the expression of globin genes.

The first transcription regulator that was shown to influence the formation of the ACH was Eklf. In the absence of Eklf, a fully functional ACH can not be formed (Drissen, Palstra et al. 2004). Together with the observation that in Eklf knockout mice loss of 5'HS3 and βmaj-promoter chromatin accessibility (Wijgerde, Gribnau et al. 1996) occurs, suggested that Eklf is necessary for hypersensitive site formation and participation of the LCR and the β-globin promoter in the ACH, probably through interactions with a SWI/SNF chromatin remodeling complex (Armstrong, Bieker et al. 1998). Gata1 was also shown to be essential for LCR-gene contacts (Vakoc, Letting et al. 2005) in contrast with Ctcf that was found dispensable for such interactions and globin gene expression (Splinter, Heath et al. 2006).

A potential role of the LCR loop formation in RNAPolII loading to the promoters of the globin genes has been proposed (Johnson, Christensen et al. 2001). Still, in Eklf knockout erythrocytes, RNAPolII is loaded on the promoter of β-globin but the levels of Ser2 phosphorylated PolII, as a mark of active transcription, are reduced. This explains the decrease in β-globin expression (Bottardi, Ross et al. 2006). Thus, it is more likely that the recruitment of RNAPolII, at least to the β-globin promoter, is LCR independent while the transition from the initiation to the elongation step of active transcription is LCR dependent (Sawado, Halow et al. 2003).

Nuclear Compartmentalization of transcription

If there is a systematic organization to the metazoan nucleus, chromosomes must be arranged in such a way so they accommodate cell-type specific and temporal patterns of gene expression. It appears that transcriptionally active genes are generally oriented towards the center of the nucleus within individual chromosome territories, while silenced genes are distributed near the periphery of the nucleus (Andrulis, Neiman et al. 1998; Verschure, van Der Kraan et al. 1999; Cremer and Cremer 2001). In concordance with that, the β-globin locus was shown to move towards the nuclear center during maturation of erythroid cells and associate with hyperphosphorylated PolII transcription factories in a LCR dependent manner, even though transcription is still taking place in the periphery (Ragoczy, Bender et al. 2006). Additionally, other erythroid genes transcribed in *cis* with the active

βmaj-globin gene location were found in considerable proximity to RNAPolII, in colocalization experiments, suggesting that transcription takes place in selected areas in the nucleus, the transcription factories (Osborne, Chakalova et al. 2004). Insight in intrachromosomal interactions of the mouse β-globin locus has been achieved by applying 4C technology demonstrating that the active β-globin gene was in contact with other active loci, tens of megabases away but when inactive, interactions mainly were restricted to non-transcribed loci (Simonis, Klous et al. 2006).

Supporting evidence of a subterritorial division of the nucleus that provides more control during different synchronous stimuli, comes from the dynamic and transient association of transcription factors with their cognate DNA recognition sites and cofactor targets. If the cell were to inactivate the entire cellular pool of a given activator or coactivator in response to one signal, such a mechanism would preclude responsiveness by other factors or cooperation at other genes in response to additional signals. Instead, if inactivation of a few molecules of a particular transcription factor takes place within a designated compartment without affecting the same factors in other compartments associated with different genes, then certain procedures are more efficiently carried out. Deciphering the subnuclear distribution of transcriptional regulators and coregulators would probably give a lot of information about the functionality of segregated compartments.

Scope of the thesis

In order to gain more insight in the γ-globin gene regulation process, I generated a human β-globin locus reporter mouse model in which the γ- and β-globin genes are modified by introducing two fluorescent proteins, GFP and DsRed respectively. Fetal liver cell lines were established from these transgenic mice and were used as an erythroid *in vitro* model to study responses to genetic manipulation, such as modulating the expression of potential regulators of γ-globin gene expression by shRNA mediated knockdown (Chapter 1).

Basal transcription mechanisms include interactions of regulators with coregulators combining distal and proximal regulatory elements in different promoters in order to initiate RNAPolII transcription. The specificity of different basal transcription factors of the TFIID complex for different promoters is poorly understood in erythroid cells. Here, I found stoichiometric differences of many TFIID members during the differentiation process in erythroid cells and I characterized distinct SAGA complexes (TBP-free TAF containing complexes) which I linked to γ-globin expression (Chapter 2).

The role of one crucial transcriptional regulator, EKLF, was studied since it was linked by single nucleotide polymorphism (SNP) analysis to increased fetal hemoglobin (HbF) amongst members of a family. *In vitro* studies in human erythroid progenitors from healthy donors focused on addressing the role of EKLF when reduced expression is achieved or a mutant form without the zinc fingers of the protein is overexpressed (Chapter 3).

In our efforts to identify crucial proteins participating in the regulation of γ–globin gene expression, I constructed a lentivirus-based naïve library of llama antibodies (HCAbs) that have been shown to be highly diverse in terms of identifying new conformational epitopes, and additionally to be strong binders with high affinities for their targets. By using the reporter transgenic mouse cell lines, I tried to identify potential antibodies that could reactivate the γ-globin gene, by blocking repressors or repressor complexes upon expression in erythroid cells (Chapter 4).

References

Aerbajinai, W., J. Zhu, et al. (2009). «SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII.» Blood **114**(1): 187-94.

Agalioti, T., S. Lomvardas, et al. (2000). "Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter." Cell **103**(4): 667-78.

Albright, S. R. and R. Tjian (2000). "TAFs revisited: more data reveal new twists and confirm old ideas." Gene **242**(1-2): 1-13.

Amrolia, P. J., L. Ramamurthy, et al. (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**(19): 10051-6.

Andrulis, E. D., A. M. Neiman, et al. (1998). "Perinuclear localization of chromatin facilitates transcriptional silencing." Nature **394**(6693): 592-5.

Armstrong, J. A., J. J. Bieker, et al. (1998). "A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro." Cell **95**(1): 93-104.

Bank, A. (2006). "Regulation of human fetal hemoglobin: new players, new complexities." Blood **107**(2): 435-43.

Bell, B. and L. Tora (1999). "Regulation of gene expression by multiple forms of TFIID and other novel TAFII-containing complexes." Exp Cell Res **246**(1): 11-9.

Bender, M. A., M. Bulger, et al. (2000). "Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region." Mol Cell **5**(2): 387-93.

Benesch, R. and R. E. Benesch (1969). "Intracellular organic phosphates as regulators of oxygen release by haemoglobin." Nature **221**(5181): 618-22.

Berghella, L., L. De Angelis, et al. (2008). "A highly conserved molecular switch binds MSY-3 to regulate myogenin repression in postnatal muscle." Genes Dev **22**(15): 2125-38.

Berk, A. J. (2000). "TBP-like factors come into focus." Cell **103**(1): 5-8.

Bernaudin, F., G. Socie, et al. (2007). "Long-term results of related myeloablative stem-cell transplantation to cure sickle cell disease." Blood **110**(7): 2749-56.

Berry, M., F. Grosveld, et al. (1992). "A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin." Nature **358**(6386): 499-502.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 60-3.

Blom van Assendelft, G., O. Hanscombe, et al. (1989). "The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner." Cell **56**(6): 969-77.

Bottardi, S., J. Ross, et al. (2006). "Lineage-specific activators affect beta-globin locus chromatin in multipotent hematopoietic progenitors." Embo J **25**(15): 3586-95.

Brand, M., J. A. Ranish, et al. (2004). "Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics." Nat Struct Mol Biol **11**(1): 73-80.

Briehl, R. W. and P. Nikolopoulou (1993). "Kinetics of hemoglobin S polymerization and gelation under shear: I. Shape of the viscosity progress curve and dependence of delay time and reaction rate on shear rate and temperature." Blood **81**(9): 2420-8.

Brotherton, T. W., D. H. Chui, et al. (1979). "Hemoglobin ontogeny during normal mouse fetal development." Proc Natl Acad Sci U S A **76**(6): 2853-7.

Bulger, M., M. A. Bender, et al. (2000). "Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters." Proc Natl Acad Sci U S A **97**(26): 14560-5.

Bulger, M. and M. Groudine (1999). "Looping versus linking: toward a model for long-distance gene activation." Genes Dev **13**(19): 2465-77.

Bulger, M., D. Schubeler, et al. (2003). "A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus." Mol Cell Biol **23**(15): 5234-44.

Bultman, S. J., T. C. Gebuhr, et al. (2005). "A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in beta-globin expression and erythroid development." Genes Dev **19**(23): 2849-61.

Burley, S. K. and R. G. Roeder (1996). "Biochemistry and structural biology of transcription factor IID (TFIID)." Annu Rev Biochem **65**: 769-99.

Cai, S. P., B. Eng, et al. (1992). «Two novel beta-thalassemia mutations in the 5' and 3' noncoding regions of the beta-globin gene.» Blood **79**(5): 1342-6.

Chada, K., J. Magram, et al. (1986). "An embryonic pattern of expression of a human fetal globin gene in transgenic mice." Nature **319**(6055): 685-9.

Charache, S., M. L. Terrin, et al. (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." N Engl J Med **332**(20): 1317-22.

Clements, A., A. N. Poux, et al. (2003). «Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase.» Mol Cell **12**(2): 461-73.

Collins, F. S., C. J. Stoeckert, Jr., et al. (1984). "G gamma beta+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G gamma gene." Proc Natl Acad Sci U S A **81**(15): 4894-8.

Collis, P., M. Antoniou, et al. (1990). "Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression." Embo J **9**(1): 233-40.

Conley, C. (1980). Sickle cell anemia.The first molecular disease, in Blood, Pure and Eloquent. New York, McGraw-Hill.

Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nat Rev Genet **2**(4): 292-301.

Crossley, M., A. P. Tsang, et al. (1994). «Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1.» J Biol Chem **269**(22): 15440-4.

Deato, M. D. and R. Tjian (2007). "Switching of the core transcription machinery during myogenesis." Genes Dev **21**(17): 2137-49.

Di Bartolomeo P, D. G. D., Olioso P, et al (1997). "The Pescara experience of allogenic bone marrow transplantation in thalassemia." Bone Marrow Transplant **19**(48).

Dillon, N. and F. Grosveld (1991). "Human gamma-globin genes silenced independently of other genes in the beta-globin locus." Nature **350**(6315): 252-4.

Donze, D., T. M. Townes, et al. (1995). "Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching." J Biol Chem **270**(4): 1955-9.

Driscoll, M. C., C. S. Dobkin, et al. (1989). "Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites." Proc Natl Acad Sci U S A **86**(19): 7470-4.

Drissen, R., R. J. Palstra, et al. (2004). "The active spatial organization of the beta-globin locus requires the transcription factor EKLF." Genes Dev **18**(20): 2485-90.

Enver, T., N. Raich, et al. (1990). "Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice." Nature **344**(6264): 309-13.

Fang, X., P. Xiang, et al. (2007). "Cooperativeness of the higher chromatin structure of the beta-globin locus revealed by the deletion mutations of DNase I hypersensitive site 3 of the LCR." J Mol Biol **365**(1): 31-7.

Farrell, C. M., A. G. West, et al. (2002). "Conserved CTCF insulator elements flank the mouse and human beta-globin loci." Mol Cell Biol **22**(11): 3820-31.

Fermi, G., M. F. Perutz, et al. (1984). "The crystal structure of human deoxyhaemoglobin at 1.74 A resolution." J Mol Biol **175**(2): 159-74.

Forrester, W. C., S. Takegawa, et al. (1987). "Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids." Nucleic Acids Res **15**(24): 10159-77.

Forrester, W. C., C. Thompson, et al. (1986). "A developmentally stable chromatin structure in the human beta-globin gene cluster." Proc Natl Acad Sci U S A **83**(5): 1359-63.

Fraser, P., J. Hurst, et al. (1990). "DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression." Nucleic Acids Res **18**(12): 3503-8.

Freiman, R. N., S. R. Albright, et al. (2001). "Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development." Science **293**(5537): 2084-7.

Fujiwara, Y., C. P. Browne, et al. (1996). "Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1." Proc Natl Acad Sci U S A **93**(22): 12355-8.

Gaziev, D., P. Polchi, et al. (1997). "Graft-versus-host disease after bone marrow transplantation for thalassemia: an analysis of incidence and risk factors." Transplantation **63**(6): 854-60.

Gelinas, R., B. Endlich, et al. (1985). "G to A substitution in the distal CCAAT box of the A gamma-globin gene in Greek hereditary persistence of fetal haemoglobin." Nature **313**(6000): 323-5.

Gershenzon, N. I. and I. P. Ioshikhes (2005). "Synergy of human *Pol II* core promoter elements revealed by statistical sequence analysis." Bioinformatics **21**(8): 1295-300.

Giglioni, B., C. Casini, et al. (1984). "A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia." Embo J **3**(11): 2641-5.

Graziano, V., S. E. Gerchman, et al. (1994). "Histone H1 is located in the interior of the chromatin 30-nm filament." Nature 368(6469): 351-4.

Grosveld, F., G. B. van Assendelft, et al. (1987). "Position-independent, high-level expression of the human beta-globin gene in transgenic mice." Cell **51**(6): 975-85.

Groudine, M., T. Kohwi-Shigematsu, et al. (1983). "Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus." Proc Natl Acad Sci U S A **80**(24): 7551-5.

Haley, J. D., D. E. Smith, et al. (2003). "Identification and characterization of mechanistically

distinct inducers of gamma-globin transcription." Biochem Pharmacol **66**(9): 1755-68.

Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-3.

Hanna, J., S. Markoulaki, et al. (2008). "Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency." Cell 133(2): 250-64.

Hanna, J., K. Saha, et al. (2009). "Direct cell reprogramming is a stochastic process amenable to acceleration." Nature. (Ahead of print).

Hanscombe, O., M. Vidal, et al. (1989). "High-level, erythroid-specific expression of the human alpha-globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes." Genes Dev **3**(10): 1572-81.

Hanscombe, O., D. Whyatt, et al. (1991). "Importance of globin gene order for correct developmental expression." Genes Dev **5**(8): 1387-94.

Hatzis, P. and I. Talianidis (2002). "Dynamics of enhancer-promoter communication during differentiation-induced gene activation." Mol Cell **10**(6): 1467-77.

Hebbel, R. P. (1992). "Endothelial adhesivity of sickle red blood cells." J Lab Clin Med **120**(4): 503-4.

Isern, J., S. T. Fraser, et al. (2008). „The fetal liver is a niche for maturation of primitive erythroid cells." Proc Natl Acad Sci U S A **105**(18): 6662-7.

Iwasaki, H., S. Mizuno, et al. (2006). "The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages." Genes Dev **20**(21): 3010-21.

Jaenisch, R. (2004). "Human cloning - the science and ethics of nuclear transplantation." N Engl J Med **351**(27): 2787-91.

Jiang, J., S. Best, et al. (2006). "cMYB is involved in the regulation of fetal hemoglobin production in adults." Blood **108**(3): 1077-83.

Johnson, K. D., H. M. Christensen, et al. (2001). "Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter." Mol Cell **8**(2): 465-71.

Kim, S. I., S. J. Bultman, et al. (2007). «Dissecting molecular steps in chromatin domain activation during hematopoietic differentiation.» Mol Cell Biol **27**(12): 4551-65.

Kim, S. I., S. J. Bultman, et al. (2009). "BRG1 requirement for long-range interaction of a locus control region with a downstream promoter." Proc Natl Acad Sci U S A **106**(7): 2259-64.

Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." Cell **128**(6): 1231-45.

Kioussis, D., E. Vanin, et al. (1983). "Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia." Nature **306**(5944): 662-6.

Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." Science **184**(139): 868-71.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Lee, C. H., C. Chinpaisal, et al. (1998). "A novel nuclear receptor heterodimerization pathway mediated by orphan receptors TR2 and TR4." J Biol Chem **273**(39): 25209-15.

Lemon, B. and R. Tjian (2000). "Orchestrated response: a symphony of transcription factors for gene control." Genes Dev **14**(20): 2551-69.

Lettre, G., V. G. Sankaran, et al. (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**(33): 11869-74.
Lewis, B. A., T. K. Kim, et al. (2000). "A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts." Proc Natl Acad Sci U S A **97**(13): 7172-7.
Li, Q., M. Zhang, et al. (2002). "Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice." Nucleic Acids Res **30**(11): 2484-91.
Liu, P., J. R. Keller, et al. (2003). «Bcl11a is essential for normal lymphoid development.» Nat Immunol **4**(6): 525-32.
Lucarelli, G., C. Giardini, et al. (1995). "Bone marrow transplantation in thalassemia." Semin Hematol **32**(4): 297-303.
Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.
Magram, J., K. Chada, et al. (1985). "Developmental regulation of a cloned adult beta-globin gene in transgenic mice." Nature **315**(6017): 338-40.
Majumder, P., J. A. Gomez, et al. (2008). "The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions." J Exp Med 205(4): 785-98.
Mankidy, R., D. V. Faller, et al. (2006). "Short-chain fatty acids induce gamma-globin gene expression by displacement of a HDAC3-NCoR repressor complex." Blood **108**(9): 3179-86.
Menzel, S., C. Garner, et al. (2007). "A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15." Nat Genet **39**(10): 1197-9.
Mucenski, M. L., K. McLain, et al. (1991). "A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis." Cell **65**(4): 677-89.
Mulder, K. W., A. B. Brenkman, et al. (2007). "Regulation of histone H3K4 tri-methylation and PAF complex recruitment by the Ccr4-Not complex." Nucleic Acids Res **35**(7): 2428-39.
Muller, A. M., A. Medvinsky, et al. (1994). "Development of hematopoietic stem cell activity in the mouse embryo." Immunity **1**(4): 291-301.
Muller, F., M. A. Demeny, et al. (2007). "New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors." J Biol Chem **282**(20): 14685-9.
Nuez, B., D. Michalovich, et al. (1995). "Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene." Nature **375**(6529): 316-8.
Olivieri, N. F. and D. J. Weatherall (1998). "The therapeutic reactivation of fetal haemoglobin." Hum Mol Genet **7**(10): 1655-8.
Oner, R., S. Agarwal, et al. (1991). "The G----A mutation at position +22 3' to the Cap site of the beta-globin gene as a possible cause for a beta-thalassemia." Hemoglobin **15**(1-2): 67-76.
Orkin, S. H., H. H. Kazazian, Jr., et al. (1982). "Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster." Nature **296**(5858): 627-31.
Orphanides, G., T. Lagrange, et al. (1996). "The general transcription factors of RNA polymerase II." Genes Dev **10**(21): 2657-83.
Osborne, C. S., L. Chakalova, et al. (2004). "Active genes dynamically colocalize to shared

sites of ongoing transcription." Nat Genet **36**(10): 1065-71.
Palstra, R. J., B. Tolhuis, et al. (2003). "The beta-globin nuclear compartment in development and erythroid differentiation." Nat Genet **35**(2): 190-4.
Palstra, R. J., W. de Laat, et al. (2008). "Beta-globin regulation and long-range interactions." Adv Genet 61: 107-42.
Parise, L. V. and M. J. Telen (2003). "Erythrocyte adhesion in sickle cell disease." Curr Hematol Rep **2**(2): 102-8.
Patrinos, G. P., M. de Krom, et al. (2004). "Multiple interactions between regulatory regions are required to stabilize an active chromatin hub." Genes Dev **18**(12): 1495-509.
Pauling, L., H. A. Itano, et al. (1949). "Sickle cell anemia a molecular disease." Science **110**(2865): 543-8.
Perkins, A. C., A. H. Sharpe, et al. (1995). "Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF." Nature **375**(6529): 318-22.
Perutz, M. F. (1978). "Hemoglobin structure and respiratory transport." Sci Am **239**(6): 92-125.
Peschle, C., F. Mavilio, et al. (1985). "Haemoglobin switching in human embryos: asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage." Nature **313**(5999): 235-8.
Pevny, L., M. C. Simon, et al. (1991). "Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1." Nature **349**(6306): 257-60.
Philipsen, S., D. Talbot, et al. (1990). "The beta-globin dominant control region: hypersensitive site 2." Embo J **9**(7): 2159-67.
Ragoczy, T., M. A. Bender, et al. (2006). "The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation." Genes Dev **20**(11): 1447-57.
Reik, A., A. Telling, et al. (1998). "The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells." Mol Cell Biol **18**(10): 5992-6000.
Sankaran, V. G., T. F. Menne, et al. (2008). "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A." Science **322**(5909): 1839-42.
Sankaran, V. G., J. Xu, et al. (2009). "Developmental and species-divergent globin switching are driven by BCL11A." Nature **460**(7259): 1093-7.
Satterwhite, E., T. Sonoki, et al. (2001). "The BCL11 gene family: involvement of BCL11A in lymphoid malignancies." Blood **98**(12): 3413-20.
Sawado, T., J. Halow, et al. (2003). "The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation." Genes Dev **17**(8): 1009-18.
Sengupta, T., N. Cohet, et al. (2009). "Distinct modes of gene regulation by a cell-specific transcriptional activator." Proc Natl Acad Sci U S A **106**(11): 4213-8.
Shivdasani, R. A., E. L. Mayer, et al. (1995). "Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL." Nature **373**(6513): 432-4.
Shivdasani, R. A. and S. H. Orkin (1995). "Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2." Proc Natl Acad Sci U S A **92**(19): 8690-4.
Simonis, M., P. Klous, et al. (2006). "Nuclear organization of active and inactive chromatin

domains uncovered by chromosome conformation capture-on-chip (4C)." Nat Genet **38**(11): 1348-54.

Singleton, B. K., N. M. Burton, et al. (2008). "Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype." Blood **112**(5): 2081-8.

Spangrude, G. J., S. Cho, et al. (2006). "Mouse models of hematopoietic engraftment: limitations of transgenic green fluorescent protein strains and a high-performance liquid chromatography approach to analysis of erythroid chimerism." Stem Cells **24**(9): 2045-51.

Splinter, E., H. Heath, et al. (2006). "CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus." Genes Dev **20**(17): 2349-54.

Sripichai, O., C. M. Kiefer, et al. (2009). "Cytokine-mediated increases in fetal hemoglobin are associated with globin gene histone modification and transcription factor reprogramming." Blood **114**(11): 2299-306.

Strouboulis, J., N. Dillon, et al. (1992). "Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice." Genes Dev **6**(10): 1857-64.

Struhl, K. (1987). "Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast." Cell **49**(3): 295-7.

Stryer, L. (1995). Biochemistry. New York, WH Freeman and Company.

Swank, R. A. and G. Stamatoyannopoulos (1998). "Fetal gene reactivation." Curr Opin Genet Dev **8**(3): 366-70.

Swenson, E. S., J. G. Price, et al. (2007). "Limitations of green fluorescent protein as a cell lineage marker." Stem Cells **25**(10): 2593-600.

Tanabe, O., F. Katsuoka, et al. (2002). "An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer." Embo J **21**(13): 3434-42.

Tanabe, O., Y. Shen, et al. (2007). "The TR2 and TR4 orphan nuclear receptors repress Gata1 transcription." Genes Dev **21**(21): 2832-44.

Tanimoto, K., Q. Liu, et al. (2000). "Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice." Genes Dev **14**(21): 2778-94.

Taramelli, R., D. Kioussis, et al. (1986). "Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster." Nucleic Acids Res **14**(17): 7017-29.

Tatarakis, A., T. Margaritis, et al. (2008). "Dominant and redundant functions of TFIID involved in the regulation of hepatic genes." Mol Cell **31**(4): 531-43.

Thein, S. L., S. Menzel, et al. (2007). "Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults." Proc Natl Acad Sci U S A **104**(27): 11346-51.

Thevenin, B. J., I. Crandall, et al. (1997). "Band 3 peptides block the adherence of sickle cells to endothelial cells in vitro." Blood **90**(10): 4172-9.

Tolhuis, B., R. J. Palstra, et al. (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-65.

Tora, L. (2002). "A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription." Genes Dev **16**(6): 673-5.

Townes, T. M., J. B. Lingrel, et al. (1985). "Erythroid-specific expression of human beta-globin genes in transgenic mice." Embo J **4**(7): 1715-23.

Tyuma, I. and K. Shimizu (1970). “Effect of organic phosphates on the difference in oxygen affinity between fetal and adult human hemoglobin.” Fed Proc **29**(3): 1112-4.
Uda, M., R. Galanello, et al. (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**(5): 1620-5.
Vacher, J. and S. M. Tilghman (1990). “Dominant negative regulation of the mouse alpha-fetoprotein gene in adult liver.” Science **250**(4988): 1732-5.
Vakoc, C. R., D. L. Letting, et al. (2005). “Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.” Mol Cell **17**(3): 453-62.
Vermeulen, M., K. W. Mulder, et al. (2007). “Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.” Cell **131**(1): 58-69.
Verrijzer, C. P. (2001). “Transcription factor IID--not so basal after all.” Science **293**(5537): 2010-1.
Verrijzer, C. P. and R. Tjian (1996). “TAFs mediate transcriptional activation and promoter selectivity.” Trends Biochem Sci **21**(9): 338-42.
Verschure, P. J., I. van Der Kraan, et al. (1999). “Spatial relationship between transcription sites and chromosome territories.” J Cell Biol **147**(1): 13-24.
Wai, A. W., N. Gillemans, et al. (2003). “HS5 of the human beta-globin locus control region: a developmental stage-specific border in erythroid cells.” Embo J **22**(17): 4489-500.
Wall, L., N. Destroismaisons, et al. (1996). “CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells.” J Biol Chem **271**(28): 16477-84.
Wallace, J. A. and G. Felsenfeld (2007). “We gather together: insulators and genome organization.” Curr Opin Genet Dev 17(5): 400-7.
Warren, A. J., W. H. Colledge, et al. (1994). “The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development.” Cell **78**(1): 45-57.
Wawrzyniak, C. J. and R. A. Popp (1987). “Expression of the two adult beta-globin genes in mouse yolk sac and fetal liver erythrocytes.” Dev Biol **119**(1): 299-301.
Weatherall, D. J. (2003). “Pharmacological treatment of monogenic disease.” Pharmacogenomics J **3**(5): 264-6.
Welch, J. J., J. A. Watts, et al. (2004). “Global regulation of erythroid gene expression by transcription factor GATA-1.” Blood **104**(10): 3136-47.
West, A. G., M. Gaszner, et al. (2002). “Insulators: many functions, many mechanisms.” Genes Dev **16**(3): 271-88.
Whitelaw, E., S. F. Tsai, et al. (1990). “Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse.” Mol Cell Biol **10**(12): 6596-606.
Wieczorek, E., M. Brand, et al. (1998). “Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II.” Nature **393**(6681): 187-91.
Wijgerde, M., J. Gribnau, et al. (1996). “The role of EKLF in human beta-globin gene competition.” Genes Dev **10**(22): 2894-902.
Wong, P. M., S. W. Chung, et al. (1985). “Ontogeny of the mouse hemopoietic system.” Prog Clin Biol Res **193**: 17-28.
Yu, J., M. A. Vodyanik, et al. (2007). “Induced pluripotent stem cell lines derived from human somatic cells.” Science 318(5858): 1917-20.
Zafarana, G., R. Rottier, et al. (2000). “Erythroid overexpression of C/EBPgamma in

transgenic mice affects gamma-globin expression and fetal liver erythropoiesis." Embo J **19**(21): 5856-63.

Zhang, M. Q. (1998). "Identification of human gene core promoters in silico." Genome Res **8**(3): 319-26.

Zhang, W., S. Kadam, et al. (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**(7): 2413-22.

Zhao, Q., W. Zhou, et al. (2006). "Repression of human gamma-globin gene expression by a short isoform of the NF-E4 protein is associated with loss of NF-E2 and RNA polymerase II recruitment to the promoter." Blood **107**(5): 2138-45.

CHAPTER 1

A DUAL REPORTER MOUSE MODEL FOR IN VIVO AND EX VIVO STUDIES OF HUMAN γ-GLOBIN GENE EXPRESSION

PETROS PAPADOPOULOS, LAURA GUTIÉRREZ, REINIER VAN DER LINDEN, JOHN KONGASAN, ALEX MAAS, DUBRAVKA DRABEK, GEORGE PATRINOS, SJAAK PHILIPSEN AND FRANK GROSVELD

A dual reporter mouse model for *in vivo* and *ex vivo* studies of human γ–globin gene expression

Petros Papadopoulos, Laura Gutiérrez, Reinier van der Linden, John Kongasan, Alex Maas, Dubravka Drabek, George Patrinos, Sjaak Philipsen and Frank Grosveld

Abstract

Studying the regulation of human γ-globin gene expression has been a difficult task, since transcriptional regulation occurs at multiple levels. A reliable mouse model that facilitates high-throughput assays would simplify such studies, especially with the emergence of new libraries of chemical compounds, antibodies or shRNA target clones that might be of help to reveal crucial information about the contribution of protein complex(es) to the orchestration of γ-globin developmental expression patterns.

The expression of fluorescent proteins under specific regulatory sequences is a legitimate strategy that has been widely used in transcriptional reporter assays. We have generated a transgenic dual reporter mouse model by introducing at the γ- and β-globin genes the cDNA of GFP and DsRed fluorescence proteins respectively. This was done in the context of the whole human β-globin locus in order to meet the needs of a reliable *in vivo* and *ex vivo* reporter system. With this mouse model we aim to identify agents or proteins that could reactivate the γ-globin gene in the adult stage and thus provide novel therapeutic solutions for β-hemoglobinopathies.

Introduction

The human β-globin locus spans approximately 70kb containing the regulatory sequences of the Locus Control Region (LCR) and the β-like globin genes situated in the same order as they are expressed throughout ontogeny (5'- ε γ^G γ^A δ β -3'). Mice carrying a "minilocus" containing the essential distal regulatory elements surrounding the β-globin gene, express it at levels equivalent to the endogenous mouse β-globin and have given valuable information related to regulatory regions, position-independent and copy number-dependent expression (Grosveld, van Assendelft et al. 1987). Mice bearing the entire human β-globin locus have been a very useful model to understand developmental expression patterns of the five functional human globin genes (Strouboulis, Dillon et al. 1992). Combined studies on human and mouse globins have revealed common and different aspects of human and mouse hematopoiesis. While in human there are two globin switches (ε to γ, occurring in the transition from primitive to definitive erythropoiesis, and γ to β, occurring in definitive erythropoiesis around the time of birth), there is one main switch in mouse occurring at the time of transition between primitive and definitive erythropoiesis. Expression of the murine βminor and βmajor genes starts in the primitive cells and peaks at definitive stage when the embryonic ε and βh1 are not expressed any more (Trimborn, Gribnau et al. 1999). Early studies on the β-globin locus mice demonstrated expression of γ-globin in the embryonic stage as well as in the early fetal liver of the mouse whereas the β-globin gene was expressed in the fetal and adult stage (Strouboulis, Dillon et al. 1992).

The transition from γ- to β-globin during development involves both genes in a competitive model (Hanscombe, Whyatt et al. 1991; Patrinos, de Krom et al. 2004) of alternating transcription (Wijgerde, Gribnau et al. 1996). A great number of studies have confirmed the complexity of this process and have pointed out the need of a flexible and reliable model that facilitates high throughput analyses. This has been attempted by substituting the globin genes with fluorescent proteins under the regulatory elements of the β-globin locus (Skarpidi, Vassilopoulos et al. 2000). However, due to the length of the complete human β-globin locus, most of the reporter constructs used to date are partial representations of the locus that can be easily manipulated and introduced in erythroid cells by standard transfection techniques. Missing genomic sequences may affect the expression of the globins genes, limiting the value of transcription competition or reactivation studies, which

are best performed in the context of the complete locus (Vadolas, Wardan et al. 2002).

Here we describe the generation of two lines of transgenic mice carrying a full-length β-globin locus with dual fluorescent reporter genes, (γ-GFP | β-DsRed) and (γ-GPA-GFP | β-DsRed). These mice allow *in vivo* tracing of γ-globin gene expression during development by flowcytometry or fluorescent microscopy. They can be used to test potential treatments aimed at reactivating the expression of γ-globin in the adult stage. Additionally, we have generated fetal liver cell lines derived from transgenic mice for *in vitro* experiments, especially for functional screens with libraries of chemical compounds and shRNA clones.

Materials and Methods

Modification of the human β-globin locus in a PAC vector and generation of transgenic mice

The Aγ-globin and β-globin genes were modified at the first ATG of the transcript by introducing the EGFP-N2 or GPA-(EGFP-N2) and DsRed2 (700bp) cDNA respectively followed by a stop codon. Mouse Glycophorin A (GPA, 507bp) cDNA was cloned and modified by introducing the EGFP-N2 cDNA (720bp) just after 114bp downstream from the ATG by mutating a single base (Stratagene mutation kit) and thus creating a BamHI site. The modified globin genes subsequently replaced the endogenous ones of the human β-globin locus in the PAC2 vector by homologous recombination as described (Imam, Patrinos et al. 2000).

Fertilized oocytes from C57/Bl6 mice were injected with linearized DNA (Sce-I digest) of the modified β-globin locus and three transgenic lines that transmitted the transgene were generated, two of them from the EGFP-N2 and one from the GPA-(EGFP-N2) construct.

Flowcytometry analysis

Analysis of embryonic or adult blood and primary fetal liver cells or cell lines was performed with the FACSAria and FACScan (Becton Dickinson, BD). Primary cells or erythrocytes where resuspended in 1% BSA/PBS and in the case of fetal liver cell lines 7AAD or Hoechst (Molecular Probes) was used to stain for dead cells. Analysis of the data obtained was performed with FlowJo software. Antibodies CD71-PECy7 and CD117-APCCy7 were purchased from BD Biosciences.

Culture of primary fetal liver cells and cell line establishment

Fetal liver cells were cultured in Stem Pro media (Invitrogen) in the presence of erythropoietin (Epo, 1U/ml), stem cell factor (SCF, 100ng/ml) and dexamethasone (Dex, 10^{-6}M). All transgenic mice were bred with p53 knockout mice to homozygosity to generate fetal liver cell lines and WT cell lines from litter mates as described (von Lindern, Deiner et al. 2001). Differentiation assays were performed in the presence of Epo (10U/μl), human transferrin (SciPac, 500μg/ml) and cell size was monitored with a CASY cell Counter (Schärfe System). Within 48h the cells homogeneously reduced

their size and started producing hemoglobin. Globin expression of transgenic mice and cell lines was followed by flowcytometry, fluorescence microscopy IX-70 (Zeiss) or Confocal microscopy (Zeiss) and standard RNA analysis techniques.

Azacytidine treatment of transgenic mice

Transgenic and wild type mice from the same genetic background C57/Bl6 were treated as described (Rupon, Wang et al. 2006). Blood from all mice was resuspended in PBS and analysed with the FACSAria. Bone marrow was cultured in hanging drop cultures (Gutierrez, Lindeboom et al. 2005) in the presence of 5-Azacytidine (500nM) or hydroxyurea (100μM) (Mabaera, Greene et al. 2008) and analyzed by flowcytometry as aforementioned.

Transduction with virus

Knockdown experiments were performed in transgenic fetal liver cell lines and primary cells with concentrated virus for different target genes. The shRNA plasmids are part of the TRC1 Mission shRNA library (Sigma). 20μg of DNA plasmid together with 15μg of PAX-2 (gag) and 5μg of VSV-G were cotransfected in 293T cells with Polyethylenimine (PEI, Sigma) and supernatant was harvested three times consecutively every 24h followed by filtration and centrifugation at 20K rpm for 2h at 4^{o}C.

Fetal liver cells were transduced with virus and grown under selection of puromycin (1μg/ml) for a maximum of one week.

Western blots

Nuclear or whole cell protein extracts were semidry blotted to PVDF (Millipore) membrane, incubated for 1 hour at room temperature (RT) in 3% milk powder PBS-0.2%Tween20 blocking buffer and 1 hour RT incubation with the specific antibody in 1% milk PBS-0,2% Tween20, two washes in PBS-0,2%Tween20 and incubation with secondary antibody coupled to HRP in the same buffer as for the primary antibody in dilution 1/15000 with final four washes of 10min each. Finally ECL kit (Amersham Biosciences) was used to develop the membrane.

The antibodies used were cMyb (SantaCruz-516), Bcl11a (Santa Cruz-56013) Hdac3 (AbCam-32369), GFP (AbCam 250), DsRed (BDBiosciences, 632393) and Nucleophosmin (Abcam-10530).

Southern and Northern Blotting

Southern blotting was performed for mapping the modified β-globin locus after each recombination step to ensure integrity of the construct. DNA was digested with different restriction enzymes and run on 0.6% agarose gel. After semidry transfer of the DNA to Hybond N^+ membrane, hybridization was performed at 65°C with the two cosmid probes, cosLCR-ε and cosγγδβ (Strouboulis, Dillon et al. 1992) spanning the β-globin locus.

Northern blotting was performed to quantitate mRNA levels of GFP protein at different developmental stages in blood and fetal liver cells. RNA was extracted with Trizol (Invitrogen) and run on 1.2% formamide gel. Transfer of RNA to nylon membrane (Hybond N^+) was carried out in alkaline buffer (0.01N NaOH, 3M NaCl). Hybridization was performed at 65°C in SDS buffer (0.5M sodium phosphate pH 7.2, 7% SDS, 1mM EDTA pH 7.0). Consecutive washes of 20min each were done with 2x SSC, 0.5x SSC and 0.1% SDS. The radioactive filter was exposed and scanned in Typhoon Imager (Molecular Probes).

Results

Modification of the human β-globin locus and generation of transgenic lines carrying the dual reporter

In order to generate a mouse model where the expression of human globins can be followed by fluorescence, we introduced the cDNA of GFP protein (EGFP-N2, Clontech) in the first ATG codon of the Aγ-globin gene and the DsRed cDNA (DsRed2, Clontech) in the first ATG codon of the β-globin gene, followed both by a stop codon (Fig. 1A). The same strategy was followed to generate the second transgenic line, which has as γ-globin reporter a GFP fusion protein with the erythroid-specific membrane protein Glycophorin A (GPA) (Fig.1A-B). GPA was chosen in order to express the GFP protein on the red cell membrane rather than in the cytoplasm, since we observed fluorescence quenching due to high hemoglobin concentrations in the cytoplasm (data not shown). Since GPA-GFP has to be efficiently transported to the plasma membrane we tested the functionality of the fusion protein GPA-GFP by transducing a fetal liver cell line and MEL cells with a GPA-GFP lentivirus (Fig. 2B and data not shown). Expression of GFP in the plasma membrane of the cells was observed by confocal microscopy (Fig. 2B).

The modified human β-globin locus devoid of any sequences of the PAC vector (Sce-I digest) was injected in fertilized C57/Bl6 oocytes to generate transgenic mice. Extensive mapping of the modified locus was performed after each modification in bacteria and after birth of the founders by Southern Blotting (Fig. 1C) using as probes the two cosmids spanning the β-globin locus (70kb), cosLCRε and cosγγδβ, as described previously (Strouboulis, Dillon et al. 1992). Two transgenic lines γ-GFP β-DsRed (from here on referred to as GFPγ) were obtained with two and one copy of the transgene respectively. Three founders of the γGPA-GFP β-DsRed construct (from here on referred to as GPA-GFPγ) were obtained, one of which transmitted the transgene properly to the next generation and was found to carry two copies of the transgene.

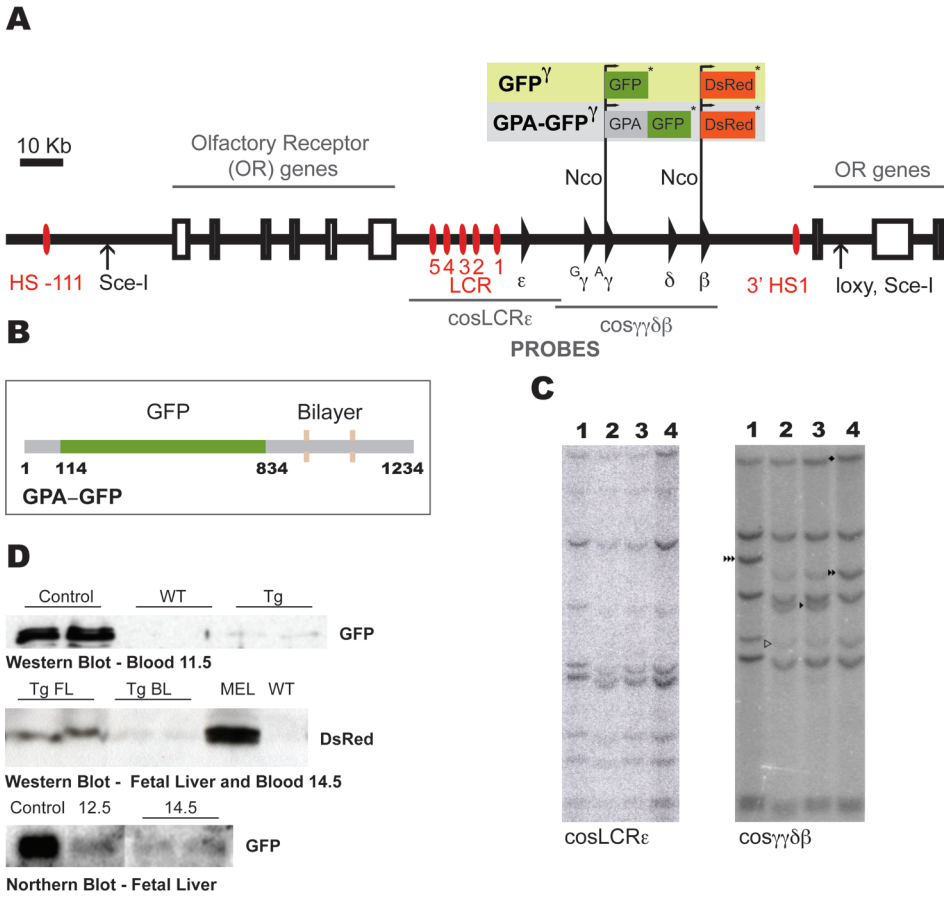


Figure 1

A. Genomic sequence of the human β-globin locus (Sce-I flanked) used as the basis for the modifications made in the γ- and β-globin genes. GFP and DsRed were introduced in the ATG (+1) position of the transcripts followed by a stop codon (*) respectively.

B. Schematic representation of the GPA-GFP construct.

C. Southern blot of both mouse transgenic lines (GFPγ and GPA-GFPγ). Tail genomic DNA was digested with SacI restriction enzyme and hybridized with cosLCRε (left) and cosγγδβ (right). Lane 1: GPA-GFPγ tail DNA, Lanes 2, 3: Cell line PAC8.1 carrying the human β-globin locus and Lane 4: GFPγ tail DNA. Symbol ▷ indicates end fragments, ▶ Aγ 3.6kb SacI fragment, ▶▶ Aγ-GFP 4.3kb SacI fragment, ▶▶▶ AγGPA-GFP 4.9kb SacI fragment, ◆ β-DsRed modification (16.4 to 17kb fragment).

D. Western blot for GFP protein from WT and GFPγ embryonic blood at 11.5*dpc* (control is I11 cells expressing GFP). Western blot for DsRed protein from WT and GFPγ embryonic blood and fetal liver at 14.5*dpc* (MEL are MEL cells expressing DsRed). Northern blot of GFP transcript from fetal livers at 12.5*dpc* and 14.5*dpc* of GFPγ embryos (control is I11 cells expressing GFP). 10μg RNA were loaded of each sample.

Human γ-globin gene expression in the transgenic mouse lines during development

RNA and protein were extracted from blood and fetal liver of embryos at different time points of development (11.5, 12.5 and 14.5*dpc*) from all transgenic lines and were analyzed for the presence of GFP transcript or protein. GFP transcript was detected in fetal liver, with a peak of γ-globin (GFP) expression at 11.5*dpc* that declines toward 14.5*dpc* (Fig. 1D). Western blot analysis of protein levels from the same developmental stages showed that there is detectable GFP protein in the blood at 11.5*dpc* (Fig. 1D) and 12.5*dpc*, but not at 14.5*dpc* (data not shown). DsRed protein was also detected by Western blot analysis in blood and fetal liver at 14.5*dpc*. These data confirmed that expression of the reporter (GFP) followed the developmental pattern of human γ-globin expression in the mouse. Therefore, we set up to study reporter expression by flowcytometry during development.

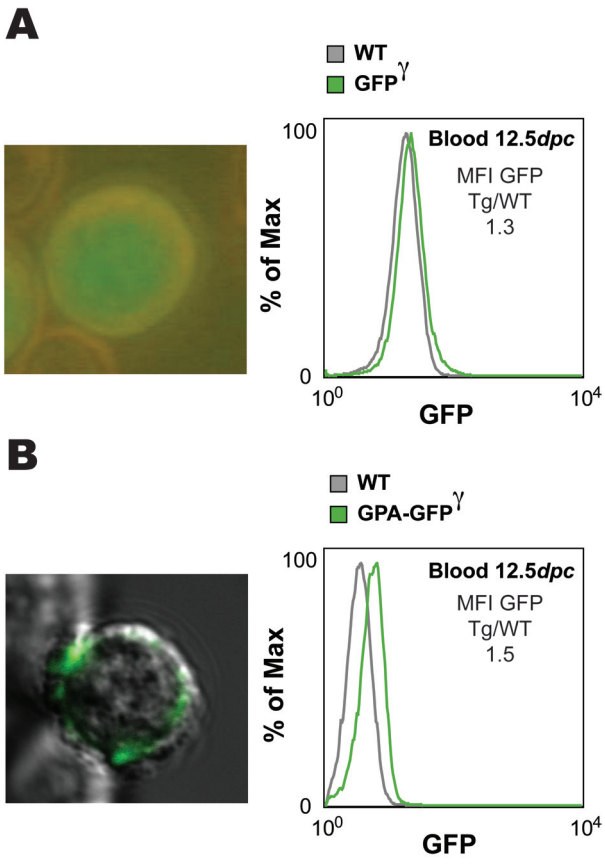
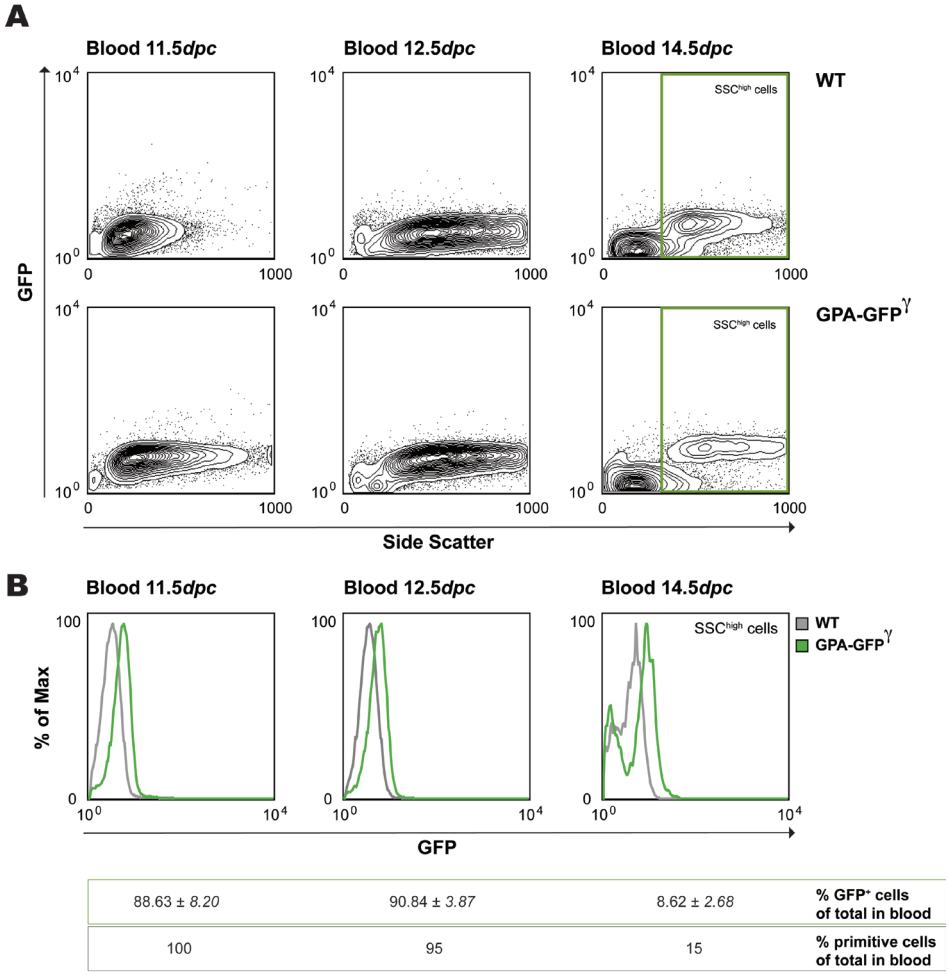


Figure 2

A. Representative picture of K562 cells transfected with the GFPγ modified human β-globin locus to check proper expression of γ-globin (left) and flowcytometry analysis of GFP expression in 12.5*dpc* embryonic blood of GFPγ transgenic embryos (right).

B. Representative picture of fetal liver cells transduced with the GPA-GFP construct to check expression of GFP protein in the cell membrane (left) and flowcytometry analysis of GFP expression in 12.5*dpc* embryonic blood of GPA-GFPγ transgenic embryos (right). Mean Fluorescence Intensity -MFI- ratio is indicated in both graphs.

As seen by flowcytometry analysis in Fig. 2 the difference between the two constructs (GFPγ and GPA-GFPγ) was significant with respect to GFP detection in 12.5*dpc* blood cells. There is a better separation between the WT and the transgenic GFP positive (γ-globin) cells as measured by the mean fluorescence intensity (MFI) ratio (1.5>1.3). This strengthens the notion that expressing the GFP protein on the cell surface, *i.e.* away from the cytoplasmic environment of the erythrocyte, facilitates the detection of fluorescence signals. For this reason, we chose to perform the developmental analysis of reporter expression in GPA-GFPγ transgenic mice.



88.63 ± *8.20*	90.84 ± *3.87*	8.62 ± *2.68*	**% GFP⁺ cells of total in blood**
100	95	15	**% primitive cells of total in blood**

Figure 3

A. Contour plots depicting embryonic blood during pregnancy time course of GPA-GFPγ transgenic mice (SSC: Side Scatter).

B. Histogram overlay of embryonic blood from transgenic and WT embryos in the GFP axis. The percentages of positive GFP cells and primitive cells for each developmental stage - as calculated by counting embryonic blood with the CASY counter - are included.

The flowcytometry analysis that was carried out in blood and fetal liver from embryos at different developmental stages was in concordance with previous studies of the human γ-globin gene expression in the mouse (Strouboulis, Dillon et al. 1992). We could detect GFP expression in practically all circulating erythroid cells at 11.5*dpc* and 12.5*dpc*, after which the percentage of GFP-positive cells started to decline. At 14.5*dpc* we could still detect GFP positive cells although their percentage was low (8.62%) and they were Side Scatter high (SSC^{high}) pointing out that these cells could be primitive cells still in the circulation (Fig. 3). We could not detect GFP protein in the fetal liver and we think that is due to its diverse cellular composition at the stages examined.

DsRed protein, which we included in our constructs as a control, was difficult to detect by flowcytometry. A small fraction of fetal liver cells (Fig. 4, arrow) is positive for DsRed at 11.5*dpc* and 12.5*dpc*. At 14.5*dpc,* when expression of β-globin gene is high, it is extremely difficult to detect DsRed positive cells. This result can be explained by the fact that hemoglobin absorbance also overlaps with the emission wavelength of DsRed.

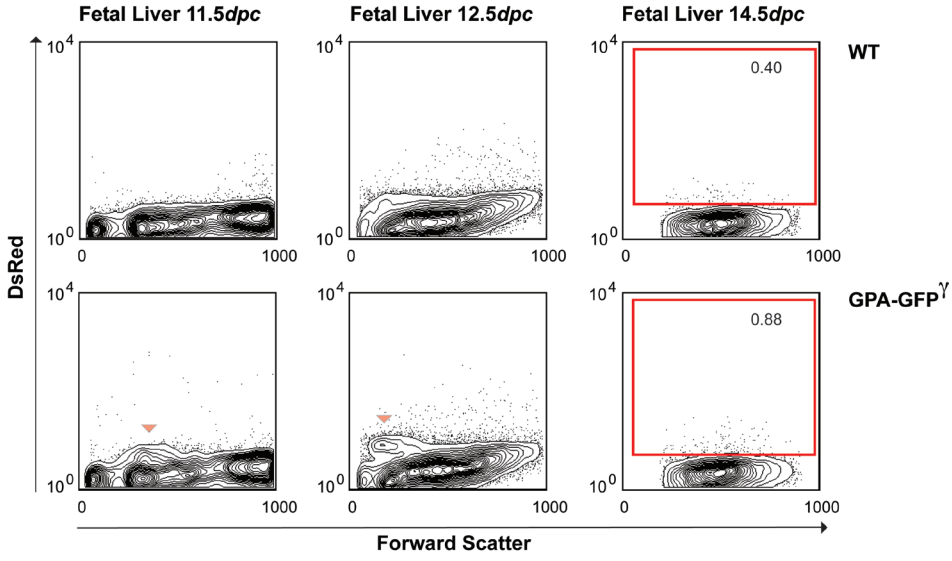


Figure 4

Flowcytometry analysis of fetal liver of GPA-GFPγ transgenic mice during development. Arrow at 11.5*dpc* and 12.5*dpc* indicates the DsRed positive population.

In vivo treatment of transgenic mice with 5-Azacytidine

Mice from the single copy GFPγ and double copy GPA-GFPγ as well as WT mice from the same genetic background C57/bL6 were treated with phenylhydrazine (PHZ) and 5-azacytidine (AZA) according to a previously published procedure (Rupon, Wang et al. 2006). The treatment consisted of PHZ intraperitoneal injections the first two days followed by AZA intraperitoneal injections the next five consecutive days. Control WT and transgenic mice were either injected with PHZ followed by PBS instead of AZA, or PBS only throughout the seven days of treatment.

At the end of the treatment, blood was analysed by flowcytometry and bone marrow cells were differentiated *ex vivo* in hanging drop cultures (Gutierrez, Lindeboom et al. 2005) prior to flowcytometry analysis.

Treated transgenic mice displayed GFP and DsRed positive cells as opposed to control transgenic mice, indicating that stress erythropoiesis caused by induced hemolytic anemia followed by AZA administration boosts the production of human globins in the mouse. In the GFPγ mice we detect separate emerging populations expressing either GFP or DsRed (Fig. 5A, green and red gates) whereas in the GPA-GFPγ mice the DsRed positive population contains a potential double positive (GFP/DsRed) subpopulation (Fig. 5A, yellow gate). This difference can be explained by the fact that an agent like AZA, which affects the morphology of the red cells *in vivo* and *in vitro* deregulating the expression of plasma membrane proteins (Appendix I), could possibly contribute to the difference observed between the two transgenic lines regarding the GFP positive population, since in the latter line GFP is fused to GPA.

The low percentage of F-cells (Boyer, Belding et al. 1975) expected in blood during a stress response correlates with the numbers of GFP positive cells detected by flowcytometry (Table 1). Furthermore, we are able to detect DsRed and GFP by flowcytometry directly in blood tissue, a very challenging task due to quenching and autofluorescence, and despite moderate expression levels of our reporter genes. This result proves our transgenic mouse model as a legitimate tool for the study of human γ-globin reactivation, with the advantage of avoiding tedious techniques for the analysis of globin expression.

Our results were further verified culturing the bone marrow cells of treated mice in hanging drop (HD) cultures where terminal differentiation occurs and globin expression is favored. The bone marrow cells were cultured in the presence of AZA

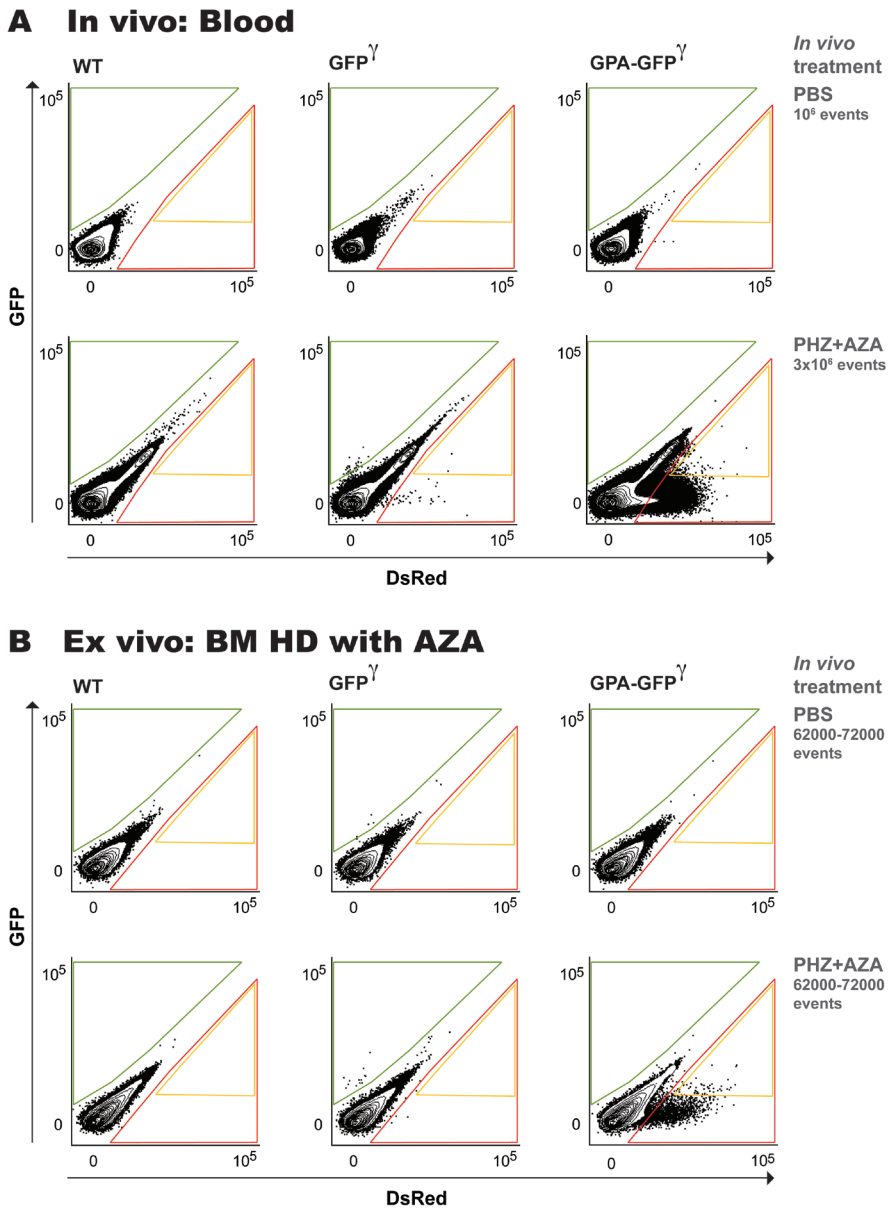


Figure 5

A. Flowcytometry analysis of AZA administration in transgenic mice (GFPγ and GPA-GFPγ). The upper three contour plots show background levels of fluorescence in peripheral blood upon PBS administration. The lower panel shows the response of WT, GFPγ and GPA-GFPγ mice upon PHZ and AZA administration.

B. Bone marrow hanging drop culture of the treated mice, presented in the same order as above (A). AZA was added to the culture medium in order to continue the treatment *ex vivo*. See tables 1 and 2 for the events recorded on each gate.

or HU (Fig. 5B and Table 2). On day two, the cells were analyzed by flowcytometry. The pattern was similar to that obtained with the peripheral blood (Table 2). Compared to whole blood, we required up to 50-fold fewer cells for flowcytometry analysis of the HD cultures, and the background or autofluorescence was reduced. The advantage that we gained using the HD culture is that fewer cells were required in order to get the same pattern of globin expression as measured *in vivo* but most importantly we foresee its application in high-throughput assays, such as screening for the effects of multiple chemical compounds using cells from one mouse.

Cell nr./million	DsRed		GFP		DsRed + GFP	
	PBS	PHZ -AZA	PBS	PHZ -AZA	PBS	PHZ -AZA
WT	0	0	0	0	0	0
	0	0	0	0	0	0
	0	0	0	0	0	0
GFP$^{\gamma}$	7	12	3	15	1	1
	1	14	0	10	0	0
	26	5	1	6	15	0
		38		8		1
GPA-GFP$^{\gamma}$	9	722	0	0	1	6
	15	390	0	0	4	6
	17	22346	3	3	1	92
		1055		0		29

Table 1: *In vivo* treatments

Transgenic fetal liver cell lines and experimental applications

With the purpose of having stable cell lines for molecular studies, we generated mouse fetal liver cell lines from transgenic embryos (11.5*dpc*-14.5*dpc*) and control (WT) cell lines from their litter mate embryos as described (von Lindern, Deiner et al. 2001). The cells can be kept in culture for long periods with serum free media supplemented with erythropoietin (Epo), stem cell factor (SCF) and dexamethasone (Dex). They divide as normal primary mouse fetal liver cultures and can be easily driven to differentiation pathway upon replacement of the aforementioned factors by

Epo and transferrin. Most importantly, the differentiation of the cells is homogeneous as reflected by size reduction towards terminal differentiation and production of adult mouse globins (βmaj, data not shown). Accordingly, cell lines generated from our transgenic mouse lines reduce their size (Fig. 6A) and express DsRed upon differentiation induction (Fig. 6B).

Flowcytometry analysis of specific erythroid cell membrane markers confirmed the early erythroid progenitor stage of the cells in culture ($CD117^{+}$ $CD71^{low}$). Upon differentiation, cKit (CD117) was downregulated and transferrin receptor (CD71) expression upregulated (Fig. 6A) whereas TER119 was almost undetectable before and after differentiation assays (data not shown).

We have been using the fetal liver cells (primary cells or the established cell lines) from both transgenic lines in several assays, in order to see their response upon treatment in culture and then try to reproduce the results *in vivo.* One of these assays involves shRNA clones for specific downregulation of potential target genes implicated in the regulation of γ-globin gene expression. We have tried several genes,

Cell nr./100000		DsRed		GFP		DsRed + GFP	
	HD	PBS	PHZ-AZA	PBS	PHZ-AZA	PBS	PHZ-AZA
WT	ST	0	0	0	0	0	0
	AZA	0	0	0	0	0	0
	HU	0	0	0	0	0	0
GFP^{γ}	ST	1	6	0	6	0	0
		3	11	0	2	0	3
	AZA	4	12	5	15	0	2
		3	11	7	13	0	4
	HU	3	5	0	3	1	0
		4	10	0	1	1	1
$GPA\text{-}GFP^{\gamma}$	ST	56	805	2	0	0	48
		5	2496	1	1	2	230
	AZA	107	718	0	0	4	41
		4	1692	2	1	1	167
	HU	52	622	0	1	0	49
		4	1600	0	1	0	114

Table 2: *Ex vivo* cultures and treatments

a few of them presented in Fig. 7. Knockdown (KD) of cMyb (Jiang, Best et al. 2006), Bcl11a (Sankaran, Menne et al. 2008) and Hdac3 (Mankidy, Faller et al. 2006) resulted in low or moderate induction of GFP expression, while Ikaros knockdown did not induce globin expression (data not shown). Knockdown efficiencies were confirmed by Western blot analysis (Fig. 7C).

cMyb is known to act as an inhibitor of terminal erythroid differentiation

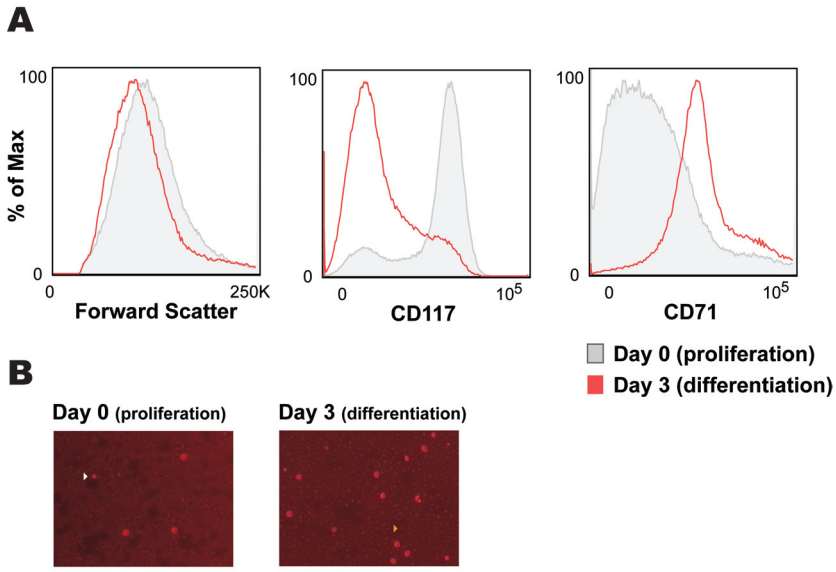


Figure 6

A. Flowcytometry analysis of transgenic fetal liver cell lines before and after differentiation. Histograms against forward scatter and erythroid surface markers CD117 (cKit) and CD71 (transferrin receptor) are depicted.

B. Representative pictures taken during erythroid differentiation of transgenic fetal liver cell lines. Arrows indicate spontaneously differentiating cells expressing DsRed protein (left) and differentiated cells with much smaller size that are not as bright as the bigger ones, most probably as a consequence of the continuous production of endogenous hemoglobins.

(Vegiopoulos, Garcia et al. 2006) . Knockdown of cMyb resulted in low levels of GFP (γ-globin) induction but also DsRed (β-globin) because the cells underwent differentiation as a result of reduced protein levels of cMyb. Conversely, overexpression of the cDNA of cMyb in the same cells delayed terminal differentiation *in vitro* (data not shown).

Bcl11a knockdown was the least efficient in terms of GFP level upregulation.

However, the GFP induction (2 fold) was not accompanied by DsRed expression, which makes it specific for γ-globin reactivation.

Hdac3 knockdown resulted in considerable upregulation of GFP levels in both transgenic cell lines and induction of DsRed to a lower extent. DsRed expression in cMyb and Hdac3 knockdown (KD) experiments is confined to different populations of cells when comparing the Forward Scatter contour plots. In the case of the cMyb KD cells, a distinct DsRed-positive population emerges in the low Forward Scatter area, a clear indication of cell differentiation. Distinctly, the DsRed positive cells in the Hdac3 KD stand on the mid-high Forward Scatter area, as occurs with GFP positive cells. This indicates that Hdac3 KD results in expression of globins desynchronized from cell maturation.

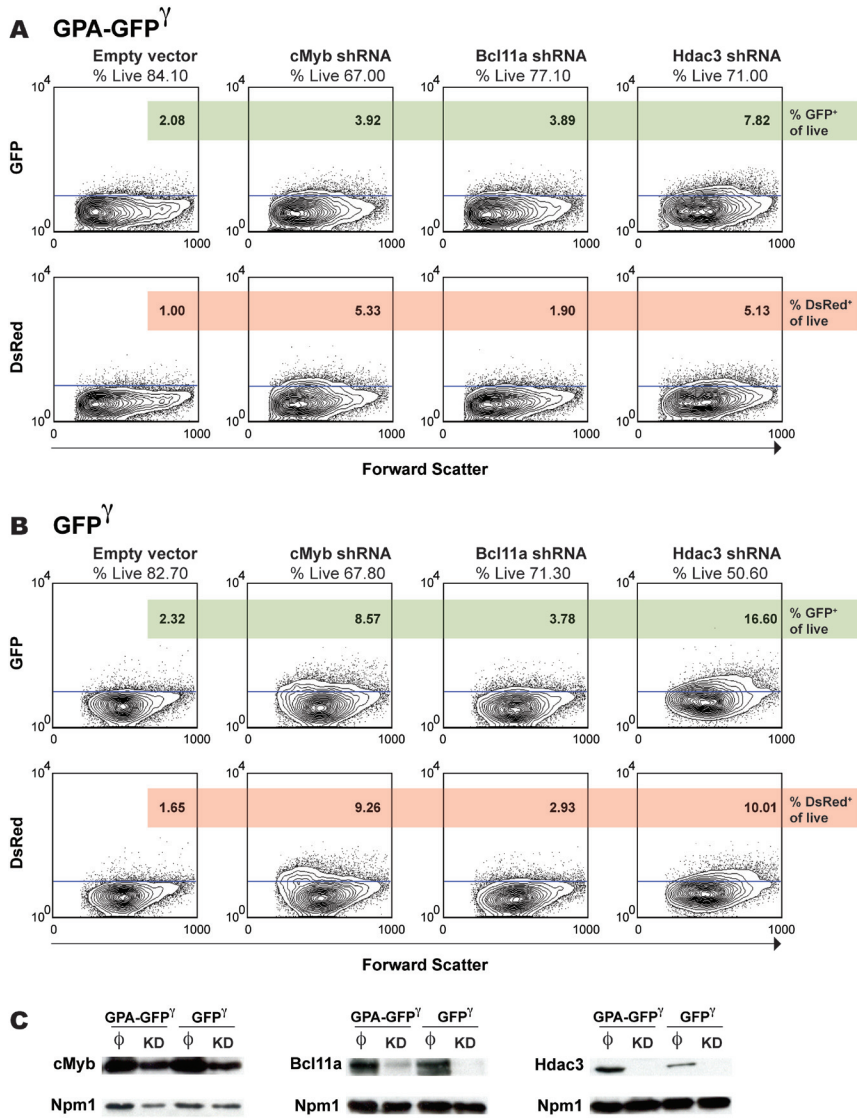


Figure 7

A. Flowcytometry analysis of the knockdown of cMyb, Bcl11a and Hdac3 in the GPA-GFPγ cell line. The same vector with a non-specific shRNA sequence was used as a control. Percentages of cells positive for GFP (upper panel) and DsRed (lower panel) are shown.

B. As in (A), Flowcytometry analysis was performed in the same way and for the knockdown of the same genes in the GFPγ cell line.

C. Western blots of the knockdown experiments in both transgenic cell lines. Equal numbers of cells are loaded on each lane. In the case of cMyb KD, due to induced differentiation after cMyb downregulation, nucleophosmin (Npm1) as an endogenous control is underrepresented (cells enucleate).

Discussion

In order to understand how globins are expressed during each developmental stage several *in vivo* and *in vitro* models have been generated. The complexity of the human globin switching from the fetal γ-globin to the adult β-globin has raised multiple questions that need to be simplified by flexible models and high throughput techniques. The erythroid cell lines available such as MEL and K562 have been extensively used and genetically manipulated to provide evidence of globin gene regulation in mouse and human *in vitro* models. The generation of transgenic mice bearing the human β-globin locus allowed dissection of the developmental pattern of globin gene expression and regulation, displaying many similarities but also differences to globin gene regulation in man (Strouboulis, Dillon et al. 1992; Sankaran, Xu et al. 2009). The use of reporter assays to measure the activity of specific promoters under certain conditions that mimic developmental programs of specific cell types is a classical strategy to tackle transcriptional regulation.

In the past, several constructs using fluorescent reporters have been used in the context of a mini-locus construct in mouse (GM979 (Skarpidi, Vassilopoulos et al. 2000)) and human cell lines (K562 (Vadolas, Wardan et al. 2002; Haley, Smith et al. 2003), and murine erythroleukemia cells (MEL (Tewari, Gillemans et al. 1996)) by standard transfection techniques. The main disadvantage of these studies lies on the position effect variegation of the transgene and in the integrity of the constructs after the transfection procedure when the whole β-globin locus is used. Another major point is the selective use of promoter regions and hypersensitive sites of the LCR for the generation of reporter constructs that in the case of the complex regulation of globin genes can be critical and suboptimal. A recent example is Bcl11a protein, which does not appear to bind the γ promoters but has a major effect in human γ-globin gene expression when it is knocked down in human proerythroblasts (Sankaran, Menne et al. 2008) and in knockout mice (Sankaran, Xu et al. 2009).

Here we tried to circumvent these problems by generating transgenic mice carrying a dual reporter β-globin locus, from which we derived fetal liver cell lines. Mice and cell lines were analyzed for proper globin gene regulation during development and integrity of the transgenes. Even more, we tried to avoid quenching of fluorescence due to the highly absorbent hemoglobin environment in the red cell by generating a new reporter mouse line, the GPA-GFPγ and its fetal liver cell line. The mouse GPA-GFPγ expression in the red cell outer membrane gave us many clues of what are the

limits and the sensitivity of our system *in vivo* and *in vitro*.

Expression of the human homologue of GPA was already tested in transgenic mice with no apparent phenotype for the mouse erythrocytes but underscoring the constant levels of GPA in the membrane of mature red cells as a result of tight coupling of Band 3 and GPA expression (Auffray, Marfatia et al. 2001). Furthermore studies on the role of cytoskeletal connectivity during erythroblast enucleation indicated that GPA was more connected to the membrane cytoskeleton in erythroblasts and young reticulocytes than in mature cells and that the presence of the negatively charged GPA along with CD47 possibly protects reticulocytes from phagocytosis following enucleation (Lee, Gimm et al. 2004). These observations were important for the analysis of the developmental expression of the transgenic GPA-GFPγ mice and the *in vivo* and *in vitro* treatment with chemical compounds.

The major problem of fluorescence quenching in the red cells is already known (Heck, Ermakova et al. 2003; Spangrude, Cho et al. 2006) and hard to overcome especially when moderate levels of expression are achieved. We experienced such problems when toxic agents like AZA or hydroxyurea (HU) were used or even with Hdac inhibitors. These agents changed the morphology of the cells dramatically and exerted significant levels of autofluorescence. This is a limitation that has to be always taken in consideration when such experiments have to be performed. But even in these cases we were able to detect real fluorescence as in the *in vivo* experiment with PHZ/AZA. That is why both constructs were crucial in order to get the most information out of any treatment *in vivo* or *in vitro* always having in consideration the tremendous changes taking place in the cytoplasm and plasma membrane of the red cells during the treatment/maturation process. To summarize, the GPA-GFP protein placed in the plasma membrane when expressed *in vivo* in embryonic blood was better for detection, but less helpful when cytotoxic agents like AZA or HU were used *in vivo* and *in vitro* (data not shown). Thus, a combination of the two cell lines upon treatment with chemical agents would be more informative, whereas for *in vivo* experiments the GPA-GFPγ mouse is a more sensitive model for fluorescence signal detection.

The potential use of our system is broad and has the unique advantage that allows sorting of positive GFP (γ-globin) cells and further analysis at the single cell level, which is essential if only a few cells respond to a treatment *in vivo* or *in vitro*. Screening of libraries of chemical compounds, shRNA clones or synthetic antibodies are unbiased approaches that can be applied to our dual reporter β-globin locus

cell lines, which provide the tools to identify changes at the molecular level in a population of cells or at single cell level using high-throughput assays.

Additionally, the generated transgenic fetal liver cell lines allow the study of transcription factor balance during erythroid differentiation with assays that require a large amount of cells. This is of great advantage, since these type of studies involving weak interactions of low abundance transcription factors with the basal transcriptional machinery have been of great interest and have revealed new concepts about pre- and postnatal gene regulation in different tissues (Deato, Marr et al. 2008; Sengupta, Cohet et al. 2009). With mouse and cell line models presented here, such studies on the whole human β-globin locus are feasible and offer a user-friendly read-out thanks to the fluorescent reporters.

References

Auffray, I., S. Marfatia, et al. (2001). «Glycophorin A dimerization and band 3 interaction during erythroid membrane biogenesis: in vivo studies in human glycophorin A transgenic mice.» Blood 97(9): 2872-8.

Boyer, S. H., T. K. Belding, et al. (1975). "Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults." Science 188(4186): 361-3.

Deato, M. D., M. T. Marr, et al. (2008). "MyoD targets TAF3/TRF3 to activate myogenin transcription." Mol Cell 32(1): 96-105.

Grosveld, F., G. B. van Assendelft, et al. (1987). "Position-independent, high-level expression of the human beta-globin gene in transgenic mice." Cell 51(6): 975-85.

Gutierrez, L., F. Lindeboom, et al. (2005). "A hanging drop culture method to study terminal erythroid differentiation." Exp Hematol 33(10): 1083-91.

Haley, J. D., D. E. Smith, et al. (2003). "Identification and characterization of mechanistically distinct inducers of gamma-globin transcription." Biochem Pharmacol 66(9): 1755-68.

Hanscombe, O., D. Whyatt, et al. (1991). "Importance of globin gene order for correct developmental expression." Genes Dev 5(8): 1387-94.

Heck, S., O. Ermakova, et al. (2003). "Distinguishable live erythroid and myeloid cells in beta-globin ECFP x lysozyme EGFP mice." Blood 101(3): 903-6.

Imam, A. M., G. P. Patrinos, et al. (2000). "Modification of human beta-globin locus PAC clones by homologous recombination in Escherichia coli." Nucleic Acids Res 28(12): E65.

Jiang, J., S. Best, et al. (2006). "cMYB is involved in the regulation of fetal hemoglobin production in adults." Blood 108(3): 1077-83.

Lee, J. C., J. A. Gimm, et al. (2004). "Mechanism of protein sorting during erythroblast enucleation: role of cytoskeletal connectivity." Blood 103(5): 1912-9.

Mabaera, R., M. R. Greene, et al. (2008). "Neither DNA hypomethylation nor changes in the kinetics of erythroid differentiation explain 5-azacytidine's ability to induce human fetal hemoglobin." Blood 111(1): 411-20.

Mankidy, R., D. V. Faller, et al. (2006). "Short-chain fatty acids induce gamma-globin gene expression by displacement of a HDAC3-NCoR repressor complex." Blood 108(9): 3179-86.

Patrinos, G. P., M. de Krom, et al. (2004). "Multiple interactions between regulatory regions are required to stabilize an active chromatin hub." Genes Dev 18(12): 1495-509.

Rupon, J. W., S. Z. Wang, et al. (2006). «Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice.» Proc Natl Acad Sci U S A 103(17): 6617-22.

Sankaran, V. G., T. F. Menne, et al. (2008). "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A." Science 322(5909): 1839-42.

Sankaran, V. G., J. Xu, et al. (2009). "Developmental and species-divergent globin switching are driven by BCL11A." Nature 460(7259): 1093-7.

Sengupta, T., N. Cohet, et al. (2009). "Distinct modes of gene regulation by a cell-specific transcriptional activator." Proc Natl Acad Sci U S A 106(11): 4213-8.

Skarpidi, E., G. Vassilopoulos, et al. (2000). "Novel in vitro assay for the detection of pharmacologic inducers of fetal hemoglobin." Blood 96(1): 321-6.

Spangrude, G. J., S. Cho, et al. (2006). "Mouse models of hematopoietic engraftment: limitations of transgenic green fluorescent protein strains and a high-performance liquid chromatography approach to analysis of erythroid chimerism." Stem Cells

24(9): 2045-51.
Strouboulis, J., N. Dillon, et al. (1992). “Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice.” Genes Dev 6(10): 1857-64.
Tewari, R., N. Gillemans, et al. (1996). “The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene.” Development 122(12): 3991-9.
Trimborn, T., J. Gribnau, et al. (1999). “Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci.” Genes Dev 13(1): 112-24.
Vadolas, J., H. Wardan, et al. (2002). “Development of sensitive fluorescent assays for embryonic and fetal hemoglobin inducers using the human beta -globin locus in erythropoietic cells.” Blood 100(12): 4209-16.
Vegiopoulos, A., P. Garcia, et al. (2006). “Coordination of erythropoiesis by the transcription factor c-Myb.” Blood 107(12): 4703-10.
von Lindern, M., E. M. Deiner, et al. (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(28): 3651-64.
Wijgerde, M., J. Gribnau, et al. (1996). “The role of EKLF in human beta-globin gene competition.” Genes Dev 10(22): 2894-902.

CHAPTER 2

Characterization of canonical and non-canonical TFIID complexes during erythroid cell differentiation

Petros Papadopoulos, Jeroen Demmers, Ellen van Drunen, Laszlo Tora and Frank Grosveld

Characterization of canonical and non-canonical TFIID complexes during erythroid cell differentiation

Petros Papadopoulos, Jeroen Demmers, Ellen van Drunen, Laszlo Tora and Frank Grosveld

Abstract

Transcription is an orchestrated process that requires the sequential recruitment of general transcription factors (GTFs) and the RNAPolII in the core promoter which serves as a platform for the assembly of the preinitiation complex (PIC) to collectively specify the transcription start site. TFIID is the first GTF recruited during the PIC formation by binding to the TATA box, initiatiator and/or downstream promoter element (DPE) found in most core promoters and is composed by the TATA-binding protein (TBP) and the 14 TBP-associated factors (TAFs). TAFs are providing interacting surfaces for activator dependent transcription and some of them exert tissue and promoter specificity. Recently it was shown that one of the TAFs, TAF10, has a role in the postnatal regulation of genes expressed in the liver and thus, we were prompted to study TAF10 in the regulation of fetal hemoglobin. Knockdown of TAF10 in mouse and human erythroid progenitors indicated a possible role in a hypothetical repressor complex responsible for the silencing of γ-globin in the adult stage. TAF10 immunoprecipitations (IPs) also revealed interactions with various regulators and coregulators in different stages of development and during erythroid differentiation that should be further characterized in respect to globin expression.

Introduction

Temporal and spatial transcriptional activation of different genes in eukaryotes and especially in metazoans is very diverse, as illustrated by the high level of control for each step required, starting from the formation of the preinitiation complex (PIC) to elongation and reinitiation of transcription. PIC formation requires the recruitment of the TFIID complex to the core promoter region, followed by the sequential entry of TFIIA and TFIIB, that stabilize the TFIID-promoter assembly, and then PolII/TFIIF, TFIIE, and TFIIH (Thomas and Chiang 2006). TFIID is a multisubunit complex, containing the central TATA binding protein (TBP) and at least 14 associated factors (TAFs) (Dynlacht, Hoey et al. 1991; Tora 2002). The interplay of different TAFs with transcriptional regulators and coregulators defines promoter specificity and tissue specific gene expression (Hochheimer and Tjian 2003; Hiller, Chen et al. 2004). In addition, TAF1 is known to function as a histone acetyltransferase (HAT) that appears to be promoter specific when acetylating histone H3 and H4 in human, Drosophila and yeast. TAF1 also functions as a kinase that phosphorylates histone H2B and as a histone ubiquitin-activating /conjugating enzyme that mediates monoubiquitylation of linker H1 both *in vivo* and *in vitro* in Drosophila (Thomas and Chiang 2006).

TAF10 is one of the smallest subunits (30KD) of the TFIID complex, and other multiprotein complexes, named SAGA complexes that include the TBP-free TAF-containing complex (TFTC), p300/CBP-associated factor (PCAF) complex and the SPT3-TAF9-GCN5-containing complex (STAGA) in mammalian cells and SPT-ADA-GCN5 acetylase (SAGA) coactivator complex in yeast (Nagy and Tora 2007). The TFTC complex was shown to support transcription in TATA-containing and TATA-less promoters *in vitro*, presumably via recognition of the core promoter by TAFs and their interaction with other components of the general transcriptional machinery (Wieczorek, Brand et al. 1998). Apart from the HAT component of these complexes which confer acetylation of histones and nucleosomes, E3 ligase activity has been reported for PCAF and recently a subcomplex of TFTC/STAGA which removes the ubiquitin moieties from histones H2A and H2B (Zhao, Lang et al. 2008). All these histone modifications are tightly connected with the establishment of a euchromatic state that promotes interactions with coregulators able to remodel the chromatin architecture of specific promoters and thus assemble a stable PIC complex to carry out transcription efficiently. Furthermore, recent genome-wide studies have revealed specific active histone marks, including trimethylation of histone H3K4 (H3K4me3)

and acetylation of H3K9 and K14 (Kim, Barrera et al. 2005; Heintzman, Hon et al. 2009) that have been shown to interact with transcription factors via different domains, *e.g.* the plant homeodomain (PHD) of TAF3, which binds to H3K4me3 mark, leading to transcriptional activation (Vermeulen, Mulder et al. 2007).

The existence of alternatives beyond the 'canonical' TFIID supported transcription, like the TBP-free complexes, broadened the limits of diversity in transcriptional activation and control. The evidence came through the demonstration of a gradual take-over process of TFIID by a TBP-related factor complex (TAF3-TRF3) driving specific transcriptional programs during differentiation of myoblasts to myotubes (Deato and Tjian 2007). TBP-related factor 2 (TRF2), also known as TBP-like factor (TLF) or TBP-like protein (TLP), and the TBP-related factor 3 (TRF3) are two TBP-homologues (TRFs) present in eukaryotes. TRF3 is unique to vertebrates and ubiquitously expressed, and is involved in TATA recognition, while TRF2 binds an undefined element, and is expressed in selective tissues (Thomas and Chiang 2006). In addition, TRF3 bears a divergent N-terminal domain that may execute distinct gene selective functions by patterning with different associated subunits (TAFs) to acquire differential promoter selectivity, as was shown for skeletal myogenesis.

Here, we explore the role of canonical and non-canonical TFIID complexes during the differentiation of erythroid cells from the perspective of TAF10, as a member of the TFIID and SAGA complexes, as well as for its unique role in TFIID structural stability and postnatal implications in gene expression (Tatarakis, Margaritis et al. 2008). We foresee a role for TAF10 in γ-globin expression in mouse and human erythroid progenitors and we aim at the identification of distinct partners, regulators and coregulators that orchestrate the transcriptional activity of globin promoters and, in particular, of the γ-globin gene.

Materials and Methods

Knockdown of TAF10 in mouse fetal liver cells and human blood erythroid progenitors

Different shRNA plasmids (TRC1 Mission shRNA Library, Sigma) against mouse and human TAF10 were tested. 20μg of DNA plasmid together with 15μg of PAX-2 (gag) and 5μg of VSV-G were cotransfected in 293T cells with Polyethylenimine (PEI, Sigma) and supernatant was harvested three consecutive times every 24hr, followed by filtration and centrifugation at 20K rpm for 2hr at 4°C.

Fetal liver cells were transduced with virus and grown under selection of puromycin (1μg/ml) for a maximum of one week. Human cells were selected at lower concentrations of puromycin (250ng/ml).

Flowcytometry

Mouse fetal liver cells were washed twice in 0.5% Fetal Calf Serum (FCS)- Phosphate Buffer Saline (PBS), and resuspended in the same buffer with Hoeschst (Molecular Probes) for dead cell staining prior to acquisition by FACSAria.

DNA FISH analysis

Chromosome preparations were made according to standard conditions. FISH was carried out as described (Mulder, Wilke et al. 1995). DNA was counterstained with DAPI and LCRεcos and cosγγδβ as probes were coupled to FITC and Texas Red.

S1 nuclease protection assay

As described in (Wijgerde, Grosveld et al. 1995).

Human peripheral blood erythroid progenitor cell culture

Buffy coat from human peripheral blood was Ficoll (Lymphoprep) fractionated and erythrocytes were separated from the mononuclear cells. The mononuclear cell fraction was collected, washed in HBSS medium and cultured under high cellular density in serum free media (StemSpan, StemCell Technologies) supplied with stem cell factor (SCF, 100ng/ml), erythropoietin (Epo, 1U/ml) and dexamethasone (Dex, Sigma, 10^{-6}M). Erythroid progenitors grow out after 3-4 days and cells can be

expanded for up to 3 weeks. In addition, cells can be driven to terminal differentiation changing the factors provided in the media by high Epo concentration (5-10U/μl) and transferrin (50μg/ml) within 2-4 days. Cultured erythroid progenitors were analyzed by flowcytometry to verify their homogeneity and erythroid origin and their phenotype was found to be $CD71^{high}$, $CD34^{low}$, $TER119^{low}$, $B220^{neg}$ (data not shown).

Extract preparation, immunoprecipitation assays (IPs) and Mass Spectrometry

Nuclear extracts from mouse and human fetal liver cells were prepared as described (Tatarakis, Margaritis et al. 2008) and IPs and westerns were performed as described (Kouskouti, Scheer et al. 2004). The TAF10 antibody was provided by Dr. L. Tora (Jacq, Brou et al. 1994) and anti-GST (sc-80004, Santa Cruz) was used as mock antibody. Other antibodies used for western blot analysis are: TAF4 (32TA-2B9, Eurogentec), TBP (3TI-3G3, Eurogentec), BCL11A (sc-56013 ,Santa Cruz), Actin (ab1801, Abcam) and Nucleophosmin (ab10530, Abcam).

Mass spectrometry was performed in LTQ-Orbitrap (Thermo).

AFP and GIP treatment of mouse and human erythroid cells

Cells were cultured in media with α-fetoprotein (AFP, ab3819, AbCam) or growth inhibitory peptide (GIP), kindly provided by Dr. G.J. Mizejewski (Wadsworth Center, Albany, NY), up to one week with daily addition of fresh medium and recombinant proteins in different dilutions (10^{-6}M, 10^{-8}M, 10^{-10}M).

Results

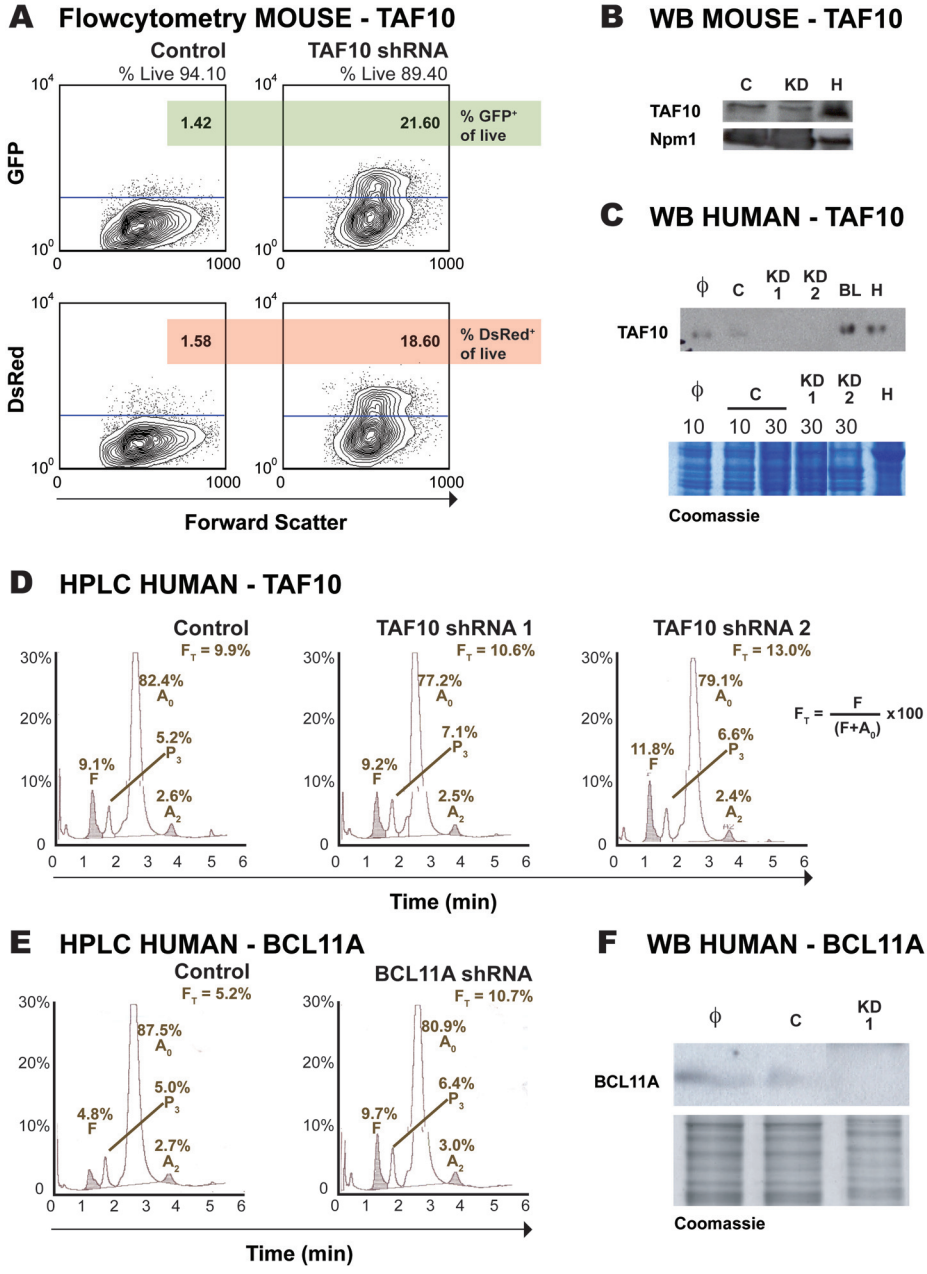
Knockdown (KD) of TAF10 affects globin gene expression in mouse and human erythroid progenitors

Transgenic fetal liver cell lines (DsRedβCL) carrying a modified human β-globin locus with two fluorescent proteins, GFP and DsRed, introduced in the start site of the γ- and β-globin genes respectively (Chapter 1), were transduced with several shRNA plasmids against TAF10. At least one of them resulted in downregulation of TAF10 protein levels (Fig. 1B) accompanied by the previously observed increase in apoptosis and cell cycle arrest when compared to the control (no-target shRNA) transduced cells and non-transduced cells (Metzger, Scheer et al. 1999). Flowcytometry (FACS) analysis showed upregulation of GFP (γ-globin) and DsRed (β-globin) fluorescence in gated live cells consistently (Fig. 1A).

Slightly different results were obtained when knockdown (KD) of TAF10 was obtained in human peripheral blood progenitors. Protein prepared from nuclear extracts was analyzed by western blot (Fig. 1C) to check levels of TAF10 protein. Equal cell numbers were loaded on each lane and up to three times more in separate lanes for the knockdown samples for comparison purposes. TAF10 was downregulated at the protein level efficiently in human erythroid cells. In concordance with the mouse fetal liver progenitors, after the third day of transduction, the majority of TAF10 KD human erythroid cells underwent growth arrest and apoptosis. The high performance liquid chromatography analysis (HPLC) showed at best a small increase of fetal hemoglobin (HbF) levels in TAF10 KD cells compared to the control at the expense of adult hemoglobin (HbA) (Fig.1D). A considerable fraction of hemoglobin was not classified (P_3) which might represent modified hemoglobin. These results coincide with the microarray data derived from TAF10KO fetal liver at 18.5*dpc*, in which fetal hemoglobin (Hbb-bh1) is 2-fold upregulated in KO fetal liver cells (Supplementary

Figure 1

A. Flowcytometry of mouse TAF10 knockdown and **B.** Western blot of TAF10 in DsRedβCL. **C.** Human TAF10 knockdown in blood erythroid cells was analyzed by Western blot**. D.** HPLC analysis was performed for HbF levels**. E.** As a control, BCL11A knockdown in human cells was analyzed in the same way, HPLC and **F.** Western blot**.** Non-transduced cells (Φ), control shRNA of TRC Mission Library (c), shRNA plasmids (KD), Hela cells (H), blood erythroid cells (BL). Coomassie blue staining for protein loading was performed and three times the control volume (30μl/10μl) was loaded in the knockdown samples from extracts with equal number of cells per volume made (nr. cells/vol).


A Flowcytometry MOUSE - TAF10
Control
% Live 94.10
TAF10 shRNA
% Live 89.40
1.42
21.60
% GFP+ of live
GFP
1.58
18.60
% DsRed+ of live
DsRed
Forward Scatter
B WB MOUSE - TAF10
C KD H
TAF10
Npm1
C WB HUMAN - TAF10
ϕ C KD 1 KD 2 BL H
TAF10
ϕ C KD 1 KD 2 H
10 10 30 30 30
Coomassie
D HPLC HUMAN - TAF10
Control
F_T = 9.9%
82.4% A0
5.2% P3
9.1% F
2.6% A2
TAF10 shRNA 1
F_T = 10.6%
77.2% A0
7.1% P3
9.2% F
2.5% A2
TAF10 shRNA 2
F_T = 13.0%
79.1% A0
6.6% P3
11.8% F
2.4% A2
F_T = F/(F+A0) x100
Time (min)
E HPLC HUMAN - BCL11A
Control
F_T = 5.2%
87.5% A0
5.0% P3
4.8% F
2.7% A2
BCL11A shRNA
F_T = 10.7%
80.9% A0
6.4% P3
9.7% F
3.0% A2
Time (min)
F WB HUMAN - BCL11A
ϕ C KD 1
BCL11A
Coomassie

data, Table 2 (Tatarakis, Margaritis et al. 2008)).

In order to compare the effect of TAF10KD with a known HbF regulator, we performed knockdown of BCL11A using the recently published shRNA target sequence (Sankaran, Menne et al. 2008) in parallel experiments and protein levels were analysed by western blot (Fig. 1F). We observed a two fold increase of HbF levels in BCL11A KD in human blood progenitor cells by HPLC analysis (Fig. 1E, F). These results suggest a potential role of TAF10 in γ-globin gene expression regulation.

Generation of a GFP (γ-globin) spontaneously expressing fetal liver cell clone (GFPγCL)

Wild type (WT) fetal liver cells from p53-/- mice were transfected (Nucleofector,Lonza) with linear DNA of the 165Kb modified β-globin reporter locus PAC (Chapter 1) and after serial dilutions several clones were obtained for further analysis of transgene integrity. Most of the lines expressed DsRed and the clone used for further experiments was named DsRedβCL. Exceptionally, one of the clones exerted an interesting expression pattern in proliferating conditions with a moderate 5-15% of GFP positive cells (γ-globin expression). Upon differentiating conditions almost hundred percent (total shift in the histogram) of the cells express GFP in contrast to the mouse derived fetal liver cell line DsRedβCL (Fig. 2). Therefore, the GFP expressing cell line is named GFPγCL. The clone was mapped for the transgene and appeared to have the correct DNA restriction digestion pattern (Fig. 3A) and single copy integration after Southern blotting analysis. In addition, DNA fluorescence in situ hybridization (FISH) confirmed the single integration site of the transgene (Fig. 3B). PCR analysis for 3'HS1 failed to confirm its presence and because the region is rich in repetitive Alu sequences we could only define a deletion of approximately 10kb by PCR analysis. However, this deletion does not explain the γ-globin expression because previous reports of generation of transgenic mice with the human β-globin locus devoid of the 3'HS1 region (Strouboulis, Dillon et al. 1992) showed proper expression of globins during development. The integration site of the transgene is not known, but when endogenous globin expression was tested upon differentiation by S1 analysis, adult βmaj-globin and not embryonic bh1 was expressed as expected (Fig. 3C) for the mouse globins, indicating a "local" effect in the γ-globin promoter.

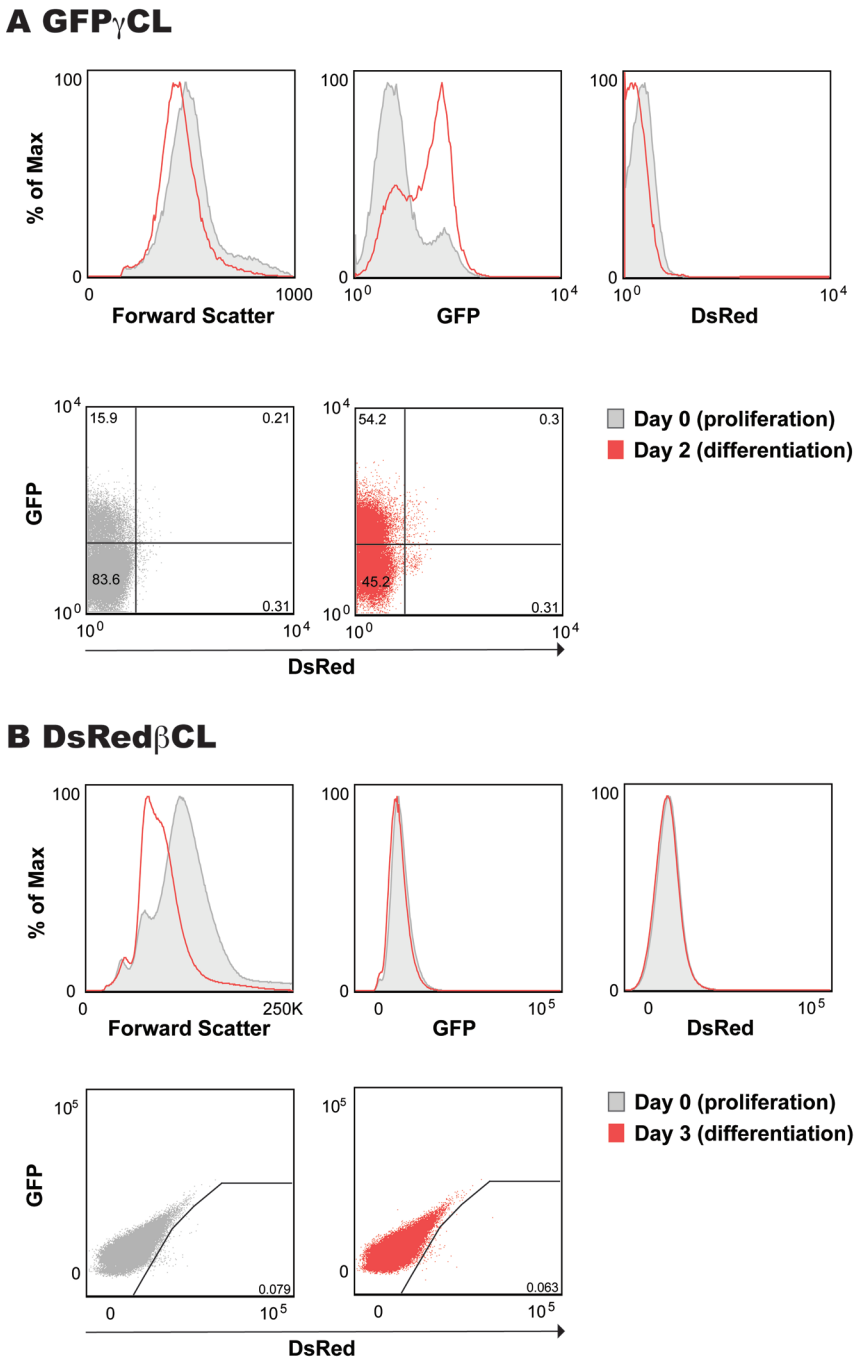


Figure 2

A. Flowcytometry analysis of GFPγCL versus **B.** DsRedβCL. GFPγCL expresses GFP at almost 100% upon differentiation in contrast to the DsRedβCL.

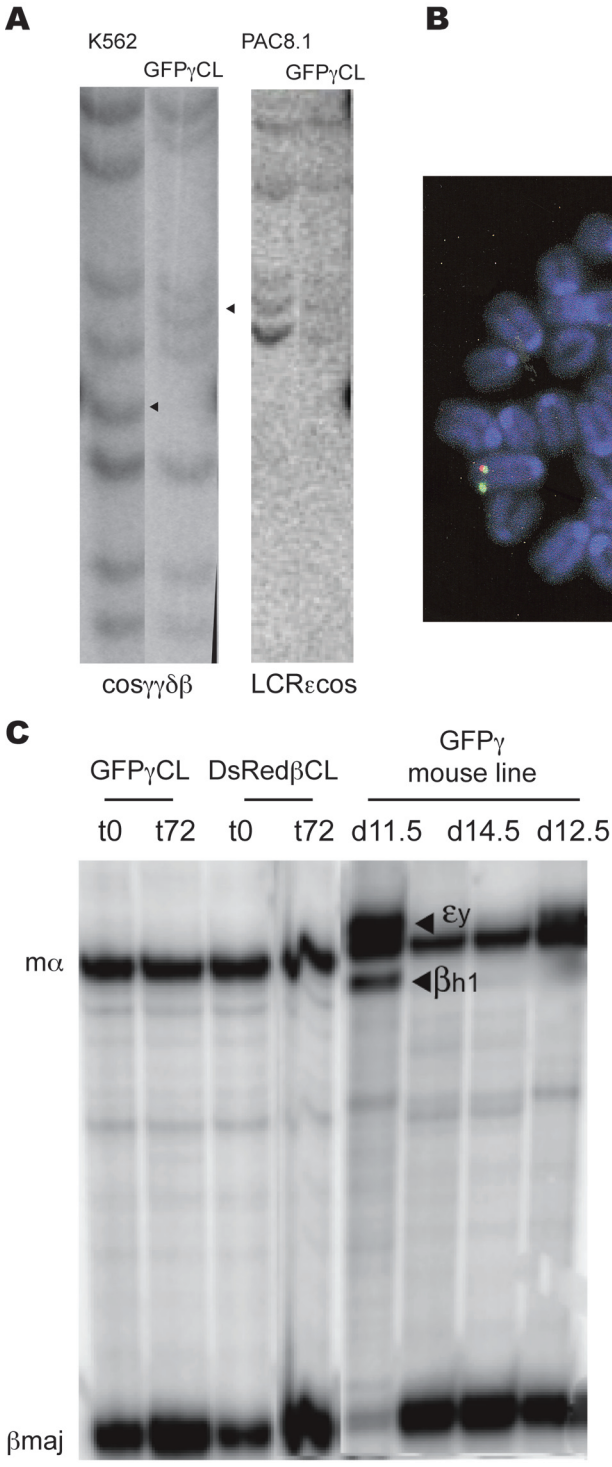


Figure 3

A. Southern blot of the GFPγCL with the two cosmids (LCRε and γγδβ) spanning 70kb of the β-globin locus. Erythroid cell lines K562 and PAC8.1 (carrying the human β-globin locus) are used as controls. **B.** Fluorescence *In situ* hybridization (FISH) was performed with the aforementioned probes coupled to FITC and Texas Red respectively in metaphase spreads of GFPγCL to identify the integration site of the modified β-globin locus transgene. **C.** S1 nuclease analysis of GFPγCL and DsRedβCL before (t=0) and after differentiation (t=72hr), and GFPγ transgenic mouse embryonic blood from different developmental stages. The mouse globins produced are indicated.

α-fetoprotein (AFP) is deregulated in GFPγCL

In order to identify transcriptional changes characterizing GFPγCL, microarray analysis was performed against several fetal liver cell lines generated from transgenic mice with the modified β-globin locus during standard growing conditions and upon stimulation with different γ-globin reactivating chemical agents (Azacytidine, Hydroxyurea) (APPENDIX 1). We observed extremely high levels (~200 fold) of α-fetoprotein (AFP) in GFPγCL when compared to the other cell lines. AFP is a postnatally silenced gene, as γ-globin. Interestingly, AFP is derepressed in TAF10KO fetal liver at day 30 after birth (Tatarakis, Margaritis et al. 2008). Even though the mouse fetal liver cell lines are established in p53-/- background, a repressor of AFP (Nguyen, Cho et al. 2005), it was tempting to test whether the high levels of AFP protein would have any effect on HbF production in human blood progenitors. We cultured human fetal liver and blood progenitors in the presence of recombinant AFP or the smaller growth inhibitory peptide (GIP) buried in a molecular crevice of the AFP, which exhibits growth inhibitory properties (Mizejewski and MacColl 2003) at different concentrations (10^{-6}M, 10^{-8}M, 10^{-10}M) as described. HPLC analysis after three and seven days of culture did not show any significant change at the levels of HbF in the cells (data not shown). Interestingly, when GFPγCL was treated with the GIP peptide in the same concentrations, we observed an increase in GFP positive cells in proliferating conditions, however those cells expressed reduced levels of GFP as measured by the Mean Fluorescence Intensity (MFI) (Fig. 4). Furthermore, we did not observe inhibition of growth as measured by counting cells through cultures or measuring cell death by flowcytometry (data not shown).

Immunoprecipitation of TAF10 reveals differences in TAF-containing complexes upon differentiation of erythroid cells and expression of specific globin genes

We asked whether TAF10, as a component of different TAF-containing complexes, would bring up specific interactions with other regulators and coregulators reflecting the developmental stage and the globin genes transcriptional activation. With this purpose, immunoprecipitation assays (IPs) were performed in the two cell lines of mouse transgenic fetal liver cells, GFPγCL and DsRedβCL cell line, with TAF10 antibody and GST antibody for the mock IP. Nuclear extracts from both cell lines were used in non-induced (proerythroblast stage) and induced state and the Protein

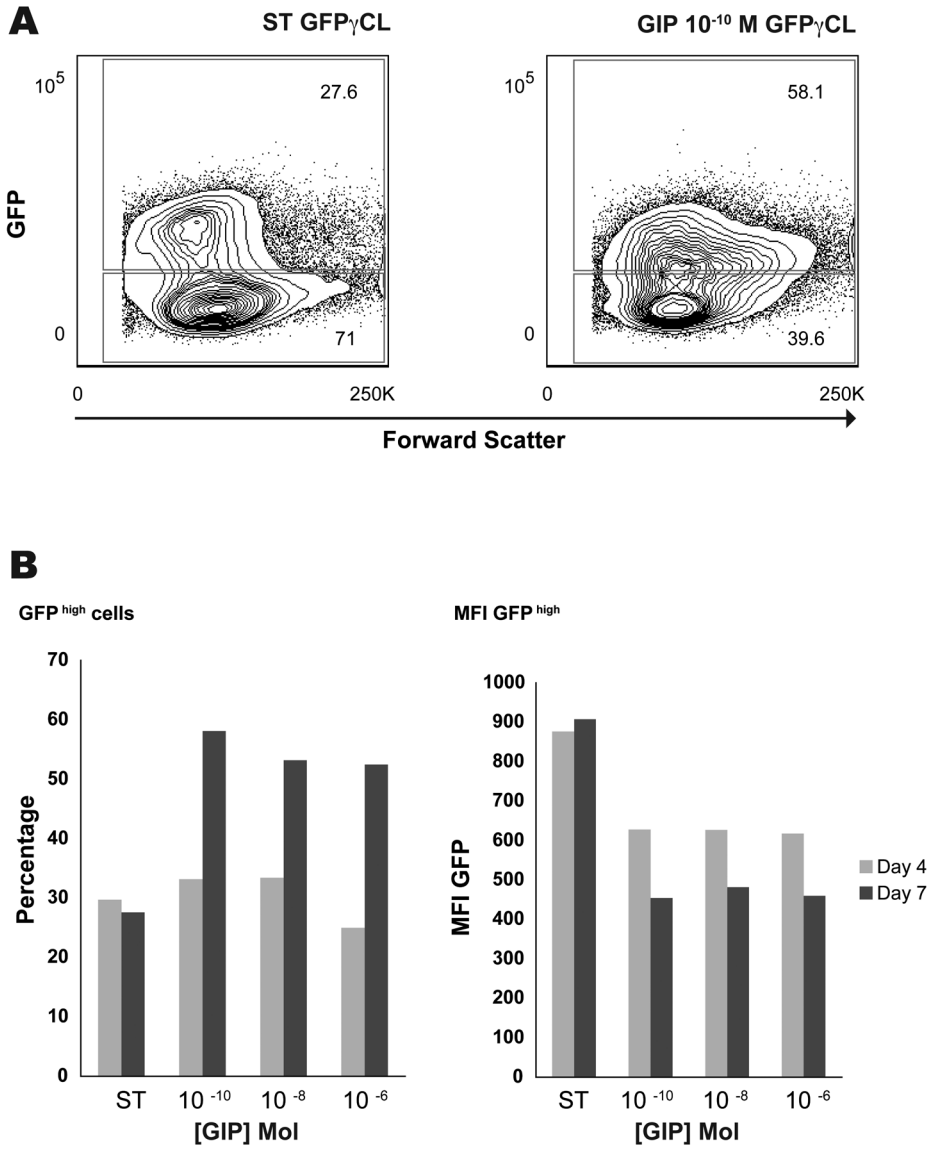


Figure 4

A. Flowcytometry analysis of GFPγCL in proliferative conditions (ST) and with 10^{-10}M growth inhibitory peptide (GIP). More cells appear to be GFP positive for the selected gate when cells were treated with GIP compared to the standard conditions (57.1%/27.6%).

B. Mean fluorescence intensity analysis (MFI) of GFP gated cells (GFP^{high}) showed that the higher percentage of the gated GFP positive cells upon GIP treatment (day7) had a lower MFI value compared to non-treated cells (almost half the value of the ST conditions at day7).

G bound fraction was analyzed by Mass Spectrometry.

GFPγCL, as mentioned before, was of particular interest, because it expresses GFP (γ-globin) under proliferative conditions with an exclusive expression of GFP protein upon differentiation. This is not the case for DsRedβCL, which expresses human β-globin and endogenous βmaj-βmin globins after induction (Fig. 2 and 3).

The first striking result (Fig. 5A) is that the TATA-binding protein (TBP) is missing in the GFPγCL TAF10 IP in both induced and non-induced state as opposed to the DsRedβCL TAF10 IP. However, Western blot analysis of nuclear extracts of different erythroid cell lines, showed that TBP protein is present and levels decrease upon differentiation (Fig. 6A). It is also interesting to note that the stoichiometry of TBP compared to other TFIID components is low when present in the IP samples, an observation that is in concordance with previous reports (Perletti, Dantonel et al. 1999; Frontini, Soutoglou et al. 2005). TAF4 was also analysed by Western blot (Fig. 6B) showing that the protein levels of these factor does not dramatically change in differentiating erythroid cells as compared to more immature progenitors, and is present on the IPs of both cell lines. TBP-related factor 3 (TRF3) (TbpL2) instead, appears in the IP of non-induced and induced state of the DsRedβCL only and remains to be confirmed by western blot that the TBP-homologue is expressed in erythroid cells.

The next TFIID component that clearly differs between the two cell lines is TAF4b. Upon differentiation, TAF4b is non-detectable in the IP in DsRedβCL, in contrast with the lower score of TAF4b protein present in both induced and non-induced GFPγCL (Fig. 5A). TAF4b is one of the most interesting TFIID factors because it is selectively expressed in differentiating lymphocytes, testis and granulosa cells of the ovary (Dikstein, Zhou et al. 1996; Freiman, Albright et al. 2001; Falender, Freiman et al. 2005).

TAF9b is another factor that is absent in the GFPγCL as compared to DsRedβCL. TAF9b as TAF4b are paralogues of the TAF9 and TAF4 respectively and expression of both of them at the same time could have redundant, overlapping functions but also unique effects in transcription activation of specific genes. It is interesting to note that we observe no TAF9b and TBP in the GFPγCL since TAF9b has been shown to be a *bona fide* TAF, present in TFIID and TBP-Free-TAF-Containing complexes (Frontini, Soutoglou et al. 2005) meaning that it should be detected also in the case of a non-canonical TFIID formation.

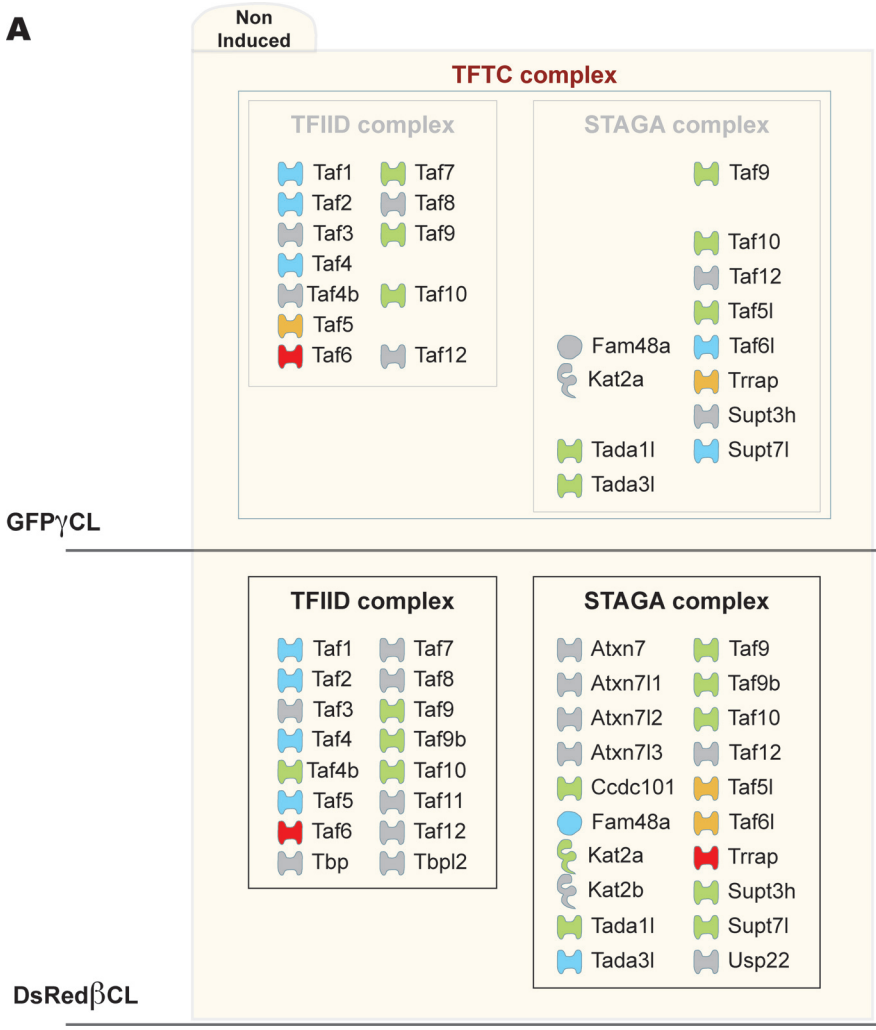
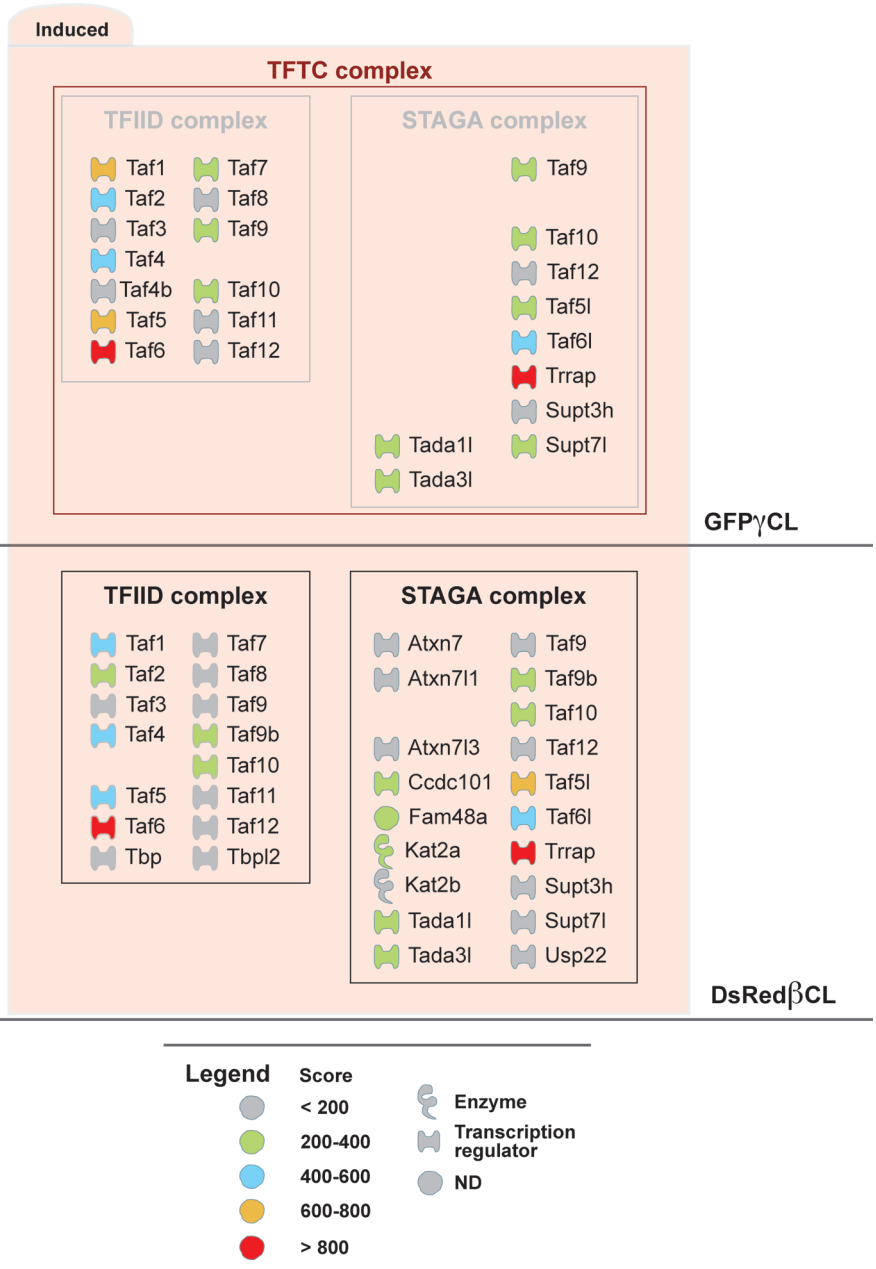


Figure 5

A. Analysis of TAF10 IP mass spectrometry data in GFPγCL and DsRedβCL in proliferating and differentiating conditions. The categorization of proteins in complexes (TFTC/STAGA) is based on published data.


Induced
TFTC complex
TFIID complex
Taf1
Taf2
Taf3
Taf4
Taf4b
Taf5
Taf6
Taf7
Taf8
Taf9
Taf10
Taf11
Taf12
STAGA complex
Taf9
Taf10
Taf12
Taf5l
Taf6l
Trrap
Supt3h
Supt7l
Tada1l
Tada3l
GFPγCL
TFIID complex
Taf1
Taf2
Taf3
Taf4
Taf5
Taf6
Tbp
Taf7
Taf8
Taf9
Taf9b
Taf10
Taf11
Taf12
Tbpl2
STAGA complex
Atxn7
Atxn7l1
Atxn7l3
Ccdc101
Fam48a
Kat2a
Kat2b
Tada1l
Tada3l
Taf9
Taf9b
Taf10
Taf12
Taf5l
Taf6l
Trrap
Supt3h
Supt7l
Usp22
DsRedβCL
Legend
Score
< 200
200-400
400-600
600-800
> 800
Enzyme
Transcription regulator
ND

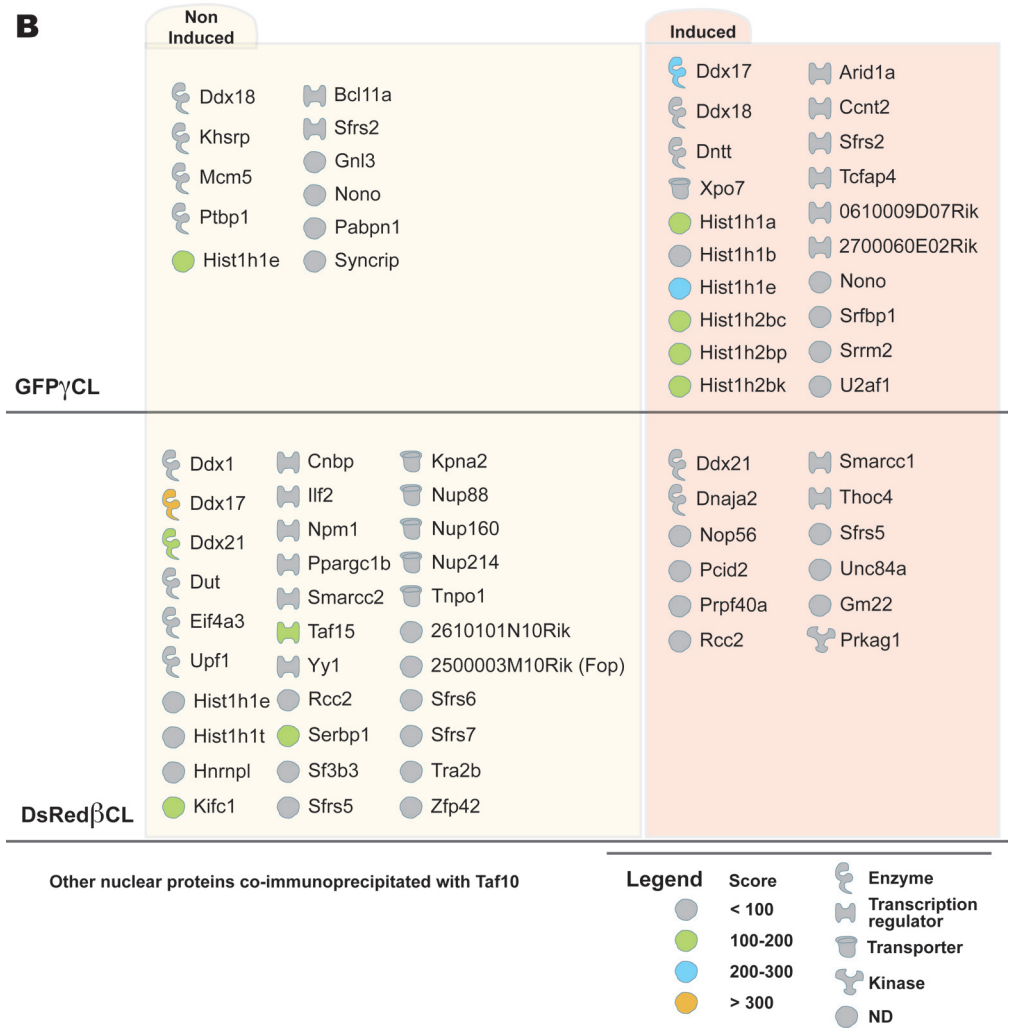


Figure 5

B. The rest of transcriptional regulators, coregulators and histone variants bound to TAF10 are summarized.

Apart from the several stoichiometric discrepancies that account for most of the TAF10 interacting proteins and can be due to a time-response delay between the two cell lines during the process of differentiation, there is a group of proteins forming part of the human TFTC/STAGA complexes, including Usp22, Atxn7, Atxn7l1/2/3, Ccdc101, Kat2b in non-induced and additionally the Kat2a, Fam48a in the induced state that are totally absent in the GFPγCL (Nagy and Tora 2007). Recently it was shown that the Usp22 subunit is part of a module of the TFTC/STAGA complex that specifically removes the ubiquitin moiety from mono-ubiquitylated histones H2A and H2B, linking the histone acetyltransferase activity and the de-ubiquitylation as essential steps during the process of full activation of nuclear receptors on chromatin templates (Zhao, Lang et al. 2008).

To our surprise, there was a significant change in numbers of transcriptional regulators either during the course of differentiation of the cells or between the two cell lines with different globin gene expression (Fig. 5B). The reasons for such an observation can be multiple, starting from technical specifications of the immunoprecipitation assays to limited interaction of TAF10 *per se* with DNA-binding transcription factors. A few of them have been involved in transcriptional regulation of the globins. Bcl11a is one interesting transcription factor that was found only in the non-induced state of the GFPγCL and in the low score category in concordance with lower levels present in the nucleus of the GFPγCL clone (Fig. 5B and 6C). It was recently shown that BCL11A regulates fetal hemoglobin (HbF) (Sankaran, Menne et al. 2008; Sankaran, Xu et al. 2009), possibly through interactions with other repressors. However, BCL11A did not show robust DNA-binding in the β-blobin locus nor compromised the expression levels of master transcriptional regulators, including GATA1, FOG1, NF-E2 and EKLF, suggesting that other crucial interactions might be the key of its repressive function on the γ-globin gene, that could imply members of the TFIID complex. Other regulators, including YY1, ZNF9N (CNBP), AP-4 (TFAP4), HP1-binding protein 3 and Fop (Friend of Prmt1) (van Dijk, Gillemans et al. 2009) were also found (Fig. 5B).

Interactions with other coregulators are also detected, as indicated by the Brg1 associated factors, Smarcc1/2 (BAF155/BAF170), Arid1A (BAF250) but without a clear connection to the differentiation process or to the globin expression pattern of the cells.

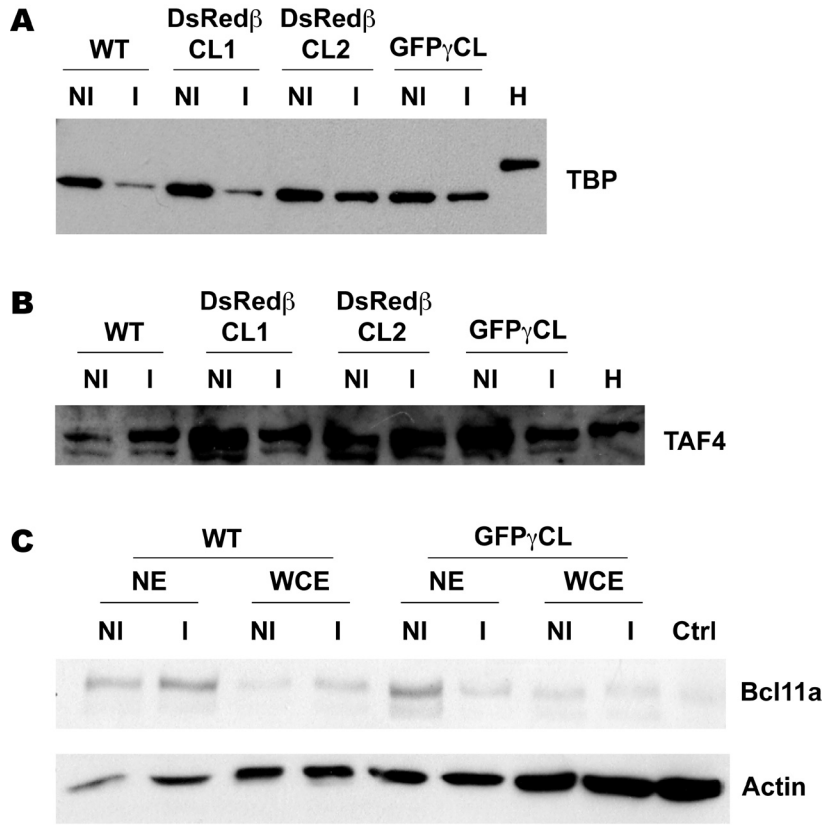


Figure 6

A. Western blot analysis of wild type (WT), DsRedβCL (Lines 1&2), GFPγCL for TBP and **B.** TAF4 in proliferating (NI) and differentiating conditions (I). **C.** Bcl11a western blot analysis of WT and GFPγCL in nuclear extracts (NE) and whole cell extracts (WCE) in both proliferating and differentiating conditions. Hela cell extracts (H) and A20 (mouse B cell lymphoma cell line) extracts (Ctrl) were used as controls. Same number of cells were loaded on each lane.

Discussion

TAF10, one of the smallest subunits of the TFIID, participates in canonical and non-canonical TFIID formations and is crucial for the stability of the TFIID structure (Mohan, Scheer et al. 2003) and the postnatal repression of specific genes (Tatarakis, Margaritis et al. 2008). It was tempting to investigate the role of TAF10 in two main directions. First we wanted to know whether there are significant changes in the TFIID multisubunit composition during erythroid differentiation and interactions with known transcriptional regulators and coregulators and secondly to identify interactions of TAF10 or the TFIID, related with potential repressor activity on globin expression and particularly the γ-globin gene, since it is silenced postnatally.

First, we checked levels of TAFs and TBP in mouse and human fetal liver and peripheral blood erythroid progenitors to confirm their expression. Secondly, we performed TAF10 knockdown experiments and we found that there is a consistent but moderate induction of γ-globin gene expression. In the mouse, γ-globin upregulation was accompanied by β-globin upregulation, while in human, fetal hemoglobin was induced at the expense of adult hemoglobin. Furthermore, when compared to reported effects on γ-globin gene expression after BCL11A knockdown in human erythroid progenitors (Sankaran, Menne et al. 2008) we achieved similar ratios in induction levels of HbF using the same target sequence as described previously. Curiously, studies of the function of Bcl11a in the mouse, revealed species differences in globin regulation (Sankaran, Xu et al. 2009), just like we show for TAF10. This observation increased the significance of a proteomic analysis in order to obtain insights about the function of TAF10 and TFIID in erythroid cells.

We generated mouse cell lines of fetal liver cells from transgenic mice carrying the modified human β-globin locus in p53-/- background in order to collect enough material for the IP assay. We used a transfected clone with the β-globin locus that for unknown reasons expresses GFP protein (γ-globin) at levels that reach almost hundred percent upon differentiation, named GFPγCL.

Mass Spectrometry analysis revealed many differences, starting from the TATA-binding protein (TBP) that even in low stoichiometry exists as core nucleating factor of the TFIID in both induced and non-induced state in DsRedβCL. In the GFPγCL the absence of TBP indicates the possible formation and exclusive incorporation of TAF10 in a SAGA complex. Sap130 and TAF13 were not detected in either cell line, even though both have been reported to coimmunoprecipitate with TAF4, that is

shared between TFIID and TFTC complexes (Mohan, Scheer et al. 2003). Whether this is connected with γ-globin gene expression remains to be shown with chromatin immunoprecipitations (ChIP) assays.

Another critical factor was present in the TAF10 IP in DsRedβCL in both proliferative and differentiating state, the Tbpl2 or TRF3 (TBP-related factor 3). The TRF3-TAF3 complex has been shown to play a key role during the differentiation process of myoblasts to myotubes, taking over the role of TFIID as a core promoter complex (Deato and Tjian 2007). The contribution of TFIID to cell proliferation and cell cycle has been shown (Wang and Tjian 1994; Metzger, Scheer et al. 1999; Martianov, Viville et al. 2002; Mohan, Scheer et al. 2003) and removing such transcriptional factors as TFIID could potentially provide the means of exiting cell cycle and entering the terminal differentiation process. It is of interest to find out the composition of a possible complex of TRF3-TAF10 that appears here and its function in the cell. In addition, TRF3 coexists with the TBP in similar stoichiometry in the TAF10 bound fraction that could simply mean distribution of regulatory tasks between distinct groups of genes responding differently to the differentiation stimuli, as reflected by the differential levels of the housekeeping genes and the erythroid specific transcription factors during the course of the differentiation process.

Amongst the other TAFs pulled down with TAF-10, TAF4b/TAF4 and TAF9b/TAF9 are of particular interest. TAF4b has been demonstrated to form TAF4b/TAF4 TFIID complexes apart from the canonical TAF4/TAF4 TFIID and selectively support transcription in specific promoters, even in the absence of activators (Liu, Coleman et al. 2008). Additionally there is some evidence that TAF4b-containing TFIID can contact DNA (Shao, Revach et al. 2005) which could explain promoter specificity, with subsequent activator recruitment. The presence of TAF4b in the GFPγCL in lower stoichiometry than TAF4 could imply the formation of both canonical and non-canonical TAF4/4b-containing complexes regulating the expression of different sets of genes. Once more, this could be the case for the γ-globin gene specifically and that needs to be investigated further.

TAF9 and its paralogue TAF9b are ubiquitously expressed as TAF4, in contrast to the TAF4b which is a tissue-specific factor (Freiman, Albright et al. 2001). Elucidation of the 3D structure of the TFIID holo-complex will give more insight of the existence of different TAF9/9b-containing complexes as previously mentioned for TAF4. Stoichiometric indications towards the prevalence of TAF9b-containing TFIID complexes is probably tissue related but expression profiling studies have

shown that there are distinct roles for each of the TAF9 paralogues within the cell (Frontini, Soutoglou et al. 2005). A functional relevance of the TAF9b absence in the γ-globin expressing clone GFPγCL, especially when other known interacting TAFs are present, including TAF6 with which it forms heterodimers, would be necessary for more conclusions regarding the globin expression in erythroid cells. Importantly, TAF9b is not present in the TAF10 IP in human fetal liver cells, raising a serious indication of its involvement in the regulation of fetal hemoglobin. Regarding the post translational modifications (PTMs), it is thought that they play an important role in the activities of TAF10 also, but our experiments did not show any modifications of TAF10 during erythroid differentiation (data not shown). Though, other TAFs where found differentially modified, including TAF9b, TAF12, TAF6l, TAF6, but these results should be repeated and validated.

A clear contrast between the two cell lines appears to be a group of proteins directly associated with histone acetyltransferase and deubiquitinase activity. Components of TFTC/STAGA complexes, including the specific subunit ATXN7, ATXN7L1, ATXN7L2, ATXN7L3, Usp22, Ccdc101 (hSgf29), appear selectively in the DsredβCL in induced and non-induced conditions. Instead, GFPγCL lacks these proteins as well as TBP. This seems to be a contradiction since they are part of TBP-free complexes (TFTC/STAGA) and we would expect that they are present in GFPγCL keeping in mind that TAF10, shared amongst TFIID and SAGA complexes, efficiently pulled them down in the DsredβCL even in the induced state, where the nucleus protein content is reduced significantly. Although ENY2 was not found in our mouse cell line IPs, the reported deubiquitinase activity of a sub-complex of the TFTC/STAGA involving ATXN7L3, Usp22 and ENY2 (Zhao, Lang et al. 2008) implicates a distinct role of histone modification, apart from acetylation, possibly related with the expression of the γ-globin gene. Still, the presence of these proteins throughout differentiation of the erythroid cells points out the importance of counteracting heterochromatin environment that is necessary for the diverse function of erythroid transcriptional regulators and expression of globins.

Interactions between TAF10 and erythroid specific transcriptional regulators were not uncovered as expected. Still, Bcl11a, recently described as a key regulator of HbF, is present in the GFPγCL clone IP in the non induced state when GFP (γ-globin) is moderately expressed and disappears upon induction when cells express γ-globin exclusively. Unfortunately, Bcl11a was not bound to TAF10 in the DsRedβ CL expressing cells, which would be supporting evidence of its reported repressor

activity on HbF expression.

TAF10 IPs on human fetal liver cells, that express predominantly HbF, showed interactions with GATA1, LDB1, TAL1, and other transcriptional regulators (APPENDIX II), implying that there is possibly specific recruitment of TAFs through interactions with regulators that initiate transcription of globins as has been reported for the selective recruitment of TAF9 in a EKLF dependent manner, essential for β-globin expression (Sengupta, Cohet et al. 2009). BCL11A was not found bound to TAF10 in human fetal liver, which makes more sense if such an interaction between TAF10 and BCL11A would be related to the almost hundred percent expression of fetal hemoglobin in fetal liver cells. Parallel IPs in human adult erythroid progenitors have to be performed to conclude whether a possible interaction of TAF10 with BCL11A is related to γ-globin silencing. It is encouraging that the human TAF10 IP has revealed many interactions with transcriptional regulators and coregulators to a much larger extent than in mouse cells. This is probably due to technical reasons and may be improved. We would like to perform TAF10 IP in human blood erythroid progenitors that express mainly adult hemoglobin (HbA) and compare that with the human fetal liver IP. Interactions with transcriptional activators and/or repressors is likely to uncover links to activation of globin promoters on each developmental stage.

The role of TAF10 as a corner stone of TFIID integrity with further implications on postnatal repression of several hepatic genes (Tatarakis, Margaritis et al. 2008) is an attractive model that could lead to new transcriptional regulation concepts taking place through the globin switching from fetal to adult stage. The direct involvement of TAF10 in the regulation of γ-globin gene is not fully supported yet but it emphasizes the role of TFIID as a sensor of the diverse interactions taking place in erythroid cells.

Acknowledgements

We are grateful to Elisabeth Scheer of IGBMC in Strasbourg for guidance to perform the TAF10 IP experiments and Arnaud Krebs for help with IP data analysis. Also we are thankful to Dr. Yiannis Talianidis (B.S.R.C Al. Fleming, Athens) for providing us with protocols and suggestions for optimizing the TAF10 IP in erythroid cells.

References

Deato, M. D. and R. Tjian (2007). "Switching of the core transcription machinery during myogenesis." Genes Dev **21**(17): 2137-49.

Dikstein, R., S. Zhou, et al. (1996). "Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130." Cell **87**(1): 137-46.

Dynlacht, B. D., T. Hoey, et al. (1991). "Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation." Cell **66**(3): 563-76.

Falender, A. E., R. N. Freiman, et al. (2005). "Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID." Genes Dev **19**(7): 794-803.

Freiman, R. N., S. R. Albright, et al. (2001). "Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development." Science **293**(5537): 2084-7.

Frontini, M., E. Soutoglou, et al. (2005). "TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9." Mol Cell Biol **25**(11): 4638-49.

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-12.

Hiller, M., X. Chen, et al. (2004). "Testis-specific TAF homologs collaborate to control a tissue-specific transcription program." Development **131**(21): 5297-308.

Hochheimer, A. and R. Tjian (2003). "Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression." Genes Dev **17**(11): 1309-20.

Jacq, X., C. Brou, et al. (1994). „Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor." Cell **79**(1): 107-17.

Kim, T. H., L. O. Barrera, et al. (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-80.

Kouskouti, A., E. Scheer, et al. (2004). "Gene-specific modulation of TAF10 function by SET9-mediated methylation." Mol Cell **14**(2): 175-82.

Liu, W. L., R. A. Coleman, et al. (2008). «Structural changes in TAF4b-TFIID correlate with promoter selectivity.» Mol Cell **29**(1): 81-91.

Martianov, I., S. Viville, et al. (2002). "RNA polymerase II transcription in murine cells lacking the TATA binding protein." Science **298**(5595): 1036-9.

Metzger, D., E. Scheer, et al. (1999). "Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes." Embo J **18**(17): 4823-34.

Mizejewski, G. J. and R. MacColl (2003). "Alpha-fetoprotein growth inhibitory peptides: potential leads for cancer therapeutics." Mol Cancer Ther **2**(11): 1243-55.

Mohan, W. S., Jr., E. Scheer, et al. (2003). "TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice." Mol Cell Biol **23**(12): 4307-18.

Mulder, M. P., M. Wilke, et al. (1995). "Positional mapping of loci in the DiGeorge critical region at chromosome 22q11 using a new marker (D22S183)." Hum Genet **96**(2): 133-41.

Nagy, Z. and L. Tora (2007). "Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation." Oncogene **26**(37): 5341-57.

Nguyen, T. T., K. Cho, et al. (2005). "Transcription factor interactions and chromatin modifications associated with p53-mediated, developmental repression of the alpha-fetoprotein gene." Mol Cell Biol **25**(6): 2147-57.

Perletti, L., J. C. Dantonel, et al. (1999). "The TATA-binding protein and its associated factors are differentially expressed in adult mouse tissues." J Biol Chem **274**(22): 15301-4.

Sankaran, V. G., T. F. Menne, et al. (2008). "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A." Science **322**(5909): 1839-42.

Sankaran, V. G., J. Xu, et al. (2009). “Developmental and species-divergent globin switching are driven by BCL11A.” Nature **460**(7259): 1093-7.
Sengupta, T., N. Cohet, et al. (2009). “Distinct modes of gene regulation by a cell-specific transcriptional activator.” Proc Natl Acad Sci U S A **106**(11): 4213-8.
Shao, H., M. Revach, et al. (2005). “Core promoter binding by histone-like TAF complexes.” Mol Cell Biol **25**(1): 206-19.
Strouboulis, J., N. Dillon, et al. (1992). “Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice.” Genes Dev **6**(10): 1857-64.
Tatarakis, A., T. Margaritis, et al. (2008). “Dominant and redundant functions of TFIID involved in the regulation of hepatic genes.” Mol Cell **31**(4): 531-43.
Thomas, M. C. and C. M. Chiang (2006). “The general transcription machinery and general cofactors.” Crit Rev Biochem Mol Biol **41**(3): 105-78.
Tora, L. (2002). “A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription.” Genes Dev **16**(6): 673-5.
van Dijk, T. B., N. Gillemans, et al. (2009). “Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases.” Mol Cell Biol.
Vermeulen, M., K. W. Mulder, et al. (2007). “Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.” Cell **131**(1): 58-69.
Wang, E. H. and R. Tjian (1994). “Promoter-selective transcriptional defect in cell cycle mutant ts13 rescued by hTAFII250.” Science **263**(5148): 811-4.
Wieczorek, E., M. Brand, et al. (1998). “Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II.” Nature **393**(6681): 187-91.
Wijgerde, M., F. Grosveld, et al. (1995). “Transcription complex stability and chromatin dynamics in vivo.” Nature **377**(6546): 209-13.
Zhao, Y., G. Lang, et al. (2008). “A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing.” Mol Cell **29**(1): 92-101.

CHAPTER 3

A genomic region on chromosome 19 is associated with persistent fetal hemoglobin production in adults

Joseph Borg*, Petros Papadopoulos*, Marianthi Georgitsi, Godfrey Grech, Marios Phylactides, Laura Gutiérrez, Annemieke Verkerk, Peter J. van der Spek, Jeroen Demmers, Christian A. Scerri, Wilhelmina Cassar, Ruth Galdies, Wilfred van Ijcken, Jun Hou, Alex E. Felice, Frank G. Grosveld, Marieke von Lindern, Sjaak Philipsen# and George P. Patrinos#

*: These authors contributed equally to this work
#: Corresponding authors

A genomic region on chromosome 19 is associated with persistent fetal hemoglobin production in adults

Joseph Borg*, Petros Papadopoulos*, Marianthi Georgitsi, Godfrey Grech, Marios Phylactides, Laura Gutiérrez, Annemieke Verkerk, Peter J. van der Spek, Jeroen Demmers, Christian A. Scerri, Wilhelmina Cassar, Ruth Galdies, Wilfred van IJcken, Jun Hou, Alex E. Felice, Frank G. Grosveld, Marieke von Lindern, Sjaak Philipsen# and George P. Patrinos#

*: These authors contributed equally to this work

#: Corresponding authors

Abstract

Hereditary Persistence of Fetal Haemoglobin (HPFH) is an inherited condition characterized by elevated levels of fetal hemoglobin (HbF). Some contributory factors, both genetic and environmental, are well understood, but others remain elusive. We studied a Maltese family of which ten members presented HPFH and the genetic cause was unknown. Seventeen other members from the same family carried normal levels of HbF. A genome-wide scan performed on all twenty seven family members, revealed a candidate region on chromosome 19p13.12-19p13.13. One haplotype was observed in all members with HPFH that was absent in normal individuals. Sequencing of candidate genes in this region revealed two novel mutations in the *KLF1* gene (p.M39L, p.K288X) and a sequence variant in the *ASF1B* gene (NT_011295.10:g.5498442G>T). Both mutations appear to be absent in the Maltese population and could possibly affect HbF regulation. Functional assays, including knockdown experiments using KLF1 and ASF1B shRNAs in blood erythroid progenitors from healthy donors and the Maltese family members, are currently in progress.

Introduction

Fetal hemoglobin (HbF) is the predominant type of hemoglobin in fetal life, but around birth there is a switch from fetal to adult globin gene expression. HbF is gradually replaced by adult hemoglobin (HbA), and at 6 months of age the major circulating hemoglobin is HbA (G. Stamatoyannopoulos 2001). Residual amounts of HbF, however, continue to be synthesized throughout adult life and the amounts vary considerably, with the majority of adults having less than 1% of HbF. Our understanding of hemoglobin control is historically based on Mendelian models of inheritance of natural mutants. Indeed, a series of mutations of the β-globin cluster have been discovered that impair the fetal-to-adult hemoglobin switch leading to persistent γ-globin expression and elevated HbF throughout adult life. This condition is termed Hereditary Persistence of Fetal Hemoglobin (HPFH) (Forget 1998). There are two types of HPFH mutations in the β–globin locus: point mutations in the promoter of the γ-globin gene (*HBG1* or *HBG2*) and deletions removing substantial regions of the β-globin cluster, often including the β-globin gene (*HBB*) (Craig, Rochette et al. 1996). A range of HbF levels that do not fit clear Mendelian inheritance models or the typical HPFH phenotype, *i.e*. high HbF levels accompanied by concomitant lower HbA_2 ($\alpha_2\delta_2$) levels, has been noted. Although some of this variability can be explained by the HBB cluster chromosomal background, a substantial proportion of the HbF increase is not linked to the HBB cluster.

It is now clear that common HbF variation is a quantitative genetic trait, shaped by common polymorphisms in genes that are not related to the HBB gene cluster. There are only a few examples of such SNPs in genes, such as *BCL11A* (Liu, Keller et al. 2003) residing on chromosome 2 (Menzel, Garner et al. 2007; Lettre, Sankaran et al. 2008; Uda, Galanello et al. 2008) and the intergenic region between the *MYB* (Vegiopoulos, Garcia et al. 2006) and *HBS1L* genes that reside on chromosome 6q22.3 (Craig, Rochette et al. 1996; Garner, Mitchell et al. 1998; Thein, Menzel et al. 2007). These genes have been identified using either genome-wide association or traditional linkage analysis approaches.

Here, we report the identification of a novel genomic region on chromosome 19p13.12-13, associated with increased HbF production in a Maltese family.

Materials and Methods

Patient recruitment

A Maltese family comprising a total of 27 members was referred to the University Hospital of Malta for routine screening (Fig. 1). Some of them presented a phenotype of mild microcytic anemia (Mean Corpuscular Volume MCV; 76fl, normal range 80-100 fl), acanthocytosis, poikilocytosis and polychromasia. Total hemoglobin levels were in the lower normal range (11 to 12 g/dl). Of particular interest was the high HbF level of 20% and high F cell number (53%) in one proband, and hence additional individuals within the same family were encouraged to undergo hematological analytics. Blood samples were obtained with informed consent and hematological indices were determined.

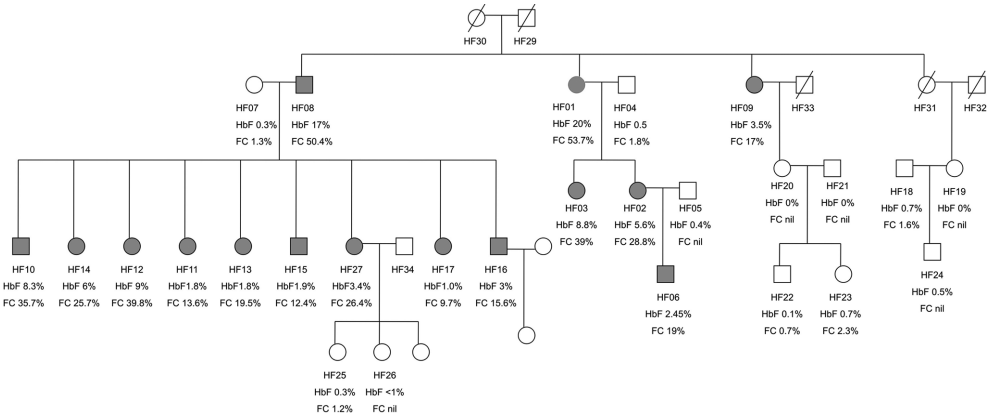


Figure 1

The pedigree of the Maltese family. The original parents are known to be consanguineous from six generations before. Percentage of HbF and F cells are indicated on each member. The HF1, HF2, HF3, HF8, HF9, HF10, HF12, HF14, HF16, HF27 individuals exhibited morphological malformations in their blood smears, a sign of membrane/cytoskeletal instability of red blood cells and mild microcytic anemia.

Molecular genetic analysis

Total genomic DNA (gDNA) was extracted from 10^6 cells from each blood sample culture using a modified salting out procedure (Miller, Dykes et al. 1988). Additional blood samples were taken from fifteen out of twenty seven family members for further experimental hematology explained in detail below. Control gDNA from 400 random cord bloods were available from the Laboratory of Molecular Genetics, Biomedical Sciences Building, University of Malta.

The family members were genotyped in the *HBB*, *HBD* genes and both HBG gene promoters to detect point mutations and small indels leading to β-, δ-thalassemia or HPFH, respectively. Gap PCR was carried out to detect possible genomic rearrangements leading to deletional HPFH or δβ-thalassemia.

The *Nsp*I mapping 250K set (array 4.0; Affymetrix, Santa Clara, CA, USA) was used to analyze the 27 samples of genomic DNA starting with 250ng of genomic DNA per array.

DNA Linkage Analysis

Multipoint parametric and non-parametric linkage analyses were performed by EasyLinkage v5.05 Beta (http://www.uni-wuerzburg.de/nephrologie/molecular_genetics/molecular_genetic.htm) (Hoffmann and Lindner, 2005) using the Merlin v1.0.1 software to calculate exponential LOD scores using the Kong and Cox model (Kong and Cox 1997). The exponential model algorithm provided a better linkage test than others available since a large increase in allele sharing among individuals carrying HbF-raising determinants was assumed. Parametric analysis was carried out using variable penetrances for both dominant and recessive models of inheritance. Penetrances used for the dominant model were 0.01, 0.90 and 0.90 for the wild-type homozygote, mutant heterozygote and mutant homozygote, respectively. The recessive model was defined by penetrances 0.01, 0.01 and 0.90 for the wild-type homozygote, mutant heterozygote and mutant homozygote, respectively. For both dominant and recessive models, the disease allele prevalence frequency assumed was 0.0001, and phenocopy rate 1%. A codominant allele frequency algorithm was used for the analysis. These analyses were carried out using the sex-averaged 500K Marshfield genetic map. Mendelian and non-Mendelian errors, such as double or triple recombination events were detected and omitted from the analysis. This increased the power and accuracy. The QTL file contained quantitative hematological

values that included HbF%, HbF per F cell and F cells %. Estimated haplotypes were constructed and visualized using Haplopainter (http://haplopainter.sourceforge.net/) version 29.5 (Thiele and Nurnberg 2005) (Fig. 2).

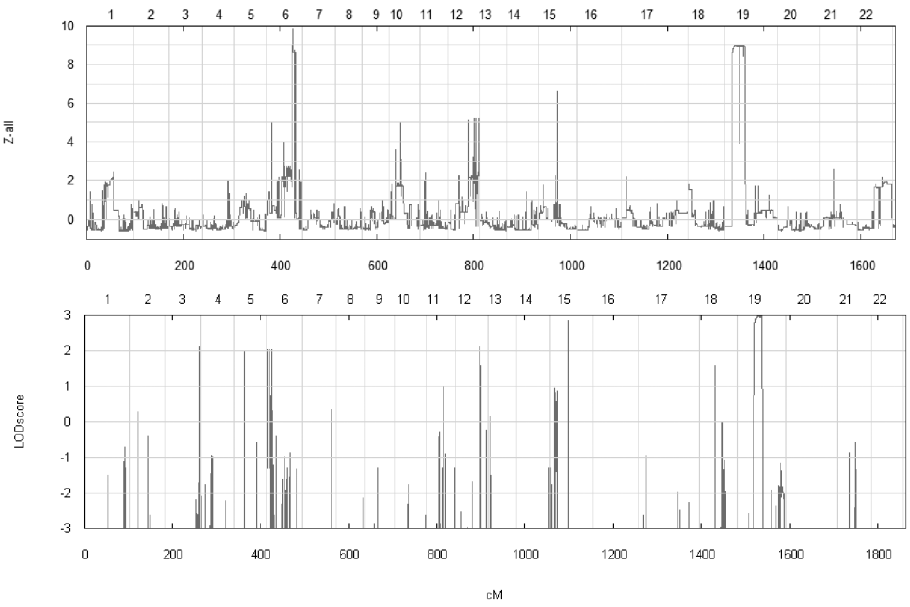


Figure 2
NPL (Z-score; model-free) and LOD scores (autosomal co-dominant), with penetrances being 0.01, 0.90, 0.90 for the wildtype homozygote, mutant heterozygote and mutant homozygote, respectively (dominant model).

Cell cultures

Erythroid progenitor cells were cultured as described (von Lindern, Zauner et al. 1999) in the presence of recombinant human Epo (0.5 U/ml Stem Cell Technologies), recombinant human SCF (100 ng/ml, Stem Cell Technologies) and dexamethasone (10^{-6} M, Sigma). Cells were cultured at 1.5 - 3 × 10^6/ml through daily dilutions or medium refreshment. Cells were counted with an electronic cell counter (CASY-1, Schärfe-System, Germany). To induce terminal differentiation, cells were washed and transferred to medium with recombinant human Epo (10 U/ml), human insulin (1 U/ml Actrapid, Bayer-Leverkusen) and a high concentration of iron-loaded transferrin (0.5 mg/ml). To determine hemoglobin accumulation, three 50-µl aliquots of the cultures were processed for photometric determination of hemoglobin (E. Kowenz 1987). To analyze cell morphology, cells were centrifuged onto slides and stained

with histological dyes and neutral benzidine for hemoglobin (Beug, Palmieri et al. 1982). Cells were harvested for HPLC analysis and RNA extraction with TRizol (Invitrogen).

Transcription profiling

A minimum of 1.5 x10^6 erythroid cells were harvested at day 12 of culture and RNA was extracted with TRizol (Invitrogen) and purified using the RNeasy Mini Kit (Qiagen, Crawley, UK), including an on-column DNaseI digestion, according to the manufacturer's instructions. RNA yield was determined using the 2100 Bioanalyzer from Agilent.

8-10µg of total RNA was analysed by microarrays. Fragmented biotinylated cRNA was prepared and 15µg hybridized to HG-U133plus2 GeneChips, according to the manufacturer's protocols (Affymetrix, Santa Clara, CA, USA). Single Array Expression Analysis was performed using the Affymetrix GeneChip Operating Software (GCOS). A global scaling strategy was used to give an average target intensity of 500 for each array. Data from all 13 arrays were filtered in Microsoft Excel to exclude all probe sets called either Absent or Marginal in all arrays. Control probe sets with the prefix AFFX were also removed prior to subsequent data analysis. Filtered data were transformed to a log2 scale and analysed to determine differentially expressed genes. A 3-fold change threshold and test statistic of $P<0.05$ were used as cut-off.

Total RNA (1µg) from each harvested erythroid cell sample was converted to cDNA using SuperScript II reverse transcriptase according to the manufacturer's instructions (Invitrogen, Paisley, UK). Globin mRNAs were analysed by quantitative real-time PCR (Q-PCR). Amplification of HBA, HBB and total HBG were performed with primers designed with Primer Express software v2.0 (Applied Biosystems, Warrington, UK). All amplifications used SYBR Green PCR Master Mix (Applied Biosystems). Q-PCR was performed on the Bio RAD Optical IQ Thermal Cycler (Bio-RAD) with the following conditions: 50°C for 2 minutes and 95°C for 10 minutes, followed by 45 cycles of 95°C for 15 seconds and 62°C for 45 seconds. All reactions were performed in triplicate. Target gene expression was normalized to *GAPDH* expression.

KLF1 Constructs

Human KLF1 clone (AccNo : BC040000, Imagenes GmbH) was amplified by PCR

with *att*-specific set of primers (Invitrogen) in order to fuse the cDNA with a V5 tag at the C-terminus of the protein (Fwr: *GGGG ACA AGT TTG TAC AAA AAA GCA GGC TAC CAT GGC CAC AGC CGA GAC C*, Rev: *GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC TGA AAG GTG GCG CTT CAT GT*). In parallel, we amplified part of the clone with *att*-specific primers (Fwr primer as described, Rev: *GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC CTT GCC GCA ACC CGG GTG*), mimicking the single base mutation at amino acid 288 which creates a stop codon, eliminating the zinc-fingers of the protein. Both PCR products (1089nt+V5tag or 864nt+V5 tag) were subject to two rounds of recombination according to the Gateway system (BP& LR, Invitrogen) and finally introduced to a lentivirus expression vector pLenti6.3-V5 DEST (V533-06, Invitrogen). Both clones were verified by sequencing analysis.

Virus transduction of human proerythroblasts

293FT cells (R700-07, Invitrogen) were transfected with the expression clones made of the KLF1 cDNA clone and the control pLenti6.3-LacZ-V5 vector by standard methodology and virus was produced in T-175cm^2 flasks. Two days after transfection, the supernatant was collected, filtrated and centrifuged at 20K (Ultracentrifuge, Beckman) for 2hr at 4^{o}C. Human proerythroblasts cultured for one week were subsequently transduced in a 24well-plate. After two days of culture, the antibiotic Blasticidin (5μg/ml final concentration) was added to the cells and was kept for at least 2-3 days. At day seven, cells were harvested and nuclear extracts and lysates for High Performance Liquid Chromatography (HPLC) were prepared, and RNA extracted with Trizol (Invitrogen).

In the case of the shRNA knockdown experiments for KLF1, we used the plasmids from the TRC Mission shRNA Library (Sigma). Virus was made in 10cm plates, and 293T cells were transfected with polyelthylenimine (PEI, Sigma) using 20μg of shRNA plasmid DNA, 15μg of PAX-2 and 5μg of VSV-G plasmids. The procedure followed regarding the concentration of virus and human proerythroblast transduction is the same as described above. At least two plasmids resulted in satisfactory knockdown of KLF1 protein as assessed by Western blot analysis.

Western Blot

Nuclear extracts were semidry blotted to PVDF (Millipore) membrane, incubated for 1 hour at room temperature (RT) in 3% milk powder PBS-0.2%Tween20 blocking buffer

and 1hr RT incubation with the specific antibody in 1% milk PBS-0.2% Tween20, two washes in PBS-0.2%Tween20 and incubation with secondary antibody coupled to HRP in the same buffer as for the primary antibody in dilution 1/15000 with final four washes of 10min each. In the case of antiV5-HRP, a second incubation was not necessary. Finally ECL kit (Amersham Biosciences) was used to develop the membrane. The antibodies used are Bcl11a (sc-56013, Santa Cruz), Nucleophosmin (ab10530, AbCam), KLF1 (in-house) and anti-V5-HRP (R961-25, Invitrogen).

Results

Mapping of the candidate genomic region

Screening for genetic variations in the human HBB gene cluster that have been previously demonstrated to lead to high HbF levels such as deletional or non-deletional HPFH yielded negative results. Also, the XmnI-HBG2 genetic variation correlated with heterocellular HPFH was not found in any of the family members with high HbF levels. This prompted us to initiate a whole-genome scan to pinpoint the genetic determinant of the high HbF levels on other chromosomes.

Fig. 2 shows NPL (Z-all) and LOD scores observed following the genome-wide scan in this family, for the dominant model. Amongst all models tested, chromosome 19p13.12-13 showed the highest LOD scores when using multi-point parametric analysis at a penetrance of 90%. The region on chromosome 6q and 15q15.5, showed a lower LOD score than that of chromosome 19p13.12, but a higher NPL on chromosome 6q. Analysis was performed using a dominant model with penetrances ranging from 90-50% and 1% phenocopy rate. No evidence of linkage was observed to other regions including the well known HPFH linked loci of chromosomes 2q33 and 6q22.3 when using both dominant and recessive models.

One inherited haplotype was observed on chromosome 19p13.12-13, with a number of recombinations very close to the indicated markers. The linked locus resides somewhere between SNP_A-2106977 and SNP_A-1786368, which is present in all HPFH members and absent in normal members of the family. Only one stands out as highly likely to be causative out of many putative candidate genes in the area, the *KLF1* gene.

A number of sequence variants and two novel mutations were identified by direct DNA sequencing in two genes sequenced at this region (*ASF1B* and *KLF1*) when compared to reference sequences in the NCBI and Ensembl databases. SNP_A-1848826 SNP mutant allele was found in all HPFH members but also in at least two other normal family members. This SNP resides in the 3' UTR of *ASF1B*. SNP rs2072597 is the other common variant that was found in all HPFH members albeit it appears to be linked with two novel mutations. SNP rs2072597 although classified as a polymorphism, causes an amino acid change p.S102P and is classified as a natural variant, and quite frequent in Asian populations. The two novel mutations, that are linked together, affect the *KLF1* gene and are identified as p.M39L and

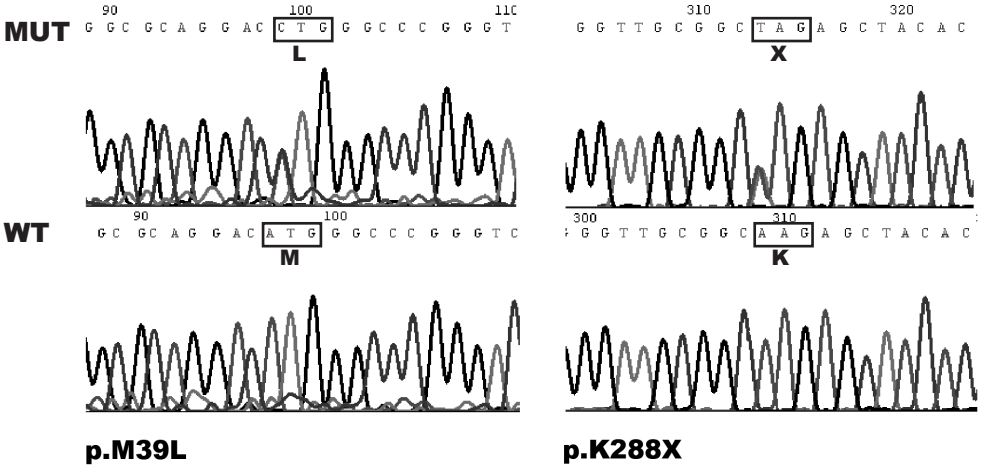


Figure 3
Sequencing analysis of the mutations identified for KLF1 gene in the high HbF members of the Maltese family.

p.K288X (Fig. 3). PCR–RFLP with Bfa I was used to determine the frequency of the p.K288X variant in a random sample from the general population (n=400), and was absent. The p.M39L was always present in family members carrying the p.K288X variant, so it could be inferred that the frequency of this mutation is equally rare due to linkage disequilibrium.

Gene expression profiling

Gene expression profiles were compared between the two groups of family members, namely high HbF (n=5) and low HbF members (n=8) to identify differentially expressed genes. A t-test for each gene was conducted to identify significant changes in expression values between the HPFH samples and normal samples. 172 genes are found to be downregulated, including heme synthesis enzymes (*ALAS1, UROD and FECH*), adult β-globin (*HBB*), and alpha hemoglobin stabilizing protein (*ERAF*). Confirming previous studies derived from mouse models, we also found differentially expressed a vast array of cell membrane proteins and markers such as the erythrocyte protein band 4.9 (*EPB49*), Glycophorin A (*GYPA*), *CD44*, the Kell and Lutheran blood group surface antigens and of course the Erythropoietin and Transferrin receptors (*CD71*).

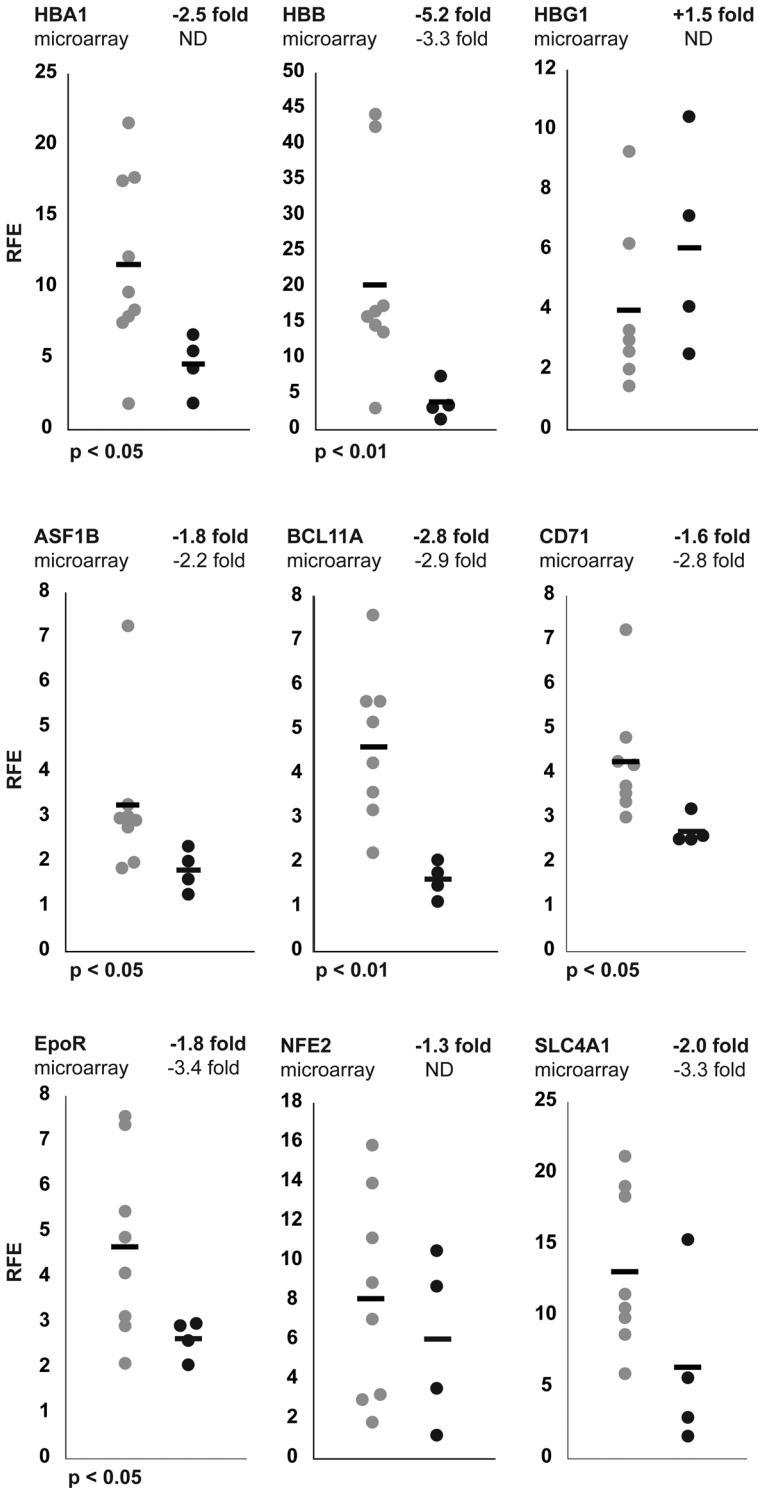
Quantitative PCR

The levels of *ASF1B* in HPFH family members correlated proportionally with *HBA* and *HBB* mRNA levels. Although downregulation of *ASF1B* did not correlate significantly with an increase in *HBG*, probably due to the small number of tested individuals (Fig. 4). All qPCR for deregulated genes, according to the expression profile data obtained, were performed twice in triplicates and the relative fold enrichment (RFE) was calculated. Fig. 4 summarizes the results from our qPCR analysis and their direct comparison with the transcription profiling data. We confirmed by qPCR the microarray results.

Functional studies

Human proerythroblasts were transduced with the KLF1 cDNA clone or the mutant form lacking the zinc-fingers. After 6 days of culture including 4 days under selection with Blasticidin (5μg/ml) cells were harvested and nuclear extracts were analyzed by western blot and whole cell lysates analyzed by HPLC. Expression of exogenous KLF1 and the mutant form were detected by anti-V5 mouse antibody together with the control LacZ-V5 protein. HPLC analysis did not show any significant difference in the fetal hemoglobin levels between the clones overexpressing KLF1 and the control cells (Fig. 5A), even though the viral cDNA expression was not optimal. The difference observed between the non-transduced cells (Control) and the control LacZ (HbF 6.6%, 9.9%) is a normal effect of the stress upon the viral infection.

In knockdown experiments in blood erythroid progenitors, we used five different shRNA clones against the KLF1 transcript and at least two resulted in efficient reduction of protein levels (Fig. 5B). There was a moderate difference in the fetal hemoglobin content (HPLC analyzed-7.5% vs 6.1%)) of the cells with reduced levels of KLF1 compared to the control cells transduced with a shRNA containing a no-target sequence or the non-transduced cells (Control).


HBA1
microarray
-2.5 fold
ND
HBB
microarray
-5.2 fold
-3.3 fold
HBG1
microarray
+1.5 fold
ND
RFE
p < 0.05
p < 0.01
ASF1B
microarray
-1.8 fold
-2.2 fold
BCL11A
microarray
-2.8 fold
-2.9 fold
CD71
microarray
-1.6 fold
-2.8 fold
EpoR
microarray
-1.8 fold
-3.4 fold
NFE2
microarray
-1.3 fold
ND
SLC4A1
microarray
-2.0 fold
-3.3 fold

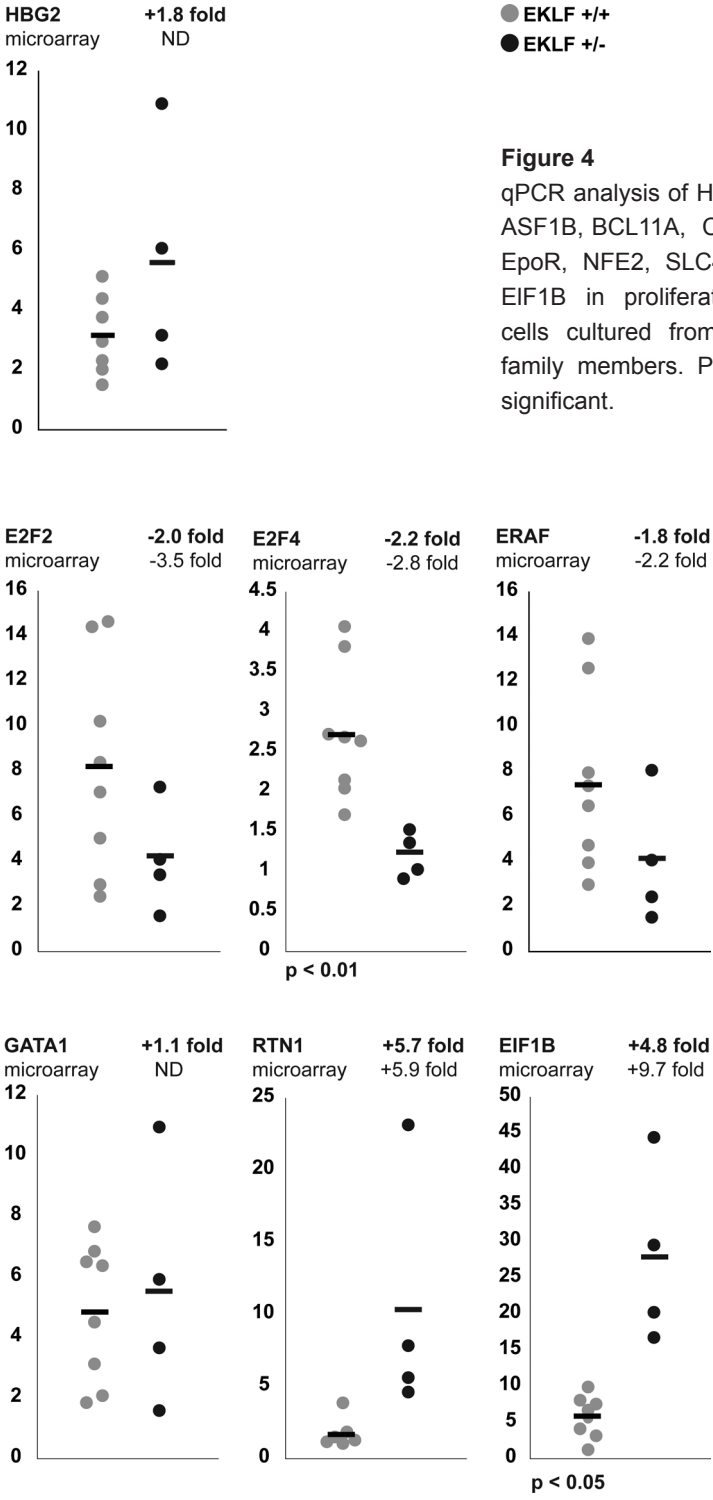


Figure 4

qPCR analysis of HBA1, HBB, HBG1, HBG2, ASF1B, BCL11A, CD71, E2F2, E2F4, ERAF, EpoR, NFE2, SLC4A1, GATA1, RTN1 and EIF1B in proliferating erythroid progenitor cells cultured from high HbF and normal family members. P-value is indicated when significant.

Discussion

Here, we report a novel candidate region responsible for increased HbF levels in adults, identified in a Maltese family. In this study, linkage analysis has pointed a high LOD score on chromosome 19p13.12-13, which harboured the two primary candidate genes in this research, *ASF1B* and *KLF1*.

ASF1B is known to function as a chaperone protein important in chromatin remodelling (Galvani, Courbeyrette et al. 2008), nucleosome assembly and disassembly and when mutated or knocked down it is also known to play dual roles both as a repressor and as an activator of gene expression. The 3' UTR SNP identified in this study SNP_A-1848826, although not very rare in the Maltese population, may influence ASF1B expression with an end effect on globin gene expression. Our results from quantitative PCR analysis indicated that family members with high HbF levels had approximately 2-fold lower *ASF1B* gene expression levels compared to family members with normal (lower) HbF levels (Fig. 4), suggesting that reduction of *ASF1B* expression alone or in combination with reduced *KLF1* levels could potentially upregulate *HBG* gene transcription. Additionally immunoprecipitation assays (IP) of Asf1b in mouse fetal liver cells (Appendix II) revealed interactions with PRMT5 that was previously shown to mediate transcriptional silencing of the HBG gene (Zhao, Rank et al. 2009), and FANCA, a protein involved in cell cycle progression and causative of Fanconi Anemia (complementation Group A) (Collins, Kupfer. 2005).

The two novel mutations in *KLF1* may play a direct role in HBG gene expression. As previously demonstrated (Singleton, Burton et al. 2008), the mutated KLF1 is responsible for the Indian Lutheran (InLu(a-)(b-)) phenotype, with a drastic decrease of cell membrane protein levels, heme and globin synthesis enzymes as well as transcription factor levels. Importantly, our qPCR data show a significant 5-fold decrease of *HBB* gene transcription in those family members with high HbF levels (Fig. 4), similarly to what has been documented for cis-acting mutations in the *HBB* gene promoter CACCC box (Treisman, Orkin et al. 1983) (see also http://globin.bx.psu.edu/hbvar). However, this decrease is accompanied by only marginal increase of HBG gene transcription that is also not significant for the group of individuals tested, as quantified by qPCR analysis but indicates a trend in that direction. Additionally, our preliminary results, obtained from knockdown experiments of *KLF1* in adult erythroid progenitors from healthy donors were in agreement with such an observation, indicating a marginal, though constant in different experiments, increase of HbF

levels compared to the control shRNA (7.5% vs 6.1%, respectively; Fig. 5B).

Trans-acting loci controlling HbF and F cell levels have now been mapped to several genomic regions on chromosomes 6q23 (Jiang, Best et al. 2006; Thein, Menzel et al. 2007), Xp22 (Miyoshi, Kaneto et al. 1988; Dover, Smith et al. 1992), 8q12 (Garner, Silver et al. 2004) and 2p16 (Menzel, Garner et al. 2007; Uda, Galanello et al. 2008) and a number of genes are known within the putative region. Additionally, the role of the Xmn I-HBG2 polymorphism (Gilman and Huisman 1985), within the β-globin locus region, has been insufficient to explain high HbF levels and does not correlate always in genetic studies, underlining the necessity of additional factors to synergistically produce high levels of HbF. Our genetic analysis in the human HBB gene cluster in these family members indicate that the XmnI-HBG2 sequence variation is not present in the family members with high HbF levels neither other already reported genetic traits.

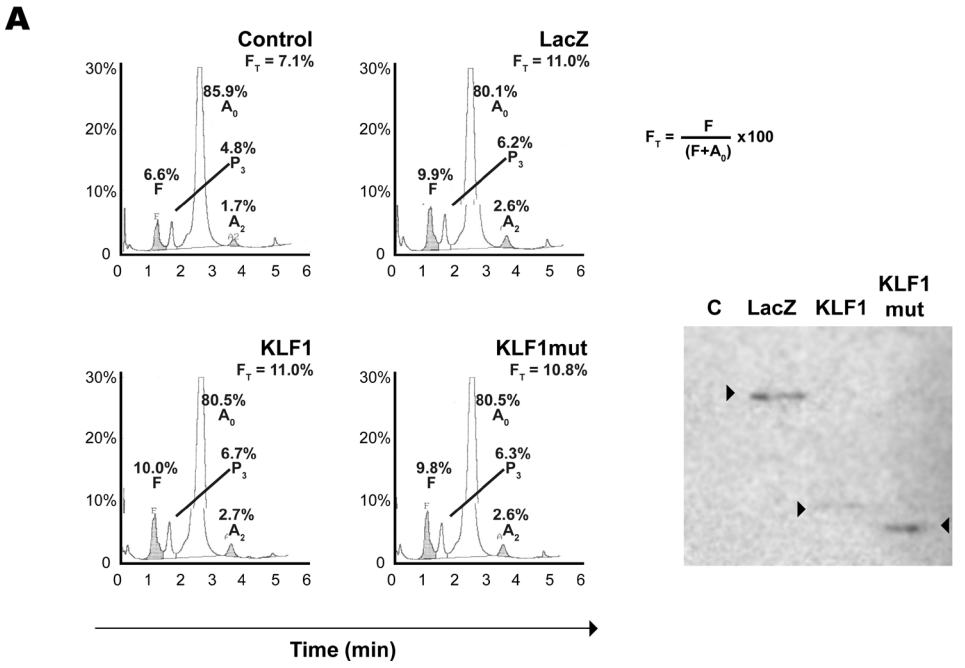


Figure 5A

Overexpression of *KLF1* and mutant *KLF1* (KLF1mut) in blood erythroid progenitors. Anti-V5 mouse antibody was used in Western blot to detect all different proteins expressed, the LacZ control, the *KLF1* and the mutant *KLF1* (KLF1mut). The size of each band corresponds to the correct MW of the proteins expressed, indicated with an arrow (►). HbF levels are depicted in the HPLC graphs. Control stands for untransduced cells, and LacZ stands for cells transduced with a lentivirus containing Lac-V5.

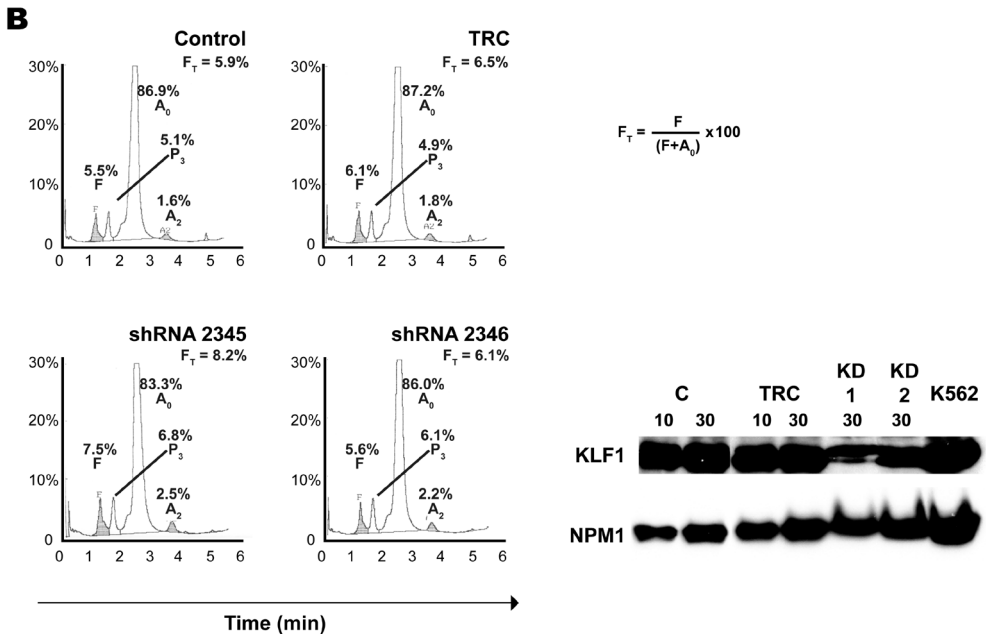


Figure 5B

Knockdown of *KLF1* in blood erythroid progenitors. The KD1 was more efficient than KD2 in reducing the protein levels of *KLF1*. In order to achieve equal loading between samples, same numbers of cells were harvested and resuspended in equal volumes of loading buffer. Knockdown samples were loaded with three times more extract volume per lane than the controls (10μl and 30μl as indicated on top of each lane). Nucleophosmin (NPM1) was used as an internal control for protein content of each sample. Control stands for untransduced cells, and TRC stands for cells transduced with a lentivirus containing a non-target shRNA (scrambled).

Transcription profiling also shed light on genes that could potentially constitute *KLF1* targets and hence be possibly involved in HBG gene silencing. Some of these genes were also previously reported (Drissen, von Lindern et al. 2005; Singleton, Burton et al. 2008). These genes include *E2F2*, *E2F4*, *EpoR* and *AHSP* (ERAF). In order to validate the GeneChip expression data serving to replicate the findings as well as to provide a more concise expression level of the interesting genes, they were quantified by qPCR. All results matched with the probe sets present in the GeneChip expression set, and presented with similar fold change difference between carriers of the *KLF1* gene variation and normal family members.

Of particular interest are the two cell cycle genes, E2F2 and E2F4. E2F2 and E2F4 are transcription activators that bind DNA cooperatively with DP proteins through the E2 recognition site, 5'-TTTC[CG]CGC-3' found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication. The DRTF1/E2F complex functions in the control of cell cycle progression from G1 to S phase. E2F2 binds specifically to RB1 protein, in a cell cycle dependent manner (Ivey-Hoyle, Conroy et al. 1993), whilst E2F4 binds with high affinity to RBL1 and RBL2. The E2F2 and E2F4 genes are expressed 2 fold lower on average in *KLF1* +/- versus cells that are *KLF1* wildtype. This leads to the speculation that primary *KLF1* +/- erythroid progenitors have a delayed cell cycle progression from G1 to S phase. E2F4 mainly regulates fetal erythropoiesis through the promotion of cellular proliferation (Kinross, Clark et al. 2006), and an impaired level of E2F4 generally perturbs the normal physiological development and maturation of erythrocytes. Same results are obtained from expression profile data of Eklf1 -/- fetal liver cells 13.5*dpc*, where the aforementioned E2f2, E2f4, Rbl1, Rbl2 together with different cyclins (A, E) are downregulated, pointing out to a cell cycle block in the G1 phase (Pilon, Arcasoy et al. 2008).

Two other genes that appear to be 2-fold lower by qPCR in *KLF1* +/- compared to normal cells, are Alpha Hemoglobin Stabilizing Protein (*AHSP* or ERAF) and *SLC4A1*. *GYPA,* which was not verified by qPCR, was also found to be reduced in the expression profile data. AHSP is an abundant erythroid-specific protein that forms a stable complex with free alpha-globin, thus, enabling it to prevent alpha globin denaturation and precipitation (Kihm, Kong et al. 2002). As it was previously shown (Drissen, von Lindern et al. 2005) Ahsp expression appears to be consistently down-regulated in Eklf1 -/- murine erythroid cells. SLC4A1, also known as Band3, is the major integral glycoprotein of the erythrocyte membrane. Band3 has two functional domains. Its integral domain mediates a 1:1 exchange of inorganic anions across the membrane, whereas its cytoplasmic domain provides binding sites for cytoskeletal proteins, glycolytic enzymes, and hemoglobin (Lux, John et al. 1989). In addition, it has been shown that Band3 regulates the levels of GYPA in the red cell membrane (Auffray, Marfatia et al. 2001). It is shown to be affected in *KLF1* gene mutations in humans (Singleton, Burton et al. 2008), altering most probably the differentiation process of the erythroid cells. This notion is supported by the cell cycle defects indicated by the deregulation of many cyclins (A, E) and cell cycle checkpoint proteins (RBL1) we encountered in the expression profile of the Maltese high HbF

family members. GYPA has been shown to be essential during the differentiation process and erythroblast enucleation (Lee, Gimm et al. 2004), and even more to protect reticulocytes from phagocytosis along with CD47 (Oldenborg, Zheleznyak et al. 2000), while extruded nuclei lacking the negative charge imparted by GYPA, are readily engulfed by macrophages following enucleation. EpoR and CD71 (Transferrin Receptor) levels are also decreased (Fig. 4), a sign of possible loss or delay of important signalling events necessary for proper erythropoiesis (Drissen, von Lindern et al. 2005; Pilon, Arcasoy et al. 2008).

BCL11A was also among the genes that were found to be downregulated in family members bearing the *KLF1* mutation. *BCL11A* has recently been implicated in globin gene switching (Lettre, Sankaran et al. 2008; Uda, Galanello et al. 2008). Most of these studies have pinpointed a number of SNPs that appear to be linked with varying HbF levels. These SNPs generally fall in intronic regions, whilst others give rise to *BCL11A* variants (Menzel, Garner et al. 2007; Uda, Galanello et al. 2008), however the exact reason why these variants cause an increase in HbF is poorly understood. In this study, *BCL11A* was 2.8 fold lower in *KLF1* +/- than in *KLF1* normal erythroid cells. Whether there is an interaction between *KLF1* and *BCL11A* to coregulate HBG gene expression remains to be shown by *KLF1* IP, although *BCLL11A* IP did not support such an interaction neither a BCL11A appears to regulates *KLF1* expression (Sankaran, Menne et al. 2008). Our data suggest that KLF1 could, on the contrary, regulate *BCL11A* levels, and this should be proven with chromatin IP (ChIP) experiments. Additionally, ChIP would be very informative when comparing *KLF1* occupancy of normal and mutant *KLF1* erythroid cells in HBG and HBB promoters since the affinity normally is higher for the HBB (Donze, Townes et al. 1995).

In essence, our data suggest that the genomic region on chromosome 19p13.12-13 controls HbF production in adult life. At present, it is unclear to which extent the *ASF1B* and *KLF1* genes contribute to the effect observed in these family members. It remains to be seen whether these genes can hold promise for the design of novel therapeutic modalities for β-type hemoglobinopathies.

References

Auffray, I., S. Marfatia, et al. (2001). «Glycophorin A dimerization and band 3 interaction during erythroid membrane biogenesis: in vivo studies in human glycophorin A transgenic mice.» Blood 97(9): 2872-8.

Beug, H., S. Palmieri, et al. (1982). "Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus." Cell 28(4): 907-19.

Collins, N. and G. M. Kupfer (2005). "Molecular pathogenesis of Fanconi anemia." Int J Hematol 82(3): 176-83.

Craig, J. E., J. Rochette, et al. (1996). "Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach." Nat Genet 12(1): 58-64.

Donze, D., T. M. Townes, et al. (1995). "Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching." J Biol Chem 270(4): 1955-9.

Dover, G. J., K. D. Smith, et al. (1992). "Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2." Blood 80(3): 816-24.

Drissen, R., M. von Lindern, et al. (2005). "The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability." Mol Cell Biol 25(12): 5205-14.

E. Kowenz, A. L., G. Doderlein , T. Graf , H Beug (1987). Modern Trends in Human Leukemia VII: 199–209.

Forget, B. G. (1998). "Molecular basis of hereditary persistence of fetal hemoglobin." Ann N Y Acad Sci 850: 38-44.

G.Stamatoyannopoulos , F. G. (2001). The Molecular Basis of Blood Diseases. Philadelphia PA, WB Saunders Company.

Galvani, A., R. Courbeyrette, et al. (2008). "In vivo study of the nucleosome assembly functions of ASF1 histone chaperones in human cells." Mol Cell Biol 28(11): 3672-85.

Garner, C., J. Mitchell, et al. (1998). "Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production, on chromosome 6q23." Am J Hum Genet 62(6): 1468-74.

Garner, C., N. Silver, et al. (2004). „Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin." Blood 104(7): 2184-6.

Gilman, J. G. and T. H. Huisman (1985). "DNA sequence variation associated with elevated fetal G gamma globin production." Blood 66(4): 783-7.

Ivey-Hoyle, M., R. Conroy, et al. (1993). "Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F." Mol Cell Biol 13(12): 7802-12.

Jiang, J., S. Best, et al. (2006). "cMYB is involved in the regulation of fetal hemoglobin production in adults." Blood 108(3): 1077-83.

Kihm, A. J., Y. Kong, et al. (2002). "An abundant erythroid protein that stabilizes free alpha-haemoglobin." Nature 417(6890): 758-63.

Kinross, K. M., A. J. Clark, et al. (2006). "E2f4 regulates fetal erythropoiesis through the promotion of cellular proliferation." Blood 108(3): 886-95.

Kong, A. and N. J. Cox (1997). "Allele-sharing models: LOD scores and accurate linkage tests." Am J Hum Genet 61(5): 1179-88.

Lee, J. C., J. A. Gimm, et al. (2004). "Mechanism of protein sorting during erythroblast enucleation: role of cytoskeletal connectivity." Blood 103(5): 1912-9.

Lettre, G., V. G. Sankaran, et al. (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(33): 11869-74.
Liu, P., J. R. Keller, et al. (2003). "Bcl11a is essential for normal lymphoid development." Nat Immunol 4(6): 525-32.
Lux, S. E., K. M. John, et al. (1989). «Cloning and characterization of band 3, the human erythrocyte anion-exchange protein (AE1).» Proc Natl Acad Sci U S A 86(23): 9089-93.
Menzel, S., C. Garner, et al. (2007). "A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15." Nat Genet 39(10): 1197-9.
Miller, S. A., D. D. Dykes, et al. (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." Nucleic Acids Res 16(3): 1215.
Miyoshi, K., Y. Kaneto, et al. (1988). "X-linked dominant control of F-cells in normal adult life: characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome." Blood 72(6): 1854-60.
Oldenborg, P. A., A. Zheleznyak, et al. (2000). "Role of CD47 as a marker of self on red blood cells." Science 288(5473): 2051-4.
Pilon, A. M., M. O. Arcasoy, et al. (2008). "Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2." Mol Cell Biol 28(24): 7394-401.
Sankaran, V. G., T. F. Menne, et al. (2008). "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A." Science 322(5909): 1839-42.
Singleton, B. K., N. M. Burton, et al. (2008). "Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype." Blood 112(5): 2081-8.
Thein, S. L., S. Menzel, et al. (2007). "Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults." Proc Natl Acad Sci U S A 104(27): 11346-51.
Thiele, H. and P. Nurnberg (2005). "HaploPainter: a tool for drawing pedigrees with complex haplotypes." Bioinformatics 21(8): 1730-2.
Treisman, R., S. H. Orkin, et al. (1983). "Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes." Nature 302(5909): 591-6.
Uda, M., R. Galanello, et al. (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(5): 1620-5.
Vegiopoulos, A., P. Garcia, et al. (2006). "Coordination of erythropoiesis by the transcription factor c-Myb." Blood 107(12): 4703-10.
von Lindern, M., W. Zauner, et al. (1999). "The glucocorticoid receptor cooperates with the erythropoietin receptor and c-Kit to enhance and sustain proliferation of erythroid progenitors in vitro." Blood 94(2): 550-9.
Zhao, Q., G. Rank, et al. (2009). "PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing." Nat Struct Mol Biol 16(3): 304-11.

CHAPTER 4

Generation of a naïve llama Heavy-Chain Antibody (HcAb) library as a tool for in vitro reporter assays

Petros Papadopoulos, Frank Grosveld and Dubravka Drabek

Generation of a naïve llama Heavy-Chain Antibody (HCAb) library as a tool for *in vitro* reporter assays

Petros Papadopoulos, Frank Grosveld and Dubravka Drabek

Abstract

Antibodies based on the heavy chain (H) only, devoid of the light chain (L), are found in camels, dromedaries, llamas and sharks as part of their humoral immune response together with the conventional antibodies that contain both heavy (H) and light (L) chains. The variable region of the heavy chain only antibodies (VHH) have unique physical properties and most importantly they exert a diverse antigen-binding capacity due to their enlarged CDR3 region accompanied by strong affinities for their antigens. Randomization of the CDR3 is a common strategy to generate synthetic libraries of a camelid VHH scaffold, or a 'camelised' human heavy chain variable region (VH) or a mouse VH. Here, we report the generation of a natural heavy chain antibody (HCAb) library from peripheral blood of three llamas in a lentivirus vector as a tool for *in vitro* reporter assays.

Introduction

Antibodies are excellent protein-based high binding affinity tools. Innovative recombinant DNA technologies, including chimerisation and humanization, have enhanced the clinical efficiency of murine monoclonal antibodies, and have led to regulatory approval of immunoglobulins and monovalent antibody fragments (Fab) molecules as treatment for cancer and infectious and inflammatory diseases. The objective is to produce highly antigen-specific antibodies (Abs) with the lowest immunogenicity for humans at the lowest costs. The use of transgenic mice carrying human immunoglobulin loci (Green, Hardy et al. 1994) was the first step towards the production of immuno-compatible human Abs that could be obtained by hybridoma technology. Recently, recombinant Abs have been dissected into minimal binding fragments. They can be rebuilt into multivalent high-avidity reagents and fused to a range of molecules (Lu, Jimenez et al. 2003).

Heavy chain only antibodies (HCAbs), naturally occurring in camelids (camels and llamas), are a unique type of antibodies devoid of light chains and the first constant region of the heavy chain (CH1). Together with conventional antibodies with heavy and light chains they are part of the antibody repertoire of the camelid immune response to antigens. The variable domain of the heavy immunoglobulin chain, referred to as VHH (Muyldermans, Atarhouch et al. 1994), is the smallest antigen-binding fragment (15kD) described. Due to their small size, high expression yields (Frenken, van der Linden et al. 2000) and ease of purification, high solubility and stability combined with high affinity for the antigens and, specifically, the ability to recognize unique conformational epitopes, makes them excellent tools for genetic engineering. Importantly, they share close homology to human VH which makes it straightforward to generate 'humanized' VHH antibodies (Vu, Ghahroudi et al. 1997; Janssens, Dekker et al. 2006).

It has been reported that HCAbs can be potent competitive inhibitors of enzymes (Lauwereys, Arbabi Ghahroudi et al. 1998), viral protein assembly in the cytoplasm (Dekker, Toussaint et al. 2003), in the nucleus (Zemel, Berdichevsky et al. 2004) and in the secretory compartments (Paganetti, Calanca et al. 2005). Many therapeutic targets are intracellular proteins and molecules designed to interact with them, must effectively bind to their target inside the cell. Here, we report the generation of an intracellular heavy chain Ab (HCAb) library made from peripheral blood of three non-immunized llamas. This naïve antibody library contains antibodies that are not

directed specifically to a limited number of antigenic epitopes during an immune response, but those produced under steady state physiological conditions.

We aim to use this library of HCAbs in fetal liver cell lines with the dual reporter β-globin locus, in which two fluorescent proteins, the GFP and DsRed, have been introduced to the transcription start site of γ- and β-globin genes respectively. The objective is to induce expression of GFP (γ-globin) in the cells by blocking potential repressors or complexes that keep the γ-globin gene silenced in the steady state.

Materials and Methods

HCAb library from Peripheral Blood Cells (PBC)

Peripheral blood (100-150ml) from three llamas (2 males and 1 female) was collected and Ficoll fractionated. The lymphocytes located at the interphase were collected and RNA was extracted with Trizol (Invitrogen). cDNA was made with oligo dT primers and Reverse Transcriptase (Invitrogen) as indicated by the manufacturer. VHHs and VH cDNAs were amplified with a forward degenerate primer fused to a nuclear localization signal (NLS) sequence and the specific AttB1 sites for Gateway recombination (Invitrogen), Fwr: 5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TCC ACC ATG GAT CCA AAA AAG AAG AGA AAG GTA GCC SAG GTS MAR CTG CAG SAG TCW GG-3' and Reverse primer specific for the hinge sequence carrying the AttB2 sites (Gateway Tech), Rev:' 5'-GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC TGG TTG TGG TTT TGG TGT CTT GGG TT-3'. The following program was used: 94°C for 3min, 94°C for 1min, 58°C for 1min, 72°C for 1min, followed for 28-30 cycles and a final step of amplification at 72°C for 10min.

The PCR product was gel purified to exclude primer dimers and used directly in the following steps of Gateway recombination.

Gateway recombination and Lentiviral transduction

Following the Gateway System protocols (Invitrogen), the PCR product (VHH) was subject to two rounds of recombination creating an Entry pool of VHH clones (BP recombination) and subsequently an Expression pool of VHH clones (LR recombination). pDON221 (Invitrogen) was used as a host vector for the BP reaction (Entry clones) and the pLenti6.3-V5 DEST (Invitrogen) was the expression host vector for the pool of VHH antibody clones.

Alternatively, a different home-made lenti-vector (pAD5) was modified by introducing the att-ccdB-att cassette (Invitrogen) to fit with the Gateway System and used to compare transduction efficiencies of individual clones and pools of VHH antibodies on mouse fetal liver cells. The VHH antibodies expressed had the c-Myc tag (EQKLISEEDL) at their C-terminus.

Virus production was done according to the protocol of Invitrogen but scaled up to T-175cm^2 flasks using 293FT cells to transfect the Expression vector-VHH-V5 pool DNA together with the packaging mix (Invitrogen). Virus was harvested after 2

days and centrifuged for 2hr at 20Krpm at 4°C (Ultra-Centrifuge, Beckman).

Transduction of the fetal liver cells was carried out in 24-well plates using 10μg/ml Polybrene (Sigma) as suggested by the manufacturer (Invitrogen).

Fingerprinting of VHH clones

The DNA isolated from single colonies of the VHH clones introduced in the Expression vector pLenti6.3-V5 was amplified using the following primers (Fwr: 5'- CGC AAA TGG GCG GTA GGC GTG-3' and Rev: 5'-ACC GAG GAG AGG GTT AGG GAT-3') and the PCR products were digested with HinfI restriction enzyme and run on a 4% NuSieve agarose gel in TBE buffer.

Single-Cell PCR

Multiplex PCR reactions were performed on single fetal liver cells spotted on a glass slide (AmpliGrid, ADvalytix) by the FACSAria Sorter. Primers for a housekeeping gene (Calmodulin) are: Fwr: 5'-AAG GAT TCC GTG TGT TTG ATA AG-3', Rev:5'-TCA CTT CGC TGT CAT CAT TTG TA-3'. Specific primer set on sequences flanking the VHH fragments are: Fwr: 5'-CGC AAA TGG GCG GTA GGC GTG-3', Rev: 5'-ACC GAG GAG AGG GTT AGG GAT-3'.

The reaction was performed in a Thermal Cycler of Advalytix (Advalytix, Munich) in a total volume of 1μl 2μM primer concentration, according to the protocol of Advalytix. The PCR program used was: 95°C for 3min, 95°C for 25sec, 60°C for 75sec, 72°C for 75sec, followed by 45 cycles and a final step of 72°C for 1min.

Flowcytometry

Transduced cells were washed in phosphate buffer saline (PBS) and fixed with formaldehyde for 20min in the dark at 4°C using the Cytofix/Cytoperm kit (Becton Dickinson). Fixed cells were incubated with V5-FITC (Invitrogen) antibody in saponin buffer for 20min at 4°C in the dark and then washed three times in the same buffer. Finally, the cells resuspended in saponin buffer were analyzed in the FACScan or FACSAria (Becton Dickinson).

Western blotting

Nuclear extracts or whole cell lysates were prepared from transduced mouse fetal liver cells and run on 15% SDS-polyacrylamide gels. The proteins were transferred

to a PVDF (Millipore) membrane and blocked with 3% Milk dry powder-0.2% Tween20 in PBS for 1hr at room temperature (RT). Incubation with the V5-HRP (Invitrogen) antibody (1/5000 dilution) for 2hr at RT was followed by four washes of 10min with 0.2% Tween20 in PBS. The membrane was developed with Enhanced Chemiluminescent Western blotting detection (ECL) kit (Amersham Biosciences).

When the pAD5 lenti-vector was used to produce virus and transduce fetal liver cells, incubation with mouse c-Myc antibody (Covance- MMS-150P (9E10), 1/1000 dilution) was followed by incubation with anti-mouse HRP conjugated IgG antibody (1/15000 dilution) prior to ECL kit detection as described before.

Results and Discussion

The PCR primers used to amplify the variable region of the llama heavy immunoglobulin chain, amplify both heavy-chain only (VHH) and conventional antibodies' heavy chain (VH). Thus, we calculated from a Real-Time PCR analysis the proportion and number of molecules of VHH amplified in a mixed cDNA prepared from the isolated lymphocytes of the three llamas as compared to a single clone of plasmid VHH cDNA.

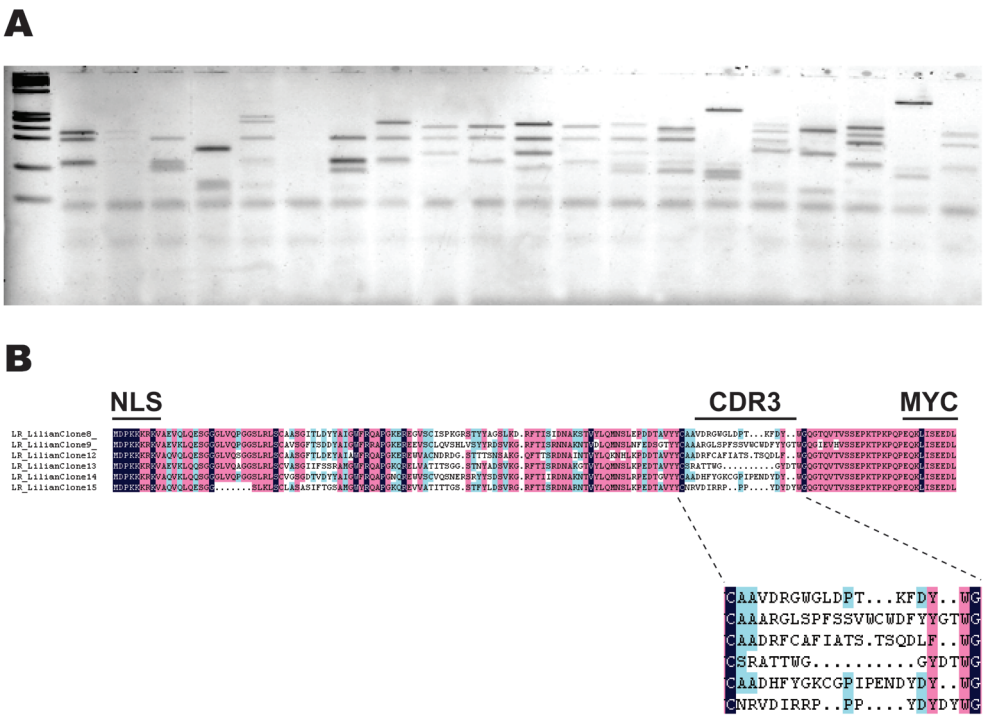


Figure 1

A. Fingerprinting of VHH clones analyzed after the second recombination reaction (LR recombination, Gateway System) as expression clones in pLenti-6.3-V5DEST vector. Here, twenty clones are represented out of a hundred analyzed.

B. Sequence analysis of six VHH clones with the NLS, CDR3 and Myc tag shown.

The significant number of VHH molecules present in the mixed cDNA preparation indicated that a complexity of at least 10^6 molecules could be achieved with the library. However, recent data (Kastelic, Frkovic-Grazio et al. 2009) show that specific, good affinity and efficiently expressed VH domains of conventional antibodies can be selected from co-amplification products of VH and VHH cDNAs. VHHs differ from VHs by V37/F, G44E/Q, L45R, W47F/G/S/L substitutions on conserved VH residues belonging to the second framework (FR2). Because llama VHs show higher sequence and structural homology with human VHIII group than llama VHHs (Vu, Ghahroudi et al. 1997) they constitute very interesting agents for therapeutic applications in humans. Fingerprinting analysis of hundred clones was in concordance with our prediction and showed only two clones with the same HinfI restriction pattern (Fig. 1 and data not shown).

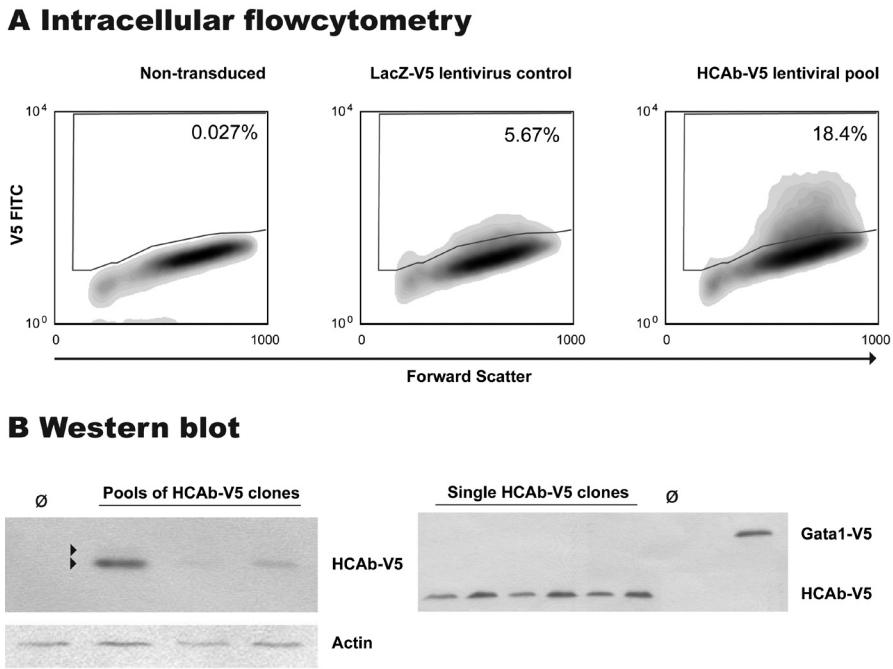


Figure 2

A. V5-FITC intracellular flowcytometry analysis of a fetal liver cell line transduced with a pool of HCAbs expressed in the pLenti6.3 DEST-V5 vector. LacZ-V5 was expressed in the same vector and used as a control for the V5-tag expression.

B. Western blot analysis for the V5 tag of single clones and pools of HCAbs expressed in the pLenti6.3 DEST-V5 vector in transduced fetal liver cells, (∅) stands for non-transduced cells.

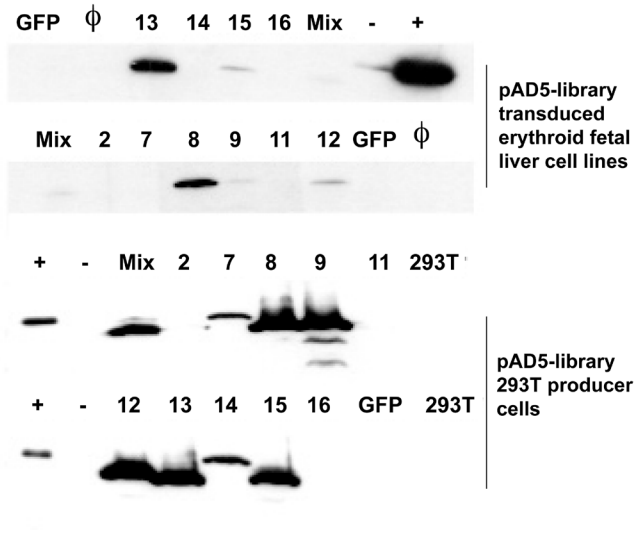


Figure 3: Western blot of single clones and a pool of HCAbs with a c-myc tag
Producer cell line 293T was also analyzed for the presence of antibodies to verify the virus production in the case of non-expressing clones in the transduced cells, *i.e.* clones 2, 14 and 16. (Mix) stands for virus of combined DNA of all single clones of antibodies, (+) and (-) for positive and negative control, (GFP) for cells transduced with a GFP containing lentivirus and (∅) for non-transduced cells.

Furthermore, a total of 10 clones were sent for sequencing and 6 clones were verified to be VHH fragments with a considerable variability in the CDR3 domain (Fig. 1B), which is mainly responsible for the diversity of antigen-binding surface of the HCAbs (Xu and Davis 2000; K.E Conrath 2003). Amongst the other 4 clones, clone Nb16 contains a premature stop codon before the Myc tag. Clones Nb2 and Nb11 failed to display a contig sequence like the other VHHs and together with Nb16 did not express the Myc tag (see Fig. 3).

Fetal liver cells, primary or immortalized cell lines, were transduced with virus made from DNA of single clones or a pool of HCAbs using the two different expression vectors in order to compare the efficiency of transduction. Both constructs were comparable when single clones of HCAb were expressed, subject to some variability probably due to the integration site in the genome. However, when the pool of HCAbs was expressed in the cells, the vector system made a significant difference. The

pLenti6.3-V5 DEST showed more consistency in maintaining detectable levels of the pool VHH clones compared to the pAD5 vector (Fig. 2B and 3 and data not shown) in several experiments.

We decided to continue with the pLenti6.3-V5 DEST vector and we tested different virus concentrations per given cell number to check the optimal conditions in transduction efficiency and cell viability of the target erythroid fetal liver cell lines. We observed no significant alterations in cell morphology when single clones or pools of HCAb were expressed as compared to non-transduced cells or cells transduced with pLenti6.3-V5-LacZ virus, which was used as a control in every experiment. However, we noticed that there was significantly more cell death with the control (LacZ) virus as compared to HCAb virus. This explains why the V5 positive fraction of live cells after V5-intracellular FACS analysis was always lower in the LacZ-V5 than in the HCAb-V5 transduced cells (5.67% vs 18.4%, Fig. 2), indicating that the difference observed is not mainly due to the different virus titer of the two preparations, but mostly to potential cell toxicity caused by LacZ expression/overexpression.

Expression of the VHH antibodies was measured after maximum three weeks in culture, under the selection of Blasticidin (5-10μg/ml). Nuclear extracts were

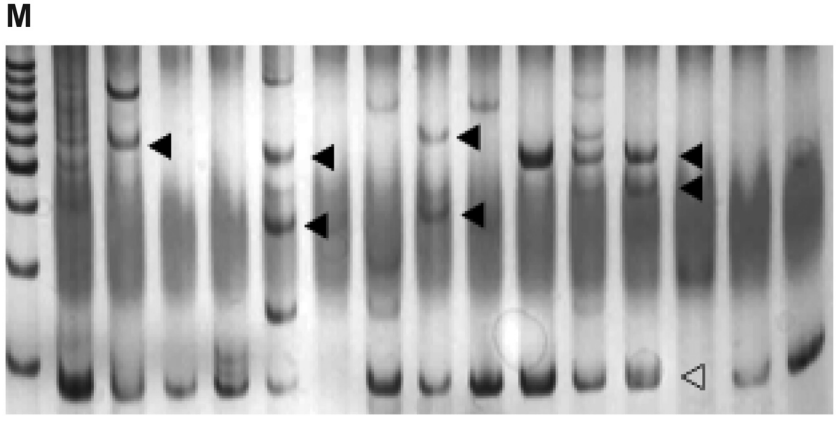


Figure 4
Multiplex Single-cell PCR to amplify VHH (◀) using the Calmodulin gene (◁) as a control, after transduction of fetal liver cells.

analyzed by Western blot with V5-Ab to verify expression of the VHH. At the same time we confirmed 100% viability of the cells in culture by daily microscopical and flowcytometry analysis.

Transduction of the HCAb lentivirus library to reporter cell lines is a useful tool aiming at the identification of an antibody that has a direct molecular effect in the cells. Additionally, the possibility of sorting single cells when the effect takes place in a very small population or if clonal studies need to be performed after transduction, prompted us to do so with fetal liver cells transduced with a pool of VHH-V5 tagged HCAbs. The result (Fig. 4) shows that after sorting live cells we were able to amplify a housekeeping gene (Calmodulin) and a potential VHH fragment in every single cell. There were also cells in which no other band than the one of the housekeeping gene was detected, because only live (Hoechst negative) cells were selected and not virus-transduced cells (V5 positive). A combination of fluorescence gates for the expression of a gene of interest (GFP for γ-globin) and/or the expression of V5 tag can easily make the sorting more specific for VHH expressing cell clones. If cells have to be sorted depending on the phenotype and grown in culture, then no V5-positive cells can be gated and consequently more cells have to be further analysed/ processed including the false-positive cells sorted due to autofluorescence.

We are currently aiming at the optimization of the assay for our dual reporter cell lines of GPA-GFPγ. We are in the process of isolating cells expressing the GFP (γ-globin) after transduction with the HCAb library of clones. We would like to identify the VHH/VHHs responsible for the expression of the gene (GFP), acting possibly by blocking a potential repressor protein or complex involved in the regulation of the γ-globin. We have already performed experiments that allowed us to calculate the range of transduced cells (10^6 cells) to be sorted in order to isolate sufficient positive cells for further analysis. False-positives and real positives will be verified by single-cell PCR and subsequently read the sequence of the antibody in order to subclone it and re-express it in human erythroid cells.

Acknowledgements

We are grateful to the veterinarian dr. H. Westerveld and the farmers of Hellevoetsluis (Zuid Holland) who were very helpful and kind enough to provide us with blood samples from llamas used in this study.

References

Dekker, S., W. Toussaint, et al. (2003). «Intracellularly expressed single-domain antibody against p15 matrix protein prevents the production of porcine retroviruses.» J Virol 77(22): 12132-9.

Frenken, L. G., R. H. van der Linden, et al. (2000). "Isolation of antigen specific llama VHH antibody fragments and their high level secretion by Saccharomyces cerevisiae." J Biotechnol 78(1): 11-21.

Green, L. L., M. C. Hardy, et al. (1994). "Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs." Nat Genet 7(1): 13-21.

Janssens, R., S. Dekker, et al. (2006). "Generation of heavy-chain-only antibodies in mice." Proc Natl Acad Sci U S A 103(41): 15130-5.

K.E Conrath, U. W., S. Muyldermans, V.K. Nguyen (2003). "Emergence and evolution of functional heavy-chain antibodies in Camelidae." Developmental & Comparative Immunology 27
87-103.

Kastelic, D., S. Frkovic-Grazio, et al. (2009). "A single-step procedure of recombinant library construction for the selection of efficiently produced llama VH binders directed against cancer markers." J Immunol Methods.

Lauwereys, M., M. Arbabi Ghahroudi, et al. (1998). "Potent enzyme inhibitors derived from dromedary heavy-chain antibodies." Embo J 17(13): 3512-20.

Lu, D., X. Jimenez, et al. (2003). "Di-diabody: a novel tetravalent bispecific antibody molecule by design." J Immunol Methods 279(1-2): 219-32.

Muyldermans, S., T. Atarhouch, et al. (1994). "Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains." Protein Eng 7(9): 1129-35.

Paganetti, P., V. Calanca, et al. (2005). "beta-site specific intrabodies to decrease and prevent generation of Alzheimer's Abeta peptide." J Cell Biol 168(6): 863-8.

Vu, K. B., M. A. Ghahroudi, et al. (1997). «Comparison of llama VH sequences from conventional and heavy chain antibodies.» Mol Immunol 34(16-17): 1121-31.

Xu, J. L. and M. M. Davis (2000). "Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities." Immunity 13(1): 37-45.

Zemel, R., Y. Berdichevsky, et al. (2004). "Inhibition of hepatitis C virus NS3-mediated cell transformation by recombinant intracellular antibodies." J Hepatol 40(6): 1000-7.

DISCUSSION

The globin switch from fetal to adult in human, is a process closely related to the gestational age of the infant and not affected by the changes in the environment and oxygen tension that occur at the time of birth (Bard 1973). After birth the decrease of fetal haemoglobin is accelerated, which suggests that there must be reasons associated with the 'new conditions of life' of the infant that accelerate this process. In the case of haemoglobin, a substitution in expression of γ-globin gene with β-globin gene in the adult stage generates obvious questions about the functional differences between the two globins and the possibility of a "common" cell program that controls the expression of both genes during each stage of development.

The study of human globin gene regulation has been largely carried out in transgenic mice containing the human β-globin locus. In the mouse there is one main globin switch at 11.5*dpc* when the embryonic genes εy and βh1 are replaced by βmaj and βmin. Thus, mice lack a fetal-specific switch and hemoglobin type. In human there are two main globin switches: embryonic to fetal, and fetal to adult. When human β-globin genes are expressed in the mouse in the context of the entire β-globin locus, expression levels of the human globins range from 39-53% of total mouse β-like globin RNA at different time points of mouse gestation 8.5*dpc*-16.5*dpc* (Strouboulis, Dillon et al. 1992). Interestingly, the two switches occur of which the switch from γ-globin to β-globin gene expression is delayed in comparison with the mouse main switch.

The importance of the human γ- and β-globin gene flanking regulatory sequences for their stage specific expression in the mouse, was already assessed before the generation of the complete β-globin locus transgenic mice (Kollias, Wrighton et al. 1986; Kollias, Hurst et al. 1987). To investigate the role of different transcription factors bound to specific regulatory sequences or present in complexes found in the respective globin promoters, studies have been conducted in transgenic mice and erythroid cell lines (MEL, K562). The combination of loss of function and mutated DNA-binding sites of transcriptional regulators, such as EKLF (Nuez, Michalovich et al. 1995; Perkins, Sharpe et al. 1995; Wijgerde, Gribnau et al. 1996), NF-E2 (Zhao, Zhou et al. 2006; Aerbajinai, Zhu et al. 2009), GATA1 (Berry, Grosveld et al. 1992; Chen, Luo et al. 2008), DRED (Tanabe, Katsuoka et al. 2002), C/EBPγ (Zafarana, Rottier et al. 2000), COUP-TF (Tanabe, Katsuoka et al. 2002) and coregulators, such as Brg1 (Bultman, Gebuhr et al. 2005) indicates that more than one protein is responsible for the silencing of the γ-globin gene in the adult stage. Loss of function studies of candidate single factors either reactivate γ-globin gene expression at

a low level (cell lines), or do not affect postnatal silencing despite upregulating γ-globin gene expression during gestation (*in vivo*). These parameters are crucial for providing evidence of a stage- and gene specific regulator of γ-globin.

Testing the loss of function of candidate proteins for γ-globin gene regulation is now easier with the existence of siRNA/shRNA libraries. In order to generate a high-throughput system that mimics the mouse developmental expression pattern of human globin genes we have generated a reporter β-globin locus transgenic mouse with two fluorescent proteins, GFP and DsRed, introduced at the start site of the γ- and β-globin genes respectively. The advantage of this reporter compared to others generated is that the established fetal liver cell lines derived from the transgenic mouse avoid possible position effect variegation (PEV) compared to transfected erythroid cells (Grosveld F 1996; Vadolas, Wardan et al. 2002). Furthermore, the use of shorter constructs lacking genomic sequences of the β-globin locus (Skarpidi, Vassilopoulos et al. 2000) can be informative but falls short as a model to study gene regulatory mechanisms within the context of the complete β-globin locus. A recent example is provided by the identification of a regulator of fetal haemoglobin, BCL11A, in human cells (Sankaran, Menne et al. 2008). BCL11A binds to genomic sequences in the β-globin locus but it does not interact directly with the γ-globin promoter.

Two different constructs were introduced in the start site of the Aγ globin gene, one with GFP cDNA (GFPγ) and another as a fusion protein with the erythroid-specific plasma membrane protein glycophorin A (GPA-GFPγ). The reason to generate the second mouse line was the low percentage of GFP positive cells detected in embryonic blood or fetal liver cells at 11.5-13.5*dpc* that are expected to express mainly γ-globin. There was a possibility that the low detection levels were due to fluorescence quenching by haemoglobin, since it had been reported in various cases (Spangrude, Cho et al. 2006; Swenson, Price et al. 2007). In an effort to optimize the expression and detection of the fluorescent protein, we generated a new transgenic mouse in which the expression of γ-globin was monitored by GFP located in the surface of the erythrocyte membrane (GFP-GPA fusion). During development, GPA-GFPγ was better detected by flowcytometry (higher MFI) compared to the GFPγ. Loss of function experiments of reported candidate genes, including HDAC3 (Mankidy, Faller et al. 2006), Myb (Jiang, Best et al. 2006) and Bcl11a (Sankaran, Xu et al. 2009) confirmed the validity of the reporter cell lines. Considering the species-specific function of BCL11A described recently (Sankaran, Xu et al. 2009), it is necessary to combine the studies in the mouse model with human cells in order to avoid possible

developmental differences among species regarding the time frame of expression of transcription factors.

Our results from studies with chemical compounds (*e.g.* HDAC inhibitors, azacytidine, hydroxyurea, short-chain fatty acids) suggest that caution should be taken when the morphology and cytoplasmic density of the cells change dramatically due to the effects of the drug. The inclusion of several controls in the flowcytometry analysis to exclude autofluorescence is necessary to avoid false positive results. This discernment becomes essential when the number of expected positive events is very low, as it is the case of fetal haemoglobin expression in a restricted low percentage of cells, named the F cells. In the *in vivo* experiment of azacytidine treatment of our transgenic mice, we were able to distinguish autofluorescence from real fluorescence of GFP by implementing state of the art flowcytometry analysis (FACSAria sorter Advance) and several controls that revealed reactivation of the γ-globin gene in a small percentage of blood cells.

The quest for γ-globin gene repressor candidates and an extensive search of the literature revealed a report on the regulation of postnatal silencing of liver genes by the transcription factor TAF10 (Tatarakis, Margaritis et al. 2008), a member of the TFIID basal transcriptional machinery complex. Thus, we studied the effects of TAF10 knockdown in erythroid cells. Initially, we carried out the experiments in the reporter mouse cell lines, where we detected induction of GFP and DsRed proteins, indicating a general induction in globin production due to loss of TAF10. This was followed by similar studies in human blood erythroid progenitors where the downregulation of TAF10 resulted in moderate induction levels of fetal haemoglobin when compared to BCL11A knockdown control cells (Sankaran, Menne et al. 2008). The implication of TFIID in the regulation of fetal haemoglobin has not been explored yet. The fact that TAF10 is necessary for correct postnatal silencing of liver genes and that knockdown of TAF10 affected globin expression prompted studies on the role of TAF10/TFIID during erythroid differentiation and at specific developmental stages, when erythroid progenitor cells express fetal or adult haemoglobin. Additionally, we investigated interactions of the TAF10/TFIID with erythroid transcriptional regulators and coregulators to provide information upon activation of the γ-globin promoter.

We performed TAF10 immunoprecipitation assays (IPs) in mouse reporter cell lines, one spontaneously expressing GFP up to 100% (GFPγCL) and a second one expressing DsRed (DsRedβCL) upon differentiation. We encountered differences in the complexes formed and stoichiometries of several members of the TFIID complex,

and also posttranslational modifications of members of SAGA complexes formed, including Kat2a, Supt7l, Fam48a, and several TAFs (TAF9b,TAF6,TAF6l,TAF12) that should be verified and functionally characterized. Even more interesting was the comparison between GFPγCL and DsRedβCL, in both induced and non-induced state. While a canonical TFIID complex was immunoprecipitated with TAF10 antibody in the DsRedβCL, TAF10 IPs in the GFPγCL lacked TBP and a group of three other proteins (Usp22, Atxn7, Eny2). These proteins form part of a subcomplex of the TFTC/STAGA complex described previously in Hela cells (Zhao, Lang et al. 2008) to remove ubiquitin moieties from H2A and H2B. This consistent difference sparked the hypothesis that distinct complexes are formed to drive specific globin expression in erythroid cells. We also observed the presence of a TBP-homologue protein (Tbpl2 or TRF3) in the DsRedβCL only, that shares high homology with TBP and has been reported to interact with specific TAFs and substitute TFIID in promoter recognition and tissue specific developmental activation of genes (Deato and Tjian 2007; Deato, Marr et al. 2008). It would be very interesting to encounter analogous complexes during erythroid differentiation. Furthermore, there are reports that exclusively restrict the expression of TRF3 in ovarian cells (Gazdag, Santenard et al. 2009) in contrast to other reports (Persengiev, Zhu et al. 2003) that show ubiquitous expression. Verification of the presence of TRF3 in combination or not with TBP is an interesting issue since we envisage control of different groups of genes by each factor, present at the same time or by mutually exclusive expression and direct involvement in the activation of the same promoters.

Erythroid-specific transcription factor interactions with the TAFs are also important in order to understand the order of events prior to gene activation of globins. It has been reported that EKLF interacts with TAF9 (Sengupta, Cohet et al. 2009) and NF-E2 with TAF4 (Amrolia, Ramamurthy et al. 1997) regulating β-globin expression, and this emphasizes the importance of the identification of such interactions with other TAFs in globin expression as we propose to pursue with the TAF10 IP in mouse and human erythroid cells. We were able to detect more interactions with transcription factors when performing the TAF10 IP in human cells (*i.e.* GATA1, LDB1, TAL1) compared to mouse cells (*i.e.* Bcl11a, HP1, Yy1, Fop) that need to be verified. This suggests important interactions between regulators, coregulators and the basal transcriptional machinery. For example, the BCL11A interaction is missing in human fetal liver cells that express HbF. The suppressor activity of BCL11A on the γ-globin promoter it is thought to occur through its participation in a repressor complex. This

model is based on the observation that BLC11A does not interact directly with the γ-promoter (Sankaran, Menne et al. 2008). In addition, our TAF10 IP in human fetal liver erythroid progenitors identified most of the reported key interactors of BCL11A (GATA1, CHD4, RBBP4, MTA2, HDAC1/2, MATRIN3), indicating communication between a repressor complex (including BCL11A) and TFIID subunits.

The hypothesis for TFIID involvement in the regulation of HbF expression is also supported by the knockdown experiments of a recently identified protein, the friend of Prmt1 (Fop) (van Dijk, Gillemans et al. 2009) and (T.B. van Dijk, unpublished data) about the possible repressor activity on γ-globin gene expression. Fop was identified in our TAF10 IP in the DsRedβCL (expressing β-globin) and human Hela cell control, but not in the GFPγCL, neither in the human fetal liver cells that both express γ-globin at high levels. Furthermore Fop knockdown experiments performed in the DsRedβCL (data not shown) resulted in induction of GFP (γ-globin) at comparable levels as with the TAF10 knockdown. These results verify previous findings by T.B. van Dijk on parallel studies on human blood erythroid progenitors.

Genomic studies on a family with high HbF levels and identification of novel mutations of transcription factors

To identify transcription factors involved in the regulation of HbF in humans, we studied members of a Maltese family with high levels of fetal haemoglobin in adult life. Single nucleotide polymorphism (SNP) analysis revealed genomic alterations in the KLF1 and ASF1B genes and expression profiles of blood erythroid progenitors showed many deregulated genes, including downregulation of the β-globin gene. These genes are targets of the Klf1 transcription factor, as previously shown in the mouse (Nuez, Michalovich et al. 1995; Perkins, Sharpe et al. 1995; Wijgerde, Gribnau et al. 1996; Patrinos, de Krom et al. 2004). This is the first report linking the levels of HbF with mutations in the KLF1 gene in humans and opens a new line of research in understanding the function of the mutant forms and how reduced levels of KLF1 affect the levels of HbF. Knockdown of KLF1 in blood erythroid progenitors from healthy individuals did induce moderate levels of HbF but the effect of overexpression of the mutant form remains to be assessed, preferably in combination with KLF1 knockdown. A dominant-negative effect could be the mechanism, and since the transactivation domains of the KLF1 are maintained in the mutant form we assume that interactions with other transcription factors, in the

absence of DNA-binding, do occur. Such interactions with repressors (*e.g.* BCL11A) could result in the reactivation of the γ-globin gene by keeping them away from the promoter and altering the composition/function of a complex responsible for the suppression of γ-globin. It is interesting to find out which interactions of KLF1 are maintained or lost in the erythroid cells of the Maltese family members as this could lead to connections to other regulators or coregulators and members of the TFIID complex as we have observed with TAF10. Importantly, expression of BCL11A is reduced in the KLF1+/- erythroid progenitors consistent with Klf1KO expression profiling data in mice, suggesting that KLF1 regulates BCL11A expression by direct binding to its promoter through its potential KLF1 binding sites.

A common feature of the expression profiles of the cells from the high HbF family members and Eklf-/- mouse fetal livers at 13.5*dpc* (Pilon, Arcasoy et al. 2008) is deregulation of the cell-cycle genes, including E2F2, E2F4, E2F7, Rbl1 and cyclins A and E. Similar effects are observed in the TAF10 *null* F9 cells (murine embryonic carcinoma) that accumulate in G_1/G_o phase with a strong reduction in cyclin E expression and phosphorylation levels of Rbl1 (Metzger, Scheer et al. 1999). Interestingly, Taf10 appears to be downregulated together with Taf4b, Taf5 and Taf9 in the Eklf-/- mouse fetal liver cells, suggesting involvement of TFIID in the cell cycle defect and a role in the activation of the γ-globin gene promoter.

Another gene with an interesting expression profile is the α-fetoprotein (AFP). The initial observation that the GFPγCL under proliferation conditions expresses AFP at very high levels (200-fold increase compared to the control cell line, Appendix I) as well as GFP (γ-globin), is akin to the derepression of AFP in the liver of adult TAF10 knockout mice (Tatarakis, Margaritis et al. 2008). In the Eklf-/- mouse fetal liver expression profile, AFP shows an average of 40 fold upregulation compared to the control wild type cells as well as Hbb1-bh1 a 400 fold increase (supplementary data, (Pilon, Arcasoy et al. 2008)). The effect of AFP and its derivative peptide Growth Inhibitory Peptide (GIP) were tested in mouse fetal liver cell lines (GFPγCL and DsRedβCL) and human cells (fetal liver and blood erythroid progenitors) in order to assess whether AFP / GIP could directly induce an increase in fetal haemoglobin levels, or whether *AFP* is just a gene regulated in the same manner. We observed no significant difference in the human cells, and thus we assume that there must be similarities in the way that both AFP and HbF proteins are regulated and are silenced in the adult stage rather than regulate each other.

Taking together all the data, the hypothesized role of TFIID in fetal haemoglobin

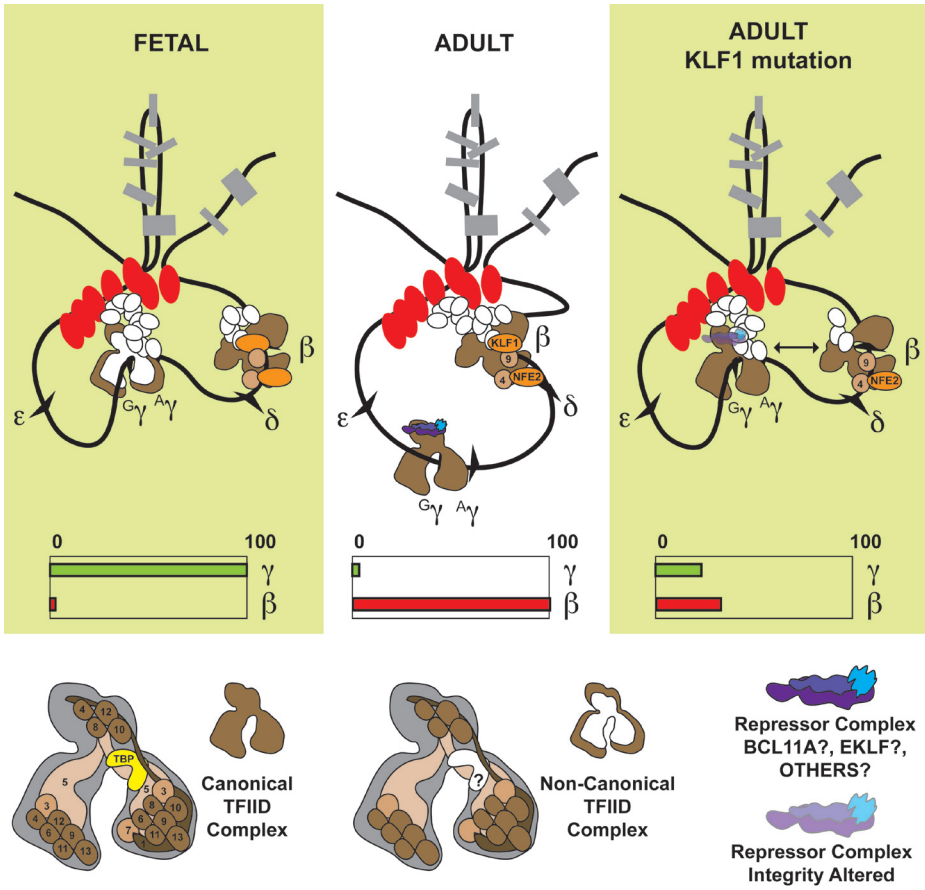


Figure 1: Model of γ-globin gene regulation by TFIID

TFIID is the sensor and executor of conformational changes that affect gene transcription. The interaction of different factors with members of the TFIID complex is different during different developmental stages. These interactions result in structural changes directly affecting further interactions of its subunits that accommodate or facilitate DNA contacts that may stabilize the binding of TFIID to DNA during PIC assembly.

One of these interactions involves the repressor complex that is present in the adult stage in the γ-globin promoter. It contains known proteins implicated in the γ-globin gene regulation, including BCL11A, KLF1 (not yet directly involved) and others. Its composition will be altered as a consequence of transcription factor imbalance (*e.g.* caused by KLF1 mutation) that results in BCL11A downregulation and produces further changes in the interactions of TFIID subunits (TAFs, *e.g.* TAF10) with regulators and coregulators of the repressor complex. Ultimately they may lead to the displacement of the repressor and full activation of the silenced promoter of the γ-globin gene.

regulation is gaining strength. A working model is presented in Figure 1.

When KLF1 binds to the CACC sites and the DPE element of the β-globin promoter, the chromatin state favours the recruitment of other transcriptional activators and coregulators that promote assembly of a stable preinitiation complex (PIC) and finally efficient transcription of the β-globin gene occurs. At that time the LCR is establishing the appropriate chromatin architecture by introducing chromatin remodellers and loading transcription factors and chromatin modifying enzymes, enhancing the transcriptional activity of the β-globin promoter (Fig. 1 adult). Total loss of Klf1 in mice results in the interruption of the stable interactions between the LCR and the β-globin promoter (Drissen, Palstra et al. 2004) and consequently loss of its hypersensitivity (Wijgerde, Gribnau et al. 1996) and transcriptional activity, similar to what occurs when the proximal β-promoter is deleted in the context of the whole β-globin locus (Patrinos, de Krom et al. 2004). As TAF9 is recruited in a KLF1-dependent manner (Sengupta, Cohet et al. 2009), loss of KLF1 initiates structural changes in TFIID, reducing transcription of the β-globin gene. In the case of the KLF1 heterozygote mutant, the frequency of interactions of the β-globin promoter with the LCR may be reduced (Fig. 1 KLF1 mutation). Interaction of the LCR and the γ-promoter now are happening in a “suppressor-free” environment due to the altered composition of the repressor complex or even its displacement from the promoter. In any case many interactions involving the TAFs are lost, similarly to a TAF10 loss, provoking structural TFIID changes which enable PolII recruitment and initiation of transcription.

Evidence of the structural changes in TFIID upon activator binding was recently shown (Liu, Coleman et al. 2009) as has been reported for other cofactors facilitating the communication between gene-specific transcription factors and components of the general transcriptional machinery (Naar, Taatjes et al. 2002; Taatjes, Naar et al. 2002; Grob, Cruse et al. 2006; Liu, Coleman et al. 2008). If new interactions would be established, reversion of repressor(s) binding to different TAFs may occur in a similar manner as described when TAF10 is lost and TFIID is disassembled to its subunits (Tatarakis, Margaritis et al. 2008) taking away the repressor protein/complex and reactivating silenced genes. As a result, the γ-globin gene may be transcribed with the LCR shared between β- and γ-globin genes in terms of frequency of efficient interactions. That suggests the existence of a complex rather than a single repressor mediating the silencing of the γ-globin gene. Conceptually, crucial TFIID (TAF)-repressor (part of the repressor complex) interactions would be key interactions to

be targeted in order to reverse suppression of the gene (Fig. 1 KLF1 mutation). The observation that TAF10 interacts with most of the BCL11A interactors reported so far, supports the idea that different ways of disrupting interactions of the “repressor complex” with the TFIID exist.

The broad diversity of canonical and non-canonical TFIID complexes in core promoter recognition and transcription mediation appears to be involved in the complex developmental expression patterns of several postnatally silenced genes, including the globin genes as suggested in this thesis. More evidence is needed to reinforce this idea and future investigation might provide logical scenarios on the role of TFIID in the pathways regulating transcription of the globin genes.

References

Aerbajinai, W., J. Zhu, et al. (2009). «SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII.» Blood 114(1): 187-94.

Amrolia, P. J., L. Ramamurthy, et al. (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(19): 10051-6.

Bard, H. (1973). "Postnatal fetal and adult hemoglobin synthesis in early preterm newborn infants." J Clin Invest 52(8): 1789-95.

Berry, M., F. Grosveld, et al. (1992). "A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin." Nature 358(6386): 499-502.

Bultman, S. J., T. C. Gebuhr, et al. (2005). "A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in beta-globin expression and erythroid development." Genes Dev 19(23): 2849-61.

Chen, Z., H. Y. Luo, et al. (2008). "A T-to-G transversion at nucleotide -567 upstream of HBG2 in a GATA-1 binding motif is associated with elevated hemoglobin F." Mol Cell Biol 28(13): 4386-93.

Deato, M. D., M. T. Marr, et al. (2008). "MyoD targets TAF3/TRF3 to activate myogenin transcription." Mol Cell 32(1): 96-105.

Deato, M. D. and R. Tjian (2007). "Switching of the core transcription machinery during myogenesis." Genes Dev 21(17): 2137-49.

Drissen, R., R. J. Palstra, et al. (2004). "The active spatial organization of the beta-globin locus requires the transcription factor EKLF." Genes Dev 18(20): 2485-90.

Gazdag, E., A. Santenard, et al. (2009). "TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte." Genes Dev 23(18): 2210-23.

Grob, P., M. J. Cruse, et al. (2006). "Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues." Structure 14(3): 511-20.

Grosveld F, A. M., Berry M, de Boer E, Dillon N, Drabek D, Ellis J, Fraser P, HAley J, Philipsen S, Pruzina S, Raguz-Bolognesi S, Smith D, Trimbon T, Wijgerde M (1996). "Drug discovery and the transcriptional control of the human beta globin locus." Genomes, molecular biology and drug discovery: 117-27.

Jiang, J., S. Best, et al. (2006). "cMYB is involved in the regulation of fetal hemoglobin production in adults." Blood 108(3): 1077-83.

Kollias, G., J. Hurst, et al. (1987). "The human beta-globin gene contains a downstream developmental specific enhancer." Nucleic Acids Res 15(14): 5739-47.

Kollias, G., N. Wrighton, et al. (1986). "Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns." Cell 46(1): 89-94.

Liu, W. L., R. A. Coleman, et al. (2008). "Structural changes in TAF4b-TFIID correlate with promoter selectivity." Mol Cell 29(1): 81-91.

Liu, W. L., R. A. Coleman, et al. (2009). «Structures of three distinct activator-TFIID complexes.» Genes Dev 23(13): 1510-21.

Mankidy, R., D. V. Faller, et al. (2006). "Short-chain fatty acids induce gamma-globin gene expression by displacement of a HDAC3-NCoR repressor complex." Blood 108(9): 3179-86.

Metzger, D., E. Scheer, et al. (1999). "Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes." Embo J 18(17): 4823-34.

Naar, A. M., D. J. Taatjes, et al. (2002). "Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation." Genes Dev 16(11): 1339-44.

Nuez, B., D. Michalovich, et al. (1995). "Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene." Nature 375(6529): 316-8.

Patrinos, G. P., M. de Krom, et al. (2004). "Multiple interactions between regulatory regions

are required to stabilize an active chromatin hub." Genes Dev 18(12): 1495-509.
Perkins, A. C., A. H. Sharpe, et al. (1995). «Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF.» Nature 375(6529): 318-22.
Persengiev, S. P., X. Zhu, et al. (2003). "TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed." Proc Natl Acad Sci U S A 100(25): 14887-91.
Pilon, A. M., M. O. Arcasoy, et al. (2008). "Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2." Mol Cell Biol 28(24): 7394-401.
Sankaran, V. G., T. F. Menne, et al. (2008). "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A." Science 322(5909): 1839-42.
Sankaran, V. G., J. Xu, et al. (2009). "Developmental and species-divergent globin switching are driven by BCL11A." Nature 460(7259): 1093-7.
Sengupta, T., N. Cohet, et al. (2009). "Distinct modes of gene regulation by a cell-specific transcriptional activator." Proc Natl Acad Sci U S A 106(11): 4213-8.
Skarpidi, E., G. Vassilopoulos, et al. (2000). "Novel in vitro assay for the detection of pharmacologic inducers of fetal hemoglobin." Blood 96(1): 321-6.
Spangrude, G. J., S. Cho, et al. (2006). "Mouse models of hematopoietic engraftment: limitations of transgenic green fluorescent protein strains and a high-performance liquid chromatography approach to analysis of erythroid chimerism." Stem Cells 24(9): 2045-51.
Strouboulis, J., N. Dillon, et al. (1992). "Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice." Genes Dev 6(10): 1857-64.
Swenson, E. S., J. G. Price, et al. (2007). «Limitations of green fluorescent protein as a cell lineage marker.» Stem Cells 25(10): 2593-600.
Taatjes, D. J., A. M. Naar, et al. (2002). "Structure, function, and activator-induced conformations of the CRSP coactivator." Science 295(5557): 1058-62.
Tanabe, O., F. Katsuoka, et al. (2002). "An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer." Embo J 21(13): 3434-42.
Tatarakis, A., T. Margaritis, et al. (2008). "Dominant and redundant functions of TFIID involved in the regulation of hepatic genes." Mol Cell 31(4): 531-43.
Vadolas, J., H. Wardan, et al. (2002). "Development of sensitive fluorescent assays for embryonic and fetal hemoglobin inducers using the human beta -globin locus in erythropoietic cells." Blood 100(12): 4209-16.
van Dijk, T. B., N. Gillemans, et al. (2009). "Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases." Mol Cell Biol.
Wijgerde, M., J. Gribnau, et al. (1996). "The role of EKLF in human beta-globin gene competition." Genes Dev 10(22): 2894-902.
Zafarana, G., R. Rottier, et al. (2000). "Erythroid overexpression of C/EBPgamma in transgenic mice affects gamma-globin expression and fetal liver erythropoiesis." Embo J 19(21): 5856-63.
Zhao, Q., W. Zhou, et al. (2006). "Repression of human gamma-globin gene expression by a short isoform of the NF-E4 protein is associated with loss of NF-E2 and RNA polymerase II recruitment to the promoter." Blood 107(5): 2138-45.
Zhao, Y., G. Lang, et al. (2008). "A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing." Mol Cell 29(1): 92-101.

APPENDIX I

MICROARRAYS

Methods

Microarray analysis was performed by using GeneChip® Mouse Genome 430 2.0 Array (Affymetrix). Quantification and normalization was performed using GeneSpring GX software.

The first set of microarray analysis was performed to identify differentially regulated genes in the GFPγCL expressing GFP (γ-globin) spontaneously, in comparison to the DsRedβCL and wildtype (WT) I11 cells. Fold ratios of the intensity values were calculated as follows: GFPγCL vs DsRedβCL and GFPγCL vs WT. We chose 2-fold as cut-off.

Filter 1: genes similarly affected were selected, either upregulated or downregulated in both comparisons.

Filter 2: genes affected at the same extent were selected by the condition

(|Average fold ratio (FR)| / |SD|) > 2

Data were analyzed through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com), and a list was generated separating the genes on subcellular location.

The second set of microarray analysis was performed to identify differentially regulated genes upon treatment with Azacytidine (AZA) or Hydroxyurea (HU) compared to standard conditions (ST), in order to validate the reporter cell lines. Fold ratios of the intensity values were calculated as follows: DsRedβCL 1 AZA vs ST, DsRedβCL 2 AZA vs ST, WT AZA vs ST and DsRedβCL 1 HU vs ST, DsRedβCL 2 HU vs ST, WT HU vs ST. We chose 2-fold as cut-off.

Filter 1: genes similarly affected were selected, either upregulated or downregulated in both comparisons.

Filter 2: genes affected at the same extent were selected by the condition

(|Average fold ratio (FR)| / |SD|) > 1.5

Data were analyzed through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com), and a drawlist was generated separating the genes on subcellular location. The list was graphically polished with Adobe Illustrator.

Extracellular space

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Afp	alpha-fetoprotein	194.71	192.11
Penk1	proenkephalin	51.71	36.55
Serpinb2	serpin peptidase inhibitor	22.35	18.77
Vegfc	vascular endothelial growth factor C	15.03	13.66
Jag1	jagged 1 (Alagille syndrome)	13.23	12.72
Tgfb2	transforming growth factor, beta 2	11.83	9.33
Thbs1	thrombospondin 1	9.93	10.10
Sulf2	sulfatase 2	6.42	5.25
Lipg	lipase, endothelial	6.23	6.84
Lrrn3	leucine rich repeat neuronal 3	6.19	7.44
Plxdc2	plexin domain containing 2	6.04	12.13
Adamts1	ADAM metallopeptidase	4.89	6.16
Fgf3	fibroblast growth factor 3	4.82	5.07
Jag2	jagged 2	3.89	5.10

DOWNREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Cela1	chymotrypsin-like elastase family	-30.72	-23.88
C1galt1	ore 1 synthase, glycoprotein-N-acetylga-lactosamine 3-beta-galactosyltransferase, 1	-8.95	-5.82

Plasma membrane

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Cap2	CAP, adenylate cyclase-associated	74.74	93.27
Slc6a15	solute carrier family 6, member 15	43.68	42.58
Hrasls3	histocompatibility 60a	36.44	39.39
Slamf1	signaling lymphocytic activation	29.98	36.46
Tspan6	tetraspanin 6	28.21	19.07
Cap1	CAP, adenylate cyclase-associated	21.81	32.91
Pard6g	par-6 partitioning defective 6 homolog	14.68	18.47
Lifr	leukemia inhibitory factor receptor	11.75	16.78
Il6st	interleukin 6 signal transducer	10.80	14.63
Adra2a	adrenergic, alpha-2A-, receptor	9.91	11.99
Gp1bb	glycoprotein Ib (platelet), beta	9.64	11.46
Fyn	oncogene related to SRC, FGR, YES	9.22	5.78
Itga2b	integrin, alpha 2b (platelet CD41)	9.12	4.86
Vangl2	vang-like 2 (van gogh, Drosophila)	8.60	9.32
Fzd2	frizzled homolog 2 (Drosophila)	8.26	7.90
Dok2	docking protein 2, 56kDa	7.28	5.58
Amotl1	angiomotin like 1	6.98	8.42
Perp	PERP, TP53 apoptosis effector	6.66	3.65
Cd9	CD9 molecule	6.66	11.57
Dtnbp1	dystrobrevin binding protein 1	6.44	4.90
Slco3a1	solute carrier organic anion transporter	6.41	7.39
Cd84	CD84 molecule	5.80	7.29
F11r	F11 receptor	5.72	6.04
Gnb1	guanine nucleotide binding protein	5.67	6.30
Crip2	cysteine-rich protein 2	5.52	5.72
Serinc5	serine incorporator 5	5.46	9.79
Ly75	lymphocyte antigen 75	4.96	4.44
Atp10d	ATPase, class V, type 10D	4.91	4.31
Ripk3	receptor-interacting ser-thr kinase 3	4.68	4.77
Gnb4	guanine nucleotide binding protein	4.63	9.31
Anxa4	annexin A4	4.62	4.08
Jub	jub, ajuba homolog (Xenopus laevis)	4.61	3.59
Ptpru	protein tyrosine phosphatase, receptor	4.56	4.99

Plasma membrane (continuation)

UPREGULATED		GFPγCL vs	GFPγCL vs
Symbol	**Entrez Gene Name**	**WT**	**DsRedβCL**
Abcg1	ATP-binding cassette, sub-family G	4.53	7.08
Pcdh7	protocadherin 7	4.26	3.79
Ctnnal1	catenin (cadherin-associated), alpha	4.04	7.74
Abca5	ATP-binding cassette, sub-family A	3.85	4.24
Enpp1	pyrophosphatase/phosphodiesterase 1	3.41	5.23
Trhde	thyrotropin-releasing hormone degrading	3.32	4.67
Csf2rb2	colony stimulating factor 2 receptor	3.14	6.42

DOWNREGULATED		GFPγCL vs	GFPγCL vs
Symbol	**Entrez Gene Name**	**WT**	**DsRedβCL**
Aqp1	aquaporin 1 (Colton blood group)	-67.08	-42.54
Kcnn4	potassium calcium-activated channel	-35.63	-42.44
Slc38a1	solute carrier family 38, member 1	-30.95	-45.79
Cldn13	claudin 13	-14.94	-9.31
Lphn2	latrophilin 2	-14.16	-13.42
P2ry14	purinergic receptor P2Y, G-protein coupled	-13.14	-7.63
Slc1a3	solute carrier family 1	-12.23	-22.44
Hpn	hepsin	-11.05	-17.20
Trfr2	transferrin receptor 2	-10.05	-12.75
Prokr1	prokineticin receptor 1	-7.47	-15.13
Tbc1d8	TBC1 domain family, member 8	-7.31	-3.73
Sgce	sarcoglycan, epsilon	-7.17	-8.58
Sh3d19	SH3 domain containing 19	-5.79	-12.10
Adrb2	adrenergic, beta-2-, receptor, surface	-5.50	-3.52
Aqp9	aquaporin 9	-5.43	-2.96
Tfrc	transferrin receptor (p90, CD71)	-5.15	-3.54
Lancl2	LanC lantibiotic synthetase component	-4.96	-3.60
Fcer1a	Fc fragment of IgE, high affinity I, receptor	-4.19	-5.20
Lgr4	leucine-rich repeat-containing G protein-coupled receptor 4	-3.87	-5.50

Cytoplasm

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Lpgat1	lysophosphatidylglycerol acyltransferase	40.45	50.59
Serpina3g	serine (or cysteine) peptidase inhibitor	34.82	29.19
Tpm4	tropomyosin 4	24.37	45.78
Olfm3	olfactomedin 3	20.54	20.54
Mylip	myosin regulatory light chain interacting	20.48	25.43
Tpd52	tumor protein D52	15.93	12.26
Pfn2	profilin 2	13.53	16.41
Tubb6	tubulin, beta 6	13.52	12.65
Rasgrp2	RAS guanyl releasing protein 2	13.44	18.77
Kif13a	kinesin family member 13A	12.15	10.31
Agpat3	O-acyltransferase 3	11.77	5.64
Gca	grancalcin, EF-hand calcium binding	11.77	7.30
Mapt	microtubule-associated protein tau	11.53	10.92
Rab34	RAB34, member RAS oncogene family	10.61	13.93
Kdelr3	ER protein retention receptor 3	9.62	7.97
Pldn	pallidin homolog (mouse)	9.22	9.05
Rab27b	RAB27B, member RAS oncogene family	8.92	17.12
B4galt4	galactosyltransferase, polypeptide 4	8.51	9.04
Syt7	synaptotagmin VII	8.48	6.51
Daam1	dishevelled associated activator of morphogenesis 1	8.32	7.31
Tuba1a	tubulin, alpha 1a	7.97	8.11
Chst13	carbohydrate sulfotransferase 13	7.88	8.50
Ssh1	slingshot homolog 1 (Drosophila)	7.19	14.52
Rbm45	RNA binding motif protein 45	7.05	10.24
Osbpl3	oxysterol binding protein-like 3	6.42	8.02
Irs1	insulin receptor substrate 1	6.36	6.13
Trim2	tripartite motif-containing 2	6.20	5.09
Ifi47	interferon gamma inducible protein 47	6.17	6.64
Bex2	brain expressed, X-linked 1	6.02	7.21
Actn1	actinin, alpha 1	5.98	6.40
Ifi27l1	interferon, alpha-inducible protein 27	5.98	6.00
Fbxo2	F-box protein 2	5.71	5.55

Cytoplasm (continuation)

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Lss	lanosterol synthase	5.52	2.98
Tpmt	thiopurine S-methyltransferase	5.39	4.21
Sytl4	synaptotagmin-like 4	5.38	6.10
Akap2	A kinase (PRKA) anchor protein 2	5.35	6.64
Kif21a	kinesin family member 21A	5.25	5.59
Prkg1	protein kinase, cGMP-dependent, type I	5.19	5.17
Ndufs4	NADH-coenzyme Q reductase	5.17	3.47
Espn	espin	5.06	4.99
Shank3	SH3 and multiple ankyrin repeat domains	4.81	3.53
Tec	tec protein tyrosine kinase	4.58	5.02
Kif1b	kinesin family member 1B	4.55	3.88
Paip1	poly(A) binding protein interacting protein	4.55	6.98
Tubb5	tubulin, beta	4.21	4.63
Pfkp	phosphofructokinase, platelet	4.12	3.98
Gch1	GTP cyclohydrolase 1	3.94	8.06
Cnn2	calponin 2	3.85	6.62
Pafah2	platelet-activating factor acetylhydrolase	3.70	4.29
Scamp1	secretory carrier membrane protein 1	3.55	6.20
Akt3	v-akt murine thymoma viral oncogene	3.23	6.51
Atp2a3	ATPase, Ca++ transporting, ubiquitous	3.17	5.25

Cytoplasm (continuation)

DOWNREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Dynlt3	dynein, light chain, Tctex-type 3	-111.57	-152.20
Myh10	myosin, heavy chain 10, non-muscle	-86.52	-116.65
Insig1	insulin induced gene 1	-69.43	-60.22
Enah	enabled homolog (Drosophila)	-36.02	-25.65
Serinc3	serine incorporator 3	-35.50	-30.41
Rpgrip1	retinitis pigmentosa GTPase regulator	-33.50	-30.03
Rps9	ribosomal protein S9	-32.58	-31.40
Mfhas1	malignant fibrous histiocytoma amplified	-26.45	-20.94
Igf2bp3	insulin-like growth factor 2 mRNA binding	-20.87	-23.62
Galc	galactosylceramidase	-12.69	-17.94
Minpp1	multiple inositol polyphosphate	-9.61	-5.05
Lman1	lectin, mannose-binding, 1	-9.24	-8.97
Padi3	peptidyl arginine deiminase, type III	-7.82	-3.90
Apobec3	apolipoprotein B mRNA editing enzyme	-7.71	-7.79
Prdx2	peroxiredoxin 2	-7.60	-7.62
Parvb	parvin, beta	-7.31	-6.51
Rpl17	ribosomal protein L17	-6.69	-3.87
Cyp2b10	cytochrome P450, family 2, subfamily B	-5.98	-5.26
Ncf2	neutrophil cytosolic factor 2	-5.68	-6.33
Spna1	spectrin, alpha, erythrocytic 1	-4.95	-7.67
Sgk1	serum/glucocorticoid regulated kinase 1	-4.80	-9.22
Selenbp2	selenium binding protein 1	-4.80	-3.59
Agap1	ArfGAP with GTPase domain, ankyrin	-4.76	-4.86
Papss2	3'-phosphoaden 5'-phosphosulf synthase 2	-4.70	-5.64
Arpp21	cyclic AMP-regulated phosphoprotein	-4.50	-4.18
Pde4b	phosphodiesterase 4B, cAMP-specific	-4.43	-7.79
Scfd1	sec1 family domain containing 1	-4.37	-5.20
Amacr	alpha-methylacyl-CoA racemase	-4.34	-5.32
Mtm1	myotubularin 1	-4.26	-4.44
Steap3	STEAP family member 3	-4.03	-5.40
Abcb6	ATP-binding cassette, sub-family B	-3.76	-5.88
Tbc1d4	TBC1 domain family, member 4	-3.65	-5.73
Bcl2l2	BCL2-like 2	-3.39	-6.41
Trim10	tripartite motif-containing 10	-3.32	-6.14
Spg20	spastic paraplegia 20 (Troyer syndrome)	-2.90	-5.72

Nucleus

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Ccnb1ip1	cyclin B1 interacting protein 1	188.87	167.44
Skiv2l2	superkiller viralicidic activity 2-like 2	25.77	21.18
Cited1	Cbp/p300-interacting transactivator	20.70	11.29
Lhx2	LIM homeobox 2	16.89	16.20
Elk3	ELK3, ETS-domain protein	16.35	9.70
Pftk1	PFTAIRE protein kinase 1	15.51	16.38
Dusp4	dual specificity phosphatase 4	12.59	9.46
Pbx1	pre-B-cell leukemia homeobox 1	11.65	13.38
Nfib	nuclear factor I/B	10.37	8.63
Peg3	paternally expressed 3	10.18	12.34
Lrrfip1	leucine rich repeat (in FLII) interacting	8.80	8.35
Wasf1	WAS protein family, member 1	8.18	9.24
Rhox5	reproductive homeobox 5	7.88	6.31
Dnmt3b	DNA (cytosine-5-)-methyltransferase 3	6.64	5.35
Ipo11	importin 11	6.56	5.44
Litaf	lipopolysaccharide-induced TNF factor	6.24	4.98
Egr1	early growth response 1	6.17	2.99
Jarid2	jumonji, AT rich interactive domain 2	6.11	6.45
2610018G03Rik	RIKEN cDNA 2610018G03 gene	5.69	10.59
Ssbp2	single-stranded DNA binding protein 2	5.50	3.62
Pla2g16	phospholipase A2, group XVI	5.39	5.02
Luzp1	leucine zipper protein 1	5.25	3.19
Supt16h	suppressor of Ty 16 homolog	5.00	5.87
Nab2	NGFI-A binding protein 2	4.55	5.43
Glis2	GLIS family zinc finger 2	4.07	4.28
Psip1	PC4 and SFRS1 interacting protein 1	3.99	4.56
Dip2a	DIP2 disco-interacting protein 2 homolog	2.97	5.89
Ctdspl	CTD small phosphatase-like	2.82	5.69

Nucleus (continuation)

DOWNREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Rgs13	regulator of G-protein signaling 13	-78.93	-115.51
Stag3	stromal antigen 3	-13.90	-26.28
Rgs2	regulator of G-protein signaling 2, 24kDa	-11.34	-20.94
2810417H13Rik	RIKEN cDNA 2810417H13 gene	-10.13	-9.08
Tia1	TIA1 cytotoxic granule-associated RNA binding protein	-10.07	-6.29
Cdkn1c	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	-8.24	-6.47
Trib3	tribbles homolog 3 (Drosophila)	-7.74	-12.55
Ndel1	nudE nuclear distribution gene E homolog (A. nidulans)-like 1	-7.52	-4.57
D10627	cDNA sequence D10627	-6.67	-5.44
Gabpb2	GA binding protein transcription factor, beta subunit 2	-6.51	-6.88
Fyb	FYN binding protein (FYB-120/130)	-6.23	-5.96
Imp4	IMP4, U3 small nucleolar ribonucleoprotein, homolog (yeast)	-6.00	-9.35
Klf6	Kruppel-like factor 6	-5.94	-3.07
Med7	mediator complex subunit 7	-5.62	-4.42
Per2	period homolog 2 (Drosophila)	-5.59	-6.77
Isg20	interferon stimulated exonuclease gene 20kDa	-5.43	-8.45
Fos	v-fos FBJ murine osteosarcoma viral oncogene homolog	-5.30	-10.11
Prmt3	protein arginine methyltransferase 3	-4.95	-3.27
Adi1	acireductone dioxygenase 1	-4.87	-6.77
Atm	ataxia telangiectasia mutated	-4.40	-3.88
Nrip1	nuclear receptor interacting protein 1	-4.40	-4.18
Bpnt1	3'(2'), 5'-bisphosphate nucleotidase 1	-4.00	-6.07
Cenpk	centromere protein K	-3.90	-4.44

Not located

UPREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Rnf144b	ring finger protein 144B	54.01	58.06
2410076I21Rik	RIKEN cDNA 2410076I21 gene	34.98	38.53
Acpl2	acid phosphatase-like 2	28.03	30.59
Gdpd3	glycerophosphodiester phosphodiesterase	26.10	35.13
Gpr177	G protein-coupled receptor 177	22.70	32.05
Armcx1	armadillo repeat containing, X-linked 1	13.65	18.30
Rcbtb2	regulator of chromosome condensation	11.81	7.22
Lrp11	low density lipoprotein receptor-related	11.13	11.64
Esrp2	epithelial splicing regulatory protein 2	9.73	11.32
Rundc3b	RUN domain containing 3B	7.64	8.72
Mtus2	microtubule associated tumor suppressor	7.19	10.16
Sel1l3	sel-1 suppressor of lin-12-like 3	6.81	6.44
Irg1	immunoresponsive 1 homolog (mouse)	6.67	7.43
Tmem87a	transmembrane protein 87A	6.04	7.99
Vsig10	V-set and immunoglobulin domain	5.86	6.15
Aof1	amine oxidase (flavin containing) domain 1	5.85	5.34
Med12l	mediator complex subunit 12-like	5.74	3.36
2010107G23Rik	RIKEN cDNA 2010107G23 gene	5.71	4.80
Dtx4	deltex homolog 4 (Drosophila)	5.39	4.26
Grrp1	glycine/arginine rich protein 1	5.39	3.81
Tmem30a	transmembrane protein 30A	5.15	2.92
Dleu2	deleted in lymphocytic leukemia 2	5.04	7.70
Ccdc122	coiled-coil domain containing 122	4.89	4.51
Prkag2	protein kinase, AMP-activated, gamma 2	4.69	4.20
Ubash3b	ubiquitin associated and SH3 domain	4.45	4.58
Zcchc14	zinc finger, CCHC domain containing 14	4.18	4.76
Inpp4b	inositol polyphosphate-4-phosphatase	4.09	3.95
Phlda3	pleckstrin homology-like domain, family A	4.02	4.61
Fry	furry homolog (Drosophila)	3.72	5.61
Gm672	predicted gene 672	3.71	4.58
Plekhg2	pleckstrin homology domain containing	3.16	4.97
Tnnt1	troponin T type 1 (skeletal, slow)	3.05	5.51
Mfsd2	major facilitator superfamily domain	2.96	5.48

Not located (continuation)

DOWNREGULATED		GFPγCL vs WT	GFPγCL vs DsRedβCL
Symbol	**Entrez Gene Name**		
Smpx	small muscle protein, X-linked	-235.87	-136.85
Zcchc3	zinc finger, CCHC domain containing 3	-20.82	-23.55
Zc3h4	zinc finger CCCH-type containing 4	-17.27	-19.42
Cep192	centrosomal protein 192kDa	-11.48	-7.40
Evi5	ecotropic viral integration site 5	-9.14	-9.96
March6	membrane-associated ring finger (C3HC4)	-8.68	-6.35
Acyp2	acylphosphatase 2, muscle type	-8.34	-4.39
Fhdc1	FH2 domain containing 1	-6.82	-3.41
Tns3	tensin 3	-6.82	-3.31
Map4k3	mitogen-activated protein kinase	-6.82	-8.79
Snx7	sorting nexin 7	-6.77	-5.82
Sdsl	serine dehydratase-like	-6.50	-3.96
Tspan33	tetraspanin 33	-6.50	-3.48
Spata5l1	spermatogenesis associated 5-like 1	-6.29	-7.96
Wwc2	WW and C2 domain containing 2	-6.24	-9.88
Kif26b	kinesin family member 26B	-6.08	-5.32
Nfu1	NFU1 iron-sulfur cluster scaffold homolog	-6.00	-4.77
Aqp11	aquaporin 11	-5.87	-3.81
1700017B05Rik	RIKEN cDNA 1700017B05 gene	-5.79	-5.83
Dph5	DPH5 homolog (S. cerevisiae)	-5.63	-6.97
Prei4	hypothetical protein KIAA1434	-5.31	-3.18
Ehbp1	EH domain binding protein 1	-5.02	-3.62
Asns	asparagine synthetase	-4.64	-3.41
Pcgf1	polycomb group ring finger 1	-4.16	-4.06
Cgnl1	cingulin-like 1	-3.80	-6.34
Cnnm2	cyclin M2	-3.50	-5.12
Camsap1l1	calmodulin regulated spectrin-associated	-2.92	-6.03

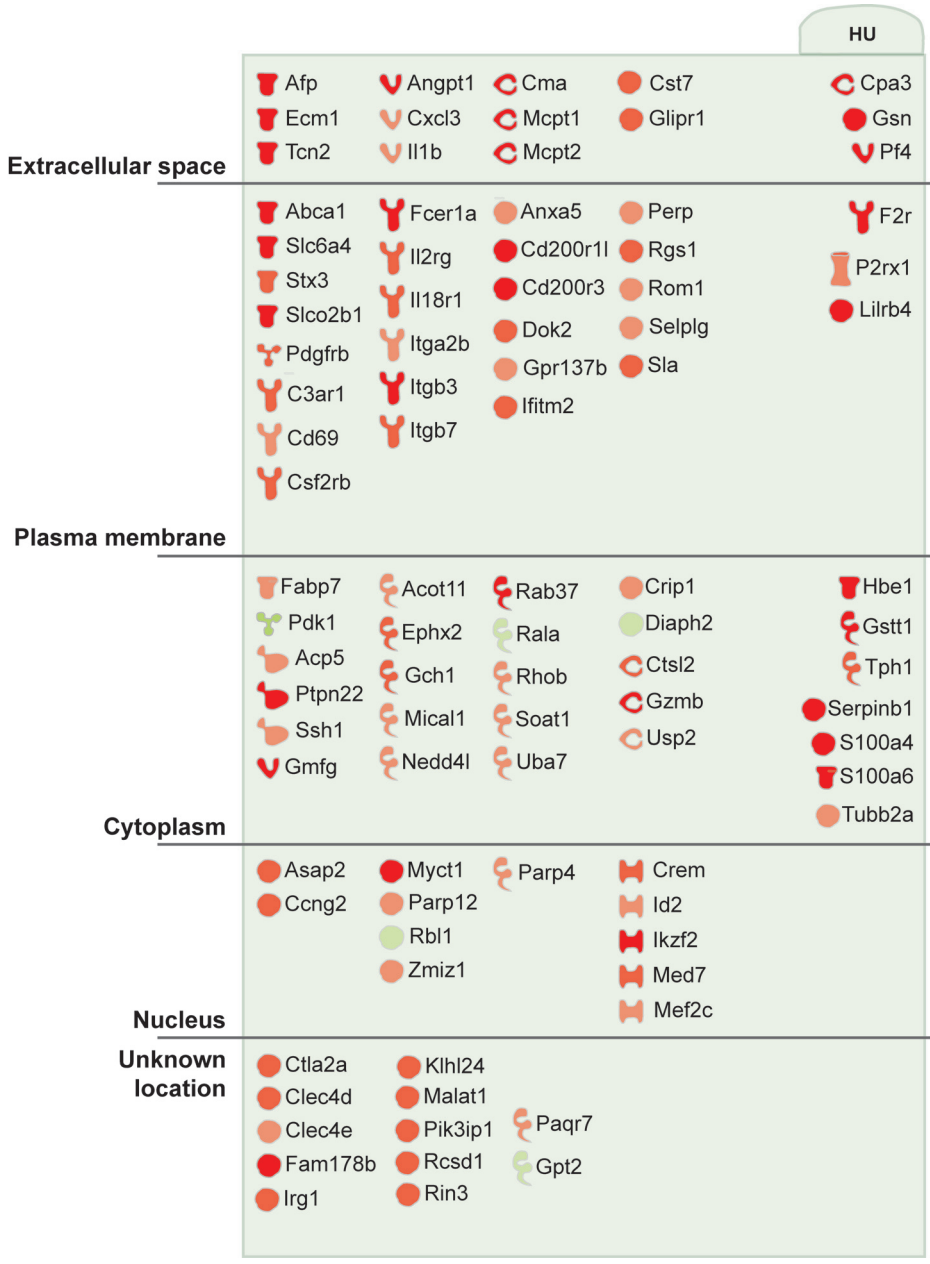
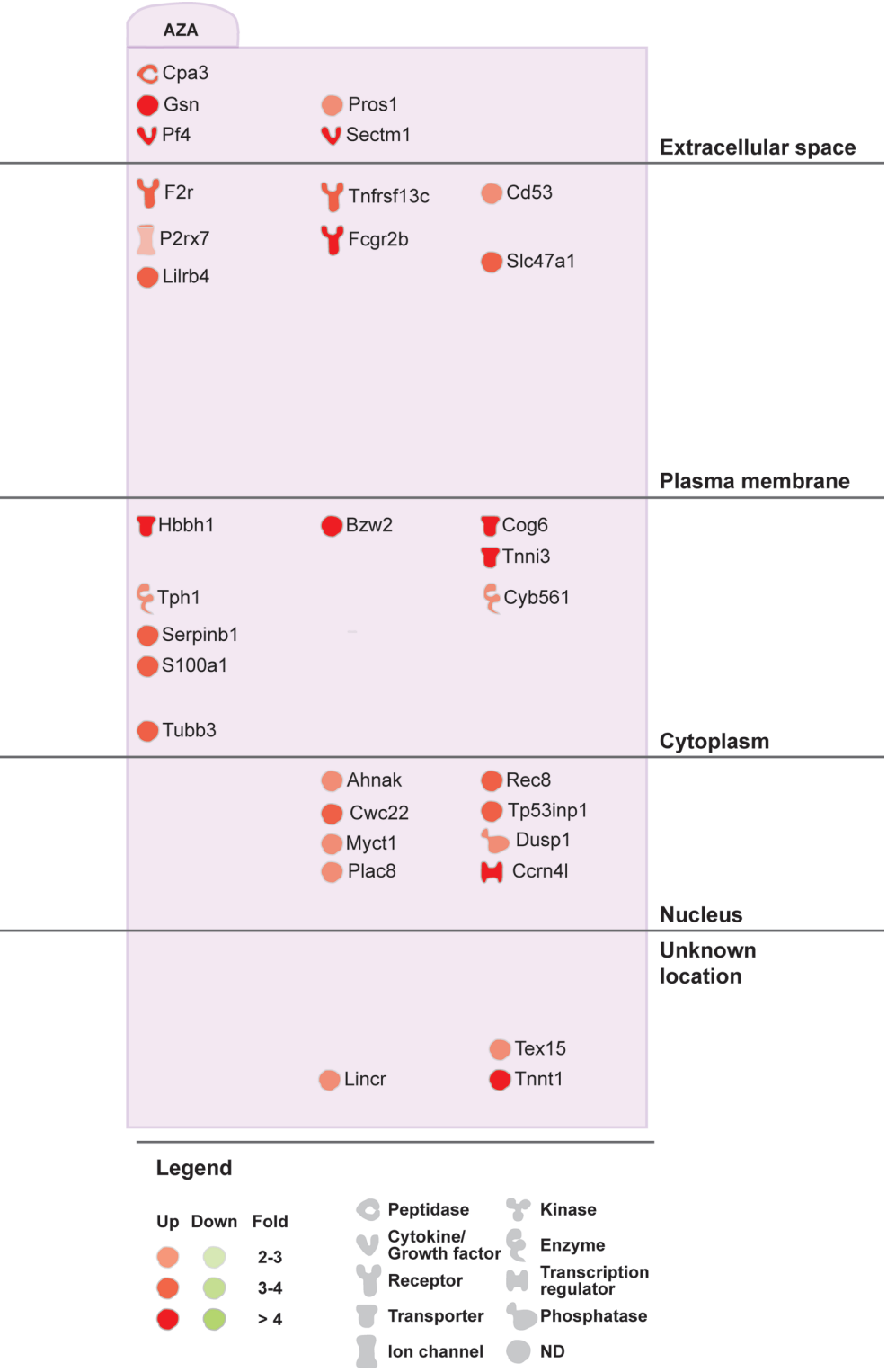
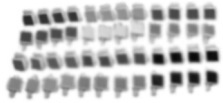


Illustration of differentially expressed genes after Azacytidine (AZA, right) or Hydroxyurea (HU, left) treatment of fetal liver erythroid cell lines. GPA-GFPγ, GFPγ and wildtype fetal liver erythroid cell lines were used. Genes commonly and similarly (up- or downregulated) affected by respective treatments in the three cell lines are depicted. The fold ratio indicated corresponds to the average of the three cell lines upon each respective treatment. Genes that are affected by both treatments are depicted in the center of the spread. Note that most of the affected genes are upregulated.


AZA
Cpa3
Gsn
Pf4
Pros1
Sectm1
Extracellular space
F2r
P2rx7
Lilrb4
Tnfrsf13c
Fcgr2b
Cd53
Slc47a1
Plasma membrane
Hbbh1
Tph1
Serpinb1
S100a1
Tubb3
Bzw2
Cog6
Tnni3
Cyb561
Cytoplasm
Ahnak
Cwc22
Myct1
Plac8
Rec8
Tp53inp1
Dusp1
Ccrn4l
Nucleus
Unknown location
Tex15
Lincr
Tnnt1
Legend
Up Down Fold
2-3
3-4
> 4
Peptidase
Cytokine/ Growth factor
Receptor
Transporter
Ion channel
Kinase
Enzyme
Transcription regulator
Phosphatase
ND

APPENDIX II

MASS SPECTROMETRY



Asf1b IP DsRedβCL 1

Score	Symbol*	Entrez Gene Name
243	ASF1B	Anti-silencing function protein 1 homolog B
377	HNRNPA1L2	heterogeneous nuclear ribonucleoprotein A1-like
358	ACTG2	actin, gamma 2, smooth muscle, enteric
189	EIF4A3	eukaryotic translation initiation factor 4A, isoform 3
155	CALR	calreticulin
148	EIF2S2	eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa
140	FXR2	fragile X mental retardation, autosomal homolog 2
128	DDX4	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4
127	MSN	moesin
126	HIST1H2BL	histone cluster 1, H2bl
126	HIST3H2BB	histone cluster 3, H2bb
126	HIST1H2BD	histone cluster 1, H2bd
126	HIST1H2BK	histone cluster 1, H2bk
122	ACTR1A	ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)
120	RAB1A	RAB1A, member RAS oncogene family
120	RAB10	RAB10, member RAS oncogene family
120	RAB1B	RAB1B, member RAS oncogene family
113	MCM4	minichromosome maintenance complex component 4
102	PUF60	poly-U binding splicing factor 60KDa
99	MED23	mediator complex subunit 23
93	FAM69B	family with sequence similarity 69, member B
93	DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1
93	PDZD2	PDZ domain containing 2
92	C12ORF51	chromosome 12 open reading frame 51
92	UNC13B	unc-13 homolog B (C. elegans)
89	HIST1H3I	histone cluster 1, H3i
86	SMARCA2	SWI/SNF related
85	CENPF	centromere protein F, 350/400ka (mitosin)
83	ESF1	ESF1, nucleolar pre-rRNA processing protein
82	ABCE1	ATP-binding cassette, sub-family E (OABP), member 1
82	IPO4	importin 4
80	SNTB1	syntrophin, beta 1
80	STXBP2	syntaxin binding protein 2
79	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3
78	DNM3	dynamin 3
76	PHLDB1	pleckstrin homology-like domain, family B, member 1
76	UBR3	ubiquitin protein ligase E3 component n-recognin 3 (putative)
75	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
75	NEDD4L	neural precursor cell expressed, developmentally down-regulated
74	SAFB	scaffold attachment factor B
74	BAG2	BCL2-associated athanogene 2
72	LUZP1	leucine zipper protein 1
72	MYOF	myoferlin
70	GCC2	GRIP and coiled-coil domain containing 2
70	PRSS2	protease, serine, 2 (trypsin 2)
68	DENND1A	DENN/MADD domain containing 1A
67	UTP20	UTP20, small subunit (SSU) processome component
66	HYOU1	hypoxia up-regulated 1
66	MLL	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog)
65	MLLT1	MLL (trithorax homolog); translocated
64	MYL6B	myosin, light chain 6B, alkali, smooth muscle and non-muscle
64	MORC2B	microrchidia 2B
64	DDX42	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
63	DNAH9	dynein, axonemal, heavy chain 9
63	ABCA2	ATP-binding cassette, sub-family A (ABC1), member 2
62	FAM35A	family with sequence similarity 35, member A
61	TSHZ1	teashirt zinc finger homeobox 1
60	C15ORF55	chromosome 15 open reading frame 55
60	SEPSECS	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase
60	PCSK5	proprotein convertase subtilisin/kexin type 5
60	DNAH7	dynein, axonemal, heavy chain 7
60	ARPC1B	actin related protein 2/3 complex, subunit 1B, 41kDa
59	MITF	microphthalmia-associated transcription factor

Asf1b IP DsRedβCL 1 (cont.)

Score	Symbol*	Entrez Gene Name
58	C2CD2	C2 calcium-dependent domain containing 2
58	XPO4	exportin 4
58	PEX3	peroxisomal biogenesis factor 3
57	SUPT6H	suppressor of Ty 6 homolog (S. cerevisiae)
57	PRPF4	PRP4 pre-mRNA processing factor 4 homolog (yeast)
57	EXOC2	exocyst complex component 2
56	TRPC7	transient receptor potential cation channel, subfamily C, member 7
56	AKNA	AT-hook transcription factor
56	MEI1	meiosis inhibitor 1
55	CCDC25	coiled-coil domain containing 25
55	NLRX1	NLR family member X1
55	HDGF	hepatoma-derived growth factor (high-mobility group protein 1-like)
55	AI324046	expressed sequence AI324046
54	CTNNA1	catenin (cadherin-associated protein), alpha 1, 102kDa
54	S1PR2	sphingosine-1-phosphate receptor 2
54	RASIP1	Ras interacting protein 1
53	CNTN1	contactin 1
53	DMXL2	Dmx-like 2
52	VPS16	vacuolar protein sorting 16 homolog (S. cerevisiae)
52	AMD1	adenosylmethionine decarboxylase 1
52	REXO1	REX1, RNA exonuclease 1 homolog (S. cerevisiae)
52	ACADVL	acyl-Coenzyme A dehydrogenase, very long chain
52	FRMPD4	FERM and PDZ domain containing 4
51	ME2	malic enzyme 2, NAD(+)-dependent, mitochondrial
51	CLP1	CLP1, cleavage and polyadenylation factor I subunit
51	POLH	polymerase (DNA directed), eta
51	FANCA	Fanconi anemia, complementation group A
51	APOL3	apolipoprotein L, 3
51	KIAA1841	KIAA1841
51	SHROOM2	shroom family member 2
50	PDCL	phosducin-like
50	SYTL2	synaptotagmin-like 2
50	C17ORF49	chromosome 17 open reading frame 49
50	ARID4A	AT rich interactive domain 4A (RBP1-like)
50	STARD9	StAR-related lipid transfer (START) domain containing 9
50	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)
50	PIGN	phosphatidylinositol glycan anchor biosynthesis, class N
50	KCNH7	potassium voltage-gated channel, subfamily H (eag-related)
50	ZNF142	zinc finger protein 142
50	GPATCH2	G patch domain containing 2
50	CAND1	cullin-associated and neddylation-dissociated 1
49	AKT3	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B)
49	INTS4	integrator complex subunit 4
49	DHFR	dihydrofolate reductase
49	SECISBP2	SECIS binding protein 2
49	DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)
49	C3ORF15	chromosome 3 open reading frame 15
49	FYTTD1	forty-two-three domain containing 1
48	PRMT5	protein arginine methyltransferase 5
48	ETV3	ets variant 3
48	FREM1	FRAS1 related extracellular matrix 1
48	PI4KB	phosphatidylinositol 4-kinase, catalytic, beta
47	TGFBR1	transforming growth factor, beta receptor 1
47	IQCE	IQ motif containing E
47	KLC4	kinesin light chain 4
47	CORO1B	coronin, actin binding protein, 1B
47	GBP2	guanylate binding protein 2, interferon-inducible
47	SAMD4B	sterile alpha motif domain containing 4B

Asf1b IP DsRedβCL 1 (cont.)

Score	Symbol *	Entrez Gene Name
46	CBY1	chibby homolog 1 (Drosophila)
46	GEMIN4	gem (nuclear organelle) associated protein 4
46	DAPK1	death-associated protein kinase 1
46	CLSTN1	calsyntenin 1
46	MYOM1	myomesin 1, 185kDa
46	ANKZF1	ankyrin repeat and zinc finger domain containing 1
46	TXK	TXK tyrosine kinase
46	PTPRC	protein tyrosine phosphatase, receptor type, C
45	APBB3	amyloid beta (A4) precursor protein-binding, family B, member 3
45	SVIL	supervillin
45	RIBC2	RIB43A domain with coiled-coils 2
45	LEPR	leptin receptor
45	RTL1	retrotransposon-like 1
45	UBR2	ubiquitin protein ligase E3 component n-recognin 2
45	ADCY5	adenylate cyclase 5
45	KIAA0430	KIAA0430
45	STIL	SCL/TAL1 interrupting locus
44	AOX4	aldehyde oxidase 4
44	ETFB	electron-transfer-flavoprotein, beta polypeptide
44	ZNF346	zinc finger protein 346
44	CLCNKB	chloride channel Kb
44	SLC1A1	solute carrier family 1
44	CDYL	chromodomain protein, Y-like
44	SFI1	Sfi1 homolog, spindle assembly associated (yeast)
44	MYBPC3	myosin binding protein C, cardiac
43	KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
43	SNX19	sorting nexin 19
43	IQCB1	IQ motif containing B1
43	OLA1	Obg-like ATPase 1
43	KTN1	kinectin 1 (kinesin receptor)
43	ABCC9	ATP-binding cassette, sub-family C (CFTR/MRP), member 9
43	NFIX	nuclear factor I/X (CCAAT-binding transcription factor)
43	TMED9	transmembrane emp24 protein transport domain containing 9
43	KIFC1	kinesin family member C1
43	LAMA2	laminin, alpha 2
43	DNM1L	dynamin 1-like
43	LYST	lysosomal trafficking regulator

* Human gene symbol depicted
(*i.e.* Ingenuity Software default)

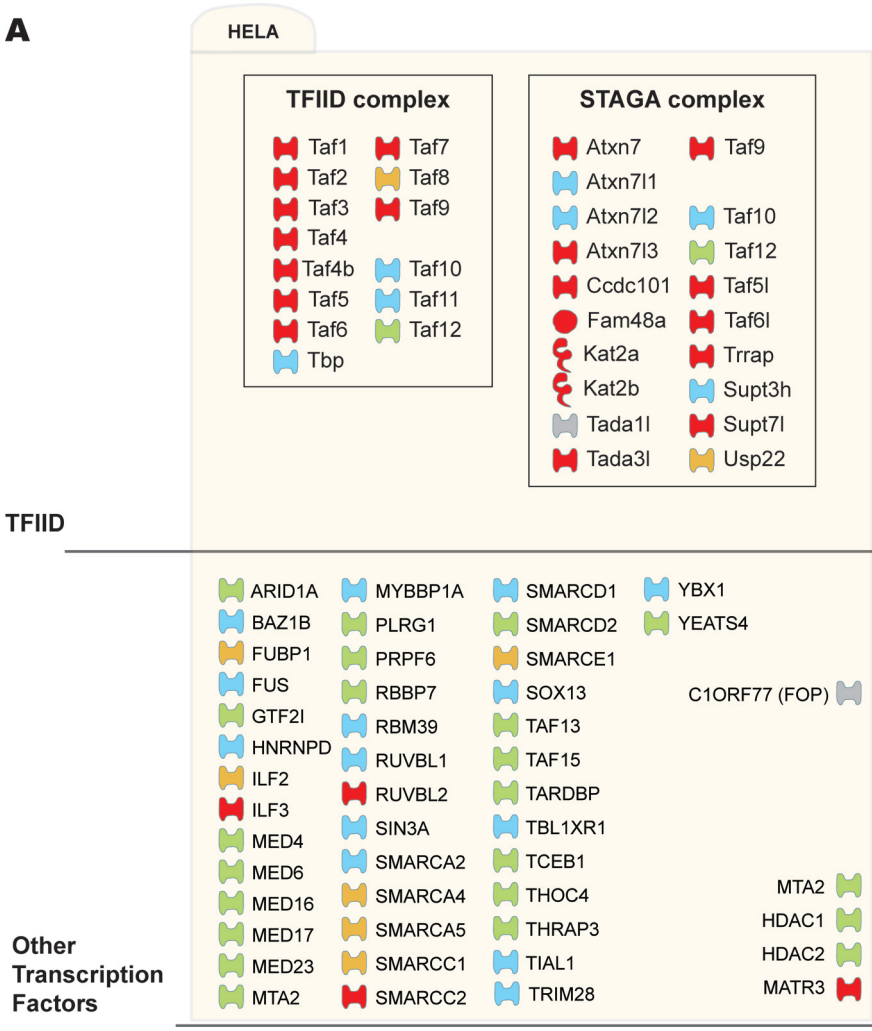
Asf1b IP DsRedβCL 2

Score	Symbol*	Entrez Gene Name
254	ASF1B	Anti-silencing function protein 1 homolog B
837	HBB	hemoglobin, beta
509	RPS17	ribosomal protein S17
215	ANXA7	annexin A7
179	LOC119358	similar to hCG2040270
166	EIF2S3Y	eukaryotic translation initiation factor 2, subunit 3
140	RBM16	RNA binding motif protein 16
121	NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)
110	HIST1H1A	histone cluster 1, H1a
107	RAB15	RAB15, member RAS onocogene family
105	RAB14	RAB14, member RAS oncogene family
104	KIF1B	kinesin family member 1B
93	PSMA6	proteasome (prosome, macropain) subunit, alpha type, 6
93	ALMS1	Alstrom syndrome 1
91	U2AF2	U2 small nuclear RNA auxiliary factor 2
88	GOLGA1	golgi autoantigen, golgin subfamily a, 1
87	ARF2	ADP-ribosylation factor 2
87	ARF5	ADP-ribosylation factor 5
86	SMARCA2	SWI/SNF related
85	VAPA	VAMP -associated protein A, 33kDa
82	EVPL	envoplakin
80	PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform
75	RGS3	regulator of G-protein signaling 3
71	XPO7	exportin 7
70	KIF2B	kinesin family member 2B
69	NAV1	neuron navigator 1
68	CABIN1	calcineurin binding protein 1
68	ZNF608	zinc finger protein 608
67	MYL6B	myosin, light chain 6B, alkali, smooth muscle and non-muscle
64	CDC2	cell division cycle 2, G1 to S and G2 to M
62	KIAA0947	KIAA0947 protein
62	BAT1	HLA-B associated transcript 1
62	FAM98B	family with sequence similarity 98, member B
61	GSS	glutathione synthetase
61	RPL36A	ribosomal protein L36a
61	FABP5L2	fatty acid binding protein 5-like 2
60	MST1R	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
60	SMARCD2	SWI/SNF related
60	RNF214	ring finger protein 214
59	SERPINA1	serpin peptidase inhibitor, clade A
59	DNAH1	dynein, axonemal, heavy chain 1
59	CIRH1A	cirrhosis, autosomal recessive 1A (cirhin)
59	TPTE	transmembrane phosphatase with tensin homology
59	ARHGEF12	Rho guanine nucleotide exchange factor (GEF) 12
57	IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)
57	GLCCI1	glucocorticoid induced transcript 1
56	BCL11B	B-cell CLL/lymphoma 11B (zinc finger protein)
55	NEUROG2	neurogenin 2
55	ATP2C1	ATPase, Ca++ transporting, type 2C, member 1
55	RECQL	RecQ protein-like (DNA helicase Q1-like)
54	NUDT5	nudix (nucleoside diphosphate linked moiety X)-type motif 5
54	DNAJB11	DnaJ (Hsp40) homolog, subfamily B, member 11
54	CENPE	centromere protein E, 312kDa
54	TMED10	transmembrane emp24-like trafficking protein 10 (yeast)
53	HIST1H3D	histone cluster 1, H3d
53	RYR1	ryanodine receptor 1 (skeletal)
53	EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
52	TPD52L2	tumor protein D52-like 2
52	ZFHX3	zinc finger homeobox 3
51	SCN3A	sodium channel, voltage-gated, type III, alpha subunit
51	TFEB	transcription factor EB
51	CKAP5	cytoskeleton associated protein 5

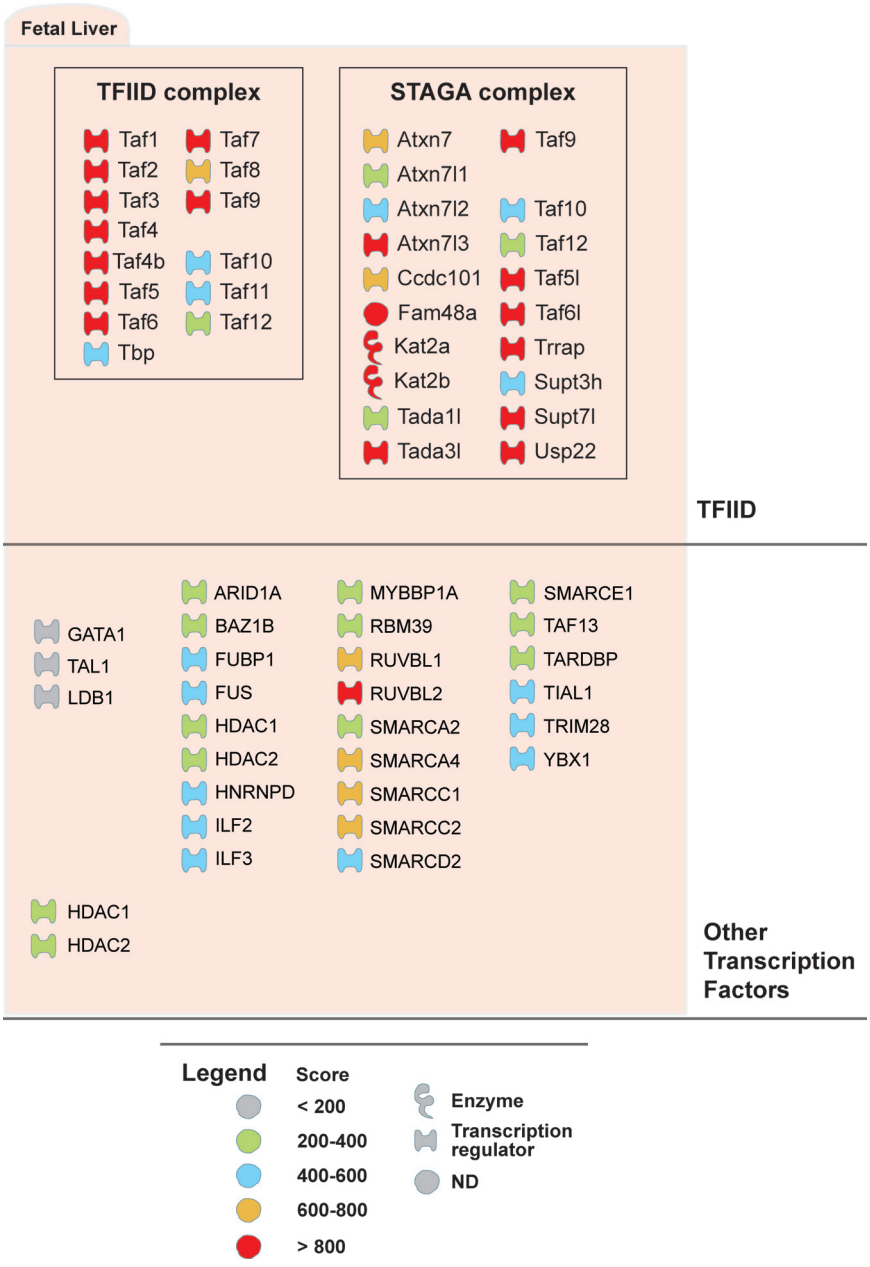
Asf1b IP DsRedβCL 2 (cont.)

Score	Symbol*	Entrez Gene Name
50	SLC22A6	solute carrier family 22 (organic anion transporter), member 6
50	C18ORF25	chromosome 18 open reading frame 25
49	ISG20L2	interferon stimulated exonuclease gene 20kDa-like 2
49	CCDC85A	coiled-coil domain containing 85A
49	FMNL2	formin-like 2
49	AP2S1	adaptor-related protein complex 2, sigma 1 subunit
48	THOP1	thimet oligopeptidase 1
48	DAAM2	dishevelled associated activator of morphogenesis 2
47	MUG1	murinoglobulin 1
47	PLXNA1	plexin A1
47	ZNF366	zinc finger protein 366
47	ZNF710	zinc finger protein 710
46	TSTA3	tissue specific transplantation antigen P35B
46	TAF2	TAF2 RNA polymerase II, TBP-associated factor, 150kDa
46	CCDC88B	coiled-coil domain containing 88B
45	CES2	carboxylesterase 2
45	CES5	carboxylesterase 5
45	FRRS1	ferric-chelate reductase 1
45	EEA1	early endosome antigen 1
45	DNAJC10	DnaJ (Hsp40) homolog, subfamily C, member 10
45	AKAP6	A kinase (PRKA) anchor protein 6
45	PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform
45	ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)
45	NPHP3	nephronophthisis 3 (adolescent)
44	NRG1	neuregulin 1
44	FAM165B	family with sequence similarity 165, member B
44	PCDH12	protocadherin 12
44	GALNT11	N-acetylgalactosaminyltransferase 11 (GalNAc-T11)
44	IQGAP2	IQ motif containing GTPase activating protein 2
44	RNF20	ring finger protein 20
44	ATF7IP2	activating transcription factor 7 interacting protein 2
44	SCLT1	sodium channel and clathrin linker 1
44	ZSCAN20	zinc finger and SCAN domain containing 20
44	RBBP7	retinoblastoma binding protein 7
44	KCNJ15	potassium inwardly-rectifying channel, subfamily J, member 15
43	RABGEF1	RAB guanine nucleotide exchange factor (GEF) 1
43	PARD3B	par-3 partitioning defective 3 homolog B (C. elegans)
43	HERC1	hect (homologous to the E6-AP (UBE3A) carboxyl terminus)
43	PCM1	pericentriolar material 1
42	AOX4	aldehyde oxidase 4
42	AHNAK	AHNAK nucleoprotein
42	CASC5	cancer susceptibility candidate 5
41	GARS	glycyl-tRNA synthetase
41	CYP27B1	cytochrome P450, family 27, subfamily B, polypeptide 1
41	KRT85	keratin 85
41	DGKK	diacylglycerol kinase, kappa
41	SSFA2	sperm specific antigen 2
41	IFI27	interferon, alpha-inducible protein 27
40	ADCY8	adenylate cyclase 8 (brain)
40	YARS	tyrosyl-tRNA synthetase

* Human gene symbol depicted
(*i.e.* Ingenuity Software default)



Analysis of TAF10 IP mass spectrometry data in human fetal liver erythroid progenitors (right) and HELA cells as control (left) in proliferating conditions. The categorization of proteins in complexes (TFTC/STAGA) is based on published data (upper panel of the spread).
The rest of transcriptional regulators bound to TAF10 are summarized. Proteins with a score > 200 were filtered. On the fetal liver IP, transcription factors GATA1, LDB1 and TAL1 were detected bound to TAF10, and despite their lower score they are still depicted.


Fetal Liver
TFIID complex
Taf1
Taf2
Taf3
Taf4
Taf4b
Taf5
Taf6
Tbp
Taf7
Taf8
Taf9
Taf10
Taf11
Taf12
STAGA complex
Atxn7
Atxn7l1
Atxn7l2
Atxn7l3
Ccdc101
Fam48a
Kat2a
Kat2b
Tada1l
Tada3l
Taf9
Taf10
Taf12
Taf5l
Taf6l
Trrap
Supt3h
Supt7l
Usp22
TFIID
GATA1
TAL1
LDB1
ARID1A
BAZ1B
FUBP1
FUS
HDAC1
HDAC2
HNRNPD
ILF2
ILF3
MYBBP1A
RBM39
RUVBL1
RUVBL2
SMARCA2
SMARCA4
SMARCC1
SMARCC2
SMARCD2
SMARCE1
TAF13
TARDBP
TIAL1
TRIM28
YBX1
HDAC1
HDAC2
Other Transcription Factors
Legend
Score
< 200
200-400
400-600
600-800
> 800
Enzyme
Transcription regulator
ND

SUMMARY

SAMENVATTING

Summary

The red blood cell is one of the most important blood cells in the human body. Hemoglobin is the major protein in the red blood cell, and functions as a gas transporter (oxygen and carbon dioxide) to all parts of the body. Changes in the structure or the amount of hemoglobin (hemoglobinopathy) affect the function of red blood cells with consequences for human health. An example are sickle cell anemia and thalassemia.

During the human development, different types of hemoglobin are present in the red blood cells. In the fetus there is mainly fetal hemoglobin (HbF) and in the adult there is mainly adult hemoglobin (HbA). HbF and HbA consist of four polypeptide chains. Both of them have two α chains. The difference lies on the other pair of chains, which are two γ chains in the HbF and two β chains in the HbA. Just after birth HbF is replaced by HbA. The mechanism of this process is not completely understood. There are cases in which persistance of fetal hemoglobin (HPFH) occurs in adults without causing health complications. Unravelling the molecular mecanism by which HPFH occurs could lead to the development of therapies for hemoglobinopathies. Part of our research is to identify the proteins involved in the expression of the different types of hemoglobin.

The approach of generating transgenic mice with the human β-globin locus has been extensively used in developmental studies of the human globin gene expression regulation. Aiming to generate an *in vivo* reporter of the β-globin locus, we modified the γ- and β-globin genes by introducing two fluorescent proteins, the GFP and DsRed respectively, and generated transgenic mice. In particular, we generated an additional transgenic line in which γ-globin promoter activity was measured by the expression of a fusion Glycophorin A-GFP (GPA-GFP) protein in the plasma membrane of the erythroid cells in order to avoid quenching of the GFP in the cytoplasm. Indeed, the developmental expression of the human γ-globin gene (GFP) showed that the expression of the GFP in the plasma membrane of the cells improved the detection of fluorescence in embryonic blood as was demonstrated by FACS analysis.

The reporter β-globin locus transgenic mice were used also to derive fetal liver cell lines and establish an *in vitro* system to perform experiments in a high-throughput manner. The established cell lines were used in several assays, including *in vitro* differentiation, chemical agent treatment, knockdown and overexpression

of different genes validating the potential of these cells to be used as a model for globin expression and biochemical studies. The main advantage of the cell lines is the ease of monitoring expression of the γ-globin (GFP or GPA-GFP) by FACS analysis avoiding gene expression assays (RNA assays) that are laborious and limit the experimental duration when compared to the FACS sorting options that permit further studies even at single cell level. We observed high levels of autofluorescence when chemical agents were tested (azacytidine, hydroxyurea, histone deacetylase inhibitors) both *in vivo* and *in vitro* but we were able to distinguish the real fluorescence using several controls and state of the art flowcytometry equipment even when small populations of cells reactivated the globin genes.

We were interested to identify transcription factors that play a crucial role in the silencing of γ-globin genes and in parallel with knockdown experiments of different genes known to be important in erythropoiesis we initiated a study of the basal transcriptional machinery and the TFIID complex. It is already known that subunits (TAFs) of the TFIID exert tissue and promoter specificity and extensive interactions with a range of transcriptional regulators and coregulators. The involvement of TAF10 in postnatal regulation of genes in the liver was a hint to investigate its role on γ-globin gene regulation. Knockdown of TAF10 in human and mouse erythroid progenitors upregulated moderately the levels of fetal haemoglobin (HbF) and revealed interactions with transcription regulators that should be further investigated for their contribution to the γ-globin promoter activity during development as well as formation of distinct complexes of canonical and non-canonical TFIID during differentiation of erythroid cells.

A transcription factor that is well studied for its importance in β-globin gene expression, KLF1, was identified to be mutated in genetic studies of a family presenting high levels of HbF amongst some of its members. Knockdown of KLF1 in adult erythroid progenitors induced moderate levels of HbF and remains to be seen whether overexpression of the mutant form will increase further the HbF levels due to a dominant-negative effect. This is the first case linking a mutation of KLF1 with high levels of HbF in human.

Our strategy to identify proteins that could directly block the repressor activity of transcription factors and coregulators or potentially a repressor complex in the γ-

promoter and ultimately reactivate it, was investigated by the generation of a heavy chain antibody (HCAb) library from peripheral blood of llamas. Pools of HCAbs were used to transduce the reporter β-globin locus cell lines and sort single cells that express GFP (γ-globin) with the FACSAria. Single-cell PCR of the antibody(ies) and identification of the antibody sequence before or after culturing the single cells sorted would reveal antibody(ies) that bind to potential repressor(s) and/or repressor complexes. It is our aim to identify those with the potential of future therapeutic use in patients with β-thalassemia and sickle cell anaemia.

Samenvatting

De rode bloedcel is een van de belangrijkste bloedcellen in het menselijk lichaam. Hemoglobine is het belangrijkste eiwit in de rode bloedcel, vanwege zijn rol als transporteur van gassen (zuurstof en koolzuur) naar het overige lichaamsweefsel. Veranderingen in de structuur van of het gehalte aan hemoglobine (hemoglobinopathie) tasten het vermogen van de rode bloedcel om zijn taken uit te voeren aan. Voorbeelden van aandoeningen waarin dit het geval is zijn sikkelcelziekte en thalassemie.

Gedurende de menselijke ontwikkeling zijn er verschillende types hemoglobine verantwoordelijk voor bovengenoemde processen. In de foetus is dit HbF (foetaal hemoglobine), bij de volwassene HbA (adult hemoglobine). HbF en HbA bestaan uit 4 polypeptideketens. Beide bevatten 2 α ketens. Het onderscheid wordt gemaakt door de aanwezigheid van 2 β ketens in HbA versus 2 γ ketens in HbF. Na de geboorte wordt het HbF vervangen door HbA (adult hemoglobine). Het mechanisme van dit proces is niet geheel bekend. Echter er zijn gevallen bekend van volwassenen waarin een persisterend hoog gehalte aan HbF wordt gezien zonder dat zij hier last van hebben. Ontrafeling van het moleculaire mechanisme dat hiervoor verantwoordelijk is, zou kunnen leiden tot een behandeling van hemoglobinopathieën. Doel van ons onderzoek is om de eiwitten die verantwoordelijk zijn voor de expressie van verschillende types hemoglobine te identificeren.

Het genereren van transgene muizen met de menselijke β-globine locus wordt veelvuldig toegepast in onderzoek naar de expressie van het menselijke globine gen.

Ten einde een *in vivo* marker (reporter) te maken van een β-globine locus, wijzigden we de γ-globine en β-globine genen door insertie van twee fluorescente eiwitten, respectievelijk GFP (Green Fluorescent Protein) en DsRed (‘Red Fluorescent Protein), waarmee we een nieuw soort transgene muis genereerden. Bovendien maakten we een extra transgene lijn waarin de activiteit van γ-globine gemeten werd door insertie van een fusie-eiwit, glycophorine A-GFP (GPA-GFP), dat terecht komt in het plasmamembraan van de erytroïde cel. Door meting van GPA-GFP aan de oppervlakte van de cel wordt storende inmenging van hemoglobine of detectie van GFP in het cytoplasma voorkomen. De expressie van GPA-GFP in het plasmamembraan van de cellen verbeterde inderdaad de detectie van fluorescentie in embryonaal bloed, zoals wij hebben aangetoond met FACS analyse.

De transgene muizen met de β-globine locus werden gebruikt om foetale lever cellijnen te kweken om zo een *in vitro* systeem op te zetten voor het uitvoeren van experimenten op grotere schaal. Deze cellijnen werden gebruikt in verschillende proeven zoals *in vitro* differentiatie, behandeling met chemische agentia, en onder- dan wel over-expressie van verschillende genen die globine expressie reguleren. Hiermee werd het gebruik van deze cellen gevalideerd als model voor globine-expressie en biochemisch onderzoek. Het belangrijkste voordeel van de cellijnen is het gemak waarmee de γ-globine expressie (GFP of GPA-GFP) gedetecteerd kan worden met FACS analyse. Hierdoor kunnen "klassieke" genexpressie-experimenten vermeden worden. Deze zijn bewerkelijk en beperken de mogelijkheden van een experiment in vergelijking met de FACS sorteer opties. Op deze manier is het mogelijk om aanvullende studies, zelfs op het niveau van één enkele cel, uit te voeren. Zoals reeds bekend zagen we een hoge mate van autofluorescentie wanneer chemische stoffen werden getest (azacytidine, hydroxyureum, en verscheidene remmers van histone deacetylases), zowel *in vivo* als *in vitro*. Echer nu waren wij ook in staat om de werkelijke fluorescentie te onderscheiden door gebruik te maken van meerdere controles en "state-of-the-art" flowcytometrie-apparatuur, zelfs als kleine aantallen van de cellen γ-globine weer geactiveerd hadden.

We waren geïnteresseerd in het identificeren van transcriptie factoren die een cruciale rol spelen bij de regulering van γ-globine genen. In combinatie met inactivering van verschillende genen waarvan bekend is dat ze belangrijk zijn voor de erytropoëse hebben we basale transcriptionele mechanismen, waaronder het TFIID complex onderzocht. Het is bekend dat subeenheden (TAFs) van het TFIID complex een belangrijke weefsel-specifieke rol spelen bij de activiteit van specifieke promotoren en bij interacties met een groot aantal transcriptionele regulatoren en co-regulatoren. De betrokkenheid van één van deze subeenheden, TAF10, in de postnatale regulatie van genen in de lever was aanleiding om de rol van TAF10 bij γ-globine genregulatie te onderzoeken. Inactivering van TAF10 in voorlopercellen van rode bloedcellen wekte in enige mate expressie van HbF op. Tevens werd aangetoond dat interacties met andere transcriptie regulatoren nader onderzocht moeten worden op hun bijdrage aan de γ-globine promotor activiteit tijdens de ontwikkeling, en dat er waarschijnlijk verschillende vormen van TFIID zijn tijdens de differentiatie van rode bloedcellen.

De transcriptiefactor KLF1 speelt een belangrijke rol in β-globine expressie. In genetische studies van een familie met aangedane familieleden met een hoog HbF-

gehalte is KLF1 in gemuteerde vorm aangetroffen. Verminderde expressie van KLF1 zorgde voor verhoogde niveaus van HbF. Het is nog niet duidelijk of overexpressie van de gemuteerde vorm zal leiden tot een verdere toename van HbF. Dit is het eerste geval waarin een verband is gevonden tussen een mutatie in KLF1 en hoge niveaus van HbF in de mens.

Ten slotte hebben wij ons gericht op het samenstellen van een bibliotheek van een enkeldomein antilichaam (HCAb) afkomstig van lama's. Het doel was om eiwitten te blokkeren die de expressie van γ-globine onderdrukken, zodat dit gen opnieuw actief zou worden. Verzamelingen van HCAbs werden tot expressie gebracht in de reporter-cellijnen met als doel om het GFP (γ-globine) gen te activeren. Met FACS analyse zijn individuele cellen gesorteerd die GFP (γ-globine) tot expressie brachten. In verder onderzoek zal getracht moeten worden uit deze cellen met behulp van PCR (polymerase kettingreactie) de verantwoordelijke antilichamen te isoleren en verder te testen. Dit opent nieuwe mogelijkheden voor toekomstig therapeutisch gebruik bij patiënten met β-thalassemie en sikkelcelziekte, omdat γ-globine de afwezige of defecte β-globine ketens kan vervangen.

ACKNOWLEDGEMENTS

Acknowledgements

It is so amazing looking back at the time when I came to Rotterdam and through these years of the PhD because everything is so vivid. It was a wise choice and I'm grateful to many people that helped me in many different ways to adapt, grow up and become a better person.

Frank, there are so many things I can say about you. I remember the first time we met and you asked me if I would be able to cope with the cloudy sky of Holland. Only after a few years I realized you were not joking… You are an example of a hard worker and a fresh mind. I appreciate your support in and out of the lab. I consider myself lucky to have met you. Thank you from my heart.

Sjaak, you've been such a great help for me all these years. You were always supportive and available when I needed you. Reading my thesis and correcting it step by step with patience and constructive comments from your side has been an experience for me.

Dies and Niels, my small committee and also my 'advisory board' thank you for reading my thesis, your understanding and your supportive comments.

Lab 702 was my first lab, work and great fun, always something big was going on. It was great to have around Robert-Jan and Bas, Erik (my bench neighbour), Miyata, Arnie, Wouter, Ernie, Farzin, George, Natalia and Mariette. Miyata, I have to thank you for these special moments we shared in the beginning of my life in Rotterdam. You also taught me how to do nice Southern blots, but mostly how to be patient and persistent at work. The few Japanese words we practised in the lab will be always remembered (Domo!!). Farzin, we shared a lot of happy and exciting moments but also frustrating times due to the difficulties of our 'common' project. We managed to understand each other and face the problems with humour. I'm glad it was with you that I shared my desk, worries and thoughts. George, you've been helpful since the beginning even though you were also in a transition period yourself. It is very important that our communication was kept and resulted in an interesting project that is about to finish.

Dubi, Rick, Alex, Ali (Imam), Sylvia, John Kongasan, Ton, it has been a wonderful time being with you in lab 710. Everybody was very helpful and each one so special. Ali, I enjoyed our discussions about history and I appreciate your help throughout all these years. Rick, you showed me many tricks on how to be fast and efficient in cloning and work with PAC vectors. You were always interested on resolving difficult

tasks, especially cloning ‘mysteries’. Alex, your presence in the lab was catalytic to create a cheerful atmosphere. Thank you for injecting again and again (even at Christmas period) the PAC constructs that I gave you. John, thank you for your precise work and your efficiency to generate transgenic animals. Without you my work would have been impossible. Dubi, it has been a pleasure to be a friend and a colleague with you. You were always helpful with experiments and encouraging when needed. I am very happy you are my paranimf.

Reinier, its impossible to forget your enormous help with the FACS analysis. Thank you so much for your patience and your input to improve my reporter assays. I am sure we did the most out of it.

Yianni, I don’t forget the late-evening hours in the lab and your help anytime I asked for that. We had a very good time here in Rotterdam. Wouter, I do appreciate the discussions we had about work and beyond that, you were always available to listen to my struggles and provide me with ways to overcome them.

I would also like to thank Jasperina and Bep for their invaluable help all these years and especially Marike that has been an indispensable person, arranging every single detail that had to be done efficiently and on time. Special thanks go to all the members of the recent lab 702, Athina, Daan, Sanja, Charlotte, Eric, Dorota, Boet, Ralph, Marianna and many more from other labs that would make this one a very long list. General thanks to all the mouse caretakers. For the people that are not mentioned here and they are many, thank you all for the good times we had together in and out of the lab.

My two paranimfen, Jurjen and Rob, helped me a lot to complete many of my thesis tasks (especially the Dutch Summary) but most importantly I thank them for being my friends.

There is a person that has contributed as nobody else for the completion of this thesis. Without this person this thesis would not have been as it appears right now, neither in its contents nor in its design. Actually, I can not imagine how it would be, just because she enjoyed it as much as I did. Thank you for being yourself, Laura.

CURRICULUM VITAE

PERSONAL DATA

Name:	Petros Papadopoulos
Date of Birth:	2nd November 1976
Place of Birth:	Athens, Greece

EDUCATION

1991-1994	High School of Ano Liosia, Athens, Greece.
1995-2000	Master degree at the School of Biology, Aristotle University of Thessaloniki, Greece.
2000-2001	Training program at the Thalassemic Center of 'Laiko' Hospital, Department of Cytogenetics, Athens.
2001-2002	Interrupted PhD at the Department of Immunology of 'Alexander Fleming Institute' BSRC, Athens, Greece.
2002-2009	PhD student at the Department of Cell Biology ErasmusMC, Rotterdam. Project: 'Studies on human γ-globin gene regulation', under the supervision of Prof.dr. F.G. Grosveld.

PUBLICATIONS

Functional analysis of an arthritogenic synovial fibroblast.

Aidinis V, Plows D, Haralambous S, Armaka M, Papadopoulos P, Kanaki MZ, Koczan D, Thiesen HJ, Kollias G. **Arthritis Res Ther. 2003;5(3):R140-57**.

Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry.

Grosveld F, Rodriguez P, Meier N, Krpic S, Pourfarzad F, Papadopoulos P, Kolodziej K, Patrinos GP, Hostert A, Strouboulis J. **Ann N Y Acad Sci. 2005;1054:55-67**