PhD

Program in Molecular and Translational Medicine
DIMET

(XXVII Cycle, year 2013-2014)

University of Milano-Bicocca
Department of Biotechnology & Biosciences

"Identification and characterization of protein complexes including transcription factor Sox6 in erythroid cells"

Isaura Font Monclús Matr: 761049

PhD Coordinator: Prof. Andrea Biondi
Principal Investigator: Prof. Antonella Ronchi

The research presented in this thesis was performed at the department of Biotechnology and Biosciences, University of Milano-Bicocca in the laboratory of genetics headed by Prof. Antonella Ronchi and co-tutored by Prof. John Strouboulis

“Si vols anar ràpid, camina sol; però si vols arribar lluny, camina acompanyat” Proverbi africà.

"Se vuoi arrivare primo corri da solo. Se vuoi arrivare lontano, camminiamo insieme" Proverbio africano.

TABLE OF CONTENTS

Chapter 1

General Introduction

HEMATOPOIESIS:

The term Hematopoiesis describes the formation of all blood cellular components. Blood is a unique tissue capable to produce a great variety of cells. Each blood cell type has its own specific identity, essential to fulfill its biological function. All blood cells are divided in two main lineages, myeloid and lymphoid. Within the myeloid lineage, there are several cell types derived from the Common Myeloid Progenitor (CMP), such us erythrocytes, megakaryocytes, platelets, granulocytes, macrophages, mast cells and monocytes. Among them, erythrocytes are the most abundant: in the adult human body approximately 2,4 million red blood cells (erythrocytes) are produced per second and circulate with a lifetime of 100-120 days, representing about 42% of the blood total volume. Erythrocytes are small and enucleated, with a biconcave discoid shape and are completely filled with hemoglobin, the molecule that carries oxygen and carbon dioxide throughout all the body. The other cell types are less abundant, about 1% of the total blood volume. Platelets are rod-shaped cell fragments derived from the differentiation of large cells denominated megakaryocytes, and have an essential role in blood coagulation. Neutrophils, eosinophils and basophils are derived from granulocytes. They are playing a role in the innate immune system, by participating in the antigen recognition and destruction and by promoting inflammatory response and phagocytosis. Macrophages and mast cells are not present in the bloodstream but they are infiltrating all other tissues.

Macrophages arise from monocytes, which are circulating in the blood before they migrate to tissues, where they terminally differentiate. They digest cellular debris, foreign substances, microbes and cancer cells by phagocytosis. They play a critical role in innate immunity, and also help adaptive immunity by recruiting other immune cells, such as lymphocytes. Mast cells originate from an unidentified progenitor in the Bone marrow (BM); according to some speculations, mast cells would be basophils that have been established within other tissues, because of its similarity in both appearance and function to basophils. Mast cells are mostly present in connective and mucosa tissue, and are commonly found surrounding blood vessels and nerves. The lymphoid linage is formed by T, B-lymphocytes and natural killer cells, derived from Common Lymphoid Progenitor (CLP). Those cells are the crucial components of the adaptive immune system. T-lymphocytes recognize the presence of antigens coming from any infection and develop sophisticated defense mechanisms, acting as a direct killer or activating other cells, such us natural killers, to eliminate infected cells. These lymphocytes have also the ability to create immunological memory to produce a faster immunological defense for future infections. By the other site, the main function of B-lymphocytes is the production of antibodies that help antigens to be eliminated (Figure 1.1).

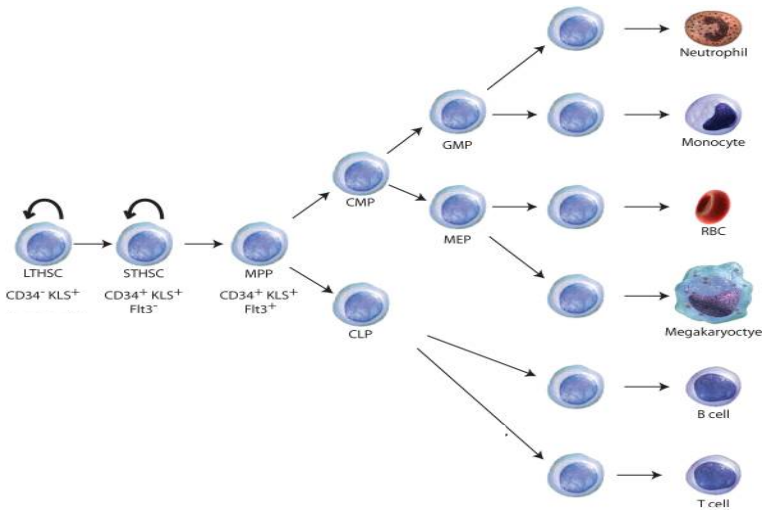


Figure 1.1 *Overview of the hematopoietic hierarchy. LTHSC, Long-term repopulating hematopoietic stem cell; STHSC, short-term repopulating hematopoietic stem cell; MPP, multipotent progenitor cell; CMP, common myeloid progenitor cell; MEP, megakaryocyte-erythroid progenitor cell; GMP, granulocyte monocyte progenitor cell; CLP, common lymphoid progenitor cell; RBC, red blood cell; (Mol Endocrinol. 2010 Jan; 24(1): 1–10)*

All blood different specialized cells are originated by a common single cell, the hematopoietic stem cell (HSC). HSCs are the only bone marrow cells capable to differentiate to all blood cell lineages. In the adult mouse, the frequency of HSCs is approximately 1 in 10000 bone marrow cells (0.01%). Self-renewal and differentiation of HSCs are tightly regulated by both cell intrinsic and extrinsic factors to maintain a functional homeostasis.

The ontogeny of HSCs is still a controversial. During development, hematopoiesis takes place in different organs at different stages. The first intraembryonic tissue identified to contain HSCs was

AGM (Aorta-Gonad-Mesonephros), where the first emergence of small numbers of HSCs begins at embryonic day 10.5 (E10.5) and 5 weeks, in murine and human embryos, respectively. However, blood cells are required for embryonic survival prior to the emergence of these adult-repopulating HSCs. The first stage of hematopoiesis, called primitive hematopoiesis, takes place in the blood islands of the extra-embryonic yolk sac starting at approximately at mouse E7.0 and E7.5, or day 15-18 in humans. Primitive hematopoiesis produces a first transient wave of a unique erythroid progenitor (EryP-CFC) (Figure 1.3). Primitive erythroblasts are large-sized, nucleated and contain embryonic hemoglobin. They enter into the vascular system of the embryo where they continue to divide for several days and undergo progressive morphological maturation to finally enucleate into the bloodstream. The yolk sac generates a second transient wave of erythroid progenitors, which will became definitive erythrocytes. These progenitors enter into the bloodstream and seed the fetal liver starting from mouse E12.5. In humans, the fetal liver becomes the site of definitive hematopoiesis around day 42 ***(McGrath, Bushnell, & Palis, 2008)***. At the same time, hematopoietic stem cells emerging from the AGM region within the embryo, seed the liver and are the presumed source of the long-term erythroid potential cell population ***(Mikkola & Orkin, 2006)***.

The fetal liver is the site of definitive hematopoiesis, which is marked by the appearance of functional HSCs capable to produce

enucleated red blood cells expressing adult globins, the full complement of myeloid cells, and, at later steps, lymphoid cells. Finally, around birth, HSCs from fetal liver colonize the bone marrow (starting at approximately mouse embryo E16.5, and around the 22th week of human gestation), which becomes the final principal site of definitive hematopoiesis ***(McGrath et al., 2008; Palis & Segel, 1998)***, except for T cells, which are formed in the thymus from progenitors derived from HSCs. It has also been reported that the placenta is another major hematopoietic tissue during embryonic development ***(Alvarez-Silva, Belo-Diabangouaya, Salaun, & Dieterlen-Lievre, 2003; Gekas, K, & H, 2008)***. In mice under stress conditions, hematopoiesis also occurs in the spleen.

It is hypothesized that there must be specific microenvironments existing within the bone marrow area that contain HSCs in order to sustain specific aspects of hematopoiesis, such as HSC survival, self-renewal, and differentiation. These special hematopoietic microenvironments are called "niche," a term proposed by Schofield in 1978 to describe "areas producing a complex and dynamic molecular crosstalk between HSC and their endogenous microenvironment ("niche") what directs its fate." The erythroid niche is the structural unit where terminal erythroid differentiation takes place; it consists of a central macrophage (also termed "nurse cell") surrounded by differentiating erythroid progenitors. However, an analogous structure has not been identified in the yolk sac, probably, because primitive

erythroblast enters in the bloodstream when they are still nucleated.

HSCs can be enriched from whole bone marrow by using flow cytometry (FACS) on the basis of specific cell-surface markers. The lineage-negative, or lin^- fraction contains all HSCs and hematopoietic progenitors. In mice, the population of lin^- cells that highly express the cell surface proteins c-kit and Sca-1 (Lin^-, $c\text{-}kit^{HI}$, $Sca\text{-}1^{HI}$ or LSK) contains all HSCs ***(Goodell, Brose, Paradis, Conner, & Mulligan, 1996; Okada et al., 1992; Osawa, Hanada, Hamada, & Nakauchi, 1996)***. Additional cell surface proteins such as Thy1.1, IL-7Rα, Flt3, CD150, and endoglin or the distinct dye-efflux profiles of Hoescht or Rhodamine can also be used to further enrich for HSCs ***(Christensen & Weissman, 2001; Park et al., 2002)***. Moreover, the cell surface antigen CD34 can discriminate between long-term HSCs, which are $CD34^-$, and short-term HSCs, which are $CD34^+$(Figure 1.2) ***(Foudi et al., 2009; Kiel et al., 2005; Matsuzaki, Kinjo, Mulligan, & Okano, 2004; A. Wilson et al., 2008)***. Regardless of the FACS methods used to isolate HSCs, functional assays (transplantation into irradiated mice) are required to truly determine the ability of the cell to self-renew and repopulate forming mature cell lineages.

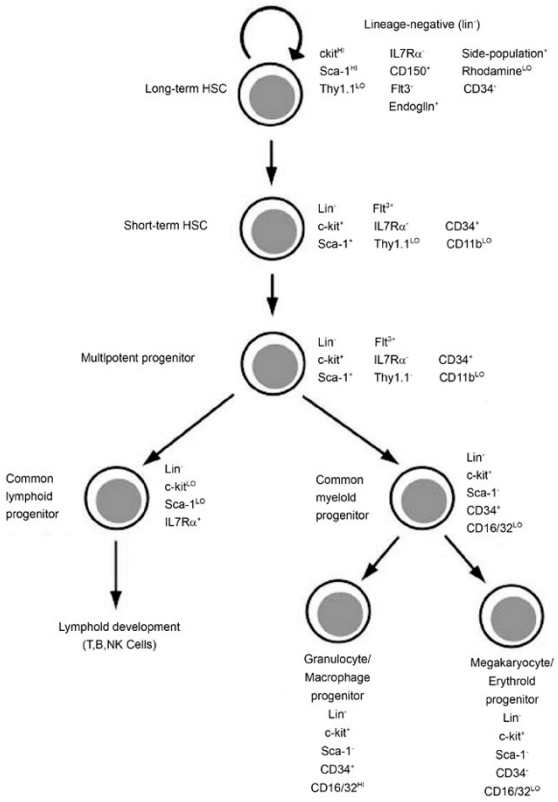


Figure 1.2 *Model of cell surface markers for hematopoiesis used to discriminate HSCs and progenitors in adult mice.*

During hematopoietic differentiation HSCs lose their pluripotency in a stepwise fashion, ultimately giving rise to unipotent progenitor cells. This progressive lineage restriction is also referred to as a 'cell fate decision' or 'lineage commitment', which progressively generates multilineage precursors (MPL). Then, the progenitors differentiation process occurs through a series of commitment steps, each leading to further restriction to a certain hematopoietic lineage. The differentiation from HSCs to progenitors and then to differentiated cells has been largely studied. However, it is still a controversial issue, and several models are proposed. According to the most widely accepted

model, called “classical” model, HSCs firstly generate a common myeloid progenitor (CMP) and a common lymphoid progenitor (CLP). CMPs will produce myeloid or erythroid cells (granulocytes, megakaryocytes, macropahges and erythrocytes). CLP will give rise T-, B-cells and natural killer cells. CMPs undergo a further lineage restriction when the granulocytes/monocytes potential split from the erythroid/megakaryocyte potential ***(Akashi, Reya, Dalma-Weiszhausz, & Weissman, 2000; Graf & Enver, 2009; Pronk et al., 2007)***. An alternative “myeloid-based” model postulates that the HSCs first diverge into the common myeloid-erythroid progenitor (CMEP) and a common myelo–lymphoid progenitor (CMLP). The CMLP will then generate T and B cell progenitors through a bipotential myeloid–T progenitor (MTP) and a myeloid–B progenitor (MBP) stage, respectively. According to this model, the myeloid potential is separated form the megakaryocytic/erythroid and B- and T-cell potential ***(Kawamoto & Katsura, 2009)***. During the last years, the rigid hierarchy between the HSCs and differentiated cells has been largely questioned. Several publications suggest the occurrence of transdifferentiation between different hematopoietic precursors ***(Graf, 2002)***, suggesting a previously unexpected plasticity of the system.

The concept that hematopoiesis occurs in orderly, predetermined path it is itself a controversial issue. In fact, some studies support a deterministic theory where the stem cell compartment contains a series of closely related cells maturing in a stepwise process,

each with distinct properties ***(Morrison, Uchida, & Weissman, 1995)***. However, other studies suggest that hematopoiesis is a random, stochastic process. This notion derives from the observation that, in *ex vivo* culture, multi-lineages colonies develop into variable combinations of lineages, suggesting that pluripotent stem cells differentiate and lose potential randomly in tissue culture ***(Graf, 2002)***. A similar controversy exists regarding how cytokines influence the development of specific cell lineages. The instructive model suggests that cytokines force primitive uncommitted cells to commit to particular hematopoietic lineages. This model reinforces the importance of the stroma and the niche in the cell differentiation. The alternative permissive model postulates that cell-fate decisions are independent from extracellular signals, and that cytokines serve only to allow certain lineages to survive and proliferate ***(D'Andrea, 1994)***.

ERYTHROPOIESIS*:*

Erythropoiesis is a dynamic complex multistep process that involves differentiation from early erythroid progenitors to enucleated red blood cells (RBCs).

Erythrocytes represent the most common cell type in adult blood. Approximately $5x10^6$ erythrocytes per microliter are present in the human blood. The average of their lifespan is of 60 days in mouse and 120 days in human ***(Franco, 2012)***. Mature enucleated erythrocytes have a biconcave-disked shape and contain large

amount of hemoglobin required to bind and delivery O_2, to all the body tissues.

Common megakaryocyte-erythroid progenitors (MEP) are early multipotent progenitors able to generate committed erythroid progenitors (erythroblasts), which will give rise to mature erythrocytes (RBCs). ***(Novershtern et al., 2011; J. G. Wilson & Tavassoli, 1994).*** These immature progenitors are capable to generate multilineage colonies in response to specific factors. In cells cultures, upon Erythropoietin treatment they differentiate into responsive erythroid burst forming unit-erythroid (BFU-E) and colony forming unit-erythroid (CFU-E) ***(Allen & Dexter, 1984; Allen, Dexter, & Simmons, 1990) (Pronk et al., 2007)***. CFU-Es differentiate then into orthochromatic normoblasts, and finally, into enucleated reticulocytes and then to RBCs (Figure 1.3). BFU-Es are large cells forming colonies containing up to several thousand of hemoglobinized cells, which appear after 5-8 days (mouse) or 10-14 days (human) in semi-solid methylcellulose cultures. Their growth depends on several factors, such us stem cell factor (SCF), thrombopoietin (TPO), interleukin 3 (IL3), IL11, and FLT3-ligand. The earliest erythroid progenitors are sensitive to several cytokines (SCL, TPO, GM-SCF, IL3, IL11). In particular, SCL binds to the Kit receptor activating several signaling pathways and acting synergistically with EPO to sustain the proliferation and expansion of erythroid progenitors ***(Bunn, 2013).***

The human CFU-Es are more mature erythroid progenitors that form small colonies of 16-125 cells that appear after 2-3 days (mouse) or 5-8 days (human) in semi-solid methylcellulose cultures ***(Whyatt et al., 2000; Wong, Chung, Reicheld, & Chui, 1986)***. They are five to ten times more abundant in bone marrow than BFU-E, but they are not present in circulation. Finally, during the subsequent maturation process the cells gradually lose their potential to proliferate and became mature enucleated cells (Figure 3.1).

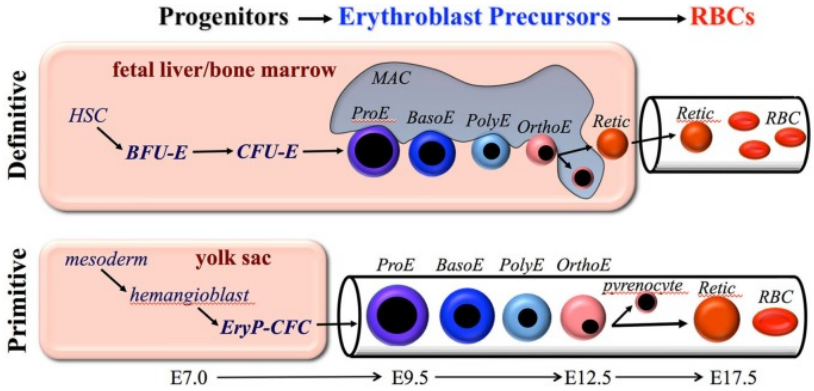


Figure 1.3 *Primitive and definitive erythropoiesis in mammals. (Palis, 2014).*

The maturation from pro-erythroblast to reticulocytes involves four to five rapid cell divisions (occurring in 48- 72 hours) progressively reducing the cell size. Pro-erythroblasts are large cells with a large nucleus and a basophilic cytoplasm. Basophilic erythroblasts are slightly smaller (12-17μm and 9-10μm in human

and mice respectively). At this stage occurs a strong hemoglobinization. The smallest nucleated red cells precursors are the orthochromatic erythroblasts (8-12μm and 7-8μm in human and mice, respectively). Their nucleus undergoes pycnotic degeneration: chromatin becomes very condensed and, as a consequence, the nucleus becomes smaller. Finally, nucleus is extruded from the cell, which then becomes a reticulocyte and enter into the bloodstream. Reticulocyte is a bit larger than mature RBCs (7-8μm and 3,5-4,5μm in human and mice respectively). These cells have the characteristically biconcave shape, and still retain few cytoplasmatic organelles, necessary to complete the hemoglobin protein synthesis ***(Allen & Dexter, 1982; Bondurant & Koury, 1986)***.

The erythroid maturation process takes place within the erythroblastic islands located in the fetal liver during the embryonic development and then, in the bone marrow, during adult life. This compartment offers a proper niche to sustain erythroid maturation. It is formed by a central macrophage surrounded by erythroid precursors at different stages of maturation. The most immature precursors are located in the center, and, as they mature, they separate form the central body of the macrophage. Erythroid precursors keep in contact with the cytoplasmatic extensions of the central macrophage during all stages of their maturation. Several proteins on the surface of the macrophage and of erythroblasts mediate the interactions between these cell types. Moreover macrophages are the

responsible to phagocyte the nuclei extruded from reticulocytes ***(Adams & Scadden, 2006; Allen & Testa, 1991)***.

HEMOGLOBIN SWITCHING:

The hemoglobin protein:

Hemoglobin is the metalloprotein responsible for the oxygen delivery carried out by red blood cells (RBCs). Each red blood cell (RBC) contains approximately 280 millions of hemoglobin molecules (***Sears, 1999***). The main function of hemoglobin is to transport oxygen from the lungs to the tissues and then to transport CO_2 back from the tissues to the lungs. One hemoglobin molecule has the ability to transport up to 4 oxygen molecules. Hemoglobin protein is a quaternary structure formed by four globular subunits (Figure 1.4a). Each subunit is composed by a protein chain (2 protein chains is tightly associated with a non-protein heme group (Figure 1.4b). The α-globin chain is composed of 141 amino acids and the β-globin chain is composed of 146 amino acids ***(Perutz, 1978)***. Both α and βglobin proteins share similar secondary and tertiary structures (***Keates, 2004***). Each protein arranges the globin fold structure using a set of α-helix segments connected together; this entire binding pattern strongly binds the heme group.

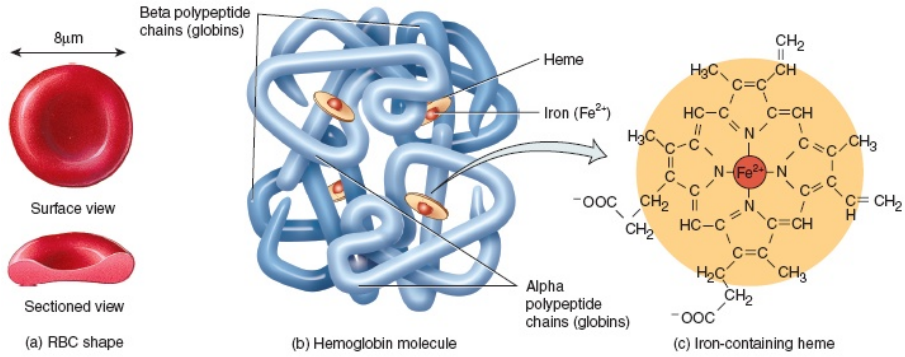


***Figure 1.4** . **Schematic representation of the hemoglobin molecule within RBCs.** a) The shape of a red blood cells (RBC). b) The hemoglobin molecule structure. c) Molecular structure of Heme group cointaining the iron ion (Fe^{2+}) (John Wiley and sons, 2011)*

The heme group consists of an iron (Fe) ion held in a heterocyclic ring, known as a porphyrin. This porphyrin ring consists of four pyrrole molecules cyclically linked together with the iron ion bound in the center. Oxygen binds to the iron ion interacting with the four nitrogens in the middle of the ring. Iron is bound covalently in the center of the globular protein, thanks to the imidazole ring. Iron must exist in the Fe^{2+} oxidation state to bind oxygen. If superoxide ion associated to Fe^{3+} is protonated, the hemoglobin iron will remain oxidized and incapable of binding oxygen (Figure 1.4c) ***(Linberg, Conover, Shum, & Shorr, 1998a, 1998b).***

Hemoglobin transports oxygen molecules from de lungs to body tissues. This process is highly regulated, and one of the most important regulators is the concentration of oxygen itself. The

hemoglobin molecule can be saturated with oxygen molecules (oxyhemoglobin), or de-saturated with oxygen molecules (deoxyhemoglobin). Hemoglobin exists in two forms, a *taut (Christensen & Weissman) form* (T) and a *relaxed form* (R) ***(Perutz, 1978)***. Under high partial pressure of the system the relaxed (high affinity, R) state is favoured. Inversely, at low partial pressures, the (low affinity, T) tense state is favoured (***Keates, 2004***). Additionally, the binding of oxygen to the Fe^{2+} heme pulls iron into the plane of the porphyrin ring, causing a slight conformational shift. The shift encourages oxygen to bind to the three remaining hemes groups within hemoglobin.

THE HUMAN GLOBINS GENES STRUCTURE

The human α-globin locus:

The α-globin genes encoding the α-globin chains are duplicated and localized in the telomeric region of chromosome 16, in a cluster containing embryonic zeta2 gene, encoding the embryonic zeta globin chains, three pseudogenes (pseudo zeta1, pseudo alpha1, and pseudo alpha2) and one gene (theta1) of unknown function (5'-ζ2-ψζ1-ψα2-ψα1-α1-α2-Θ-3'). The globin ζ2 is expressed during the embryonic stage, whereas α1 and α2 genes are expressed during fetal and adult stages. Both α1 and α2 encode identical α-globin chains ***(Liebhaber, Cash, & Ballas, 1986)***. The functional genes are arranged in the order of their expression and their expression is regulated by four remote highly

conserved noncoding regions (named multispecies conserved sequences [MCS]) located about 40 kb upstream in the gene cluster ***(Higgs et al., 1990)***. This region is a DNase hypersensitive site and a transcription factors binding site, firstly named as HS-40 and later it was named as α-MRE (Major Regulatory element), containing several binding sites of erythroid-specific transcription factors. This region is a powerful erythroid-specific enhancer containing several binding sites of erythroid-specific transcription factors (Figure 1.5) ***(De Gobbi et al., 2007) (H. Chen, Lowrey, & Stamatoyannopoulos, 1997)***.

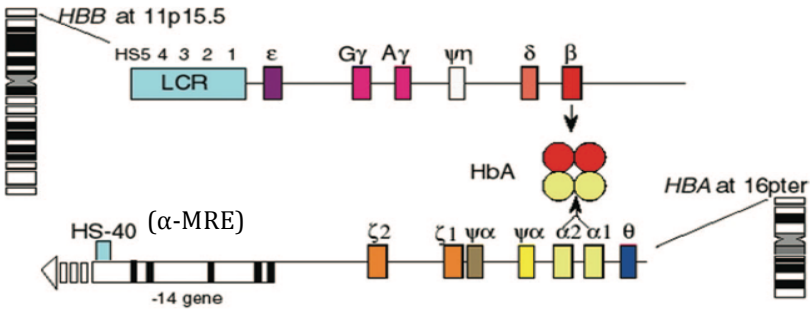


Figure 1.5 *. Chromosome localization and genomic structure of the human α-like and β-like globin genes clusters. (Galanello & Origa, 2010).*

The human β-globin locus:

The β-globin locus is localized in the chromosome 11, and it encodes the five different β-like globin chains expressed during development; the embryonic (ε), fetal (^{G}γ and ^{A}γ) and the adult (δ and β) globins genes. The subsets of globins genes are actively

expressed or repressed during development in the same linear order in which they are located within the chromosome (5'-ε-^{G}γ-^{A}γ-ψβ-δ-β-3') ***(Bank, 2006; Sankaran & Nathan, 2010; Sankaran, Xu, & Orkin, 2010)*** (Figure 1.5).

The globin-coding region was first cloned as cDNA ***(Rabbitts, 1976)***, and one of the first complete genes to be cloned and sequenced ***(Proudfoot & Longley, 1976; Rougeon, Kourilsky, & Mach, 1975).*** The cluster contains six genes, five of them encode functional proteins, while the sixth one (β-globin pseudogene – HBBP1) does not produce a recognizable protein product. In the past was considered as "genomic fossil" or "junk DNA". However, it has been probed that pseudogenes play essential roles in the regulation of other genes and are transcribed into a variety of functional RNAs. The HBBP1 sequence is well conserved between humans and apes, suggesting its functionally importance ***(Moleirinho et al., 2013)***. Other evidences of its functionality are the presence of a single base mutation in the pseudogene associated with β-thalassemia ***(Giannopoulou et al., 2012; Nuinoon et al., 2010; Roy et al., 2012)***. HBBP1 encodes two regulatory RNAs, and multiple transcripts. Moreover, the key regulatory regions of the pseudogene present marks commonly associated with active regulatory regions ***(Jeffrey P. Tomkins 2013).***

The β-globin locus is surrounded by a large cluster of olfactory receptor (OR) genes ***(Molete, Petrykowska, Sigg, Miller, &***

Hardison, 2002; Sheffield et al., 2013), which are not expressed in erythroid cells ***(Bulger et al., 2000)***. The embryonic-to-adult transitions are accompanied by a differential gene expression of the β-like globin genes. To ensure the proper expression in each stage there are a variety of regulatory elements that can be detected as DNaseI hypersensitive sites (HSs) in the chromatin of erythroid cells and include cis-elements (promoter, enhancers and silencers), which flank the genes and are sufficient and necessary for the developmentally-regulated and tissue-specific expression profile of the individual genes. Each globin gene has its own promoter. All globins promoters contain three consensus motifs the CACCC (Kruppel-like factor 1 (KLF1) recognition motif), CCAAT and TATA boxes. The most important distant cis-acting element controlling the differential gene expression during development is the Locus Control Region (LCR) ***(Grosveld, van Assendelft, Greaves, & Kollias, 1987)***. LCR is a 15 kb regulatory element upstream of the globin structural genes. The LCR contains four highly conserved subdomains, HS1-HS4, which were originally identified as DNase I hypersensitive sites (HSs). During erythroid differentiation there is a structural transition where LCR interacts with the active genes ***(Palstra et al., 2003)***.

This shows the functional significance of long-range contacts between regulatory DNA elements and the formation of an active chromatin hub (ACH) spatial structure, in which DNA containing inactive genes would be looped out from this configuration ***(Palstra, de Laat, & Grosveld, 2008)***. Despite of LCR has been

largely studied, the molecular mechanism that serially activates and represses globins genes expression is still unclear.

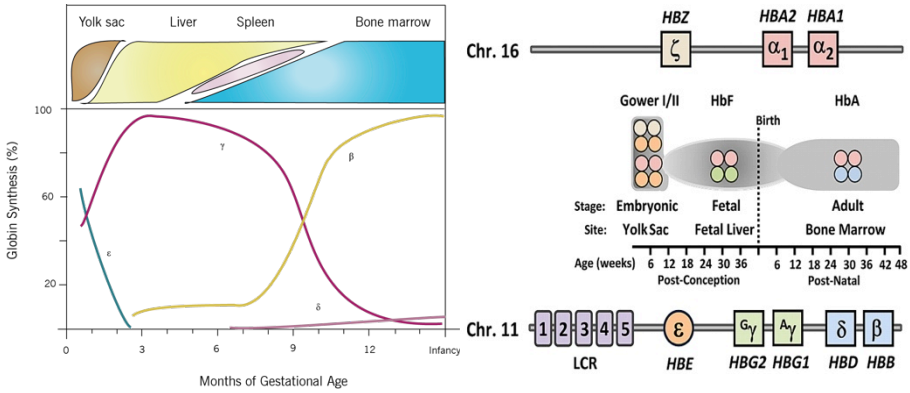


Figure 1.6 *Left panel: Globin expression during erythroid development (Copyright © 2012-2015 McMaster Pathophysiology Review) . Right panel: Schematic of genomic structural organization of the human α-globin and β-globin loci and temporal expression of the various hemoglobin types* **(Wilber et al., 2011)**.

The ε-globin gene is expressed in the embryonic yolk sac during the first few months of pregnancy. Starting from the second month of gestation, when definitive erythroid cells emerge, the ε-gene is silenced and the fetal γ-globin expression is activated in the fetal liver. Around the time of birth, fetal γ-globin expression is decreased and the adult β-globin gene is activated. This process, in which all the β-globin genes are differentially expressed at different developmental stages, is called "hemoglobin switching" (Figure 1.6).

The two γ-globin genes ($^{G}\gamma$ and $^{A}\gamma$) are the result of a 5 Kb tandem duplication. The coding region of these two regions differs by one nucleotide in exon 3, where the $^{G}\gamma$ gene codon is GGA and the $^{A}\gamma$ gene codon is GCA ***(Schroeder & Kurth, 1971)*** (Figure 1.5 and 1.6).

The first hemoglobin produced in the yolk sac is formed by Gower I ($\zeta_2\varepsilon_2$) and Gower II ($\alpha_2\varepsilon_2$) tetramers. When the first switch takes place, around 13th week of gestation, $^{G}\gamma$ and $^{A}\gamma$ genes are activated in the fetal liver and they remain expressed till the time of birth. The β- and δ-globin genes are first activated in the fetal liver and they increase their expression after birth, completely replacing fetal hemoglobin (HbF; $\alpha_2\,{}^{G}\gamma_2$ and $g_2\,{}^{A}\gamma_2$) with adult hemoglobin (HbA ($A_2\beta_2$) and HbA_2 ($_2\delta_2$)) producing the second switch. After the first year of life the amount of residual HbF is less than 1%. The d switch-globin genes are expressed in erythroid cells derived from the BM. The delta globin is expressed at very low level; so, the "major adult hemoglobin" is HbA whereas the "minor adult hemoglobin" is HbA_2, expressed between 1,5% and 3,5% of the total hemoglobin in adult life ***(Rochette, Craig, & Thein, 1994)*** (Figure 1.6).

MOUSE GLOBIN GENES:

In mouse, four different globin genes constitute the β-like globin chain (εϒ, βh1, β-major and β-minor globin genes). During primitive erythropoiesis εϒ and βh1 are sequentially expressed. At the beginning of definitive erythropoiesis, that takes place in fetal liver between days E11.5 and E13.5, the β-major and β-minor adult genes start to be expressed. This process has several molecular mechanisms in common with the switching of human globins genes. Several studies demonstrated that in transgenic mice carrying the human β-locus, human genes have a proper temporal regulation of gene expression. In embryonic fetal liver cells at E10.5 stage the most abundant globin is the fetal γ-globin gene, but then during the embryonic stages (E11.5 and E13.5) the hemoglobin switching takes place (the fetal γ-globin is silenced and the expression of the adult β-globin is established (Figure 1.7) **(McGrath et al., 2008)**.

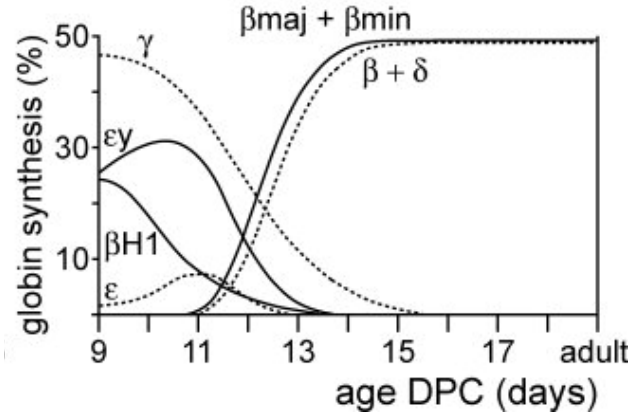


Figure 1.7 β -like globin protein content during development mouse as the percentage of the total globin content in the cells*. The graph shows, both the globin content of the mouse globins (straight line) and human transgenically expressed globins (dashed line) are shown* ***(Noordermeer & de Laat, 2008)***.

β-GLOBIN DISORDERS:

Failure in hemoglobin production causes a group of diseases collectively called "hemoglobinopathies". These diseases are among the most common inherited recessive pathologies. Globins genes mutations occurred numerous times during human history and some of them have been selected to high frequencies in malarial endemic regions. Although the acquisition of heterozygous mutation in the globin genes there confers protection against malaria, the co-inheritance of the 2 mutant β-globin alleles, in homozygous or in heterozygous combination, causes the severe β-globin disorders. Sickle cell disease (SCD) and β-thalassemias are the most common inherited chronic diseases with considerable morbidity and mortality. In countries with limited resources, most of the affected individuals die in early childhood. Nearly 300.000 infants are estimated born with SCD each year (the majority in Africa), and another 40.000 affected by β-thalassemia. Moreover, these disease are expected to rise over the years in parallel with the expected increase of worldwide population.

Sickle cell disease (SCD):

Sickle cell disease is a heritable blood disorder that can have devastating effects. It is the result of a single mutation: the amino acid, acid glutamic (a charged residue) is replaced by another amino acid, a valine (a hydrophobic residue), at the sixth amino acid position in the β-globin gene ***(Ingram, 1956; Pauling, Itano, & et al., 1949)***. This small change has major consequences; the presence of HbS can convert the flexible RBCs shape into a rigid shape during deoxygenation, giving them a distorted sickle-like appearance.

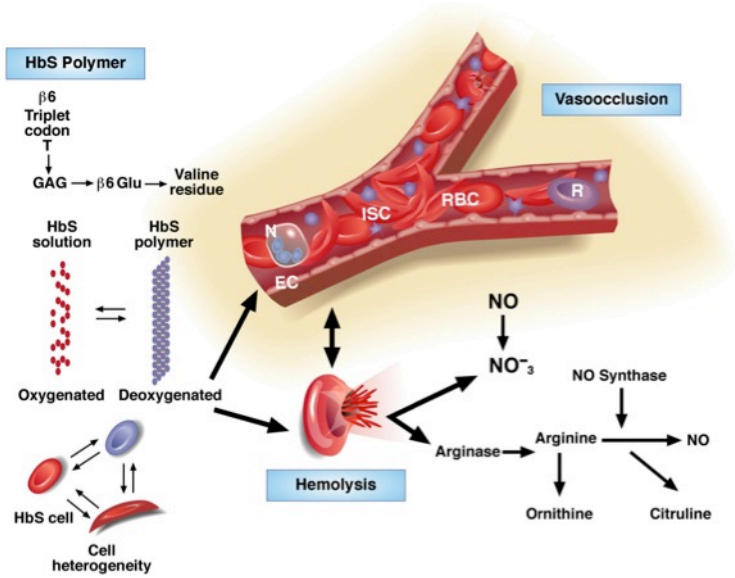


Figure 1.8 Sickle Cell Disease. *Effects of Sickle Cell Disease (SCD) on Red Blood Cells (RBCs). Under deoxygenation conditions HbS have the ability to form polymers, what produced vasoocclusion and hemolysis (Martin Steinberg, Murphylaboratory).*

These cells are also stickier than normal RBCs, what causes them to adhere to each other and to plug up narrow blood vessels, reducing blood flow and preventing adequate oxygen delivery, finally resulting in the necrosis of the tissue (Figure 1.8) (Okpala, 2004; Okpala, Ugochukwu, Tawil, & Greaves, 2004). Sickle cells have a lifespan of only 10-20 days. People with SCD commonly experience episodes of crises, what produces an acute pain that can last for few hours to few days. Other complications include anemia, leg ulcers, jaundice, kidney damage, high blood pressure, gallstones, increased susceptibility to infections and stroke. Nowadays adequate therapies increase life expectancy of SCD patients, such as penicillin preventing infections, folic acid stimulating RBCs production, hydroxyurea reducing the number of painful crisis, and blood transfusions tempering acute cases of the disease. However, in low-income countries at least 50% of children born with the illness die within five years ***(Grosse et al., 2011)***.

β-thalassemias:

β-thalassemias are a group of hereditary blood disorders caused by anomalies in the production of β-globin chains, resulting in a variable phenotypes ranging from severe anemia to clinically asymptomatic conditions. β-thalassemia is the most common autosomal recessive disorders worldwide; a to clinically asymptomatiction of β-globin ch^{+}) or absence (β^{0}) of the β-globin

chains. The resulting excess of unbound excess o makes them to precipitate within the RBC causing oxidative damage to the cell membrane of erythroid precursors in the bone marrow ***(Rund & Rachmilewitz, 2005; Weatherall, Williams, Allen, & O'Donnell, 2010)***. Such damaged cells are recognized by macrophages and quickly removed from the circulation, leading to ineffective erythropoiesis. β-thalassemia comes in several diverse forms (Table 1). Depending on the clinical phenotype, thalassemia can be classified into three main forms: Thalassemia Major (known also as "Cooley's Anemia" and "Mediterranean Anemia"), Thalassemia intermedia and thalassemia minor (also called "β-thalassemia carrier", "β-thalassemia trait" or "heterozygous β-thalassemia"). These diseases are caused by different point mutations or deletions affecting the β-globin gene. So far, more than 200 mutations causing β-thalassemia have been described. Deletions in the β-globin gene are less common cause of β-thalassemia ***(Galanello & Origa, 2010)***.

Peripheral hemolysis occurs when insoluble α-globin chains induce membrane damage to the peripheral erythrocytes. The resulting anemia, in turn, stimulates the production of erythropoietin with consequent intensive but ineffective expansion of the bone marrow. Prolonged anemia also causes hepatosplenomegaly and extramedullary erythropoiesis.

The most common treatment for β-thalassemia major is blood transfusion plus chelation therapy. The goals of the transfusion

therapy are the correction of anemia, the suppression of ineffective erythropoiesis and the inhibition of gastrointestinal iron absorption, which occurs in non-transfused patients as a consequence of ineffective erythropoiesis. However, also patients maintained with a regular transfusion therapy develop iron overload, and need a chelation treatment.

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

Table 1. Different types of th thalassemia*(Galanello & Origa, 2010)*.

Hereditary Persistence of Fetal Hemoglobin (HPFH):

Early in the 1960s, in some people from both Afro-American and Greek population a persistence of the fetal hemoglobin (HbF) during adult life was observed ***(Wheeler & Krevans, 1961)*** ***Fessas (Fessas & Stamatoyannopoulos, 1962; Fessas, Stamatoyannopoulos, & Karaklis, 1962)***. This condition was named Hereditary Persistence of Fetal Hemoglobin (HPFH), and it was subsequently shown to be the result of a variety of mutations HPFH can be cause by two types of mutations: "linked-HPFH" involves large deletions at the (HPFH), and it was subsequently shown to be the result of a variety of mutations HPFH can be cause by two types of mutations: "linked-HPFH" involves large deletions at the β-globin cluster or point mutations in the promoters of the γ-globin genes. "Non-linked-HPFH", where elements unlinked to β-globin gene cluster are able to upregulate HbF expression (for example polymorphism reducing the repression of γ-globin produced by BCL11a, and polymorphisms on the long arm of chromosome 6, located in a non-protein-coding region between the genes HBS1L and MYB) ***(Uda et al., 2008)***.

HPFH can be subdivided into deletion and non-deletion HPFH. Non-deletional HPFH are associated by a change in the DNA-protein pattern interactions, and have been located between the -260 and +25 region of the γ-globin promoter, suggesting that repressive factors act in the promoter ***(Gumucio et al., 1988; Ikuta & Kan, 1991; Jane, Ney, Vanin, Gumucio, & Nienhuis,***

1992; Mantovani et al., 1987; Mantovani et al., 1988; Martin, Tsai, & Orkin, 1989; O'Neill, Kaysen, Donovan-Peluso, Castle, & Bank, 1990). The point mutations in the γ-globin promoter are located at positions -114, -175 and -200 in the 5' promoter regions HBG2 and HBG1 ***(Chakalova, Carter, et al., 2005; Chakalova, Osborne, et al., 2005; Gibney et al., 2008; Hoyer, McCormick, Snow, Kwon, et al., 2002; Hoyer, McCormick, Snow, Lawler, et al., 2002; Hoyer, Rachut, et al., 2002; L. R. Liu et al., 2005)***. Moreover, large deletions in the HPFH can also be a consequence of mutations occurred outside the locus, such as some mutations in the KLF1 gene ***(Borg et al., 2010; Satta, Perseu, Maccioni, Giagu, & Galanello, 2012; Satta et al., 2011)***. Recently, quantitative trait loci (QTL) study reveals genetic elements unlinked to β-globin gene cluster able to upregulate HbF expression (for example polymorphism within intron 2 of the zinc-finger transcription factor BCL11A producing a reduced expression of this gene, and polymorphism on the long arm of chromosome 6 located in a non-protein-coding region between the genes HBS1L and MYB) ***(Uda et al., 2008)***.

Interestingly, the elevated levels of HbF observed in HPFH condition, ameliorate the severity of SCD and β-thalassemia ***(Z. Chen et al., 2009; Clegg & Gagnon, 1981; Stamatoyannopoulos et al., 1994)***. In fact, the presence of HbF interferes with the polymerization and increases the solubility of HbS under deoxygenated conditions ***(Sunshine, Hofrichter, & Eaton, 1978)***. HbF expression during adult life reduces the painful crises in SCD

(Platt, 1994; Platt et al., 1994). Moreover, the persistence of HbF in patients with β-thalassemia corrects the globin chain imbalance, rendering erythropoiesis more effective.

On the above basis, reactivation of γ-globin expression is one of the main focuses to treat β-hemoglobin disorders. So, HPFH mutations have been largely studied in order to better understand the mechanisms to reactivate fetal hemoglobin. During 1980s it was used 5-azacytadine, a DNA demethylating agent able to increase the HbF levels. However, this treatment produces highly toxic profile ***(Ley et al., 1982)***. Hydroxyurea (HU) was then discovered as HbF inducer. HU treatment is beneficial for many patients with SCD, although it has a variable efficacy, requires careful monitoring with dose-limiting myelosupression, and it is of limited utility for β-thalassemia ***(Platt, 2008; Taher et al., 2010)***. A potent, uniform and safe HbF reactivator remains to be discovered. The poor understanding of hemoglobin switching mechanisms makes more difficult the development of new therapies for HbF reactivation.

TRANSCRIPTIONAL REGULATION OF βRANSCRI LOCUS:

Erythropoiesis is highly regulated by a transcriptional program controlled by lineage-specific transcription factors (TFs) ***(Cantor, Katz, & Orkin, 2002; Kim, Zhao, Ifrim, & Dean, 2007)***. Erythroid TFs were first identified through their binding to the promoters of

β-globin genes. The analysis of TFs null hematopoietic cells has shown their molecular and cellular functionally relevance. Furthermore, biochemical and proteomic analysis reveal multiprotein interactions and signaling pathways regulating erythroid maturation ***(Akashi et al., 2003)*** (Figure 1.9).

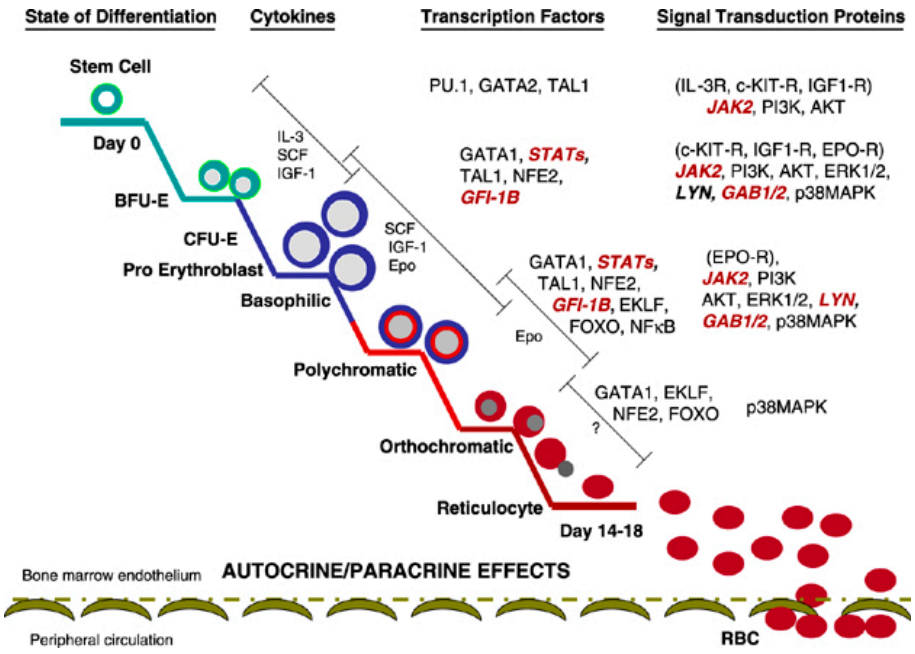


Figure 1.9 Transcriptional regulation during erythropoiesis. *There are also outlined the cytokines requirement in each stage, and the signal transduction pathway activated during differentiation*

Here below is a short description of the main TFs involved in the transcriptional control of erythropoiesis.

GATA-1: is a member of the GATA family of TF, whose members are playing critical roles during development in different tissues. GATA-1 is crucial for erythroid differentiation. All the GATA members contain two conserved zing-finger domains and bind to the (A/T)GATA(A/G) consensus DNA binding motif. GATA-1 is expressed in erythroid cells, megakaryocytes, mast cells, eosinophils, and dendritic cells ***(Ferreira, Ohneda, Yamamoto, & Philipsen, 2005; Gutierrez & Sung, 2007)***. GATA-1 contains three functional domains: the N-terminal domain, and two zing-finger domains located in the C-terminal part of the protein. The N-terminal domain has been reported as the transactivational domain. The C-terminal zing-finger is responsible for the DNA binding, whereas the N-terminal one is responsible for protein-protein interaction. GATA-1 null mice die early during gestation (around day E10.5-E11.5) due to severe anemia. GATA-1 null embryonic stem (ES) cells fail to produce RBCs. Pro-erythroblasts are unable to differentiate due to apoptotic cell death ***(Pan et al., 2005; Pevny et al., 1995; Weiss, Keller, & Orkin, 1994)***. In mouse, as little as 5% of physiological level of GATA-1 can prevent apoptosis, but it is not sufficient to end up into terminal maturation. GATA-1 promotes survival of erythroid cells upregulating anti-apoptotic genes such as Bcl-X_L ***(Gregory et al., 1999)*** and by activating the transcription of EPO receptor ***(Chiba et al., 1993)***, whose signaling is known to be important for progenitors survival ***(C. Lacombe & Mayeux, 1999a, 1999b; F. Lacombe et al., 1999)***.

GATA-1 form several complexes interacting with many other TFs, such as FOG-1, EKLF, TAL1/SCL, PU.1 and with cofactors such as CBP/p300, Brg1, MeCP1/NuRD, and others ***(Lowry & Mackay, 2006)***. Several analyses suggest that different GATA-1 protein sub-complexes play a differ role on different target genes. For example, GATA-1/FOG1/MeCP1 complex represses early hematopoietic and alternative lineage transcription programs ***(Rodriguez et al., 2005)***. Meanwhile, GATA-1/Gif-1b represses the Myc gene, and GATA-1/FOG-1 and GATA-1/TAL1/SCL up-regulate the erythroid transcription program. ***(Crispino, Lodish, MacKay, & Orkin, 1999; Tsang et al., 1997; Wadman et al., 1997)***.

FOG-1: FOG-1 is a zing-finger protein close-interacting partner of GATA-1, and for that reason it was named Friend Of GATA (FOG-1). It contains nine zing-fingers, which do not bind to DNA directly ***(A. N. Chang et al., 2002)***. Four of these zing-fingers contribute to the GATA-1/FOG-1 interaction ***(Fox et al., 1999)***. FOG-1 expression in hematopoiesis is mirroring GATA-1 expression; it is highly expressed in erythroid cells and megakaryocytes. FOG-1 null mice have similar phenotype as GATA-1 knockout mice: embryonic lethality due to severe anemia ***(Tsang, Fujiwara, Hom, & Orkin, 1998)***. GATA-1/FOG-1 interaction is essential for erythroid differentiation ***(Crispino et al., 1999)***. The repressive complexes MeCP/NuRD interact with GATA-1/FOG-1 to repress GATA-1 targets ***(Hong et al., 2005)(Rodriguez et al., 2005)***. It is still unclear how FOG-1/GATA1 interaction mediates gene

activation. One hypothesis is that FOG-1 facilitates the GATA-1 induced bending of DNA, bringing together distal enhancer elements and gene promoters ***(Vakoc et al., 2005)***.

TAL-1/SCL/LMO2/Ldb1/E2A complex: TAL-1/SCL is a basic helix-loop-helix TF that binds to the E-box consensus sequence (CANNTG). TAL-1/SCL is expressed in erythroid cells, megakaryocytes and mast cells ***(Cantor et al., 2002; Cantor & Orkin, 2002; Kim et al., 2007)***. TAL-1/SCL null mice result in a total absence of hematopoiesis in the yolk sac, and its conditional erythroid knockout produce a failure in erythropoiesis ***(Mikkola et al., 2003; Porcher et al., 1996; Shivdasani, Mayer, & Orkin, 1995)***.

TAL-1/SCL forms a multiprotein complex with E2A, LMO2 and Ldb1, interacting with GATA-1 to form a pentameric complex that binds to the E-box/GATA-1 DNA motifs ***(Wadman et al., 1997)***, which are separated by 9 to 12 nucleotides. This regulatory module is present in many regulatory elements of erythroid genes ***(Soler et al., 2010)***. TAL-1 is a TF that can be both activator and repressor ***(Schuh et al., 2005)***. The repression function of TAL-1 is primarily controlled though its interaction with ETO-2 ***(Soler et al., 2010; Tripic et al., 2009)***. However, the precise function of TAL-1/SCL together with ETO-2 in early erythropoiesis is still unclear. Ldb1 complex is essential to establish the interactions of the β-globin gene and the LCR to provide transcriptional

activation ***(Song et al., 2010; Z. Xu, Huang, Chang, Agulnick, & Brandt, 2003)***.

KLF1/EKLF: The Krüppel-like factor 1 (EKL1) is an erythroid specific zing-finger TF with a crucial role in globins gene regulation ***(Miller & Bieker, 1993; Nuez, Michalovich, Bygrave, Ploemacher, & Grosveld, 1995; Perkins, Sharpe, & Orkin, 1995)***. It binds to the CACC box motifs present in the β-globin promoter and in the promoters of many erythroid genes. Mutations in the CACC box in the β-globin promoter are related with occurrence of β-thalassemia ***(Bieker & Southwood, 1995; Feng, Southwood, & Bieker, 1994)***. KLF1 null mice die due to severe anemia during gestation (around E14.5-E15.5) because of a failure in the β-globin gene activation.

KLF1 acts as an activator and as a repressor depending on its post-translational modifications that promote protein-protein interaction. KLF1 binds specifically to adult globin gene and also to HS1, HS2, HS3 and HS4 throughout all development stages. The expression level of KLF1 increases during erythroid development ***(Zhou, Pawlik, Ren, Sun, & Townes, 2006)***. KLF1 also strongly binds to the BCL11A (binds to the BCL11A promoter activating its expression ***(Sankaran, Menne, et al., 2008; Sankaran, Orkin, & Walkley, 2008)***. During the early stages of the development, low levels of KLF1 results in low levels of BCL11A and β-globin, but high levels of γ-globin expression ***(Borg, Patrinos, Felice, & Philipsen, 2011)***. As the expression levels of KLF1 increase during

development, consequently the expression levels of BCL11A and on levels also increase ***(Siatecka & Bieker, 2011)***.

NF-E2: is an erythroid specific transcription factor ***(Romeo et al., 1990)***. It belongs to the small *Maf* protein family ***(Andrews, Erdjument-Bromage, Davidson, Tempst, & Orkin, 1993; Andrews, Kotkow, et al., 1993)***. NF-E2 is expressed in erythroid and megakaryocytic cells. It interacts with Maf recognition elements (MARE) within HS2 and HS3 sub-regions of the LCR in the β-globin locus ***(Forsberg, Downs, & Bresnick, 2000; Kang, Vieira, & Bungert, 2002; Talbot & Grosveld, 1991)***. NF-E2 null mice have no significantly effect in erythropoiesis or in in β-globin transcription. NF-E2 contains two different domains: p45 that is the hematopoietic subunit, and p18 (MafK), the ubiquitous subunit. P45 interacts with TAFII30, and together they mediate looping between the LCR and the globin genes promoters ***(Amrolia et al., 1997)***. Moreover, NF-E2 cooperates with GATA1 to recruit RNA-polymerase II to the β-globin promoters ***(Johnson, Grass, et al., 2002; Johnson, Norton, & Bresnick, 2002)***.

BCL11A: BCL11A (B-cell lymphoma/leukemia 11A) is a C2H2 type zinc-finger transcription factor with a crucial role in the silencing of γ-globin gene and hemoglobin switching ***(Suzuki et al., 2002)***. There exists four different alternative isoforms of BCL11A, known as: eXtra long (XL), long (L), short (S) and eXtra-short ***(Hitoshi et al.)***. BCL11A is required for normal lymphoid development as the

knockout mice embryos lack B-cells and have alterations in several types of T-cells ***(J. J. Liu, Hou, & Shen, 2003)***.

A quantitative trail loci (QTL) approach revealed that a polymorphisms within the BCL11a locus reducing its expressing causes upregulation of HbF ***(Uda et al., 2008)***. Furthermore, down-regulation of BCL11A in adult human erythroid precursors induces HbF expression ***(Sankaran, Menne, et al., 2008; Sankaran, Orkin, et al., 2008)***. BCL11A interacts with GATA-1, FOG-1 and with some subunits of NuRD complex including CHD4, HDAC1-2 and MTA2 ***(H. Chang & Shih, 2009)***. Sox6 and BCL11A physically interact in order to silence γ-globin expression ***(J. Xu et al., 2010)***.

COUP-TFII: (Chicken ovalbumin Upstream promoter-transcription factor II), also known as NR2F2/ARP-1. Orphan nuclear receptor TF with no ligands known to bind it. COUP-TFII forms homodimers and binds direct repeats elements, composed by two single consensus sites (AGGTCA) ***(Cooney, Tsai, O'Malley, & Tsai, 1992)(Kliewer et al., 1992)***. Knockout mice have early gestation lethality due to a failure in several developmental programs demonstrating that COUP-TFII is implicated in many vital processes ***(Tang, Alger, & Pereira, 2006)***. In hematopoiesis, COUP-TFII is expressed during the embryonic/fetal developmental stages, in the yolk sac and in the fetal liver ***(Filipe et al., 1999)***. COUP-TFII was described to bind in vitro to the ε- and γ-globin promoters, where it binds to the double CCAAT box

region ***(Liberati, Sgarra, Manfioletti, & Mantovani, 1998; Ronchi, Bellorini, Mongelli, & Mantovani, 1995)***. COUP-TFII might contribute to the HPFH phenotype. In fact, HPFH mutations mapped in the double CCAAT box region of the γ-globin promoter are associated with the perturbation of the COUP-TFII binding in vitro ***(Liberati et al., 2001)***.

Ex vivo human CD34 erythroid cultures from peripheral blood show that Stem Cell Factor (SCF) reduces COUP-TFII expression at the mRNA level and protein level, resulting in the up-regulation of γ-globin expression ***(Aerbajinai, Zhu, Kumkhaek, Chin, & Rodgers, 2009)***. Recent studies suggest that the NF-Y transcription factor recruits and stabilizes the binding of BCL11A with COUP-TFII on the proximal γ-globin promoter region forming the repression complex silencing γ-globin expression ***(Zhu et al., 2012)***.

Sox6: The Sox family (Sry-related HMG box) is a group of TFs formed by 20 different Sox proteins classified into eight groups (A-H, with two B groups, B1 and B2) depending on their degree of identity **(Bowles, Schepers, & Koopman, 2000)**. This family comprises important transcriptional regulators with a key role during embryonic development and in the determination of cell fate of many cell types. Their common HMG box domain can bind to DNA, producing DNA bending, thus suggesting a chromatin architectural role for these proteins. Moreover, Sox proteins can establish a complex network of protein interactions, crucial for

their function **(Lefebvre, Dumitriu, Penzo-Mendez, Han, & Pallavi, 2007)**. In particular, Sox6 was originally isolated from adult mouse testis **(Connor, Nornes, & Neuman, 1995)**. Sox6 is essential for the development of the central nervous system **(Hamada-Kanazawa, Ishikawa, Nomoto, et al., 2004; Hamada-Kanazawa, Ishikawa, Ogawa, et al., 2004)**, for chondrogenesis **(Han & Lefebvre, 2008; Ikeda et al., 2004; Lefebvre, Behringer, & de Crombrugghe, 2001; Lefebvre & Smits, 2005; Smits, Dy, Mitra, & Lefebvre, 2004; Smits et al., 2001)** and for cardiac and skeletal muscle formation **(Cohen-Barak et al., 2003; Hagiwara et al., 2000)**. Furthermore, Sox6 has an important role in the erythropoietic development, where it stimulates cell survival, proliferation and terminal maturation of definitive mature RBCs. Sox6 null mice die at birth or within two weeks of age due to skeletal, glial and cardiac developmental defects. They also bear a large portion of nucleated definitive RBCs and are anemic **(Dumitriu, Dy, Smits, & Lefebvre, 2006; Dumitriu, Patrick, et al., 2006)**. Moreover, Sox6 is necessary for efficient erythropoiesis in adult mice under both basal and stress conditions **(Dumitriu et al., 2010)**, as shown by the conditional KO where mice show compensated anemia, erythroid cell developmental defects, and short-lived RBCs. Sox6 overexpression in both erythroleukemic cell lines and in *ex vivo* primary human $CD34^+$ cells purified from cord blood causes cell cycle withdrawal and terminal maturation, together with a general increase of several erythroid specific genes, including globins ***(Cantu, Ierardi, et al., 2011).*** Beside a general role on enhancing

erythroid differentiation, Sox6 plays a specific role in the differential regulation of globin genes. In fact, it directly binds into the εϒ gene promoter silencing its expression in definitive adult erythroid cells from mice fetal liver E15.5. Furthermore, these cells in Sox6-null mice show a persistence of the embryonic εϒ-globin expression in comparison to WT cells **(Yi et al., 2006)**. Furthermore, Sox6 cointeract with BCL11a-XL in *ex* vivo human $CD34^{+}$ cells from peripheral blood, they collaborate on the γ-globin promoter silencing its expression (***Xu et al., 2013***).

SCOPE OF THE THESIS:

The aim of my thesis is to better characterize the molecular mechanisms of Sox6 function in erythropoiesis and, more specifically, its role in regulating globins genes by identifying Sox6 interacting proteins

In fact, Sox6 can act both as an activator or a repressor during erythroid development, but it lacks any conventional transactivator or transrepressor domain. This evidence suggests that it must to interact with other proteins to exert its function.

Mapping Sox6 partners has a potential clinical relevance because it could lead to the identification of new target genes that could be manipulated to improve ineffective erythropoiesis, often associated with anemic syndromes, as well as to reactivate γ-globin gene expression in beta hemoglobinopathies.

In **chapter 2**, I will describe the single-step biotin-tagged technique used to isolate and identify Sox6 multiprotein complexes. I will present in detail i) the cell lines established to perform this experiment; ii) the technique used to purify the Sox6 complexes; iii) the Mass Spectrometry analysis of Sox6 complexes performed by the *Proteomics Center in the Erasmus University (Rotterdam, The Netherlands); iv)* the filtering criteria, and v) the validation approaches used to select the most promising candidate partners for further functional analysis

- In **chapter 3**, I will describe in detail the validation of three Sox6 interacting proteins: the two subunits of the NuRD complex MTA1 and HDAC1, the COUP-TFII transcription factor and the cyclin dependent kinase 13 (CDK13). In particular:
    - In **chapter 3.1,** I will describe the role of the repression NuRD complex in collaboration with Sox6.
    - In **chapter 3.2,** I will describe the functional role of the COUP-TFII and Sox6 in modulating the γ- and β-globins ratio. These studies were performed by lentiviral-mediated overexpression of these two transcription factors both in a human erythroid cell lien (β-K562)and in human *ex vivo* erythroid cultures from CD34+ cells isolated from peripheral blood
    - In **chapter 3.3**, I will describe the potential role of CDK13 in erythroid cells by down regulating its expression by using short hairpin RNA shRNA in β-K562 cells.

- In **chapter 4**, I will describe a new experimental platform to perform first step high-content screening (HCS) based on multiplex imaging on β-K562 cells. This assay is able to

detect changes in the γ/β ratio in response to drugs or gene modulators (siRNA). We took advantage of this method to analyze whether any of the putative interactors of Sox6 identified by the Mass Spectrometry analysis described in Chapter 2 has a potential role in regulating γ/β globins ratio.

- In **chapter 5,** the results of this thesis will be briefly discussed as well as the future perspectives of this research

REFERENCES:

Adams, G. B., & Scadden, D. T. (2006). The hematopoietic stem cell in its place. *Nat Immunol, 7*(4), 333-337. doi: 10.1038/ni1331

Aerbajinai, W., Zhu, J., Kumkhaek, C., Chin, K., & Rodgers, G. P. (2009). SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII. *Blood, 114*(1), 187-194. doi: 10.1182/blood-2008-07-170712

Akashi, K., He, X., Chen, J., Iwasaki, H., Niu, C., Steenhard, B., . . . Li, L. (2003). Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood, 101*(2), 383-389. doi: 10.1182/blood-2002-06-1780

Akashi, K., Reya, T., Dalma-Weiszhausz, D., & Weissman, I. L. (2000). Lymphoid precursors. *Curr Opin Immunol, 12*(2), 144-150.

Allen, T. D., & Dexter, T. M. (1982). Ultrastructural aspects of erythropoietic differentiation in long-term bone marrow culture. *Differentiation, 21*(2), 86-94.

Allen, T. D., & Dexter, T. M. (1984). The essential cells of the hemopoietic microenvironment. *Exp Hematol, 12*(7), 517-521.

Allen, T. D., Dexter, T. M., & Simmons, P. J. (1990). Marrow biology and stem cells. *Immunol Ser, 49*, 1-38.

Allen, T. D., & Testa, N. G. (1991). Cellular interactions in erythroblastic islands in long-term bone marrow cultures, as studied by time-lapse video. *Blood Cells, 17*(1), 29-38; discussion 39-43.

Alvarez-Silva, M., Belo-Diabangouaya, P., Salaun, J., & Dieterlen-Lievre, F. (2003). Mouse placenta is a major hematopoietic organ. *Development, 130*(22), 5437-5444. doi: 10.1242/dev.00755

Amrolia, P. J., Ramamurthy, L., Saluja, D., Tanese, N., Jane, S. M., & Cunningham, J. M. (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), 10051-10056.

Andreani, M., Testi, M., Gaziev, J., Condello, R., Bontadini, A., Tazzari, P. L., . . . Lucarelli, G. (2011). Quantitatively different red cell/nucleated cell chimerism in patients with long-term, persistent hematopoietic mixed chimerism after bone marrow transplantation for thalassemia major or sickle cell disease. *Haematologica, 96*(1), 128-133. doi: 10.3324/haematol.2010.031013

Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., & Orkin, S. H. (1993). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature, 362*(6422), 722-728. doi: 10.1038/362722a0

Andrews, N. C., Kotkow, K. J., Ney, P. A., Erdjument-Bromage, H., Tempst, P., & Orkin, S. H. (1993). The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc Natl Acad Sci U S A, 90*(24), 11488-11492.

Bank, A. (2006). Regulation of human fetal hemoglobin: new players, new complexities. *Blood, 107*(2), 435-443. doi: 10.1182/blood-2005-05-2113

Bieker, J. J., & Southwood, C. M. (1995). The erythroid Kruppel-like factor transactivation domain is a critical component for cell-specific inducibility of a beta-globin promoter. *Mol Cell Biol, 15*(2), 852-860.

Bondurant, M. C., & Koury, M. J. (1986). Anemia induces accumulation of erythropoietin mRNA in the kidney and liver. *Mol Cell Biol, 6*(7), 2731-2733.

Borg, J., Papadopoulos, P., Georgitsi, M., Gutierrez, L., Grech, G., Fanis, P., . . . Philipsen, S. (2010). Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet, 42*(9), 801-805. doi: 10.1038/ng.630

Borg, J., Patrinos, G. P., Felice, A. E., & Philipsen, S. (2011). Erythroid phenotypes associated with KLF1 mutations. *Haematologica, 96*(5), 635-638. doi: 10.3324/haematol.2011.043265

Bowles, J., Schepers, G., & Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on

sequence and structural indicators. *Dev Biol, 227*(2), 239-255. doi: 10.1006/dbio.2000.9883

Bulger, M., Bender, M. A., van Doorninck, J. H., Wertman, B., Farrell, C. M., Felsenfeld, G., . . . Hardison, R. (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), 14560-14565. doi: 10.1073/pnas.97.26.14560

Bunn, H. F. (2013). Erythropoietin. *Cold Spring Harb Perspect Med, 3*(3), a011619. doi: 10.1101/cshperspect.a011619

Cantor, A. B., Katz, S. G., & Orkin, S. H. (2002). Distinct domains of the GATA-1 cofactor FOG-1 differentially influence erythroid versus megakaryocytic maturation. *Mol Cell Biol, 22*(12), 4268-4279.

Cantor, A. B., & Orkin, S. H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene, 21*(21), 3368-3376. doi: 10.1038/sj.onc.1205326

Cantu, C., Grande, V., Alborelli, I., Cassinelli, L., Cantu, I., Colzani, M. T., . . . Ronchi, A. (2011). A highly conserved SOX6 double binding site mediates SOX6 gene downregulation in erythroid cells. *Nucleic Acids Res, 39*(2), 486-501. doi: 10.1093/nar/gkq819

Cantu, C., Ierardi, R., Alborelli, I., Fugazza, C., Cassinelli, L., Piconese, S., . . . Ronchi, A. (2011). Sox6 enhances erythroid differentiation in human erythroid progenitors. *Blood, 117*(13), 3669-3679. doi: 10.1182/blood-2010-04-282350

Chakalova, L., Carter, D., Debrand, E., Goyenechea, B., Horton, A., Miles, J., . . . Fraser, P. (2005). Developmental regulation of the beta-globin gene locus. *Prog Mol Subcell Biol, 38*, 183-206.

Chakalova, L., Osborne, C. S., Dai, Y. F., Goyenechea, B., Metaxotou-Mavromati, A., Kattamis, A., . . . Fraser, P. (2005). The Corfu deltabeta thalassemia deletion disrupts gamma-globin gene silencing and reveals post-transcriptional regulation of HbF expression. *Blood, 105*(5), 2154-2160. doi: 10.1182/blood-2003-11-4069

Chang, A. N., Cantor, A. B., Fujiwara, Y., Lodish, M. B., Droho, S., Crispino, J. D., & Orkin, S. H. (2002). GATA-factor dependence of the multitype zinc-finger protein FOG-1 for

its essential role in megakaryopoiesis. *Proc Natl Acad Sci U S A, 99*(14), 9237-9242. doi: 10.1073/pnas.142302099

Chang, H., & Shih, L. Y. (2009). Concurrence of multiple myeloma and idiopathic erythrocytosis. *Acta Clin Belg, 64*(5), 434-435. doi: 10.1179/acb.2009.071

Chen, H., Lowrey, C. H., & Stamatoyannopoulos, G. (1997). Analysis of enhancer function of the HS-40 core sequence of the human alpha-globin cluster. *Nucleic Acids Res, 25*(14), 2917-2922.

Chen, Z., Luo, H. Y., Steinberg, M. H., & Chui, D. H. (2009). BCL11A represses HBG transcription in K562 cells. *Blood Cells Mol Dis, 42*(2), 144-149. doi: 10.1016/j.bcmd.2008.12.003

Chiba, T., Nagata, Y., Kishi, A., Sakamaki, K., Miyajima, A., Yamamoto, M., . . . Todokoro, K. (1993). Induction of erythroid-specific gene expression in lymphoid cells. *Proc Natl Acad Sci U S A, 90*(24), 11593-11597.

Christensen, J. L., & Weissman, I. L. (2001). Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc Natl Acad Sci U S A, 98*(25), 14541-14546. doi: 10.1073/pnas.261562798

Clegg, J. B., & Gagnon, J. (1981). Structure of the zeta chain of human embryonic hemoglobin. *Proc Natl Acad Sci U S A, 78*(10), 6076-6080.

Cohen-Barak, O., Yi, Z., Hagiwara, N., Monzen, K., Komuro, I., & Brilliant, M. H. (2003). Sox6 regulation of cardiac myocyte development. *Nucleic Acids Res, 31*(20), 5941-5948.

Connor, H., Nornes, H., & Neuman, T. (1995). Expression screening reveals an orphan receptor chick ovalbumin upstream promoter transcription factor I as a regulator of neurite/substrate-cell contacts and cell aggregation. *J Biol Chem, 270*(25), 15066-15070.

Cooney, A. J., Tsai, S. Y., O'Malley, B. W., & Tsai, M. J. (1992). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol, 12*(9), 4153-4163.

Crispino, J. D., Lodish, M. B., MacKay, J. P., & Orkin, S. H. (1999). Use of altered specificity mutants to probe a specific protein-

protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell, 3*(2), 219-228.

D'Andrea, A. D. (1994). Hematopoietic growth factors and the regulation of differentiative decisions. *Curr Opin Cell Biol, 6*(6), 804-808.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J. A., Sharpe, J. A., Koch, C. M., . . . Higgs, D. R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood, 110*(13), 4503-4510. doi: 10.1182/blood-2007-06-097964

Dore, L. C., Amigo, J. D., Dos Santos, C. O., Zhang, Z., Gai, X., Tobias, J. W., . . . Weiss, M. J. (2008). A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci U S A, 105*(9), 3333-3338. doi: 10.1073/pnas.0712312105

Dumitriu, B., Bhattaram, P., Dy, P., Huang, Y., Quayum, N., Jensen, J., & Lefebvre, V. (2010). Sox6 is necessary for efficient erythropoiesis in adult mice under physiological and anemia-induced stress conditions. *PLoS One, 5*(8), e12088. doi: 10.1371/journal.pone.0012088

Dumitriu, B., Dy, P., Smits, P., & Lefebvre, V. (2006). Generation of mice harboring a Sox6 conditional null allele. *Genesis, 44*(5), 219-224. doi: 10.1002/dvg.20210

Dumitriu, B., Patrick, M. R., Petschek, J. P., Cherukuri, S., Klingmuller, U., Fox, P. L., & Lefebvre, V. (2006). Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood, 108*(4), 1198-1207. doi: 10.1182/blood-2006-02-004184

Feng, W. C., Southwood, C. M., & Bieker, J. J. (1994). Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem, 269*(2), 1493-1500.

Ferreira, R., Ohneda, K., Yamamoto, M., & Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol, 25*(4), 1215-1227. doi: 10.1128/MCB.25.4.1215-1227.2005

Fessas, P., & Stamatoyannopoulos, G. (1962). Absence of haemoglobin A2 in an adult. *Nature, 195*, 1215-1216.

Fessas, P., Stamatoyannopoulos, G., & Karaklis, A. (1962). Hemoglobin "pylos": study of a hemoglobinopathy resembling thalassemia in the heterozygous, homozygous and double heterozygous state. *Blood, 19*, 1-22.

Filipe, A., Li, Q., Deveaux, S., Godin, I., Romeo, P. H., Stamatoyannopoulos, G., & Mignotte, V. (1999). Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *EMBO J, 18*(3), 687-697. doi: 10.1093/emboj/18.3.687

Forsberg, E. C., Downs, K. M., & Bresnick, E. H. (2000). Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells. *Blood, 96*(1), 334-339.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J. W., Jaenisch, R., Carey, V., & Hock, H. (2009). Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat Biotechnol, 27*(1), 84-90. doi: 10.1038/nbt.1517

Fox, A. H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., & Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *EMBO J, 18*(10), 2812-2822. doi: 10.1093/emboj/18.10.2812

Franco, R. S. (2012). Measurement of red cell lifespan and aging. *Transfus Med Hemother, 39*(5), 302-307. doi: 10.1159/000342232

Galanello, R., & Origa, R. (2010). Beta-thalassemia. *Orphanet J Rare Dis, 5*, 11. doi: 10.1186/1750-1172-5-11

Gekas, C., K, E. R., & H, K. A. M. (2008). Isolation and analysis of hematopoietic stem cells from the placenta. *J Vis Exp*(16). doi: 10.3791/742

Giannopoulou, E., Bartsakoulia, M., Tafrali, C., Kourakli, A., Poulas, K., Stavrou, E. F., . . . Patrinos, G. P. (2012). A single nucleotide polymorphism in the HBBP1 gene in the human beta-globin locus is associated with a mild beta-thalassemia disease phenotype. *Hemoglobin, 36*(5), 433-445. doi: 10.3109/03630269.2012.717515

Gibney, G. T., Panhuysen, C. I., So, J. C., Ma, E. S., Ha, S. Y., Li, C. K., . . . Chui, D. H. (2008). Variation and heritability of Hb F and F-cells among beta-thalassemia heterozygotes in Hong Kong. *Am J Hematol, 83*(6), 458-464. doi: 10.1002/ajh.21150

Goodell, M. A., Brose, K., Paradis, G., Conner, A. S., & Mulligan, R. C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med, 183*(4), 1797-1806.

Graf, T. (2002). Differentiation plasticity of hematopoietic cells. *Blood, 99*(9), 3089-3101.

Graf, T., & Enver, T. (2009). Forcing cells to change lineages. *Nature, 462*(7273), 587-594. doi: 10.1038/nature08533

Gregory, T., Yu, C., Ma, A., Orkin, S. H., Blobel, G. A., & Weiss, M. J. (1999). GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-xL expression. *Blood, 94*(1), 87-96.

Grosse, S. D., Odame, I., Atrash, H. K., Amendah, D. D., Piel, F. B., & Williams, T. N. (2011). Sickle cell disease in Africa: a neglected cause of early childhood mortality. *Am J Prev Med, 41*(6 Suppl 4), S398-405. doi: 10.1016/j.amepre.2011.09.013

Grosveld, F., van Assendelft, G. B., Greaves, D. R., & Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell, 51*(6), 975-985.

Gumucio, D. L., Rood, K. L., Gray, T. A., Riordan, M. F., Sartor, C. I., & Collins, F. S. (1988). Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. *Mol Cell Biol, 8*(12), 5310-5322.

Gutierrez, E., & Sung, L. A. (2007). Interactions of recombinant mouse erythrocyte transglutaminase with membrane skeletal proteins. *J Membr Biol, 219*(1-3), 93-104. doi: 10.1007/s00232-007-9065-5

Hagiwara, N., Klewer, S. E., Samson, R. A., Erickson, D. T., Lyon, M. F., & Brilliant, M. H. (2000). Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death. *Proc Natl Acad Sci U S A, 97*(8), 4180-4185.

Hamada-Kanazawa, M., Ishikawa, K., Nomoto, K., Uozumi, T., Kawai, Y., Narahara, M., & Miyake, M. (2004). Sox6 overexpression causes cellular aggregation and the neuronal differentiation of P19 embryonic carcinoma cells

in the absence of retinoic acid. *FEBS Lett, 560*(1-3), 192-198. doi: 10.1016/S0014-5793(04)00086-9

Hamada-Kanazawa, M., Ishikawa, K., Ogawa, D., Kanai, M., Kawai, Y., Narahara, M., & Miyake, M. (2004). Suppression of Sox6 in P19 cells leads to failure of neuronal differentiation by retinoic acid and induces retinoic acid-dependent apoptosis. *FEBS Lett, 577*(1-2), 60-66. doi: 10.1016/j.febslet.2004.09.063

Han, Y., & Lefebvre, V. (2008). L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol Cell Biol, 28*(16), 4999-5013. doi: 10.1128/MCB.00695-08

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I. M., & Ayyub, H. (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev, 4*(9), 1588-1601.

Hitoshi, S., Seaberg, R. M., Koscik, C., Alexson, T., Kusunoki, S., Kanazawa, I., . . . van der Kooy, D. (2004). Primitive neural stem cells from the mammalian epiblast differentiate to definitive neural stem cells under the control of Notch signaling. *Genes Dev, 18*(15), 1806-1811. doi: 10.1101/gad.1208404

Hong, W., Nakazawa, M., Chen, Y. Y., Kori, R., Vakoc, C. R., Rakowski, C., & Blobel, G. A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J, 24*(13), 2367-2378. doi: 10.1038/sj.emboj.7600703

Hoyer, J. D., McCormick, D. J., Snow, K., Kwon, J. H., Booth, D., Duarte, M., . . . Fairbanks, V. F. (2002). Four new variants of the alpha2-globin gene without clinical or hematologic effects: Hb Park Ridge [alpha9(alpha7)Asn-->Lys (alpha2)], Hb Norton [alpha72(EF1)His-->Asp (alpha2)], Hb Lombard [alpha103(G10)His-->Tyr (alpha2)], and Hb San Antonio [A113(GH2)Leu-->Arg (A2)]. *Hemoglobin, 26*(2), 175-179.

Hoyer, J. D., McCormick, D. J., Snow, K., Lawler, J., Jadick, M., Grageda, R., . . . Fairbanks, V. F. (2002). Three new variants of the alpha1-globin gene without clinical or hematologic effects: Hb Hagerstown [alpha44(CE2)Pro-->Ala (alpha1)]; Hb Buffalo [alpha89(FG1)His-->Gln (alpha1)], a hemoglobin variant from Somalia and Yemen; Hb Wichita

[alpha95(G2)Pro-->Gln (alpha1)]; and a second, unrelated, case of Hb Roubaix [alpha55(E4)Val-->Leu (alpha1)]. *Hemoglobin, 26*(3), 291-298.

Hoyer, J. D., Rachut, E., Kubik, K. S., Jones, R. T., Honig, G. R., Vida, L. N., & Fairbank, V. F. (2002). Hb Rampa [alpha 95(G2)pro-->Ser (alpha 2)] in a family of European ancestry: DNA analysis confirms the CCG-->TCG mutation at codon 95 of the alpha 2-globin gene; clinical and laboratory features. *Hemoglobin, 26*(4), 397-403.

Ikeda, T., Kamekura, S., Mabuchi, A., Kou, I., Seki, S., Takato, T., . . . Chung, U. I. (2004). The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage. *Arthritis Rheum, 50*(11), 3561-3573. doi: 10.1002/art.20611

Ikuta, T., & Kan, Y. W. (1991). In vivo protein-DNA interactions at the beta-globin gene locus. *Proc Natl Acad Sci U S A, 88*(22), 10188-10192.

Ingram, V. M. (1956). A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature, 178*(4537), 792-794.

Jane, S. M., Ney, P. A., Vanin, E. F., Gumucio, D. L., & Nienhuis, A. W. (1992). Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *EMBO J, 11*(8), 2961-2969.

Johnson, K. D., Grass, J. A., Boyer, M. E., Kiekhaefer, C. M., Blobel, G. A., Weiss, M. J., & Bresnick, E. H. (2002). Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A, 99*(18), 11760-11765. doi: 10.1073/pnas.192285999

Johnson, K. D., Norton, J. E., & Bresnick, E. H. (2002). Requirements for utilization of CREB binding protein by hypersensitive site two of the beta-globin locus control region. *Nucleic Acids Res, 30*(7), 1522-1530.

Kang, S. H., Vieira, K., & Bungert, J. (2002). Combining chromatin immunoprecipitation and DNA footprinting: a novel method to analyze protein-DNA interactions in vivo. *Nucleic Acids Res, 30*(10), e44.

Kawamoto, H., & Katsura, Y. (2009). A new paradigm for hematopoietic cell lineages: revision of the classical concept of the myeloid-lymphoid dichotomy. *Trends Immunol, 30*(5), 193-200. doi: 10.1016/j.it.2009.03.001

Kiefer, J. C. (2007). Back to basics: Sox genes. *Dev Dyn, 236*(8), 2356-2366. doi: 10.1002/dvdy.21218

Kiel, M. J., Yilmaz, O. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., & Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell, 121*(7), 1109-1121. doi: 10.1016/j.cell.2005.05.026

Kim, A., Zhao, H., Ifrim, I., & Dean, A. (2007). Beta-globin intergenic transcription and histone acetylation dependent on an enhancer. *Mol Cell Biol, 27*(8), 2980-2986. doi: 10.1128/MCB.02337-06

Kliewer, S. A., Umesono, K., Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., & Evans, R. M. (1992). Retinoid X receptor-COUP-TF interactions modulate retinoic acid signaling. *Proc Natl Acad Sci U S A, 89*(4), 1448-1452.

Lacombe, C., & Mayeux, P. (1999a). Erythropoietin (Epo) receptor and Epo mimetics. *Adv Nephrol Necker Hosp, 29*, 177-189.

Lacombe, C., & Mayeux, P. (1999b). The molecular biology of erythropoietin. *Nephrol Dial Transplant, 14 Suppl 2*, 22-28.

Lacombe, F., Lacoste, L., Vial, J. P., Briais, A., Reiffers, J., Boisseau, M. R., & Bernard, P. (1999). Automated reticulocyte counting and immature reticulocyte fraction measurement. Comparison of ABX PENTRA 120 Retic, Sysmex R-2000, flow cytometry, and manual counts. *Am J Clin Pathol, 112*(5), 677-686.

Lefebvre, V., Behringer, R. R., & de Crombrugghe, B. (2001). L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthritis Cartilage, 9 Suppl A*, S69-75.

Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., & Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol, 39*(12), 2195-2214. doi: 10.1016/j.biocel.2007.05.019

Lefebvre, V., & Smits, P. (2005). Transcriptional control of chondrocyte fate and differentiation. *Birth Defects Res C Embryo Today, 75*(3), 200-212. doi: 10.1002/bdrc.20048

Ley, T. J., DeSimone, J., Anagnou, N. P., Keller, G. H., Humphries, R. K., Turner, P. H., . . . Nienhuis, A. W. (1982). 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. *N Engl J Med, 307*(24), 1469-1475. doi: 10.1056/NEJM198212093072401

Liberati, C., Cera, M. R., Secco, P., Santoro, C., Mantovani, R., Ottolenghi, S., & Ronchi, A. (2001). Cooperation and competition between the binding of COUP-TFII and NF-Y on human epsilon- and gamma-globin gene promoters. *J Biol Chem, 276*(45), 41700-41709. doi: 10.1074/jbc.M102987200

Liberati, C., Sgarra, R., Manfioletti, G., & Mantovani, R. (1998). DNA binding of NF-Y: the effect of HMGI proteins depends upon the CCAAT box. *FEBS Lett, 433*(1-2), 174-178.

Liebhaber, S. A., Cash, F. E., & Ballas, S. K. (1986). Human alpha-globin gene expression. The dominant role of the alpha 2-locus in mRNA and protein synthesis. *J Biol Chem, 261*(32), 15327-15333.

Linberg, R., Conover, C. D., Shum, K. L., & Shorr, R. G. (1998a). Hemoglobin based oxygen carriers: how much methemoglobin is too much? *Artif Cells Blood Substit Immobil Biotechnol, 26*(2), 133-148.

Linberg, R., Conover, C. D., Shum, K. L., & Shorr, R. G. (1998b). Increased tissue oxygenation and enhanced radiation sensitivity of solid tumors in rodents following polyethylene glycol conjugated bovine hemoglobin administration. *In Vivo, 12*(2), 167-173.

Liu, J. J., Hou, S. C., & Shen, C. K. (2003). Erythroid gene suppression by NF-kappa B. *J Biol Chem, 278*(21), 19534-19540. doi: 10.1074/jbc.M212278200

Liu, L. R., Du, Z. W., Zhao, H. L., Liu, X. L., Huang, X. D., Shen, J., . . . Zhang, J. W. (2005). T to C substitution at -175 or -173 of the gamma-globin promoter affects GATA-1 and Oct-1 binding in vitro differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. *J Biol Chem, 280*(9), 7452-7459. doi: 10.1074/jbc.M411407200

Lowry, J. A., & Mackay, J. P. (2006). GATA-1: one protein, many partners. *Int J Biochem Cell Biol, 38*(1), 6-11. doi: 10.1016/j.biocel.2005.06.017

Mantovani, R., Malgaretti, N., Giglioni, B., Comi, P., Cappellini, N., Nicolis, S., & Ottolenghi, S. (1987). A protein factor binding to an octamer motif in the gamma-globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells. *Nucleic Acids Res, 15*(22), 9349-9364.

Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., . . . Ottolenghi, S. (1988). An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. *Nucleic Acids Res, 16*(10), 4299-4313.

Martin, D. I., Tsai, S. F., & Orkin, S. H. (1989). Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature, 338*(6214), 435-438. doi: 10.1038/338435a0

Matsuzaki, Y., Kinjo, K., Mulligan, R. C., & Okano, H. (2004). Unexpectedly efficient homing capacity of purified murine hematopoietic stem cells. *Immunity, 20*(1), 87-93.

McGrath, K. E., Bushnell, T. P., & Palis, J. (2008). Multispectral imaging of hematopoietic cells: where flow meets morphology. *J Immunol Methods, 336*(2), 91-97. doi: 10.1016/j.jim.2008.04.012

Mikkola, H. K., Klintman, J., Yang, H., Hock, H., Schlaeger, T. M., Fujiwara, Y., & Orkin, S. H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature, 421*(6922), 547-551. doi: 10.1038/nature01345

Mikkola, H. K., & Orkin, S. H. (2006). The journey of developing hematopoietic stem cells. *Development, 133*(19), 3733-3744. doi: 10.1242/dev.02568

Miller, I. J., & Bieker, J. J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol, 13*(5), 2776-2786.

Moleirinho, A., Seixas, S., Lopes, A. M., Bento, C., Prata, M. J., & Amorim, A. (2013). Evolutionary constraints in the beta-

globin cluster: the signature of purifying selection at the delta-globin (HBD) locus and its role in developmental gene regulation. *Genome Biol Evol, 5*(3), 559-571. doi: 10.1093/gbe/evt029

Molete, J. M., Petrykowska, H., Sigg, M., Miller, W., & Hardison, R. (2002). Functional and binding studies of HS3.2 of the beta-globin locus control region. *Gene, 283*(1-2), 185-197.

Morrison, S. J., Uchida, N., & Weissman, I. L. (1995). The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol, 11*, 35-71. doi: 10.1146/annurev.cb.11.110195.000343

Noordermeer, D., & de Laat, W. (2008). Joining the loops: beta-globin gene regulation. *IUBMB Life, 60*(12), 824-833. doi: 10.1002/iub.129

Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., . . . Ebert, B. L. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell, 144*(2), 296-309. doi: 10.1016/j.cell.2011.01.004

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., & Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature, 375*(6529), 316-318. doi: 10.1038/375316a0

Nuinoon, M., Makarasara, W., Mushiroda, T., Setianingsih, I., Wahidiyat, P. A., Sripichai, O., . . . Fucharoen, S. (2010). A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E. *Hum Genet, 127*(3), 303-314. doi: 10.1007/s00439-009-0770-2

O'Neill, D., Kaysen, J., Donovan-Peluso, M., Castle, M., & Bank, A. (1990). Protein-DNA interactions upstream from the human A gamma globin gene. *Nucleic Acids Res, 18*(8), 1977-1982.

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y., & Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood, 80*(12), 3044-3050.

Okpala, I. (2004). The intriguing contribution of white blood cells to sickle cell disease - a red cell disorder. *Blood Rev, 18*(1), 65-73.

Okpala, I., Ugochukwu, C. C., Tawil, A., & Greaves, R. (2004). Steady-state plasma lactoferrin levels in relation to infections and complications of sickle cell disease. *Hematol J, 5*(4), 371-372. doi: 10.1038/sj.thj.6200372

Osawa, M., Hanada, K., Hamada, H., & Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science, 273*(5272), 242-245.

Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. *Front Physiol, 5*, 3. doi: 10.3389/fphys.2014.00003

Palis, J., & Segel, G. B. (1998). Developmental biology of erythropoiesis. *Blood Rev, 12*(2), 106-114.

Palstra, R. J., de Laat, W., & Grosveld, F. (2008). Beta-globin regulation and long-range interactions. *Adv Genet, 61*, 107-142. doi: 10.1016/S0065-2660(07)00004-1

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet, 35*(2), 190-194. doi: 10.1038/ng1244

Pan, X., Ohneda, O., Ohneda, K., Lindeboom, F., Iwata, F., Shimizu, R., . . . Yamamoto, M. (2005). Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *J Biol Chem, 280*(23), 22385-22394. doi: 10.1074/jbc.M500081200

Park, I. K., He, Y., Lin, F., Laerum, O. D., Tian, Q., Bumgarner, R., . . . Li, L. (2002). Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood, 99*(2), 488-498.

Pauling, L., Itano, H. A., & et al. (1949). Sickle cell anemia a molecular disease. *Science, 110*(2865), 543-548.

Perkins, A. C., Sharpe, A. H., & Orkin, S. H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature, 375*(6529), 318-322. doi: 10.1038/375318a0

Perutz, M. F. (1978). Hemoglobin structure and respiratory transport. *Sci Am, 239*(6), 92-125.

Pevny, L., Lin, C. S., D'Agati, V., Simon, M. C., Orkin, S. H., & Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. *Development, 121*(1), 163-172.

Platt, O. S. (1994). Easing the suffering caused by sickle cell disease. *N Engl J Med, 330*(11), 783-784. doi: 10.1056/NEJM199403173301110

Platt, O. S. (2008). Hydroxyurea for the treatment of sickle cell anemia. *N Engl J Med, 358*(13), 1362-1369. doi: 10.1056/NEJMct0708272

Platt, O. S., Brambilla, D. J., Rosse, W. F., Milner, P. F., Castro, O., Steinberg, M. H., & Klug, P. P. (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med, 330*(23), 1639-1644. doi: 10.1056/NEJM199406093302303

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F. W., & Orkin, S. H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell, 86*(1), 47-57.

Pronk, C. J., Rossi, D. J., Mansson, R., Attema, J. L., Norddahl, G. L., Chan, C. K., . . . Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell, 1*(4), 428-442. doi: 10.1016/j.stem.2007.07.005

Proudfoot, N. J., & Longley, J. I. (1976). The 3' terminal sequences of human alpha and beta globin messenger RNAs: comparison with rabbit globin messenger RNA. *Cell, 9*(4 PT 2), 733-746.

Rabbitts, T. H. (1976). Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA. *Nature, 260*(5548), 221-225.

Rochette, J., Craig, J. E., & Thein, S. L. (1994). Fetal hemoglobin levels in adults. *Blood Rev, 8*(4), 213-224.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., . . . Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J, 24*(13), 2354-2366. doi: 10.1038/sj.emboj.7600702

Romeo, P. H., Prandini, M. H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., . . . Uzan, G. (1990). Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature, 344*(6265), 447-449. doi: 10.1038/344447a0

Ronchi, A., Bellorini, M., Mongelli, N., & Mantovani, R. (1995). CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the

minor groove and distorts DNA. *Nucleic Acids Res, 23*(22), 4565-4572.

Rougeon, F., Kourilsky, P., & Mach, B. (1975). Insertion of a rabbit beta-globin gene sequence into an E. coli plasmid. *Nucleic Acids Res, 2*(12), 2365-2378.

Roy, P., Bhattacharya, G., Lahiri, A., Dasgupta, U. B., Banerjee, D., Chandra, S., & Das, M. (2012). hsa-miR-503 is downregulated in beta thalassemia major. *Acta Haematol, 128*(3), 187-189. doi: 10.1159/000339492

Rund, D., & Rachmilewitz, E. (2005). Beta-thalassemia. *N Engl J Med, 353*(11), 1135-1146. doi: 10.1056/NEJMra050436

Sankaran, V. G., Menne, T. F., Xu, J., Akie, T. E., Lettre, G., Van Handel, B., . . . Orkin, S. H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science, 322*(5909), 1839-1842. doi: 10.1126/science.1165409

Sankaran, V. G., & Nathan, D. G. (2010). Reversing the hemoglobin switch. *N Engl J Med, 363*(23), 2258-2260. doi: 10.1056/NEJMcibr1010767

Sankaran, V. G., Orkin, S. H., & Walkley, C. R. (2008). Rb intrinsically promotes erythropoiesis by coupling cell cycle exit with mitochondrial biogenesis. *Genes Dev, 22*(4), 463-475. doi: 10.1101/gad.1627208

Sankaran, V. G., Xu, J., & Orkin, S. H. (2010). Transcriptional silencing of fetal hemoglobin by BCL11A. *Ann N Y Acad Sci, 1202*, 64-68. doi: 10.1111/j.1749-6632.2010.05574.x

Satta, S., Perseu, L., Maccioni, L., Giagu, N., & Galanello, R. (2012). Delayed fetal hemoglobin switching in subjects with KLF1 gene mutation. *Blood Cells Mol Dis, 48*(1), 22-24. doi: 10.1016/j.bcmd.2011.10.003

Satta, S., Perseu, L., Moi, P., Asunis, I., Cabriolu, A., Maccioni, L., . . . Galanello, R. (2011). Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica, 96*(5), 767-770. doi: 10.3324/haematol.2010.037333

Schroeder, T. M., & Kurth, R. (1971). Spontaneous chromosomal breakage and high incidence of leukemia in inherited disease. *Blood, 37*(1), 96-112.

Schuh, A. H., Tipping, A. J., Clark, A. J., Hamlett, I., Guyot, B., Iborra, F. J., . . . Porcher, C. (2005). ETO-2 associates with SCL in

erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol, 25*(23), 10235-10250. doi: 10.1128/MCB.25.23.10235-10250.2005

Sheffield, N. C., Thurman, R. E., Song, L., Safi, A., Stamatoyannopoulos, J. A., Lenhard, B., . . . Furey, T. S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res, 23*(5), 777-788. doi: 10.1101/gr.152140.112

Shivdasani, R. A., Mayer, E. L., & Orkin, S. H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature, 373*(6513), 432-434. doi: 10.1038/373432a0

Siatecka, M., & Bieker, J. J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood, 118*(8), 2044-2054. doi: 10.1182/blood-2011-03-331371

Smits, P., Dy, P., Mitra, S., & Lefebvre, V. (2004). Sox5 and Sox6 are needed to develop and maintain source, columnar, and hypertrophic chondrocytes in the cartilage growth plate. *J Cell Biol, 164*(5), 747-758. doi: 10.1083/jcb.200312045

Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J. M., Behringer, R. R., . . . Lefebvre, V. (2001). The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev Cell, 1*(2), 277-290.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J. C., Thongjuea, S., Stadhouders, R., . . . Grosveld, F. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev, 24*(3), 277-289. doi: 10.1101/gad.551810

Song, S. H., Kim, A., Ragoczy, T., Bender, M. A., Groudine, M., & Dean, A. (2010). Multiple functions of Ldb1 required for beta-globin activation during erythroid differentiation. *Blood, 116*(13), 2356-2364. doi: 10.1182/blood-2010-03-272252

Stamatoyannopoulos, G., Blau, C. A., Nakamoto, B., Josephson, B., Li, Q., Liakopoulou, E., . . . Dover, G. (1994). Fetal hemoglobin induction by acetate, a product of butyrate catabolism. *Blood, 84*(9), 3198-3204.

Sunshine, H. R., Hofrichter, J., & Eaton, W. A. (1978). Requirement for therapeutic inhibition of sickle haemoglobin gelation. *Nature, 275*(5677), 238-240.

Suzuki, N., Ohneda, O., Takahashi, S., Higuchi, M., Mukai, H. Y., Nakahata, T., . . . Yamamoto, M. (2002). Erythroid-specific expression of the erythropoietin receptor rescued its null mutant mice from lethality. *Blood, 100*(7), 2279-2288. doi: 10.1182/blood-2002-01-0124

Taher, A. T., Musallam, K. M., El-Beshlawy, A., Karimi, M., Daar, S., Belhoul, K., . . . Cappellini, M. D. (2010). Age-related complications in treatment-naive patients with thalassaemia intermedia. *Br J Haematol, 150*(4), 486-489. doi: 10.1111/j.1365-2141.2010.08220.x

Talbot, D., & Grosveld, F. (1991). The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J, 10*(6), 1391-1398.

Tang, L. S., Alger, H. M., & Pereira, F. A. (2006). COUP-TFI controls Notch regulation of hair cell and support cell differentiation. *Development, 133*(18), 3683-3693. doi: 10.1242/dev.02536

Tripic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C. R., Gregory, G. D., . . . Blobel, G. A. (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood, 113*(10), 2191-2201. doi: 10.1182/blood-2008-07-169417

Tsang, A. P., Fujiwara, Y., Hom, D. B., & Orkin, S. H. (1998). Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev, 12*(8), 1176-1188.

Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., . . . Orkin, S. H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell, 90*(1), 109-119.

Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V. G., Chen, W., . . . Cao, A. (2008). Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc*

Natl Acad Sci U S A, 105(5), 1620-1625. doi: 10.1073/pnas.0711566105

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., . . . Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell, 17*(3), 453-462. doi: 10.1016/j.molcel.2004.12.028

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A., & Rabbitts, T. H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J, 16*(11), 3145-3157. doi: 10.1093/emboj/16.11.3145

Weatherall, D. J., Williams, T. N., Allen, S. J., & O'Donnell, A. (2010). The population genetics and dynamics of the thalassemias. *Hematol Oncol Clin North Am, 24*(6), 1021-1031. doi: 10.1016/j.hoc.2010.08.010

Weiss, M. J., Keller, G., & Orkin, S. H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev, 8*(10), 1184-1197.

Wheeler, J. T., & Krevans, J. R. (1961). The homozygous state of persistent fetal hemoglobin and the interaction of persistent fetal hemoglobin with thalassemia. *Bull Johns Hopkins Hosp, 109*, 217-233.

Whyatt, D., Lindeboom, F., Karis, A., Ferreira, R., Milot, E., Hendriks, R., . . . Philipsen, S. (2000). An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature, 406*(6795), 519-524. doi: 10.1038/35020086

Wilber, A., Hargrove, P. W., Kim, Y. S., Riberdy, J. M., Sankaran, V. G., Papanikolaou, E., . . . Persons, D. A. (2011). Therapeutic levels of fetal hemoglobin in erythroid progeny of beta-thalassemic CD34+ cells after lentiviral vector-mediated gene transfer. *Blood, 117*(10), 2817-2826. doi: 10.1182/blood-2010-08-300723

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., . . . Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal

during homeostasis and repair. *Cell, 135*(6), 1118-1129. doi: 10.1016/j.cell.2008.10.048

Wilson, J. G., & Tavassoli, M. (1994). Microenvironmental factors involved in the establishment of erythropoiesis in bone marrow. *Ann N Y Acad Sci, 718*, 271-283; discussion 283-274.

Wong, P. M., Chung, S. W., Reicheld, S. M., & Chui, D. H. (1986). Hemoglobin switching during murine embryonic development: evidence for two populations of embryonic erythropoietic progenitor cells. *Blood, 67*(3), 716-721.

Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W., & Orkin, S. H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev, 24*(8), 783-798. doi: 10.1101/gad.1897310

Xu, Z., Huang, S., Chang, L. S., Agulnick, A. D., & Brandt, S. J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. *Mol Cell Biol, 23*(21), 7585-7599.

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P. D., Fuchs, D. A., Erickson, D. T., . . . Brilliant, M. H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet, 2*(2), e14. doi: 10.1371/journal.pgen.0020014

Zhou, D., Pawlik, K. M., Ren, J., Sun, C. W., & Townes, T. M. (2006). Differential binding of erythroid Krupple-like factor to embryonic/fetal globin gene promoters during development. *J Biol Chem, 281*(23), 16052-16057. doi: 10.1074/jbc.M601182200

Zhu, X., Wang, Y., Pi, W., Liu, H., Wickrema, A., & Tuan, D. (2012). NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human gamma-globin gene. *PLoS One, 7*(10), e47175. doi: 10.1371/journal.pone.0047175

Chapter 2

Identification of Sox6 interacting complexes

INTRODUCTION

HMG proteins including Sox proteins bind to the minor groove on DNA, producing a bending of about 75º, that in turn, introduces local conformational changes. Sox proteins bind to DNA with low sequence specificity and also with a lower affinity than most other TFs ***(Kamachi, Uchikawa, & Kondoh, 2000; Wilson & Koopman, 2002)***. Sox proteins beside of binding to DNA also are able to interact with other proteins, through their HMG box domain, thus increasing their DNA binding affinity. Sox proteins can thus act as "architectural proteins" by promoting the assembly of biologically active multiprotein complexes. Sox6 has the ability to activate and repress gene expression, depending on its interactions and on its target sequences **(Kamachi et al., 2000; Kiefer, 2007; Lefebvre, Dumitriu, Penzo-Mendez, Han, & Pallavi, 2007)**. Despite its dual function of activator and repressor, Sox6 lacks any conventional transactivation or transrepression domain.

This suggests that Sox6 needs to interact with other proteins to exert its function. The aim of this study is to identify the Sox6 interacting multiprotein complexes in erythropoietic cells. To this aim we set up a proteomic Tag-affinity screening by using a metabolic biotin tagging approach **(E. de Boer et al., 2003; Rodriguez et al., 2005)** in mouse (MEL) and Human (HEL) erythroleukemic cells, stably expressing BirA (biotin ligase) alone (in basal conditions) and under Sox6 overexpression.

Biotinylation (Biotin-Streptavidin binding) is the strongest noncovalent binding that exists in nature ($K_d=10^{-15}$ M). In tag-affinity protein purification approaches, the biotin–streptavidin binding offers several advantages: its high affinity allows to purify proteins under high stringent conditions, highly reducing the possible background. Moreover, only very few proteins are naturally biotinylated, and this reduces the chances to immuoprecipitate proteins not specifically interacting with the protein of interest.

In our study, we adopted the single-step streptavidin affinity purification approach followed by Mass Spectrometry sequencing. In order to perform this study, we used nuclear extracts from human (HEL) and mouse (MEL) erythroleukemic cells overexpressing the biotin-tagged Sox6 cDNA, to the end of using the single-step streptavidin protein purification, followed by SDS/PAGE fractionation, liquid chromatography and tandem Mass Spectrometry (MS/MS) peptide sequencing. Finally, we identify several Sox6 peptides and a long list of Sox6 interacting proteins.

EXPERIMENTAL PROCEDURES

Constructs:

The construct expressing the *Escherichia coli* BirA biotin-protein ligase (containing 3' HA epitope) was kindly given by Prof. John Strouboulis **(de Boer et al., 2003)**. The Sox6 murine cDNA (obtained from Prof. Michiko Hamada-Kanazawa, Japan) was cloned in frame with a 3' FLAG epitope to generate a Sox6-FLAG cassette. 3' to the FLAG epitope, a sequence coding the 23 amino acids necessary for the biotin-tag was cloned to generate the bioSox6-FLAG cassette, flanked by two BglII sites. This cassette was then cloned immediately upstream to the IRES - Emerald GFP cassette (blunted BamHI site) of the CSIemerald derived from pHR SIN CSGW, **(Demaison et al., 2002)** lentiviral vector, a kind gift from Prof. Tariq Enver, UCL, London. In this vector, the SFFV promoter drives the expression of the exogenous cDNA, highly active in hematopoietic cells. The bioSox6-FLAG cassette flanked by EcoRI and XhoI sites was cloned into SparQ Dual Promoter under the Cumate switch promoter (System biosciences - QM511B-1). Cumate repressor (CymR) cDNA was cloned under EF1 promoter and containing Neomicin resistant cassette (System biosciences - QM400PA/VA-2, pCDH-EF1-CymR-T2A-Neo). A Kozak consensus followed by the ATG starting codon and then the FLAG epitope were cloned in frame to the 5' CymR cDNA (FLAG-CymR). The two packaging plasmids psPAX2 and pMD-VSVG were used to produce Lentiviral pseudo-particles in HEK 293T cells (www.lentiweb.com).

Lentiviral vector production:

Viral stocks were produces by transient cotransfection of the three different vectors in the Human Embryonic Kidney (HEK) 293T packaging cell line. 72 hours after transfection, the supernatant containing the recombinant viruses was collected, filtered (0.45n, 2 and centrifuged at 20,000g for 8 hours at 4ºC. The viral pellet was resuspended in PBS 1x and stored at -80ºC. Viral titers were determined by transduction of HEL cells with serial dilutions of the vector stock and by scoring the GFP transgene expression by fluorescence-activated cell sorter (FACS) analysis.

Cell Cultures and transduction:

Human (HEL) and mouse (MEL) erytholeukemic cell lines, stably expressing the biotin ligase BirA were cultured in RPMI 1640 medium (Lonza) supplemented with 10% heat inactivated fetal bovine serum (Lonza), L-glutamine (Euroclone), antibiotics Penicillin-Streptomycin 100U/100ug/ml (Euroclone) in a humidified 5% CO_2 atmosphere at 37°C. In the cells stably expressing BirA or CymR, the medium was supplied with 3µg/ml of puromycin (Sigma P8833), and 500µg/ml of Neomycin (G418 – Roche), respectively. Transduction was performed overnight with a multiplicity of infection (MOI) of 30.

Nuclear extracts:

Cells were harvested and centrifuged at 640g for 15 min at 4ºC . The pellet washed three times in ice-cold 1x PBS (phosphate buffered saline, pH 7.4) and gently resuspended in ice-cold NP-40 lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.5% v/v NP-40, supplemented with protease inhibitors (complete EDTA-free, Roche) added just before use). Cells were lysed by incubation on a rotating wheel for 10 min at 4ºC and nuclei were then centrifuged at 3300g. Nuclei were resuspended in nuclear lysis buffer (10mM HEPES-KOH pH 7.9, 100mM KCl, 3mM $MgCl_2$, 0,1 mM EDTA, 20% glycerol and protease inhibitors, as above). Nuclear proteins were extracted by drop-wise addition of 4M KCl with gentle agitation on ice, until the final salt concentration was approx. 350-400mM. Nuclear lysis and protein extraction were then allowed to proceed by incubation on a rotating wheel for 40 min at 4ºC, followed by centrifugation. The supernatant, which correspond to the soluble nuclear extract fraction, was retained.

Immunoblotting analysis:

To confirm for the presence of Sox6 overexpression and *in vivo* biotinylation, nuclear extracts (30-50µg/lane) were resolved by SDS/PAGE in a 7% gel and blotted onto Hybond ECL Nitrocellulose membrane (GE healthcare life science - Amersham) by "wet blotting." carried out under constant voltage at 100V for 90-120 min at 4ºC (Transblot apparatus, Biorad). Membranes were blocked for 1-2h at room temperature with Milk 5% in TTBS

(1x TBS (Tris buffered saline, pH 7.6) 0,1% Tween 20 (Sigma)) and incubated with the appropriate primary antibody diluted in Milk 3% TTBS overnight at 4ºC. Membranes were washed for three times with TTBS and incubated with the secondary antibody (diluted in Milk 3%). In particular, the Streptavidin-HRP (Cell signaling) antibody was incubated for 2h at room temperature (no secondary antibody was required). Antibodies binding was detected by using appropriate horseradish peroxidase-conjugated IgG and revealed by ECL (Millipore).

Benzonase treatment:

Nuclear extracts were diluted by adding 3 volumes of HENG buffer (0 mM KCl, 10 mM HEPES pH=9, 1.5 mM MgCl2, 0.25 mM EDTA, 20 % glycerol, PMSF or PI) and 25 U of Benzonase (Viscolase nuclease (A&A Biotechnology)) per mg of nuclear extract were added. The concentration of Mg^{+2} was adjusted to 2mM. Diluted lysates were incubated with Benzonase for 2-4 hours at 4ºC on a rotating wheel. DNA digestion and removal was checked on gel on a phenol-extracted aliquot.

Binding to Streptavidin beads:

Paramagnetic streptavidin beads (Dynabeads M-280, Dynal (Life Technologies) 50 µl per 1 mg of protein), were washed three times with PBS at room temperature and blocked with HENG/BSA (Sigma-Aldrich) 200µg/ml (10 mM HEPES pH 9, 1.5 mM MgCl2, 0.25 mM EDTA, 20 % glycerol, PMSF) for 1h at room temperature under rotation. The binding was carried out with Nuclear Extracts

diluted in 150 mM KCl and 0.3 % NP40), overnight on a rotating wheel. After 5 washes with wash HENG solution (HENG buffer with 300mM KCl) and 2 washes with PBS 1x at room temperature, the bound material was eluted by boiling for 10 min in Laemmli protein simple loading buffer (62.5mM Tris–HCl pH6.8, 25% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.1% bromophenol blue) and analyzed by immunoblotting.

Mass-Spectrometry:

Proteins eluted from streptavidin beads were resolved by SDS-PAGE, and gel lanes were cut into slices by using an automatic gel slicer and subjected to in-gel trypsinization, ***(Shevchenko, Wilm, Vorm, & Mann, 1996)***. Bound proteins were treated with trypsin on the beads after resuspending in 50 mM ammonium bicarbonate and adding trypsin (sequencing grade; Promega) to approximately 60 ng/mg of total protein, followed by overnight incubation at 37°C ***(Rybak et al., 2005)***. The supernatant containing the trypsin-treated peptides was then recovered by magnetically removing the beads. Peptides released by in-gel or on-bead trypsinization were analyzed by nano-LC-MS/MS performed on either a CapLC system (Waters, Manchester, United Kingdom) coupled to a Q-ToF Ultima mass spectrometer (Waters), operating in positive mode and equipped with a Z-spray source, or on a 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap or LTQ-FT-MS mass spectrometer (both from Thermo Scientific) operating in positive mode and equipped with a nanospray source. Peptides were trapped and separated on a Jupiter C_{18} reversed-phase column (Phenomenex)

using a linear gradient from 0 to 80% medium B (where medium A = 0.1 M acetic acid and medium B = 80% [vol/vol] acetonitrile, 0.1 M acetic acid) using a splitter. The column eluate was directly sprayed into the electrospray ionization source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode.

Data analysis and protein identification:

Peak lists were automatically created from raw data files by using ProteinLynx Global Server software (version 2.0; Waters, Manchester, United Kingdom) for Q-ToF spectra and Mascot Distiller software (version 2.0; MatrixScience, London, United Kingdom) for LTQ-Orbitrap and LTQ-FT-MS spectra. The Mascot search algorithm was used for searching the National Center for Biotechnology Information (NCBI) database. The Mascot score cutoff value for a positive hit was set to 65. Individual peptide MS/MS spectra with Mowse scores below 40 were checked manually and either interpreted as valid identifications or discarded. Identified proteins listed as NCBI database entries were screened to identify proteins that were also identified in mass spectrometry experiments from control BirA-expressing cells ***(E. de Boer et al., 2003)***. These were removed as background binding proteins. The remaining proteins were classified according to gene ontology criteria using the Panther software (www.pantherdb.org) and were then grouped according to their molecular function.

RESULTS:

Biotinylation tagging and Sox6-complexes purification

The biotinylation tagging approach is based on the use of artificial short peptide tags that are specifically and efficiently biotinylated by the bacterial BirA biotin ligase, an enzyme responsible for the covalent attachment of biotin to metabolic enzymes ***(Schatz, 1993).*** A number of these short sequence tags can be very efficiently biotinylated *in vitro* ***(Beckett, Kovaleva, & Schatz, 1999)***. In order to perform purification of the multiprotein complexes interacting with Sox6, we set up the single-step purification approach based on Biotin-Streptavidin binding. In our strategy: i) We overexpressed the biotin-tagged Sox6-FLAG in the appropriated erythroleukemic cell line. ii) We isolated nuclear protein extracts by a single-step purification using Streptavidin coated beads. iii) The purified peptides were sequenced by MS.

As final result, we obtained a list of putative candidates interactors of Sox6 (Figure 2.1).

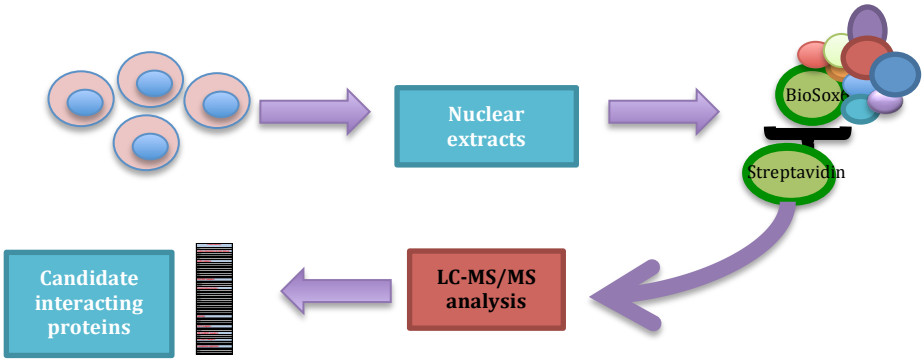


Figure 2.1. Biotinylation strategy. *Single step streptavidin affinity purification followed by tandem Mass Spectrometry peptide sequencing. Schematic diagram of the experimental approach used.*

Generation of a cell line expressing a Biotin-tagged-Sox6 protein

In order to set up the metabolic biotin tagging approach to identify and characterize the multiprotein complexes containing Sox6, we used short-tag peptide sequence (23 amino acids) fused to the C-terminal domain of Sox6, in frame with the FLAG epitope (bioSox6-FLAG)(Figure 2.2). The biotin-tag is efficiently biotinylated by the bacterial biotin ligase *BirA*, which is stably coexpressed in both Mouse (MEL) and Human (HEL) erythroleukemic cell lines. MEL cells stably expressing BirA (Figure 2.3a) were a gift of Prof. John Strouboulis, whereas the HEL cell line stably expressing BirA was created in our laboratory. To this end, we transiently transfected a BirA-HA tagged expression vector and we selected single expressing clones by using puromycin. The BirA expressing clones were confirmed and selected on the basis of their level of expression by using as a reference the BirA protein level expressed by the already established MEL-BirA cell line (Figure 2.4). We chose three different clones (Clone 2, 4 and 12) to proceed to the biotinylation analysis. To select one of those three clones we transduced these cells with the lentivirus expressing bioSox6-FLAG and we analyzed the level of biotinylated Sox6 protein by using the streptavidin-horseradish peroxidase (HRP) conjugate. We observed robust biotinylation of the tagged Sox6 in all the three selected clones (Figure 2.4). For the following experiments we chose clone 12 that gave the highest level of Sox6 biotinylation.

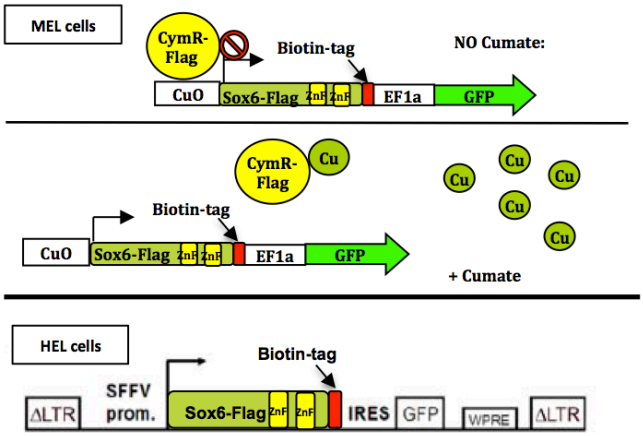


Figure 2.2 Schematic representation of the lentiviral vector expressing bioSox6-FLAG. *The two different systems used to overexpress bioSox6-FLAG in MEL cells (upper panel) and in HEL cells (lower panel). (Upper panel lentiviral vector used for the inducible system expressing bioSox6-FLAG under the cumate operator promoter (CuO). In the absence of cumate the bioSox6-FLAG gene expression is repressed. Upon the addition of cumate into the cell culture medium the promoter is de-repressed and bioSox6-FLAG can start to be expressed. (lower panel) lentivector expressing bioSox6-FLAG under the SFFV (Spleen focus forming virus) promoter.*

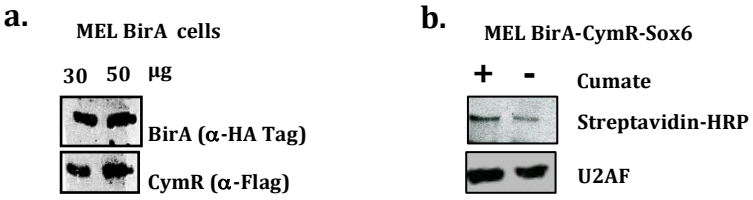


Figure 2.3. The mouse erythroleukemic (MEL) cell line expressesing the enzyme biotin ligase (BirA) and properly biotinylating*. a) Western Blot (WB) of mouse erytholeukemic (MEL) cells nuclear extracts by using the HA-antibody to check for the experssion of BirA, and by using the Flag-antibody for the expression of the cumate repressor (CymR). Different amounts of protein were loaded: first line 30 µg; second line 50 µg of nuclear extracts b) WB for the biotinylated Sox6 nuclear extracts using the cumate inducible system. Nuclear extracts from MEL-BirA-CymR cells transduced with bioSox6-FLAG upon (+/-) addition of cumate in the cell cultures. U2AF was used as loading control.*

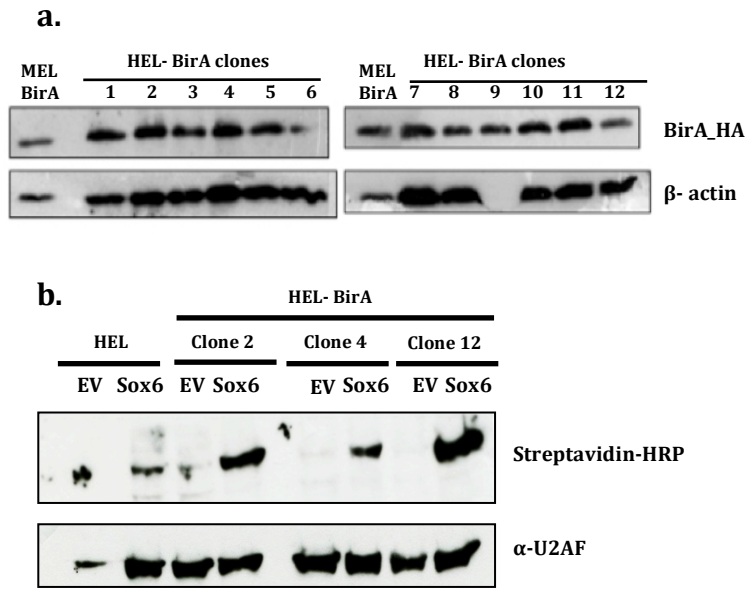


Figure 2.4. Human erythroleukemic (HEL) cell line clones expressing the enzyme BirA and properly biotinylating Sox6 (bioSox6-FLAG). *a) WB of HEL-BirA nuclear extracts by using the HA-antibody to detect the expression of BirA in twelve different clones. MEL BirA extracts were used as positive control. Anti-β-actin was used as loading control. b) WB for the biotinylated Sox6 protein in HEL nuclear extracts by using Streptavidin-HRP conjugated, U2AF was used as loading control.*

The overexpression of Sox6 in erythroid cells is known to induce cell cycle withdrawal, differentiation and terminal maturation with cells dying within few days ***(Cantu et al., 2011)***.

In order to overcome this problem, that is an obstacle to the preparation of the large amount of cells required for Sox6 complexes purification, we designed two different strategies. The first one was the use of a Cumate-base inducible system. This inducible system is based on the repression of the inducible promoter by a Cumate repressor (CymR). Upon the addition of Cumate into the cell culture medium, the repressor is removed from the promoter due to its binding by Cumate molecules,

leading the promoter free to activate gene expression. In order to establish this system we set up a MEL cell line, stably coexpressing the biotin ligase (BirA) and the cumate repressor (CymR) (Figure 2.2 (upper panel)). Cells were then transduced with the bioSox6-Flag lentivirus, expanded and treated with Cumate. The amount of the Sox6 biotinylated protein was assessed by using streptavidin-HRP to validate the system (Figure 2.3 b).

The second approach was based on HEL (Human erythroleukemic) cells (Figure 2.2 (lower panel)). HEL cells do not die as fast as MEL cells upon Sox6 overexpression, although they do terminally differentiate. In fact, HEL cells carry 8 copies of the Jak2 V617F mutation that render them insensitive to the Suppressor of cytokine signaling 3 (SOCS-3), a direct target of Sox6, responsible for the Sox6-mediated block of cell proliferation ***(Cantu et al., 2011).***

We found that the inducible system used to set up the overexpression of Sox6 in MEL cells had several problems. First, the amount of Sox6 obtained after the induction was not decent enough to accomplish a successful purification of the Sox6 partners by Mass Spectrometry. Moreover, the Mass Spectrometry results showed that there was not significant difference between the list obtained from MEL/BirA, and MEL/BirA-bioSox6-FLAG nuclear extracts. Thus, we decided to continue the experiments by using the overexpression of bioSox6-FLAG in HEL cells.

Pull down and identification of Sox6-containing multiprotein complexes

Large-scale nuclear extracts from HEL cells overexpressing biotinylated Sox6 (bioSox6-FLAG) were used to perform single step streptavidin affinity-tag purification. Co-purified nuclear proteins were then fractionated by using SDS/PAGE followed by liquid chromatography and analyzed by tandem Mass Spectrometry peptide sequencing (strategy described on Figure 2.1) (*Mass spectrometry was performed by Dr. Demmers at the Proteomics Center in Erasmus MC, Rotterdam*).

As a negative control, we also performed in parallel a similar pull down in cells only expressing BirA and a the lentiviral empty vector, to be considered as background for the Mass Spectrometry peptides sequencing.

Two biological replicates of the Mass Spectrometry sequencing analysis were performed for HEL cells overexpressing bioSox6-FLAG.

Filtering criteria

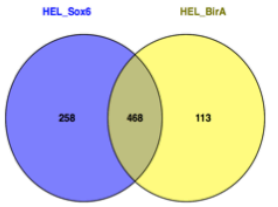
From the list of the potential Sox6 interacting candidates, we filtered out several proteins previously identified as a common biotinylated background proteins, such carboxylases and their interacting enzymes, factors involved in mRNA processing and ribosomal proteins **(E. de Boer et al., 2003).**

As a final output from the Mass Spectrometry analysis carried out in duplicate in HEL cells we obtained a list of 839 putative

interacting proteins. However, some of these proteins were also present in the negative control. In order to select for proteins specifically interacting with the bioSox6-FLAG, we used a software based on a Venn diagram (http://bioinfogp.cnb.csic.es/tools/venny/) to select the proteins specifically interacting with bioSox6-FLAG. We found that between those 839 proteins just 258 were only present in the bioSox6-FLAG pull down (Figure 2.5).

The proteins identified from the Venn diagram to be specifically pulled down by bioSox6-FLAG were then classified according to the molecular function and the biological process, as defined by the Gene Ontology Consortium (www.geneontology.org). Following GO analysis, proteins were classified using the Panther software (www.pantherdb.org). According to GO and Panther analyses, most of the Sox6 interacting proteins are relevant for chromatin binding (nucleic acid binding, chromatin binding and protein binding), catalytic activity (hydrolase, transferase and enzyme regulator activity) and also for structural functions (in particular proteins with a role in the cytoskeleton modulation)(Figure 2.5). Among them, we decided to exclude proteins with catalytic activity because they are common background (27,7% of the analyzed proteins, see figure 2.5b) ***(de Boer, Goemans, Ghezavat, van Ooijen, & Maes, 2003)***.

a.



b.

Molecular function (HEL cells)	Percentage (%)
transporter activity (GO:0005215)	5,3
translation regulator activity (GO: 0045182)	2,7
protein binding transcription factor activity (GO:0000988)	2,0
enzyme regulator activity (GO:0030234)	4,7
catalytic activity (GO:0003824)	27,7
receptor activity (GO:0004872)	2,0
nucleic acid binding transcription factor activity (GO:0001071)	7,0
antioxidant activity (GO:0016209)	0,3
structural molecule activity (GO: 0005198)	10,7
binding (GO:0005488)	42,0

c.

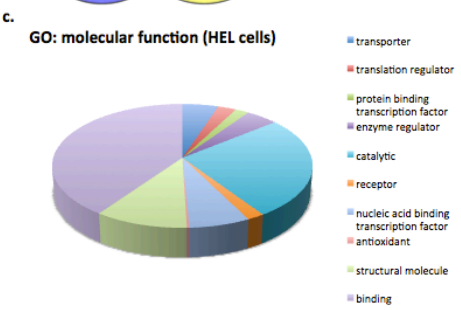


Figure 2.5. Filtering criteria for selecting Sox6 interacting proteins in HEL cells. *a) Venn diagram used to dicriminate the outcome from the Mass Spectrometry protein sequence analysis between overexpressing and non-overexpressing bioSox6-FLAG protein. b) Table corresponding to the Gene Ontology (GO) analysis of proteins identified in the Mass Spectrometry outcome as specific for bioSox6-FLAG fraction. c) Pie chart representing the different percentages in gene categories found in the Gene Ontology (GO) analysis.*

A positive selection of candidates was done on the basis of data from literature, pointing to a role of these proteins in hematopoiesis or erythropoiesis. A further criterion was the comparison of the retrieved list of peptides with data from a DNA microarray experiment performed previously in our laboratory on mouse fetal liver erythroid cells. In this experiment, mouse primary erythroid cells were purified at different developmental stages (E11.5, E12.5 and E13.5) and at different stages of differentiation (on the basis of the expression of cell surface specific antigens). cKit^{+}-TER119^{-} cells represent pluripotent hematopoietic progenitors, cKit^{+}-TER119^{+} cells are the erythroid

committed early progenitors, and cKit$^-$-TER119$^+$ are differentiated erythroblast and mature erythrocytes. This approach provided a database of global expression profile in erythroid freshly purified fetal liver cells during the mouse hemoglobin switching. During these developmental stages the fetal liver is the main erythropoietic organ within the embryo. Thus, the erythroid population is higher than 95% of the total fetal liver cells.

Sox6 candidates interactors after filtering

In the Mass Spectrometry readout, we identified several peptides of Sox6 present in cells overexpressing bioSox6-FLAG but absent in HEL/BirA cells, which proves that the pull down experiment was successfully working.

Numerous previously described proteins with a role in erythroid development were also identified as candidate of Sox6 interacting proteins: among them, i) known erythroid TFs, such us COUP-TFII, GATA1, GATA2 and Gfi1b. ii) chromatin remodeling factors belonging to the NuRD (Nucleosome remodeling and deacetylase) and to the LSD1/CoRest (lysine specific demethylase 1 and repressor element-1 TF corepressor 1) complexes.

However, novel corepressors or coactivators were identified as possible Sox6 interactors during erythroid development, BCOR (Bcl-6 co-repressor), FACT complex, SWI/SNF and PcG (Figure 2.6).

Furthermore, among these novel potential interactors, we found particularly interesting several cyclin dependent kinases (CDKs), such us CDK11, CDK12, CDK13 and CycK. These kinases are not fully related with cell cycle control, but they also have a role in splicing, genome stability and DNA damage response (DDR). However, nowadays only CDK13 has been found related with hematopoiesis, but few it is known about its specific function (Figure 2.6).

To directly confirm and validate the interaction between some of the novel Sox6 partners we performed some co-immunoprecipitation analyses/pull down by using streptavidin beads followed by immunoblotting detection in human erythroleukemic cells overexpressing and non-overexpressing Sox6 (Figure 2.6).

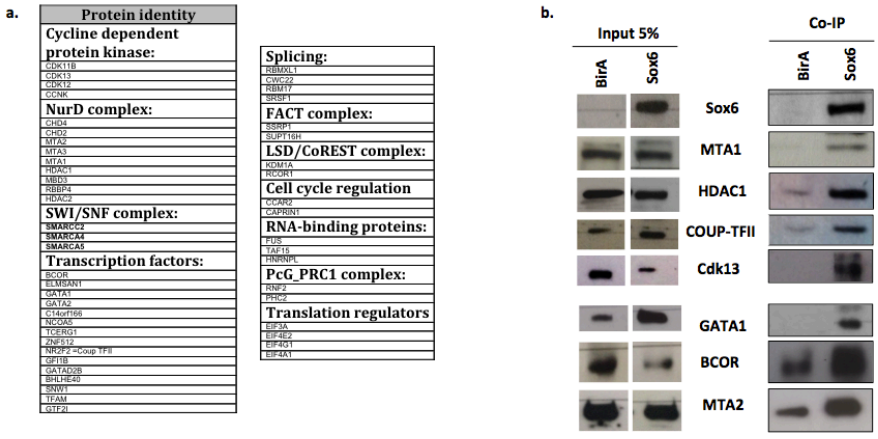


Figure 2.6. Identification of Sox6 interacting proteins by affinity tag purification followed by Mass Spectrometry sequencing*. a) The readout of Mass Spectrometry sequencing was filtered following different criteria to end with a list of the most interesting candidates. Chromatin remodeling factors (such us several subunits of the NuRD complex), transcriptions factors with a role in the globins genes regulation and cyclin dependent kinases (CDK) are found the most abundant and specific identified proteins. b) Co-precipitation of Sox6 interacting proteins by using streptavidin beads to confirm the results obtained by Mass Spectrometry. 5% of the total nuclear lysate was used as input.*

DISCUSSION

In general, tissue specific TFs are flexible factors with the ability to induce lineage specific differentiation, what involves not only transcriptional activation of lineage specific genes, but also the repression of other TFs. However it is still unclear how they could display both functions. Sox6 can act as an activator or as a repressor, although it lacks any transactivator or transrepressor domain. Hence we used the biotinylation approach to isolate and characterize Sox6 multiprotein complexes. This method was published to be useful, not only to successfully purify low-abundance TFs complexes, but also to better understand the molecular mechanisms implicated in the γ-globin gene silencing ***(Grosveld et al., 2005; Rodriguez et al., 2005)***. In fact, our final aim is to use biotin-streptavidin approach is to identify all the Sox6 multiprotein complexes that potentially regulate gene expression in the β-globin locus. The identification of these proteins might be of potential therapeutic interest for the treatment of β-thalassemia and Sickle Cell Disease (SCD), by prevention or reversing γ-globin silencing.

As expected, among the entire list of the putative Sox6 interacting proteins, we identified several erythroid TFs, with a known function in the regulation of the globin genes (COUP-TFII, GATA1, GATA2 and Gif-1b). Several studies described COUP-TFII as TF that represses γ-globin expression in collaboration with other TFs forming repressing complexes ***(Aerbajinai, Zhu, Kumkhaek, Chin, & Rodgers, 2009; Liberati, Ronchi, Lievens, Ottolenghi, & Mantovani, 1998; Liberati, Sgarra, Manfioletti, & Mantovani, 1998; Ronchi, Bottardi, Mazzucchelli, Ottolenghi, & Santoro, 1995).*** Recently, Zhu *et al.* show that NF-Y recruits and stabilizes the binding of BCL11a, thanks to COUP-TFII, on the proximal γ-globin promoter region, thus forming the repression complex that silences γ-globin expression in human adult primary cells. Moreover, also BCL11a-XL has been reported to cointeract with Sox6 in the γ-globin promoter, in $CD34^{+}$ purified human adult primary cells from peripheral blood, where these two TFs collaborate to silence γ-globin expression ***(Xu et al., 2010)***. Furthermore, Sox6 and BCl11a are critical TFs implicated in the erythroid differentiation and in the regulation of the hemoglobin switching. Thus, these data open the question that Sox6, COUP-TFII, BCL11a and maybe, also other interacting proteins might form a repression hub on the γ-globin promoter. We will discuss more in detail the function of these TFs in the next chapter.

Other TFs identified as putative Sox6 partners were GATA-1 and GATA-2 that also have important roles in erythropoiesis and hemoglobin switching. GATA-1 was described as master regulator of erythropoiesis, working as an activator or repressor during the

differential stages of development. It is also implicated in the regulation of GATA-2, by silencing its expression in early erythropoiesis to ensure proper erythroid differentiation. Both, GATA-1 and Sox6 bind to the γ-globin promoter and other erythroid regulatory regions in adjacent binding in close proximity to each other. Thus, we could not be sure if they really physically interact or we co-immunoprecipitate them together due to their DNA-binding proximity. In fact, despite we treated the protein lysates with benzonase to degradate the DNA and the RNA, it is difficult to remove nucleic acids from proteins. Consequently, we can not exclude that this incomplete digestion can produce false positive interactors (Figure 2.6b).

Another erythroid candidate TF interacting with Sox6 is Gfi1b (The Growth factor independent 1b). Gfi1b is a transcriptional repressor with several roles in hematopoiesis: it is necessary for erythroid maturation, it regulates the differentiation from pro-erythroblast to erythrocytes, and for the silencing of globin genes during embryonic development, in both mouse primary fetal liver and bone marrow cells ***(Vassen et al., 2014)***. Moreover, its conditional KO in erythroid cells shows that *Gfi1b* deficient mice show upregulated levels of the embryonic εY- and βh1-globin genes in late embryonic stages and that these increased levels are also maintained in erythroid adult cells from *Gfi1b* deficient mice. This conditional KO also shows strong down regulation of GATA-1 and Sox6 genes. As both these TFs are largely implicated in the silencing of the embryonic globin genes, Gfi1b depletion might

induce, direct and indirect, upregulation of embryonic globins. Gfi1b is necessary for early erythroid maturation and to stabilize the repression hub at the moment of the switching. As Sox6 binds to the εϒ-globin promoter to induce its silencing ***(Yi et al., 2006)***, it might be possible that both TFs collaborate to silence embryonic globins.

Besides TFs, we also found several chromatin remodeling complexes to interact with Sox6. Between these chromatin remodelers, we could identify several complexes that act as repressors (such as Nucleosome Remodeling (NuRD) complex) ***(Allen, Wade, & Kutateladze, 2013)***, and others that act as activators (such as SWI/SNF complex) ***(Bultman, Gebuhr, & Magnuson, 2005; Lee, Murphy, Lee, & Chung, 1999)***, accordingly with the double function of Sox6. Both remodelers interact with several TFs to regulate the hemoglobin switching. The interaction between TFs and chromatin remodelers is crucial for the massive chromatin reorganization required for cellular differentiation.
NuRD complex interact with GATA-1 and FOG-1 to promote γ-globin silencing ***(Rodriguez et al., 2005),*** however the complex do not bind directly to the promoter ***(Rodriguez et al., 2005)***. The interaction between NuRD and FOG1 is mediated by a short, conserved N-terminal motif that is also found in other transcriptional repressors, including Bcl11A and Bcl11B, SALL1-4, Ebf-associated zinc finger protein (EBFAZ), and Evi3 ***(Hong et al., 2005)***. NuRD binding by FOG-1 is essential for normal erythroid and megakaryocyte development. The NuRD complex contains

histone deacetylases and interacts with diverse transcriptional repressors, leading to its classification as a corepressor complex. In the next chapters we will discuss the possible NuRD/Sox6 complex function. In fact, we hypothesize that NuRD is the co-factor required by Sox6 to repress γ-globin genes. Moreover, several subunits of the NuRD complex were also identified to interact with BCL11a ***(Xu et al., 2010).***

The SWI/SNF are complexes with a core enzymatic subunit with helicase-like ATPase activity. These complexes regulate DNA accessibility in chromatin fibers, thus they are important regulators of gene expression and genome stability. These complexes together with their co-factors are implicated in the regulation of erythropoiesis. Several subunits interact with erythroid TFs, being their expression critical for the activation of β-globin expression ***(Bultman et al., 2005)***. Furthermore, the KLF1 (Krüppel-like factor 1) TF interacts with several subunits of the SWI/SNF complex *in vitro* promoting γ-globin expression ***(Lee et al., 1999)***.

Between the CDKs identified by Mass Spectrometry sequencing, the most interesting are CDK12 and CDK13. The sequences of both proteins are highly similar, as well as they share common functions. There are different points that make them interesting as putative novel partners of Sox6: i) they have a role in regulating alternative splicing, ii) they are transcriptional regulators and iii)

CDK13 is upregulated in ringed sideroblasts anemia, suggesting its specific role in erythropoiesis.

1) Role in alternative splicing

In fact, a study published by Kenji Ohe show that Sox6 is playing a role in alternative splicing: Sox6 colocalizes in the nuclear speckles domains with other splicing factors, and its disruption impairs several steps of the splicing process ***(Ohe, Lalli, & Sassone-Corsi, 2002***). Moreover, in our lab we found evidences that Sox6 could mediate splicing of the BCL11a transcription factor. Both, CDK12 and CDK13 have been described implicated in inducing alternative splicing on splicing reporter constructs ***(Chen, Wong, Geneviere, & Fann, 2007; Even et al., 2006)***. Moreover, both kinases colocalize in the nuclear speckles with other splicing factors, as also does Sox6 ***(Ohe et al., 2002)***. Finally, CDK13 is involved in the phosphorylation of ASF/SF and in the alternative splicing of HIV RNA ***(Berro et al., 2008)***.

2) Transcriptional role of CDK13

There are some evidences that CDK12 and CDK13 are transcriptional regulators: they have the ability *in vitro* to phosphorylate Ser2 in the CTD domain of the RNA-polymerase II (RNAPII), a hallmark that releases RNAPII from the paused mode and allow productive elongation ***(Bartkowiak & Greenleaf, 2015; Blazek et al., 2011; Peterlin & Price, 2006).***

3) Putative specific role in erythropoiesis

Missregulation of CDK12 and CDK13 have been related with several diseases, but in particular, CDK13 was found upregulated in some patients with refractory anemia with ringed sideroblasts associated with marked thrombocytosis (RARS-T) ***(Malcovati & Cazzola, 2013; Malcovati et al., 2009)***. Accordingly, CDK13 promotes megakaryocytic development in bone marrow cell cultures ***(Lapidot-Lifson et al., 1992)***, suggesting that its inappropriate activation could force differentiation toward megakaryocytes at expenses of erythrocytes.

In the next chapters we will discuss the functional role of several Sox6 interacting proteins. We will focus our study on the nucleosome remodeling complex (NuRD), the COUP-TFII TF and in the cyclin dependent kinase 13 (CDK13).

REFERENCES:

Aerbajinai, W., Zhu, J., Kumkhaek, C., Chin, K., & Rodgers, G. P. (2009). SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII. *Blood, 114*(1), 187-194. doi: 10.1182/blood-2008-07-170712

Allen, H. F., Wade, P. A., & Kutateladze, T. G. (2013). The NuRD architecture. *Cell Mol Life Sci, 70*(19), 3513-3524. doi: 10.1007/s00018-012-1256-2

Bartkowiak, B., & Greenleaf, A. L. (2015). Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J Biol Chem, 290*(3), 1786-1795. doi: 10.1074/jbc.M114.612226

Beckett, D., Kovaleva, E., & Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci, 8*(4), 921-929. doi: 10.1110/ps.8.4.921

Berro, R., Pedati, C., Kehn-Hall, K., Wu, W., Klase, Z., Even, Y., . . . Kashanchi, F. (2008). CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. *J Virol, 82*(14), 7155-7166. doi: 10.1128/JVI.02543-07

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., . . . Peterlin, B. M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev, 25*(20), 2158-2172. doi: 10.1101/gad.16962311

Bultman, S. J., Gebuhr, T. C., & Magnuson, T. (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), 2849-2861. doi: 10.1101/gad.1364105

Cantu, C., Grande, V., Alborelli, I., Cassinelli, L., Cantu, I., Colzani, M. T., . . . Ronchi, A. (2011). A highly conserved SOX6 double binding site mediates SOX6 gene downregulation in erythroid cells. *Nucleic Acids Res, 39*(2), 486-501. doi: 10.1093/nar/gkq819

Chen, H. H., Wong, Y. H., Geneviere, A. M., & Fann, M. J. (2007). CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochem Biophys Res Commun, 354*(3), 735-740. doi: 10.1016/j.bbrc.2007.01.049

de Boer, D., Goemans, W. P., Ghezavat, V. R., van Ooijen, R. D., & Maes, R. A. (2003). Seizure of illicitly produced para-fluorofentanyl: quantitative analysis of the content of capsules and tablets. *J Pharm Biomed Anal, 31*(3), 557-562.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., . . . Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A, 100*(13), 7480-7485. doi: 10.1073/pnas.1332608100

Even, Y., Durieux, S., Escande, M. L., Lozano, J. C., Peaucellier, G., Weil, D., & Geneviere, A. M. (2006). CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *J Cell Biochem, 99*(3), 890-904. doi: 10.1002/jcb.20986

Grosveld, F., Rodriguez, P., Meier, N., Krpic, S., Pourfarzad, F., Papadopoulos, P., . . . Strouboulis, J. (2005). Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry. *Ann N Y Acad Sci, 1054*, 55-67. doi: 10.1196/annals.1345.008

Hong, W., Nakazawa, M., Chen, Y. Y., Kori, R., Vakoc, C. R., Rakowski, C., & Blobel, G. A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J, 24*(13), 2367-2378. doi: 10.1038/sj.emboj.7600703

Kamachi, Y., Uchikawa, M., & Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet, 16*(4), 182-187.

Kiefer, J. C. (2007). Back to basics: Sox genes. *Dev Dyn, 236*(8), 2356-2366. doi: 10.1002/dvdy.21218

Lapidot-Lifson, Y., Patinkin, D., Prody, C. A., Ehrlich, G., Seidman, S., Ben-Aziz, R., . . . Soreq, H. (1992). Cloning and antisense oligodeoxynucleotide inhibition of a human homolog of cdc2 required in hematopoiesis. *Proc Natl Acad Sci U S A, 89*(2), 579-583.

Lee, C. H., Murphy, M. R., Lee, J. S., & Chung, J. H. (1999). Targeting a SWI/SNF-related chromatin remodeling complex to the beta-globin promoter in erythroid cells. *Proc Natl Acad Sci U S A, 96*(22), 12311-12315.

Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., & Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol, 39*(12), 2195-2214. doi: 10.1016/j.biocel.2007.05.019

Liberati, C., Ronchi, A., Lievens, P., Ottolenghi, S., & Mantovani, R. (1998). NF-Y organizes the gamma-globin CCAAT boxes region. *J Biol Chem, 273*(27), 16880-16889.

Liberati, C., Sgarra, R., Manfioletti, G., & Mantovani, R. (1998). DNA binding of NF-Y: the effect of HMGI proteins depends upon the CCAAT box. *FEBS Lett, 433*(1-2), 174-178.

Malcovati, L., & Cazzola, M. (2013). Refractory anemia with ring sideroblasts. *Best Pract Res Clin Haematol, 26*(4), 377-385. doi: 10.1016/j.beha.2013.09.005

Malcovati, L., Della Porta, M. G., Pietra, D., Boveri, E., Pellagatti, A., Galli, A., . . . Cazzola, M. (2009). Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis. *Blood, 114*(17), 3538-3545. doi: 10.1182/blood-2009-05-222331

Ohe, K., Lalli, E., & Sassone-Corsi, P. (2002). A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proc Natl Acad Sci U S A, 99*(3), 1146-1151. doi: 10.1073/pnas.022645899

Ottolenghi, S., Nicolis, S., Taramelli, R., Malgaretti, N., Mantovani, R., Comi, P., . . . et al. (1988). Sardinian G gamma-HPFH: a T----C substitution in a conserved "octamer" sequence in the G gamma-globin promoter. *Blood, 71*(3), 815-817.

Peterlin, B. M., & Price, D. H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol Cell, 23*(3), 297-305. doi: 10.1016/j.molcel.2006.06.014

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., . . . Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J, 24*(13), 2354-2366. doi: 10.1038/sj.emboj.7600702

Ronchi, A. E., Bottardi, S., Mazzucchelli, C., Ottolenghi, S., & Santoro, C. (1995). Differential binding of the NFE3 and

CP1/NFY transcription factors to the human gamma- and epsilon-globin CCAAT boxes. *J Biol Chem, 270*(37), 21934-21941.

Rybak, J. N., Ettorre, A., Kaissling, B., Giavazzi, R., Neri, D., & Elia, G. (2005). In vivo protein biotinylation for identification of organ-specific antigens accessible from the vasculature. *Nat Methods, 2*(4), 291-298. doi: 10.1038/nmeth745

Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y), 11*(10), 1138-1143.

Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem, 68*(5), 850-858.

Vassen, L., Beauchemin, H., Lemsaddek, W., Krongold, J., Trudel, M., & Moroy, T. (2014). Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. *PLoS One, 9*(5), e96636. doi: 10.1371/journal.pone.0096636

Wilson, M., & Koopman, P. (2002). Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev, 12*(4), 441-446.

Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W., & Orkin, S. H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev, 24*(8), 783-798. doi: 10.1101/gad.1897310

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P. D., Fuchs, D. A., Erickson, D. T., . . . Brilliant, M. H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet, 2*(2), e14. doi: 10.1371/journal.pgen.0020014

Chapter 3

Validation and functional analysis of the putative interacting proteins

INTRODUCTION*:*

During the last decade proteomic research have made great advances in the efficient isolation and in peptide identification o nuclear protein complexes involved in transcriptional regulation, including TFs an their chromatin associated cofactors thanks to the development of powerful Mass Spectrometry (MS) technique. TFs are sequence-specific DNA-binding proteins, which can act as transcriptional activators and repressors. The low abundance of the TFs is the main limitation to their proteome analysis because they tend to be masked by the most abundant proteins ***(Rusconi, Guillonneau, & Praseuth, 2002)***.

In the previous chapter, I described a successful technique based on a single-step purification of biotin tagged-TFs, followed by Mass Spectrometry sequencing. Taking advantage of this approach, we set up an experimental pipeline to purify multiprotein complexes containing the Sox6 TF in erythroid cells (see chapter 2). The readout of the experiment was a long list of peptide sequences, which was further filtered by molecular function, on the basis of previous data from literature, and by looking at proteins with a known role in erythropoiesis. Regarding this last aspect, we compare the MS-derived list with data obtained from a DNA microarray analyses done in mouse erythroid primary cells. Based on these criteria, we finally selected a short list of candidate Sox6 interactors to be functionally studied. In order to validate all the candidates, we first set up immunoprecipitation (IP) to analyze its physical

interaction; and then, to obtain more evidences of a putative interaction at the physiological level of the proteins we performed gel filtration experiments from nuclear extracts of mouse primary fetal liver cells at days E12.5 and E13.5. Finally, we designed functional experiments to understand the molecular function of the complexes.

This chapter is focused on three different Sox6 interacting candidates (Proteins/complexes):

3.1. The Nucleosome Remodeling and Deacetylase (NuRD) complex.

3.2. The COUP-TFII transcription factor.

3.3. The cyclin dependent kinase 13 (CDK13).

EXPERIMENTAL PROCEDURES:

Constructs:

The construct expressing the *Escherichia coli* BirA biotin-protein ligase (containing 3' HA epitope) was kindly given by Prof. John Strouboulis **(de Boer et al., 2003)**. The Sox6 murine cDNA (obtained from Prof. Michiko Hamada-Kanazawa, Japan) was cloned in frame with a 3' FLAG epitope to generate a Sox6-FLAG cassette. 3' to the FLAG epitope, a sequence coding the 23 amino acids necessary for the biotin-tag was cloned to generate the bioSox6-FLAG cassette, flanked by two BglII sites. This cassette was then cloned immediately upstream to the IRES - Emerald GFP cassette (blunted BamHI site) of the CSIemerald derived from pHR SIN CSGW, **(Demaison et al., 2002)** lentiviral vector, a kind gift from Prof. Tariq Enver, UCL, London. In this vector, the SFFV promoter drives the expression of the exogenous cDNA, highly active in hematopoietic cells. Starting from this construct, the N-terminal 9 amino acids of Sox6 were deleted by cutting at a NotI (upstream the ATG starting codon) and a BsrgI (in Sox6 cDNA) restriction sites. A kozak consensus followed by the ATG starting codon was reconstituted. The N-terminus truncated bioSox6-FLAG cassette (BglII fragment) was then cloned immediately upstream to the IRES-GFP cassette (BglII-compatible BamHI site) of the CSIemerald. The two packaging plasmids psPAX2 and pMD-VSVG were used to produce Lentiviral pseudo-particles in HEK 293T cells (www.lentiweb.com).

Short-hairpin RNA (shRNA) interference oligonucleotides were

designed against human and mouse CDK13. ShRNA targeting the human and mouse CDK13 contains the sequence "5'-TAGGACAAGACCGAGCCTAATTTTCAAGAGAAATTAGG CTCGGTCTTGTCCTTTTTTTC -3'". Lentilox 3.7 lentiviral vector (pLlx3.7) was used to cloned the shRNA sequence as described (web.mit.edu/ccr/labs/jacks/protocols/pll37cloning.htm). The vector carries a GFP cassette under the CMV promoter. Lentilox 3.7 empty vector was used as control.

Lentiviral vector production:

Viral stocks were produced by transient cotransfection of the three different vectors in the Human Embryonic Kidney (HEK) 293T packaging cell line. 72 hours after transfection, the supernatant containing the recombinant viruses was collected, filtered (0.45µm), and centrifuged at 20,000g for 8 hours at 4ºC. The viral pellet was resuspended in PBS 1x and stored at -80ºC. Viral titers were determined by transduction of HEL cells with serial dilutions of the vector stock and by scoring the GFP (for CSIemerald_EV/CSI_bioSox6FLAG/CSI_Sox6$_{GFP}$/LLX_EV/LLX_CDK 13) or by staining them with Anti-NGFR (CD271) antibody conjugated PE/APC antibodies (for CSI$_{\Delta NGFR}$ and COUP-TFII$_{\Delta NGFR}$) transgene expression by fluorescence-activated cell sorter (FACS) analysis.

Cell Cultures and transduction:

β-K562 and Human erytholeukemic (HEL-BirA, stably expressing the biotin ligase BirA) cell lines, were cultured in RPMI 1640

medium (Lonza) supplemented with 10% heat inactivated fetal bovine serum (Lonza), L-glutamine (Euroclone), antibiotics Penicillin-Streptomycin 100U/100ug/ml (Euroclone) and in a humidified 5% CO_2 atmosphere at 37°C. In the cells stably expressing BirA the medium was supplied with 3µg/ml of puromycin (Sigma P8833). Transduction was performed overnight with a multiplicity of infection (MOI) of 30.

RNA isolation and Real Time PCR

Total RNA from >10^6 cells (from HEL-BirA cells, β-K562 cells, from mouse fetal liver E11.5, E12.5 and E13.5 dpc erythroid cells or from human $CD34^+$ erythroid culture purified from peripheral blood) was purified with TRIzol Reagent (Applied Biosystem), treated with RQ1 DNase (Promega) for 30 min at 37°C and retrotranscribed (High Capacity cDNA Reverse Transcription Kit, Applied Biosystem). Negative control reactions (without Reverse Transcriptase) gave no PCR amplification. Real time analysis was performed using ABI Prism 7500, (Applied Biosystems). Primers were designed to amplify 100 to 200bp amplicons, spanning an exon-exon junction when possible, on the basis of sequences from the Ensembl database (http://www.ensembl.org). Samples from each experiment were analyzed in triplicate. Specific PCR product accumulation was monitored by SYBR Green dye fluorescence in 12-µl reaction volume. Dissociation curves confirmed the homogeneity of PCR products.

All primers used are listed in the table below:

Human:

Primers	**Sequence (5'-3')**
GAPDH Fw	ACGGATTTGGTCGTATTGGG
GAPDH Rw	TGATTTTGGAGGGATCTCGC
α-globin Fw	GAGGCCCTGGAGAGGATGTTCC
α-globin Rw	ACAGCGCGTTGGGCATGTCGTC
β-globin Fw	TACATTTGCTTCTGACACAAC
β-globin Rw	ACAGATCCCCAAAGGAC
γ-globin Fw	CTTCAAGCTCCTGGGAAATGT
γ-globin Rw	GCAGAATAAAGCCTACCTTGAAAG
ε-globin Fw	GCCTGTGGAGCAAGATGAAT
ε-globin Rw	GCGGGCTTGAGGTTGT
Sox6 Fw	GAGGCAGTTCTTTACTGTGG
Sox6 (endogenous) Rw	CCGCCATCTGTCTTCATAC
Sox6 (Flag) Rw	CTTATCGTCGTCATCCTTGTA
Coup-TFII (endogenous) Fw	TTGACTCAGCCGAGTACAGC
Coup-TFII (endogenous) Rw	AAAGCTTTCCGAATCTCGTC
Coup-TFII (exogenous) Fw	TCCAAGAGCAAGTGGAGAAG
Coup-TFII (exogenous) Rw	CTTCCAAAGCACACTGGGAC
CD71 Fw	AAAATCCGGTGTAGGCACAG
CD71 Rw	CCTTTAAATGCAGGGACGAA
CD235 Fw	CAGCTCATGATCTCAGGATG
CD235 Rw	CACCTCAGTGGTACTTAATGC
Bcl-xL Fw	GAATGACCACCTAGAGCCTTGG
Bcl-xL Rw	TGTTCCCATAGAGTTCCACAAAAG
Bcl-2 Fw	ATGTGTGTGGAGAGCGTCAACC
Bcl-2 Rw	TGAGCAGAGTCTTCAGAGACAGCC

IGF-1 Fw	ATGCTCTTCAGTTCGTGTGTG
IGF-1 Rw	GCACTCCCTCTACTTGCGTTC
SOCS3 Fw	GGAGACTTCGATTCGGGACC
SOCS3 Rw	GAAACTTGCTGTGGGTGACC
FECH Fw	ATCCAGCAGCTGGAGGGTCT
FECH Rw	TGAATCTTGGGGGTTCGGCG
ALAS2 Fw	CAACATCTCAGGCACCAGTA
ALAS2 Rw	CTCCACTGTTACGGATACCT
GPIIB Fw	CTCCACAACAATGGCCCTGG
GPIIB Rw	CTTGAGAGGGTTGACAGGAG
GPIIIA Fw	AAGTGTGAATGTGGCAGCTG
GPIIIA Rw	GTGGCCTCTTTATACAGTGG
Fli-1 Fw	AATGGATCCAGGGAGTCTCC
Fli-1 Rw	CCTTGCCATCCATGTTCTGG

Mouse:

GAPDH Fw	TGTGTCCGTCGTGGATCTGA
GAPDH Rev	CCTGCTTCACCACCTTCTTGA
Sox6 Fw	TTCCTCCTGCATGGAAAAAC
Sox6 Rev	GATGCTGCCAGCTTTTTCTG
Coup-TFII Fw	AAGCAAGCCACCTCTCCATT
Coup-TFII Rw	GGTGTTGATCACTGCCCTCT
CDK13 Fw	ATTCTTCGGCAACTCACCCA
CDK13 Rev	AAGCCTGATTCCAGCAGTCC

Nuclear extracts:

Cells were harvested and centrifuged at 640g for 15 min at 4ºC. The pellet washed three times in ice-cold 1x PBS (phosphate buffered saline, pH 7.4) and gently resuspended in ice-cold NP-40 lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.5%

v/v NP-40, supplemented with protease inhibitors (complete EDTA-free, Roche) added just before use). Cells were lysed by incubation on a rotating wheel for 10 min at 4ºC and nuclei were then centrifuged at 3300g. Nuclei were resuspended in nuclear lysis buffer (10mM HEPES-KOH pH 7.9, 100mM KCl, 3mM $MgCl_2$, 0,1 mM EDTA, 20% glycerol and protease inhibitors, as above). Nuclear proteins were extracted by drop-wise addition of 4M KCl with gentle agitation on ice, until the final salt concentration was approx. 350-400mM. Nuclear lysis and protein extraction were then allowed to proceed by incubation on a rotating wheel for 40 min at 4ºC, followed by centrifugation. The supernatant, which correspond to the soluble nuclear extract fraction, was retained.

Benzonase treatment:

Nuclear extracts were diluted by adding 3 volumes of HENG buffer (0 mM KCl, 10 mM HEPES pH=9, 1.5 mM MgCl2, 0.25 mM EDTA, 20 % glycerol, PMSF or PI) and 25 U of Benzonase (Viscolase nuclease (A&A Biotechnology)) per mg of nuclear extract were added. The concentration of Mg^{+2} was adjusted to 2mM. Diluted lysates were incubated with Benzonase for 2-4 hours at 4ºC on a rotating wheel. DNA digestion and removal was checked on gel on a phenol-extracted aliquot.

Binding to Streptavidin beads:

Paramagnetic streptavidin beads (Dynabeads M-280, Dynal (Life Technologies); 50 µl per 1 mg of protein) were washed three times with PBS at room temperature and blocked with HENG/BSA

(Sigma-Aldrich) 200µg/ml (10 mM HEPES pH 9, 1.5 mM MgCl2, 0.25 mM EDTA, 20 % glycerol, PMSF) for 1h at room temperature under rotation. The binding was carried out with Nuclear Extracts diluted in 150 mM KCl and 0.3 % NP40), overnight on a rotating wheel. After 5 washes with wash HENG solution (HENG buffer with 300mM KCl) and 2 washes with PBS 1x at room temperature, the bound material was eluted by boiling for 10 min in Laemmli protein simple loading buffer (62.5mM Tris–HCl pH6.8, 25% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.1% bromophenol blue) and analyzed by immunoblotting.

Superose 6 gel filtration:

Size fractionation of protein complexes was carried out on AKTA FLPC apparatus with a Superose 6 10/300 GL Tricorn™ column (GE Healthcare). Fractions were precipitated with 100% acetone and analyzed by immonoblotting. Molecular size standards were thyroglobulin (670KDa), Ferritin (440 KDa), BSA (67 KDa) and ribonuclease (14 kDa).

Immunoprecipitations:

Nuclear extracts were precleaned at 4ºC using protein G sepharose beads and affinity purified IgG (mouse- Santa Cruz, CA, SC-2025). Anti-FLAG M2 affinity gel (A220 – Sigma) was used to immunoprecipitate the proteins. Immunoprecipitations were performed at 4ºC for 3 hours. Washes were carried out at room temperature in PBS 1x (phosphate buffered saline, pH 7.4). Bound material was eluted by boiling in 1x Laemmli protein simple

loading buffer (62.5mM Tris–HCl pH6.8, 25% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.1% bromophenol blue) and analyzed by immunoblotting.

Immunoblotting analysis:

Nuclear proteins were resolved by SDS/PAGE in a 7% gel and blotted onto ProTran Nitrocellulose membrane (brand) by "wet blotting." carried out under constant voltage at 100V for 90-120 min at 4ºC (Transblot apparatus, Biorad). Membranes were blocked for 1-2h at room temperature with Milk 5% in TTBS (1x TBS (Tris buffered saline, pH 7.6) 0,1% Tween 20 (Sigma)) and incubated with the appropriate primary antibody diluted in Milk 3% TTBS overnight at 4ºC. Membranes were washed for three times with TTBS and incubated with the secondary antibody (diluted in Milk 3%). The antibodies used were: anti-FLAG antibody (abcam ab125243); anti-Sox6 (abcam ab64956); anti-MTA1 (Santa Cruz sc17779); anti-HDAC1 (Santa Cruz sc81598); anti-BCL11a (all isoforms) (Novus, Ctip1 Antibody [NB600-258]), anti-COUP-TFII (abcam H7147); anti-CDK13 (anti-CDC2L5 Bethyl laboratories). Protein loading control was checked with an anti-CPSF73 homemade antibody (kindly given from Prof. Silvia Barabino laboratory). Antibodies binding was detected by using appropriate horseradish peroxidase-conjugated IgG and revealed by ECL (Millipore).

Flow cytometry analysis and FACS:

Transduced β-K562 cells were washed, fixed in 1%

Paraformaldehyde and stained with PE-anti NGFR conjugated antibody (Biolegend) for 15 min at 4ºC and analyzed by flow cytometer (Becton-Dickinson FACS Calibur).

Freshly extracted mouse fetal liver cells from E11.5, E12.5 and E13.5 cells were disaggregated to single cells and washed in PBS 1x (phosphate buffered saline, pH 7.4), and incubated with conjugated antibodies: allophycocyanin (Schuh, Kinast, Mezzomo, & Kapczinski) conjugated anti-mouse CD117 (c- Kit), Phycoerythrin (PE) anti-mouse CD71 and FITC anti-mouse Ter119 all from Becton-Dickinson. Cells were analyzed by flow cytometer (Becton-Dickinson FACS Calibur). The erythroid population within all the fetal livers cells were >95%.

Chapter 3.1: the NuRD (Nucleosome Remodeling and Deacetylase) chromatin remodeling complex

INTRODUCTION:

The Mass Spectrometry peptide sequencing of Sox6 interactors identified almost all the subunits present in the NuRD (Nucleosome Remodeling and histone Deacetylase) complex, involved in gene repression. The eukaryotic genome is highly regulated by several mechanisms involving chromatin reorganization and epigenetic modulation. These mechanisms determine the accessibility/inaccessibility status of DNA to nuclear proteins and complexes.

The NuRD complex was first identified in 1998, and it remains the only known chromatin remodeler with a double function of ATPase and of histone deacetylase ***(Xue et al., 1998)***. This complex has been implicated in a wide variety of nuclear processes, including gene transcription, DNA damage repair, maintenance of genome stability and chromatin assembly. Aberrant activity of NuRD is related with aging and cancer ***(Lai & Wade, 2011; Pegoraro et al., 2009)***. The main subunits forming the complex are CHD3/4 (ATPase), HDAC1/2 (histone deacetylases), MBD2/2 (methyl-CpG-binding domain), RBBP7/4 (retinoblastoma binding proteins) MTA1/2/3 (metastasis associated proteins) and p66α/β (transcriptional repressor proteins). Although the subunits composition of the complex is variable, depending on the type of the locally introduced

epigenetic changes, several studies identify NuRD as a repressive complex. This repression is obtained by protein-DNA interactions (mediated by the MBD2/3 subunit interacting with methylated DNA) or by protein-protein interactions between the complex and other transcriptional repressors. Two different mode of action have been described for the complex: i) a "static" mode, where NuRD acts as a repressor by its association to the silenced genes; and ii) a "dynamic" mode, in which NuRD maintains a dynamic equilibrium between gene activation and repression, mediated by histone acetylation at the active genes loci and resulting in fine-tuning of gene expression ***(Allen, Wade, & Kutateladze, 2013)***.

The N-terminal domain of several TFs contains a repression motif, necessary and sufficient for the recruitment of all NuRD components ***(Lin, Roche, Wilk, & Svensson, 2004)***. This motif consists in a highly conserved 12-amino acids peptide, common to several zing finger TFs, including all four Sall family members, FOG1 and FOG2, BCL11a and BCL11b. This consensus motif consists in an aminoacidic MSRRKQaKPqhF sequence, where the amino acids in capital letters mean conservation throughout all the zinc fingers TFs containing it. Single amino acid substitutions within this motif abolish the interaction with NuRD and thus the repression activity of these TFs. In particular, arginine 3 and 4, lysine 5 and proline 9 (RRKqxxP) are the crucial amino acids whose substitution results in the loss or reduction of NuRD recruitment and in the subsequent loss of repression activity ***(Hong et al., 2005; Lauberth & Rauchman, 2006)***.

The NuRD complex has been reported to have an important role in the regulation of erythroid development. MBD2 and CHD4 (also called Mi2-β) subunits play a critical role in silencing of the human embryonic ε- and fetal γ-globin genes, in coordination with several TFs (BCL11a-XL and KLF1) ***(Amaya et al., 2013; Gnanapragasam et al., 2011; Rupon, Wang, Gaensler, Lloyd, & Ginder, 2006; Rupon, Wang, Gnanapragasam, Labropoulos, & Ginder, 2011)***. Finally, in transgenic mice carrying the entire human beta locus, the GATA1/FOG1/NuRD complex is implicated in the negative regulation of the γ-globin gene by binding to its distal promoter ***(Harju-Baker, Costa, Fedosyuk, Neades, & Peterson, 2008)***.

Within the N-terminus of the Sox6 gene, we identified a consensus very similar to the NuRD interaction motif described above. We then hypothesized that Sox6 could recruit the NuRD complex in order to repress gene expression. To test this hypothesis, we generated a truncated N-terminal version of Sox6 with the final goal to possibly perturb the Sox6 ability to recruit NuRD and to test whether this could result in the derepression of genes regulated by Sox6.

RESULTS:

The NuRD complex interacts with Sox6 in human erythroid HEL cells.

The NuRD complex has been identified as a repressive complex containing double catalytic activity: ATPase and histone deacetylase. This complex binds in a sequence specific manner to some TFs known as repressors. In our Mass Spectrometry experiment we identified several subunits of the NuRD complex among Sox6 interactors, suggesting that this repressive multiprotein complex could be the responsible for the Sox6-mediated repression of some of its target genes. To verify this interaction we transduced human (HEL) and mouse (MEL) erythroleukemic cells with a lentivirus carrying the bioSox6-FLAG cDNA expression cassette. Three days after transduction, we performed a nuclear proteins extraction, followed by the pull down of Sox6 interacting proteins by using streptavidin beads. We then confirmed by Western blot that MTA1 and HDAC1 physically interact with Sox6 in HEL cells (Figure 3.1.1). In our analysis, we were unable to identify MDB2, (neither by Mass Spectrometry and by Western blot), suggesting that Sox6 interacts with NuRD but not with MeCP-1, a complex formed by all NuRD subunits together with MBD2. These data demonstrate that Sox6 can interact with diverse subunits of the NuRD complex.

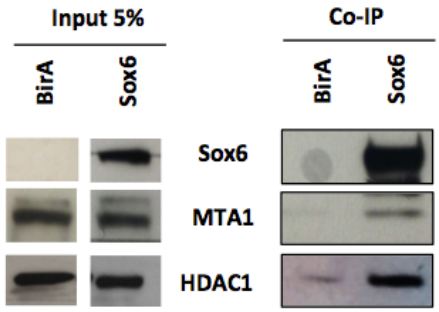


Figure 3.1.1 Physical interaction between Sox6, MTA1 and HDAC1, two subunits of the NuRD complex*. Co-immunoprecipitation of the bioSox6-FLAG, MTA1 and HDAC1 proteins from HEL cells transduced and non-transduced with bioSox6-FLAG. Immunoprecipitation was done by using the biotin-streptavidin approach. Co-purified proteins were analyzed by western blot using the anti-Sox6, anti-MTA1 and anti-HDAC1 antibodies. 5% of the total nuclear lysate was used as input.*

Although we validated the physical interaction between Sox6 and the different subunits of the NuRD complex, the interaction was demonstrated in HEL cells upon bioSox6-FLAG overexpression, what is a forced non-natural system.

To explore whether the NuRD subunits and Sox6 are coexpressed during mouse adult erythroid development, and to support the possibility of their physiological interaction, we performed a gel filtration size-exclusion Superose 6 chromatography of nuclear extracts freshly purified from E12.5 (Figure 3.1.2a) and E13.5 (Figure 1c) primary mouse fetal liver cells (The gel filtration was done in the Laboratory headed by Prof. Angela Bachi in IFOM (Milan)). The elution patterns of Sox6, MTA1 and HDAC1 were substantially overlapping, and the molecular mass of the positive fractions for the three proteins was around 157 KDa. Recent study demonstrated that Sox6 interact with BCL11a-XL and cooperate

with it in the silencing of the γ-globin transcription in adult human erythroid progenitors ***(Xu et al., 2010)***. We thus tested the eluted fractions for the presence of BCL11a-XL in E13.5 mouse fetal liver cells. Interestingly, we found that whereas the shorter isoforms of BCL11a (BCL11a-S) co-elute with Sox6 in the same fractions, the larger BCL11a-L/-XL isoforms (which interacts with Sox6 in human adult progenitors silencing γ-globin expression) is in the same fractions as Mi2-β (Figure 3.1.2b). Accordingly with a recent study describing that BCL11a-XL might also interact with different subunits of the NuRD complex (such as HDAC1 and Mi2-β) (Xu et al., 2013).

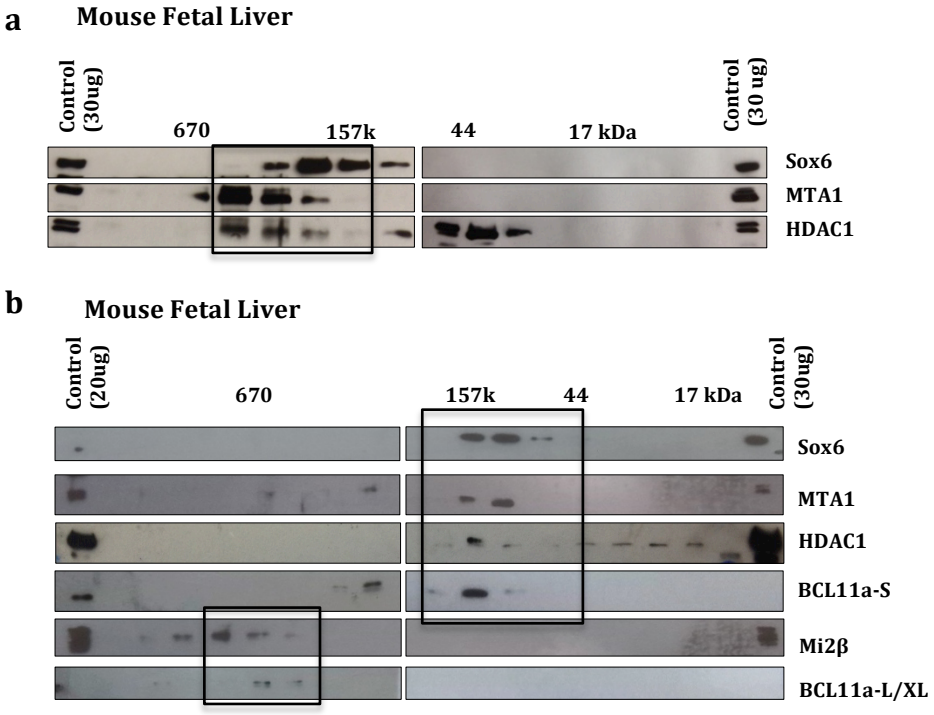


Figure 3.1.2 Gel filtration size-exclusion Superose 6 chromatography.
a) Gel filtration of nuclear extracts from mouse fetal liver E12.5 cells. Fractions were analyzed by western blot for the presence of Sox6, MTA1 and HDAC1. There is a significant overlap between Sox6 and the two components of the NuRD complex. b) Gel filtration fraction of nuclear extracts from mouse fetal liver E13.5 cells. Fractions were analyzed by western blot for the presence of Sox6, MTA1, HDAC1, Mi2-β and for the different isoforms of BCL11a. There is a significant overlap between Sox6 and the two components of the NuRD complex and the shorter isoforms of BCL11a (S). Whereas, the larger isoforms of BCL11a (L/XL) overlap with Mi2-β. Elution positions of molecular mass standards are shown in the top of the figures (from right to left: 670 KDa, 157 KDa, 44 KDa and 17 KDa).

The Sox6 N-terminal domain contains a motif with high similarity with the NuRD interacting motif.

A highly conserved 12-amino acids N-terminal motif is necessary for the recruitment of the NuRD complex by several TFs. The MSRRKQaKPqhF repression motif is indeed present in different zinc finger TFs, such us Sall1, FOG1, FOG2, BCL11a and BCL11b, known to recruit NuRD. Interestingly, the N-terminal domain of Sox6 shows a high percentage of identity with the consensus repressive motive (MSSKQATSP): Serine 2, lysine 5 and glutamate 6 (the spacing in the Sox6 motif corresponds to lysine 4 and glutamate 5) and proline 9 (Figure 3.1.3a). Although this motif is not perfectly identical to the canonical consensus, the two crucial amino acids for NuRD binding (and the only ones able to completely abolish NuRD recruitment) are conserved in the Sox6 N-terminal domain: Lysine 5 and Proline 9. Moreover, the N-terminal sequence is evolutionary conserved among Sox6 orthologs in vertebrates from *Danio rerio* to *Homo sapiens,* whereas the flanking region is not (Figure 3.1.3b).

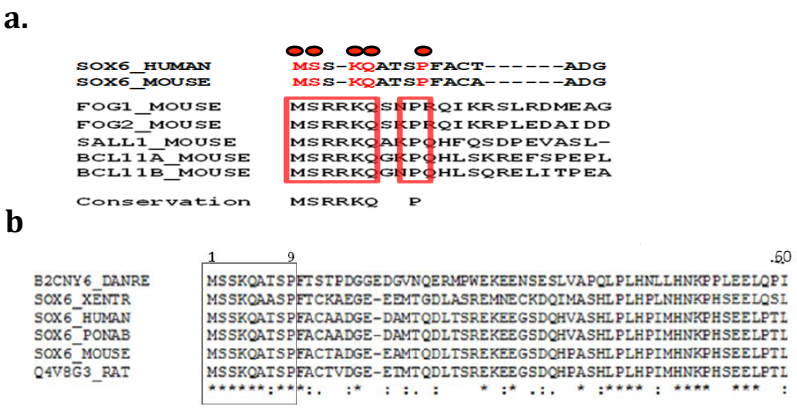


Figure 3.1.3 The conserved N-terminal motif of Sox6. *a) Alignment of the extreme N-terminus of the different mammalian transcription factors, Sox6, FOG1, FOG2, Sall1, BCL11A and BCL11B, reveals a high sequence homology within the 12 first amino acids. The conserved residues are shown in red boldface type. Sequences were obtained from NCBI database. b) Evolutionary conservation of the N-terminus aminoacidic sequence of Sox6 among vertebrates. The putative N-terminus NuRD repressive motive is boxed.*

The deletion of the Sox6 N-terminal nine amino acids does not abolish the ability of Sox6 to bind to the NuRD complex subunits.

In order to test *in vivo* the ability of the N-terminal Sox6 motif to recruit NuRD, we transduced human erythroleukemic β-K562 cells (a variant K562 subclone cells expressing β-globin) with a lentivirus carrying a bioSox6-FLAG cDNA cassette lacking the first nine amino acids of the N-terminal domain (ΔN-Sox6) (Figure 3.1.4a) and with the corresponding biotin tagged Sox6 full length - cDNA (FL-Sox6). 3 days after transduction, we assessed both the

ΔN-Sox6 and FL-Sox6 mRNA (Figure 3.1.4b) and protein expression (Figure 3.1.4c).

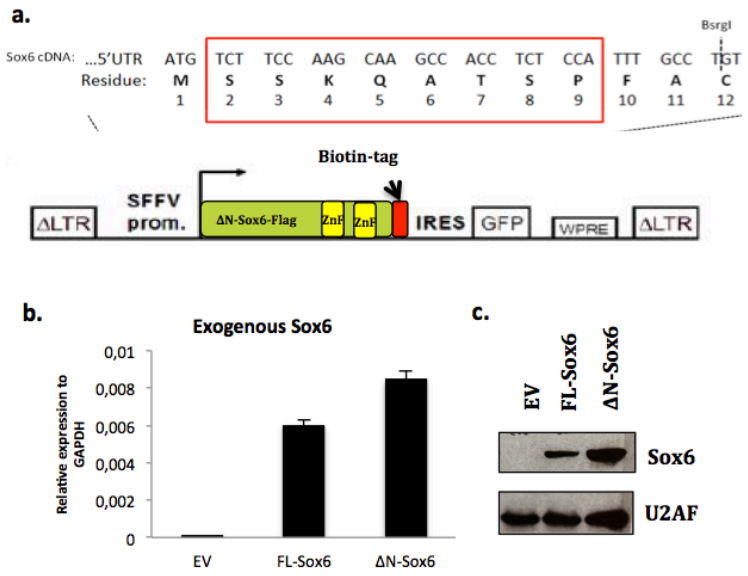


Figure 3.1.4 Expression of FL-Sox6 and ΔN-Sox6. *a) Schematic representation of the lentiviral vector expressing FL-Sox6 and ΔN-Sox6 under SFFV (Spleen focus forming virus) promoter and representation of the deletion sequence within the ΔN-Sox6 construct (marked by the red box). b) RT-qPCR shows the mRNA level of exogenous Sox6. Error bars show SD. c) Western Blot for nuclear extracts of β-K562 overexpressing FL-Sox6 and ΔN-Sox6 by using anti-Sox6 antibody. U2AF was used as loading control.*

We performed RT-qPCR analyses in order to identify any differential expression in genes regulated by Sox6 upon the overexpression of either FL-Sox6 or ΔN-Sox6 in β-K562 cells. To the end of to detect a presumable derepression of Sox6 target genes upon ΔN-Sox6 overexpression in comparison to FL-Sox6 overexpression, we analyzed several genes that normally are downregulated by Sox6 overexpression during erythroid differentiation in β-K562 cells. First of all, we analyzed the

endogenous Sox6 mRNA levels. In fact, Sox6 has the ability to repress its own transcription by binding to its promoter during terminal erythroid maturation. Thus, the overexpression of exogenous Sox6 reduces the expression of the endogenous transcript ***(Cantu et al., 2011)***. However, our results show that both exogenous FL-Sox6 and ΔN-Sox6 retain the ability to reduce the endogenous Sox6 transcript (Figure 3.1.5a). Also, the anti-apoptotic gene Bcl-2 and the insulin growth factor 1 (IGF1) are equally repressed by ΔN-Sox6 and FL-Sox6 overexpression (Figure 3.1.5b/c).

Finally, we also check the expression of erythroid genes that are normally upregulated by Sox6 during the erythroid differentiation (such as β-like globin genes (Figure 3.1.6a), the transferrin receptor (CD71) (Figure 3.1.6b) and the δ-aminolevulinate synthase 2 enzyme (ALAS2) (Figure 3.1.6c) genes, both implicated in the heme synthesis pathway. No major differences were found between FL-Sox6- and ΔN-Sox6-mediated overexpression of these genes.

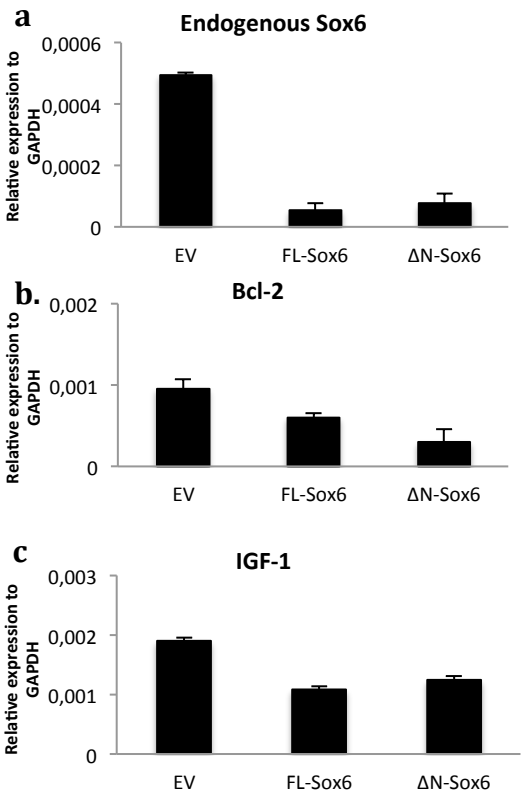


Figure 3.1.5 RT-qPCR analysis upon overexpression of FL-Sox6 and ΔN-Sox6 on genes normally downregulated by Sox6. *Reduced mRNA levels upon overexpression of FL-Sox6 and ΔN-Sox6 of a) endogenous Sox6. b) Bcl-2 pro-apoptotic gene. c) Insulin growth factor 1 (IGF1-1). Error bars show SD. Relative to technical replicates representative experiment of N≥3 is shown.*

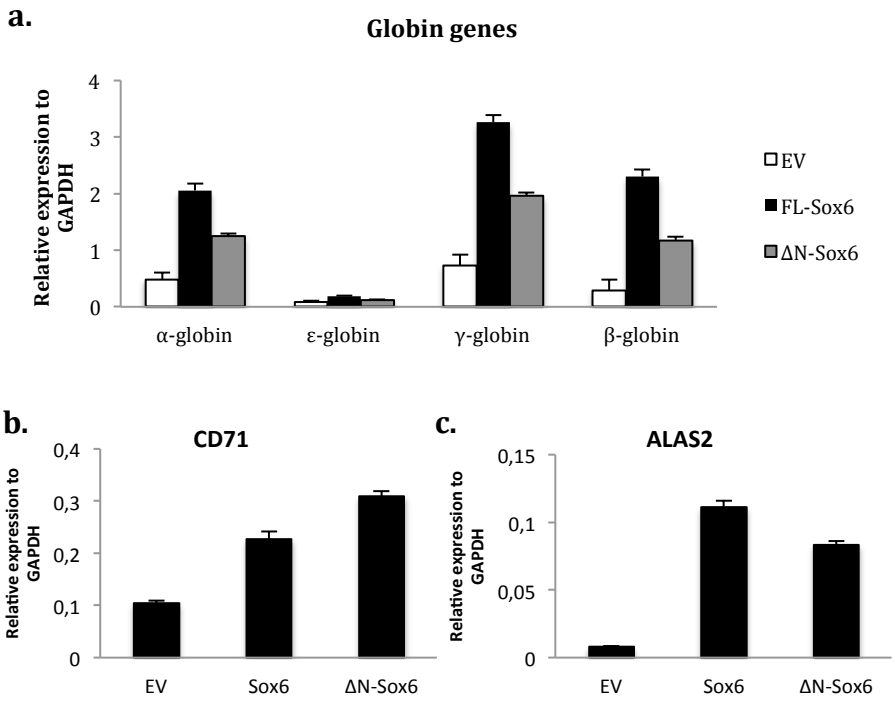


Figure 3.1.6 RT-qPCR analysis upon overexpression of FL-Sox6 and ΔN-Sox6 genes normally by Sox6. *Increased mRNA levels upon overexpression of FL-Sox6 and ΔN-Sox6 of a) all the β-like globin genes, b) The transferrin receptor (CD71), and c)* The δ-aminolevulinate synthase 2 enzyme (ALAS2). *Error bars show SD. Relative to technical replicates representative experiment of N≥3 is shown.*

Despite β-K562 cell line is a good model to assess the variation between γ- and β-globin genes ratio, and also to assess the genetic differences in erythroid differentiation, this model has a limitation to study protein-protein interaction. These cells start to die three days after of Sox6 transduction due to block in proliferation and terminal erythroid differentiation elucidated by Sox6. Thus, in order to asses the physical interaction between NuRD complex and Sox6 we used Human erytholeukemic (HEL) cells. As I previously mentioned, these cells carry 8 copies of the JAK2

V617F mutation making them insensitive to Suppressor Of Cytokine Signaling 3 (SOCS3). SOCS3 is a direct target of Sox6, which is upregulated when Sox6 is overexpressed, thus producing a block in proliferation and the withdrawal from the cell cycle. Hence, upon Sox6 overexpression, HEL cells do not exit from the cell cycle as fast as β-K562, but they still retain the ability to differentiate. So, we used HEL-BirA cells to assess the interaction between biotin-Tagged ΔN-Sox6 (bioΔN-Sox6) and the subunits of the NuRD complex, MTA1 and HDAC1. To this end we performed a single-step purification of bioΔN-Sox6 by using streptavidin beads, and we then analyzed the pulled down complexes by Western blot. Against our expectations, we could still detect MTA1 and HDAC1, suggesting that the deletion of the Sox6 N-terminal nine amino acids is not sufficient to impair the ability of the NuRD subunits to bind to Sox6 (Figure 3.1.7).

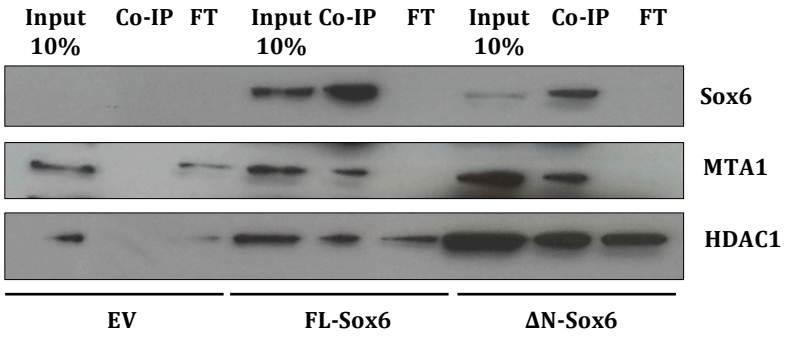


Figure 3.1.7 Co-immunoprecipitation (Co-IP) by using the biotin-streptavidin pull-down approach on nuclear lysates from HEL-BirA cells transduced with EV, FL-Sox6 and ΔN-Sox6. *Both, FL-Sox6 and ΔN-Sox6 retain the ability to immunoprecipitate the subunits of the NuRD complex, MTA1 and HDAC1. 10% of the total lysate is used as input. Flow through (FT) was loaded was equivalent to the 10% input.*

These data suggest that, in the case of Sox6 the N-terminal repression motif might not be responsible for the interaction between Sox6 and the NuRD complex.

As an alternative explanation, since it is known that Sox6 has the ability to homo- and heterodimerize through its coiled-coiled domain; it might be possible that ΔN-Sox6 dimerize with the endogenous FL-Sox6, thus forming a Sox6 dimer retaining the Sox6 WT properties. This explanation is, however, unlikely because the large amount of the ΔN-Sox6 overexpressed protein should promote the formation of ΔN-Sox6/ ΔN-Sox6 dimers.

DISCUSSION:

NuRD (Nucleosome Remodeling and histone Deacetylase) complex is a multiprotein complex with a role in chromatin reorganization and epigenetic modulation; it regulates the accessibility/inaccessibility status of DNA to nuclear proteins and complexes ***(Allen et al., 2013).***

Sox6 is a TF that can act either as an activator or a repressor **(Kamachi, Uchikawa, & Kondoh, 2000; Kiefer, 2007; Lefebvre, Dumitriu, Penzo-Mendez, Han, & Pallavi, 2007)**. In this study we validated by immunoprecipitation the interaction between Sox6 and with MTA1 and HDAC1, two of the subunits of the NuRD complex, in human erythroleukemic (HEL) cells (Figure 3.1.1). of interest, it is known that Sox6 and BCL11a collaborate in the silencing of the γ-globin promoter to switch from HbF to HbA ***(Xu***

et al., 2010). Moreover, also the NuRD complex has been described to interact with GATA-1 and FOG-1 on the γ-promoter to repress γ-globin gene expression ***(Ginder, Gnanapragasam, & Mian, 2008; Sankaran, Xu, & Orkin, 2010b; Stamatoyannopoulos et al., 1994)***. So, it might be possible that Sox6 thanks to the NuRD complex recruitment could exert its function in silencing γ-globin expression. Several studies in humanized transgenic mice (carrying the human β-locus) show that the hemoglobin switching takes place from the embryonic E10.5 to E13.5 stage. In order to support the possible Sox6/NuRD complex interaction in physiological conditions during the hemoglobin switching we performed the Superose6 size-exclusion chromatography of nuclear extracts from E12.5 (Figure 3.1.2a) and E13.5 (Figure 3.1.2b) mouse fetal liver primary cells, which revealed an overlapping pattern between Sox6, MTA1 and HDAC1, suggesting a possible interaction.

Sox6 was described to interact and cooperate with BCL11a-XL on the γ-globin promoter to silence γ-globin expression ***(Xu et al., 2010)***. Recently, BCL11a-XL was described also to interact with the NuRD subunit Mi2-β and HDAC1 ***(Xu et al., 2013)***. Hence, we were expecting an overlapping elution pattern of both BCL11a-XL and Mi2β with Sox6, MTA1 and HDAC1 in the gel filtration experiment. Surprisingly, the elution pattern of the larger isoforms of BCL11a (L/XL) and Mi2β was totally different from the pattern of Sox6, MTA1 and HDAC1. Whereas Sox6 was eluted around 157 KDa, the elution pattern of BCL11a-(L/XL) and Mi2-β

was around 670 KDa (Figure 3.1.2b). However, we identified the shorter isoforms of BCL11a (S) in the same elution fractions as Sox6. The BCL11a isoforms are differentially expressed during development, starting with high expression of the shorter isoforms during primitive erythropoiesis and in the fetal stage, and then moving to high expression of the larger isoforms (XL and L) in the adult. These isoforms, differentially expressed during development, have a direct relation with the transition from "high HbF" to "low HbF", with the BCL11a-L/XL isoforms being implicated in γ-globin silencing. It was described that the most abundant isoforms at the second trimester of human fetal liver development, and in the primitive circulating erythroblasts are the shorter BCL11a isoforms, where there is still a robust expression of γ-globin ***(Sankaran et al., 2008)***. The overlapping patter between Sox6, MTA1, HDAC1 and the shorter isoforms of BCL11a could be due to at the mouse embryonic stage E13.5 the expression of γ-globin has been reduced but not totally silenced. So, Sox6 could probably interact with the different BCL11a isoforms during the hemoglobin switching time, interacting before with the shorter isoforms and then shifting toward the larger ones. Thus, resulting in the silencing of γ-globin expression in collaboration of the NuRD complex.

Regarding the protein domain of Sox6 mediating the NuRD recruitment, Sox6 presents a highly conserved N-terminal motif (MSRRKQaKPqhF) found in several zinc finger TFs, such as BCL11a, BCL11b, FOG-1, FOG-1 and all the Sall family members

(Lin et al., 2004) (Figure 3.1.3a). The presence of this motif is sufficient and necessary to repress gene expression by the recruitment of the NuRD complex to the regulatory regions of their target genes ***(Lauberth & Rauchman, 2006)***. Point mutations analysis within the motif describes the arginine 3 and 4, lysine 5 and proline 9 (RRKqxxP) as the crucial amino acids whose mutation abolishes repression and produces significant loss or reduction in NuRD subunits binding ***(Hong et al., 2005)***. Interestingly, the Sox6 N-terminus aminoacidic sequence shows highly similar conservation to the repression motif necessary to recruit the NuRD complex (Figure 3.1.3a). These amino acids are also highly conserved among Sox6 orthologs in vertebrates (Figure 3.1.3b). Pointing an important functionality role. We then created a truncated version of Sox6 (ΔN-Sox6) lacking the 9 N-terminus amino acids (Figure 3.1.4a). Surprisingly, when we analyze the interaction between NuRD subunits and biotin-tagged ΔN-Sox6 (bioΔN-Sox6), we observed that bioΔN-Sox6 still retains the ability to bind MTA1 and HDAC1 in HEL cells (Figure 3.1.7).

These results gave evidences that either the N-terminus motif is not necessary for Sox6 to recruit the NuRD complex, or the endogenous Sox6 might dimerize with ΔN-Sox6 to form dimers still able to recruit NuRD complex and to exert gene repression.

Although the ΔN-Sox6 mutation does not perturb the interaction between MTA1 and HDAC1 and Sox6, we might expect some differential expression in the mRNA levels of genes directly or

indirectly regulated by Sox6. We then analyzed the expression of genes that are downregulated upon Sox6 overexpression (such as endogenous Sox6, the anti-apoptotic gene Bcl-2 and insulin growth factor 1 IGF1) to check whether any of these genes was derepressed upon ΔN-Sox6 overexpression in β-K562 cells. Again, we could not find any significant difference between FL-Sox6 and ΔN-Sox6 (Figure 3.1.5). Finally, to find out whether β-K562 cells overexpresing ΔN-Sox6 were differentiating normally, we tested the expression of erythroid genes that are normally upregulated during the erythroid differentiation (such as β-like globin, CD71 and ALAS2 genes, the last two implicated in the heme synthesis pathway). Again, no major differences were found between FL-Sox6 and ΔN-Sox6 (Figure 3.1.6).

All the above results suggest that the N-terminal motif is not necessary for Sox6 function.

In conclusion, the fact that we could not disrupt the interaction between ΔN-Sox6 and the NuRD complex subunits, and we could not find any differential expression in the mRNA levels of genes regulated by FL-Sox6 during erythroid differentiation, suggest that this motif is not unique and exclusively necessary to Sox6 to repress or activate its targets.

Chapter 3.2: the role of Sox6 and COUP-TFII in the regulation of the hemoglobin switching

INTRODUCTION:

The human β-globin locus consists of an upstream regulatory region, the locus control region (LCR), and five different β-like coding genes: ε, γ^G, γ^A, δ and β. The process in which β-like globin genes are differentially express during development is called hemoglobin switching. There are two switches during the erythroid development, the first, at embryonic stage, where ε-globin is replaced by γ-globin; and the second around birth, where γ-globin is replaced by β-globin ***(Bank, 2006; Grosveld, Dillon, & Higgs, 1993; Sankaran, Xu, & Orkin, 2010a)***. There are several diseases due to a failure in the β-globin expression, called β-hemoglobinopathies. Between them the most common are β-thalassemia and Sickle cell disease (SCD), they are the most frequent inherited single gene disorders in the human population ***(Weatherall & Clegg, 1996)***. The clinical phenotype of the patients is milder when there is an increase in the expression of γ-globin in response to some drugs, such as hydroxyurea (HU) ***(Steinberg, 2002, 2006; Steinberg, Nagel, & Brugnara, 1997)*** or in patients with Hereditary Persistence of Fetal Hemoglobin (HPFH) ***(Li, Duan, & Stamatoyannopoulos, 2001)***. The reactivation of the γ-globin expression has been largely studied in order to develop an attractive treatment for these diseases. The understanding the mechanism of the switching from γ- to β-globin is the basis to develop new therapeutical approaches normal

switching process. ***(Akinsheye et al., 2011; Bank, 2006; Higgs, Engel, & Stamatoyannopoulos, 2012; Stamatoyannopoulos, 2005)***.

The hemoglobin switching is a highly regulated process. Regulatory regions are present throughout the locus, the most prominent of which is the LCR in the 5' region upstream in the locus. The LCR elements are bound by TFs complexes ***(Soler et al., 2010)***. The interactions with the TF complexes change during development ***(Palstra, de Laat, & Grosveld, 2008)***, resulting in temporal and developmental stage specific globin gene expression.

COUP-TFII is a TF that binds to direct repeats elements (RE) AGGTCA ***(Cooney, Tsai, O'Malley, & Tsai, 1992; Kliewer et al., 1992)***. It binds *in vitro* the ε- and γ-globin promoters, where binds to the double CCAAT box region, interfering with NF-Y binding and possibly repressing ε- and γ-globin expression ***(Liberati, Ronchi, Lievens, Ottolenghi, & Mantovani, 1998; Ronchi, Bellorini, Mongelli, & Mantovani, 1995)***. COUP-TFII might contribute to the HPFH phenotype because several mapped mutations within the double CCAAT box region of the γ-globin promoter cause the perturbation of COUP-TFII binding *in vitro* ***(Liberati et al., 2001)***. An *in vitro* model for primary human erythroblasts development shows that with the stimulation of the cytokine Stem Cell Factor (SCF) there is a reduction of COUP-TFII expression at the mRNA level and protein level. Moreover,

knockdown of COUP-TFII in cultured *ex vivo* human erythroblasts results in up-regulation of γ-globin expression ***(Aerbajinai, Zhu, Kumkhaek, Chin, & Rodgers, 2009)***. Recent study proponed that NF-Y recruits and stabilizes the binding of BCL11A with COUP-TFII on the proximal γ-globin promoter region forming the repression complex silencing γ-globin expression ***(Zhu et al., 2012)***.

Among the number of TFs present in the list of interacting proteins of Sox6 as the outcome of the mass spectrometry sequencing, we focused our study in the interaction between Sox6 and COUP-TFII TFs. We decide to study the functional role of their interaction due to both TFs have been implicated in the repression of ε- and γ-globin expression (*See introduction pages 46-47).*

RESULTS

COUP-TFII interacts with Sox6 in human and mouse erythroid cells

In the mass spectrometry results (described in chapter 2) from human erythroleukemic cells (HEL), we identified several COUP-TFII peptides. In order to validate the interaction between Sox6 and COUP-TFII, we transduced HEL/BirA cells with a lentivirus carrying a biotin-tagged Sox6 cDNA expression cassette (bioSox6-FLAG). Three days after transduction, we performed a nuclear proteins extraction, followed by the pull down of Sox6 interacting proteins by using streptavidin beads. We then confirmed by Western blot that COUP-TFII physically interacts with Sox6 in HEL/BirA cells (Figure 3.2.1).

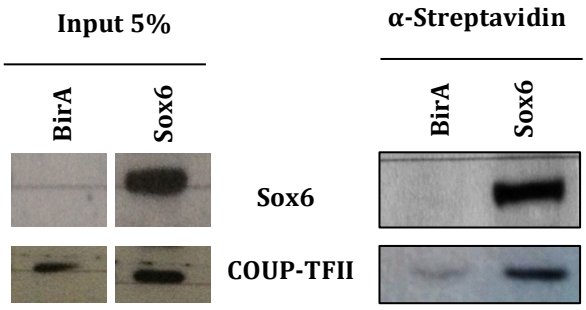


Figure 3.2.1 Physical interaction between Sox6 and COUP-TFII in HEL cells. *Co-immunoprecipitation of the bioSox6and COUP-TFII proteins from HEL cells transduced and non-transduced with bioSox6-FLAG. Immunoprecipitation was done using the biotin-streptavidin approach. Co-purified proteins were analyzed by western blot by using anti-Sox6 and anti-COUP-TFII antibodies. 5% of the total nuclear lysate was used as input.*

Profiling COUP-TFII and Sox6 expression during mouse embryonic development

In our lab previously, we performed a DNA microarray to characterize the transcriptional profile of mouse erythroid cells during hemoglobin switching. We purified erythroid cells from mouse fetal liver at days E11.5, E12.5 and E13.5, and we sorted them on the basis of the expression of cell surface specific antigens, cKit - TER119 to isolate three different stages of erythroid differentiation. The erythroid population within the fetal liver during these developmental stages is higher than 95%. Fetal liver is the main erythropoietic organ from the embryonic days E10.5 to E16.5, when the cells start to colonize the bone marrow.

In order to understand how Sox6 and COUP-TFII behave during mouse embryonic development and differentiation we moved back to above described DNA Microarray output. Whereas Sox6 expression increases during development, the expression of COUP-TFII seems to be upregulated in late differentiation at the developmental stage E11.5, and it seems to be at low levels in the stage E13.5, but the results were not clear.

To validate the data, we again purified freshly hematopoietic cells from fetal liver populations at days E11.5, E12.5 and E13.5 of embryonic development, when more than 95% of cells are erythroid. From the purified cells we obtain mRNA and nuclear protein extracts. We used the mRNA to assess the expression

levels of Sox6 and COUP-TFII during development (Figure 3.2.2a). This results confirms that the expression level of Sox6 increases, whereas the expression level of COUP-TFII decreases during the hemoglobin switching period (The expression levels were normalized on GAPDH expression). To validate their differential expression at the protein level during, we performed Western Blot analysis by using anti-Sox6 and anti-COUP-TFII antibodies, and for an internal housekeeping gene used as a loading control, anti-CPSF73 was used as a loading control. The Western Blot results were in agreement with RT-qPCR analyses: Sox6 expression increases, in parallel with COUP-TFII decrease (Figure 3.2.2b). During these embryonic days adult hemoglobin increases while embryonic hemoglobin decreases. Which seems to be in conflict with the described role of COUP-TFII, repressing γ-globin.

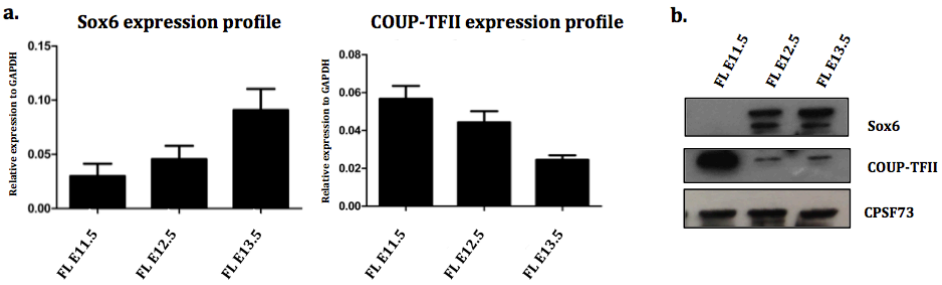


Figure 3.2.2 Expression profile of Sox6 and COUP-TFII in mouse primary fetal liver cells. *a) RT-qPCR analysis of the mRNA expression of Sox6 (left) and COUP-TFII (right) during mouse embryonic development (E11.5 – E13.5). b) WB analysis of purified nuclear extracts from the same mouse primary fetal liver cells confirms the expression of both TFs at the protein level, using anti-Sox6 and anti-COUP-TFII antibodies. Anti-CPSF73 was used as a loading control.*

COUP-TFII interacts with Sox6 in mouse primary fetal liver cells

To have more evidence of the Sox6 and COUP-TFII interaction observed in the mass spectrometry experiments (see chapter 2), we followed the elution profile of these two TFs, in size exclusion Superose 6 chromatography of nuclear extracts form E12.5 mouse primary fetal liver cells (Figure 3.2.3), as done for NuRD complex (experiment done in collaboration with the Laboratory of Prof. Angela Bachi in IFOM (Milan)). The elution patterns of Sox6 and COUP-TFII were substantially overlapping, and the molecular mass of the positive fractions was between 670 and 157 KDa.

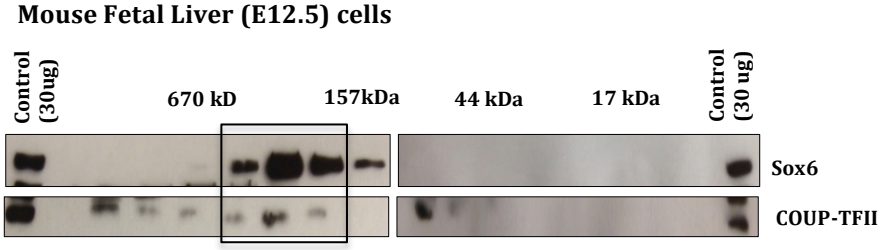


Figure 3.2.3 Gel filtration size-exclusion Superose 6 chromatography.
Gel filtration fraction of nuclear extracts from mouse fetal liverE12.5 cells. Fractions were analyzed by western blot with anti-Sox6 and anti-COUP-TFII antibodies. There is a significant overlap between both TFs. Overlapped fractions are boxed. Elution positions of molecular mass standards are shown on the top of the figure (from right to left: 670 KDa, 157 KDa, 44 KDa and 17 KDa).

These results confirm that they might interact during mouse embryonic development due to their co-elution in the same fractions. In the next sections we will try to address this question and further characterize the specific roles in this process.

Overexpression of sox6 and COUP-TFII in β-k562 cells

The human β-K562 cell line is a variant of the normal K562 cell line that express significant levels of β-globin (see chapter 4 for its characterization). Thus, this is a good model to study the effect of TFs overexpression on the regulation γ- to β-globin expression ratio. The β-K562 cells were transduced with COUP-TFII and Sox6 expressing lentiviruses (Figure 3.2.4) to characterize their role in globin gene regulation.

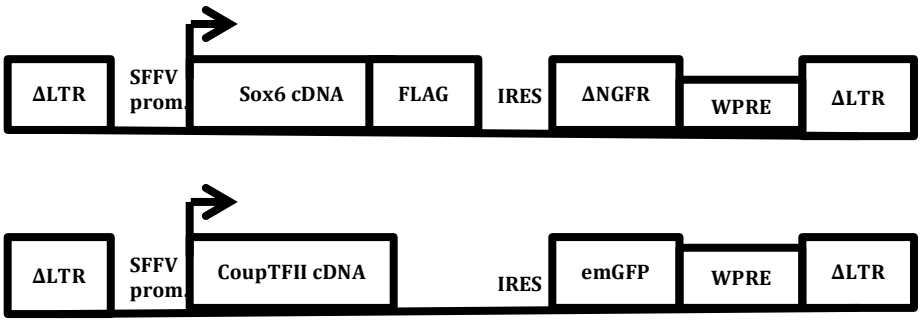


Figure 3.2.4 Lentiviral expression vectors for Sox6-FLAG and COUP-TFII. *Schematic representation of the Sox6 and COUP-TFII overexpressing lentiviral vectors used. LTR, Long Terminal Repeats; SFFV, Spleen Focus Forming Virus; WPRE, Woodchuck-Hepatitis-virus-Posttranscriptional-Regulatory Element.*

In agreement with previous results obtained in our lab, the overexpression of Sox6 in the K562 cells enhances their terminal maturation ***(Cantu et al., 2011)***. This is accompanied by an increase in the expression level of all the globins genes (α, ε, β and γ) due to a general differentiation effect (Figure 3.2.5a). The largest increase elicited by Sox6 overexpression is that of β-globin

expression (8.5x), whereas the lowest increase is observed in γ-globin expression (3.2x). When setting γ+β=1, the β/(γ+β) ratio increases from 0,2 in the EV-β-K562 to 0,7 in Sox6-β-K562.

In contrast, COUP-TFII overexpression does not induce any characteristic phenotypic change of the erythroid differentiation, as assessed by unchanged levels of CD71 and TER119 analyzed by FACS and RT-qPCR (Data not shown). However, it does induce an increase in relative levels of ε- and γ-globin genes transcription when compared to those cells transduced with the corresponding empty vector (EV) (Figure 3.2.5b). When setting γ+β=1, the β/(γ+β) ratio decreases from 0,2 in EV-β-K562 to 0,08 in COUP-TFII-β-K562. No changes were observed in the levels of β-globin gene, whereas there is a reduction in the α-globin gene expression levels (by approximately 40% of reduction). Overall, the overexpression of COUP-TFII forces an increase in γ-globin gene expression. This effect is against the previous described role of COUP-TFII, where it was proposed as a repressor of γ-globin gene (see discussion). As the overexpression of COUP-TFII has no typical erythroid phenotypic change, it might be possible that COUP-TFII plays a specific role in the hemoglobin genes regulation.

These data suggest that Sox6 and COUP-TFII have an opposite role on γ- and β-globin expression. Sox6 mainly increases β-globin, whereas COUP-TFII increases γ-globin. To better elucidate whether Sox6 and COUP-TFII could directly contribute to the γ- to β-globin switching, we co-transduced both TFs in β-K562 cells in

order to analyze globin expression upon simultaneous overexpression of different concentration of Sox6 and COUP-TFII. These conditions should mimic the changes of COUP-TFII and Sox6 expression observed during the Hemoglobin Switching (Figure 3.2.2)

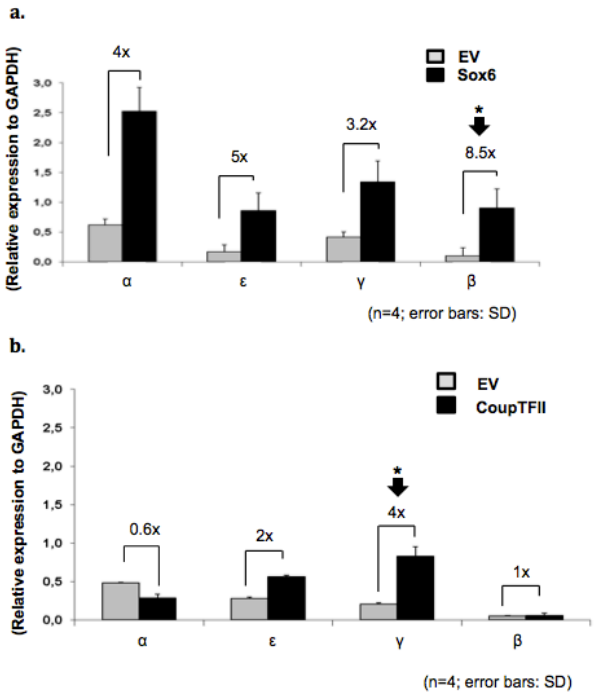


Figure 3.2.5 Globin genes expression upon Sox6 and COUP-TFII overexpression. *a) RT-qPCR analysis of all globins on cDNA from β-K562 cells transduced with the EV and Sox6 lentiviruses. Sox6 increases the levels of all the globins, specifically β-globin. b) COUP-TFII overexpression increases the level of ε and more specifically γ-globins expression. Histograms show expression levels relative to GAPDH (n=4; P<0,05).*

Simultaneous co-transduction of β-K562 cells with Sox6 and COUP-TFII

To set up the experiment of Sox6 and COUP-TFII co-transduction of both TFs we cloned their cDNA in the same lentivectors containing different reporter cassettes: Sox6 in the vector containing the emerald GFP cassette, whereas COUP-TFII was cloned upstream the ΔNGFR cassette (ΔNGFR can be detected by using anti-ΔNGFR-PE antibody, visualized in flow cytometry analysis in the red FL2 channel). The use of the two different reporter cassettes allows us to independently assess the level of transduction of Sox6 and COUP-TFII in co-transduction experiments. The corresponding empty vector (EV) containing either the GFP or the ΔNGFR cassettes were used as a control. Cells were transduced with equal Multiplicity Of Infection (MOI) of both $Sox6_{GFP}$ and $COUP\text{-}TFII_{\Delta NGFR}$ lentiviruses, and also at different relative concentrations, to mimic the changes in their relative abundance observed during the switching. In particular, we kept one TF at fixed levels and we increased the concentration of the other one by increasing the MOI of the corresponding expression vector. The efficiency of transduction it was monitored 72 hours after transduction by flow cytometry (the efficiency of transduction was higher than 90% in all the experiments) and the expression of exogenous Sox6 and COUP-TFII were tested by western blot.

In the first series of experiments, the amount of COUP-TFII was kept fixed and Sox6 was gradually increased correspondingly with the respective individual controls and the Empty vector (Figure 3.2.6). Then, mRNA from these cells was purified three days after transduction and subjected to RT-qPCR analysis of globin genes expression. The expression of exogenous Sox6 and COUP-TFII was also tested by western blot (Figure 3.2.6a). RT-qPCR results reveal an increase in the level of β-globin expression parallel to Sox6 increase (Figure 3.2.6b). When setting γ+β=1, the β/(γ+β) ratio increases from 0,2 in the EV-β-K562 to 0,56 in Sox6-β-K562. In the presence of COUP-TFII and Sox6 together, the ratio of β/(γ+β) increases from 0.50 to a maximum of 0.67 in the presence of the highest Sox6 concentration (Figure 3.2.6b).

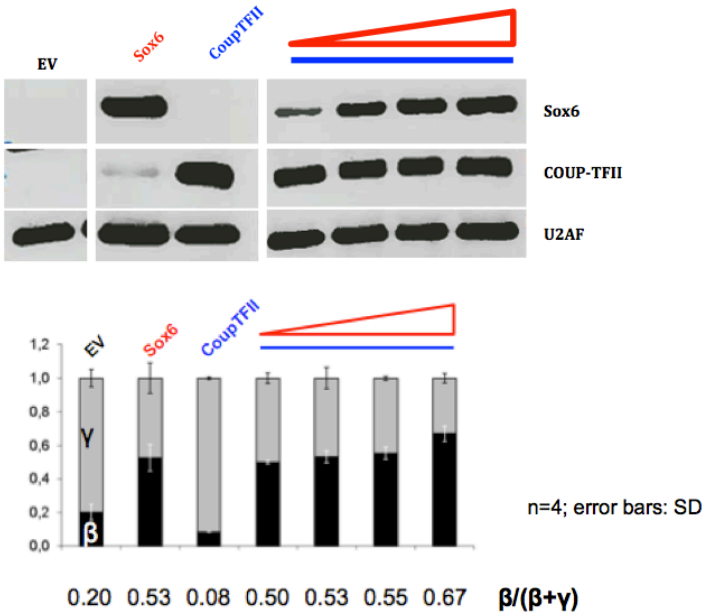


Figure 3.2.6 Co-tranduction of β-K562 cells with Sox6 and COUP-TFII.
Upper panel: WB analysis of exogenous Sox6 and COUP-TFII levels by using anti-FLAG and anti-COUP-TFII antibodies, respectively. U2AF expression was used as a loading control. Lanes 4-7 indicates the co-expression of fixed amount of COUP-TFII with increased levels of Sox6. Lower panel: RT-qPCR analysis shows the increase in the levels of β- to γ-globin ratio upon Sox6 increase with the presence of fixed COUP-TFII. Black histograms represent the ratio of β over (γ+β) and the grey bars indicate the ratio of γ over (γ+β), being γ+β set equal to 1, normalized against GAPDH (N=4).

A second set of experiments was performed in an opposite direction, maintaining the amount of Sox6 fixed and gradually increasing the expression of COUP-TFII (Figure 3.2.7). The transduction efficiency was again controlled by FACS (more than 90% of positivity of infection) and by protein levels of both TFs (Figure 3.2.7a). RT-qPCR analysis reveals a progressive dose-dependent increase in the proportion of γ-globin expression (Figure 3.2.7b). When setting γ+β=1, the β/(γ+β) ratio decreases from 0,2 in the EV-β-K562 to 0,02 in COUP-TFII-β-K562. In the presence of increased COUP-TFII together with fixed Sox6, the ratio of β/(γ+β) decreases from 0.34 to 0.08 in the presence of the highest COUP-TFII concentration (Figure 3.2.7b). The ratio changes from 0.7 when we transduced only Sox6 to 0.08 when the cells were co-transduced with the highest amount of COUP-TFII.

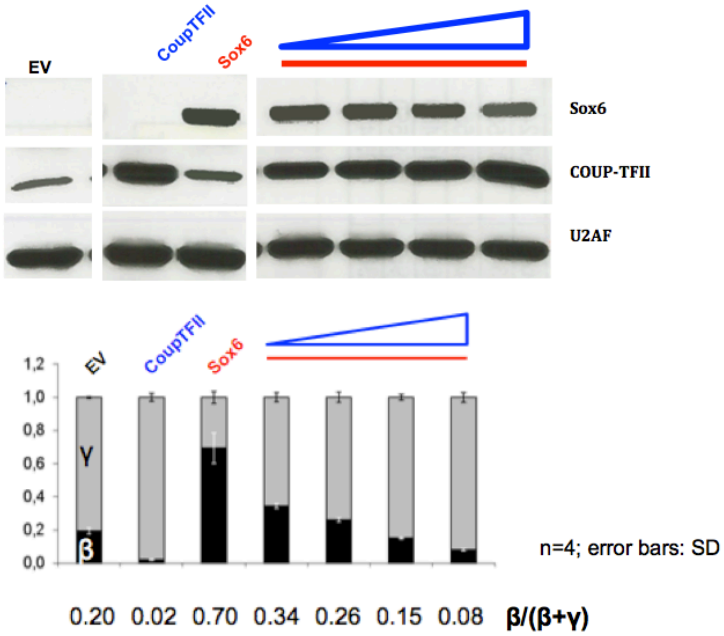


Figure 3.2.7 Co-transduction of β-K562 cells with Sox6 and COUP-TFII.
Upper panel: WB analysis of exogenous Sox6 and COUP-TFII levels by using anti-FLAG and anti-COUP-TFII antibodies, respectively. U2AF expression was used as a loading control. Lanes 4-7 indicates the co-expression of fixed amount of Sox6 with increased levels of COUP-TFII. Lower panel: RT-qPCR analysis shows the increase in the levels of β- to γ-globin ratio upon COUP-TFII increase with the presence of fixed Sox6. Black histograms represent the ratio of β over (γ+β) and the grey bars indicate the ratio of γ over (γ+β), being γ+β set equal to 1, normalized against GAPDH (N=4).

These data confirm that the relative levels of Sox6 and COUP-TFII are important modulators of the γ/β expression ratio, suggesting that changes in their level could contribute to tip the balance towards β-globin expression during the switching. Moreover, these results are in agreement with the previous results where both Sox6 and COUP-TFII showed an opposite expression profile during erythroid embryonic development in mouse fetal liver cells

from E11.5 to E13.5, where Sox6 was upregulated and COUP-TFII silenced (Figure 3.2.2).

Overexpression of COUP-TFII in human CD34$^+$ "*ex vivo*" primary cells

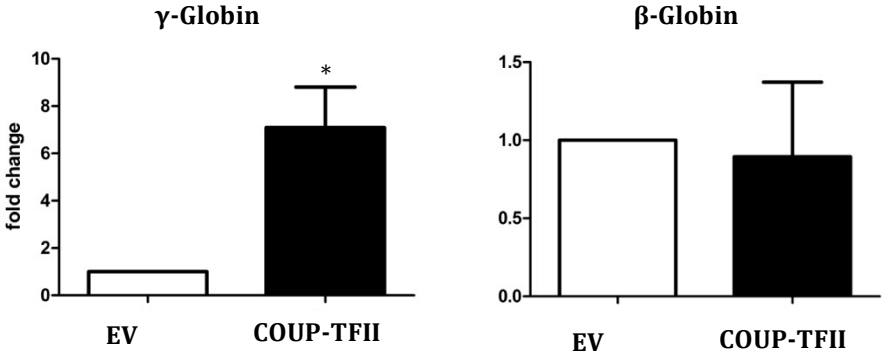
The ability of COUP-TFII to upregulate γ-globin expression in β-K562 cells suggests that by understanding its molecular mechanism of action could help to reactivate γ-globin gene expression in patients with β-thalassemia o Sickle Cell Disease (SCD). In order to confirm the role of COUP-TFII as γ-globin activator in human cells, we moved to a better model where the levels of γ-globin gene have been already mostly or completely silenced.

Primary erythroid cultures form CD34$^+$ cells purified from peripheral blood can be established from both healthy donors and β-thalassemic patients. These human *ex vivo* cultures are closer to the *in vivo* situation than cell lines. A two-Phase liquid culture system (Phase I & II, expansion and differentiation) ***(Fibach, Manor, Oppenheim, & Rachmilewitz, 1989)*** was used to perform these cultures and their transduction with COUP-TFII overexpression lentiviruses. Cultures were done in Cagliari (Italy) in the laboratory headed by our collaborator Prof. Paolo Moi & Dr. Giusi Marini.

We used human purified CD34$^+$ cells from peripheral blood (cells predominantly expressing β-globin) to test whether transduction

of COUP-TFII can increase γ-globin expression. BFUe cells from cultured Peripheral blood from five different healthy donors were transduced with the empty vector (EV-ΔNGFR) and the corresponding COUP-TFII-ΔNGFR lentiviral vector (Figure 3.2.4). Cells were infected on the day-1 of their differentiation phase with a multiplicity of infection (MOI) equal to 30. The ΔNGFR cassette expresses a truncated protein membrane that allows to sort infected cells by using magnetics beads carrying the bound NGFR monoclonal antibody. 72hrs post transduction, cells were collected and magnetically sorted. mRNA was purified from the sorted cells to analyze globin gene expression.

RT-qPCR confirmed that transduction of COUP-TFII in human *ex vivo* primary cultures increases γ-globin expression; whereas the expression of the β remains constant (Figure 3.2.8). This preliminary data is pointing at COUP-TFII as a prospective target whose modulation can increase γ-globin expression in human cells. To better characterize the ability of COUP-TFII to reactivate γ-globin expression, we will set up cultures from peripheral blood of thalassemic patients to evaluate whether COUP-TFII is also able to reactivate γ-globin in patients.



N=5; error bars: SD.
EV= Empty vector.

Figure 3.2.8 Overexpression of COUP-TFII in human "ex vivo" CD34+ cells purified from peripheral blood. *RT-qPCR analysis of cDNA from human "ex vivo" cultures from peripheral blood of healthy donors transduced with the EV and COUP-TFII. Increase levels of the γ-globin expression upon COUP-TFII overexpression in comparison to the EV. No significant difference in β-globin expression. Histograms show the relative level of expression to GAPDH (n=5; P<0,05).*

DISCUSSION

COUP-TFII emerged as a putative Sox6 interactor in the mass spectrometry described in chapter 2. COUP-TFII is an orphan nuclear receptor, playing a critical role acting in several developmental processes. COUP-TFII was proposed to have a negative role on the γ-globin expression ***(Filipe et al., 1999; Liberati et al., 1998; Ronchi et al., 1995)***.

In this chapter, we have confirmed the interaction between Sox6 and COUP-TFII in HEL cells by immunoprecipitation (Figure 3.2.1).

In mouse, as it was demonstrated by using transgenic mice carrying the complete human β-globin locus, the hemoglobin switching takes place in the embryonic stages from E10.5 to E13.5. Thus, we analyzed the protein and gene expression pattern of both Sox6 and COUP-TFII in the mouse fetal liver cells during embryonic development at the E11.5, E12.5 and E13.5 stages. Surprisingly, whereas Sox6 expression increases, COUP-TFII decreases during development (Figure 3.2.2). These unexpected results raise the question, whether they also have opposite roles in regulating the hemoglobin switching. Of interest, next experiments reveal that COUP-TFII overexpression in β-K562 cells upregulates of γ-globin expression (Figure 3.2.5b).

In contrast with the previous data describing COUP-TFII as transcriptional repressor of γ-globin gene, this results suggest that COUP-TFII instead of having a role in the silencing of the γ-globin expression, it might induces γ-globin activation. In contrast, the overexpression of Sox6 induces a general increase in all globins but specifically increases the β-globin expression when compared to the γ-globin increase (8.5x and 3.2x; Figure 3.2.5a). This supports the idea that COUP-TFII and Sox6 play an opposite role regulating globins gene.

We then co-transduced both Sox6 and COUP-TFII in β-K562 cells. We kept one TF at fixed concentration and we progressively increased the other, and vice-versa (Figure 3.2.6 and Figure 3.2.7). These experiments reveal that Sox6 and COUP-TFII have an

opposite effect on the β/(γ+β) ratio. Their differential expression is able to influence the switch between β- and γ- globin gene expression. The increased amount of Sox6 forces increased levels of β-globin expression, whereas the increased amount of COUP-TFII forces the upregulation of γ-globin expression levels. Taking into account that during hemoglobin switching γ-globin gene expression is silenced and β-globin expression is activated, these data are in agreement with the expression pattern seen in mouse primary fetal liver cells during development, where Sox6 expression increases, whereas the expression of COUP-TFII decreases (Figure 3.2.2).

Furthermore, to support a possible interaction between these two TFs at their physiological levels during hemoglobin switching, we performed a Superose 6 size-exclusion chromatography of nuclear proteins from mouse fetal liver primary cells at the embryonic stage E12.5 (when both TFs are present). The results demonstrated that both proteins have an overlapping elution pattern (Figure 3.2.3). Their co-elution demonstrates that they might be present in the same complex at their physiological levels during hemoglobin switching. However, to understand how Sox6 and COUP-TFII interact during the hemoglobin switching producing opposite phenotypes is still an opened question.

The ability of COUP-TFII to upregulate γ-globin expression in β-K562 cells suggests that if we could understand its molecular mechanism of action we could consider COUP-TFII as a putative

candidate target to reactivate γ-globin gene expression in patients with β-hemoglobinopathies, such as β-thalassemia o Sickle Cell Disease (SCD). To better understand whether COUP-TFII can really reactivate γ-globin in adult erythroid cells, we moved to a better model, where the levels of γ-globin gene have been already mostly or completely silenced. Thus, we used erythroid cultures from $CD34^+$ cells purified from human peripheral blood, expanded *in vitro* and induced to differentiate in liquid culture. The results show that the forced expression of COUP-TFII in this human *ex vivo* erythroid cultures increases the γ-globin expression level, with no changes in the expression of the β-globin (Figure 3.2.8). In agreement with all data previously described, these results are also pointing to COUP-TFII as a prospective target whose modulation can increase γ-globin expression. Of interest, another orphan nuclear receptors (TR2/TR4) complex was initially described to repress ε and γ-globin expression *in vitro* binding to their promoter in erythroid cells, but studies with *in vivo* models of transgenic mice carrying the complete human β-locus, demonstrated that forced expression of TR2/TR4 produces increased levels of γ-globin and repression of ε-globin ***(Tanabe, McPhee, et al., 2007; Tanabe, Shen, et al., 2007)***. These controversial results could be due to different model system used: *in vitro* versus *in vivo/ex vivo* models, being closer to reproduce physiological situations.

During the last decades, the regulation of the proximal γ-globin promoter during embryonic development and during adult life has been under tremendous investigation. In the γ-globin promoter there are two CCAAT box regions, whose sequences are the most conserved motif found within the globin promoters. Mutations within the double CCAAT box region cause HPFH due to the altered binding of several TFs. Recently, in our lab, we have identified a Sox6 binding site between the two CCAAT boxes, what emphasizes the need of further studies of the TFs binding to this region. Moreover, COUP-TFII binds to the region of the proximal and distal CCAAT box overlapping with the binding of NF-Y. These two different binding sites have a different functional role. COUP-TFII cooperates with NF-Y when bound to the 5' site to form a more stable complex, thus providing greater transcriptional activity ***(Liberati et al., 2001)***. However, when COUP-TFII binds to 3' site overlapping with the CCAAT box, NF-Y and COUP-TFII might compete for the binding site. In this case, the binding of COUP-TFII could either fail to activate or even to repress transcriptional activation mediated by NF-Y, which is a strong activator.

One hypothesis to explain the conflictual roles of Sox6 and COUP-TFII in regarding the hemoglobin switching, is that, in the presence of high amount of Sox6, there is no cross talk between the two CCAAT boxes bound by the activator NF-Y. Thus, resulting in the repression of the γ-globin gene. In the absence of Sox6 or at equilibrium, the two CCAAT boxes, located at 22 bp of distance

(two turns of DNA) would serve as a high affinity site for the binding of NF-Y and therefore for the activation of γ-globin expression. Alternatively, high amount of COUP-TFII would displace Sox6 from its biding site and stabilized the binding of NF-Y complex, resulting in activation of γ-globin. In order to produce γ-globin silencing, Sox6 is also cooperating with BCL11a-XL ***(Xu et al., 2013; Xu et al., 2010)***, and they provably might recruit some chromatin remodeling factors helping to repress γ-globin expression.

All these data indicate that Sox6 and COUP-TFII have a complex role, where the relative balance between their levels might induce the switching from γ- (High COUP-TFII) to β- (High Sox6) globin expression.

3.3. The role of CDK13 together with sox6 in erythropoiesis.

INTRODUCTION:

Cyclin dependent kinases (CDKs) and their associated Cyclines are crucial proteins that control the cell cycle progression and RNA transcription. In the last years it has been discovered that CDKs are also related to the regulation of RNA splicing. The family of CDKs consists in 21 proteins, whose activity usually requires their association with other specific Cyclin subunits. CDK1, CDK2, CDK4 and CDK6 were the first CDKs characterized being cell cycle regulators. Then, cyclin/CDK complexes were described to have cell cycle independent activities. Among them, Cyclin H/CDK7 and cyclin T/CDK9 complexes have a role in the phosphorylation of the C-terminal domain of RNA polymerase and of other transcriptional regulators, such as DRB (5,6-dichloro-1-β-ribofuranosylbenzimidazole), the sensitivity including factor (DSIF) or the negative elongation factor (NELF). The lately CDK complexes identified are the transcription cycle-related complexes Cyclin k/Cyclin dependent kinase 12 (CycK/CDK12) and CycK/CDK13 ***(Blazek et al., 2011; Kohoutek & Blazek, 2012)***, that work as two independent complexes.

The human CDK12 (Cell division cycle 2 (cdc2)-related protein kinase 7, CrkRS) and CDK13 (Cell division cycle 2 (cdc2)-like protein kinase 5, CDC2L5) are homologous to the *Drosophila* CDK12. Both kinases were identified as cell cycle regulators ***(Ko, Kelly, & Pines, 2001; Marques et al., 2000)*** and they belong to a

high molecular weight subfamily of CDC2 family with PITAI/VRE motifs. Both kinases have a similar domain composition (Figure 3.3.1). They possess a conserved kinase CTD domain localized in the center, and their sequences are highly similar (>93%). They contain a PITAIRE motif at the conserved position of the PSTAIRE motif found in yeast cdc2 and in the related kinases. Like the cdc2 kinases, they contain characteristic threonine tyrosine residues at the beginning of the ATP-binding region, implicated in the kinase activity regulation. There are 20 and 17 arginine/serine rich (RS) motif in the N-terminus of CDK12 and CDK13, respectively. These motifs have a role in docking protein complexes, and they are characteristically found in splicing factors and regulators of splicing. Both kinases are found in the nuclear speckles, the characteristic subnuclear structures for mRNA splicing factors. In agreement with that, modulation of CDK12 and CDK13 expression affects alternative splicing of splicing reporter constructs. ***(Blazek et al., 2011; Chen, Wang, & Fann, 2006; Ko et al., 2001)***

Both kinases have a highly similar structure. Proline-rich motif (PRM) is present in the C-terminal region of both CDK12 and CDK13; and CDK12 carries an additional motif in the center of the protein, and CDK13 in the N-terminal region. This motif has a role in transcriptional regulation, RNA processing and alternative splicing ***(Sudol, Sliwa, & Russo, 2001)***. By the other site, alanine-rich motif (A) is only present in the N-terminus of CDK13 with unknown function. There are also several nuclear localization signals (NLS) present in both proteins. Although CDK12 and

CDK13 have a direct role in alternative splicing, the molecular mechanism of their activity is still not completely understood.

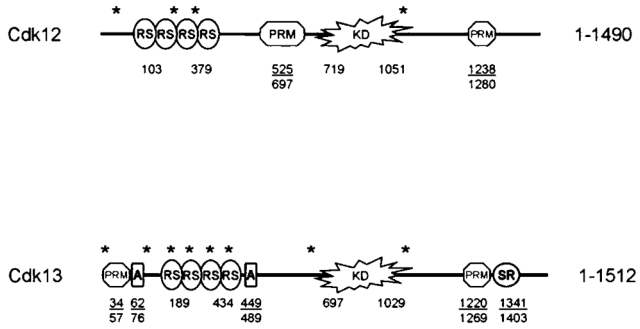


Figure 3.3.1 Domains composition of CDK12 and CDK13.
Schematic diagrams of CDK12 and CDK13 domains structure. Nuclear localization signals (NLS) are marked by asterisks. They contain Arginine/Serine-rich (RS), proline-rich (PRM), alanine-rich (A), serine-rich (SR) domains and kinase domain (KD). Numbers below the schemes indicate the amino acid position for a given domain ***(Kohoutek & Blazek, 2012)***.

As mentioned above, CDK12 and CDK13 are able to bind to overexpressed Cyclin L and regulate alternative splicing in vitro. However, recent studies reported that CDK12 and CDK13 do not bind to cyclin L *in vivo*, but rather with CycK ***(Bartkowiak et al., 2010; Blazek et al., 2011; de la Mata et al., 2003; de la Mata & Kornblihtt, 2006; Ko et al., 2001; Liang et al., 2015; Mortillaro et al., 1996)***. In particular, CDK13 has been implicated in the phosphorylation of the splicing factor ASF/SF2, and working together with the splicing factor SRSF1 to regulate alternative splicing of HIV ***(Berro et al., 2008)***. Furthermore, a recent study reported that the knockdown of CDK12 and CDK13 in mammalian cells produces RNA-processing defects ***(Liang et al., 2015)***.

Sox6 also has a role in alternative splicing, and it also co-localizes in the nuclear speckles domains with other splicing factors, and its disruption impairs several steps of the splicing process ***(Ohe, Lalli, & Sassone-Corsi, 2002)***. Moreover, in our lab we found evidences that Sox6 could mediate splicing of BCL11a. A recent study demonstrated that there is a dynamic alternative splicing program that regulates gene expression during erythroid terminal differentiation. In this study, Sox6 was found in different splicing forms during erythroblast differentiation ***(Pimentel et al., 2014)***. Thus, Sox6 has the ability to modulate the splicing process.

By the other side, CDK12 is able to phosphorylate the C-terminal domain (CTD) of RNA-polymerase II (RNAPII) *in vivo* and *in vitro*, indicating a role in transcription ***(Bartkowiak et al., 2010; Blazek et al., 2011)***. The functional relationship between CDK12 and CycK is supported by the common phenotype of genome instability upon inhibition of either CDK12 or CycK ***(Blazek et al., 2011; Kohoutek & Blazek, 2012)***. However, CDK12 was not co-immunoprecipitated with the RNA-polymerase II, suggesting that it does not belong to the core transcriptional apparatus. The phosphorylation of the CTD of the RNAPII by CDK13 has only been validated *in vitro*. So, the exact role of the CDK13/CycK complex is still unknown.

Missregulation of CDK12 or CDK13 has been associated with several diseases. Dysregulation of CDK12 was identified in

several cancers; and CDK12 is one of the most frequently somatically mutated genes in the most severe form of ovarian cancer. Moreover, some evidences point to CDK12 having an important role in breast cancer development. Also a fusion between CDK12 and ERBB2 proteins has been detected in gastric cancer ***(Kohoutek & Blazek, 2012)***.

The involvement of CDK13 in diseases is less clear. Interestingly, some patients with refractory anemia with ringed sideroblast associated with marked thrombocytosis (RARS-T) present an increased expression of CDK13 ***(Malcovati et al., 2009)***. Furthermore, there are some evidences that CDK13 promote megakaryocytic development in bone marrow cell cultures ***(Lapidot-Lifson et al., 1992)***.

Overall, CycK, CDK11, CDK12 and CDK13 were identified as putative partners of Sox6. However, only for CDK13 there are some indications that it might have a role in hematopoiesis, but nothing is known about its specific role in erythropoiesis. Therefore, we decided to study CDK13 to better understand its role during erythroid development.

RESULTS

CDK13 and Sox6 interact in human erythroid cell line

In the mass spectrometry results (described in chapter 2) from human erythroleukemic (HEL) cells we identified several peptides of different Cyclin dependent kinases (CycK, CDK11, CDK12 and CDK13). Between them the only kinase that seems to be implicated in the hematopoiesis regulation is CDK13. In order to validate the interaction between CDK13 and Sox6 we transduced HEL/BirA cells with Sox6 gene carrying the biotin tag and the FLAG epitope (bioSox6-FLAG). Three days after transduction, we performed a nuclear proteins extraction, followed by the co-Immunoprecipitation (Co-IP) analysis using the anti-Flag M2 affinity resin (Sigma). Then, we confirmed by Western blot that CDK13 physically interact with Sox6 in HEL cells overexpressing bioSox6-FLAG (Figure 3.3.2)

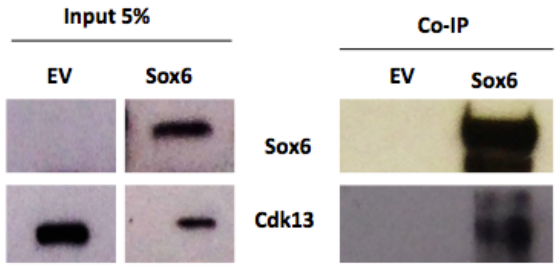


Figure 3.3.2 Physical interaction between Sox6 and CDK13.
Co-immunoprecipitation of bioSox6 and COUP-TFII proteins from HEL/BirA cells transduced with the empty vector (EV) and with bioSox6-FLAG. Immunoprecipitation was done using anti-FLAG M2 affinity gel. Co-purified proteins were analyzed by western blot by using anti-Sox6 and anti-CDK13 antibodies. 5% of the total nuclear lysate was used as input.

Profiling CDK13 expression in mouse primary fetal liver and bone marrow cells

As nothing is known about the role of CDK13 in erythropoiesis, we wanted to characterize its expression profile in hematopoiesis and erythropoiesis. We purified hematopoietic cells from *ex vivo* fetal liver populations at days E11.5, E12.5 and E13.5 of embryonic development (when the hemoglobin switching takes place). From the purified cells we obtain mRNA that we used to assess the expression levels of CDK13 during erythroid development. Figure 3.3.3a shows that the expression levels of CDK13 do not have significant change during mouse fetal liver embryonic development.

We also analyze the expression profile of CDK13 in *ex* vivo $CD34^+$ purified cells from bone marrow. We extracted total bone marrow from adult mice. The total population obtained was then purified for $CD34^+$ cells using magnetic beads (CD34 is a glycoprotein found in the cell surface of the early stage hematopoietic population). To obtain three different fractions (total bone marrow, $CD34^+$ and $CD34^-$ cells). From all these three fractions we purified mRNA that we then used to perform quantitative PCRs (RT-qPCR). The expression profile of CDK13 in the different fraction from the bone marrow reveals an enrichment of CDK13 expression in the $CD34^+$ population, which is the fraction containing the early erythroid progenitors (Figure 3.3.3b).

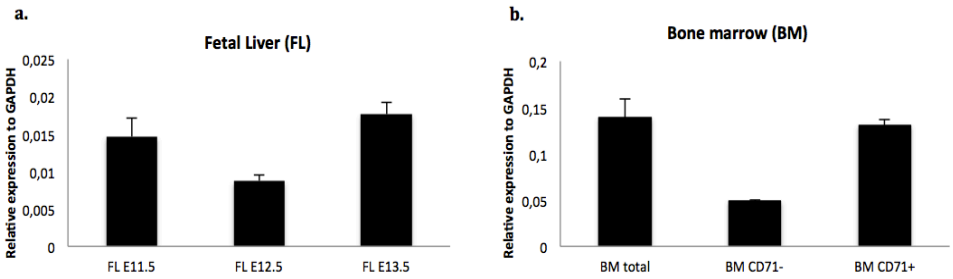


Figure 3.3.3 Expression profile of CDK13 in mouse primary fetal liver cells and in mouse bone marrow cells. *a) RT-qPCR analysis of the mRNA expression of CDK13 during mouse embryonic development (E11.5 – E13.5). b) RT-qPCR analysis of the mRNA expression of CDK13 in BONE MARROW cells. Total BM, $CD34^+$ and $CD34^-$ fractions are analyzed. There is an enrichment of the CDK13 expression in the $CD34^+$ fraction. The expression was analyzed by its relative expression to GAPDH.*

CDK13 downregulation in the β-K562 cell line forces erythroid differentiation

We used short hairpin RNA (shRNA) interference to examine the role of CDK13 in human erythroid cells.

RNAi is a useful tool to study mammalian genetics by expediting the interrogation of gene function. However, to produce more stable and efficient introduction of RNAi into cell cultures, or animal models, a more robust system is needed. Viral vectors, which can infect a wide variety of cell types and drive consistent transgene expression, provide a potential delivery vehicle for shRNA, an alternative form of double stranded RNA produced within the target cell. Within the cell the virally expressed shRNAs target endogenous genes inducing their silencing.

We cloned short-hairpin RNA (shRNA) oligonucleotides directed against human and mouse CDK13 (selected from H-silencer Select

Druggable Genome siRNA library V4, Ambion) into the lentilox 3.7 lentiviral (LLX) vector also carrying a GFP reporter (Figure 3.3.4a). Short sequences alignment show that the oligonucleotide was specifically targeting only CDK13, and not any of the other similar CDKs (Data not shown). The Knockdown virus (Cdk13_shRNA) reduced CDK13 protein expression in comparison to the empty vector (EV) in β-K562 cell line (Figure 3.3.4b). Abrogation of CDK13 expression in β-K562 cells was also confirmed by reduced CDK13 mRNA levels upon transduction of shRNA targeting for CDK13 (Figure 3.3.4c). We observed that the knockdown of CDK13 produces a block in proliferation between day 3 and 4 after transduction (Figure 3.3.4d).

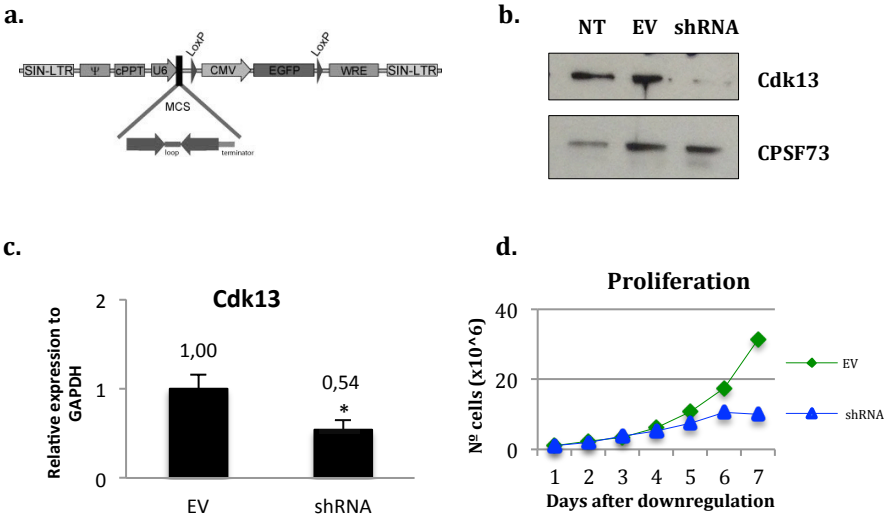


Figure 3.3.4 CDK13 downregulation in β-K562 cells. *a) Schematic representation of the Lentilox 3.7 vector used to downregulate CDK13. (Picture from http://www.sciencegateway.org) b) WB analysis of β-K562 cells nuclear extracts at day 3 after transduction shows downregulation of CDK13 protein. Anti-CPSF73 is used as a loading control. c) RT-qPCR analyses of the cDNA from β-K562 cells at day 3 after transduction confirm the downregulation of CDK13. The expression was analyzed by its relative expression to GAPDH (n=3; P<0,05). Asteriks (*) indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test. d) A block in proliferation occurs after 3-4 days after transduction.*

The overexpression of Sox6 enhances erythroid differentiation, producing cell cycle withdrawal and terminal maturation ***(Cantu et al., 2011)***. As the downregulation CDK13 produces a block in proliferation, we were wondering whether the cells stop growing due to their terminal differentiation or to other effects. In order to elucidate this point we analyzed the expression of erythroid genes related with terminal differentiation upon CDK13 knockdown. This analysis shows a reduction in the transferrin receptor (CD71)(Figure 3.3.5a), and an accumulation of glycophorin A (GpA; also referred as CD235) (Figure 3.3.5b). CD71 is an early erythroid marker for differentiation; meanwhile GpA is a late erythroid marker. Suggesting that downregulation of CDK13 forces late terminal erythroid differentiation. No major differences were found in the expression of the γ- and β-globin genes (Figure 3.3.5c). Several changes in the mRNA level of key enzymes of the heme biosynthetic pathway were also observed. There is a downregulation of ALA synthase (ALAS2) (Figure 3.3.6a), enzyme implicated in the first step of the heme biosynthesis pathway where it catalyzed the condensation of the succynil-CoA and glycine to form δ-aminolevulinic acid (ALA); ferrochelatase (FECH) is upregulated (Figure 3.3.6b). FECH is an enzyme implicated in the last step of the heme biosynthesis pathway catalyzing the reaction to incorporate ferrous ion to the protoporphyrin IX (PPIX) (Figure 3.3.6). These results are giving more evidence to the idea that downregulating CDK13 forces late terminal erythroid maturation. Of note, when we inhibit the expression of CDK13 we obtain highly similar erythroid

phenotypic effect as when we overexpress Sox6, with the exception of no direct effects on globins expression.

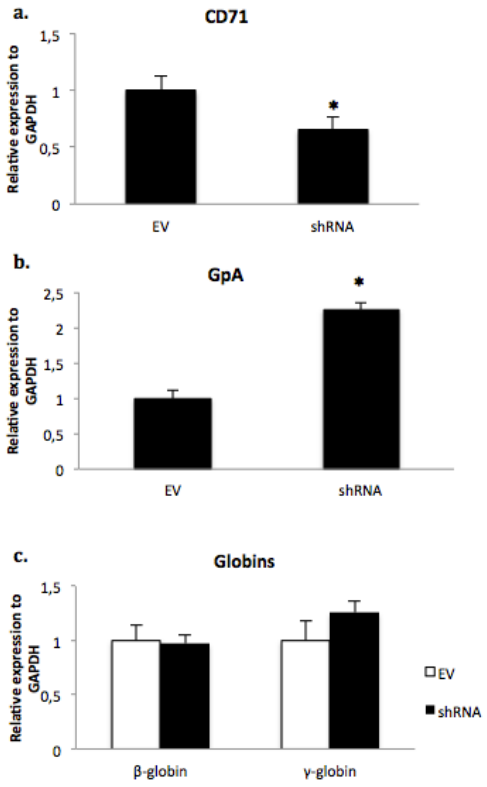


Figure 3.3.5 CDK13 knockdown forces erythroid terminal maturation in β-K562 cells. *RT-qPCR analyses of the cDNA from β-K562 cells at day 3 after transduction show a) downregulation of CD71, b) upregulation of GpA (CD235) and c) no major differences in the expression of γ- and β-globin, when compared transduced cells with the EV. The expression is relative expression to GAPDH (n=3; P<0,05). Asterisks (*) Indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test.*

Thus, we decided to analyze the expression of endogenous Sox6 upon CDK13 downregulation in the β-K562 cells. Interestingly, we found that Sox6 was greatly upregulated (2.95 times upregulated in comparison to the EV) (Figure 3.3.7a). As Sox6 induces

erythroid differentiation, we checked the expression of genes regulated by Sox6 when CDK13 is repressed. A direct target of Sox6 is the Suppressor Of Cytokine Signaling 3 (SOCS3). Interestingly, we found it positively regulated (2.78 times upregulated in comparison with the EV) (Figure 3.3.7b). Sox6 binds to the SOCS3 promoter positively regulating its expression in a dose-dependent manner and the direct SOCS3 activation by Sox6 block cell proliferation inhibiting the phosphorylation of STAT5B ***(Cantu et al., 2011)***.

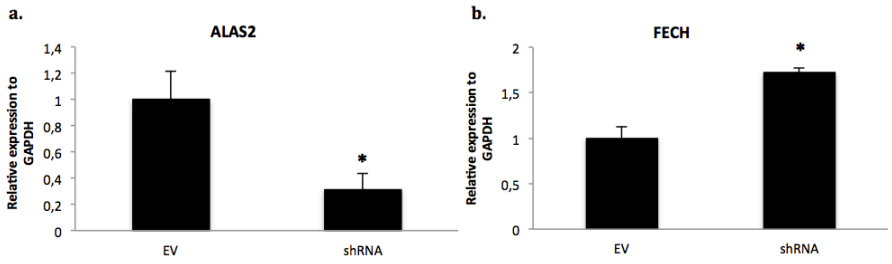


Figure 3.3.6 CDK13 knockdown alterates the heme pathway *RT-qPCR analyses of the cDNA from β-K562 cells at day 3 after transduction show a) downregulation of ALAS2, and b) upregulation of FECH, when compared transduced cells with the EV. The expression is relative expression to GAPDH (n=3; P<0,05). Asterisks (*) Indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test.*

The effect of CDK13 downregulation is controversial on the expression of the anti-apoptotic genes Bcl-2 and Bcl-xL. There is a reduction in the gene expression of Bcl-2 (Figure 3.3.8a), but a slightly increase in the Bcl-xL gene expression (Figure 3.3.8b). Accordingly with previous studies, SOCS3 has the ability to induce pro-apoptotic phenotype downregulating anti-apoptotic genes,

but in late erythroid differentiation is known that Bcl-xL expression is upregulated. This effect is also produced in erythroid cells overexpressing Sox6, since Bcl-XL is a direct Sox6 target ***(Dumitriu et al., 2010)***.

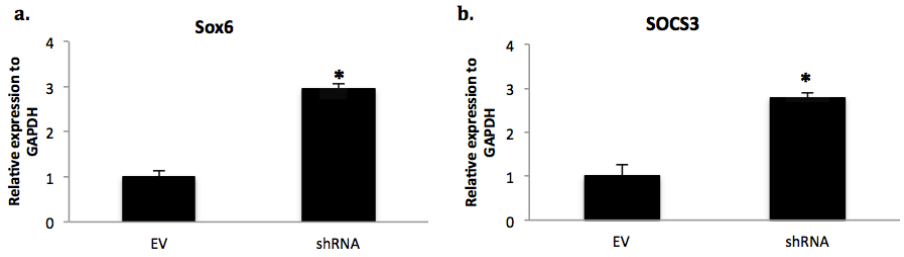


Figure 3.3.7 CDK13 knockdown upregulates Sox6 and SOCS3. *RT-qPCR analyses of the cDNA from β-K562 cells at day 3 after transduction show a) upregulation of Sox6 (2.95x compared to the EV), and b) upregulation of SOCS3 (2.78x compared with the EV). The expression is relative expression to GAPDH (n=3; P<0,05). Asterisks (*) Indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test.*

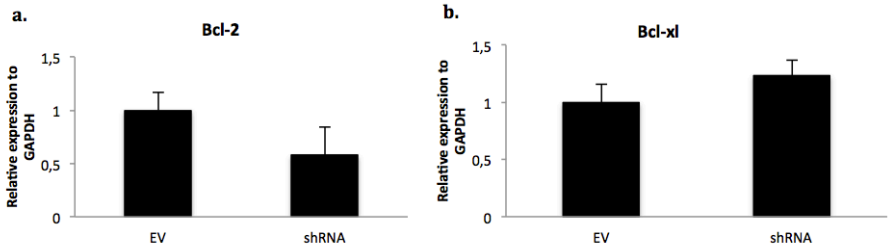


Figure 3.3.8 CDK13 knockdown alterates anti-apoptotic genes. *RT-qPCR analyses of the cDNA from β-K562 cells at day 3 after transduction show a) downregulation of Bcl-2, however there is no statistical significance in comparison to the EV; and b) slight upregulation of Bcl-xL, with not statistical significance when compared transduced cells to the EV. The expression is relative expression to GAPDH (n=3; P<0,05). Asterisks (*) Indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test.*

All these data suggest that CDK13 might be necessary to maintain the undifferentiated state of the cells. Since its downregulation enhances the genetic pathway characteristic of erythroid differentiation. However, as CDK13 was described to have a role in megakaryocytic development, we were wondering whether the abrogation of CDK13 might produce any differential expression of the megakaryocytic markers of differentiation.

CDK13 downregulation in the β-K562 cell line inhibits megakaryocytic differentiation.

CDK13 has been reported to have a role promoting megakaryocytic development ***(Lapidot-Lifson et al., 1992)***; with this information and with the previous data showing that CDK13 downregulation forces erythroid differentiation arises the question whether megakaryocytic markers would be downregulated upon CDK13 inhibition. Interestingly, the expression of some megakaryocytic markers decreased upon CDK13 downregulation. Although, the integrin platelet glycoprotein IIb/IIIa ($\alpha_{IIb}\beta_3$) complex (encoded by GpIIB and GpIIIa genes) is only slightly decreased during CDK13 downregulation (Figure 3.3.9a/b), the proto-oncogene, Fli-1 (Friend of leukemia virus integration 1) transcription factor is significantly reduced upon the CDK13 abrogation (Figure 3.3.9c).

The activation of GpIIB/GpIIIa complex induces platelet-platelet interaction through binding of soluble fibrinogen, and after controls hemostasis and thrombosis to instigate wound healing

after vascular damage. Fli-1 coordinates megakaryocyte development, and it is upregulated in the late stage of maturation. Mouse Fli-1 null embryos and Fli-1 null mouse primary cells cultures show defects in megakaryocytic development.

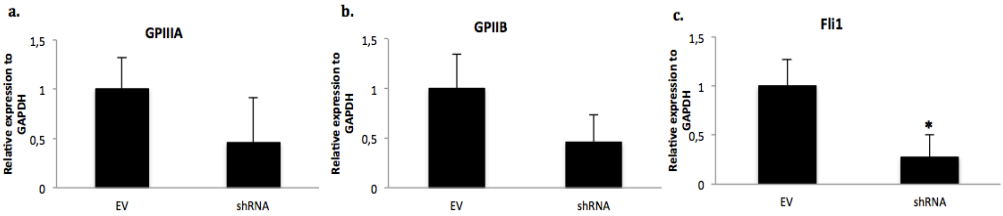


Figure 3.3.9 CDK13 knockdown downregulates megakaryocytic markers. *RT-qPCR analyses, of the cDNA from β-K562 cells at day 3 after transduction, show downreglation in several megakaryocytic markers. a) downregulation of GPIIIA and b) GPIIB, with not statistical significance when compared transduced cells to the EV. The expression is relative expression to GAPDH (n=3; P<0,05). Asterisks (*) Indicates statistical significance. Error bars indicate SEM. Statistical analysis was performed by using the paired Student t test.*

This data suggests that CDK13 might be necessary to maintain a undifferentiated state in early erythropoiesis. Therefore, downregulation of CDK13 forces erythroid maturation and blocks megakaryocytosis. Equilibrium between the CDK13 and Sox6 expression might be necessary to maintain the immature state in erythroid cells.

DISCUSSION

Among the CDKs identified by mass spectrometry sequencing (CycK, CDK11, CDK12 and CDK13) as potential Sox6 partners, the most interesting one is CDK13. The expression on this kinase is upregulated in several patients with ringed sideroblasts anemia associated with marked thrombocytosis (RARS-T) ***(Malcovati et al., 2009)*** and it also promotes megakaryocytic development in bone marrow cell cultures ***(Lapidot-Lifson et al., 1992)***.

In this chapter we validated the interaction between Sox6 and CDK13 in human erythroleukemic (HEL) cells overexpressing bioSox6-FLAG (Figure 3.3.2).

In order to characterize the expression profile of CDK13 in erythropoiesis we analyzed its expression in mouse fetal liver cells during embryonic developmental stages E11.5, E12.5 and E13.5 (Figure 3.3.3a), However, we did not found any significant change during hemoglobin switching. This suggests that this kinase is not playing a specific role in regulating hemoglobin switching. In link with this, downregulation of CDK13 in β-K562 cells has no effect in either γ- or β-globin genes (Figure 3.3.5c). Of note, CDKs are ubiquitously expressed kinases and play different roles regulating gene expression. Thus, the fact that CDK13 has not a specific profile during mouse embryonic development seems to be in agreement with their general functions in regulating gene expression.

In order to understand the role of CDK13 in erythroid cells we downregulated CDK13 expression in β-K562 cells using a lentivector (Lentilox 3.7 vector; Figure 3.3.4a) carrying a short hairpin RNA (shRNA) targeting CDK13. Surprisingly, the phenotype of β-K562 depleted of CDK13 is remarkably similar to that of cells overexpressing Sox6. In particular, the knockdown of CDK13 enhances late erythroid maturation. What was demonstrated by the upregulation of several genes, such as glycophorin A (GpA, also known as CD235)(Figure 3.3.5b) and ferrochelatase (FECH)(Figure 3.3.6b).

These data together with the decreased expression levels of transferrin receptor (CD71) (Figure 3.3.5a) and the enzyme ALA synthase (ALAS2) (Figure 3.3.6a), which are upregulated in the first steps of differentiation and later downregulated, demonstrate that the abrogation of CDK13 forces late erythroid maturation. Interestingly, also Sox6 and its direct target SOCS3 are upregulated when CDK13 is knocked down (Figure 3.3.7). Furthermore, the proapoptotic genes Bcl-2 and Bcl-xL behave as when Sox6 is overexpressed (Figure 3.3.8) with reduced levels of Bcl-2 and slight increase in Bcl-xL gene expression. ***(Cantu et al., 2011; Dumitriu et al., 2010)***.

CDK13 has been described to promote megakaryocytic development in human bone marrow cell cultures. Among the megakaryocytic markers, the GpIIB/GpIIIA complex was slightly decreased, and the proto-Oncogene Fli-1 was significantly

reduced upon the CDK13 abrogation (Figure 3.3.9). Of interest, the constitutive activation of Fli-1 in erythroblasts leads to a dramatic shift in the EPO/EPO-R signal transduction pathway, blocking erythroid differentiation, activating the Ras pathway, and resulting in massive EPO-independent proliferation of erythroblasts ***(Tamir et al., 1999; Zochodne et al., 2000)***. Thus, it seems that CDK13 is necessary to maintain an immature state in the erythroid cells, since its abrogation forces erythroid maturation and blocks megakaryocytic development.

The results obtained in this chapter suggest that the endogenous levels of Sox6 and CDK13 might contribute to maintain the equilibrium between proliferation and differentiation. The overexpression of Sox6 (when large number of Sox6 molecules are free, and several of them might not interact with CDK13) or the downregulation of CDK13 (when there are not enough molecules of CDK13 to titer all Sox6 molecules), breaks this equilibrium and forces the cells to differentiate toward the erythroid direction as demonstrated by the increased levels of the erythroid genes necessary for the terminal maturation and decreased of the megakaryocytic genes implicated in megakaryopoiesis.

Several hypotheses might explain the mechanism of action between Sox6 and CDK13.

i) As both, CDK13 and Sox6 are alternative splicing modulators ***(Berro et al., 2008; Chen, Wong, Geneviere, & Fann, 2007; Even***

et al., 2006; Ohe et al., 2002), it might be possible that they collaborate in maintaining the mRNA processing status of undifferentiated common megakaryocyte/erythrocyte progenitors (MEP)

ii) CDK13 phosphorylates the C-terminal domain (CTD) of the RNA polymerase II (RNAPII), what, in turn, lead the transcription machinery to start the elongation process and the following mRNA processing. Sox6/CDK13 balance might maintain the transcriptional machinery active for undifferentiated genes.

iv) Finally, it might be also possible that the interaction between Sox6 and CDK13 maintain the proliferation state, characteristically active in undifferentiated cells. However, this is unlikely because decreased number of CDK13 or increased number of Sox6 molecules induces cell cycle arrest, probably by the upregulation of SOCS3, the direct target of Sox6.

None of theses hypothesis are unique and exclusives and they are not incompatible one with each other.

To further confirm the postulated hypotheses, several studies are still necessary. We will set up experiments of loss of function of CDK13 in a megakaryocytic cell line, to observe whether downregulation of CDK13 is also able to enhance erythroid differentiation in this cell line. Furthermore, we are cloning the cDNA of CDK13 in a lentiviral vector to overexpress it in erythroid and megakaryocytic cell lines to test whether CDK13

overexpression is able to force differentiation toward the megakaryocytic lineage.

REFERENCES

Aerbajinai, W., Zhu, J., Kumkhaek, C., Chin, K., & Rodgers, G. P. (2009). SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII. *Blood, 114*(1), 187-194. doi: 10.1182/blood-2008-07-170712

Akinsheye, I., Alsultan, A., Solovieff, N., Ngo, D., Baldwin, C. T., Sebastiani, P., . . . Steinberg, M. H. (2011). Fetal hemoglobin in sickle cell anemia. *Blood, 118*(1), 19-27. doi: 10.1182/blood-2011-03-325258

Allen, H. F., Wade, P. A., & Kutateladze, T. G. (2013). The NuRD architecture. *Cell Mol Life Sci, 70*(19), 3513-3524. doi: 10.1007/s00018-012-1256-2

Amaya, M., Desai, M., Gnanapragasam, M. N., Wang, S. Z., Zu Zhu, S., Williams, D. C., Jr., & Ginder, G. D. (2013). Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells. *Blood, 121*(17), 3493-3501. doi: 10.1182/blood-2012-11-466227

Bank, A. (2006). Regulation of human fetal hemoglobin: new players, new complexities. *Blood, 107*(2), 435-443. doi: 10.1182/blood-2005-05-2113

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., . . . Greenleaf, A. L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev, 24*(20), 2303-2316. doi: 10.1101/gad.1968210

Berro, R., Pedati, C., Kehn-Hall, K., Wu, W., Klase, Z., Even, Y., . . . Kashanchi, F. (2008). CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. *J Virol, 82*(14), 7155-7166. doi: 10.1128/JVI.02543-07

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., . . . Peterlin, B. M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev, 25*(20), 2158-2172. doi: 10.1101/gad.16962311

Cantu, C., Ierardi, R., Alborelli, I., Fugazza, C., Cassinelli, L., Piconese, S., . . . Ronchi, A. (2011). Sox6 enhances erythroid

differentiation in human erythroid progenitors. *Blood, 117*(13), 3669-3679. doi: 10.1182/blood-2010-04-282350

Chen, H. H., Wang, Y. C., & Fann, M. J. (2006). Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol Cell Biol, 26*(7), 2736-2745. doi: 10.1128/MCB.26.7.2736-2745.2006

Chen, H. H., Wong, Y. H., Geneviere, A. M., & Fann, M. J. (2007). CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochem Biophys Res Commun, 354*(3), 735-740. doi: 10.1016/j.bbrc.2007.01.049

Cooney, A. J., Tsai, S. Y., O'Malley, B. W., & Tsai, M. J. (1992). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol, 12*(9), 4153-4163.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., . . . Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A, 100*(13), 7480-7485. doi: 10.1073/pnas.1332608100

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., . . . Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell, 12*(2), 525-532.

de la Mata, M., & Kornblihtt, A. R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol, 13*(11), 973-980. doi: 10.1038/nsmb1155

Demaison, C., Parsley, K., Brouns, G., Scherr, M., Battmer, K., Kinnon, C., . . . Thrasher, A. J. (2002). High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. *Hum Gene Ther, 13*(7), 803-813. doi: 10.1089/10430340252898984

Dumitriu, B., Bhattaram, P., Dy, P., Huang, Y., Quayum, N., Jensen, J., & Lefebvre, V. (2010). Sox6 is necessary for efficient erythropoiesis in adult mice under physiological and

anemia-induced stress conditions. *PLoS One, 5*(8), e12088. doi: 10.1371/journal.pone.0012088

Even, Y., Durieux, S., Escande, M. L., Lozano, J. C., Peaucellier, G., Weil, D., & Geneviere, A. M. (2006). CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *J Cell Biochem, 99*(3), 890-904. doi: 10.1002/jcb.20986

Fibach, E., Manor, D., Oppenheim, A., & Rachmilewitz, E. A. (1989). Proliferation and maturation of human erythroid progenitors in liquid culture. *Blood, 73*(1), 100-103.

Filipe, A., Li, Q., Deveaux, S., Godin, I., Romeo, P. H., Stamatoyannopoulos, G., & Mignotte, V. (1999). Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *EMBO J, 18*(3), 687-697. doi: 10.1093/emboj/18.3.687

Ginder, G. D., Gnanapragasam, M. N., & Mian, O. Y. (2008). The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development. *Curr Top Dev Biol, 82*, 85-116. doi: 10.1016/S0070-2153(07)00004-X

Gnanapragasam, M. N., Scarsdale, J. N., Amaya, M. L., Webb, H. D., Desai, M. A., Walavalkar, N. M., . . . Williams, D. C., Jr. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A, 108*(18), 7487-7492. doi: 10.1073/pnas.1015341108

Grosveld, F., Dillon, N., & Higgs, D. (1993). The regulation of human globin gene expression. *Baillieres Clin Haematol, 6*(1), 31-55.

Harju-Baker, S., Costa, F. C., Fedosyuk, H., Neades, R., & Peterson, K. R. (2008). Silencing of Agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. *Mol Cell Biol, 28*(10), 3101-3113. doi: 10.1128/MCB.01858-07

Higgs, D. R., Engel, J. D., & Stamatoyannopoulos, G. (2012). Thalassaemia. *Lancet, 379*(9813), 373-383. doi: 10.1016/S0140-6736(11)60283-3

Hong, W., Nakazawa, M., Chen, Y. Y., Kori, R., Vakoc, C. R., Rakowski, C., & Blobel, G. A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional

repression by GATA-1. *EMBO J, 24*(13), 2367-2378. doi: 10.1038/sj.emboj.7600703

Kamachi, Y., Uchikawa, M., & Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet, 16*(4), 182-187.

Kiefer, J. C. (2007). Back to basics: Sox genes. *Dev Dyn, 236*(8), 2356-2366. doi: 10.1002/dvdy.21218

Kliewer, S. A., Umesono, K., Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., & Evans, R. M. (1992). Retinoid X receptor-COUP-TF interactions modulate retinoic acid signaling. *Proc Natl Acad Sci U S A, 89*(4), 1448-1452.

Ko, T. K., Kelly, E., & Pines, J. (2001). CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci, 114*(Pt 14), 2591-2603.

Kohoutek, J., & Blazek, D. (2012). Cyclin K goes with Cdk12 and Cdk13. *Cell Div, 7*, 12. doi: 10.1186/1747-1028-7-12

Lai, A. Y., & Wade, P. A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer, 11*(8), 588-596. doi: 10.1038/nrc3091

Lapidot-Lifson, Y., Patinkin, D., Prody, C. A., Ehrlich, G., Seidman, S., Ben-Aziz, R., . . . Soreq, H. (1992). Cloning and antisense oligodeoxynucleotide inhibition of a human homolog of cdc2 required in hematopoiesis. *Proc Natl Acad Sci U S A, 89*(2), 579-583.

Lauberth, S. M., & Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J Biol Chem, 281*(33), 23922-23931. doi: 10.1074/jbc.M513461200

Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., & Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol, 39*(12), 2195-2214. doi: 10.1016/j.biocel.2007.05.019

Li, Q., Duan, Z. J., & Stamatoyannopoulos, G. (2001). Analysis of the mechanism of action of non-deletion hereditary persistence of fetal hemoglobin mutants in transgenic mice. *EMBO J, 20*(1-2), 157-164. doi: 10.1093/emboj/20.1.157

Liang, K., Gao, X., Gilmore, J. M., Florens, L., Washburn, M. P., Smith, E., & Shilatifard, A. (2015). Characterization of human

cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol Cell Biol, 35*(6), 928-938. doi: 10.1128/MCB.01426-14

Liberati, C., Cera, M. R., Secco, P., Santoro, C., Mantovani, R., Ottolenghi, S., & Ronchi, A. (2001). Cooperation and competition between the binding of COUP-TFII and NF-Y on human epsilon- and gamma-globin gene promoters. *J Biol Chem, 276*(45), 41700-41709. doi: 10.1074/jbc.M102987200

Liberati, C., Ronchi, A., Lievens, P., Ottolenghi, S., & Mantovani, R. (1998). NF-Y organizes the gamma-globin CCAAT boxes region. *J Biol Chem, 273*(27), 16880-16889.

Lin, A. C., Roche, A. E., Wilk, J., & Svensson, E. C. (2004). The N termini of Friend of GATA (FOG) proteins define a novel transcriptional repression motif and a superfamily of transcriptional repressors. *J Biol Chem, 279*(53), 55017-55023. doi: 10.1074/jbc.M411240200

Malcovati, L., Della Porta, M. G., Pietra, D., Boveri, E., Pellagatti, A., Galli, A., . . . Cazzola, M. (2009). Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis. *Blood, 114*(17), 3538-3545. doi: 10.1182/blood-2009-05-222331

Marques, F., Moreau, J. L., Peaucellier, G., Lozano, J. C., Schatt, P., Picard, A., . . . Geneviere, A. M. (2000). A new subfamily of high molecular mass CDC2-related kinases with PITAI/VRE motifs. *Biochem Biophys Res Commun, 279*(3), 832-837. doi: 10.1006/bbrc.2000.4042

Mortillaro, M. J., Blencowe, B. J., Wei, X., Nakayasu, H., Du, L., Warren, S. L., . . . Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A, 93*(16), 8253-8257.

Ohe, K., Lalli, E., & Sassone-Corsi, P. (2002). A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proc Natl Acad Sci U S A, 99*(3), 1146-1151. doi: 10.1073/pnas.022645899

Palstra, R. J., de Laat, W., & Grosveld, F. (2008). Beta-globin regulation and long-range interactions. *Adv Genet, 61*, 107-142. doi: 10.1016/S0065-2660(07)00004-1

Pegoraro, G., Kubben, N., Wickert, U., Gohler, H., Hoffmann, K., & Misteli, T. (2009). Ageing-related chromatin defects through loss of the NURD complex. *Nat Cell Biol, 11*(10), 1261-1267. doi: 10.1038/ncb1971

Pimentel, H., Parra, M., Gee, S., Ghanem, D., An, X., Li, J., . . . Conboy, J. G. (2014). A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res, 42*(6), 4031-4042. doi: 10.1093/nar/gkt1388

Ronchi, A., Bellorini, M., Mongelli, N., & Mantovani, R. (1995). CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA. *Nucleic Acids Res, 23*(22), 4565-4572.

Rupon, J. W., Wang, S. Z., Gaensler, K., Lloyd, J., & Ginder, G. D. (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), 6617-6622. doi: 10.1073/pnas.0509322103

Rupon, J. W., Wang, S. Z., Gnanapragasam, M., Labropoulos, S., & Ginder, G. D. (2011). MBD2 contributes to developmental silencing of the human epsilon-globin gene. *Blood Cells Mol Dis, 46*(3), 212-219. doi: 10.1016/j.bcmd.2011.01.001

Rusconi, F., Guillonneau, F., & Praseuth, D. (2002). Contributions of mass spectrometry in the study of nucleic acid-binding proteins and of nucleic acid-protein interactions. *Mass Spectrom Rev, 21*(5), 305-348. doi: 10.1002/mas.10036

Sankaran, V. G., Menne, T. F., Xu, J., Akie, T. E., Lettre, G., Van Handel, B., . . . Orkin, S. H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science, 322*(5909), 1839-1842. doi: 10.1126/science.1165409

Sankaran, V. G., Xu, J., & Orkin, S. H. (2010a). Advances in the understanding of haemoglobin switching. *Br J Haematol, 149*(2), 181-194. doi: 10.1111/j.1365-2141.2010.08105.x

Sankaran, V. G., Xu, J., & Orkin, S. H. (2010b). Transcriptional silencing of fetal hemoglobin by BCL11A. *Ann N Y Acad Sci, 1202*, 64-68. doi: 10.1111/j.1749-6632.2010.05574.x

Schuh, C., Kinast, E. J., Mezzomo, E., & Kapczinski, M. P. (2005). Effect of glazed and polished surface finishes on the friction

coefficient of two low-fusing ceramics. *J Prosthet Dent, 93*(3), 245-252. doi: 10.1016/j.prosdent.2004.12.010

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J. C., Thongjuea, S., Stadhouders, R., . . . Grosveld, F. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev, 24*(3), 277-289. doi: 10.1101/gad.551810

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Exp Hematol, 33*(3), 259-271. doi: 10.1016/j.exphem.2004.11.007

Stamatoyannopoulos, G., Blau, C. A., Nakamoto, B., Josephson, B., Li, Q., Liakopoulou, E., . . . Dover, G. (1994). Fetal hemoglobin induction by acetate, a product of butyrate catabolism. *Blood, 84*(9), 3198-3204.

Steinberg, M. H. (2002). Hydroxyurea treatment for sickle cell disease. *ScientificWorldJournal, 2*, 1706-1728. doi: 10.1100/tsw.2002.295

Steinberg, M. H. (2006). Pathophysiologically based drug treatment of sickle cell disease. *Trends Pharmacol Sci, 27*(4), 204-210. doi: 10.1016/j.tips.2006.02.007

Steinberg, M. H., Nagel, R. L., & Brugnara, C. (1997). Cellular effects of hydroxyurea in Hb SC disease. *Br J Haematol, 98*(4), 838-844.

Sudol, M., Sliwa, K., & Russo, T. (2001). Functions of WW domains in the nucleus. *FEBS Lett, 490*(3), 190-195.

Tamir, A., Howard, J., Higgins, R. R., Li, Y. J., Berger, L., Zacksenhaus, E., . . . Ben-David, Y. (1999). Fli-1, an Ets-related transcription factor, regulates erythropoietin-induced erythroid proliferation and differentiation: evidence for direct transcriptional repression of the Rb gene during differentiation. *Mol Cell Biol, 19*(6), 4452-4464.

Tanabe, O., McPhee, D., Kobayashi, S., Shen, Y., Brandt, W., Jiang, X., . . . Engel, J. D. (2007). Embryonic and fetal beta-globin gene repression by the orphan nuclear receptors, TR2 and TR4. *EMBO J, 26*(9), 2295-2306. doi: 10.1038/sj.emboj.7601676

Tanabe, O., Shen, Y., Liu, Q., Campbell, A. D., Kuroha, T., Yamamoto, M., & Engel, J. D. (2007). The TR2 and TR4 orphan nuclear receptors repress Gata1 transcription. *Genes Dev, 21*(21), 2832-2844. doi: 10.1101/gad.1593307

Weatherall, D. J., & Clegg, J. B. (1996). Thalassemia--a global public health problem. *Nat Med, 2*(8), 847-849.

Xu, J., Bauer, D. E., Kerenyi, M. A., Vo, T. D., Hou, S., Hsu, Y. J., . . . Orkin, S. H. (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A, 110*(16), 6518-6523. doi: 10.1073/pnas.1303976110

Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W., & Orkin, S. H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev, 24*(8), 783-798. doi: 10.1101/gad.1897310

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J., & Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell, 2*(6), 851-861.

Zhu, X., Wang, Y., Pi, W., Liu, H., Wickrema, A., & Tuan, D. (2012). NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human gamma-globin gene. *PLoS One, 7*(10), e47175. doi: 10.1371/journal.pone.0047175

Zochodne, B., Truong, A. H., Stetler, K., Higgins, R. R., Howard, J., Dumont, D., . . . Ben-David, Y. (2000). Epo regulates erythroid proliferation and differentiation through distinct signaling pathways: implication for erythropoiesis and Friend virus-induced erythroleukemia. *Oncogene, 19*(19), 2296-2304. doi: 10.1038/sj.onc.1203590

Chapter 4

A multiplex high-content assay for quantification the γ- and β-globin content at the single cell level.

INTRODUCTION

Imbalance or alterations in the production of hemoglobin chains due to mutations or deletions in globin genes and/or in their regulators can be deleterious for RBCs and their precursors and may lead to Hemoglobinopathies ***(J. A. Stamatoyannopoulos, 1992; J. A. Stamatoyannopoulos & Nienhuis, 1992)***. β-hemoglobinopathies (β-thalassemia and Sickle cell disease, SCD) are the most common inherited monogenic diseases affecting million people in the world are and represent a significant health problem in more than 160 countries worldwide with over 330,000 affected infants born annually (83% sickle cell disease, 17% thalassemia).

There is a group of genetic conditions, called hereditary persistence of fetal hemoglobin (HPFH), where the expression of the fetal γ-globin gene persists at high levels in adult erythroid cells.

Clinical observations show that coinheritance of HPFH with β-hemoglobinopathies significantly ameliorates the clinical phenotype of these patients by reducing sickle hemoglobin polymers in SCD and the α/non-α chain imbalance in β-Thalassemia ***(Forget, 1998; G. Stamatoyannopoulos et al., 1994; Thein, 1998; Weatherall, Higgs, et al., 1981; Weatherall, Pressley, Wood, Higgs, & Clegg, 1981)***.

"Linked" HPFH are the result of a variety of mutations or deletions of different sizes involving the β-globin locus. "Non-linked" HPFH are caused by mutations outside the β-globin gene cluster.

On the above bases, many efforts have been made to increase γ-globin expression in β-thalassemic and HbS patients However only few drugs (with limited efficacy) have been identified so far. Among them, Hydroxyurea (HU) has been approved for the treatment of SCD and β-thalassemia, but its efficacy varies among patients. Indeed, approximately half of the patients do not reach acceptable levels of HbF at HU doses of acceptable toxicity ***(Banan, 2013; Platt, 2008; Ronchi & Ottolenghi, 2013; Steinberg, 2002; Steinberg, Nagel, & Brugnara, 1997)***. Other drugs, such as butyrate and its derivates, 5-azacytabine, Decitabine and Tranylcypromine act on the epigenetic regulation od HbF, by inhibiting histone deacetylation or methylation on the γ-globin gene, but their efficacy is still very limited ***(Banan, 2013; Musallam, Cappellini, & Taher, 2013; Musallam, Rivella, Vichinsky, & Rachmilewitz, 2013; Musallam, Taher, Cappellini, & Sankaran, 2013; Perrine et al., 2010).***

The poor results obtained with these drugs encourage the study of new therapeutical approaches with a particular attention molecules that could modulate the γ-globin promoter activity.

In collaboration with Dr. Fabio Gasparri and Marta Durlak from Nerviano Medical Science (NMS) in Milano (Italy), we set up a high content screening platform based on multiplex imagining on a subclone of the human erytholeucemic K562 cell line expressing significant levels of β-globin (β-K562). Simultaneous analysis of DNA content, together with Adult Hemoglobin HbA ($\alpha_2\beta_2$) and Fetal Hemoglobin HbF ($\alpha_2\gamma_2$) immunodetection, resulted in a robust and sensitive assay, capable of detecting changes in hemoglobinization and in γ/β ratio in response to both, siRNA (small interference RNA) transfections and drugs treatments targeting genes that could affect hemoglobinitzation (***Durlak M. et al. in preparation***).

Taking advantage of this novel platform, we applied it to carry out a high-content screening for several of the most interesting partners of Sox6, with the final goal of discovering novel genes affecting hemoglobinitzation and thus to find novel putative therapeutic targets that could be exploited to reactivate γ-globin expression.

EXPERIMENTAL PROCEDURES

Cell lines and chemical treatments:

ECACC-K562 (European Collection of Cell Cultures) and β-K562 (a kind gift of Prof. G. Ferrari, HSR, Milan) were grown in standard conditions. Doubling times were calculated on cells growing in exponential phase. For morphological analysis, K562 and β-K562 cells were hematoxylin/eosin stained and images were acquired and processed on an Axioplan2 microscope (Zeiss, Thornwood, NY). β-K562 authentication was obtained by short tandem repeat (STR) fingerprinting by using the AmpFlSTR Identifiler Plus PCR Amplification kit (Applied Biosystems). For chemical treatments, 5 x 10^4 cells were exposed to increasing doses of the drug of interest in 24-well plates. Four days after treatment, cells were analyzed by RTqPCR or high-content analysis. All experiments were performed in triplicates (at least two technical replicates per experiment).

siRNA oligonucleotide transfections:

β-K562 cells were transfected with siRNA oligonucleotides (H-Silencer Select Druggable Genome siRNA Library V4, Ambion). A siRNA oligo targeting Eg5, a kinesin related with the mitotic spindle, and a non-targeting oligo (NTO) were used as positive and negative controls for transfection. siRNA oligonucleotides were transfected by using lipofectamine® RNAiMAX (Invitrogen)

siRNA oligos	**Sequence (5'-3')**
NTO	UGGUUUACAUGUCGACUAAtt
Eg5	CUGAAGACCUGAAGACAAUtt
HBB (oligo 1)	GAAAGUGCUCGGUGCCUUUtt
HBB (oligo 2)	AGGUGAACGUGGAUGAAGUtt
HBG1 (oligo 1)	GUCUACCCAUGGACCCAGAtt
HBG1 (oligo 2)	UGACCGUUUUGGCAAUCCAtt
MTA1 (oligo 1)	CACCGACUUGUUAAAAGAAtt
MTA1 (oligo 2)	GCAUCUUGUUGGACAUAUUtt
MTA1 (oligo 3)	GGAGAGAUUCGAGUAGGAAtt
HDAC1 (oligo 1)	CCGGUCAUGUCCAAAGUAAtt
HDAC1 (oligo 2)	CCAAUAUGACUAACCAGAAtt
HDAC1 (oligo 3)	CUAUGGUCUCUACCGAAAAtt
SMARCA5 (Oligo 1)	GGAGAUACUUAGUAAUAGAtt
SMARCA5 (Oligo 2)	GGGCAAAUAGAUUCGAGUAtt
SMARCA5 (Oligo 3)	GGGCGAAAGUUCACUUAGAtt
SMARCC2 (Oligo 1)	GCAAUGCACCGCUCACUAAtt
SMARCC2 (Oligo 2)	CCUCAACACCUUACACUAAtt
SMARCC2 (Oligo 3)	GCUACUAUCCUGACAGUUAtt
BHLHB2 (Oligo 1)	CAAACCUAAUUGAUCAGCAtt
BHLHB2 (Oligo 2)	CGAACAUCUCAAACUUACAtt
BHLHB2 (Oligo 3)	GGAUCGGCGCAAUUAAGCAtt
RCOR1 (Oligo 1)	CCAGAUAAAUCUAUAGCAAtt
RCOR1 (Oligo 2)	GGAAUUGGUUUCAGUCAAAtt
RCOR1 (Oligo 3)	GAGUGGACUGUGGAAGAUAtt

SSRP1 (Oligo 1)	CGUUGACUCUGAACAUGAAtt
SSRP1 (Oligo 2)	GCAAGACCUUUGACUACAAtt
SSRP1 (Oligo 3)	GGACUUAAACUGCUUACAAtt
SUBT16H (Oligo 1)	GGUUUGGGAUGGGAAUUGAtt
SUBT16H (Oligo 2)	GCAUAUACCAUCGCUGUAAtt
SUBT16H (Oligo 3)	CUCUAACCGUGGUUCCAGAtt

siRNA oligonucleotide transfections:

β-K562 cells were transfected with siRNA oligonucleotides obtained from the human Silencer Select Druggable Genome siRNA Library V4 (Ambion). For all genes analyzed, two different siRNAs were tested. For siRNA showing a promising result, transfections were performed in triplicates. In detail, siRNA oligonucleotides were dissolved in Opti-MEM® I medium (Invitrogen) at a concentration of 120 nM and gently mixed with an equal volume of Opti-MEM® I containing 0.5% v/v lipofectamine® RNAiMAX (Invitrogen). 50 ml/well of transfection mixture was transferred to U-bottom 96-well plates (Greiner Bio-One, Germany) containing 100 ml/well of cells in growth medium at a density of 1500 cells/ml. Final siRNA oligos concentration was set to 20 nM. Plates were incubated for 7 days before high-content analysis (and RTqPCR analysis, when carried out). Cell transfection was done with at least two technical replicates per experiment by using at least two targeting siRNA per gene.

Immunofluorescence and high-content analysis:

Cells were collected and fixed in 3.7% paraformaldehyde for 20' min at RT, washed and permeabilized in staining buffer (PBS with

0.05% v/v Triton® X-100 and 1% w/v powdered milk) for 30 min. After washing in PBS, cells were incubated overnight at 4°C in staining buffer containing the appropriate antibodies and 1mg/ml Hoechst 33342. After washing, cells were resuspended in PBS and transferred to 96-well CELLSTAR®, Black/µClear® plates (Greiner Bio-One). Plates were spin for 5 min at 2g to facilitate cell attachment, sealed and analyzed with the ArrayScan VTI high-content screening reader (Thermo-Fisher Scientific). At least 600 cells were acquired in each well with a 20x magnification in three fluorescence channels (blue, green and red). The Molecular Translocation Bioapplication was used to determine the cell count per field, the nuclear area and intensity (based on the Hoechst staining in the blue channel) and the cytoplasmatic fluorescence intensity of β-(green) and γ-(red) globins. For simultaneous globins/GlycophorinA staining, APC anti-CD235 antibody was added for 2 hours to cells already stained for globins.

Confocal microscopy:

K562 and β-K562 cells, stained as above, were transferred to a microscope glass slide and mounted with Mowiol (Sigma-Aldrich). Microphotographs were acquired with a confocal Zeiss microscope LSM710.

Flow cytometry:

10^6 cells were washed, fixed and permeabilized for 10 min on ice, then washed and incubated in PBS+1% milk for 20 min. After

washing, cells were stained overnight at 4°C in PBS+1% milk containing the appropriate antibodies. After washing, cells were analyzed with FACSCalibur (Becton Dickinson).

RNA Isolation and RT-PCR:

Total RNA from 10^6 cells was extracted with the TRI Reagent (Applied Biosystems), treated with RQ1 DNase (Promega) for 30 min at 37°C and retrotranscribed (Applied Biosystems). Negative control reactions (RT-) gave no signal. Real time analysis was performed using ABI Prism 7500, (Applied Biosystems).

All primers used are listed in the table below:

Primers	**Sequence (5'-3')**
GAPDH Fw	ACGGATTTGGTCGTATTGGG
GAPDH Rw	TGATTTTGGAGGGATCTCGC
Alpha globin Fw	GAGGCCCTGGAGAGGATGTTCC
Alpha globin Rw	ACAGCGCGTTGGGCATGTCGTC
Beta globin Fw	TACATTTGCTTCTGACACAAC
Beta globin Rw	ACAGATCCCCAAAGGAC
Gamma globin Fw	CTTCAAGCTCCTGGGAAATGT
Gamma globin Rw	GCAGAATAAAGCCTACCTTGAAAG
Epsilon globin Fw	GCCTGTGGAGCAAGATGAAT
Epsilon globin Rw	GCGGGCTTGAGGTTGT

RESULTS:

Identification and characterization of a variant K562 subclone expressing β-globin

The K562 cell line has been widely used over the last decades as a cellular model of human "erythroid" cells. K562 cells are human leukemia cells derived form a patient with Bcr-Abl^{+} chronic myelogenious leukemia (CML). Upon treatment with inducing agents, these cells are able to differentiate into hemoglobin producing cells. However, they are not able to produce adult hemoglobin HbA (they do not express β-globin) despite the fact that these cells carry an intact β-globin cluster. As for their differentiation stage, K562 cells behave like undifferentiated early hematopoietic multipotent progenitors.

We identified and characterized a variant of the K562 subclone expressing significant levels of the adult β-globin gene, that we thus called β-K562. These cells are morphologically similar to ECACC (European Collection of Cell Culture) K562 cells, considered as "prototypical" K562 clone (Figure 4.1). Different experiments comparing their behavior and their sensitivity to drugs treatments demonstrate that both cell lines are highly similar. The doubling time between them is practically equal (Figure 4.1b), and they also have the same sensitivity in the response to several drugs that inhibits K562 proliferation (such as imatinib mesylate, desamatinib and doxorubicin) (Figure 4.1d). Despite their *bona fide* "K562-like" profile, β-K562 do express the adult β-globin chain, as assessed by both RTqPCR (Figure 4.1c)

and FACS analysis (Figure 4.1e).

Based on these results, we concluded that β-K562 subclone is a good model to set up the multiplex immunofluorescence high-throughput, high-content screening platform to analyze novel gene/drugs able to modulate the γ-/β-globin ratio.

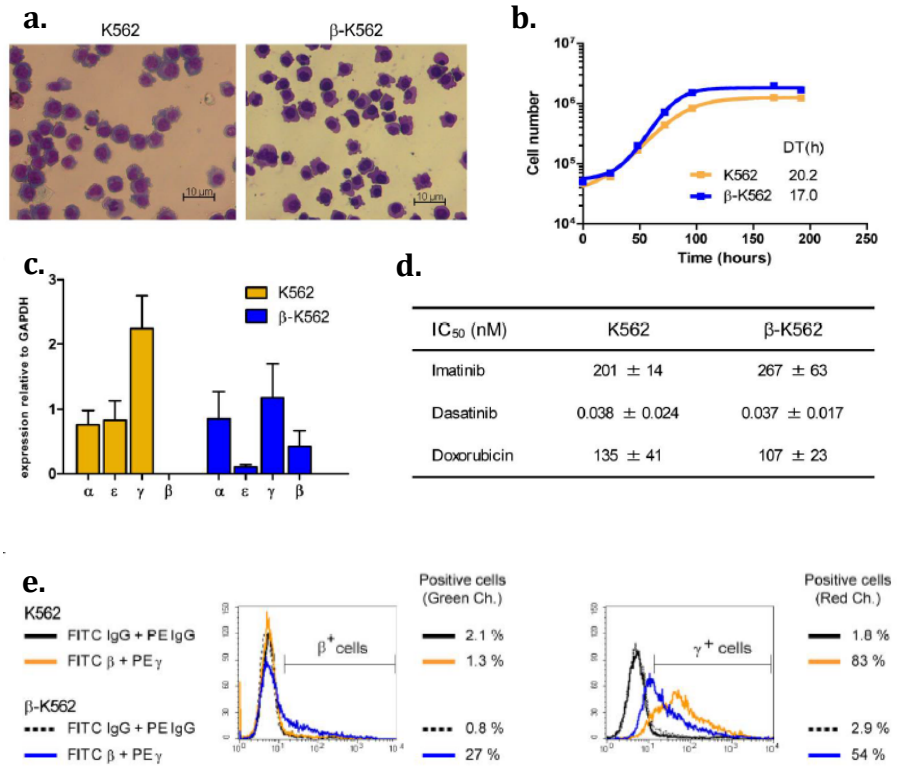


Figure 4.1 Characterization of the β-K562 subclone by its comparison with ECAAC-K562.

a) Microscopy pictures comparing both cells lines. Hematoxylin/eosin staining. b) Proliferation analysis of the cell lines show similar growth curves (n=2). c) RTqPCR on α-, ε-, γ- and β-globin genes. Histograms show the relative levels of expression normalized on GAPDH (n≥4). d) Response (IC_{50}) to imatinib mesylate, dasatinib and doxorubicin (n≥3). e) FACS analysis: cells were stained with anti γ- and anti β-globin antibodies and with the corresponding isotype controls and read in FL-1 (FITC, green channel) or in FL-2 (PE, red channel). A representative experiment is shown. ***(Durlak et al. in preparation)***

Development of a multiplex high-content assay for quantification of γ- and β-globin content in β-K562 cells at the single cell level

ECACC-K562 or β-K562 cells ($5x10^4$ cells/well) were seeded in 24-well plate and analyzed using an Array Scan VTI reader (Thermo-Fisher Scientific). Nuclei were stained with Hoechst-33342; γ and β globins were immunostained by using specific PE-anti γ- and FITC-anti β-globin antibodies, respectively. The processed data provides a quantitative fluorescence imaging at a cell single level. DNA content (Hoechst-33342), β-globin and γ-globin are read in channel 1 (Ch1), channel 2 (Ch2) and channel 3 (Ch3), respectively. The intensity of the staining is automatically converted into the corresponding intensity of colors: blue for Hoechst, green for β-globin and red for γ-globin. The detection threshold for the scoring of single γ^+-, β^+- and double $\beta^+\gamma^+$-cells was defined by using cells stained with the respective isotype controls (PE-IgG1 and FITC-IgG1) (Figure 4.3). When the signals from the three single channels are merged the double expression of γ plus β results in an orange/yellow coloration of different intensity, depending on the amount of γ and β chains (Figure 4.4). This analysis allows measuring of both the percentage of single-positive (γ^+ or β^+) and of double positive ($\gamma^+\beta^+$) cells in each field. In a standard experiment, data are acquired from a minimum of 500 cells and plotted to give an immediate visual image of cell distribution with respect to globin accumulation per cell. The majority of ECACC-K562 cells are γ^+ (≈54.1±0,4% $\gamma^+\beta^-$+≈3.1±0,3% $\gamma^+\beta^+$), the remaining being mostly $\gamma^-\beta^-$, with just a few marginally

β⁺ cells (3.1÷0,3% γ⁺β⁺+0.7±0,5% γ⁻β⁺). In contrast, about 57% of β-K562 cells are γ⁺ (38.3±4,1% γ⁺β⁻+18.7±4,3% γ⁺β⁺) and about 22% of cells are positive for β staining (18.7±4,3% γ⁺β⁺ +3.6±1,2% γ⁻β⁺)(Figure 4.2)(***Durlak et al. in preparation***). Moreover, the Mean Fluorescence Intensity (MFI, Y axis) of ECACC-K562 γ⁺ cells is higher than in β-K562.

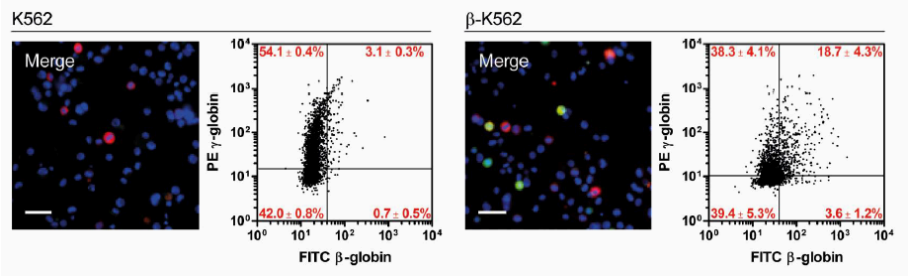


Figure 4.2 Analysis of γ/β _globin levels by immunofluorescence and automated image capture. *"Merge" image and plot of K562 (left panels) and β-K562 (right panel), confirming a significant expression of b-globin in these latter cells (n≥3). Bar=50mm. (**Durlak et al. in preparation**).*

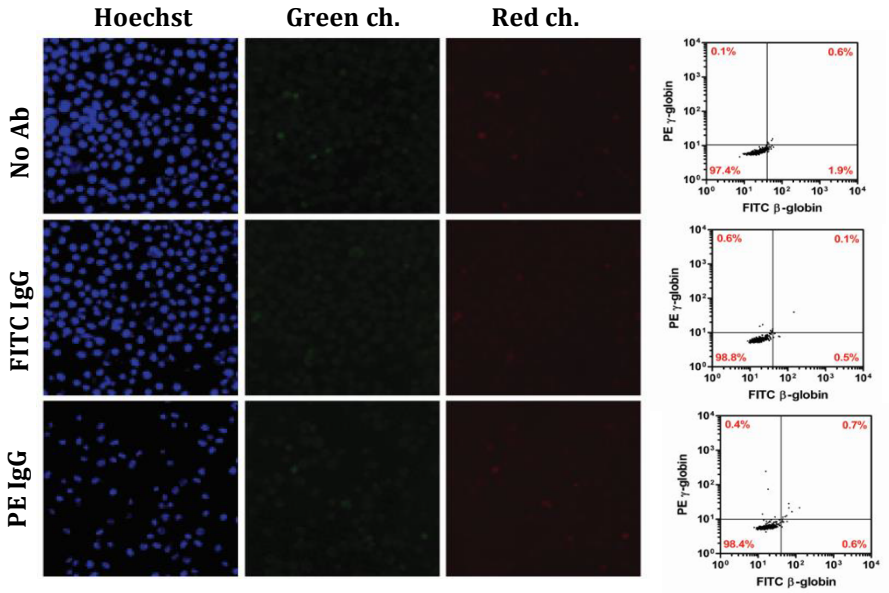


Figure 4.3 Representative array scan pictures of high content γ/β globin ratio analysis. in β-K562 cells. *Immunostaining control: (first line) unstained cells; (second and third line) isotypes controls: FITC-IgG and PE-IgG. Each representative set of pictures is accompanied by the corresponding scatter plot. Each representative set of pictures is accompanied by its scatter plot. All cells were counter-stained with Hoechst-33342. Bar: 50µm*

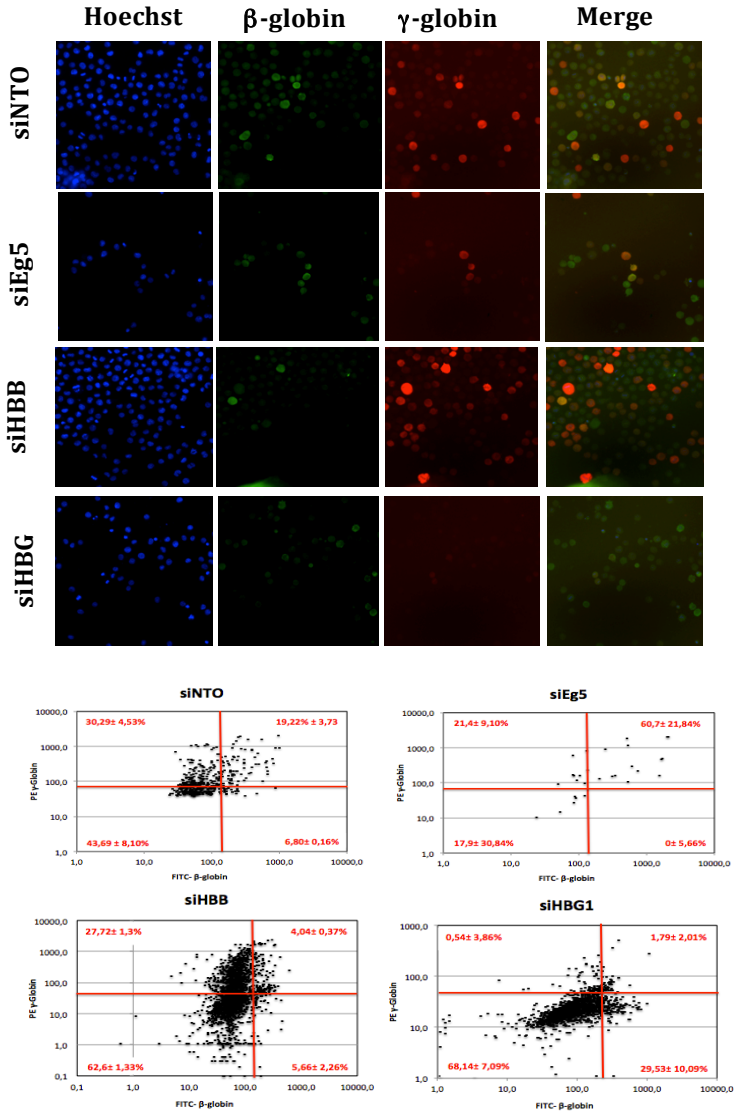


Figure 4.4 Representative array scan pictures of high content γ/β globin ratio analysis in β-K562 cells. *Transfection controls: Cells were transfected with a non-targeting oligo (siNTO) as a negative control and with a siRNA targeting Eg5, gene with an important role in proliferation, as a positive control. As a control of the specificity of the antibodies (Ab) siRNA targeting γ- (HBG) and β-globins (HBB) were used to control the proper reduction of the mean intensity of fluorescence of each globin chain. Each representative set of pictures is accompanied by the corresponding scatter plot. All cells were counter-stained with Hoechst-33342. Bar: 50μm*

These data confirm what we have seen by RT-qPCR and FACS analysis (Figure 4.1), β-K562 express high amount of β-globin, but the huge majority of the cells are $\gamma^+\beta^+$, suggesting that most of the cells have not complete the globins switch. This confirms that β-K562 cells are an optimal model to be used to perform the multiplex immunofluorescence high-content screening to analyze novel gene/drugs able to modulate the γ-/β-globin ratio.

Screening of the most promising Sox6 interacting proteins

By using the multiplex immunomicroscopy high throughput platform described above, we performed a functional screening for several Sox6 putative interacting proteins to presumably identify novel genes able to induce changes in the γ-/β- globin ratio under loss of function conditions. Knockdown in β-K562 cells of each gene was performed using three siRNA oligonucleotides particularly designed to ensure proper gene downregulation.

To validate the efficacy of the method we set up different kind of

controls: i)we transfected a non-targeting oligo as a negative control (siNTO); ii)as a positive transfection control we targeted Eg5 (kinesin-related motor protein), inhibition of Eg5 activity leads to a severe effect in cell growth; iii) we downregulated both γ- (siHbG) and β-Globin(siHbB), obtaining almost a complete depletion of respective signals (Figure 4.4).

The negative control of transfection (siNTO) also serves to set the threshold of both γ- and β- globins for each single experiment. About 50% of β-K562 cells are γ^+ (30.29%±4.53 $\gamma^+\beta^-$ +19.22%±3.73 $\gamma^+\beta^+$) and about 26% of cells are positive for β staining (19.22%±3.73 $\gamma^+\beta^+$ +6.8%±0.15 $\gamma^-\beta^+$) and approximately the 45% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.4).

From the list of putative candidate intaractors of Sox6 (described in chapter 2) we screened by siRNA knockdown 30 different genes to analyze their presumable effect on the γ/β globin ratio upon its downregulation. From the screening we found few putative partners of Sox6 with a potential role in the hemoglobin switching. However, from all genes analyzed with no effect in the γ/β globin ratio we found several of them producing strong effect in cell proliferation. The knockdown of these genes, upon four days of incubation, was lethal for the cells (more than 70% of the cells were dead), and thus we could not evaluate a possible effect of these knockdown on globins expression. Such genes affecting proliferation are several eukaryotic initiation factors (eIFs), the vascular endothelial zing finger (VEZF) and several cyclin

dependent kinases (CDKs), such as CDK11 and Cyck (See table 1). Other genes analyzed show no effects on both γ/β globin ratio and cell proliferation. Among genes with no putative effect on hemoglobinization, there are genes that have been previously described to act as regulators of the cytoskeleton, such as vimentin (VIM), villin 1 (VIL1) or plakophylin 3 (PKP3); as regulators of the chromatin organization, such as prohomeotic-like 2 (PHD2); or nuclear receptors acting as co-activators, such as thyroid hormone receptor 4 (TRIP4) and nuclear receptor co-activator 5 (NCOA5). The functional diversity of the candidate interactors of Sox6 demonstrates that Sox6 might play several different roles during erythroid development. However, the high throughput immunofluorescence screening approach is limited to analyze the role of the different genes in the γ/β ratio regulation. Although several genes show no effect using this screening, it does not mean that these genes might not have an important role during hematopoietic or erythropoietic development.

This screening is giving a first approach on how the protein interacting with Sox6 behave within the context of the control the hemoglobin genes. Thus, we will describe in the next sections the candidate interactors of Sox6 with a putative role regulating hemoglobin genes.

Gene ID	Effect on proliferation	Effect on γ/β globin ratio	No effect
BCOR			✔

BHLHB2		✔	
CCNK	✔		
CCNL			✔
CDK11	✔		
CHD2			
CHD4	✔		
DNM2	✔		
EIF4A1	✔		
EIF4G1	✔		
HDAC1	✔	✔	
LIMA1			✔
MART3	✔		
MBD3			✔
MTA1		✔	
MTA2			✔
MTA3			✔
NCOA5			✔
PHC2			✔
PKP3			✔
RCOR1		✔	
SMARCA5		✔	
SMARCC2		✔	
SRSP1	✔		
SSRP1	✔		
SUPT16H		✔	
TRIP4			✔
VEZF	✔		
VIL			✔
VIM	✔		

Table 1. List of the 30 genes analyzed by high-content screening. *All the genes are classifies depending in their phenotype upon their downregulation.*

1. NuRD Complex

We performed systematic transfection screening on the β-K562 cell line with the siRNA of the majority of the NuRD complex subunits (HDAC1, MTA1/2/3, MBD3, CDH2/4). As it was previously described (see Chapter 3.1) the NuRD complex has a role in repressing gene expression. We hypothesized that, as Sox6 is a transcription factor with a critical role in the repression of γ-globin gene, NuRD subunits could be involved in helping Sox6 to exert its role. Among genes belonging to the NuRD complex we targeted by siRNA MTA1 and HDAC1 to test whether they have any effect in modulating the γ/β globins ratio. Interestingly, the downregulation of these two proteins seems to have an opposite effect: whereas siRNAs targeting MTA1 result in a large increase of γ-globin expression, siRNAs targeting HDAC1 increases in β-globin, accompanied by a strong reduction of cell proliferation. MTA1 knockdown results in a robust increase in the percentage of both $\gamma^+\beta^-$ and $\gamma^+\beta^+$ expressing cells, suggesting that also β-globin expression levels are increased. About 80% of β-K562 cells are γ^+ (50.1±12.55% $\gamma^+\beta^-$+29.7±4.87% $\gamma^+\beta^+$) and about 30% of the cells are β^+ (29.7±4.87% $\gamma^+\beta^+$ +4.9±2.94% $\gamma^-\beta^+$) and approximately 20% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.5). The upregulation of γ-globin in both the single (γ+β-) and the double positive ($\gamma^+\beta^+$) cells suggests that β expression is also increased, but only in the double positive cells fraction ($\gamma^+\beta^+$), whereas single $\gamma^-\beta^+$ positives cells remain constant. Taken together these results, suggest that the cells do not have the ability to specifically upregulate β expression.

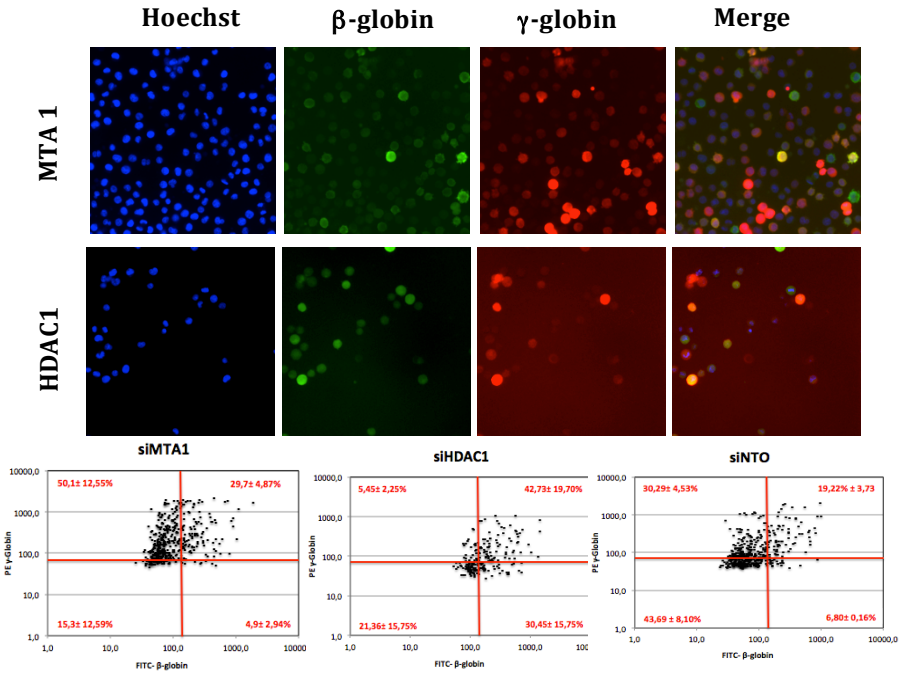


Figure 4.5 High content screening for the γ/β globins ratio analysis as readout of siRNA targeting NuRD subunits in β-K562 cells*. siRNAs targeting the MTA1 subunit show an increase in the γ-globin expression. In contrast, siRNAs targeting for HDAC1 show increase of β-globin, and strong effect in block of proliferation. For each gene, three different siRNA were tested (The immunofluorescence images shown are representative experiments). Each representative set of pictures is accompanied by the corresponding scatter plot. . The scatter plot corresponding to siNTO represents the transfection negative control (see the corresponding images in the figure 4.3). All cells were counter-stained with Hoechst-33342. Bar: 50μm*

In contrast, siRNAs targeting HDAC1 show a large increase in the percentage of β^+ cells. However, we have to keep in mind that its downregulation has also a strong effect on proliferation, with a high percentage of cells dying during the incubation time. In fact, when cells are collected four days after transduction, about 60% of β-K562 cells are γ^+ (5.45±2.25% $\gamma^+\beta^-$+42.73±19.7% $\gamma^+\beta^+$) and about 75% of cells are positive for β staining (42.73±19.7% $\gamma^+\beta^+$

+30.45±15.75% $\gamma^{-}\beta^{+}$) and approximately 20% of the cells are negative for both stainings ($\gamma^{-}\beta^{-}$) (Figure 4.5). This data shows that most of the cells have increased levels of β-globin expression, thus suggesting a potential role of HDAC1 in silencing β-globin expression, although this effect could be an artifact due to the high mortality induced by HDAC1 siRNAs .

The other NuRD subunits analyzed (MTA2, MTA3, MBD3, CHD2 and CHD4) do not show any effect on hemoglobinization and on influencing the γ/β ratio. In particular, siRNA targeting CHD2 and CHD4 have only a strong effect on proliferation, with a mortality of about 75% of cells (Data not shown).

2. SWI/SNF Complex

Between the Sox6 interacting proteins we also identified some proteins belonging to the SWI/SNF complexes (See chapter 2). This complex regulates DNA accessibility of chromatin fibers, in ATP-dependent mode, and is important regulators of gene expression and genome stability. Moreover, they have been implicated in the regulation of both, hematopoiesis and erythropoiesis. KLF1 requires the SWI/SNF complex for transactivation of β-globin gene ***(Armstrong, Bieker, & Emerson, 1998)***. The KLF1-SWI/SNF complex generates a transcriptionally active β-globin promoter *in vitro*. The SWI-SNF complexes are formed by eight or more different proteins. In particular, KLF1 recruits the subunits SMARCA4 (also known as BRG1) and SMARCC2 near the transcription initiation site on the β-globin

promoter. Also GATA1 has been described to interact with SMARCA4 ***(Kadam & Emerson, 2003; Kadam et al., 2000)***. Both TFs can recruit SWI/SNF-related complexes to the β-globin LCR. Once recruited, SMARCA4 is required for the LCR to adopt an open chromatin structure ***(Bultman, Gebuhr, & Magnuson, 2005)***.

The SWI/SNF complexes were initially described to be activators of gene expression due to their ability to increase the chromatin accessibility. However, whole-genome mRNA expression studies suggest that SWI/SNF also repress transcription, as almost half of the genes affected in SWI/SNF mutants have increased mRNA levels ***(Holstege et al., 1998; Sudarsanam & Winston, 2000)***. Nowadays there are no evidences of SWI/SNF complexes working as repressors in the β-globin locus.

To clarify this point, we performed a transfection screening on β-K562 cells with siRNAs targeting the two candidates interactors of Sox6. Both SMARCC2 and SMARCA5 knockdown have no effect on the γ-/β- globins ratio, but there is an overall increase in the percentage of γ^+ cells, both single ($\gamma^+\beta^-$) and double ($\gamma^+\beta^+$) positives. In particular, in SMARCC2 knockdown cells, about 76% of β-K562 cells are γ^+ (35.82±8.24% $\gamma^+\beta^-$+41.36±14.63% $\gamma^+\beta^+$) and about 46% of cells are positive for β staining (41.36±14.63% $\gamma^+\beta^+$+6.61±1.77% $\gamma^-\beta^+$) and approximately the 16% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.6).

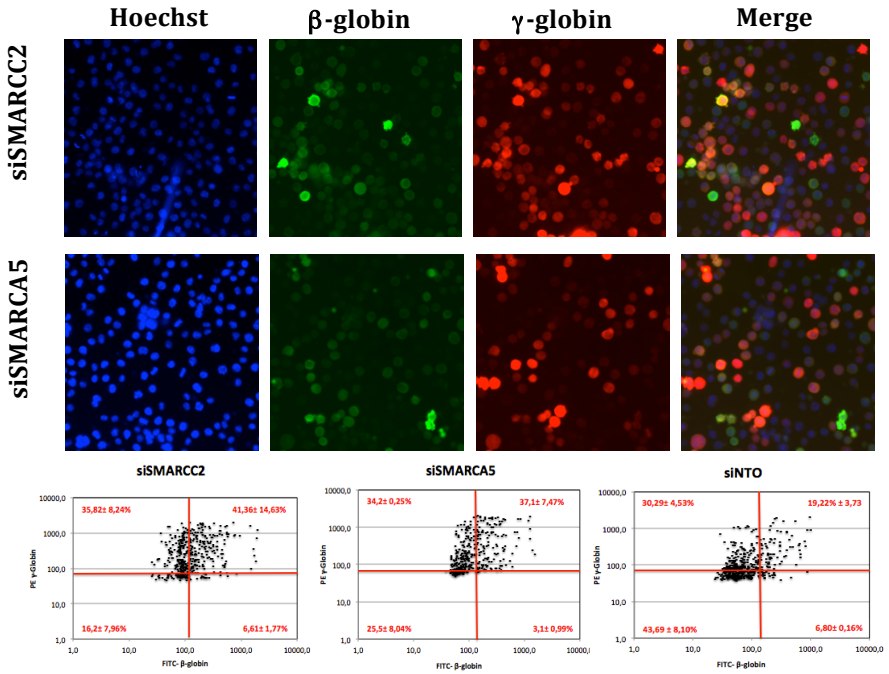


Figure 4.6 High content screening for the γ/β globins ratio analysis as readout of siRNA targeting of SWI/SNF subunits in β-K562 cells*. siRNAs targeting the SMARCC2 and SMARCA5 subunits show increase in both γ- and β-globin expression. For each gene, three different siRNA were tested (The immunofluorescence images shown are representative for the experiments). Each representative set of pictures is accompanied by the corresponding scatter plot. . The scatter plot corresponding to siNTO represents the transfection negative control (see the corresponding images in the figure 4.3). All cells were counter-stained with Hoechst-33342. Bar: 50μm*

SMARCA5 β-K562 knockdown cells are about 71% γ^+ (34.02±0.25% $\gamma^+\beta^-$+37.1±7.47% $\gamma^+\beta^+$) and about 40% positive for β staining (37.1±7.47% $\gamma^+\beta^+$+3.1±0.99% $\gamma^-\beta^+$) and approximately 25% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.6). Both SMARCC2 and SMARCA5 knockdown results in a large increase in γ-globin expression (both $\gamma^+\beta^-$ and $\gamma^+\beta^+$ cells) whereas the percentage of cells expressing only β-globin remains

unchanged, or even slightly reduced (in the case of single β^+ ($\gamma^-\beta^+$) cells upon SMARCA5 downregulation). Accordingly with the previous published studies, this data suggests that these chromatin-remodeling factors might play a role in maintaining the chromatin accessibility in the β-globin promoter.

3. BHLHB2

Bhlhb2 (also known as Dec1, Stra13, Sharp2, or Bhlhe40) belongs to a family of basic helix-loop-helix transcriptional regulators. Members of this family are known to respond to environmental stimuli and regulate several physiological processes in diverse cell types, including cell cycle, apoptosis, and differentiation via their actions as both transcriptional activators and repressors. bHLHB2 is key regulator of lymphocyte activation responding to cytokine stimulus. Nothing is known about the role of this TF in erythropoiesis, but having the ability to activate and repress gene expression is a putative interesting Sox6 interacting protein to act regulating globin gene expression.

Of interest, siRNAs targeting BhlhB2 shows a more marked increased in the percentage of β^+ cells (both $\gamma^+\beta^+$ and $\gamma^-\beta^+$) than of γ^+ cells (both $\gamma^+\beta^-$ and $\gamma^+\beta^+$). In fact, about 66% of β-K562 cells are γ^+ (25.8±2.56% $\gamma^+\beta^-$+41.4±8.12% $\gamma^+\beta^+$) and about 53% of cells are positive for β staining (41.4±8.12% $\gamma^+\beta^+$+12.2±1.74% $\gamma^-\beta^+$) and approximately the 20% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.7). These results show that is the percentage of β^+ cells increased by 2.06 times in comparison to the γ^+ cells that increased just 1.32 times. Thus, BhlhB2 might have a role in the

repression of β-globin expression.

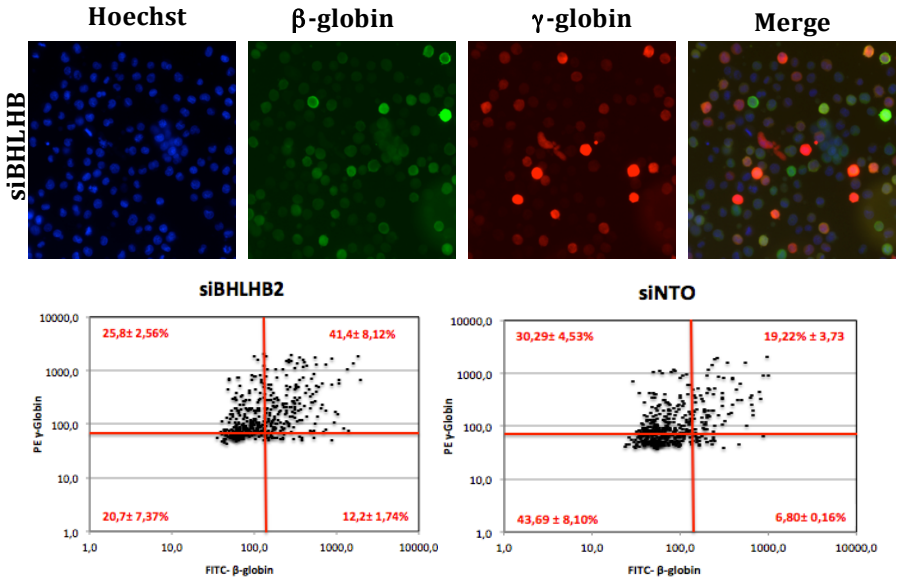


Figure 4.7 High content screening for the γ/β globins ratio analysis as readout of siRNA targeting of bHLHB2 in β-K562 cells line*. siRNAs targeting the TF bHLHb2 show marked increase in β-globin expression. For each gene, three different siRNA were tested (The immunofluorescence images shown are representative for the experiments). Each representative set of pictures is accompanied by the corresponding scatter plot. The scatter plot corresponding to siNTO represents the transfection negative control (see the corresponding images in the figure 4.3). All the cells were counter-stained with Hoechs-33342. Bar: 50μm.*

4. RCOR1

RCOR1 (also called CoREST) is a repressor with a crucial role in erythropoiesis, Rcor1 null mice die late in gestation due to severe anemia ***(Yao et al., 2014)***. Rcor1 interacts with LSD (Lysine – specific demethylase 1) to regulate gene repression by epigenetic modifications ***(Lee, Wynder, Cooch, & Shiekhattar, 2005; Shi et al., 2005)***. This complex associates with the erythroid transcription factors Gfi1 and Gfi1b (Growth factor independence

1/1b) repressing the majority of its gene targets in erythropoiesis ***(Saleque, Kim, Rooke, & Orkin, 2007)***, with Scl1/Tal1 ***(Hu et al., 2009; Li et al., 2012)*** and with BCL11a. BCL11a/LSD/Rcor1 complex acts silencing the mouse embryo β-like globin genes and human γ-globin genes in adult erythroid cells *in vivo* ***(Xu et al., 2013)***. As Sox6 and BCL11a cooperate in the silencing of γ-globin promoter, BCL11a/LSD/RCOR1/Sox6 might form a complex to silence the expression of γ-globin gene.

RCOR1 silencing results in an increase in both γ^+ and β^+ cells (Figure 4.8). In fact, about 72% of β-K562 cells are γ^+ (32.19±0.61% $\gamma^+\beta^-$+39.04±8.72% $\gamma^+\beta^+$) and about 49% of cells are positive for β staining (39.04±8.72% $\gamma^+\beta^+$+9.82±1.7% $\gamma^-\beta^+$) and approximately the 20% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.7). These data suggest that, in this specific situation, RCOR1 works as a general repressor of both γ- and β-globins. Thus, RCOR1 might have a role regulating hemoglobin genes independently from BCL11a. As in the β-K562 cells the expression of BCL11a is almost absent, but there is still low expression of Sox6, it might be possible that RCOR1 cooperates with Sox6 to repress gene expression.

5. FACT complex

FACT (facilitates chromatin transcription) is a heterodimeric protein complex that affects RNA polymerase II (RNAPII) transcription elongation. FACT consists in two subunits, SUPT16H and SSRP1. The complex was suggested to contribute to the fidelity of Pol II transcription by linking the processes of initiation

and elongation. A recent study published that the Transcriptional Intermediary Factor 1 gamma (TIF1γ) plays an essential role in erythroid differentiation. TIF1γ and the FACT complex interact to activate transcriptional elongation of erythroid genes in human erythroid cells ***(Bai et al., 2010)***.

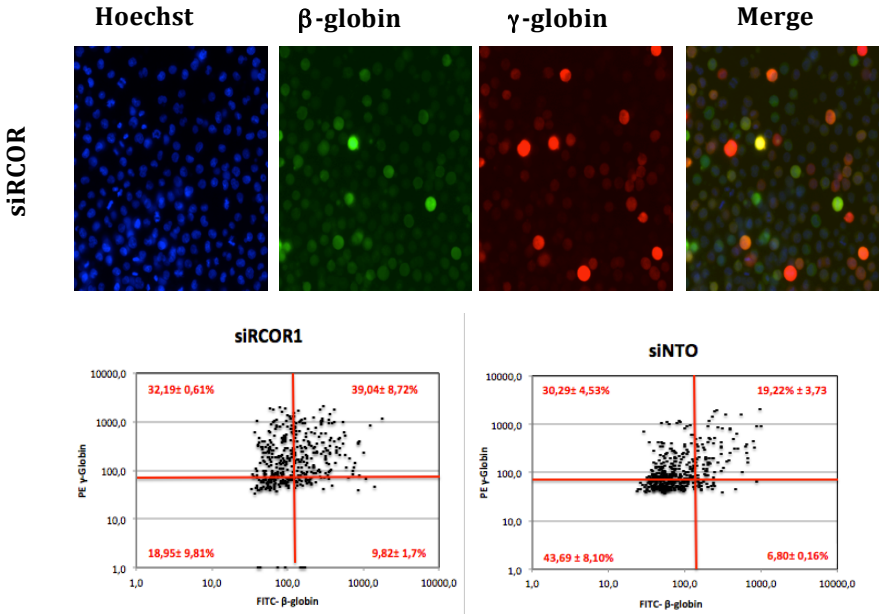


Figure 4.8 High content screening for the γ/β globins ratio analysis as readout of siRNA targeting RCOR1 in β-K562 cells*. siRNAs targeting the repressor RCOR1 show marked increase in both γ- and β-globin expression. For each gene, three different siRNA were tested (The immunofluorescence images shown are representative for the experiments). Each representative set of pictures is accompanied by the corresponding scatter plot. The scatter plot corresponding to siNTO represents the transfection negative control (see the corresponding images in the figure 4.3). All the cells were counter-stained with Hoechst-33342. Bar: 50μm.*

Both subunits of the FACT complex were found to interact with the Sox6 (See chapter 2). The siRNA targeting for the SSRP1 subunit have a strong effect on proliferation: upon four days of

incubation, only few cells were still alive (approximately 70% of the cells were dead) and thus we could not evaluate a possible effect of its knockdown on globins expression. The targeting siRNA of SUPT16H also inhibits cellular proliferation (approximately 20% of the cells were death), but its effects are milder than those of SSRP1 knockdown. SUPT16H downregulation induces an upregulation of both γ- and β-globins. About 77% of β-K562 cells are γ^+ (32.56±2.82% $\gamma^+\beta^-$ +45.17±7.14% $\gamma^+\beta^+$) and about 53% of cells are positive for β staining (45.17±7.14% $\gamma^+\beta^+$+8.19±0.71% $\gamma^-\beta^+$) and approximately the 15% of the cells are negative for both stainings ($\gamma^-\beta^-$) (Figure 4.9)

These results are in agreement with previous studies showing that the conditional inactivation of the SUPT16H subunit of the FACT complex results in increased Pol II density, transcription, and TATA-binding protein (TBP) occupancy in the 3′ portion of certain coding regions, indicating that FACT suppresses inappropriate initiation from cryptic promoters within coding regions ***(Mason & Struhl, 2003)***. Thus, the FACT complex might play a role controlling the transcriptional regulation of γ- and β-globins.

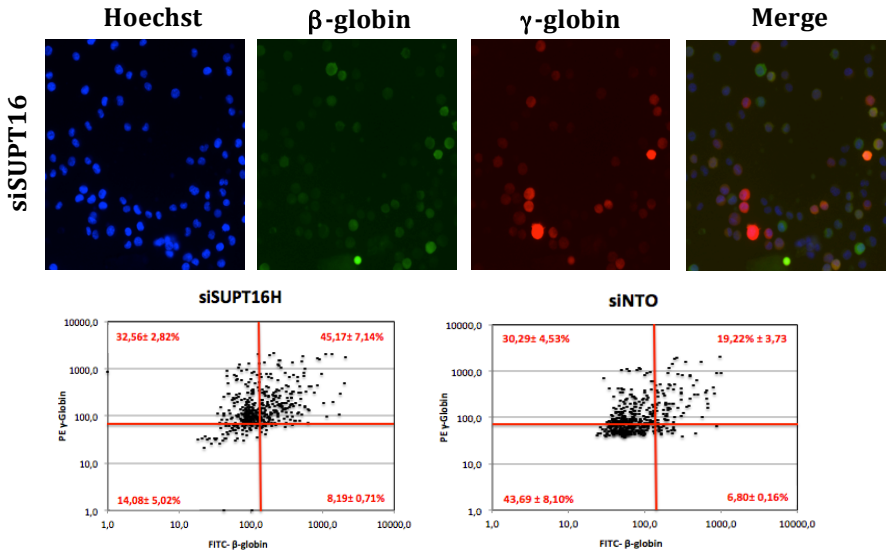


Figure 4.9 High content screening for the γ/β globins ratio analysis as readout of siRNA targeting FACT complex in β-K562 cells*. siRNAs targeting the FACT complex subunit SUPT16H show marked increase in both γ- and β-globin expression. For each gene, three different siRNA were tested (The immunofluorescence images shown are representative for the experiments). Each representative set of pictures is accompanied by the corresponding scatter plot. The scatter plot corresponding to siNTO represents the transfection negative control (see the corresponding images in the figure 4.3). All the cells were counter-stained with Hoechst-33342. Bar: 50μm.*

DISCUSSION

In this chapter I described the novel assay that we set up to preform first-step high-throughput screening (HTS) to identify gene/drugs influencing the γ-/β-globin ratio. The usage of the variant β-K562 cell line that express both the fetal γ- and the adult β-globin from the intact β-locus, overcomes the major limitation of the available human erythroid cell lines to study the hemoglobin switching. In fact, the main fist-step HTS published so far are almost exclusively based on cell lines transfected with a variety of reporters, under the control of γ- and β- globin

promoters in the context of artificial genes/genomic arrangements. Mice carrying the artificial chromosome construct containing the entire human β-locus are an alternative and more physiological model to perform HTS. However these models are more difficult to handle than cell lines, thus they are not suitable to first-step HTS.

The fact that β-K562 cells express spontaneously γ- and β-globin from their intact β-locus becomes a useful and valid tool for this first-step HTS (Figure 4.1). By using specific immunofluorescent antibodies for γ- and β-globins led to produce a final picture of the analysis outcome, which validates the efficiency of the system. The final output illustrates the heterogeneity of the cells upon modulators (siRNA transfection) or drugs treatments. The robustness of β-K562 cells allows efficiently transfection-based screening.

On this basis, the novel multiplex imagining platform presented here can represent a very useful first-step tool for HTS of chemical compounds/siRNA libraries to identify new γ-globin inducers.

From the Mass Spectrometry approach used to search for Sox6 interactors (See chapter 2) we obtained a list of putative interacting proteins. In this chapter, we used the novel assay to perform a first-step HTS of several candidate interactors of Sox6 in order to identify proteins with a potential role in the differential regulation of globins gene expression. Among the 30

genes analyzed so far, 7 emerged as potential regulators of globins genes. Most of these genes are chromatin remodeling factors (such as the NuRD complex, SWI/SNF complex, FACT complex and LSD/CoREST complex) and one (BhlhB2) is a transcription factor

Modulating several subunits of the repressor NuRD complex we found that MTA1 and HDAC1 have an effect on the γ- and β-globin genes expression. Loss of MTA1 increases γ-globin expression (Figure 4.5). Whereas, loss of HDAC1 induces an increase in the expression of β-globin, but also strong effect in proliferation, being highly lethal for the cells (Figure 4.5). In the chapter 3.1 NuRD was already described and discuss about its potential role together with Sox6. The complex also interacts with other TFs such as with GATA-1 and FOG-1 regulating hemoglobin gene expression. Thus, it is not surprising that the modulation of its subunits produces an alteration on the γ/β globins ratio.

Downregulation of the subunits belonging to the SWI/SNF complexes, identified as putative Sox6 interacting proteins (SMARCC2 and SMARCA5), produces overall increase in either γ- and β-globin, but the increase is stronger for the γ^+ cells (Figure 4.6). The cells that are only β^+ remain unchanged, or even slightly reduced. This data is in agreement with the role of the SWI/SNF complexes together with KLF1 maintaining the chromatin accessibility in the β-globin promoter ***(Armstrong et al., 1998; Bultman et al., 2005; Kadam & Emerson, 2003; Kadam et al., 2000)***. Thus, upon the loss of its function some transcription

factors could not have accessibility to the promoter giving an advantage to those factors acting in the γ-globin promoters. Despite of its known role in the β-globin promoter, as the SWI/SNF complex has the ability to activate and to repress gene expression, it might act as cofactor of Sox6 in repressing γ-globin expression ***(Holstege et al., 1998; Sudarsanam & Winston, 2000)***.

siRNAs targeting the BhlhB2 transcription factor results in the upregulation in the β-globin gene (Figure 4.7). This TF is known to regulate lymphocytosis ***(Seimiya et al., 2002; Sun, Lu, Li, Flavell, & Taneja, 2001)***, but nothing is known of its role in erythropoiesis. This result suggests that BhlhB2 might be a repressor of β-globin gene.

LSD/RCOR1 have been described having a role in silencing the mouse embryonic β-like genes (εY and βh1) and the human fetal γ-globin gene ***(Xu et al., 2013)***. In our results, loss of RCOR1 does not show any specific effect on γ-globin, but it produces upregulation of both γ- and β-globin genes (Figure 4.8), suggesting that RCOR1 might not be a specific repressor for γ-globin. Moreover, a recent study published that RCOR1 is an essential corepressor for murine terminal erythropoiesis ***(Yao et al., 2014)***. Abrogation of RCOR1 produces blocking in erythroid progenitor maturation and in the differentiation toward to myeloid lineages. Thus, RCOR1 plays several important roles during erythropoiesis, not only in regulation hemoglobin genes.

Despite of its role repressing mouse embryonic and human fetal globins in collaboration with BCL11a ***(Xu et al., 2013)***, the fact that in β-K562 cells BCL11a is almost absent, but Sox6 is still expressed (at very low levels) gave indications that RCOR1 retains the ability of inhibit fetal γ-globin independently of BCL11a.

The FACT complex is formed by two different subunits, SSRP1 and SUPT16H. Although the loss of SSRP1 subunit seems to be lethal for the cells, siRNA targeting the SUPT16H subunit produce an increased expression of both γ- and β-globin expression (Figure 4.9). FACT complex plays a role in controlling transcriptional activation. Its downregulation produces loss of the transcriptional control and thus general transcriptional activation ***(Mason & Struhl, 2003)***. This complex might play a role with Sox6 in controlling and maintaining the proper transcriptional activation of globin genes.

This high throughput screening approach led us to perform a first-step analysis of the most interesting putative Sox6 interacting proteins in order to figure out if they have any specific role in the hemoglobin switching. The results from this screening is just giving the opportunity to open new research lines to better understand the molecular mechanism and function of all the factors interplaying a role in the β-globin locus. Our final goal is to draw a map up with all the functional interactome of Sox6. This screening provides a small map of candidates interactors of Sox6 with a potential functional role in the hemoglobin switching.

REFERENCES

Armstrong, J. A., Bieker, J. J., & Emerson, B. M. (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.

Bai, X., Kim, J., Yang, Z., Jurynec, M. J., Akie, T. E., Lee, J., . . . Zon, L. I. (2010). TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell, 142*(1), 133-143. doi: 10.1016/j.cell.2010.05.028

Banan, M. (2013). Hydroxyurea treatment in beta-thalassemia patients: to respond or not to respond? *Ann Hematol, 92*(3), 289-299. doi: 10.1007/s00277-012-1671-3

Bultman, S. J., Gebuhr, T. C., & Magnuson, T. (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), 2849-2861. doi: 10.1101/gad.1364105

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

Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., . . . Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell, 95*(5), 717-728.

Hu, X., Li, X., Valverde, K., Fu, X., Noguchi, C., Qiu, Y., & Huang, S. (2009). LSD1-mediated epigenetic modification is required for TAL1 function and hematopoiesis. *Proc Natl Acad Sci U S A, 106*(25), 10141-10146. doi: 10.1073/pnas.0900437106

Kadam, S., & Emerson, B. M. (2003). Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Mol Cell, 11*(2), 377-389.

Kadam, S., McAlpine, G. S., Phelan, M. L., Kingston, R. E., Jones, K. A., & Emerson, B. M. (2000). Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev, 14*(19), 2441-2451.

Krishna, R., Anderson, M. S., Bergman, A. J., Jin, B., Fallon, M., Cote, J., . . . Wagner, J. A. (2007). Effect of the cholesteryl ester transfer protein inhibitor, anacetrapib, on lipoproteins in

patients with dyslipidaemia and on 24-h ambulatory blood pressure in healthy individuals: two double-blind, randomised placebo-controlled phase I studies. *Lancet, 370*(9603), 1907-1914. doi: 10.1016/S0140-6736(07)61813-3

Lee, M. G., Wynder, C., Cooch, N., & Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature, 437*(7057), 432-435. doi: 10.1038/nature04021

Li, Y., Deng, C., Hu, X., Patel, B., Fu, X., Qiu, Y., . . . Huang, S. (2012). Dynamic interaction between TAL1 oncoprotein and LSD1 regulates TAL1 function in hematopoiesis and leukemogenesis. *Oncogene, 31*(48), 5007-5018. doi: 10.1038/onc.2012.8

Mason, P. B., & Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol Cell Biol, 23*(22), 8323-8333.

Musallam, K. M., Cappellini, M. D., & Taher, A. T. (2013). Iron overload in beta-thalassemia intermedia: an emerging concern. *Curr Opin Hematol, 20*(3), 187-192. doi: 10.1097/MOH.0b013e32835f5a5c

Musallam, K. M., Rivella, S., Vichinsky, E., & Rachmilewitz, E. A. (2013). Non-transfusion-dependent thalassemias. *Haematologica, 98*(6), 833-844. doi: 10.3324/haematol.2012.066845

Musallam, K. M., Taher, A. T., Cappellini, M. D., & Sankaran, V. G. (2013). Clinical experience with fetal hemoglobin induction therapy in patients with beta-thalassemia. *Blood, 121*(12), 2199-2212; quiz 2372. doi: 10.1182/blood-2012-10-408021

Olivieri, N. F., & Weatherall, D. J. (1998). The therapeutic reactivation of fetal haemoglobin. *Hum Mol Genet, 7*(10), 1655-1658.

Perrine, S. P., Castaneda, S. A., Chui, D. H., Faller, D. V., Berenson, R. J., Siritanaratku, N., & Fucharoen, S. (2010). Fetal globin gene inducers: novel agents and new potential. *Ann N Y Acad Sci, 1202*, 158-164. doi: 10.1111/j.1749-6632.2010.05593.x

Platt, O. S. (2008). Hydroxyurea for the treatment of sickle cell anemia. *N Engl J Med, 358*(13), 1362-1369. doi: 10.1056/NEJMct0708272

Ronchi, A., & Ottolenghi, S. (2013). To respond or not to respond to hydroxyurea in thalassemia: a matter of stress adaptation? *Haematologica, 98*(5), 657-659. doi: 10.3324/haematol.2013.084392

Saleque, S., Kim, J., Rooke, H. M., & Orkin, S. H. (2007). Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. *Mol Cell, 27*(4), 562-572. doi: 10.1016/j.molcel.2007.06.039

Seimiya, M., Bahar, R., Wang, Y., Kawamura, K., Tada, Y., Okada, S., . . . J, O. W. (2002). Clast5/Stra13 is a negative regulator of B lymphocyte activation. *Biochem Biophys Res Commun, 292*(1), 121-127.

Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., & Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell, 19*(6), 857-864. doi: 10.1016/j.molcel.2005.08.027

Stamatoyannopoulos, G., Blau, C. A., Nakamoto, B., Josephson, B., Li, Q., Liakopoulou, E., . . . Dover, G. (1994). Fetal hemoglobin induction by acetate, a product of butyrate catabolism. *Blood, 84*(9), 3198-3204.

Stamatoyannopoulos, J. A. (1992). Future prospects for treatment of hemoglobinopathies. *West J Med, 157*(6), 631-636.

Stamatoyannopoulos, J. A., & Nienhuis, A. W. (1992). Therapeutic approaches to hemoglobin switching in treatment of hemoglobinopathies. *Annu Rev Med, 43*, 497-521. doi: 10.1146/annurev.me.43.020192.002433

Steinberg, M. H. (2002). Hydroxyurea treatment for sickle cell disease. *ScientificWorldJournal, 2*, 1706-1728. doi: 10.1100/tsw.2002.295

Steinberg, M. H., Nagel, R. L., & Brugnara, C. (1997). Cellular effects of hydroxyurea in Hb SC disease. *Br J Haematol, 98*(4), 838-844.

Sudarsanam, P., & Winston, F. (2000). The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. *Trends Genet, 16*(8), 345-351.

Sun, H., Lu, B., Li, R. Q., Flavell, R. A., & Taneja, R. (2001). Defective T cell activation and autoimmune disorder in Stra13-

deficient mice. *Nat Immunol, 2*(11), 1040-1047. doi: 10.1038/ni721

Thein, S. L. (1998). Beta-thalassaemia. *Baillieres Clin Haematol, 11*(1), 91-126.

Weatherall, D. J., Higgs, D. R., Bunch, C., Old, J. M., Hunt, D. M., Pressley, L., . . . Bebbington, D. (1981). Hemoglobin H disease and mental retardation: a new syndrome or a remarkable coincidence? *N Engl J Med, 305*(11), 607-612. doi: 10.1056/NEJM198109103051103

Weatherall, D. J., Pressley, L., Wood, W. G., Higgs, D. R., & Clegg, J. B. (1981). Molecular basis for mild forms of homozygous beta-thalassaemia. *Lancet, 1*(8219), 527-529.

Xu, J., Bauer, D. E., Kerenyi, M. A., Vo, T. D., Hou, S., Hsu, Y. J., . . . Orkin, S. H. (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A, 110*(16), 6518-6523. doi: 10.1073/pnas.1303976110

Yao, H., Goldman, D. C., Nechiporuk, T., Kawane, S., McWeeney, S. K., Tyner, J. W., . . . Mandel, G. (2014). Corepressor Rcor1 is essential for murine erythropoiesis. *Blood, 123*(20), 3175-3184. doi: 10.1182/blood-2013-11-538678

Chapter 5

FINAL DISCUSSION:

Molecular and clinical considerations

SUMMARY:

The aim of this project was to identify the protein complexes interacting with Sox6 and modulating its activity in erythroid cells.

To this aim, we set up a single-step biotin-tagged approach in both Human (HEL) and mouse (MEL) erythroid cell lines (See chapter 2). This approach is particularly suitable to purify complexes containing TFs thanks to the strength of covalent binding between biotin and streptavidin molecules and to the low abundance of naturally biotinylated proteins (***De Boer et al. 2002***). These features result in a reduced background of unspecific proteins in the pulled down fractions.

The Sox6 pulled down complexes were subjected to a subsequent Mass Spectrometry analysis. The outcome of the peptides sequencing was a list of putative candidate Sox6 interactors, to which we applied several filtering criteria to draw a priority list of putative Sox6 interactors for further functional analysis.

As Sox6 is playing several different roles during erythroid maturation, we selected three different putative interactors (the NuRD complex, the COUP-TFII transcription factor and the cyclin dependent kinase CDK13) that might contribute to specific erythroid Sox6 functions. The validation of these interactions and their potential role was assessed in different cellular model

systems: in erythroleukemic cell lines and in more physiological *ex vivo* human erythroid cultures.

Moreover, in collaboration with Nerviano Medical Science, we set up a novel high-content screening platform based on multiplex imaging on a variant K562 cell line (β-K562) expressing significant levels of β-globin (***Durlak et al. in preparation***). This assay allows to detect changes in multiplex imaging on a variant K562 cell line (β-K562) expressing substantial amount of β-globin. We took advantage of this method to screen some of the previously identified Sox6 candidates for their potential role in hemoglobinization and in the control of γ/β globins switching.

NuRD as a repressor complex collaborating with Sox6:

Nucleosome Remodeling and histone Deacetylase (NuRD) complex plays role in chromatin reorganization and epigenetic modulation, and it was described as a repressor complex ***(Allen, Wade, & Kutateladze, 2013)***. Sox6 has been described to repress γ-globin gene expression ***(Ginder, Gnanapragasam, & Mian, 2008; Sankaran, Xu, & Orkin, 2010; Stamatoyannopoulos, 2005; Xu et al., 2010)***. On this basis, we hypothesized that Sox6 might recruit NuRD on the γ-promoter to silence γ-globin. Despite almost all the subunits belonging to the NuRD complex were present in the readout of the Mass Spectrometry sequencing, we could only validate the direct Sox6-MTA1 and Sox6-HDAC1 interactions in human erythroleukemic (HEL) cells (Chapter 3.1, Figure 3.1.1).

The elution pattern of Sox6, MTA1 and HDAC1 in Superose 6 size-exclusion chromatography of nuclear extracts from E12.5 (Chapter 3.1, Figure 3.1.2a) and E13.5 (Chapter 3.1, Figure 3.1.2b) mouse fetal liver primary cells is overlapping. This supports the fact that Sox6 can interact with these proteins *in vivo* at their physiological levels of expression and not only when Sox6 is overexpressed, as in the case of the experiments carried out on cell lines.

In the attempt to identify the Sox6 aminoacidic residues mediating the interaction with the NuRD complex, we identified within the N-terminal of Sox6 a motif (MSRRKQaKPqhF) very similar to the consensus present at the same position in several Zinc-finger TFs (BCL11a, BCL11b, FOG-1, FOG-1 and Sall family members). This aminoacids sequence is known to be sufficient and necessary to recruit the NuRD complex and to repress gene expression ***(Lauberth & Rauchman, 2006; Lin, Roche, Wilk, & Svensson, 2004)***. Interestingly, the N-terminal sequence of Sox6 is conserved among Sox6 orthologs in vertebrates (Chapter 3.1, Figure 3.1.3b), pointing to the functional importance of this motif. We thus created a N-terminal truncated version of Sox6 (ΔN-Sox6)(Chapter 3.1, Figure 3.1.4a) to test whether this sequence could be responsible for NuRD recruitment and thus for the Sox6-mediated transcriptional repression activity. Against our expectations, the N-terminal truncated version of Sox6 (ΔN-Sox6) still retains the ability to bind MTA1 and HDAC1 in HEL cells (Chapter 3.1, Figure 3.1.7). The expression analysis of genes

normally repressed by Sox6 confirmed that full length Sox6 (FL-Sox6) and ΔN-Sox6 behave in the same way (Chapter 3.1, Figure 3.1.5-6). These results indicate that probably the NuRD interacting domain within Sox6 does not reside in the N-terminal motif. A further detailed deletion mapping of Sox6 will be required to clarify this issue.

Regarding this point, it is important to notice that the knockdown (by siRNA transfection) of MTA1 and HDAC1 in β-K562 cells alters the γ/β globins expression in β-K562 cells. In particular, the knockdown of MTA1 is associated with increased levels of the γ-globin expression (Chapter 4, Figure 4.4), whereas the knockdown of HDAC1 results in increased β-globin expression, together with a block in cellular proliferation (Chapter 4, Figure 4.4).These data suggest that the NuRD complex plays an important role in the differential regulation of globins genes, possibly by interacting with Sox6.

Sox6 and COUP-TFII during hemoglobin switching:

COUP-TFII is an orphan nuclear receptor with unknown ligand, playing a critical role acting as transcriptional activator in several developmental processes. Both, Sox6 and COUP-TFII, have been described having a role in hemoglobin switching. Sox6 binds to the mouse εY-globin promoter repressing its expression ***(Yi et al., 2006)*** and cooperates with BCL11a-XL in repressing γ-globin expression in adult human erythroid progenitors ***(Xu et al., 2010)***. On the other hand, COUP-TFII was identified as a

modulator of ε- and γ-globin genes, where it binds to the CCAAT box regions. Once bound to these regions, it can either compete or cooperate with the binding of NF-Y ***(Liberati, Ronchi, Lievens, Ottolenghi, & Mantovani, 1998; Ronchi, Bellorini, Mongelli, & Mantovani, 1995)***. On the ε-globin promoter, COUP-TFII binds to the direct repeats DR1 (TGACCA) partially overlapping to the CCAAT box recognized by NF-Y ***(Filipe et al., 1999)***. COUP-TFII thus competes with NF-Y for this site and this competition results in a decreased promoter activity in transfection experiments ***(Liberati et al., 2001)***.

COUP-TFII was one of the erythroid TFs interacting with Sox6 according to the Mass Spectrometry results and we validated this interaction in HEL cells (Chapter 3.2, Figure 3.2.1). The mRNA and protein analysis of the expression pattern of both TFs in mouse fetal liver cells during embryonic development at the E11.5, E12.5 and E13.5 stages, reveals that both TFs have an opposite pattern of expression: whereas Sox6 expression increases, COUP-TFII decreases during development (Chapter 3.2, Figure 3.2.2) and they are simultaneously present at similar levels only at E12.5. To mimic this situation we overexpressed these two factors at different concentrations. Overexpression of COUP-TFII in β-K562 cells increases the levels of ε- and γ- globin gene expression (Chapter 3.2, Figure 3.2.5b), suggesting that COUP-TFII instead of having a role in the silencing of the γ-globin, it might be an activator. In contrast, overexpression of Sox6 induces a general

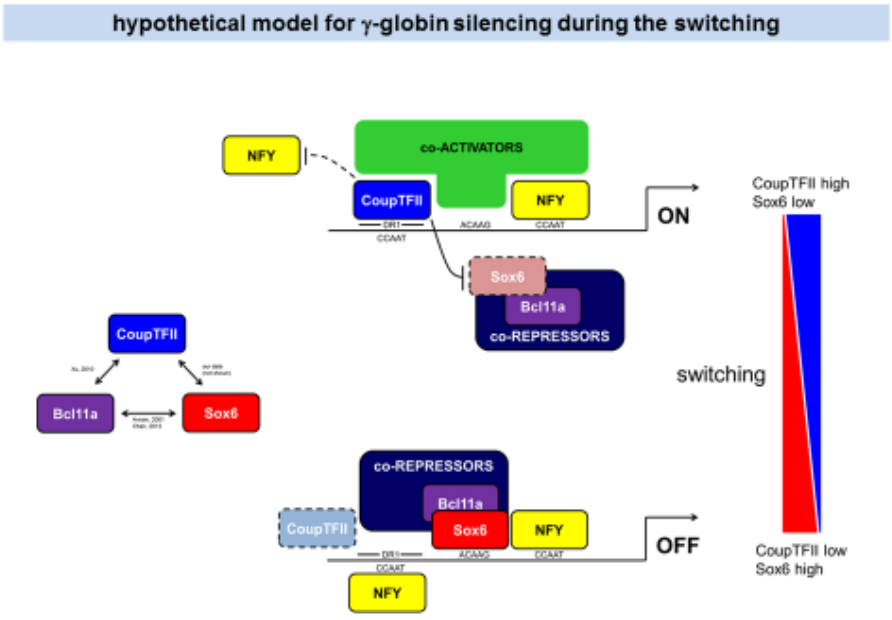
increase in all globins with a specific increase in the β-globin expression (Chapter 3.2, Figure 3.2.5a).

To confirm this unexpected result we moved to *ex vivo* erythroid cultures from $CD34^+$ cells purified from peripheral blood. These cells express γ-globin at low levels because they are almost completely switched to the adult pattern of globin expression. The lentiviral-mediated transduction of COUP-TFII of five independent cultures (from five different healthy donors) confirmed that COUP-TFII induces a sharp increase in γ-globin at low levels because they are almost completely switched to the adult pattern of globin expression (Chapter 3.2, Figure 3.2.8). In agreement with data from cell lines, these data confirm that COUP-TFII is potent γ-globin inducer and point to COUP-TFII as a prospective target whose modulation can increase γ-globin expression.

How Sox6 and COUP-TFII interact at the molecular level during the hemoglobin switching?

Given the above results, we propose here a molecular mechanism to explain how Sox6 and COUP-TFII interaction could facilitate the fetal to adult hemoglobin switching. During the fetal stage, when the level of COUP-TFII is high, it could contribute to stabilize the binding of NF-Y to the γ-globin promoter . In the absence of Sox6 or at equilibrium, the two CCAAT boxes, located at 22 bp of distance (two turns of DNA) would serve as a high affinity site for the binding of NF-Y and therefore for the maintenance of γ-globin expression. In this view, the co-presence of COUP-TFII and NF-Y

would form the activating complex required for γ-globin expression. When the level of Sox6 increases during the Switching, Sox6 would titer COUP-TFII, thus reducing the activation of the γ-globin gene. In the adult configuration, Sox6 could recruit other transcription factors, such as BCL11a-XL, to silence γ-globin, as proposed by Xu et al. ***(Xu et al., 2013; Xu et al., 2010)***. Moreover, increased levels of Sox6 could also directly favor the expression of β-globin, possibly by binding to the β-globin promoter itself. Of note, COUP-TFII is also known to interact with BCL11a (whose original name was CTIP= COUP-interacting-protein (***Avram et al., 2000***) and this further suggests that the balance between Sox6/COUP-TFII and BCL11a might be important for the regulation of γ-globin and possibly of other genes.



Cdk13 and Sox6

Between the CDKs identified by Mass spectrometry sequencing (CycK, CDK11, CDK12 and CDK13), the most interesting to study as putative Sox6 interactor was CDK13. The expression of this kinase is upregulated in several patients with ringed sideroblasts anemia associated with marked thrombocytosis (RARS-T) (***Malcovati L. et al. 2009***) and also it promotes megakaryocytic development in bone marrow cell cultures (***Lapidot-Lifson Y. et al. 1992***). Of interest, both Sox6 and CDK13 are localized in the nuclear speckles, together with splicing factors. As both, CDK13 and Sox6 are implicated in the regulation of alternative splicing our data support their possible cooperation in this process ***(Berro et al., 2008; Chen, Wong, Geneviere, & Fann, 2007; Even et al., 2006; Ohe, Lalli, & Sassone-Corsi, 2002)***. We thus validated the interaction between Sox6 and CDK13 in human erythroleukemic (HEL) by Co-IP (Chapter 3.3, Figure 3.3.2) and we characterized the expression profile of CDK13 in mouse $CD34^{+}/CD34^{-}$ cells purified from total Bone Marrow, confirming its enrichment in $CD34^{+}$ cells (Chapter 3.3, Figure 3.3.3b) the fraction containing early erythroid progenitors. Interestingly, downregulation of CDK13 in β-K562 cells forces late erythropoiesis, with a phenotype very similar to that induced by Sox6 overexpression in the same cells. In agreement with this observation, Sox6 is upregulated in CDK13 Knockdown cells, suggesting that the equilibrium between these two factors interacting at the protein level can influence the balance between proliferation and differentiation. Furthermore, the expression of megakaryocytic

markers in CDK13 knockdown cells is reduced, suggesting that the relative levels of Sox6 and CDK13 can also be important for cell fate decisions. To elucidate this point we are currently planning to modulate the expression of CDK13 (by lentiviral overexpression or shRNA downregulation) in a megakaryocytic cellular models.

High-content screening to quantify γ/β-globin expression in β-K562

The final output using this first-step HTS illustrates the heterogeneity of the cells upon modulators (siRNA transfection) or drugs treatments. The robustness of β-K562 cells allows transfection-based screening. What led to the identification of modulators as possible novel druggable target for γ-globin reactivation. On the basis of this multiplex high-content screening, we analyze several of the putative interactors of Sox6 to identify any candidate with a potential role in γ/β globin ratio regulation.

From more than 30 genes analyzed, only 7 were found with a potential role in hemoglobin switching. The readout from this screening proportionate an starting point to study new functional interactions between Sox6 and any of its putative interactors with a potential role in the regulation of the hemoglobin switching, and maybe as a potential candidates for the γ-globin reactivation.

CLINICAL RELEVANCE of the present work:

The molecular mechanism elucidating the regulation the fetal (γ) to adult (β) globin switching has been largely studied during the last decades. β-thalassemia and Sickle Cell Disease (SCD) are two of the most common inherited diseases caused by failure in the expression of the β-globin gene. Of interest, clinical evidences demonstrated that the maintenance of γ-globin expression during adult life (even at modest levels) is sufficient to ameliorate the clinical condition caused by β-thalassemia and SCD ***(Forget, 1998)***. Thus, many efforts have been made to reactivate γ-globin in β-thalassemic and SCD patients. Despite the intense efforts, only few drugs (with limited efficacy) have been identified so far. Among them, Hydroxyurea (HU) has been approved for the treatment of SCD and, recently, for β-thalassemia, but its efficacy varies among patients. Indeed, approximately half of the patients do not reach acceptable levels of HbF at HU doses of acceptable toxicity ***(Ronchi & Ottolenghi, 2013)***. Other drugs, such as butyrate and its derivates, 5-azacytabine, Decitabine and Tranylcypromine act on the epigenetic regulation of HbF, by inhibiting histone deacetylation or methylation on the γ-globin gene, but their efficacy is still very limited ***(Banan, 2013; Musallam, Rivella, Vichinsky, & Rachmilewitz, 2013; Perrine et al., 2010)***.

On this context, the High-content screening for γ/β-globin ratio in β-K562 cells described in chapter 4 represents a very promising first-step high-throughput screening (HTS) tool to identify

gene/drugs influencing the γ-/β-globin ratio.

On the other end, the understanding of the molecular mechanisms regulating the Hemoglobin switching is necessary to envisage new therapies having as a final goal γ-globin reactivation. The identification of Sox6 interactors and their functional validation, presented in this thesis, contribute to deepen the understanding of the TFs network governing the differential globins genes expression during development.

REFERENCES

Allen, H. F., Wade, P. A., & Kutateladze, T. G. (2013). The NuRD architecture. *Cell Mol Life Sci, 70*(19), 3513-3524. doi: 10.1007/s00018-012-1256-2

Banan, M. (2013). Hydroxyurea treatment in beta-thalassemia patients: to respond or not to respond? *Ann Hematol, 92*(3), 289-299. doi: 10.1007/s00277-012-1671-3

Berro, R., Pedati, C., Kehn-Hall, K., Wu, W., Klase, Z., Even, Y., . . . Kashanchi, F. (2008). CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. *J Virol, 82*(14), 7155-7166. doi: 10.1128/JVI.02543-07

Chen, H. H., Wong, Y. H., Geneviere, A. M., & Fann, M. J. (2007). CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochem Biophys Res Commun, 354*(3), 735-740. doi: 10.1016/j.bbrc.2007.01.049

Even, Y., Durieux, S., Escande, M. L., Lozano, J. C., Peaucellier, G., Weil, D., & Geneviere, A. M. (2006). CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *J Cell Biochem, 99*(3), 890-904. doi: 10.1002/jcb.20986

Filipe, A., Li, Q., Deveaux, S., Godin, I., Romeo, P. H., Stamatoyannopoulos, G., & Mignotte, V. (1999). Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *EMBO J, 18*(3), 687-697. doi: 10.1093/emboj/18.3.687

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

Ginder, G. D., Gnanapragasam, M. N., & Mian, O. Y. (2008). The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development. *Curr Top Dev Biol, 82*, 85-116. doi: 10.1016/S0070-2153(07)00004-X

Lauberth, S. M., & Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J Biol Chem, 281*(33), 23922-23931. doi: 10.1074/jbc.M513461200

Liberati, C., Cera, M. R., Secco, P., Santoro, C., Mantovani, R., Ottolenghi, S., & Ronchi, A. (2001). Cooperation and competition between the binding of COUP-TFII and NF-Y

on human epsilon- and gamma-globin gene promoters. *J Biol Chem, 276*(45), 41700-41709. doi: 10.1074/jbc.M102987200

Liberati, C., Ronchi, A., Lievens, P., Ottolenghi, S., & Mantovani, R. (1998). NF-Y organizes the gamma-globin CCAAT boxes region. *J Biol Chem, 273*(27), 16880-16889.

Lin, A. C., Roche, A. E., Wilk, J., & Svensson, E. C. (2004). The N termini of Friend of GATA (FOG) proteins define a novel transcriptional repression motif and a superfamily of transcriptional repressors. *J Biol Chem, 279*(53), 55017-55023. doi: 10.1074/jbc.M411240200

Musallam, K. M., Rivella, S., Vichinsky, E., & Rachmilewitz, E. A. (2013). Non-transfusion-dependent thalassemias. *Haematologica, 98*(6), 833-844. doi: 10.3324/haematol.2012.066845

Ohe, K., Lalli, E., & Sassone-Corsi, P. (2002). A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proc Natl Acad Sci U S A, 99*(3), 1146-1151. doi: 10.1073/pnas.022645899

Perrine, S. P., Castaneda, S. A., Chui, D. H., Faller, D. V., Berenson, R. J., Siritanaratku, N., & Fucharoen, S. (2010). Fetal globin gene inducers: novel agents and new potential. *Ann N Y Acad Sci, 1202*, 158-164. doi: 10.1111/j.1749-6632.2010.05593.x

Ronchi, A., Bellorini, M., Mongelli, N., & Mantovani, R. (1995). CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA. *Nucleic Acids Res, 23*(22), 4565-4572.

Ronchi, A., & Ottolenghi, S. (2013). To respond or not to respond to hydroxyurea in thalassemia: a matter of stress adaptation? *Haematologica, 98*(5), 657-659. doi: 10.3324/haematol.2013.084392

Sankaran, V. G., Xu, J., & Orkin, S. H. (2010). Transcriptional silencing of fetal hemoglobin by BCL11A. *Ann N Y Acad Sci, 1202*, 64-68. doi: 10.1111/j.1749-6632.2010.05574.x

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Exp Hematol, 33*(3), 259-271. doi: 10.1016/j.exphem.2004.11.007

Xu, J., Bauer, D. E., Kerenyi, M. A., Vo, T. D., Hou, S., Hsu, Y. J., . . . Orkin, S. H. (2013). Corepressor-dependent silencing of

fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A, 110*(16), 6518-6523. doi: 10.1073/pnas.1303976110

Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W., & Orkin, S. H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev, 24*(8), 783-798. doi: 10.1101/gad.1897310

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P. D., Fuchs, D. A., Erickson, D. T., . . . Brilliant, M. H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet, 2*(2), e14. doi: 10.1371/journal.pgen.0020014

ABBREVIATIONS:

ACH	-	Active chromatin hub
AGM	-	Aorta-Gonad Mesonephros
ALAS2	-	δ-aminolevulinate synthase 2
APC	-	Allophycocyanin
BCL11A	-	B-cell Lymphoma/Leukemia 11A
BFU-E	-	Blast Forming Unit – Erythroid
BM	-	Bone Marrow
CD71	-	Transferrin receptor
CDKs	-	Cyclin dependent kinases
CFU-E	-	Colony Forming Unit – Erythroid
CFU-S	-	Colony Forming Unit –Spleen
CLP	-	Common Lymphoid Progenitor
CML	-	Chronic Myelogenous Leukemia
CMP	-	Common Myeloid Progenitor
COUP-TFII	-	Chicken Ovalbumin Upstream Promoter-Transcription Factor II
CRISPR	-	clustered regularly interspaced short palindromic repeats
EKLF	-	Erythroid Kruppel-like factor
ENCODE	-	Encyclopedia of DNA Elements
EPO	-	Erythropoietin
FACS	-	Fluorescence-activated cell sorter
FECH	-	Ferrochelatase
FITC	-	Fluorescein Isothiocyanate
FL	-	Fetal Liver
Fli-1	-	Friend of leukemia virus integration 1
GFP	-	Green Fluorescence Protein
GM-CSF	-	Granulocyte Macrophage –Colony stimulating Factor
GpA	-	Glycophorin A
GpIIB/GpIIIa	-	Integrin platelet glycoprotein complex
HbF/A	-	Fetal/Adult hemoglobin

HCS - High Content Screening

HDAC - Histone DeAcetylases

HEL - Human erythroleukemic cells

HMG - High Mobility Group

HPFH - Hereditary Persistence of Fetal Hemoglobin

HS - Hypersensitive site

HSC - Hematopoietic Stem Cell

HU - Hydroxyurea

IL-3/11 - Interleukin-3/11

IRES - Internal Ribosome Entry Site

LC-MS/MS - Liquid chromatography tandem mass spectrometry

LCR - Locus Control Region

LT/ST -HSC - Long-term/ short-term repopulating hematopoietic stem cell

MBP - Myeloid B-cell Precursor

MEL - Mouse erythroleukemic cells

MEP - Megakaryocyte Erythroid Progenitor

MLP - Myeloid Lymphoid progenitor

MOI - Multiplicity Of Infection

MPL - Multilineage precursors

MPP - Multipotent progenitor cell

MRE - Major Regulatory element

MS - Mass spectrometry

MTA1 - Metastasis Associated 1

MTP - Myeloid T-cell Precursor

NGFR - Nerve Growth Factor Receptor

NuRD - Nucleosome Remodeling Deacetylases

PB - Peripheral Blood

PE - Phycoerythrin

RARS-T - Ringed sideroblasts anemia associated with marked thrombocytosis

RBC - Red Blood Cell

SCD - Sickle cell disease

SCF - Stem cell factor

SFFV - Spleen Focus Forming Virus

shRNA - Short hairpin RNA

SOCS3 - Suppressor Of Cytokine Signaling 3

TALEN - Transcription Activator Like Effector Nuclease

TFs - Transcription Factors

TPO - Thrombopoietin

WPRE - Woodchuck- hepatitis-virus Posttranscriptional Regulatory Element

ZFN - Zing Finger Nucleases

KNOWLEDGEMENTS:

I am so grateful to have been in Italy for these three years doing my PhD in Università degli studi Milano Bicocca, not only because of all the knowledge that I have got and the opportunity to develop my career, also because all the people that I have met here, people who helped me in the day-to-day life, making me feeling like at home, even being more than one thousand km away.

Firstly, I have to express my most sincere gratitude to my tutor Prof. Antonella Ronchi for giving the opportunity to do the PhD in her lab, and to participate in the amazing Marie Curie European program and in the HemID consortium. I am also really thankful for her support and patience during my PhD, trying always to help and advice me for the best. Also, I am so glad to have such a nice colleagues, Gloria, Sudhi, Serena, Cristina, Robert and Giovanni, who have been very good friends during my period in the lab, and we shared nice experiences together. Besides the people from our lab, I have to also express my gratitude to the people from the labs 3018, 3048 and 3040 and all my colleagues from the HemID consortium. I have also to thank Shruti Sinha to the help me with her knowledge with informatics tools and Dr. Rebecca Favaro for being always really nice and helpful, and for all the moments shared in MagaFurla.

Several experiments performed in this thesis were done thanks to the collaboration with other laboratories. I express my thanks to

Dr. Fabio Gasparri and Marta Durlak from NMS (Milano) to help me performing the high-content screening. I also express my gratitude to Prof. Angela Bachi and Giuseppe Filosa to allow for and help performing the gel filtration analysis in her lab (IFOM-Milano) and to Prof. Paolo Moi and Dr. Maria Giuseppe Marini (UNICA, Cagliari) to perform the human *ex vivo* cultures from peripheral blood.

Besides all the people from my lab environment, I have to show my gratitude to all the friends that have been always around, I can not mention all of them, but they are: the cantabrian-lover Victor Fifi, "*La Colla Pessigolla"* with special attention to *"les ch/xemes"*, my lovely friends from the Universitat Autònoma de Barcelona, "Ohana Milano", my lovely dutch friend Ijia, Alessio Giavazzi and, Marina and Borja.

Last but not the least, thanks to family, Monclús-Font and Font-San Jose (parents, grandparents, siblings, aunts, uncles and cousins). I owe them everything. I express my sincere gratitude, because even from the distance, I have never felt alone or being away to home, they have been always present. In particular, I have to thank the new family member, el petit Pol Guillaumet Font, to make our life happier since the past year. I love you all so much.